2024 IEEE Intelligent Vehicles Symposium (IV)
June 2-5, 2024. Jeju Island, Korea

Knowledge-based explainable pedestrian behavior predictor

Angie Nataly Melo[1] , Luis Felipe Herrera-Quintero[2] , Carlota Salinas[1], Miguel Angel Sotelo[1]

2024 IEEE Intelligent Vehicle Symposium (IV) | 979-8-3503-4881-1/24/$31.00 ©2024 IEEE | DOI: 10.1109/IV55156.2024.10588605

***Abstract*—In the context of autonomous driving, pedestrian behavior prediction is a key component for improving road safety. Presently, many existing prediction models prioritize achieving reliable results, however, they often lack insights into the explainability of each prediction. In this work, we propose a novel approach to pedestrian behavior prediction using knowledge graphs (KG), knowledge graph embeddings (KGE), and a Bayesian Inference process, enabling fully inductive reasoning on KGEs. Our approach aims to consolidate knowledge from annotated datasets through explainable pedestrian features and fuzzy rules, evaluating the importance of these two components within the KG. The entire pipeline has been trained and tested using two datasets: Joint Attention for Autonomous Driving (JAAD) and Pedestrian Situated Intent (PSI). Preliminary results demonstrate the effectiveness of this system in providing explainable clues for pedestrian behavior predictions, even improving results by up to 15% compared to other models. Our approach achieves an F1 score of 0.84 for PSI and 0.82 for JAAD.**

***Index Terms*—Autonomous driving, Explainability, Pedestrian behavior prediction, Knowledge graph, Knowledge graph embeddings, Bayesian inference**

I. Introduction

Despite significant progress, road safety remains a pressing global issue. Predicting the behavior of road users holds immense importance, particularly for autonomous driving and intelligent driving systems. According to the latest World Health Organization report on road safety [30], although road traffic deaths have decreased, road crashes persist as a global health crisis, especially for vulnerable road users (VRUs), constituting 53% of all road traffic fatalities. The report highlights a concerning 23% of fatal accidents involving pedestrians. Moreover, it is crucial to emphasize that pedestrians represent the group most severely affected on European Union roads, accounting for one out of every five fatalities [25].

This data underscores the critical need for advancements in pedestrian crossing action prediction and road safety measures to protect this vulnerable group and reduce accidents on the road. In light of this, many research communities have been developed Machine Learning (ML) models and methods to make more robust prediction systems. For a long time most of these models and methods were viewed as *'black boxes'* because they lacked the ability to explain the reasoning behind their predictions. Consequently, understanding why a machine learning model made a specific prediction was challenging.

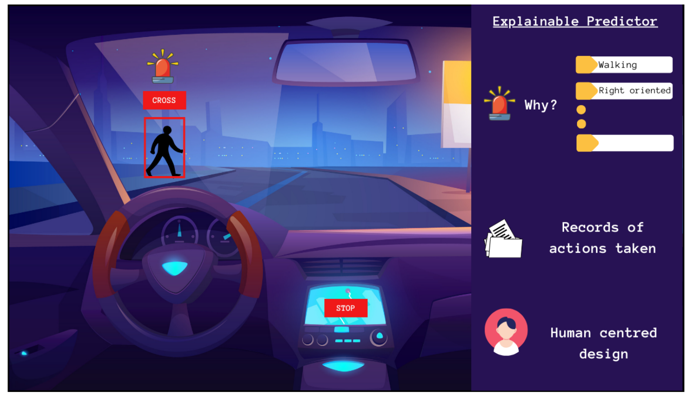

Fig. 1: Relevance of explanations in autonomous driving (Resource based on FreePick image)

Besides, in the context of autonomous driving, the need for explanations arises from psychological, sociotechnical, and philosophical perspective [28]. Understanding why a specific decision was made can have deep implications for enhancing road safety and reducing traffic accidents. Moreover, explaining ego-vehicle decisions contributes to providing descriptive information regarding the causal history of actions taken (See Figure 1).

In this study, we aim to address the need to understand pedestrian behavior and improve prediction systems by incorporating visual clues and human knowledge into a knowledge-based approach. Specifically, a novel pipeline architecture is proposed that combines pedestrian explainable features extracted from neural networks, a KG, KGE learning, and a novel process of behavior prediction based on Bayesian inference. The contributions of this work are threefold: 1) the introduction of a pedestrian behavior predictor based on a knowledge graph, explainable features, and fuzzy rules, 2) the proposed pipeline allows fully inductive reasoning based on KGEs using Bayesian inference, and 3) two KGE models have been trained, tested, and compared with other models and methods for pedestrian behavior predictions.

The rest of the article is organized as follows: Section II discusses related work, details about the proposed pipeline, and the pedestrian behavior ontology are introduced in Section III. Subsequently, the implementation details and experimental setup are described in Section IV. Section V contains the results. Finally, Section VI concludes the work and provides insights for future research.

1-Department of Computer Engineering, University of Alcalá, Madrid, Spain, [nataly.melo, carlota.salinasmaldo and miguel.sotelo]@uah.es

2- Department of Computer Engineering, University Piloto, Bogotá, Colombia, [luis-herrera]@unipiloto.edu.co

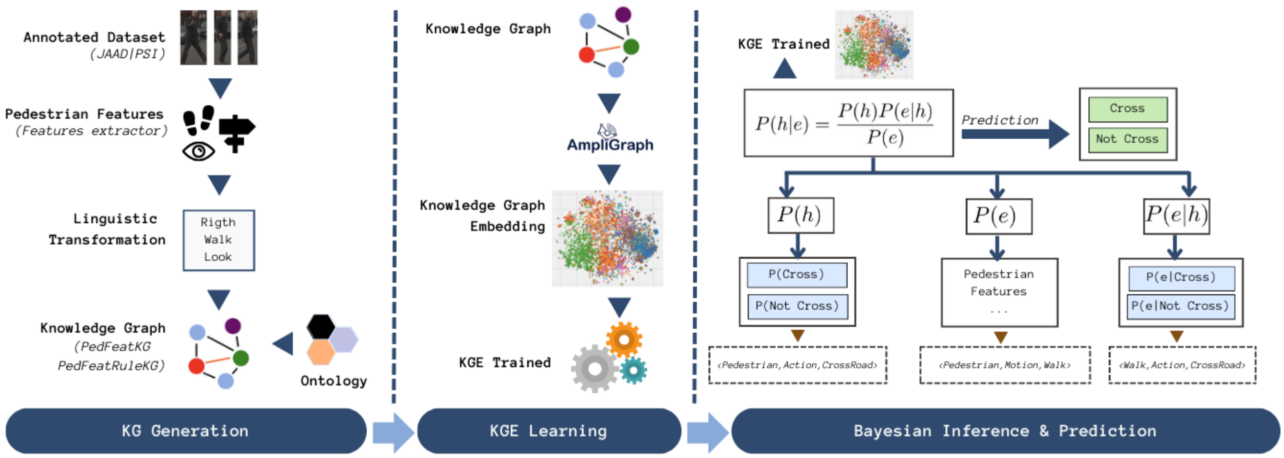

Fig. 2: Pipeline architecture for a knowledge-based explainable pedestrian behavior predictor

II. Related work

A. Pedestrian behavior prediction

Pedestrian behavior prediction, also known as crossing action prediction, is a ML task focused on forecasting if a pedestrian will cross the road at some point in the future. This task has been addressed through a diverse range of algorithms and architectures. Among these approaches, it is particularly noteworthy to highlight some methods as SingleRNN based on recurrent neural networks (RNNs) [14], CapFormer which uses a self-attention alternative based on transformer architecture [20], a 3D Convolutional model (C3D) based on spatiotemporal feature learning [7] and another group of algorithms relies on stacked with multilevel fusion RNN (SFRNN) [15] and convolutional LSTM (ConvLSTM) [6].

Despite the abundance of models and research focused on pedestrian crossing predictions, only a limited number of them provides insights into explainability or are specifically developed within the context of explainability. For instance, the research [27] highlights that Transformers offer an advantage in terms of interpretability, due to their attention mechanism. Moreover, the utilization of pedestrian location and body keypoints as features in predicting pedestrian actions results in more human-like behavior. In [21], the authors propose a dynamic Bayesian network model that takes into account the influence of interaction and social signals. This system leverages visual means and employs various inference methods to provide explanations for its predictions, with a specific focus on determining the relative importance of each feature in influencing the probability of pedestrian actions.

B. Knowledge graph and knowledge graph embedding

A KG is a graph with edges as relations and nodes as entities. It encodes triples that expose real-world facts and semantic relationships [31]. A triple is a fundamental unit in the KG and is composed of three elements: subject, predicate, and object, also known as head, relation, and tail. It is important to highlight that the direction of the relation matters and it can affect the type of KG.

The research on KG encompasses knowledge reasoning, artificial intelligence systems, knowledge acquisition, knowledge graph completion, knowledge fusion, and KGE. Regarding the last topic, the focus is on transforming a KG into a low-dimensional vector that represents entities and relationships, and applying relationship reasoning on the embedding [17]. Then, the obtained vector is used to learn through machine learning models. There are various KGE models utilizing distance-based measures to produce similarity scores for pairs of entities and their relationships. Examples include TransE [2], TransH [4], RotatE [12], and HakE [13]. On the other hand, semantic matching models in KGE focus on similarity-based scoring functions, and notable models in this category include DistMul [8], HolE [5], and ComplEx [9].

Regarding the context of autonomous vehicles, it is important to highlight that the KGE learning has been started to be implement under the entity prediction in driving scenes [22] and for situation comprehension in driving scenarios [18].

III. Knowledge-based explainable predictor

Our approach for developing an explainable pedestrian behavior predictor based on knowledge comprises a pipeline architecture, as illustrated in the figure 2 . The pipeline consists of three primary phases: 1) KG generation, 2) KGE learning and 3) Bayesian inference and prediction. This section provides a detailed explanation of these phases, starting with the generation of the knowledge graph, followed by the process of learning KGEs, and concluding with a formal description of the Bayesian inference and prediction process.

A. Knowledge graph generation

The first phase focuses on generating a knowledge graph that encapsulates data related to pedestrian behavior and crossing intentions on the road. This phase involves annotated datasets that include information about pedestrian behavior. In our study we utilized two datasets: JAAD [10] and PSI [16].

Utilizing the videos, images, and annotations from the mentioned datasets, a set of explainable features is extracted by deep learning and neural networks, as detailed in [29]. In this work, we carefully selected five features for extraction:

- **Gaze:** describes the attention of the pedestrian, indicating whether the pedestrian is looking at the ego-vehicle.
- **Body Orientation:** describes the pedestrian posture through an angle from 0 to 360º.
- **Action:** describes the motion state of the pedestrian, classifying between the following actions: stand, walk, wave, run or undefined (used when pedestrian action is not clear).
- **Proximity to the road:** describes if the pedestrian is near to the road. This feature is classified in three levels according the pedestrian closeness to the road: near, medium distance or far.
- **Distance:** represents the estimated distance to the ego-vehicle.

The mentioned features are extracted from the training set of each dataset and they are extracted for each annotated pedestrian each 2 frames and for a defined number of frames, according to the following rules: 1) Don´t consider more than 60 frames after pedestrian cross and 2) Don´t consider more than 90 frames when the pedestrian will not cross.

Once the explainable features are extracted these are taken as an input to transform from numerical values to linguistic values (See Table I). The transformation takes as inspiration the membership functions defined in [29] for neuro-symbolic approach based on fuzzy logic.

The subsequent step in this phase involves the utilization of the knowledge graph ontology. In this study, we use two ontology versions, one that includes only the pedestrian explainable features and another that additionally incorporates fuzzy rules aiming to explain pedestrian behavior (See Section III-B).

Based on these KG ontologies and the pedestrian features extracted from the dataset, transformed into linguistic values, the KG is generated using the Ampligraph 2.0.0 library [11]. The KG is formed in the shape of triples and a group of triples represents the pedestrian state in a frame. It is important to highlight that we applied some reifications on the given inputs to get reified paths, specifically on the KG which include the fuzzy rules.

B. *Pedestrian behavior ontology*

In this study, two distinct KG ontologies are used, each with differing levels of informational detail. The two versions include 1) a base KG containing only the explainable pedestrian features (PedFeatKG) and 2) a KG that encompasses both the explainable pedestrian features and the fuzzy rules for pedestrian behavior (PedFeatRulesKG).

1) KG ontology from explainable features: With the intention of generating a KG applicable to the pedestrian prediction task, the PedFeatKG ontology includes the entity **Pedestrian** as a generalization entity, which is related to all pedestrians existing in the training set of the mentioned dataset. Into the KG, each pedestrian is identify by an ID which is composed of the defined ID in the dataset and the frame number. For example, if there is a pedestrian with the ID '0-44-202b', there will be as many entities as frames considered in the following structure: '0-44-202b-30', '0-44-202b-32', '0-44-202b-34', and so on. All of this pedestrian id's entities are linked to the Pedestrian entity and this relation is considered a path reification, allowing any specific pedestrian to be linked to a general one.

On the other hand, the PedFeatKG ontology takes into consideration the five pedestrian features mentioned in section III-A and based on it, each annotated pedestrian state and behavior is represented in the following triple form <*pedestrian-ID*, feature relation, *linguistic-value* >. For each feature were defined a group of possible linguistic values as described in the Table I.

TABLE I: Linguistic values for pedestrian features

Feature	Relation KG	Linguistic Values
Action	Motion	Stand, Walk, Wave, Run, Na
Proximity	Location	NearFromCurb, MiddleDisFromCurb FarFromCurb
Distance	EgoDistance	TooNearToEgoVeh, NearToEgoVeh MiddleDisToEgoVeh, FarToEgoVeh TooFarToEgoVeh
Orientation	Orientation	VehDirection, LeftDirection OppositeVehDirection, RigthDirection
Gaze	Attention	Looking, NotLooking

Finally, into this KG ontology the final action of the pedestrian is included as two possibles entities **crossRoad** or **noCrossRoad**, and it is linked to the pedestrian's ID. The Figure 3 shows an example of a generated KG from one instance using the KG ontology from only the explainable features.

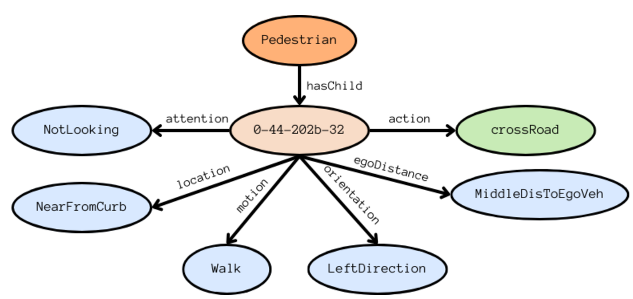

Fig. 3: KG from explainable features with 1 instance

2) KG ontology from explainable features and fuzzy rules: As an extension of the PedFeatKG, the PedFeatRulesKG utilizes a set of fuzzy rules extracted from the JAAD and PSI datasets (detailed in [29]). It is important to highlight that the data used for extracting fuzzy rules maintains a balance between the number of pedestrians who cross the street and those who do not. This balanced representation ensures that the extracted fuzzy rules are equally informed by both scenarios. The rule mining process employed the IVTURS-FARC algorithm [3] and the resulting fuzzy rules take the following form:

$$\begin{aligned} &Rule\ R_j : If\ x_1\ is\ A_{j1}\ and ... and\ x_n\ is\ A_{jn} \\ &then\ Class\ = C_j\ with\ RW_j \end{aligned} \quad (1)$$

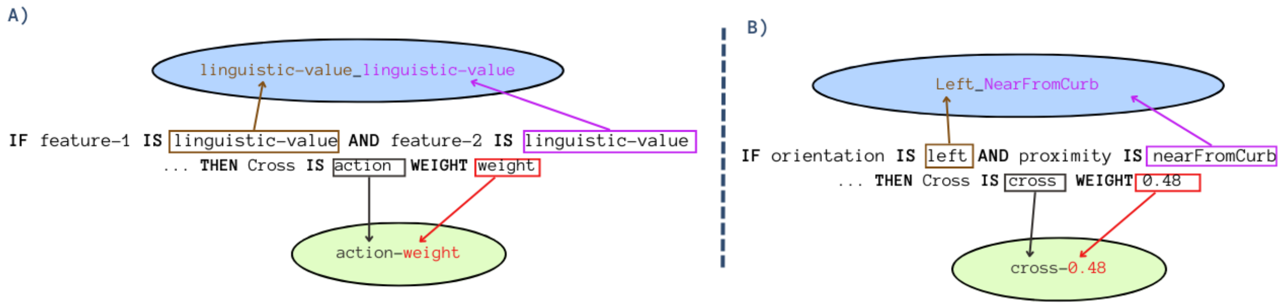

Fig. 4: A) Fuzzy rule conversion definition, B) Example from fuzzy rule to KG entities

where R_j is the label of the jth rule, $x = (x_1, ..., x_n)$ is an n-dimensional pattern vector (pedestrian features in our work), A_{ji} is an antecedent fuzzy set representing a linguistic term, C_j is the class label and RW_j is the rule weight [1].

In the case of the PSI dataset, 60 rules were generated, while in the JAAD dataset, 51 rules were created. To include these rules in the KG, a transformation process was employed, converting the fuzzy rule format into entities suitable for KG inclusion. This process, illustrated in the figure 4, involves combining all linguistic values from the antecedent part of the rule to create one entity and combining the action and fuzzy rule weight to create another entity. Therefore, for each rule is created two new entities which can be later used when the KG is generated.

In fact, the new entities are integrated into the PedFeatKG based on the pedestrian state. That means that in the process of generating the KG, the pedestrian in addition to the explainable features state is linked with the rules which apply to its state. This involves searching for rules that include the current pedestrian features and then incorporating them into the KG to generate the PedFeatRulesKG.

Likewise, the fuzzy rule weight entity is also linked with the **'crossRoad'** or **'noCrossRoad'** entity. The figure 5 illustrates an example of a generated KG instance with the explainable features and fuzzy rules.

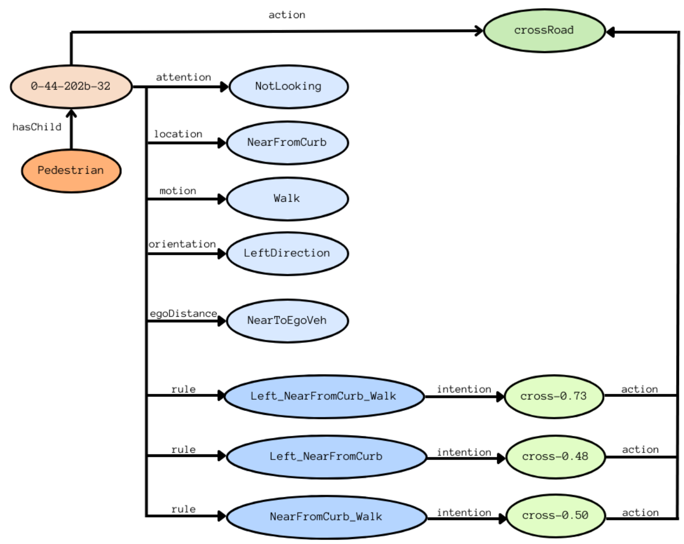

Fig. 5: KG from explainable features and rules with 1 instance

C. Knowledge graph embedding learning

In the second phase, we use Ampligraph 2.0.0 to generate a KGE model from the KG created in the previous phase. Due the type of relations from the KG, it was used the model *ComplEx*. Then, this KGE model is train, validate and test using the Ampligraph learning features.

D. Bayesian inference and prediction

Once the KGE model is trained and the embeddings are generated, we can infer the probability of a specific triple from the KG using the evaluation method provided by Ampligraph. However, considering that our proposal aims to enable inductive reasoning and predict the behavior of various pedestrians not included in the KG, a different process is required for prediction.

Therefore, we implemented path reifications (detailed in Section III-B), enabling inference about a pedestrian not present in the KG. Additionally, we employed Bayesian inference on the learned embeddings. Specifically, we utilized Bayes' rule to predict whether the pedestrian will cross or not, as shown in the following equation:

$$P(h|e) = \frac{P(h)P(e|h)}{P(e)} \tag{2}$$

where hypothesis h is the event or entity that we want to predict (pedestrian behavior), and evidence e is the information regarding the pedestrian and which was extracted from the video for the current frame. It is important to highlight that the computation of $P(h|e)$ is done for both Cross and Not Cross behavior, and the prediction is determined by the higher probability between both computations. For detailing how the inference and prediction work, we introduce the following example for the cross action:

- **Hypothesis:** pedestrian behavior action is cross,
- **Evidence:** the pedestrian is near to the road, he is walking, he is not looking the car, his orientation is left and the car is in a moderate distance from the pedestrian.

The computation of each element of the equation takes place by evaluating a single triple using the KGE evaluation method provided from Ampligraph. For instance, taking into account the example, the computation of $P(h)$ implies the evaluation of the triple <*Pedestrian*, action, *CrossRoad*>.

The computation of $P(e)$ takes into account the number of evidences extracted from the pedestrian and follows the next form:

$$P(e) = P(e_1) * P(e_2) * P(e_3)... * P(e_n) \quad (3)$$

In the mentioned example, the triples composing $P(e_n)$ are evaluated for each pedestrian feature, as shown in Table II. The multiplication of these probabilities computes $P(e)$.

TABLE II: Example of computation of the evidence $P(e)$

$P(e)$	Triples
$P(e_1)$	<*Pedestrian*, location, *NearFromCurb*>
$P(e_2)$	<*Pedestrian*, motion, *Walk*>
$P(e_3)$	<*Pedestrian*, attention, *NotLooking*>
$P(e_4)$	<*Pedestrian*, orientation, *LeftDirection*>
$P(e_5)$	<*Pedestrian*, egoDistance, *MiddleDisToEgoVeh*>

On the other hand, the computation of the $P(e|h)$ follow the next equation:

$$P(e) = P(e_1|h) * P(e_2|h) * P(e_3|h)... * P(e_n|h) \quad (4)$$

The computation of $P(e|h)$ in the mentioned example can be expressed by the triples related in the Table III. The evaluation of these triples and its multiplication gives the $P(e|h)$ result.

TABLE III: Example of computation of the $P(e|h)$

$P(e\|h)$	Triples
$P(e_1\|h)$	<*NearFromCurb*, action, *CrossRoad*>
$P(e_2\|h)$	<*Walk*, action, *CrossRoad* >
$P(e_3\|h)$	<*NotLooking*, action, *CrossRoad*>
$P(e_4\|h)$	<*LeftDirection*, action, *CrossRoad*>
$P(e_5\|h)$	<*MiddleDisToEgoVeh*, action, *CrossRoad*>

Finally, $P(h|e)$ can be calculated using Equation 1 given that all these individual probabilities are computable from the KG using the embeddings.

It is important to highlight that in our study, we predict the next 30_{th} frame after 15 frames of observation in the PSI dataset and 30 frames of observation in JAAD.

IV. Implementation and Experimental Setup

The following section presents the details regarding the implementation and configuration of the experiments.

A. Data Sampling

In this work, we split the datasets mentioned previously for two main tasks: 1) training the KGE and 2) testing the performance of the models.

In the PSI dataset, we use the default split list provided, which is composed of 104 videos for training, 34 videos for validation, and 48 videos for testing. On the other hand, in JAAD, we included an additional process that implies the selection of videos that meet specific criteria such as visibility and high quality. As a result, we carefully chose 284 videos from JAAD. From these selected videos, we use 232 videos to extract the fuzzy rules and extract the pedestrian features, which are used to generate the KG. It is important to highlight that the extracted data is balanced between pedestrians who cross the street and those who do not.

The performance of the proposed approach was evaluated using for JAAD all the selected videos $JAAD_{all}$ and for PSI only the test videos PSI_{test}.

B. Implementation details

The proposed predictor implementation can be described in two parts. Regarding the extraction of the pedestrian features from annotated datasets, we develop a modular architecture using Python and Pytorch, which allows the integration of different modules to extract each feature mentioned in Section III-A. These modules use neural networks such as YOLOv7 [26], YOLOPv2 [24], PedRecNet [23], and our own transformer to detect pedestrian activity, all the implementation details are described in [29].

On the other hand, for the KG generation, KGE Learning and Bayesian Inference and prediction, we use Python, TensorFlow and Ampligraph library. To train the proposed approach we use the scoring model ComplEx, the Adam optimizer and the SelfAdversialLoss. Regarding the training parameters we use an embedding size $k = 150$ while the number of corruptions to generate during training varies from 5 to 20, depending on the quantity of triples and the dataset. Likewise we use $learningRate = 0.0005$, $batchSize = 10000$ and an early stopping criteria using the mean reciprocal rank (MRR). All the learnable methods were trained with a CPU AMD Ryzen 5 5600X 6-Core with a GPU NVIDIA GeForce RTX 3080.

We test for the two defined pedestrian behavior ontology's into the JAAD and PSI dataset. In the table IV are described the number of triples which composed each KG according with the ontology and the dataset.

TABLE IV: Number of triples in the experimental setup

Dataset	Ontology	Triples rows
PSI	PedFeatKG	167.356
	PedFeatRulesKG	367.258
JAAD	PedFeatKG	98.000
	PedFeatRulesKG	225.350

To evaluate the performance of the proposed pipeline, we utilized precision, recall, and F1 score metrics. Where the precision is defined as the ratio of correct positive predictions to the total predicted positives. The recall is the ratio of correct positive predictions to the total positives examples and the F1 Score is the harmonic mean of precision and recall.

V. Results and Discussions

The following section presents the preliminary results on pedestrian crossing prediction task.

A. Pedestrian behavior decision

We evaluate the performance of the proposed knowledge-based explainable predictor using both the PedFeatKG and PedFeatRulesKG over the $JAAD_{all}$ and PSI_{test}. We compare the results provided with the neuro-symbolic approach based on fuzzy logic and two 'black box' methods. In JAAD we took as a reference the PCPA model proposed by the York

University [19] while in PSI we develop our own model using transformer encoding blocks and pedestrian features.

In Table V, it can be observed that in both datasets, the knowledge graph approach (PedFeatKG) improves the F1 score of the predictions compared with the *'black box'* methods and fuzzy logic approach, showing a 13% improvement in JAAD and 15% in PSI regarding the *'black box'* methods, while 6% improvement in JAAD and 12% in PSI regarding the fuzzy logic approach. Regarding to the precision and recall metrics, a slight improvement can also be observed in the JAAD dataset.

TABLE V: Comparing the pedestrian behavior predictor with various methods

(a) $JAAD_{all}$

Model	F1	Precision	Recall
PCPA[19]	0.68	-	-
Fuzzy Logic	0.75	0.69	0.81
PedFeatKG	0.81	0.73	**0.89**
PedFeatRulesKG	**0.82**	**0.79**	0.84

(b) PSI_{test}

Model	F1	Precision	Recall
Black Box	0.69	**0.81**	0.60
Fuzzy Logic	0.72	0.74	0.70
PedFeatKG	**0.84**	0.75	**0.95**
PedFeatRulesKG	0.79	0.75	0.95

In the case of JAAD, it is important to highlight that the inclusion of the fuzzy rules into the KG (PedFeatRulesKG) slightly improves the results and reach a F1 score of 0.82. These results not only demonstrate improvement from a numeric perspective, but the knowledge-based approach also offers valuable insights into the motivations and explainability of the provided predictions.

In addition, we conducted an ablation study to understand the impact of the fuzzy rules on the knowledge graph. Based on the rules that were activated more frequently during the evaluation, Figure 6 illustrates that not all rules have the same impact on the prediction. For JAAD, it is evident that certain rules have a more pronounced impact on each prediction, particularly those related to proximity to the road, orientation, and action. In the case of PSI, more rules are observed to be activated, but, similar to JAAD, the prominently activated rules are associated with proximity to the road and pedestrian orientation.

Considering the rules that are more frequently activated, we decided to reduce the number of rules included in the knowledge graph. We limited the rules to 13 for both JAAD and PSI datasets; in the figure 6, these rules are highlighted with a red arrow. Despite this reduction, we proceeded with the proposed pipeline and evaluated the performance of the predictor using only the most relevant rules. Surprisingly, the performance did not improve in either case (refer to Table VI), underscoring the importance of including all the rules extracted from the datasets. Nevertheless, the F1 score in both cases is still better than the results obtained from the fuzzy logic approach. This highlights the effectiveness of the KG and KGE approach within an explainable context.

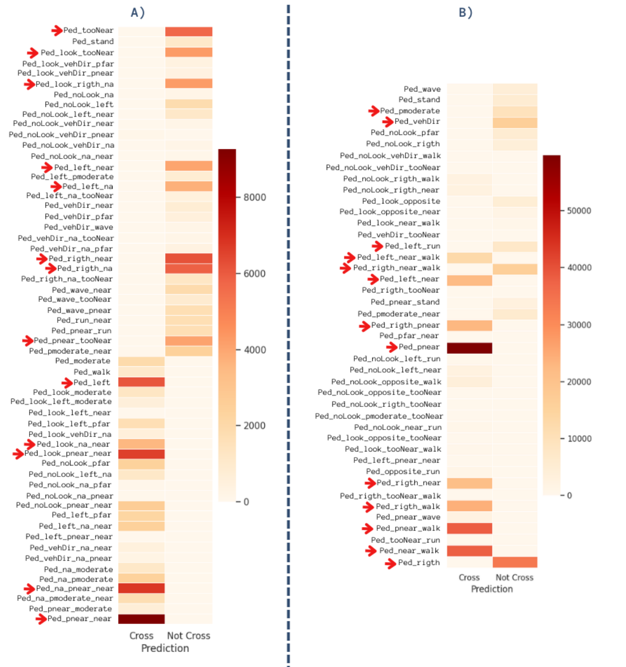

Fig. 6: A) Heat map rule activation in PSI, B) Heat map rule activation in JAAD

TABLE VI: Performance of the model with reduced fuzzy rules

Dataset	F1	Precision	Recall
$JAAD_{all}$	0.79	0.79	0.80
PSI_{test}	0.78	0.79	0.77

B. Decision explainability

The incorporation of KG and KGE into a pedestrian behavior predictor offers clear insights into the explanation of each prediction. In the PedFeatKG model, these insights focus on the explainable pedestrian features. In the PedFeatRulesKG, the rules are also included, providing a possible interpretation of the reasons for the prediction. The figure 7 illustrates an example of a prediction from a JAAD video. The sequence following the pedestrian is detailed, indicating when the predictor detects the future crossing action and associates it with specific features and rules that represent the current state of the pedestrian.

In this example, the predictor anticipates the crossing action 1 second before it happens. Analyzing the extracted explainable features and the activated fuzzy rules at the moment when the action is anticipated can provide interpretability within the context of typical road users. In this case, two fuzzy rules where activated and their meaning are as follows:

- **Ped_pnear** The pedestrian is near to the road.
- **Ped_left_run:** The pedestrian is oriented to the left and is running.

Notably, factors such as proximity to the road and the pedestrian's body orientation are considered valuable inputs in this example. In the following website *https://kg-*

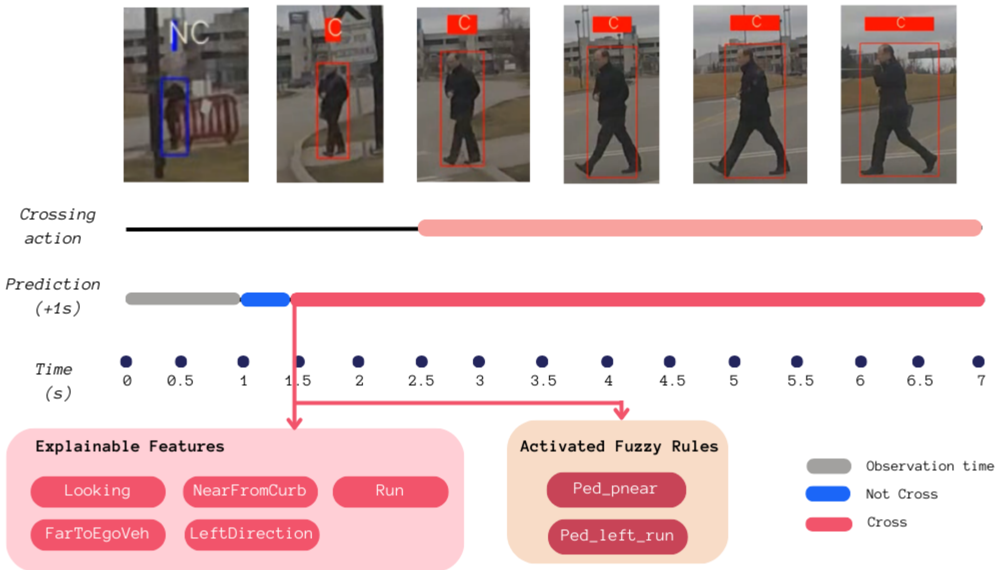

Fig. 7: Example of prediction explainability from JAAD dataset

pedestrian.s3.us-west-2.amazonaws.com/index.html, we provide some examples about the predictions over the JAAD and PSI dataset, as well as some more details about the KG ontologies.

VI. Conclusions

This work proposes a novel approach to predicting pedestrian behavior based on knowledge, utilizing knowledge graphs, knowledge graph embedding learning, and Bayesian inference. It is noteworthy that the use of Bayesian inference allows for fully inductive reasoning, enabling the prediction of pedestrian behavior even when the knowledge graph lacks specific information about the pedestrian in question. The proposed pipeline enhances prediction systems by up to 15%, achieving an F1 score of 0.84 in PSI and 0.82 in JAAD. Additionally, this approach yields preliminary features and fuzzy rules that support the explainability of predictions.

Similarly, in this study, we compare two KG ontologies: one that includes only pedestrian features and another that additionally incorporates fuzzy rules. The results obtained demonstrate that these rules provide valuable information, complementing the importance of explainable pedestrian features. Notably, proximity to the road, pedestrian orientation, and pedestrian action emerge as the most crucial features. However, it's essential to emphasize that other features are also necessary for achieving accurate results in pedestrian behavior predictions.

In future work, we plan to incorporate additional explainable features that capture the pedestrian's story over multiple frames. Furthermore, we aim to enrich the KG by integrating knowledge from both road experts and everyday road users.

Acknowledgment

This research has been funded by the HEIDI project of the European Commission under Grant Agreement: 101069538

References

[1] H. Ishibuchi and T. Nakashima. "Effect of rule weights in fuzzy rule-based classification systems". In: *IEEE Transactions on Fuzzy Systems* 9.4 (2001), pp. 506–515. DOI: 10.1109/91.940964.

[2] Antoine Bordes et al. "Translating Embeddings for Modeling Multi-relational Data". In: *Advances in Neural Information Processing Systems*. Ed. by C.J. Burges et al. Vol. 26. Curran Associates, Inc., 2013. URL: https://proceedings.neurips.cc/paper_files/paper/2013/file/1cecc7a77928ca8133fa24680a88d2f9-Paper.pdf.

[3] J. Sanz et al. "IVTURS: a linguistic fuzzy rule-based classification system based on a new Interval-Valued fuzzy reasoning method with TUning and Rule Selection". In: *IEEE Transactions on Fuzzy Systems* 21.3 (2013), pp. 399–411.

[4] Zhen Wang et al. "Knowledge Graph Embedding by Translating on Hyperplanes". In: *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*. AAAI'14. Québec City, Québec, Canada: AAAI Press, 2014, pp. 1112–1119.

[5] Maximilian Nickel, Lorenzo Rosasco, and Tomaso A. Poggio. "Holographic Embeddings of Knowledge Graphs". In: *CoRR* abs/1510.04935 (2015). arXiv: 1510.04935. URL: http://arxiv.org/abs/1510.04935.

[6] Xingjian SHI et al. "Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting". In: *Advances in Neural Information Processing Systems*. Ed. by C. Cortes et al. Vol. 28. Curran Associates, Inc., 2015. URL: https://proceedings.neurips.cc/paper_files/paper/2015/file/07563a3fe3bbe7e3ba84431ad9d055af-Paper.pdf.

[7] Du Tran et al. *Learning Spatiotemporal Features with 3D Convolutional Networks*. 2015. arXiv: 1412.0767 [cs.CV].

[8] Bishan Yang et al. *Embedding Entities and Relations for Learning and Inference in Knowledge Bases*. 2015. arXiv: 1412.6575 [cs.CL].

[9] Théo Trouillon et al. "Complex Embeddings for Simple Link Prediction". In: *CoRR* abs/1606.06357 (2016). arXiv: 1606.06357. URL: http://arxiv.org/abs/1606.06357.

[10] Amir Rasouli, Iuliia Kotseruba, and John K Tsotsos. "Are they going to cross? A benchmark dataset and baseline for pedestrian crosswalk behavior". In: *Proceedings of the IEEE International Conference on Computer Vision Workshops*. 2017, pp. 206–213.

[11] Luca Costabello et al. *AmpliGraph: a Library for Representation Learning on Knowledge Graphs*. Mar. 2019. DOI: 10.5281/zenodo.2595043. URL: https://doi.org/10.5281/zenodo.2595043.

[12] Zhiqing Sun et al. "RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space". In: *CoRR* abs/1902.10197 (2019). arXiv: 1902.10197. URL: http://arxiv.org/abs/1902.10197.

[13] Zhanqiu Zhang et al. "Learning Hierarchy-Aware Knowledge Graph Embeddings for Link Prediction". In: *CoRR* abs/1911.09419 (2019). arXiv: 1911.09419. URL: http://arxiv.org/abs/1911.09419.

[14] Iuliia Kotseruba, Amir Rasouli, and John K. Tsotsos. "Do They Want to Cross? Understanding Pedestrian Intention for Behavior Prediction". In: *2020 IEEE Intelligent Vehicles Symposium (IV)*. 2020, pp. 1688–1693. DOI: 10.1109/IV47402.2020.9304591.

[15] Amir Rasouli, Iuliia Kotseruba, and John K. Tsotsos. *Pedestrian Action Anticipation using Contextual Feature Fusion in Stacked RNNs*. 2020. arXiv: 2005.06582 [cs.CV].

[16] Tina Chen et al. "Psi: A pedestrian behavior dataset for socially intelligent autonomous car". In: *arXiv preprint arXiv:2112.02604* (2021).

[17] Shivani Choudhary et al. *A Survey of Knowledge Graph Embedding and Their Applications*. 2021. arXiv: 2107.07842 [cs.IR].

[18] Lavdim Halilaj et al. "A Knowledge Graph-Based Approach for Situation Comprehension in Driving Scenarios". In: *The Semantic Web*. Ed. by Ruben Verborgh et al. Cham: Springer International Publishing, 2021, pp. 699–716. ISBN: 978-3-030-77385-4.

[19] Iuliia Kotseruba, Amir Rasouli, and John K. Tsotsos. "Benchmark for Evaluating Pedestrian Action Prediction". In: *2021 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2021, pp. 1257–1267. DOI: 10.1109/WACV48630.2021.00130.

[20] Javier Lorenzo et al. "CAPformer: Pedestrian Crossing Action Prediction Using Transformer". In: *Sensors* 21.17 (2021). ISSN: 1424-8220. DOI: 10.3390/s21175694. URL: https://www.mdpi.com/1424-8220/21/17/5694.

[21] Nora Muscholl et al. "EMIDAS: explainable social interaction-based pedestrian intention detection across street". In: *Proceedings of the 36th Annual ACM Symposium on Applied Computing* (2021).

[22] Ruwan Wickramarachchi, Cory Henson, and Amit Sheth. "Knowledge-infused Learning for Entity Prediction in Driving Scenes". In: *Frontiers in Big Data* 4 (2021). ISSN: 2624-909X. DOI: 10.3389/fdata.2021.759110. URL: https://www.frontiersin.org/articles/10.3389/fdata.2021.759110.

[23] Dennis Burgermeister and Cristóbal Curio. "PedRecNet: Multi-task deep neural network for full 3D human pose and orientation estimation". In: *2022 IEEE Intelligent Vehicles Symposium, IV 2022, Aachen, Germany,June 4-9, 2022*. IEEE, 2022, pp. 441–448. DOI: 10.1109/IV51971.2022.9827202. URL: https://doi.org/10.1109/IV51971.2022.9827202.

[24] Cheng Han et al. "YOLOPv2: Better, Faster, Stronger for Panoptic Driving Perception". In: 2022. arXiv: 2208.11434 [cs.CV].

[25] Freya Slootmans. *European Road Safety Observatory*. Tech. rep. European Commission, 2022.

[26] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors". In: arXiv, 2022. DOI: 10.48550/ARXIV.2207.02696. URL: https://arxiv.org/abs/2207.02696.

[27] Lina Achaji et al. *Analysis over vision-based models for pedestrian action anticipation*. 2023. arXiv: 2305.17451 [cs.CV].

[28] Shahin Atakishiyev et al. "Explainable Artificial Intelligence for Autonomous Driving: A Comprehensive Overview and Field Guide for Future Research Directions". In: (2023). arXiv: 2112.11561 [cs.AI].

[29] Angie Nataly Melo, Carlota Salinas, and Miguel Angel Sotelo. *Experimental Insights Towards Explainable and Interpretable Pedestrian Crossing Prediction*. 2023. arXiv: 2312.02872 [cs.LG].

[30] World Health Organization. *Global status report on road safety 2023*. World Health Organization, 2023, ix, 81 p.

[31] Ciyuan Peng et al. *Knowledge Graphs: Opportunities and Challenges*. 2023. arXiv: 2303.13948 [cs.AI].