3 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 computer sciences

Subject

Pedestrian Trajectory Prediction Using Deep Learning and Rule Based Models

---

BY
Beggaa Siham
&
Benboubkeur Houria

In front of the jury members of

Dr. First name Lastname	President	University of Ouargla
Dr. First name Lastname	Examiner	University of Ouargla
Dr. Messaoud Mezati	Supervisor	University of Ouargla

Academic year: 2024/2025

إهداء

أولًا، نحمدُ اللهَ حمدًا كثيرًا طيبًا مباركًا فيه، فلولاه ما بلغنا هذه المرحلة، ولا تحقق هذا الإنجاز. هو الميسّر لكل خير، والمصدر الأول للقوة والثبات.

ثم، نوجّه أسمى عبارات الشكر والامتنان إلى والدينا الأعزّاء، الذين كانوا لنا عونًا وسندًا في كل خطوة، بدعائهم، وتضحياتهم، ودعمهم اللامحدود.

كما نعبّر عن امتناننا العميق لإخوتنا وأخواتنا، رفاق دربنا، وسندنا في كل الأوقات... محبتكم ودعمكم منارة في طريقنا.

ولا ننسى أصدقاءنا الذين شاركونا هذا المشوار، في لحظاته الحلوة والصعبة، وتركوا أثرًا جميلًا لا يُنسى وسام ،عزيزة ،هبة، جهينة ، ملاك،شهد، منار،ريم، منال أنتم جزء من هذا الإنجاز.

ونخص بالشكر والتقدير أستاذنا الفاضل "مزاتي مسعود"، على دعمه، وصبره، وإيمانه بنا، وتوجيهاته التي كانت نبراسًا في رحلتنا.

وأخيرًا... شكرًا لأنفسنا، لأننا اجتهدنا، صبرنا، وتمسّكنا بأحلامنا رغم كل التحديات.

نهدي هذا العمل إلى كل من كان جزءًا من هذه الرحلة... فهو ثمرة قلوب اجتهدت، وآمال كبرت، ومحبة صادقة جمعتنا.

Abstract

Pedestrian trajectory prediction is vital for intelligent transportation and human-aware autonomous systems, as it ensures safety and supports context-aware navigation. While deep learning models such as LSTM, GRU, and BiLSTM with attention mechanisms effectively model temporal dependencies, they often lack interpretability and require extensive annotated data. In contrast, rule-based models offer semantic clarity but struggle with flexibility in complex scenarios.

This work presents a hybrid framework that integrates interpretable behavioral features extracted via a rule-based module with deep learning models. Using the JAAD dataset, we derived indicators like motion states and temporal transitions to support context-aware trajectory prediction. The proposed framework was evaluated by comparing three deep learning models (LSTM, GRU, BiLSTM) and their hybrid counterparts that incorporate rule-based components.

Results on clean data demonstrate that the hybrid RBM-BiLSTM model achieves lower prediction errors (ADE = 24.36, CMSE = 437.74) compared to the pure BiLSTM model (ADE = 28.94, CMSE = 642.84), highlighting the benefit of integrating semantic cues for enhanced prediction accuracy.

Keywords: Pedestrian Trajectory Prediction, Deep Learning, Rule-Based Modeling, JAAD Dataset, Behavioral Cues, Long Short Term Memory (LSTM).

الملخص

يُعدّ التنبؤ بمسار المشاة أمرًا بالغ الأهمية لأنظمة النقل الذكية والأنظمة الذاتية الإدراك بالبشر، حيث يضمن السلامة ويدعم التنقل المُدرك للسياق. وعلى الرغم من أن نماذج التعلم العميق مثل LSTM وGRU وBiLSTM المزودة بآليات الانتباه تُظهر كفاءة في نمذجة الاعتماد الزمني، إلا أنها غالبًا ما تفتقر إلى القابلية للتفسير وتتطلب بيانات موسومة بكثافة.

في المقابل، توفّر النماذج القائمة على القواعد وضوحًا دلاليًا لكنها تواجه صعوبات في التكيّف مع السيناريوهات المعقدة.

يقدّم هذا العمل إطارًا هجينًا يدمج بين الميزات السلوكية القابلة للتفسير، والمستخرجة من خلال وحدة قائمة على القواعد، مع نماذج التعلم العميق. وبالاعتماد على مجموعة بيانات JAAD، قمنا باشتقاق مؤشرات مثل حالات الحركة والانتقالات الزمنية لدعم التنبؤ بالسياق في المسارات.

تم تقييم الإطار المقترح من خلال مقارنة ثلاثة نماذج تعلم عميق (LSTM ،GRU ،BiLSTM) ونظيراتها الهجينة التي تدمج مكونات قائمة على القواعد. أظهرت النتائج على بيانات نظيفة أن النموذج الهجين RBM-BiLSTM حقق أخطاء

، (ADE = 28.94 فقط BiLSTM مقارنة بالنموذج القائم على (CMSE = 437.74، ADE = 24.36) تنبؤ أقل
، مما يبرز فائدة دمج المؤشرات الدلالية في تحسين دقة التنبؤ(CMSE = 642.84.

الكلمات المفتاحية:

الإشارات السلوكية، الذاكرة طويلة وقصيرة المدى, توقّع مسارات المشاة، التعلم العميق، النمذجة القائمة على القواعد، مجموعة بيانات (JAAD).

Résumé

La prédiction des trajectoires des piétons est essentielle pour les systèmes de transport intelligents et les systèmes autonomes sensibles au contexte humain, car elle garantit la sécurité et favorise une navigation contextuelle.

Bien que les modèles d'apprentissage profond tels que LSTM, GRU et BiLSTM dotés de mécanismes d'attention soient efficaces pour modéliser les dépendances temporelles, ils manquent souvent d’interprétabilité et nécessitent de grandes quantités de données annotées. En revanche, les modèles basés sur des règles offrent une clarté sémantique, mais rencontrent des difficultés à s’adapter aux scénarios complexes.
Ce travail propose un cadre hybride qui intègre des caractéristiques comportementales interprétables, extraites via un module basé sur des règles, avec des modèles d'apprentissage profond. À l’aide du jeu de données JAAD, nous avons dérivé des indicateurs tels que les états de mouvement et les transitions temporelles pour soutenir une prédiction de trajectoire sensible au contexte.

Le cadre proposé a été évalué en comparant trois modèles d’apprentissage profond (LSTM, GRU, BiLSTM) et leurs homologues hybrides intégrant des composants basés sur des règles. Les résultats obtenus sur des données propres montrent que le modèle hybride RBM-BiLSTM obtient des erreurs de prédiction plus faibles (ADE = 24.36, CMSE = 437.74) par rapport au modèle BILSTM pur (ADE = 28.94, CMSE = 642.84), ce qui met en évidence l'intérêt d'intégrer des indices sémantiques pour améliorer la précision des prédictions.

Mots-clés : Prédiction de trajectoires piétonnes, Apprentissage profond, Modélisation à base de règles, Jeu de données JAAD, Indices comportementaux, Mémoire à court et long terme (LSTM)

Table of Contents

List of Figures:

List of Tables:

List of Equations:

CHAPTER I: GENERAL INTRODUCTION

1.1 Motivation:

Pedestrian trajectory prediction is an increasingly vital area of research, particularly in the context of intelligent transportation systems and autonomous driving. As urban environments grow more dynamic and interconnected, the ability to anticipate human movement in real-time is critical for ensuring safety, improving traffic flow, and supporting human-centric decision-making systems. The integration of predictive models into urban mobility infrastructure and autonomous systems promises to reduce risks and enhance the responsiveness of artificial agents in complex environments.

Traditional deep learning models, such as Long Short-Term Memory (LSTM)-based networks, Bidirectional LSTM (BiLSTM), and Gated Recurrent Unit (GRU) architectures, as well as attention-driven mechanisms, have demonstrated strong capabilities in capturing temporal and social dependencies in human trajectory. Recent advancements even include the use of graph-based and Transformer-based methods. Despite their success, these models typically require large-scale annotated datasets, struggle to generalize across diverse contexts, and offer limited interpretability an issue particularly problematic in safety-critical applications.

On the other hand, knowledge-based approaches attempt to encode prior knowledge or behavioral understanding directly into the prediction process. Among these, rule-based systems offer high interpretability by explicitly modeling behavioral constraints such as waiting, stopping, or transitioning actions. They allow for trustworthy predictions in well-understood scenarios but are often too rigid to handle the variability and unpredictability inherent in real-world pedestrian behavior.

To address these limitations, this thesis explores a hybrid approach that combines the learning power of deep models with the transparency and domain knowledge provided by rule-based reasoning. Such integration aims to produce a robust and context-aware model that can operate efficiently in dynamic environments and adapt to diverse pedestrian behaviors.

1.2 Thesis Objectives:

The main goal of this thesis is to develop a hybrid pedestrian trajectory prediction model that leverages both deep learning and rule-based reasoning to improve prediction accuracy, interpretability, and real-time applicability. The specific objectives include:

- **Hybrid Architecture Design:** Develop a hierarchical attention-based model that integrates LSTM encoding with handcrafted behavioral rule features for improved interpretability. In addition to LSTM, the architecture incorporates other recurrent units such as BiLSTM and GRU to capture diverse temporal dynamics and enhance the model's flexibility in learning complex pedestrian behaviors.
- **Robustness and Stabilization:** Implement velocity-aware smoothing techniques and enforce physics-based constraints to enhance prediction stability and realism.
- **Contextual Awareness:** Encode social interactions and environmental cues into the model via relative positioning, speed profiling, and behavioral state transitions.
- **Comprehensive Evaluation:** Benchmark the model against state-of-the-art approaches using standardized dataset (JAAD) and a multi-metric evaluation protocol covering spatial, temporal, and physical plausibility.

1.3 Major Contributions:

This thesis presents several novel contributions to the field of pedestrian trajectory prediction:

- A New Hybrid Framework: Proposes a supervised hybrid model that integrates handcrafted behavioral indicators with data-driven learning for improved accuracy and transparency.
- Interpretable Attention Mechanism: Introduces a dynamic weighting scheme that balances learned and rule-based features based on trajectory context and prediction difficulty.
- Evaluation Metrics: Establishes a composite benchmarking protocol combining spatial accuracy (ADE/FDE/MSE/CMSE/CFMSE), temporal smoothness, and physical realism (jerk analysis), standardized across datasets.
- Open-Source Implementation: Releases a modular and extensible framework including data processing, model configuration, and trajectory visualization tools to support reproducibility and further research.

1.4 Thesis Outline:

This thesis is organized as follows:

Chapter II provides an introductory overview of pedestrian trajectory prediction, distinguishing it from related tasks like motion forecasting and intention inference. It highlights its importance in fields such as autonomous driving, urban planning, robotics, and surveillance, emphasizing the value of anticipating human movement in real-world settings.

Chapter III reviews key modeling approaches, categorizing them into physics-based, pattern-based, and planning-based methods, and discussing their strengths and limitations. It also focuses on hybrid models that blend rule-based logic with deep learning for improved interpretability and adaptability. The chapter ends with an overview of major benchmark datasets essential for developing socially-aware models.

Chapter IV addresses the core problem of pedestrian trajectory prediction, highlighting the inherent challenges of forecasting complex and dynamic human motion influenced by personal intent, environmental context, and social interactions. It presents a formal definition of the task by outlining the input-output structure, relevant contextual variables, and key constraints such as spatial limitations and social norms. This structured formulation establishes a consistent foundation for developing and evaluating predictive models, emphasizing the need to incorporate behavioral indicators and interpretability to ensure reliable performance in real-world scenarios.

Chapter V introduces the JAAD dataset, detailing its structure, annotations, and behavioral data relevant to pedestrian trajectory prediction in urban settings. It emphasizes the dataset's strengths such as rich behavioral labels and contextual metadata which are key for modeling pedestrian behavior. The chapter also compares JAAD to other datasets, outlining its advantages and limitations, and underscores its role in supporting research on socially-aware autonomous driving.

Chapter VI presents a hybrid pedestrian trajectory prediction model that combines rule-based reasoning with deep learning. Behavioral features extracted from the JAAD dataset are used as inputs to a recurrent neural network incorporating LSTM, BiLSTM, and GRU units along with attention mechanisms. The model demonstrates superior performance compared to baseline approaches, showing improved prediction accuracy over both short and long time horizons.

Chapter VII concludes the thesis by summarizing the hybrid pedestrian trajectory prediction model, which combines rule-based and deep learning approaches. The model showed strong performance on both pre-processed and real-time video data. Future work will focus on improving generalization across datasets, exploring advanced models, and incorporating additional behavioral insights into predictions.

CHAPTER II: PRELIMINARIES

2.1 Introduction:

An in-depth understanding of pedestrian trajectory prediction requires a clear grasp of the key concepts and distinctions that frame this task. This chapter provides essential background that contextualizes the research. It begins by explaining the core idea of **trajectory prediction for pedestrians** and how it extends beyond related tasks such as **motion estimation**, intention recognition, and behavior modeling. The chapter then discusses a range of real-world applications where accurate trajectory forecasting plays a critical role, including **autonomous driving**, **urban design**, human-robot interaction, and **security surveillance**. By examining these domains, the chapter emphasizes the practical importance of anticipating pedestrian movement and the various environmental, social, and **semantic factors** influencing it. This foundational understanding supports the formal problem formulation and modeling approaches that follow in subsequent chapters.

2.2 Understanding the Concept of Trajectory Prediction:

Trajectory prediction can be defined as the method by which future locations of a moving object are estimated based on its previous motion patterns [1]. Applied to pedestrians, the process entails the prediction of the probable movements of an individual in the near future using a history of previous locations [1]. The future trajectory is usually represented as a continuum of spatial points along the time axis [1]. The same forecast is applicable to the direction a person is walking towards, but, on top of this, to potential alterations of their direction motivated by proximate others, obstacles, or environmental cues [2]. The nature of this activity is fundamentally probabilistic and dynamic, reflecting the non-deterministic nature of human action [1]. Forecasting pedestrian trajectories is therefore an extremely complicated and advanced discipline, in need of models that can manage uncertainty, variation, and context-sensitivity [3].

2.3 Distinction from Motion, Intention, and Behavior Prediction:

While prediction of pedestrian movement has been commonly paired with predictions of motion, intent, and activity [4] [5] [6], it is a distinct task with specific goals and methods. Motion prediction is primarily concerned with the estimation of future locations given present speed and direction and frequently makes use of physical or kinematic models[7] [8]. Trajectory prediction takes a longer timescale into account and introduces the effects of nearby agents along with the characteristics of the environment [9][10]. This distinction between motion, trajectory, intention, and behavior prediction is visually summarized in Figure 1. Moreover, intention prediction aims to infer the underlying goal or decision of an individual, such as whether a pedestrian intends to cross the road or stay still [11]. Behavior prediction goes further, attempting to understand complex actions that are often socially driven, such as group behavior, sudden stops, or deferential behavior[12][13]. These tasks rely heavily on semantic analysis, psychological models, or high-level understanding of the scene [14][15]. Trajectory prediction is midway along the spectrum of predictive modeling: more sophisticated than mere motion extrapolation, yet not necessarily involving a complete decoding of the individual's mental state or higher-level goals[16]. It rather makes a pragmatic estimation of probable future locations, which can be used to improve downstream systems such as decision-making algorithms for self-driving vehicles without necessarily underpinning any direct understanding of purpose or behavior [17].

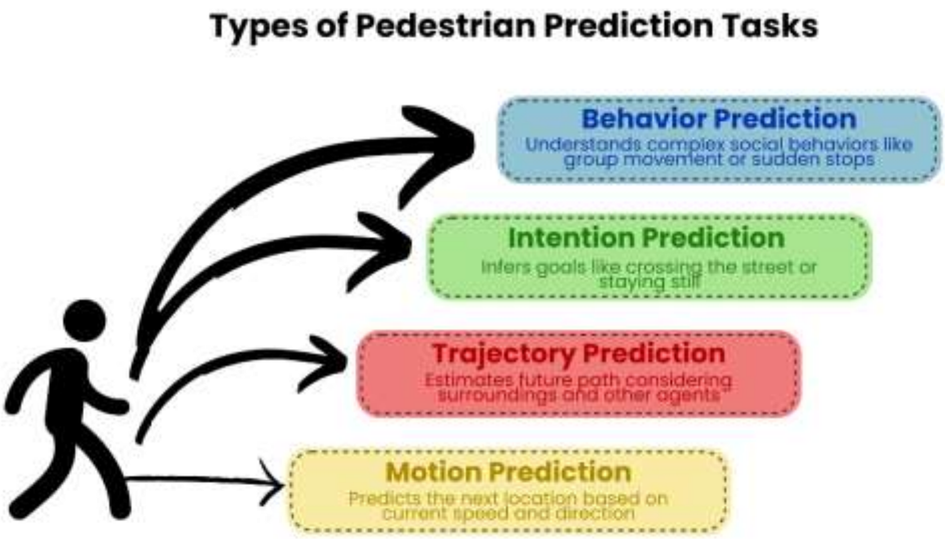

Figure 1: Type of Pedestrian Prediction Tasks

2.4 Applications of Pedestrian Trajectory Prediction:

Pedestrian trajectory prediction serves as a cornerstone for multiple real-world applications where anticipating human movement is crucial. Its integration across various domains enhances safety, efficiency, and overall system intelligence in dynamic environments [18][19][20].

2.4.1 Autonomous Driving:

Pedestrian trajectory prediction is a critical component of autonomous driving systems, enabling vehicles to interpret the intentions of nearby individuals and make safe, informed decisions[20]. For instance, when approaching an intersection, it is not enough for an autonomous vehicle to detect pedestrians; it must also anticipate whether they intend to cross or remain stationary [21][22]. This foresight supports crucial actions such as braking, stopping, or proceeding safely, thereby enhancing safety in dynamic urban environments[23].

The Figure 2 illustrates such a scenario, where the autonomous vehicle analyzes pedestrian behavior including trajectory, orientation, and gaze direction to evaluate the likelihood of road crossing [23][24]. These cues help the vehicle plan its motion and adapt its driving strategy accordingly [23] [24]. The red dashed lines represent predicted pedestrian paths, while the red paths from the vehicle indicate alternative navigation plans depending on its destination and the predicted behavior of surrounding agents. This predictive capability is essential for robust path planning and real-time decision-making in pedestrian-rich settings [25].

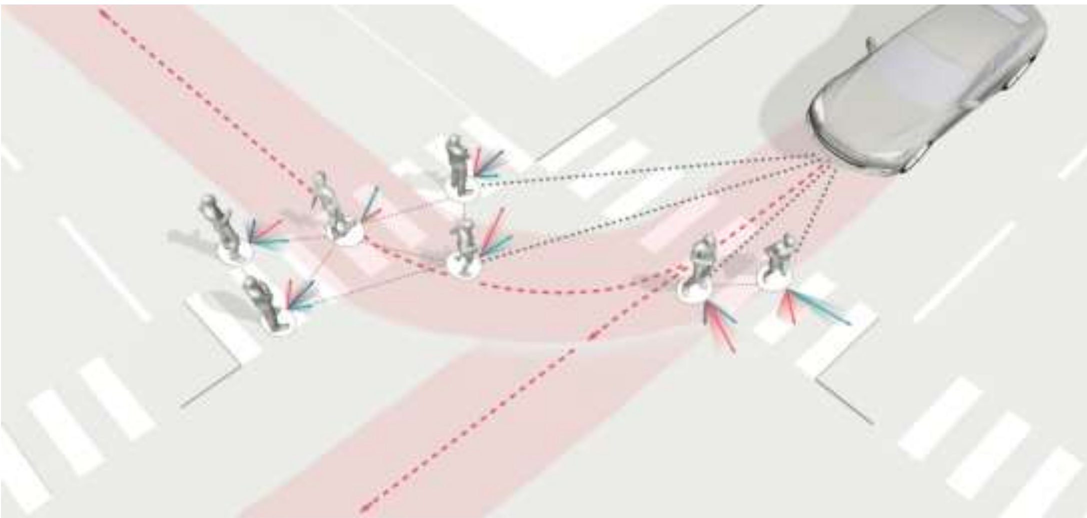

Figure 2: Predictive Motion Planning in Autonomous Driving at Pedestrian-Rich Intersections [26]

2.4.2 Urban Planning:

In planning and urban design, trajectory prediction provides valuable means to investigate pedestrian interaction and movement along infrastructure[27][28]. Engineers and planners can make predictions on the usage of sidewalks, crossings, and public space based on predictive models [29][30]. This helps to develop safer and more efficient environments. For instance, predictions enable the mapping of congestion points or unsafe intersections, allowing for improvements such as the widening of sidewalks or the installation of smart traffic lights[31].

As illustrated in Figure 3, sidewalks can be divided into functional zones. Trajectory prediction enables planners to assess how pedestrians utilize each zone, which supports the design of safer and more accessible public spaces[29].

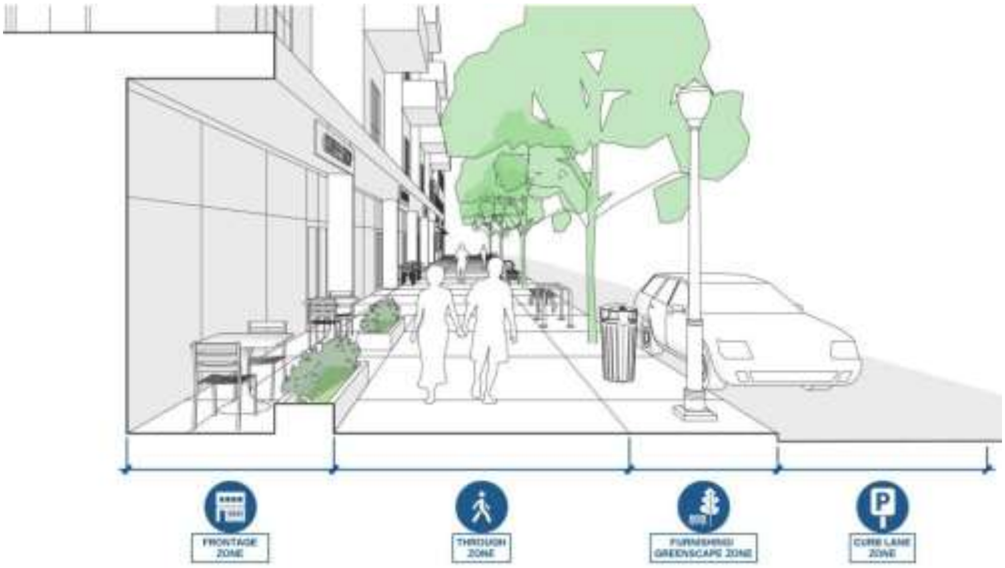

Figure 3: Functional sidewalk zoning for pedestrian space planning

[32]

2.4.3 Robotics and Human-Robot Interaction:

In robot systems, particularly those that function in human-shared environments such as shopping malls and airports, it is desirable that the robot be aware of human motion within its vicinity to move efficiently and safely [33]. Pedestrian trajectory prediction allows the robot to prevent collisions, choose the most effective routes, and communicate with individuals in a cohesive manner. This is increasingly required in dynamic environments where human activity is constantly changing [34].

As illustrated in Figure 4, robots can use predicted pedestrian trajectories to adjust their movement path either by avoiding a person immediately before contact or by maintaining a socially acceptable distance, enhancing both safety and social comfort[35].

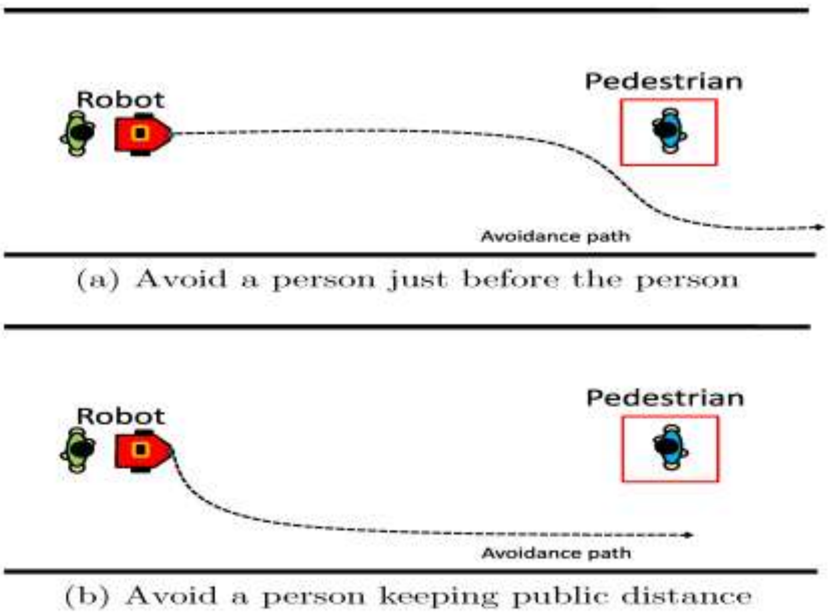

Figure 4: Comparison of pedestrian avoidance methods

[36]

2.4.4 Security Surveillance and Crowd Analysis:

In the safety and surveillance domain, trajectory prediction is utilized for analyzing the movement of crowds and picking up early signs of anomalous behavior[37]. For instance, if a person's movement in a public place greatly deviates from predicted trajectories, it might be indicating an emergency or suspicious activity [38]. These models are also used in crowd control for huge numbers of individuals visiting large events, forecasting bottlenecks and streamlining the movement of people more safely and efficiently. Intelligent surveillance systems can also utilize the predictions to enhance emergency response planning[38].

As illustrated in Figure 5, colored lines represent predicted trajectories of individuals within a crowd, allowing for real-time analysis of movement trends and detection of deviations from expected patterns.

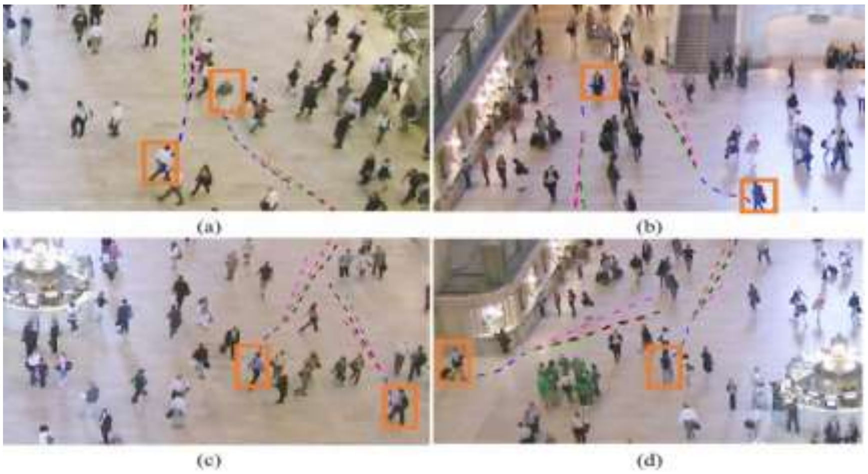

Figure 5: Prediction of pedestrian paths in a crowded environment.

[39]

2.5 Conclusion:

This chapter presented a concise overview of the field of pedestrian trajectory prediction. It began by explaining the concept itself, emphasizing its probabilistic and dynamic nature. It then clarified how trajectory prediction is distinct from motion, intention, and behavior prediction, by focusing on the prediction of future locations rather than underlying goals or mental states. The chapter also explored major real-world applications of the task, including autonomous driving, urban planning, robotics, and security surveillance. These foundations provide essential context for the following chapter, which delves into the related literature and methodologies that have shaped recent advancements in this domain

CHAPITRE III: RELATED WORKS

3.1 Introduction:

Accurate pedestrian trajectory prediction is a fundamental requirement in various intelligent systems, particularly those concerned with safety and human-aware navigation. This chapter presents a structured overview of the principal modeling approaches adopted in the literature, namely physics-based, pattern-based, and planning-based paradigms. Each category is introduced with its underlying principles, strengths, and inherent limitations, supported by visual summaries to facilitate conceptual understanding.

The discussion further extends to hybrid models that integrate rule-based reasoning with deep learning techniques, offering a balanced trade-off between interpretability and adaptability. These models have shown increasing potential in addressing complex urban scenarios where both domain knowledge and data-driven learning are essential.

The chapter concludes with a classification of widely used datasets, which serve as benchmarks for training and evaluating prediction models. Emphasis is placed on behaviorally annotated datasets that support socially-aware modeling, particularly in safety-critical contexts.

3.2 Trajectory Prediction:

Trajectory prediction aims to estimate future pedestrian positions based on their past movements,[40][19] and is commonly framed as a time series forecasting task.[41] With the rapid progress of deep learning, models such as RNNs, CNNs, GNNs, and Transformers have emerged as powerful tools in this domain,[42][43] surpassing traditional statistical methods like Kalman Filters and Hidden Markov Models[44], which struggle with complex behaviors in dynamic environments.[16][45]

Unlike knowledge-based models that rely on predefined rules and handcrafted features,[16][18]deep learning approaches learn motion patterns directly from data, enabling better adaptability and scalability[46]. This shift has made deep learning a dominant paradigm in real-world applications, from autonomous vehicles[47] to intelligent surveillance systems.{1.6}

Recent studies propose a wide range of model architectures and design objectives, focusing on aspects such as social interactions[48][49], motion intention, and environmental context.[20] These approaches highlight the increasing need for models that balance predictive accuracy with robustness[50] and real-time performance,[51] especially in safety-critical systems.[52]

In the following sections, these modeling paradigms are introduced in detail. The aim is to provide a structured overview of the fundamental approaches to pedestrian trajectory prediction and highlight their respective strengths and limitations, supported by Figure 6, which illustrates the three main model categories.

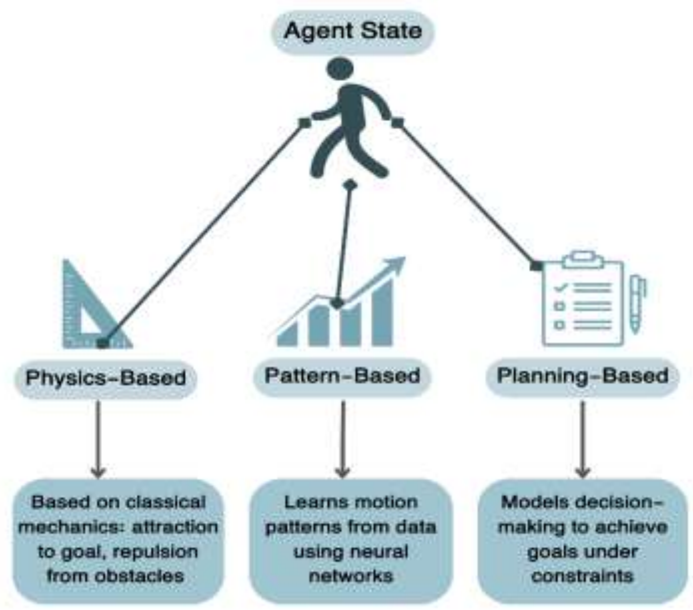

Figure 6: Trajectory Prediction Paradigms

3.2.1 Physics-Based Models:

These models replicate pedestrian movement by applying the laws of classical mechanics. An example of such a model is the Social Force Model,[12] in which pedestrians experience attractive forces towards their destination and repulsive forces from others and obstacles.[53] These models are simple and interpretable, making them preferable to apply in controlled systems such as robot simulations and emergency evacuation plans.[54] [55] They fall short in reproducing sophisticated social behaviors and need a lot of manual parameter adjustment.[56]

3.2.2 Pattern-Based Models:

Pattern-based models rely on learning motion patterns from historical data directly without any physical assumptions. [57] Empowered by deep learning architectures like RNNs, CNNs, [58] and GNNs, these models excel at capturing complex temporal and spatial dependencies. [43] Their flexibility and scalability have made them the backbone of modern trajectory prediction systems.[59][60] Nevertheless, they require large annotated datasets, are prone to overfitting, and may generalize poorly to new environments.[61]

3.2.3 Planning-Based Models:

The models above encode pedestrian movement as a decision-making process for the achievement of some goal.[62] Through methods such as Inverse Reinforcement Learning (IRL),[63] planning-based models infer the latent reward functions underlying pedestrian movement.[62] The models are especially effective at capturing long-term objectives, contextual constraints, [64] and strategic interactions,[48] and hence are most suitable for use in autonomous driving and social robotics environments.[62]

3.3 A Systematic Taxonomy of Deep Learning Approaches to Pedestrian Trajectory Prediction:

This section presents a systematic taxonomy of deep learning models applied to pedestrian trajectory prediction. The taxonomy is developed based on a comprehensive analysis of state-of-the-art contributions in the literature, and it aims to organize the diverse modeling approaches according to well-defined technical criteria. The goal is to highlight how different model types leverage large-scale datasets, architectural designs, and advanced AI techniques to capture human movement dynamics in varied environments.

a) 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).

b) Type of Algorithms: Focuses on key deep learning algorithms, including:

- LSTM (Long Short-Term Memory): Captures temporal dependencies in pedestrian trajectories.
- CNN (Convolutional Neural Networks): Extracts spatial features from scene data.
- GAN (Generative Adversarial Networks): Generates realistic pedestrian trajectories by modeling complex distributions.
- Other: Includes Graph Neural Networks (GNNs) and Reinforcement Learning-based models.

c) Network structures: Examines how the deep learning models are structured

- Sequential: Models like LSTM and RNN that process time-series pedestrian data sequentially.
- Non-Sequential: Includes CNNs, GANs, and Transformer networks that handle non-linear and parallel data processing.

d) Prediction tasks: Pedestrian trajectory prediction models serve multiple purposes, including:

- Interaction Modeling: Predicts pedestrian interactions with other pedestrians and their environment.
- Trajectory Planning: Estimates future movement paths, ensuring accurate navigation predictions.
- Intention Prediction: Infers pedestrian goals and intended destinations for proactive decision-making.

e) Use of advanced AI techniques: Focuses on the integration of advanced AI methods:

- Transformer Networks: Utilize attention mechanisms for modeling long-range dependencies in pedestrian trajectories.
- Variational Autoencoder (VAE): Learns probabilistic representations for diverse trajectory predictions.

Other: Includes Reinforcement Learning (RL), Inverse Reinforcement Learning (IRL), and Self-Supervised Learning, enhancing adaptability and decision-making capabilitie

Article	**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-[65]	x				x		PIS	EMP	APC			Attention Mechanism
Amir Rasouli-2020-[66]	x	x			x	x	PIS	IPV	CFT			Attention Mechanism
Jie Yang -2020-[67]		x				x	APC	EMF	ECM			Multi-Scale Detection
Joseph Gesnouin-2020-[68]		x		Auto-encoder		x	PIS	EMP	APC			Autoencoder-Based Feature Learning
J. Lorenzo-2020-[69]	x	x			x		APB	PP	EPU			Autoencoder-Based Feature Learning
Jun Hayakawa, Behzad Dariush-2020-[70]		x		TRN			CFT	PP	EMP			Temporal Relation Network
Dan Xiong-2021-[71]	x	x	x		x		AHI	PMP	**PAC**	x		
Arash Kalatian-2021-[72]	x				x		APB	IPV	IAV			
J. Lorenzo-2021-[73]		x		Transformer	x	x	APB	PMP	**EMP**	x		
Chi Zhang-2021-[74]		x		TCN, GCN		x	APB	EPU	**APC**			GNN

Raphael Rozenberg-2021-[75]	x			Bi-RNN	x		EMSI	PP	EMP			Asymmetrical Bi-RNN
Lina Achaji-2021-[76]				Transformer Networks		x	APB	PP	EPU	x		
Joseph Gesnouin-2021-[77]		x		GRU, Atrous Convolutions	x	x	SFT	PMP	EPU			Temporal, Self-Attention
Chi Zhang -2022-[78]	x	x	x		x	x	PMP	EMP	EMSI	x		
Riddhi Joshi -2022-[79]	x	x			x	x	EMF	PP	EMF		x	
Eric Liang -2022-[80]		x		yolo		x	CFT	PP	APB	x		
Amir Rasouli-2022-[81]	x	x		SAIM	x	x	PIS	PP	IAV	x		
Dongxu Guo -2022-[82]	x	x			x	x	PIS	PMP	IAV	x		
Jibran Ali Abbasi,2023[83]		x			x		PP	EPU	APB			X(Attention Mechanism)
Zhengming Zhang,2023[84]				Transformer		x	EMSI	PP	EBT	x		
Esteban Moreno-2023[85]				FNN	x		PP	PP	IPV			
Jia Huang-2023-[86]	x			Transformer	x		PP	PMP	IAV	x		
Chen- 2024[87]				GCN	x		PIS	EMP				Attention Mechanism

Li -2024[88]				GCN	x		PIS					Heterogeneous Graph Learning
Liang- 2024[89]				TCN	x			EMP				Visual Information Integration
Ling- 2024 [90]				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.

3.4 Hybrid Model:

Hybrid models in pedestrian trajectory prediction aim to merge the adaptability of deep learning with the structured reasoning of rule-based systems**.**[16] This combination leverages the strengths of both paradigms deep models excel in learning from large datasets and capturing nonlinear patterns, while rule-based components offer interpretability**,**[40] safety enforcement, and domain knowledge integration**.**[91]

Recent research proposes hybrid architectures that vary in structure and purpose.[92] Some approaches use rules during data preprocessing, enriching the input before feeding it into neural networks.[93] Others implement constraint-checking mechanisms that validate or adjust predictions to ensure they comply with physical or social norms.[94] In more complex settings, parallel modules process information independently where neural networks[40][95] handle pattern learning and symbolic logic modules apply reasoning or prioritization.[16]

This synergy enables robust performance in uncertain or dynamic environments, improves explainability critical in safety-sensitive domains like autonomous driving and ensures context-aware behavior even with limited data.[93][64] Hybrid models are increasingly considered a viable direction for developing reliable, scalable, and transparent prediction systems.[96]

3.5 Datasets:

To train and evaluate pedestrian trajectory prediction models, several datasets have been collected, capturing pedestrian movements and behavioral cues in diverse real-world environments. Depending on their focus and type of annotations, these datasets can generally be grouped into three main categories:

3.5.1 Pedestrian Trajectory Datasets:

Trajectory datasets provide time-stamped positional data of pedestrians moving through urban or semi-urban environments. The most commonly used datasets in the literature are ETH and UCY,[97] both of which contain video sequences and annotated trajectories of pedestrians recorded from a top-down view in public spaces such as sidewalks and plazas.[98] Despite their relatively small scale and controlled nature,[99] they remain benchmark datasets widely adopted in trajectory prediction research[98] due to their high-quality annotations and ease of use for model comparison.[1]

3.5.2 Intention-Based Datasets:

Datasets like PIE (Pedestrian Intention Estimation)[100] and JAAD (Joint Attention in Autonomous Driving)[101] go beyond tracking positions to include annotations for pedestrian intentions, behaviors, and visual cues such as gaze direction and hand gestures.[66][102] These datasets are useful for developing models that incorporate semantic understanding and early intention recognition, allowing for more proactive and human-aware predictions. [66] They often support both trajectory forecasting and behavior classification tasks, enriching model capabilities.[101]

3.5.3 Behaviorally Annotated Datasets:

This category includes datasets focused specifically on behavioral context and visual signals relevant to decision-making. **JAAD**, in particular, is notable not only for intention cues[101] but also for its rich contextual information about the scene (e.g., presence of crosswalks, traffic signs, and other contextual road elements).[102][101] Such datasets are especially valuable when developing socially-aware prediction systems or hybrid models that combine visual perception with reasoning over behavioral patterns.
Such datasets enable learning from rich urban scenes, as shown in **Figure 7**, where pedestrians are annotated with bounding boxes and behavior labels while crossing at a designated crosswalk.

Figure 7**:** Pedestrian detection using the pedestrian tag for all pedestrians

[103]

3.6 Conclusion:

This chapter provided a comprehensive survey of the main modeling paradigms used for pedestrian trajectory prediction, highlighting the diversity of approaches from physics-based formulations to data-driven and planning-based models. Special attention was given to hybrid models, which attempt to reconcile the strengths of rule-based reasoning with the adaptability of deep learning. Furthermore, relevant datasets were introduced, particularly those offering rich behavioral annotations, as they represent a key component in training and validating socially-aware systems.

Despite the progress made, several challenges persist, such as handling uncertainty, ensuring generalization across environments, and maintaining interpretability in complex settings. These limitations raise fundamental questions about how to design predictive systems that are both accurate and trustworthy. The next chapter introduces the problem statement that drives this research, outlining the central issues and motivations that guide the proposed approach.

CHAPITRE IV: PROBLEM FORMULATION

4.1 Introduction:

Pedestrian trajectory prediction is a core task in human-interactive intelligent systems, such as autonomous vehicles and service robots. Pedestrian motion is highly complex due to personal intentions, environmental conditions, and social interactions. It is therefore essential to formulate the problem properly to model and improve computational prediction systems. This chapter provides a formalization of the problem, specifying the inputs, outputs, variables, and constraints that define the parameters of pedestrian trajectory forecasting, thereby establishing the basis on which to build and compare models.

4.2 Problem Statement:

Pedestrian trajectory prediction is talking about the problem of estimating future positions of a pedestrian over a specified time horizon, based on observed motion history as reference [99][104]. This problem is inherently difficult because of the dynamic, nonlinear, and context-dependent characteristics of human movement. In contrast with motions of rigid bodies, pedestrian trajectories are affected by intentions of a person, social interactions, and environmental constraints [99].

In practical applications, correct path prediction is key to providing safety and enhancing efficiency of autonomous systems. For example, autonomous vehicles must predict pedestrian actions in order to prevent accidents and create safe navigation strategies [66]. Likewise, in similar environments of human-robot interaction, robots must navigate shared spaces while responding appropriately to humans presence and movement. Therefore, there is an urgent need to model pedestrians behavior correctly to allow intelligent systems to move safely in human-oriented environments [99][34].

Figure 8 illustrates a comparative visualization of observed trajectories (blue), predicted paths (orange), and ground truth trajectories (green) of multiple pedestrians in a real-world setting, highlighting the challenges in achieving precise prediction accuracy.

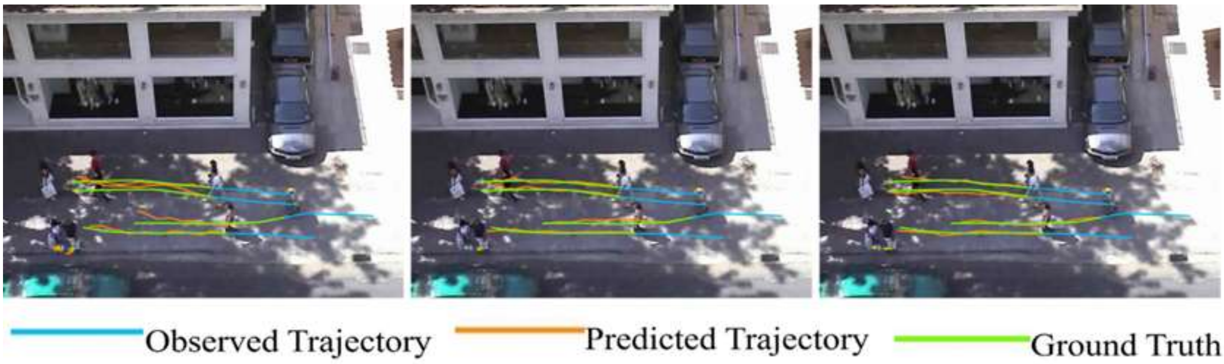

Figure 8: Comparison of Observed, Predicted, and Ground Truth Pedestrian Trajectories in a Real-World Urban Environment

[105]

4.3 Formal Problem Formulation:

The pedestrian trajectory prediction problem involves forecasting future locations of a pedestrian based on their past observed positions and the surrounding context [66][106].
Inputs are sequences of past positions over time, typically in the form of 2D coordinates[98]:

$$\{(x_{t-k}, y_{t-k}), \dots, (x_t, y_t)\}.$$

Equation 1: sequences of past positions

Outputs are the predicted future positions for the next T time steps[98]:

$$\{(x_{t+1}, y_{t+1}), \dots, (x_{t+t}, y_{t+t})\}.$$

Equation 2: predicted future positions

Contextual factors may include other pedestrians positions and movements, the semantic map of the scene (e.g., sidewalks, obstacles), and social interactions such as group movement or collision avoidance[98][107].

The goal is to learn a function that maps historical motion patterns to plausible future trajectories, taking into account environmental and social cues[98][107][94]:
Given:

$$\{(x_{t-k}, y_{t-k}), \dots, (x_t, y_t)\}.$$

Predict:

$$\{(x_{t+1}, y_{t+1}), \dots, (x_{t+t}, y_{t+t})\}.$$

This formalization enables systematic modeling, training, and evaluation of trajectory prediction methods.

4.4 Constraints and Challenges in Pedestrian Trajectory Prediction:

Predicting pedestrian trajectories is a complex task influenced by various real-world limitations and challenges [99][94]. One of the key limitations involves the constraints imposed by infrastructure and environment [99][107]. Pedestrians move in designated zones like sidewalks, crosswalks, and specified paths, and thus models need to account for these spatial limitations [99][107][94]. Further, any effective prediction system must also replicate collision avoidance between individuals or with vehicles, in addition to adhering to implicit social norms like maintaining distance between individuals or moving behind a group [99][98].

In addition, pedestrian movement is non-determined; it is inherently multimodal and unpredictable. A person can suddenly change direction or be influenced by external and often subtle factors, such as environmental factors or visual cues. As a result, this complexity challenges the model to generate a single accurate output; instead, it requires multiple plausible future options to be represented [99][94], a concept termed "multimodal prediction" [1][108].

Technically, there are issues like dealing with noisy or missing data, or generalizing to new unseen environments. Dealing with human intent and modeling social interaction are also still major challenges, especially in complicated or busy environments[99][94].

These multidimensional problems underscore the necessity for a hybrid model that merges the interpretability of rule-based models encoding structured human behaviors and the strong pattern recognition properties of deep learning[99]. This synthesis has the potential to surmount existing limitations of the field and render pedestrian motion prediction more trustworthy[99][94].

4.5 Research Questions:

RQ1: What motion characteristics and behavioral patterns can be extracted from the JAAD dataset to support pedestrian trajectory prediction?

RQ2: How effective is the proposed hybrid model combining rule-based features with deep learning in predicting pedestrian trajectories using clean data?

RQ3: How robust is the hybrid prediction model when applied to real-time, unprocessed video data from the JAAD dataset?

RQ4: Can the hybrid prediction framework operate effectively in real-time and meet the temporal constraints of practical deployment?

4.6 Methodology:

This section outlines the experimental procedures used to address each research question, drawing upon the JAAD dataset and the proposed hybrid prediction architecture.

4.6.1 RQ1 – Behavioral and Motion Analysis of JAAD Data:

The JAAD dataset was analyzed to extract motion and behavioral features relevant to pedestrian trajectory prediction. XML annotations were parsed to obtain frame-level data such as position, speed, behavior state, and context. These features informed the design of the rule-based component in the hybrid model, with a focus on subtle behaviors like hesitation and vehicle interaction.

4.6.2 RQ2 – Prediction on Clean Data:

To evaluate the model under ideal conditions, the dataset was preprocessed to remove noise and ensure consistency. Rule-based features were encoded numerically and used as inputs alongside position and motion data. A hybrid architecture combining bidirectional LSTM layers with multi-head attention was trained and evaluated using standard metrics such as ADE, FDE, and MSE to assess spatial accuracy and trajectory shape fidelity.

4.6.3 RQ3 – Prediction on Real-Time Data:

The model was deployed on raw video sequences from the JAAD dataset without extensive preprocessing. A real-time prediction **framework** was implemented to simulate deployment scenarios, incorporating smoothing, feature extraction, and on-the-fly inference. Performance was evaluated using the same metrics, with a focus on model robustness in the presence of noise, occlusion, and irregular motion.

4.6.4 RQ4 – Real-Time Inference Evaluation:

To assess real-time feasibility, the full prediction framework was executed frame-by-frame on JAAD video data. Visualization outputs were generated to overlay predicted, observed, and ground truth trajectories. The model's temporal responsiveness and accuracy under real-world constraints were analyzed, supporting its potential integration in real-time autonomous systems.

4.7 Conclusion:

In this chapter, we presented a structured and formalized presentation of the problem of pedestrian trajectory prediction. We first presented a thorough overview, emphasizing the real-world relevance and difficulty of predicting human motion in dynamic scenes. We then presented a mathematical framework to delineate the inputs, outputs, and contextual variables of interest for prediction tasks. Furthermore, we discussed the significant limitations and uncertainties which work towards the intrinsic complexity of the issue, including multimodal behaviors, interpersonal relationships, and environmental factors.

This background sets the stage for the next chapter, in which we present the JAAD dataset, an essential tool used for training and testing pedestrian trajectory prediction models in urban settings.

CHAPITRE V: JAAD DATASET

5.1 Introduction:

Pedestrian trajectory prediction in urban environments requires accurate and diverse datasets to capture the complexities of human behavior. The JAAD dataset stands out as a key resource for this task, providing a rich collection of short video clips that depict real-world interactions between pedestrians and vehicles in dynamic urban settings. Captured from a dashboard-mounted camera, these videos are annotated with detailed behavioral, contextual, and demographic information, making them invaluable for developing models that can predict pedestrian movements in the context of autonomous driving and other intelligent systems. The dataset's ability to capture subtle pedestrian behaviors, such as hesitation before crossing or eye contact with drivers, makes it particularly useful for socially-aware modeling. Despite its advantages, the JAAD dataset has certain limitations, including potential biases related to its geographical focus and the fixed perspective of the camera. Nevertheless, it offers an essential foundation for advancing pedestrian trajectory prediction models.

5.2 Dataset Overview:

The JAAD (Joint Attention for Autonomous Driving) dataset was created to support the study of pedestrian behavior in urban traffic environments,[109] with a particular emphasis on the interactions between pedestrians, vehicles, and surrounding infrastructure. It includes 346 short video clips, ranging from 5 to 10 seconds, captured using a dashboard-mounted camera in various real-world conditions.[102]

Each video is accompanied by detailed annotations provided in XML format. These annotations encompass multiple levels of information, including pedestrian actions, environmental context, and demographic attributes. The dataset also incorporates additional resources such as traffic and vehicle annotations, and appearance-based descriptions. The overall structure and content of the dataset are illustrated in Figure 8, which outlines the different components and the type of data each contains.[101]

This dataset serves as a valuable resource for tasks such as pedestrian intent prediction, behavior modeling, and the development of socially aware autonomous driving systems.

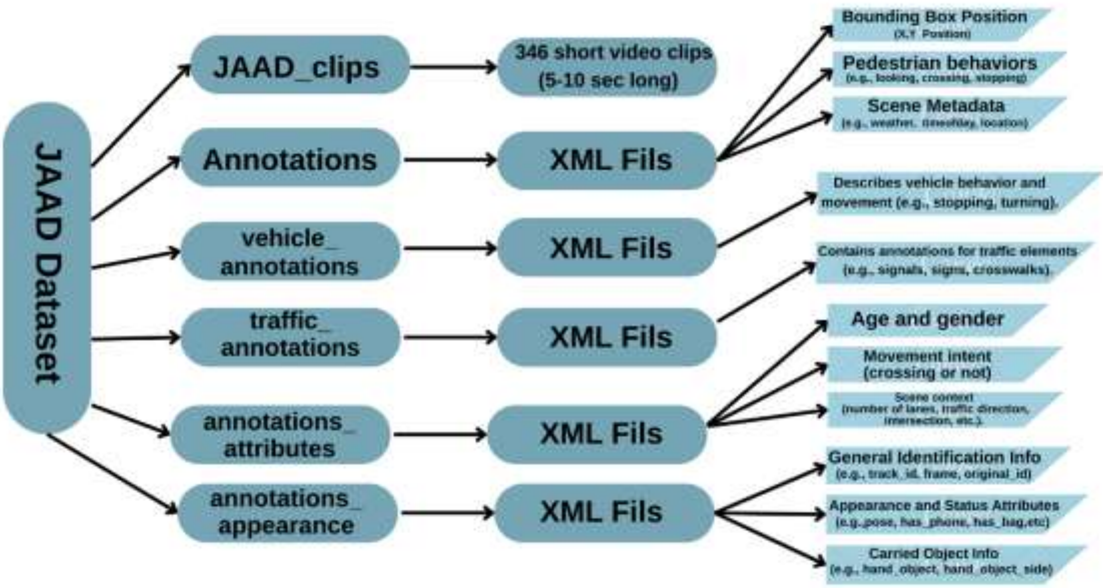

Figure 9: Structure of JAAD Dataset and its Annotated Components

One of the key strengths of the JAAD dataset is its heterogeneity, as data was collected in various urban environments across Eastern Europe and North America.[109] This diversity encompasses a wide range of pedestrian behaviors, traffic regulations, and cultural norms, making the dataset highly suitable for developing models with strong generalizability across different settings[101]. The dataset is widely used in tasks such as trajectory prediction, action recognition, and behavior modeling, and has proven especially valuable in scenarios involving subtle human actions, such as hesitation before crossing or checking for oncoming traffic. [110]

In Figure 10, a practical example of such subtle behaviors is illustrated through a temporal sequence showing the interaction between a pedestrian and a driver during a street-crossing event. The figure shows how the pedestrian initially looks toward oncoming traffic, begins to move slowly, and then crosses the street, while the driver slows down in response. This moment-by-moment coordination underscores the relevance of the JAAD dataset in capturing nuanced human behaviors in complex urban environments.[111]

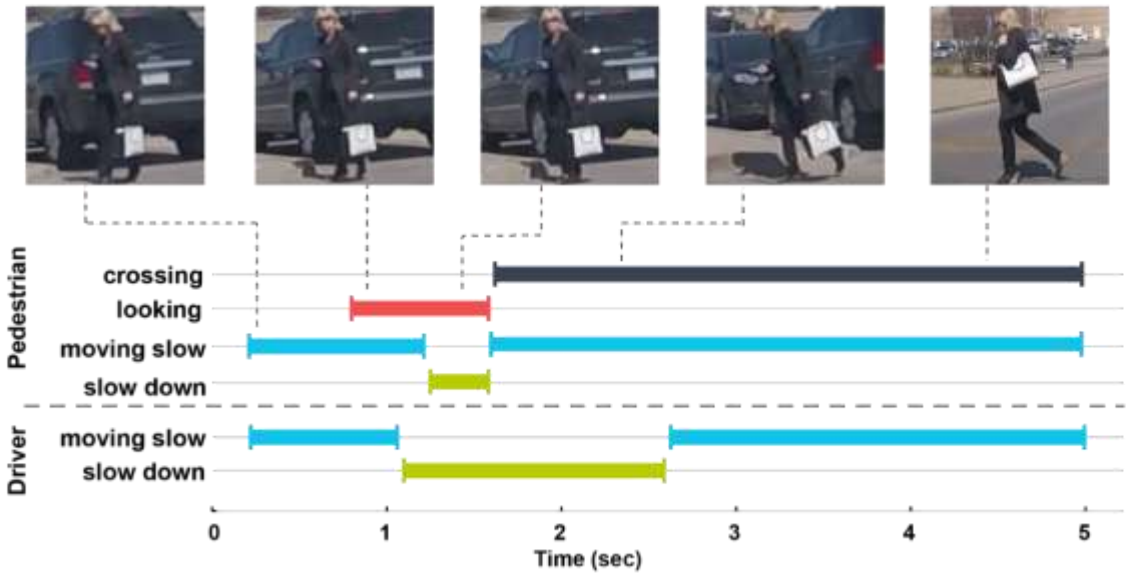

Figure 10: Timeline of Annotated Pedestrian Behaviors in the JAAD Dataset

[111]

5.3 Annotation Format and Labeling Scheme:

The JAAD dataset offers XML-formatted annotations, with one XML file per video clip[109]. The XML files contain frame-by-frame detailed information on all entities seen, predominantly pedestrians and vehicles, along with environmental contextual data and behavioral labels[110]. The metadata in each XML file comprises the video ID number, weather, geographical location, and temporal information[112]. Moreover, it is a set of annotated objects, each of which has a unique identifier for tracking between frames[113]. For every single object, the annotation specifies the bounding box coordinates (x, y, width, height), categorizes the object type (e.g., pedestrian or vehicle), and includes contextual behavioral tags such as "looking", "crossing", or "standing" [112].Temporal data is included as well, indicating the frame numbers upon which each object appears, thus facilitating a full temporal examination of their movement and behavior.[114]

5.4 Dataset Statistics:

JAAD dataset contains 346 short video segments, each lasting between 5 and 10 seconds, and totaling over 55,000 frames[101]. The video segments represent diverse urban settings and dynamic interactions between pedestrians and vehicles in real-world contexts[112]. The dataset covers a number of behavioral categories, such as "crossing," "looking," and "standing," thus providing a rich variety of pedestrian behaviors for analysis[109]. The dataset includes approximately 700 monitored entities, predominantly pedestrians, with accurate spatial coordinate and behavioral state annotations per individual frame[115]. Furthermore, the captured videos feature varying lighting and weather conditions, both day and night scenes, and fair and inclement weather, thereby adding variability and strength of the data essential for learning generalizable predictive models[116].

5.5 Advantages of the JAAD Dataset:

The JAAD dataset presents a combination of strengths that render it especially qualified for investigation into pedestrian action prediction and the development of socially-aware autonomous driving systems[101]. Unlike standard trajectory datasets, JAAD is not just supplying frame-by-frame bounding box annotations but also extensive behavioral annotations describing human actions and intentions, i.e., "looking", "crossing", and "waiting"[109]. A special feature of JAAD is that it includes contextual metadata such as weather, time of day, and scene attributes (i.e., traffic lights, crosswalks) that are essential in understanding real-world interactions[115]. JAAD also emphasizes joint attention dynamics such as when a pedestrian establishes eye contact with an approaching driver making it extremely useful for socially intelligent agent modeling[112].

Additionally, the dataset offers variation in geographical location, camera perspective, and pedestrian populations (gender, age, etc.), rendering it more generalizable to environments[109]. The videos are taken in urban and suburban regions with numerous types of lighting and weather conditions, which renders any model trained using it more realistic and stronger[102]. JAAD is characterized by its relative compactness and ease of use, which makes it a suitable candidate for both academic research and industry prototyping plans. These qualities render JAAD one of the most behaviorally rich datasets on the topic of pedestrian trajectory and intention prediction[101].

5.6 Comparison with Other Datasets:

The JAAD dataset has some distinct characteristics that set it apart from other widely used pedestrian trajectory datasets, including ETH, UCY, SDD, and nuScenes. As opposed to ETH and UCY, which were recorded by fixed overhead cameras in controlled urban environments[97], JAAD is recorded by a dashboard-mounted camera on a car, thus offering a first-person perspective of driving[102]. This characteristic renders it especially valuable for autonomous vehicle-related studies. In addition, JAAD includes rich behavioral annotations, such as pedestrian gaze, gestures, and crossing intentions annotations that are often absent in other datasets that focus on position coordinates alone[101]. While other datasets like nuScenes and Waymo provide larger-scale data and 3D lidar input[117], they generally lack the fine-grained pedestrian-centered behavioral annotation of JAAD. However, JAAD is relatively smaller in scale and geographic diversity, which may limit its generalizability[101]. Despite this, the contextual richness and practical authenticity make it a strong complement to other benchmarks in trajectory prediction[101].

5.7 Limitations and Biases:

Despite the vast size of the JAAD dataset, it is subject to several limitations and potential biases that must be carefully considered when applied to pedestrian trajectory prediction. The data is only captured from a frontal dashboard camera, resulting in a perspective bias; pedestrians are predominantly viewed from frontal or lateral angles, and their movement is restricted to the scenes within the field of view of the vehicle[102]. Besides, the dataset predominantly includes urban and suburban settings, which are restricted in terms of geographical or cultural contexts[101]. A majority of the clips are captured in North America or Europe, which cannot possibly generalize well to other traffic or pedestrian behavior patterns observed in other parts of the world. A second source of bias arises from choosing annotated behaviors; although the data set contains exhaustive labels such as "crossing" or "looking," it perhaps underrepresents unclear or not-so-common pedestrian intentions. Last, since each video has a limited duration (between 5 to 10 seconds), long-term behavior patterns or infrequent events are not possible[102].

5.8 Conclusion:

In conclusion, the JAAD dataset is a rich and challenging resource for analyzing pedestrian behavior in real urban traffic conditions. Its unique viewpoint from a car dashboard camera, combined with frame-level behavioral annotations, makes it particularly valuable for autonomous driving research and human-centered trajectory prediction. Although it offers strengths in terms of behavioral context and realism, it is limited by dataset scale and geographical diversity. Nevertheless, when paired with other datasets or carefully fine-tuned for specific tasks, JAAD can significantly contribute to the development of more context-aware and socially intelligent motion prediction models.

Building upon the behavioral insights extracted from JAAD, the next chapter introduces a hybrid prediction framework that integrates rule-based reasoning with deep learning. This approach aims to combine interpretable behavioral features with the modeling power of neural networks to improve both short-term and long-term trajectory prediction accuracy.

CHAPTER VI: PREDICTION ON CLEAN DATA

6.1 Introduction:

This chapter presents a hybrid pedestrian trajectory prediction model that combines rule-based reasoning with deep learning, Such as LSTM, BiLSTM, GRU network with attention mechanisms. Behavioral features extracted from the JAAD dataset are processed to capture complex temporal and semantic patterns. The model aims to improve prediction accuracy by leveraging both rule-based insights and deep learning capabilities. Evaluation using multiple metrics shows that this hybrid approach outperforms traditional baseline models.

6.2 Data Exploration and Pre-processing:

Rule-based models are among the earliest approaches employed in pedestrian trajectory prediction[118]. Rule-based systems are controlled by strict logical rules that define agent behaviors in a given context [119]. As opposed to data-driven systems that depend on acquiring knowledge from large data, rule-based systems operate by defining pre-established patterns of behavior based on human reasoning or observational facts[120][121].

6.2.1 Principles Overview in Rule-Based Modeling:

Rule-based modeling is founded on deterministic rules derived from behavioral studies, expert opinion, or engineering hypotheses that specify the way agents respond to environmental cues[121]. Rule-based models utilize conditional logic to simulate pedestrian behavior within organized spaces[122]. Among the key strengths of rule-based methods is interpretability since every decision is a result of clearly defined rules. Rule-based models also provide high computational efficiency, which makes them amenable to real-time application and low-resource settings[121][123]. While rule-based systems can be fitted to new situations through hand-tuning of the rules, their use of static rules limits their deployment over large populations in dynamic or uncertain environments. Without continuous human intervention or integration with learning-based approaches, performance on difficult cases might deteriorate[124][125].

6.2.2 Key Rules in Rule-Based Pedestrian Movement Models:

Rule-based pedestrian movement models simulate realistic human navigation using a set of deterministic rules that capture common behaviors in structured environments[121][126]. These rules provide interpretable and computationally efficient mechanisms for modeling pedestrian decision-making. In the context of data-driven forecasting and real-time prediction[120], certain behavioral patterns are especially relevant and are outlined below:

a) Goal-Oriented Movement:

Pedestrians are modeled to move purposefully toward a predefined destination when one exists, simulating navigation toward transit points, exits, or areas of interest[118].

b) Obstacle Avoidance:

Upon encountering a static obstacle such as a wall, vehicle, or other street infrastructure the pedestrian modifies their trajectory to avoid collisions[127].

c) Collision Avoidance with Other Pedestrians:

When another pedestrian is detected within a certain safety range and on a potential collision course, the individual alters their speed or direction, maintaining personal space and avoiding overlap[13].

d) Traffic Signal Compliance:

At designated crossings, pedestrians respond to traffic signals halting at red and proceeding on green. This behavior is essential for lawful and safe street-level interaction with vehicular traffic[128][129].

e) Density-Based Speed Adjustment:

As crowd density increases, pedestrians naturally slow down, mirroring real-world congestion effects and ensuring smoother navigation through crowded areas[130].

While numerous additional rules exist such as group synchronization, attraction-repulsion responses, and dynamic path replanning these are referenced primarily to contextualize the simulation framework. Their inclusion may be considered for future work or in broader behavioral modeling settings [130].

6.2.3 Numerical Encoding of Rule-Based Behavioral Features:

While rule-based pedestrian models are defined conceptually through deterministic behavioral rules, their practical implementation within a simulation or predictive framework requires transforming these rules into structured numerical features[131][132]. These features serve as intermediate representations that encapsulate key kinematic and directional aspects of each pedestrian agent, enabling the integration of behavior rules into data-driven or hybrid modelling[98].

The feature vector constructed for each predicted pedestrian position includes the following components:

- 2D Position: The predicted spatial location of the pedestrian at a given timestep[98].
- Velocity Magnitude: Calculated using the Euclidean norm of the agent's average velocity, scaled based on the prediction index to reflect temporal dynamics[133].
- Acceleration and Jerk: Both initialized as zeros in this context, these placeholders may be expanded in future iterations to capture second- and third-order motion dynamic[134].
- Heading Angle: Derived from the velocity vector using the arctangent of its components, this represents the orientation of movement[1].
- Turning Rate: Currently set to zero, this feature could represent changes in heading over time.
- Speed and Speed Deviation: Redundant speed information is included for clarity and possible normalization needs; deviation is set to zero, indicating steady-state motion[108].
- Relative Position: A placeholder for potential future features capturing interaction with other agents or obstacles[98].
- Movement Direction: Also represented via the heading angle, reinforcing directional consistency in the feature space.

This structured encoding bridges the gap between high-level behavior rules such as goal pursuit, collision avoidance, or group synchronization and their numerical counterparts used within simulation or learning modules. It ensures that the behavioral logic is faithfully captured in forms interpretable by computational models[135].

To further clarify the structure and computation of each feature, Table 2 summarizes the key components of the feature vector along with their mathematical formulations and relevant implementation notes.

Feature Name	Symbol / Expression	Computational Formula	Notes
2D Position	$p_i = [x, y]$	Directly taken from predicted position $\boldsymbol{pred}[\boldsymbol{i}]$	Represents spatial location at timestep i
Velocity Magnitude[136]	$\|\|v_i\|\|$	$\text{avg}_{\text{vel}} \times \min\left(1.0, 0.8 + \frac{(0.2 \times i)}{N}\right); \|\|v\|\| = \sqrt{\{v_x^2 + v_y^2\}}$	Time-scaled average velocity norm
Acceleration[136]	a_i	Currently set to 0.0; $a = \frac{\Delta v}{\Delta t}$	Placeholder; may capture second-order motion
Jerk [137]	j_i	Currently set to 0.0; later $j = \frac{\Delta a}{\Delta t}$	Placeholder; models smoothness of motion changes
Heading Angle[138]	θ_i	$\theta = \arctan 2(v_x, v_y)$	Captures movement orientation
Turning Rate[139]	θ_i	Set to 0.0; $\frac{\Delta_\theta}{\Delta_t}$	Placeholder for angular dynamics
Speed	s_i	Same as $\|\|v_i\|\|$	Repeated for clarity and model compatibility
Speed Deviation	Δs_i	Set to 0.0; future: $\Delta_s = s_i - \bar{s}$	Captures deviation from average speed
Relative Position[16]	Δp	Set to 0.0; future: $\Delta_p = p_i - p_{ref}$	Placeholder for relative distance to agents or obstacles
Movement Direction[140]	φ_i	$\varphi = \arctan 2\,(v_x, v_y)$	Ensures directional consistency across features

Table 2: Summary of Numerical Features for Rule-Based Pedestrian Modeling

6.3 Our Architecture:

Our architecture implements a hybrid framework that combines interpretable rule-based computations with a deep learning model (LSTM,GRU,BiLSTM) for pedestrian trajectory prediction. The pipeline begins by loading CSV annotations extracted from the JAAD dataset. These annotations contain pedestrian positions and other contextual data, which are processed by a Rule-Based Module (RBM) to compute motion features like velocity, acceleration, and goal-related behavior.

The architecture is modular and sequential, starting with raw positional data preprocessing, followed by handcrafted social feature extraction, sequence formatting, and ending with deep learning-based trajectory forecasting using models such as LSTM, BiLSTM, or GRU. This design balances interpretability and robustness, as shown in Figure 11, where rule-based reasoning effectively complements and enhances the neural prediction process.

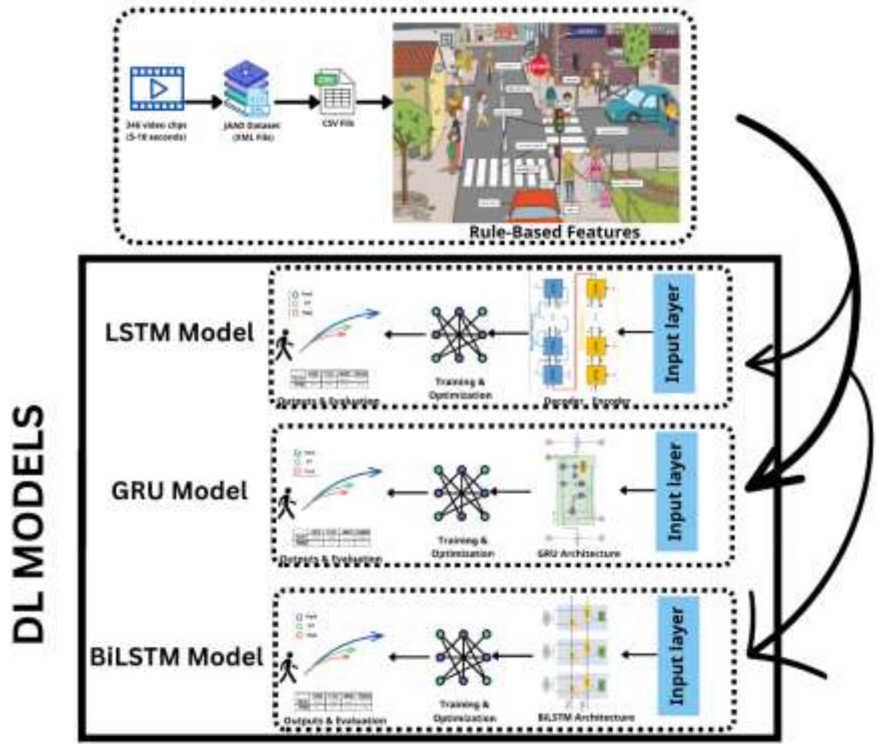

Figure 11: Hybrid architecture for pedestrian trajectory prediction

6.3.1 Input Features:

The input features represent structured temporal sequences of vectors characterizing pedestrian motion. These features include positional, kinematic, and social attributes essential for understanding and predicting pedestrian trajectories. Initially, they are directly extracted from raw time-series coordinate data (CSV format) of the JAAD dataset and subsequently enriched by a Rule-Based Module (RBM), which computes additional metrics such as velocity, acceleration, inter-pedestrian interactions, and spatial orientation relative to defined contextual cues.

a) Features Extracted from Raw JAAD Dataset:

The raw dataset provides fundamental spatial and contextual information, extracted directly from annotations, including pedestrian position, identification, and environmental context. These primary features describe each pedestrian's situational state dynamically:

- **center_x, center_y:** Pedestrian centroid coordinates (horizontal and vertical).
- **frame:** Video frame number.
- **pedestrian_id:** Unique pedestrian identifier.
- **occlusion:** Degree of pedestrian occlusion (partial or full).
- **crossing_point:** Binary indicator if pedestrian is at a crossing.
- **decision_point:** Binary indicator for decision-making zones.
- **intersection:** Categorical label of intersection presence.
- **ped_crossing:** Presence of pedestrian crossing signals.
- **ped_sign:** Presence of pedestrian signage.
- **stop_sign:** Presence of stop signs.
- **group_size:** Number of pedestrians in the pedestrian's group.
- **motion_direction:** Categorical descriptor (e.g., lateral, longitudinal).
- **vehicle_action:** Pedestrian interaction with vehicles (e.g., walking behind).

b) Behavioral states:

The RBM enhances raw data by generating binary semantic flags and scalar metrics to represent behavioral states and social context, derived through logical rules analyzing speed, acceleration, spatial relations, and traffic signals:

- **velocity:** Scalar magnitude of velocity (Euclidean norm of x and y components).

$$\mathbf{Velocity} = \frac{\Delta_{Position}}{\Delta_t} \text{ , } \Delta_{\boldsymbol{t}} = 0.033 \; seconds \; [136]$$

Equation 3: Velocity

- **acceleration:** Scalar magnitude of acceleration.

$$\boldsymbol{Acceleration} = \frac{\Delta_{velocity}}{\Delta_t} \text{ [136]}$$

Equation 4: Acceleration

- **is_moving:** Binary flag for movement based on velocity threshold.
- **is_accelerating:** Binary flag for acceleration above threshold.
- **approaching_crossing:** Binary flag indicating proximity to crossing and decision points.
- **in_intersection_zone:** Binary flag for presence inside an intersection zone.
- **is_occluded:** Binary flag for occlusion state.
- **traffic_control_active:** Binary flag for active pedestrian-related traffic signals.
- **group_context:** Normalized group size relative to largest observed group.
- **is_lateral, is_longitudinal:** Flags for movement direction type.
- **interacts_with_vehicle:** Flag for active pedestrian-vehicle interaction (excluding "none").

Feature	Description	State 1	State 2
velocity	Instantaneous pedestrian speed	velocity>0.5m/s (moving)	Velocity <= 0.5 (stationary)
acceleration	Instantaneous pedestrian acceleration	acceleration > 0.5 (accelerating)	Acceleration <= 0.5 (not accelerating)
is_moving	Movement state (moving/not moving)	1 if velocity > 0.5	0 if velocity <= 0.5
is_accelerating	Acceleration state (accelerating/not)	1 if acceleration > 0.5	0 if acceleration <= 0.5
approaching_crossing	Approaching crossing and decision point	1 if near crossing and decision point	0 if not near
in_intersection_zone	Inside an intersection zone	1 if inside intersection zone	0 if outside zone
is_occluded	Occlusion status (partial or full)	1 if occluded	0 if visible
traffic_control_active	Active pedestrian-related traffic signals	1 if traffic control signals active	0 if no active signals
group_context	Group size (normalized metric)	Value between 0 (alone) and 1 (largest group)	-
is_lateral	Lateral movement direction	1 if lateral movement	0 if not lateral
is_longitudinal	Longitudinal movement direction	1 if longitudinal movement	0 if not longitudinal
interacts_with_vehicle	Interaction with vehicles	1 if interacting with a vehicle	0 if no interaction

Table 3: Description of Input Features with Mathematical Conditions for Pedestrian Behavior Analysis

6.3.2 Deep Learning Model:

The Deep Learning (DL) model in this system is designed to predict future pedestrian positions using sequences of spatiotemporal features enriched by a Rule-Based Module (RBM). The input vectors incorporate both raw positional data and engineered physical-social descriptors, enabling the model to capture short- and long-term motion patterns.

a) Input Layer:

This layer receives preprocessed sequences of feature vectors for each pedestrian. These vectors include:

- **Raw spatial coordinates:** center_x, center_y
- **Derived physical features:** velocity_x, velocity_y, acceleration_x, acceleration_y
- **Social and goal-related descriptors:** nearest_neighbor_dist, nearest_neighbor_velocity_x, nearest_neighbor_velocity_y, goal_angle, goal_distance
- **Input shape:** `(batch_size, seq_len = 100, features_dim ≈ 11+)`

These features collectively represent both the individual motion and the surrounding context, offering a comprehensive snapshot of the pedestrian's state at each time frame.

b) LSTM Encoder:

The core of the model consists of stacked Long Short-Term Memory (LSTM) layers that capture temporal dependencies:

- **First LSTM Layer:** 128 hidden units, processes the input sequence into a compact temporal encoding.
- **Dropout (0.3):** Introduced after the first LSTM to mitigate overfitting.
- **RepeatVector Layer:** Duplicates the temporal context to fit the output sequence length.
- **Second LSTM Layer:** 64 units, returns a sequence with one output per predicted future time step.
- **Dropout (0.3):** Applied again to regularize the output dynamics.

These layers enable the model to learn both immediate transitions and longer-term trends in pedestrian behavior.

c) Decoder Layer:

This module transforms the encoded temporal features into concrete spatial predictions:

- **TimeDistributed(Dense(2))** applies a fully connected layer at each time step of the predicted sequence.
- **Output:** `(batch_size, pred_len = 50, 2)` corresponding to 50 future (x, y) positions per pedestrian.

This simple decoder design ensures frame-level prediction with minimal computational overhead.

d) Training and Optimization:

The model is trained end-to-end using the following setup:

- Loss Function: Mean Squared Error (MSE)
- Optimizer: Adam with a learning rate of 0.001
- Batch Size: 32
- Training Epochs: 50, with early stopping after 3 epochs without validation improvement

Prior to training, all inputs and targets are normalized. Sequences containing NaN or infinite values are filtered to maintain data integrity.

e) Output and Interpretation:

After training, the model produces predicted trajectories that are:

- Frame-by-frame position sequences for 1.5 seconds into the future (assuming 30 FPS).
- Post-processed using inverse normalization to restore coordinates to their original scale.

These predictions are compared against ground truth using quantitative metrics (MSE, CMSE, FDE) and are visualized to assess spatial and temporal accuracy.

To ensure robust evaluation and performance benchmarking, two additional recurrent architectures **Bidirectional LSTM (BiLSTM)** and **Gated Recurrent Unit (GRU)** were implemented using the same input structure, preprocessing pipeline, and training configuration as the baseline LSTM model. While all three models share the same foundational design and function similarly in their sequence-to-sequence prediction setup, they differ in internal mechanisms and computational characteristics, as detailed in the following comparative table.

Feature	LSTM	BiLSTM	GRU
Directionality	Unidirectional	Bidirectional	Unidirectional
Number of Gates	3 (Forget, Input, Output)	Same as LSTM	2 (Update, Reset)
Cell State	Present	Present	Not Present
Architectural Complexity	High	Very High (≈2× LSTM)	Moderate
Training Speed	Relatively slow	Slower	Faster
Performance	Good for long dependencies	Excellent for full-context	Good for moderate dependencies
Training Resources	High	Very High	Moderate

Table 4: Comparative Summary of LSTM, BiLSTM, and GRU Architectures

These differences highlight the trade-offs between model complexity, performance, and computational efficiency. By evaluating LSTM, BiLSTM, and GRU under the same conditions, we aim to identify the most suitable architecture for pedestrian trajectory prediction in terms of accuracy, generalization, and resource usage.

6.4 Evaluation metrics:

To rigorously assess the performance of pedestrian trajectory prediction models, a diverse set of metrics is utilized. These metrics evaluate both spatial accuracy and the dynamical consistency of predicted trajectories, providing a comprehensive view of model effectiveness[46][141].

6.4.1 Basic Position Metrics:

These metrics quantify how close the predicted trajectory is to the ground truth in Euclidean space.[141]

a) Average Displacement Error (ADE):

ADE measures the mean Euclidean distance between each predicted position and the corresponding ground truth across all time steps in the sequence[142].

Mathematical Expression:

$$\boldsymbol{ADE} = \frac{1}{T} \sum_{t=1}^{T} \|\hat{p}_t - p_t\| \qquad [142]$$

Equation 5: ADE(Average displacement error)

Interpretation:

- A lower ADE indicates better average positional accuracy. It is sensitive to all deviations throughout the sequence.

b) Final Displacement Error (FDE)

FDE measures the Euclidean distance between the final predicted position and the final ground truth position[142].

Mathematical Expression:

$$\mathbf{FDE} = \|\hat{p}^t - p^t\|^2 \qquad [142]$$

Equation 6: FDE(Final displacement error)

Interpretation:

- This metric is particularly important in applications where the final destination is critical.

c) Mean Squared Error (MSE):

MSE calculates the average of squared Euclidean distances between predicted and actual positions[143][144].

Mathematical Expression:

$$\boldsymbol{MSE} = \frac{1}{T} \sum_{t=1}^{T} \|\hat{p}_t - p_t\|_2^2 \qquad [143]$$

Equation 7: MSE(Mean squred error)

- **Interpretation**:
  MSE penalizes larger errors more than ADE, making it useful for identifying occasional large mistakes.

6.4.2 Centered Metrics:

Centered metrics evaluate prediction accuracy by measuring errors relative to the trajectory's central tendencies, offering a more robust assessment of spatial alignment and distribution[1].

a. CFMSE (Centered Final Mean Squared Error):

Measures the squared error between predicted and ground truth final positions, after centering both trajectories by subtracting their respective means. This emphasizes trajectory shape over absolute position[145].

Mathematical Formula:

Let p_i and r_I be the predicted and ground truth positions at time i, and N the total number of frames.

Center both trajectories:

$$\bar{\boldsymbol{r}} = \frac{1}{N}\sum_{i=1}^{N} r_i \quad , \bar{\boldsymbol{p}} = \frac{1}{N}\sum_{i=1}^{N} p_i \qquad [145]$$

Equation 8: CFMSE(Centerred final mmean squared error)

Then compute the final centered positions:

$$\boldsymbol{p^c_{final}} = \bar{p} - p_N \quad , \boldsymbol{r^c_{final}} = \bar{r} - r_N \qquad [145]$$

Equation 9:CFMSE(Centerred final mmean squared error)

Then,

$$\boldsymbol{CFMSE} = \left\| p^c_{final} - r^c_{final} \right\|^2 \qquad [145]$$

Equation 10: CFMSE(Centerred final mmean squared error)

Interpretation:

- Focuses on trajectory shape rather than spatial location
- Useful when trajectories are similar in form but translated in space.

b. CMSE (Centered Mean Squared Error):

Calculates the meansquared error between predicted and real trajectories after removing their mean positions (centering). This approach focuses on the similarity in the shape ofthe trajectories rather than their absolute positions[145].

Mathematical Formula:

$$\boldsymbol{CMSE} = \frac{1}{N}\sum_{i=1}^{N} \|(\bar{p} - p_i) - (\bar{r} - r_i)\|^2 \qquad [145]$$

Equation 11: CMSE (Centered Mean Squared Error)

Interpretation:

- Removes absolute positional bias by centering both predicted and real trajectories.
- Emphasizes alignment in trajectory *shape* and *motion pattern*.
- Useful for evaluating the model's ability to capture relative movement trends independent of global location shifts.

6.5 Evaluation results:

To evaluate the predictive performance of our proposed pedestrian trajectory models, we conducted a systematic assessment using the JAAD dataset. The evaluation involved a comparative analysis of three deep learning architectures and their corresponding hybrid counterparts:

- **Deep Learning Model:** This category includes three separate models Long Short-Term Memory (LSTM), Gated Recurrent Unit (GRU), and Bidirectional LSTM (BiLSTM) each trained using only the raw spatial coordinates (center_x, center_y) of pedestrians as input features.

- **Hybrid Model (Rule-Based + DL):** For each of the above architectures, we developed a hybrid version that integrates hand-crafted features derived from a Rule-Based Model (RBM). These features include velocity, acceleration, distance to nearby agents, and goal-oriented vectors, which are used alongside the raw spatial coordinates as inputs to the corresponding deep learning model.

This comparative evaluation allows us to quantify the contribution of rule-based behavioral components to the predictive accuracy of each deep learning architecture. The results highlight how integrating structured domain knowledge through rule-based features can enhance the model's ability to anticipate future pedestrian movements.

6.5.1 Quantitative Evaluation:

The following metrics were used to evaluate model performance:

- **MSE (Mean Squared Error):** The overall average of squared errors across the predicted sequence.
- **MSE @ t (0.5s, 1.0s, 1.5s):** Pointwise MSE evaluated at specific future time horizons corresponding to 15, 30, and 45 frames.
- **CMSE (Cumulative MSE):** The average MSE accumulated over the entire 1.5-second prediction horizon.
- **FDE (Final Displacement Error):** The Euclidean distance between the predicted and actual positions at the final prediction frame (1.5s).
- **ADE (Average Displacement Error):** The average Euclidean distance between predicted and ground truth trajectories over all time steps.
- **CFMSE (Centered Final Mean Squared Error):** Squared error at the final frame after centering both trajectories to focus on shape similarity.
- **Global MSE (X, Y):** Separate MSE values for the X and Y position coordinates across the sequence.

All results are reported in pixel-based Euclidean units, with a frame rate of 30 FPS. The table below summarizes the results of all models across these metrics, comparing three deep learning architectures (LSTM, GRU, BiLSTM) and their corresponding hybrid versions (incorporating rule-based components). The best-performing model per metric is also indicated.

Model	ADE	FDE	CMSE (1.5s)	CFMSE	Global MSE X	Global MSE Y	MSE @ 0.5s	MSE @ 1.0s	MSE @ 1.5s
LSTM	33.38	58.40	911.78	1588.6989	2036.53	46.18	569.77	389.36	1588.70
GRU	32.07	**54.33**	796.71	1667.9808	1820.89	50.40	497.19	663.87	1667.98
BILSTM	28.94	83.35	642.84	1066.4317	1610.99	34.79	588.48	416.26	1066.43
RBM_LSTM	31.34	60.40	833.18	1829.0155	1938.68	38.53	**391.70**	448.30	1829.02
RBM_GRU	35.38	66.99	913.99	2348.8633	2192.25	49.98	600.02	580.86	2348.86
RBM_BILSTM	**24.36**	78.78	**437.74**	**916.2192**	**1186.71**	**28.60**	357.36	**250.84**	**916.22**
BEST MODEL	RBM_ BILSTM	GRU	RBM_ BILSTM	RBM_ BILSTM	RBM_ BILSTM	RBM_ BILSTM	RBM_ LSTM	RBM_ BILSTM	RBM_ BILSTM

Table 5: Performance Metrics for DL and Hybrid Models

The results presented in Table 5 demonstrate the clear advantage of the hybrid RBM_BiLSTM model over both standard deep learning models and other hybrid variants. Among all models, RBM_BiLSTM achieved the lowest values in key metrics such as ADE (24.36), CMSE at 1.5s (437.74), Global MSE X (1186.71), Global MSE Y (28.60), and MSE at 0.5s, 1.0s, and 1.5s. These results indicate its strong capability to maintain accurate spatial alignment with ground-truth trajectories throughout the prediction horizon. Although the GRU model slightly outperformed others in terms of FDE (54.33), the RBM_BiLSTM model consistently delivered better performance across most metrics, particularly in early and mid-term prediction windows. This confirms the effectiveness of integrating rule-based behavioral features with deep learning architectures, as it enhances the model's generalization and sensitivity to pedestrian movement patterns achieving a more robust and precise trajectory prediction than using DL models alone.

6.5.2 Qualitative Visualization:

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.

a) Deep Learning Model:

To complement the quantitative analysis, we provide qualitative insights into the DL model's prediction behavior using three types of visual evidence:

- **Training Curve:**

The figure shows the training and validation MSE curves for the BiLSTM-DL model over 40 epochs. The model demonstrates a rapid initial decline in error, followed by a gradual convergence, indicating relatively stable learning.

However, a persistent gap between the two curves suggests limited generalization capability and a slight tendency to overfit the training data, which may affect its accuracy on unseen samples.

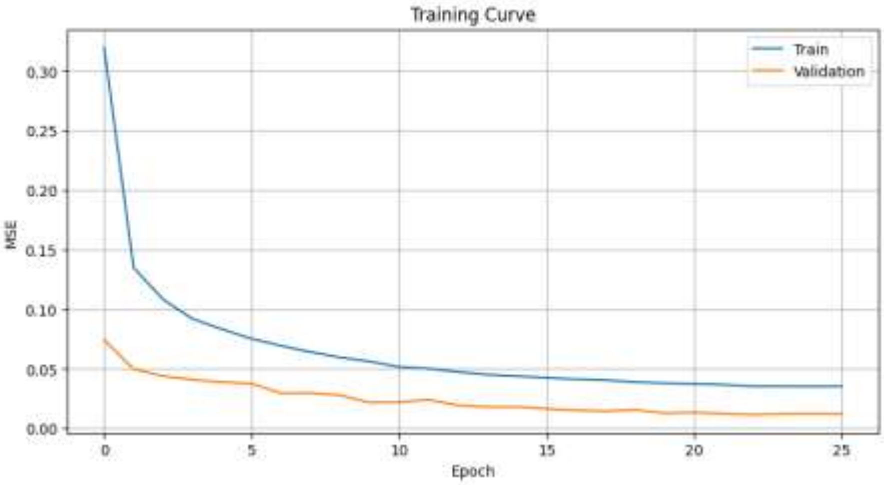

Figure 12: Training and Validation MSE Curve – BILSTM

- **Predicted Trajectories:**

The figure presents five representative pedestrian trajectories used to assess the predictive performance of the BiLSTM-DL model. The model shows strong accuracy in simple, linear motion patterns, particularly in Trajectories 2, 3, and 5, where the predicted paths (in red) closely align with the ground truth (in green).

However, in more complex cases such as Trajectories 1 and 4, the model exhibits noticeable deviations, including overshooting and misalignment with actual pedestrian paths. This indicates that the model still faces challenges in capturing socially or spatially complex movement behaviors.

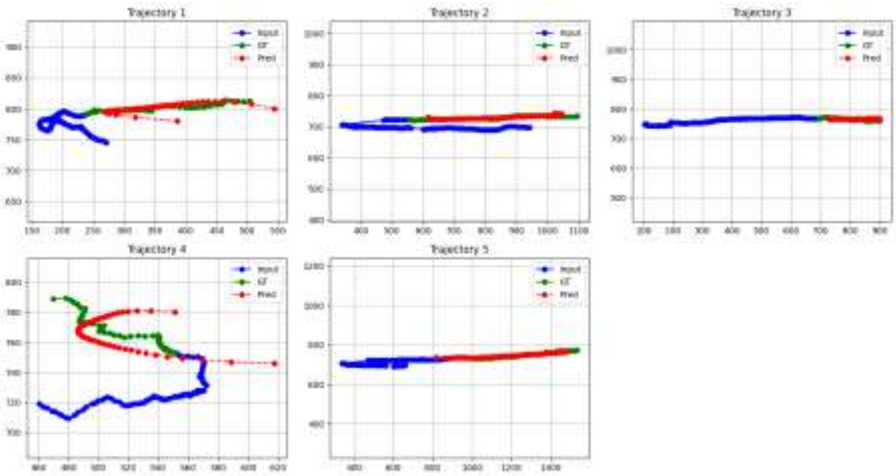

Figure 13: Predicted vs. Ground-Truth Trajectories – BILSTM (Samples 1–5)

- **Positional Distributions:**

The histograms in Figure 14 show the spatial distribution of the actual and predicted pedestrian coordinates along the X and Y axes using the BiLSTM-DL model. Although the predicted distributions follow the general trends of the actual distributions, there are noticeable differences, most notably shifts in the peaks and a reduction in variance. This reflects an excessive smoothing in the predictions and a limitation in capturing the multimodal nature of pedestrian movement.

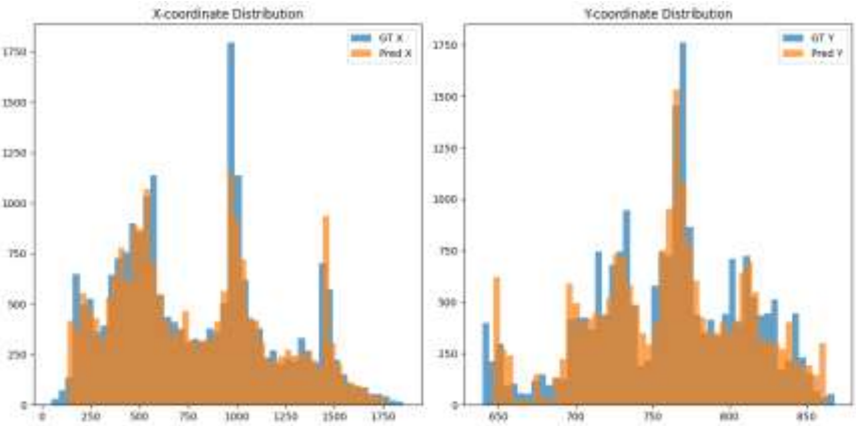

Figure 14: Positional Distribution Comparison (X and Y Axes) – BILSTM

b) Hybrid Model:

To complement the quantitative analysis, we provide qualitative insights into the hybrid model's prediction behavior using three types of visual evidence:

- **Training Curve:**

As illustrated in Figure 15, the training and validation MSE over 32 epochs shows a steep initial decrease, indicating rapid learning during the early stages. Both curves gradually converge as the number of epochs increases, with the validation error consistently lower than the training error. This stable and narrow gap suggests strong generalization capability and minimal overfitting. The behavior reflects the effectiveness of the Hybrid BiLSTM Rule-Based model in capturing essential motion patterns while avoiding excessive reliance on training data.

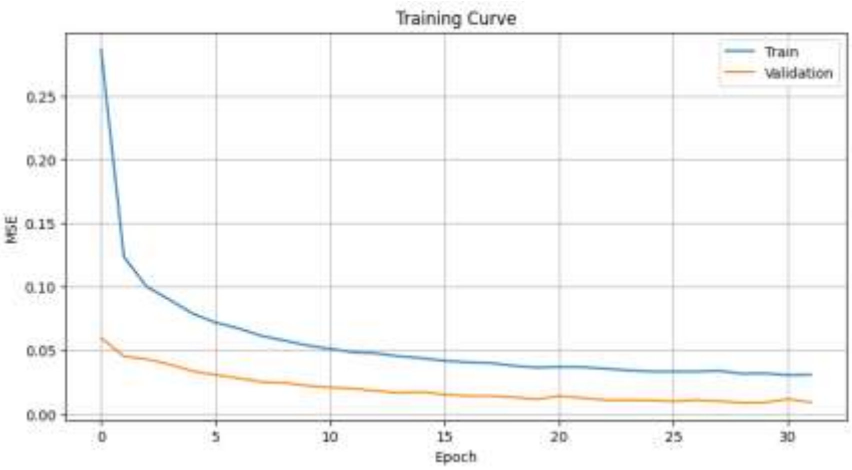

Figure 15: Training and Validation MSE Curve – RBM-BILSTM

- **Predicted Trajectories:**

In Figure 16, the predicted trajectories (red) closely follow the ground truth paths (green), highlighting the effectiveness of the Hybrid BiLSTM Rule-Based model in various scenarios. In relatively straightforward sequences such as Trajectory 2, 3, and 5, the model exhibits high accuracy and stability, with minimal deviation between the predicted and true paths. Notably, in Trajectory 4, which involves more abrupt turns and non-linear motion, the model still manages to maintain a coherent prediction with significantly reduced drift. This suggests that the rule-based correction contributes to better adherence to expected pedestrian behavior, especially under non-linear or sudden directional changes—an area where pure deep learning models often struggle.

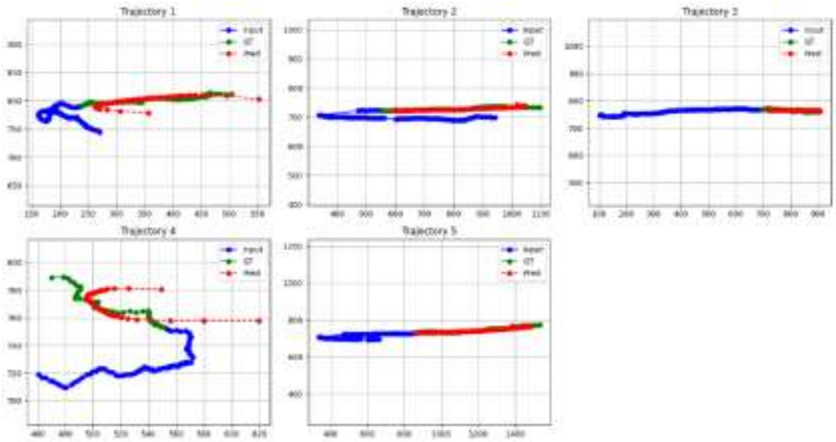

Figure 16: Predicted vs. Ground-Truth Trajectories – RBM-BILSTM (Samples 1–5)

- **Positional Distributions:**

In Figure 17, the histogram plots compare the ground truth (blue) and predicted (orange) coordinate distributions along the X and Y axes. The close overlap between the two distributions indicates that the Hybrid BiLSTM Rule-Based model successfully captures the underlying spatial tendencies of pedestrian movement. Although slight discrepancies are observed at the outer edges particularly in areas of lower density the central regions show strong alignment. This implies the model’s ability to adapt to diverse path structures and reflects enhanced awareness of multimodal and context-driven behaviors, which are often overlooked in purely data-driven models.

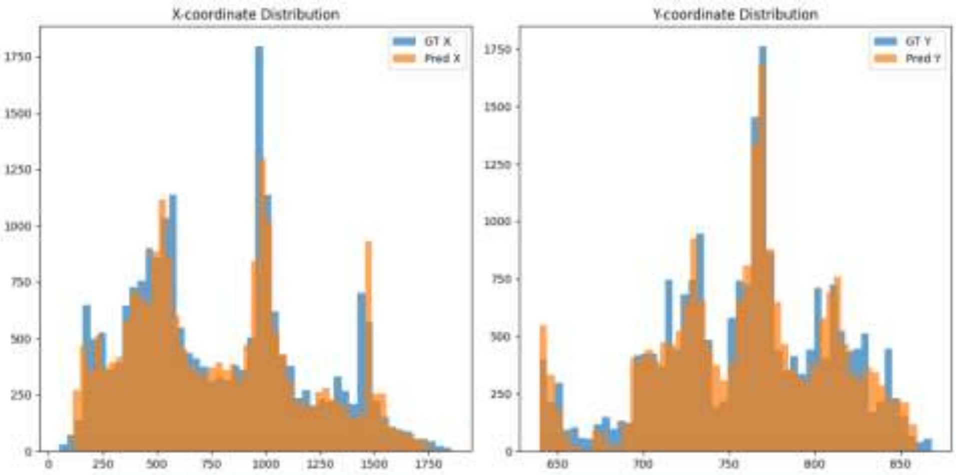

Figure 17: Positional Distribution Comparison (X and Y Axes) – RBM-BILSTM

c. Comparison of Prediction Behaviors: DL vs. Hybrid Model:

The qualitative analysis highlights clear differences in the prediction behaviors of the pure deep learning BiLSTM-DL model and the Hybrid RBM-BiLSTM model.

The BiLSTM-DL model performs well in capturing straightforward and linear pedestrian movements, as evidenced by close alignment between predicted and actual trajectories in simpler cases. However, it faces difficulties with more complex and socially or spatially influenced trajectories. These challenges manifest as spatial drift, overshooting, and misalignment with ground truth paths. Furthermore, the positional distribution analysis reveals mode shifts and reduced variance in the predictions, indicating that the model overly smooths the data and lacks the capacity to fully represent the inherently multimodal nature of pedestrian movement.

In contrast, the Hybrid RBM-BiLSTM model demonstrates enhanced predictive performance across diverse scenarios. By integrating rule-based reasoning with deep learning, it achieves stronger generalization and robustness, particularly in handling abrupt directional changes and nonlinear trajectories where the pure DL model falters. The predicted trajectories show significantly reduced drift and better adherence to actual pedestrian paths. This improvement is corroborated by the coordinate distribution histograms, which exhibit a close match between predicted and ground truth data in both X and Y axes. This suggests the hybrid model's superior ability to capture spatial nuances and multimodal, context-dependent behaviors that are often overlooked by data-driven models alone.

Overall, these findings underscore the benefits of combining data-driven and rule-based approaches. The hybrid model not only mitigates overfitting and excessive smoothing but also delivers more accurate, context-aware pedestrian trajectory predictions, making it a more reliable tool for real-world applications requiring nuanced motion understanding.

6.6 Conclusion:

This chapter evaluated a hybrid pedestrian trajectory prediction model that integrates rule-based reasoning (RBM) with deep learning techniques, including LSTM, BiLSTM, and GRU. The RBM-BiLSTM model, tested on the JAAD dataset, consistently outperformed baseline approaches across key metrics such as ADE, CMSE, and per-timestep MSE, particularly in complex scenarios. Qualitative analysis further confirmed the hybrid model's ability to better capture multimodal pedestrian behavior and maintain closer alignment with ground truth trajectories. Despite its promising results, the model still faces limitations in real-time adaptability, highlighting the need for future research focused on performance optimization and broader validation across more diverse environments.

CHAPTER VII: CONCLUSION AND FUTUR WORKS

This thesis presented a hybrid pedestrian trajectory prediction framework that integrates interpretable rule-based reasoning with deep learning architectures. The approach addresses the challenge of anticipating pedestrian trajectory by capturing both low-level physical dynamics and high-level behavioral cues. Leveraging features extracted from the JAAD dataset, we demonstrated that combining handcrafted semantic indicators with data-driven modeling yields more robust and behaviorally grounded predictions.

Our model was rigorously evaluated on clean, preprocessed data by comparing three deep learning models (LSTM, GRU, BiLSTM) and their hybrid counterparts incorporating rule-based components. The hybrid RBM-BILSTM model outperformed the pure deep learning BILSTM model, achieving an ADE of 24.36 and CMSE of 437.74 compared to ADE = 28.94 and CMSE = 642.84 for the baseline. These results confirm that incorporating interpretable behavioral features significantly improves prediction accuracy and semantic richness.

Despite these promising outcomes, limitations remain, including the rule-based component's reduced generalization to noisy or uncontrolled conditions, and the reliance on assumptions such as consistent frame rates and accurate pedestrian detection.

To enhance the model's applicability and robustness in real-world scenarios, several future directions are proposed:

- **Deployment in Dynamic Real-World Environments:** Future work aims to adapt the model for efficient operation in dynamic urban settings, handling real-time data streams, complex interactions, and unexpected behaviors to support practical applications in safety and navigation.
- **Multi-Dataset Generalization:** Extending training beyond JAAD to include diverse pedestrian datasets such as PIE, ETH, UCY, and nuScenes to improve generalization across different urban environments and cultural contexts.
- **Architectural Improvements:** While LSTM-based models are effective for sequential data, future versions could leverage modern architectures like Transformers or Graph Neural Networks (GNNs) for more flexible temporal and spatial dependency modeling.
- **Multi-Modal Output Prediction:** Expanding model outputs to jointly predict motion states (e.g., stopping, walking, running), behavioral intentions (e.g., crossing intent, yielding), and social interactions (e.g., influence from nearby agents) to enable deeper understanding of pedestrian behavior beyond trajectory paths.

BIBLIOGRAPHY

Bibliography

[1] 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.

[2] 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

[3] 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.

[4] F. Munir and T. P. Kucner, "Context-aware Multi-task Learning for Pedestrian Intent and Trajectory Prediction," *arXiv Prepr. arXiv2407.17162*, vol. 14, no. 8, pp. 1–11, 2024.

[5] 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.

[6] R. Q. Minguez, I. P. Alonso, D. Fernandez-Llorca, and M. A. Sotelo, "Pedestrian Path, Pose, and Intention Prediction Through Gaussian Process Dynamical Models and Pedestrian Activity Recognition," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 5, pp. 1803–1814, 2019, doi: 10.1109/TITS.2018.2836305.

[7] S. Lefèvre, D. Vasquez, and C. Laugier, "Lefèvre2014_Article_ASurveyOnMotionPredictionAndRi.pdf," pp. 1–14, 2014.

[8] H. Cui *et al.*, "Deep Kinematic Models for Kinematically Feasible Vehicle Trajectory Predictions," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 10563–10569, 2020, doi: 10.1109/ICRA40945.2020.9197560.

[9] M. Lee, S. S. Sohn, S. Moon, S. Yoon, M. Kapadia, and V. Pavlovic, "MUSE-VAE: Multi-Scale VAE for Environment-Aware Long Term Trajectory Prediction," *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, vol. 2022-June, pp. 2211–2220, 2022, doi: 10.1109/CVPR52688.2022.00226.

[10] T. Westny, B. Olofsson, and E. Frisk, "Diffusion-Based Environment-Aware Trajectory Prediction," pp. 1–24, 2024, [Online]. Available: http://arxiv.org/abs/2403.11643

[11] S. A. Bouhsain, S. Saadatnejad, and A. Alahi, "Pedestrian Intention Prediction: A Multi-task Perspective," 2020, [Online]. Available: http://arxiv.org/abs/2010.10270

[12] D. Helbing and P. Molnar, "Social Force Model for Pedestrian Dynamics," no. February, 2013, doi: 10.1103/PhysRevE.51.4282.

[13] M. Moussaïd, N. Perozo, S. Garnier, D. Helbing, and G. Theraulaz, "The walking behaviour of pedestrian social groups and its impact on crowd dynamics," *PLoS One*, vol. 5, no. 4, pp. 1–18, 2010, doi: 10.1371/journal.pone.0010047.

[14] M. Holzmann and S. Moroni, "Itinerant-Electron Magnetism: The Importance of Many-Body Correlations," *Phys. Rev. Lett.*, vol. 124, no. 20, 2020, doi: 10.1103/PhysRevLett.124.206404.

[15] F. Camara *et al.*, "Pedestrian Models for Autonomous Driving Part II: High-Level Models of Human Behavior," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 9, pp. 5453–5472, 2021, doi: 10.1109/TITS.2020.3006767.

[16] R. Korbmacher and A. Tordeux, "Review of Pedestrian Trajectory Prediction Methods: Comparing Deep Learning and Knowledge-Based Approaches," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 12, pp. 24126–24144, 2022, doi: 10.1109/TITS.2022.3205676.

[17] N. A. Madjid *et al.*, "Trajectory Prediction for Autonomous Driving: Progress, Limitations, and Future Directions," 2025, [Online]. Available: http://arxiv.org/abs/2503.03262

[18] R. Tamaru, P. Li, and B. Ran, “Enhancing Pedestrian Trajectory Prediction with Crowd Trip Information,” no. Pei Li, 2024, [Online]. Available: http://arxiv.org/abs/2409.15224

[19] M. Zong, Y. Chang, Y. Dang, and K. Wang, “applied sciences Pedestrian Trajectory Prediction in Crowded Environments Using Social Attention Graph Neural Networks,” 2024.

[20] Y. Wang, W. Wan, H. Zhang, C. Chen, and G. Jia, “Pedestrian Trajectory Intention Prediction in Autonomous Driving Scenarios Based on Spatio-temporal Attention Mechanism,” *2024 4th Int. Conf. Electron. Inf. Eng. Comput. Commun. EIECC 2024*, pp. 1519–1522, 2024, doi: 10.1109/EIECC64539.2024.10929534.

[21] A. S. Abdelrahman, M. Abdel-Aty, and Q. D. Tran, “VRU-CIPI: Crossing Intention Prediction at Intersections for Improving Vulnerable Road Users Safety,” 2025, [Online]. Available: http://arxiv.org/abs/2505.09935

[22] F. G. Landry and M. A. Akhloufi, “Predicting Pedestrian Crossing Intention in Autonomous Vehicles: A Review,” *Neurocomputing*, vol. 618, no. February 2024, p. 129105, 2025, doi: 10.1016/j.neucom.2024.129105.

[23] K. Kim, Y. K. Lee, H. Ahn, S. Hahn, and S. Oh, “Pedestrian intention prediction for autonomous driving using a multiple stakeholder perspective model,” *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 7957–7962, 2020, doi: 10.1109/IROS45743.2020.9341083.

[24] “PeekNet: A Prediction System for Pedestrian Behaviors and Trajectories | by Yu Huang | Medium.” Accessed: May 18, 2025. [Online]. Available: https://yuhuang-63908.medium.com/a-prediction-system-for-pedestrian-behaviors-and-trajectories-362542cd2d9b

[25] D. Zabolotnii, Y. Muhammad, and N. Muhammad, “Pedestrian motion prediction evaluation for urban autonomous driving,” 2024, [Online]. Available: http://arxiv.org/abs/2410.16864

[26] “Back to the Future: Planning-Aware Trajectory Forecasting for Autonomous Driving | SAIL Blog.” Accessed: May 12, 2025. [Online]. Available: https://ai.stanford.edu/blog/trajectory-forecasting/

[27] M. Garzón, D. Garzón-Ramos, A. Barrientos, and J. Del Cerro, “Pedestrian trajectory prediction in large infrastructures: A long-term approach based on path planning,” *ICINCO 2016 - Proc. 13th Int. Conf. Informatics Control. Autom. Robot.*, vol. 2, no. Icinco, pp. 381–389, 2016, doi: 10.5220/0005983303810389.

[28] P. Pasien and H. Studi, “PREDICTING PEDESTRIAN TRAJECTORIES IN ARCHITECTURAL SPACES: A GRAPH NEURAL NETWORK APPROACH,” vol. 4, pp. 2020–2025, 2024.

[29] A. Sevtsuk and R. Kalvo, “Modeling pedestrian activity in cities with urban network analysis,” *Environ. Plan. B Urban Anal. City Sci.*, vol. 52, no. 2, pp. 377–395, 2024, doi: 10.1177/23998083241261766.

[30] A. Sevtsuk, “Estimating Pedestrian Flows on Street Networks: Revisiting the Betweenness Index,” *J. Am. Plan. Assoc.*, vol. 87, no. 4, pp. 512–526, 2021, doi: 10.1080/01944363.2020.1864758.

[31] Z. Pan, “Analysis of Sidewalk Traffic Lights Setting Modes Optimization,” *Commun. Humanit. Res.*, vol. 45, no. 1, pp. 32–39, 2024, doi: 10.54254/2753-7064/45/20240059.

[32] “Pedestrian Environment | VTA.” Accessed: May 12, 2025. [Online]. Available: https://www.vta.org/cdt/street-design-home-page/pedestrian-environment

[33] T. Petković, “Human action and motion prediction in industrial human-robot shared environments using probabilistic decision-making methods,”

2022, [Online]. Available: https://repozitorij.unizg.hr/islandora/object/fer:7735

[34] R. Möller, A. Furnari, S. Battiato, A. Härmä, and G. M. Farinella, "A survey on human-aware robot navigation," *Rob. Auton. Syst.*, vol. 145, 2021, doi: 10.1016/j.robot.2021.103837.

[35] P. Trautman, J. Ma, R. M. Murray, and A. Krause, "Robot navigation in dense human crowds: The case for cooperation," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 2153–2160, 2013, doi: 10.1109/ICRA.2013.6630866.

[36] Y. Hiroi and A. Ito, "A pedestrian avoidance method considering personal space for a guide robot," *Robotics*, vol. 8, no. 4, pp. 1–21, 2019, doi: 10.3390/ROBOTICS8040097.

[37] P. V. K. Borges, N. Conci, and A. Cavallaro, "Video-based human behavior understanding: A survey," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 11, pp. 1993–2008, 2013, doi: 10.1109/TCSVT.2013.2270402.

[38] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," *2009 IEEE Conf. Comput. Vis. Pattern Recognition, CVPR 2009*, no. 1, pp. 935–942, 2009, doi: 10.1109/CVPRW.2009.5206641.

[39] S. Liu, X. Chen, and H. Chen, "CIRAN: extracting crowd interaction with residual attention network for pedestrian trajectory prediction," *Int. J. Mach. Learn. Cybern.*, vol. 13, no. 9, pp. 2649–2662, 2022, doi: 10.1007/s13042-022-01548-0.

[40] H. Wang, W. Zhi, G. Batista, and R. Chandra, "Pedestrian Trajectory Prediction Using Dynamics-based Deep Learning," *Proc. - IEEE Int. Conf. Robot. Autom.*, pp. 15068–15075, 2024, doi: 10.1109/ICRA57147.2024.10609993.

[41] C. Zhou, B. Balle, and J. Pineau, "Learning time series models for pedestrian motion prediction," *Proc. - IEEE Int. Conf. Robot. Autom.*, vol. 2016-June, pp. 3323–3330, 2016, doi: 10.1109/ICRA.2016.7487506.

[42] 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.

[43] C. Yu, X. Ma, J. Ren, H. Zhao, and S. Yi, "Spatio-Temporal Graph Transformer Networks for Pedestrian Trajectory Prediction," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 12357 LNCS, pp. 507–523, 2020, doi: 10.1007/978-3-030-58610-2_30.

[44] Q. Li and H. C. Lau, "A layered hidden Markov model for predicting human trajectories in a multi-floor building," *Proc. - 2015 IEEE/WIC/ACM Int. Jt. Conf. Web Intell. Intell. Agent Technol. WI-IAT 2015*, vol. 2, pp. 344–351, 2016, doi: 10.1109/WI-IAT.2015.239.

[45] J. J. Dabrowski, J. P. de Villiers, A. Rahman, and C. Beyers, "Deep learning and statistical models for time-critical pedestrian behaviour prediction," *Commun. Comput. Inf. Sci.*, vol. 1142 CCIS, pp. 458–465, 2019, doi: 10.1007/978-3-030-36808-1_50.

[46] J. Jiang, K. Yan, X. Xia, and B. Yang, "A Survey of Deep Learning-Based Pedestrian Trajectory Prediction : Challenges and Solutions," 2025.

[47] L. K. Sahoo and V. Varadarajan, "Deep learning for autonomous driving systems: technological innovations, strategic implementations, and business implications-a comprehensive review," *Complex Eng. Syst.*, vol. 5, no. 1, 2025, doi: 10.20517/ces.2024.83.

[48] J. Xie, B. Xia, Z. Xiao, and H. Jiang, "Pedestrian Trajectory Prediction Based on Social Interactions Learning with Random Weights," no. January, 2024, doi: 10.1109/TMM.2024.3368931.

[49] Q. Liu, X. Shi, R. Jiang, H. Zhang, and L. Lu, "Modeling interpretable social interactions for

pedestrian trajectory," *Transp. Res. Part C*, vol. 162, no. June 2022, p. 104617, 2024, doi: 10.1016/j.trc.2024.104617.

[50] L. Zhang, N. Xu, P. Yang, G. J. B, C. Huang, and L. Zhang, "T RAJ PAC : Towards Robustness Verification of Pedestrian Trajectory Prediction Models," pp. 8327–8339.

[51] C. T. Networks and K. Saleh, "Pedestrian Trajectory Prediction for Real-Time Autonomous," 2022.

[52] J. Li *et al.*, "Towards Benchmarking and Assessing the Safety and Robustness of Autonomous Driving on Safety-critical Scenarios," vol. 14, no. 8, pp. 1–16, 2025, [Online]. Available: http://arxiv.org/abs/2503.23708

[53] "Social Force Model – Pedestrian Dynamics." Accessed: May 16, 2025. [Online]. Available: https://pedestriandynamics.org/models/social_force_model/

[54] N. Hussain, C. W. Shiang, S. Loke, and M. A. Bin Khairuddin, "A Multi-Agent Simulation Evacuation Model Using The Social Force Model: A Large Room Simulation Study," *Int. J. Informatics Vis.*, vol. 6, no. 1–2, pp. 221–229, 2022, doi: 10.30630/joiv.6.1-2.929.

[55] 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.

[56] T. Kretz, "On oscillations in the Social Force Model," *Phys. A Stat. Mech. its Appl.*, vol. 438, pp. 272–285, 2015, doi: 10.1016/j.physa.2015.07.002.

[57] Y. Deng *et al.*, "Pedestrian Trajectory Prediction Based on Motion Pattern De-Perturbation Strategy," 2024.

[58] S. Bai, J. Z. Kolter, and V. Koltun, "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling," 2018, [Online]. Available: http://arxiv.org/abs/1803.01271

[59] A. Graser, A. Jalali, J. Lampert, A. Weißenfeld, and K. Janowicz, "MobilityDL: a review of deep learning from trajectory data," *Geoinformatica*, pp. 115–147, 2024, doi: 10.1007/s10707-024-00518-8.

[60] Q. Sun *et al.*, "Large Trajectory Models are Scalable Motion Predictors and Planners," no. i, pp. 1–20, 2023, [Online]. Available: http://arxiv.org/abs/2310.19620

[61] D. Rika Widianita, "TRAJECTORY PREDICTION LEARNING USING DEEP GENERATIVE MODELS," *AT-TAWASSUTH J. Ekon. Islam*, vol. VIII, no. I, pp. 1–19, 2023.

[62] R. Showcase *et al.*, "Planning-based Prediction for Pedestrians Recommended Citation Planning-based Prediction for Pedestrians," *Intell. Robot. …*, 2009, [Online]. Available: http://repository.cmu.edu/robotics/331http://repository.cmu.edu/robotics/331

[63] D. Gonon and A. Billard, "Inverse Reinforcement Learning of Pedestrian-Robot Coordination," *IEEE Robot. Autom. Lett.*, vol. 8, no. 8, pp. 4815–4822, 2023, doi: 10.1109/LRA.2023.3289770.

[64] X. Gao, Y. Wang, Y. Zhao, Y. Li, and G. Wu, "M2Tames: Interaction and Semantic Context Enhanced Pedestrian Trajectory Prediction," *Appl. Sci.*, vol. 14, no. 18, 2024, doi: 10.3390/app14188497.

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

[66] 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.

[67] 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.

[68] 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.

[69] 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.

[70] 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.

[71] X. Dan, "Spatial-Temporal Block and LSTM Network for Pedestrian Trajectories Prediction," 2020, [Online]. Available: http://arxiv.org/abs/2009.10468

[72] 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.

[73] J. Lorenzo, I. Parra, and M. A. Sotelo, "IntFormer: Predicting pedestrian intention with the aid of the Transformer architecture," 2021, [Online]. Available: http://arxiv.org/abs/2105.08647

[74] 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.

[75] R. Rozenberg, J. Gesnouin, and F. Moutarde, "Asymmetrical Bi-RNN for pedestrian trajectory encoding," 2021, [Online]. Available: http://arxiv.org/abs/2106.04419

[76] 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.

[77] 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.

[78] 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

[79] 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.

[80] 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.

[81] 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.

[82] 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.

[83] J. A. Abbasi, N. M. Imran, L. C. Das, and M. Won, "WatchPed: Pedestrian Crossing Intention Prediction Using Embedded Sensors of Smartwatch," *IEEE Int. Conf. Intell. Robot. Syst.*, no. 3, pp. 9574–9581, 2023, doi: 10.1109/IROS55552.2023.10341607.

[84] 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.

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

[86] 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.

[87] W. Chen, H. Sang, J. Wang, and Z. Zhao, "STIGCN: spatial–temporal interaction-aware graph convolution network for pedestrian trajectory prediction," *J. Supercomput.*, vol. 80, no. 8, pp. 10695–10719, 2024, doi: 10.1007/s11227-023-05850-8.

[88] Z. Li, J. Gong, Z. Zhang, C. Lu, V. L. Knoop, and M. Wang, "Interactive Behavior Modeling for Vulnerable Road Users with Risk-Taking Styles in Urban Scenarios: A Heterogeneous Graph Learning Approach," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 8, pp. 8538–8555, 2024, doi: 10.1109/TITS.2024.3399481.

[89] X. Liang and E. W. M. Lee, "Visual-Information-Driven Model for Crowd Simulation Using Temporal Convolutional Network," *IEEE Trans. Intell. Transp. Syst.*, vol. 25, no. 9, pp. 12297–12314, 2024, doi: 10.1109/TITS.2024.3375528.

[90] Y. Ling, Z. Ma, Q. Zhang, B. Xie, and X. Weng, "PedAST-GCN: Fast Pedestrian Crossing Intention Prediction Using Spatial-Temporal Attention Graph Convolution Networks," *IEEE Trans. Intell. Transp. Syst.*, vol. PP, pp. 1–14, 2024, doi: 10.1109/TITS.2024.3398252.

[91] J. Hu *et al.*, "A survey of decision-making and planning methods for self-driving vehicles," *Front. Neurorobot.*, vol. 19, 2025, doi: 10.3389/fnbot.2025.1451923.

[92] S. Kumar, Y. Gu, J. Hoang, G. C. Haynes, and M. Marchetti-Bowick, "Interaction-Based Trajectory Prediction over a Hybrid Traffic Graph," *IEEE Int. Conf. Intell. Robot. Syst.*, pp. 5530–5535, 2021, doi: 10.1109/IROS51168.2021.9636143.

[93] M. Baden, A. Abouelazm, C. Hubschneider, Y. Wu, D. Slieter, and J. M. Zöllner, "TPK: Trustworthy Trajectory Prediction Integrating Prior Knowledge For Interpretability and Kinematic Feasibility," 2025, [Online]. Available: http://arxiv.org/abs/2505.06743

[94] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-Feasible Trajectory Forecasting with Heterogeneous Data," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Insa. T., Ivanovic, B., Chakravarty, P., Pavone, M. (2020). Trajectron++ Dyn. Trajectory Forecast. with Heterog. Data. Lect. Notes Comp*, vol. 12363 LNCS, pp. 683–700, 2020, doi: 10.1007/978-3-030-58523-5_40.

[95] J. Yue, D. Manocha, and H. Wang, "Human Trajectory Prediction via Neural Social Physics," vol. 1, pp. 1–26.

[96] T. Wang, Y. Fu, X. Cheng, L. Li, Z. He, and Y. Xiao, "Vehicle Trajectory Prediction Algorithm Based on Hybrid Prediction Model with Multiple Influencing Factors," *Sensors*, vol. 25, no. 4, pp.

1–15, 2025, doi: 10.3390/s25041024.

[97] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, “Learning social etiquette: Human trajectory understanding in crowded scenes,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 9912 LNCS, pp. 549–565, 2016, doi: 10.1007/978-3-319-46484-8_33.

[98] K. Goel and A. Robicquet, “Neigborhood LSTM : Human trajectory Prediction in Crowded Spaces .,” 2016, [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2016/html/Alahi_Social_LSTM_Human_CVPR_2016_paper.html

[99] 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.

[100] “PIE dataset.” Accessed: May 17, 2025. [Online]. Available: https://data.nvision2.eecs.yorku.ca/PIE_dataset/

[101] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, “Joint Attention in Autonomous Driving (JAAD),” no. September, 2016, doi: 10.48550/arXiv.1609.04741.

[102] “JAAD dataset.” Accessed: May 17, 2025. [Online]. Available: https://data.nvision2.eecs.yorku.ca/JAAD_dataset/

[103] D. O. Pop, “Detection of Pedestrian Actions Based on Deep Learning Approach,” *Stud. Univ. Babeș-Bolyai Inform.*, vol. 64, no. 2, pp. 5–13, 2019, doi: 10.24193/subbi.2019.2.01.

[104] Y. Yuan and K. Kitani, “DLow: Diversifying Latent Flows for Diverse Human Motion Prediction,” pp. 1–25.

[105] D. Meng, G. Zhao, and F. Yan, “Social-STGMLP: A Social Spatio-Temporal Graph Multi-Layer Perceptron for Pedestrian Trajectory Prediction,” *Inf.*, vol. 15, no. 6, pp. 1–15, 2024, doi: 10.3390/info15060341.

[106] M. Hosseinzadeh, B. Sinopoli, and A. F. Bobick, “Toward Safe and Efficient Human – Robot Interaction via Behavior-Driven Danger Signaling”.

[107] C. V Sep, “SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints”.

[108] B. Ivanovic, “The Trajectron : Probabilistic Multi-Agent Trajectory Modeling”.

[109] 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.

[110] I. Kotseruba, A. Rasouli, and J. K. Tsotsos, “Benchmark for evaluating pedestrian action prediction,” *Proc. - 2021 IEEE Winter Conf. Appl. Comput. Vision, WACV 2021*, no. January, pp. 1257–1267, 2021, doi: 10.1109/WACV48630.2021.00130.

[111] R. Trabelsi, R. Khemmar, B. Decoux, J. Y. Ertaud, and R. Butteau, “Recent Advances in Vision-Based On-Road Behaviors Understanding: A Critical Survey,” *Sensors*, vol. 22, no. 7, pp. 1–27, 2022, doi: 10.3390/s22072654.

[112] A. Rasouli and J. K. Tsotsos, “arXiv : 2005 . 06582v1 [cs . CV] 13 May 2020 Pedestrian Action Anticipation using Contextual Feature Fusion in Stacked RNNs,” pp. 1–13.

[113] M. Vision, D. Monekosso, P. Remagnino, and S. Velastin, “Crowd analysis : A survey The original publication is available at www.springerlink.com,” no. May 2014, 2008, doi: 10.1007/s00138-008-0132-4.

[114] J. Francisco, P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, “Context-Based Pedestrian Path Prediction,” pp. 618–633, 2014.

[115] A. Rasouli and J. K. Tsotsos, “Joint Attention in Driver-Pedestrian Interaction : from Theory to

Practice," pp. 1–111, 2018.

[116] 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.

[117] K. Chen *et al.*, "WOMD-LiDAR: Raw Sensor Dataset Benchmark for Motion Forecasting".

[118] G. Antonini, "a discrete choice modeling framework for pedestrian walking behavior with application to human tracking in video sequences," vol. 3382, no. 2005, 2006.

[119] S. E. E. Profile and S. E. E. Profile, "Activity-Driven Populace: A Cognitive Approach to Crowd Simulation," no. February, 2014, doi: 10.1109/MCG.2009.58.

[120] M. Moussaïd, D. Helbing, and G. Theraulaz, "How simple rules determine pedestrian behavior and crowd disasters," pp. 1–16.

[121] "Agent Based Modeling and Simulation : An Informatics Perspective Agent Based Modeling and Simulation : An Informatics Perspective," no. October, 2009.

[122] S. Bandini, L. Crociani, A. Gorrini, and G. Vizzari, "An agent-based model of pedestrian dynamics considering groups : A real world case study An Agent-Based Model of Pedestrian Dynamics Considering Groups : A Real World Case Study," no. November, 2014, doi: 10.1109/ITSC.2014.6957751.

[123] A. Schadschneider, W. Klingsch, H. Klüpfel, T. Kretz, C. Rogsch, and A. Seyfried, "Evacuation Dynamics: Empirical Results, Modeling and Applications," in *Extreme Environmental Events*, 2011, pp. 517–550. doi: 10.1007/978-1-4419-7695-6_29.

[124] B. D. Ziebart, A. Maas, J. A. Bagnell, and A. K. Dey, "Maximum Entropy Inverse Reinforcement Learning," pp. 1433–1438, 2008.

[125] D. Helbing and A. Johansson, "Pedestrian, Crowd and Evacuation Dynamics," no. April 2010, 2014, doi: 10.1007/978-1-4419-7695-6.

[126] E. Sugawara and H. Nikaido, "STOCHASTIC TRANSPORT IN COMPLEX SYSTEMS," *Antimicrob. Agents Chemother.*, vol. 58, no. 12, pp. 7250–7257, 2014, doi: 10.1128/AAC.03728-14.

[127] I. Karamouzas, B. Skinner, and S. J. Guy, "Universal Power Law Governing Pedestrian Interactions," no. December 2014, 2015, doi: 10.1103/PhysRevLett.113.238701.

[128] A. R. Bagabaldo, "I MPROVING P EDESTRIAN S AFETY AT I NTERSECTIONS U SING P ROBABILISTIC M ODELS AND M ONTE C ARLO S IMULATIONS ∗," no. Idm, 2022.

[129] H. J. Park, D. Ph, and S. Li, "PEDESTRIAN CROSSING BEHAVIOR AND COMPLIANCE AT SIGNALIZED INTERSECTIONS," vol. 1, no. 12, pp. 1–12.

[130] K. Teknomo, Y. Takeyama, and H. Inamura, "Microscopic Pedestrian Simulation Model to Evaluate 'Lane-Like Segregation' of Pedestrian Crossing," pp. 1–4, 2016, [Online]. Available: http://arxiv.org/abs/1610.00749

[131] V. Karasev and S. Soatto, "Intent-Aware Long-Term Prediction of Pedestrian Motion".

[132] A. C. Berg, L. E. Ortiz, and T. L. Berg, "Who are you with and Where are you going ?".

[133] T. Zhao *et al.*, "Multi-Agent Tensor Fusion for Contextual Trajectory Prediction".

[134] M. Jerk, T. Generation, and M. Analysis, "Minimum Jerk Trajectory Generation for Straight and Curved Movements :," pp. 187–201.

[135] E. Polytechnique and F. De Lausanne, "Crowd-Robot Interaction: Crowd-aware Robot Navigation with Attention-based Deep Reinforcement Learning".

[136] Y. Huang, Y. Cheng, and K. Wang, "Learning Velocity and Acceleration: Self-Supervised

Motion Consistency for Pedestrian Trajectory Prediction”.

[137] C. Dias, M. Abdullah, M. Sarvi, R. Lovreglio, and W. Alhajyaseen, “Modeling and simulation of pedestrian movement planning around corners,” *Sustain.*, vol. 11, no. 19, pp. 1–15, 2019, doi: 10.3390/su11195501.

[138] F. Lygerakis, V. Dave, and E. Rueckert, “Topics in Guided Motion Control of Marine Vehicles,” *2024 21st Int. Conf. Ubiquitous Robot. UR 2024*, pp. 490–497, 2024, doi: 10.1109/UR61395.2024.10597462.

[139] G. Zhai, H. Meng, and X. Wang, “A constant speed changing rate and constant turn rate model for maneuvering target tracking,” *Sensors (Switzerland)*, vol. 14, no. 3, pp. 5239–5253, 2014, doi: 10.3390/s140305239.

[140] “Math.atan2() - JavaScript | MDN.” Accessed: May 12, 2025. [Online]. Available: https://developer.mozilla.org/en-US/docs/Web/JavaScript/Reference/Global_Objects/Math/atan2

[141] R. Korbmacher and A. Tordeux, “Toward Better Pedestrian Trajectory Predictions: The Role of Density and Time-to-Collision in Hybrid Deep-Learning Algorithms,” *Sensors*, vol. 24, no. 7, 2024, doi: 10.3390/s24072356.

[142] 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”.

[143] H. L. Rietz, *Mathematical statistics*, vol. 9780883851. 2013. doi: 10.5948/UPO9781614440031.

[144] T. O. Hodson, T. M. Over, and S. S. Foks, “Mean Squared Error, Deconstructed,” *J. Adv. Model. Earth Syst.*, vol. 13, no. 12, pp. 1–10, 2021, doi: 10.1029/2021MS002681.

[145] 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.

ANNEX

Notification of Acceptance of the PRML 2025

Chongqing, China, June 13-16, 2025

http://www.prml.org

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)
https://prml.org/IEEE%20Latex%20Template.zip (latex)

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