INCREASING EXPLAINABILITY IN TIME SERIES CLASSIFICATION BY FUNCTIONAL DECOMPOSITION

Thies de Graaff, Michael Wild, Tino Werner, Eike Möhlmann | German Aerospace Center (DLR)
Stefan Seibt, Benjamin Ebrecht | Technical University Berlin

2024-07-18

Motivation

- Current trends in Machine Learning are relying more and more on end-to-end architectures
  - This might increase the performance but decreases the explainability of the reasoning
- In mobility applications, systems must be certified before being released
  - Difficult due to black-box nature of E2E architectures

Motivation

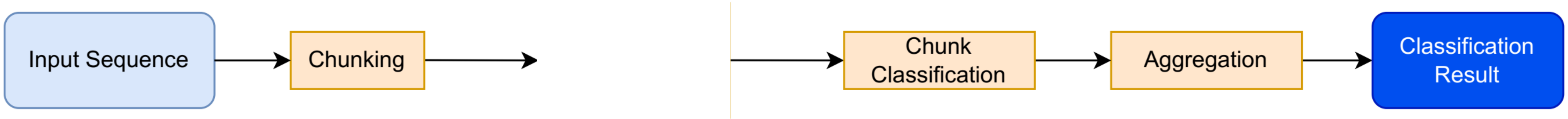
- Current trends in Machine Learning are relying more and more on end-to-end architectures
  - This might increase the performance but decreases the explainability of the reasoning
- In mobility applications, systems must be certified before being released
  - Difficult due to black-box nature of E2E architectures

- In this work, we…
  - revisit some classical approaches in Machine Learning for time series classification,
  - show how functional decomposition naturally introduces more transparency using à-priori knowledge,
  - derive easy explanations to localize the source of errors, and
  - apply these concepts to a use case in the railway domain, showing their effectiveness.

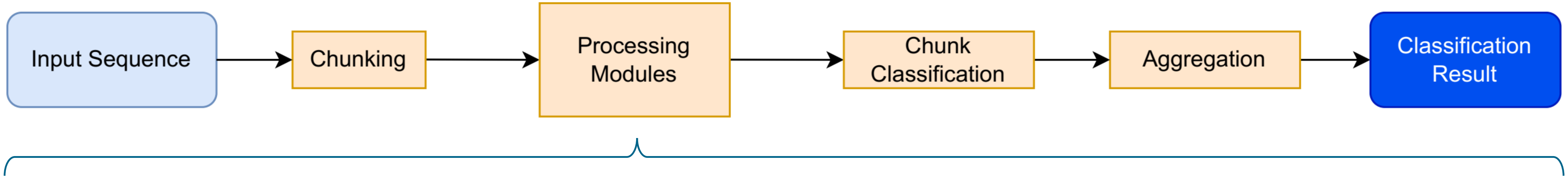
Outline

- Motivation
- General Methodology
- Case Study
- Application
- Evaluation
- Explanations

General Methodology



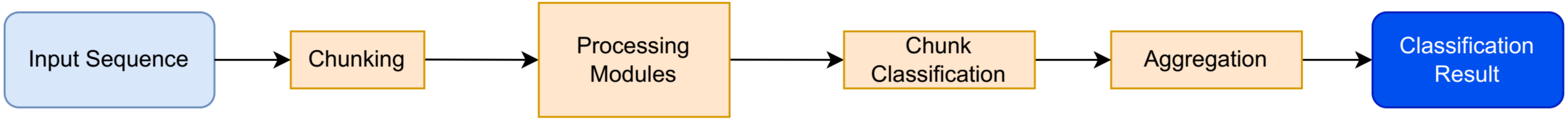
General Methodology



<u>Reducing Decompositions</u>

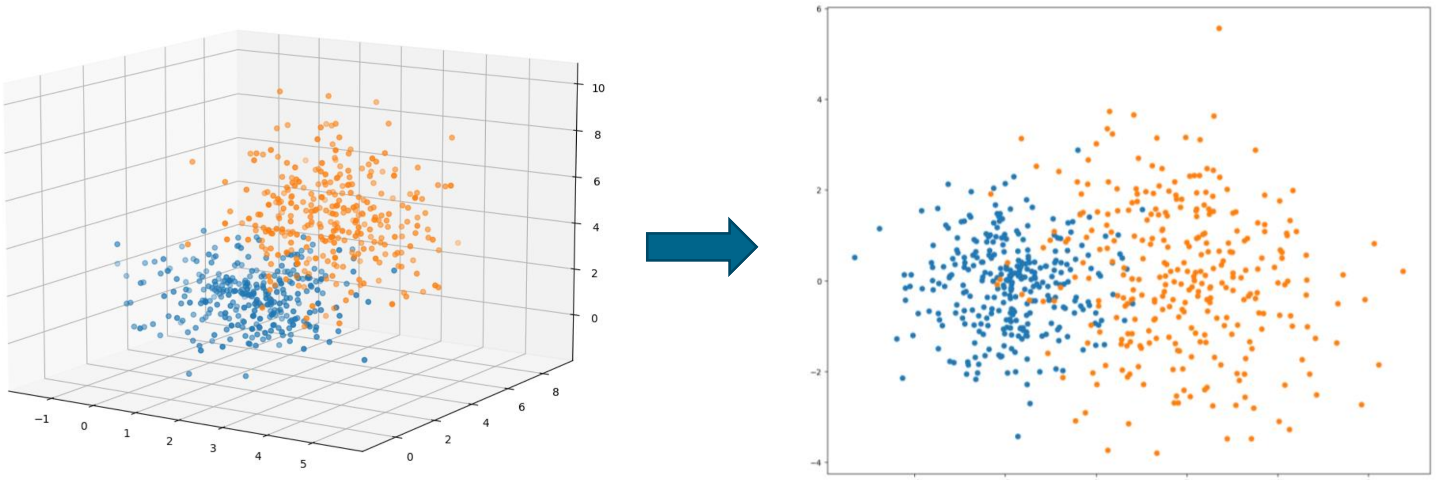
<u>Restructuring Decompositions</u>

General Methodology

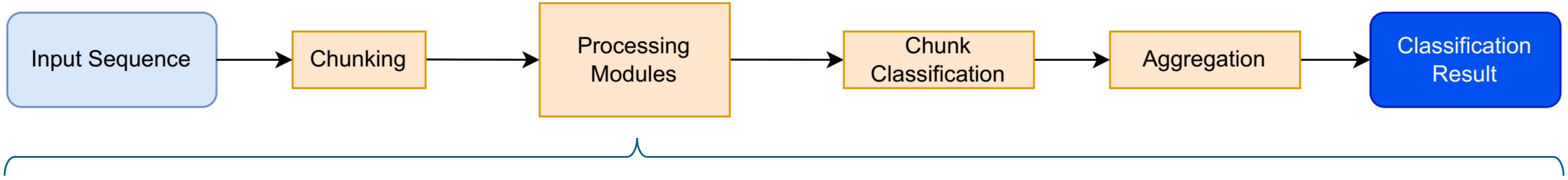


Reducing Decompositions

Restructuring Decompositions

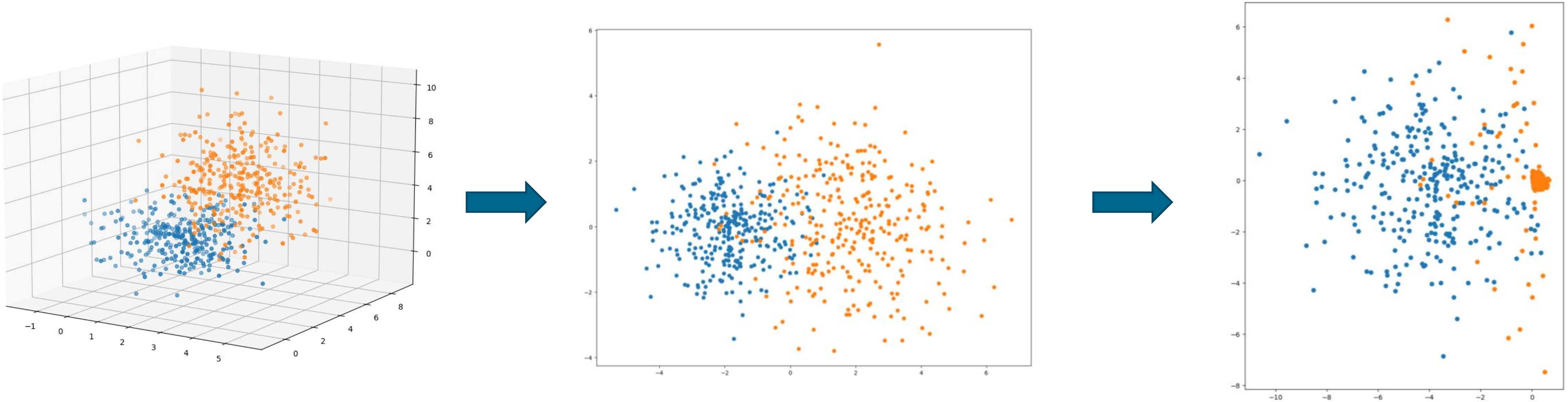


General Methodology



Reducing Decompositions

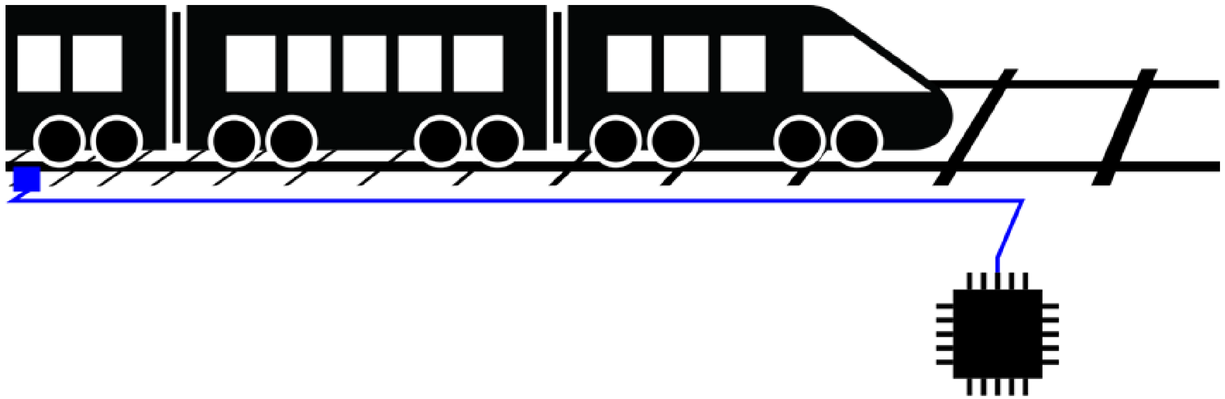
Restructuring Decompositions



Case Study

Railway domain

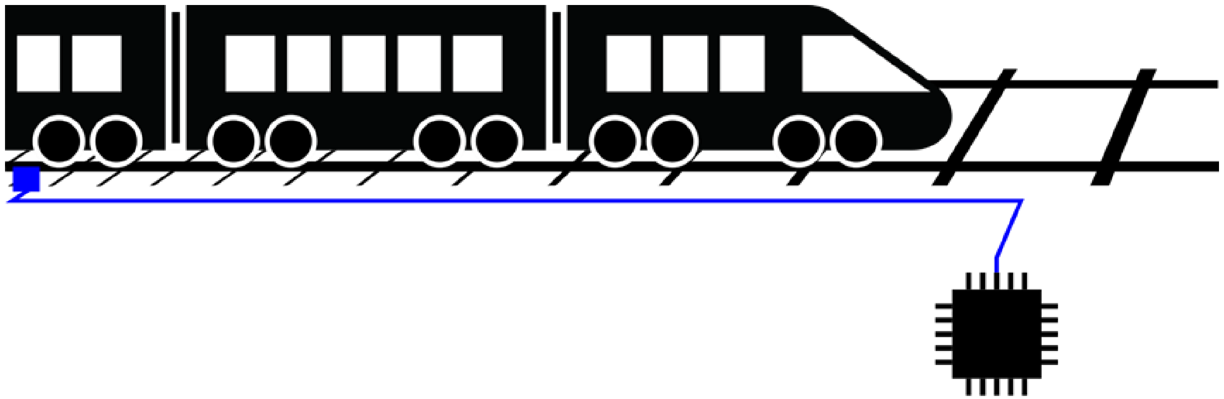
- Detection of train types based on axle counters



Case Study

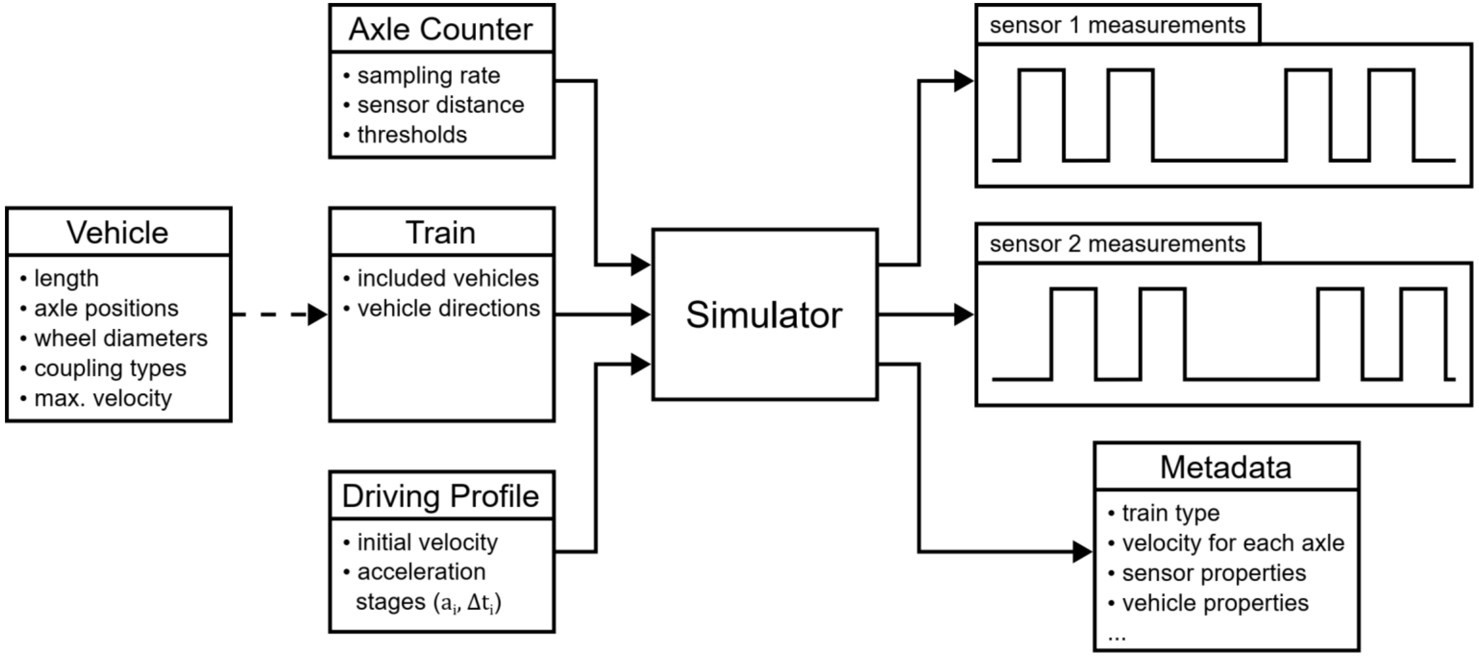
Railway domain

- Detection of train types based on axle counters

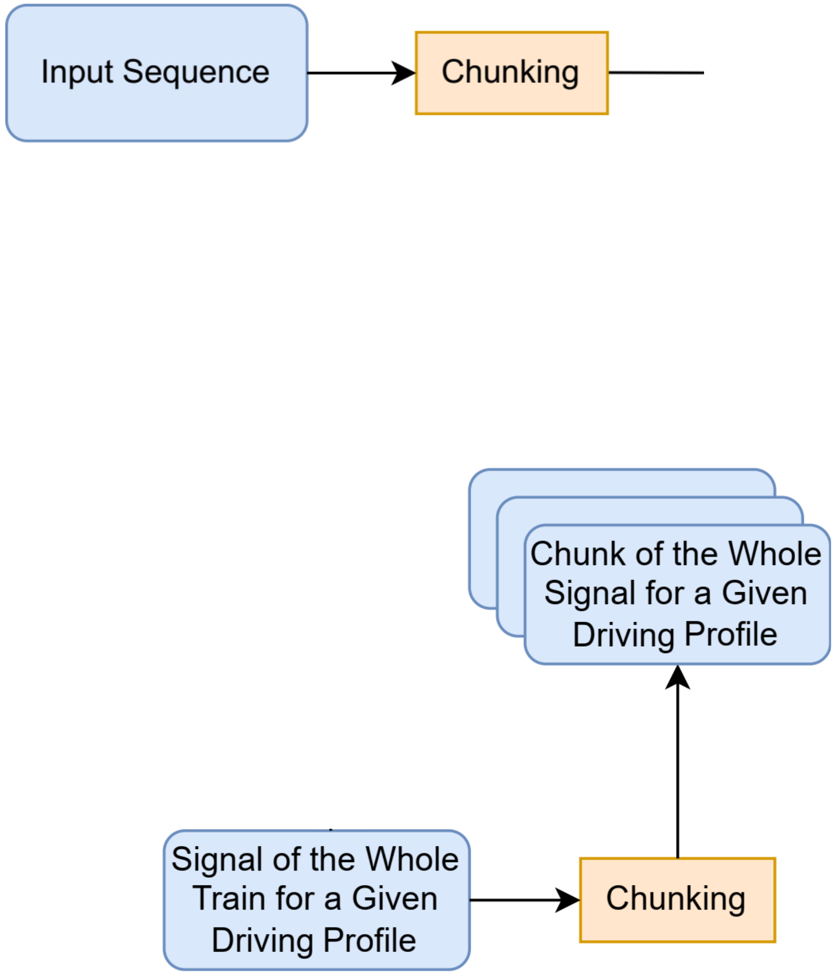


Collecting real data is difficult

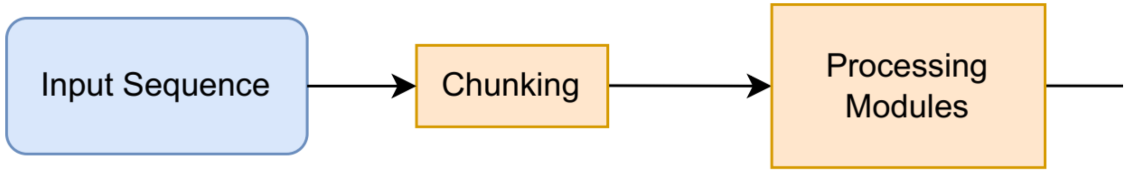
- Implementation of simulator
- Development of vehicle and train catalogues
- Enabling generation of fast and diverse datasets

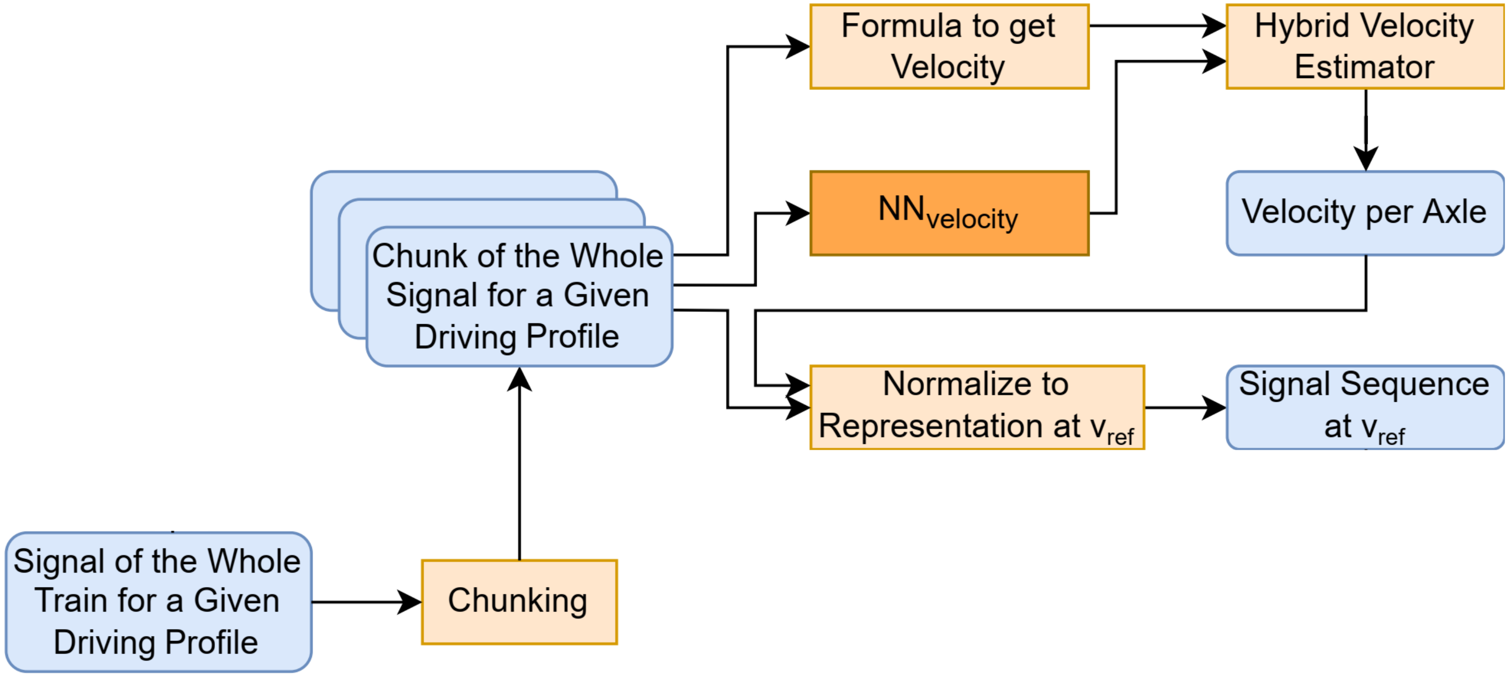


Application

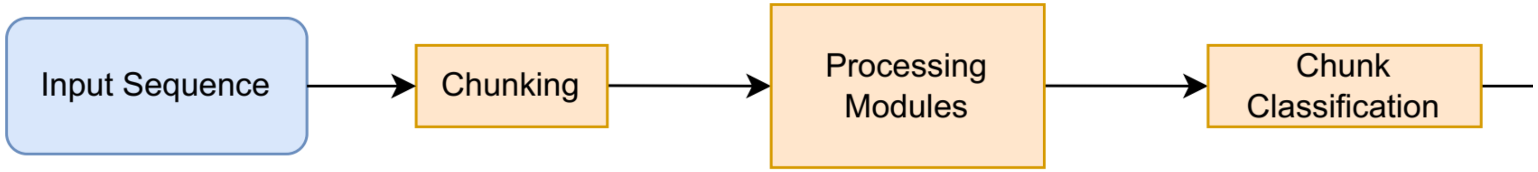


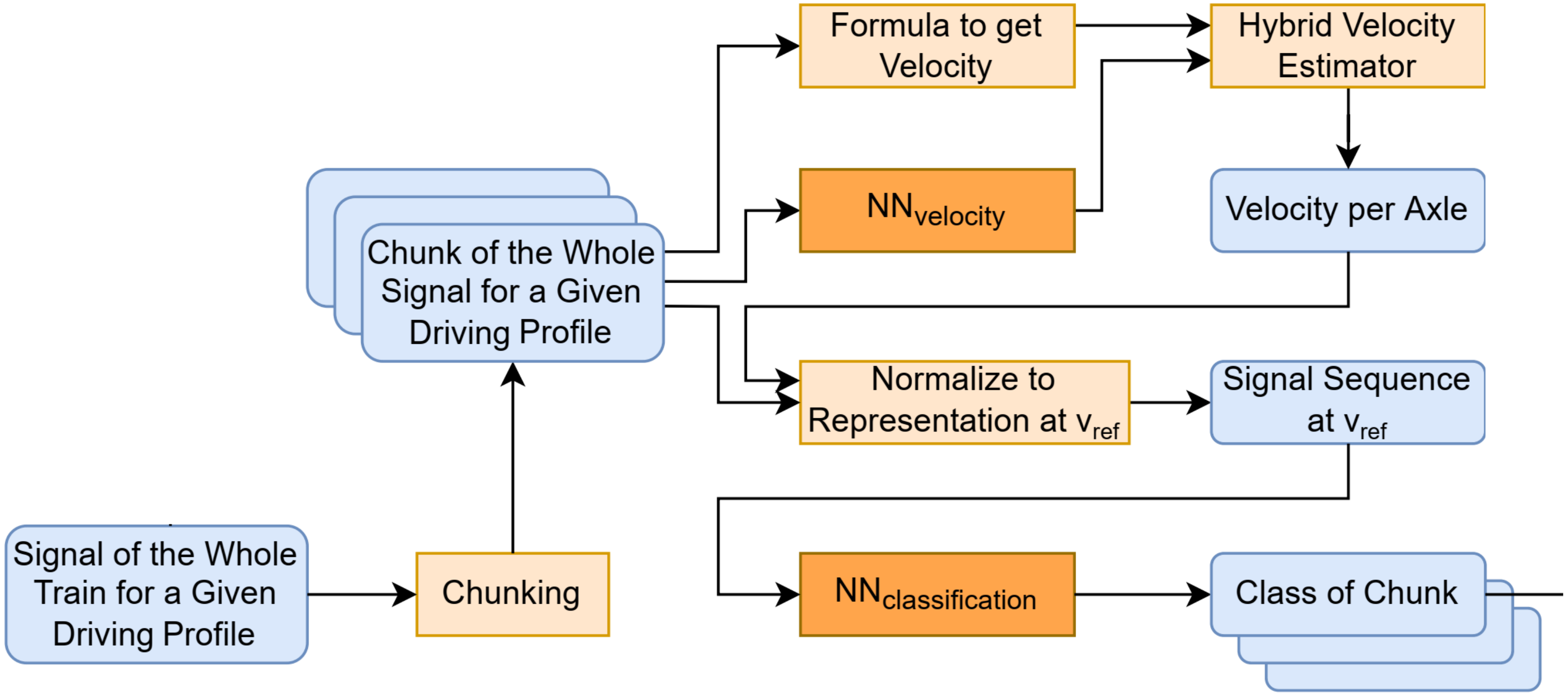
Application



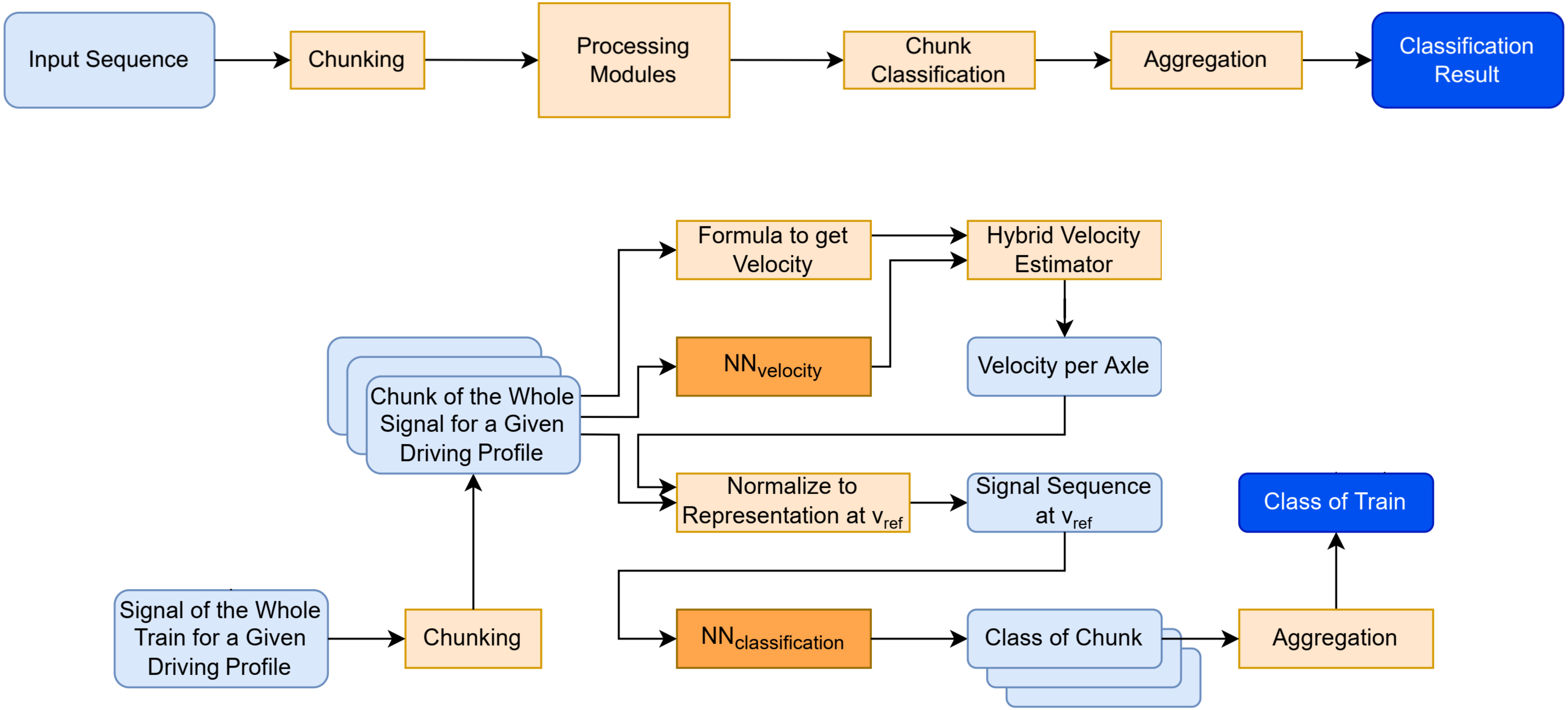


Application

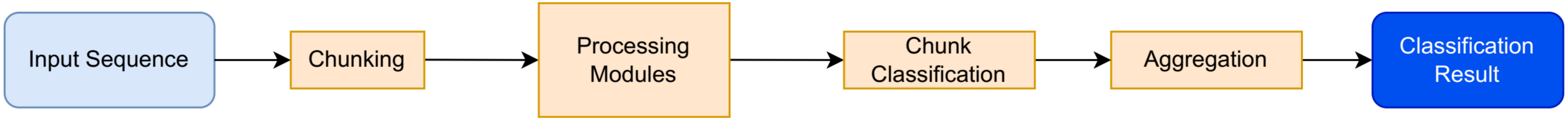


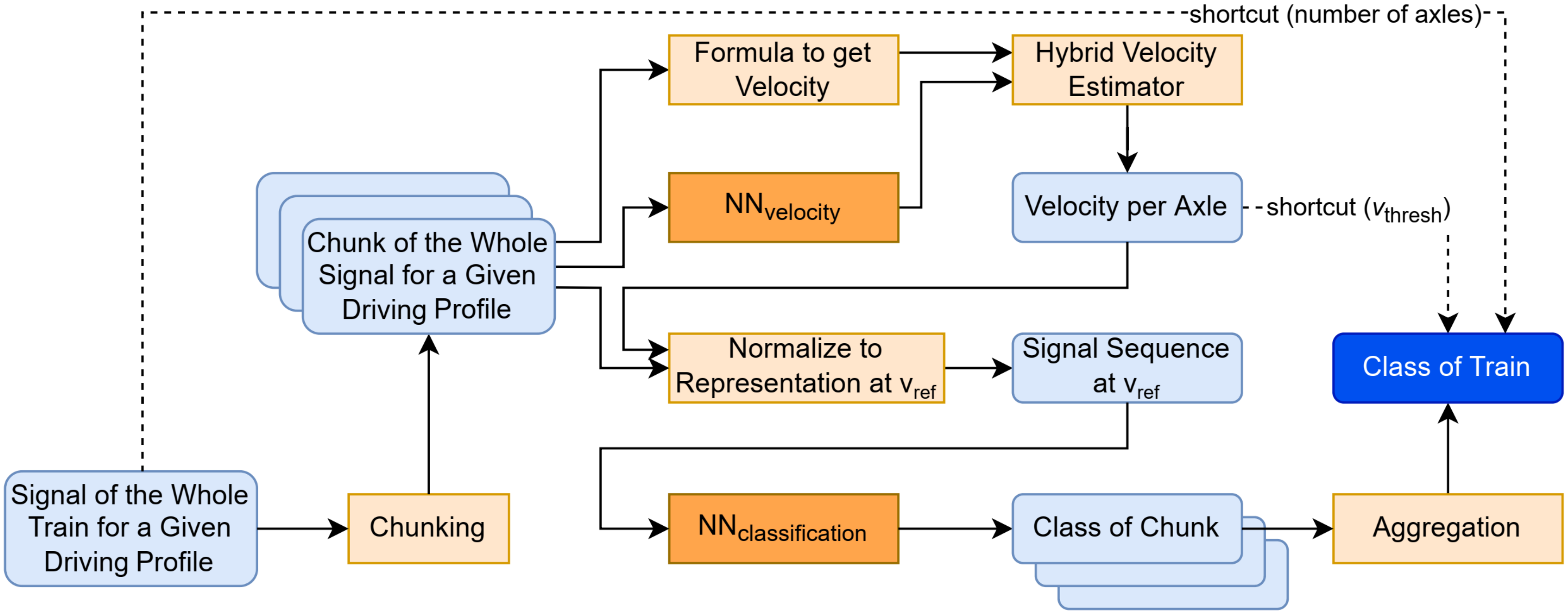


Application



Application





Evaluation

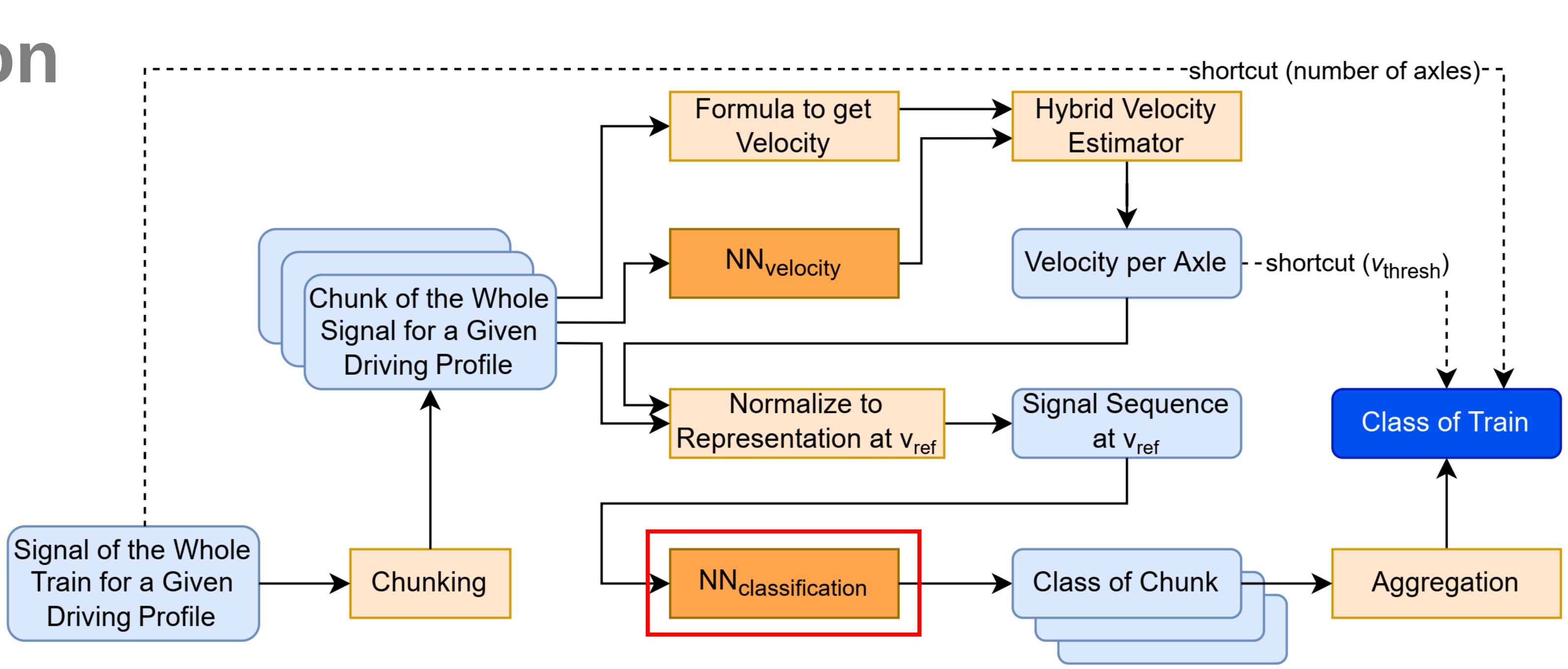
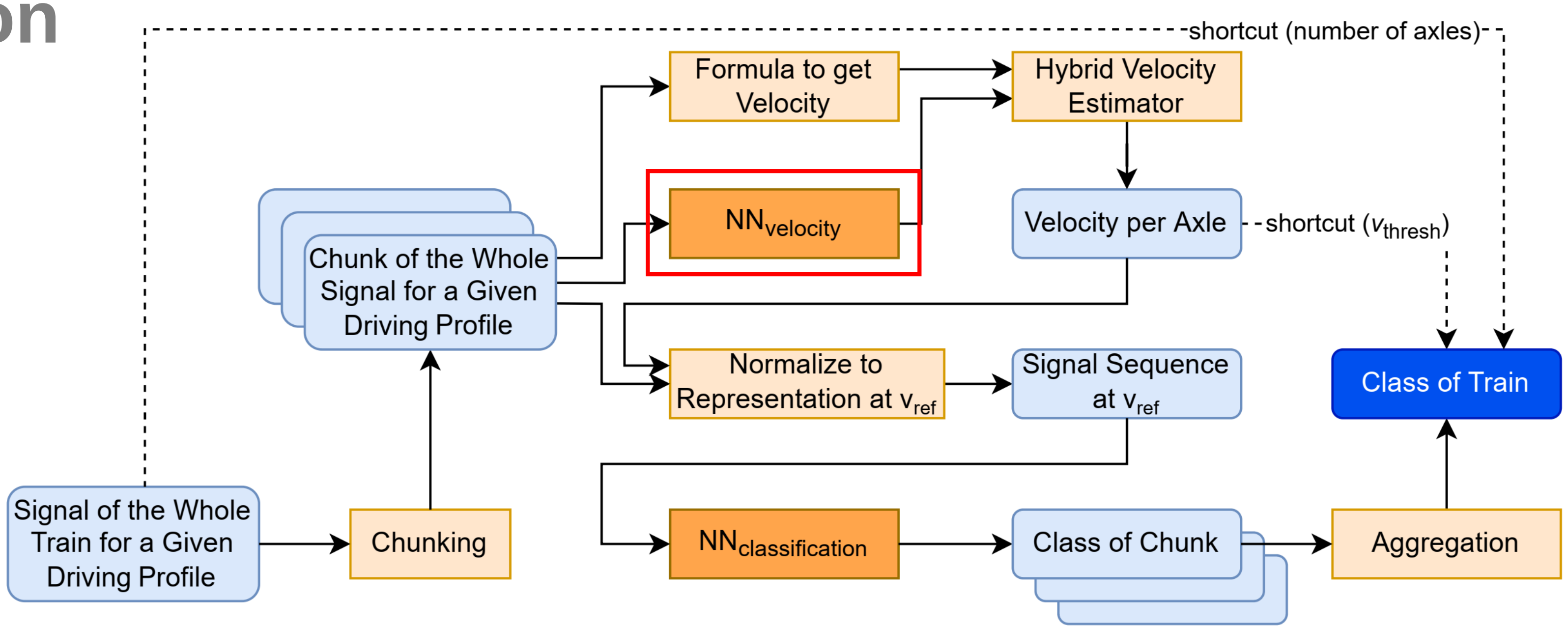


Table 1. Results of $NN_{classification}$ on its training/validation/test datasets.

	accuracy	precision	recall
$Train_{class}$	0.9835	0.9695	0.9836
$Validation_{class}$	0.9623	0.9618	0.9291
$Test_{class}$	0.9886	0.9777	0.9902

Evaluation



- NN$_{velocity}$ achieves 0.93% relative error
- Formula achieves 0.56% relative error
- But NN$_{velocity}$ is better at higher velocities, while the formula favors lower velocities
- Hybrid Velocity estimator switches between both models based on a threshold at ~49 km/h
  - → 0.35% relative error

Evaluation

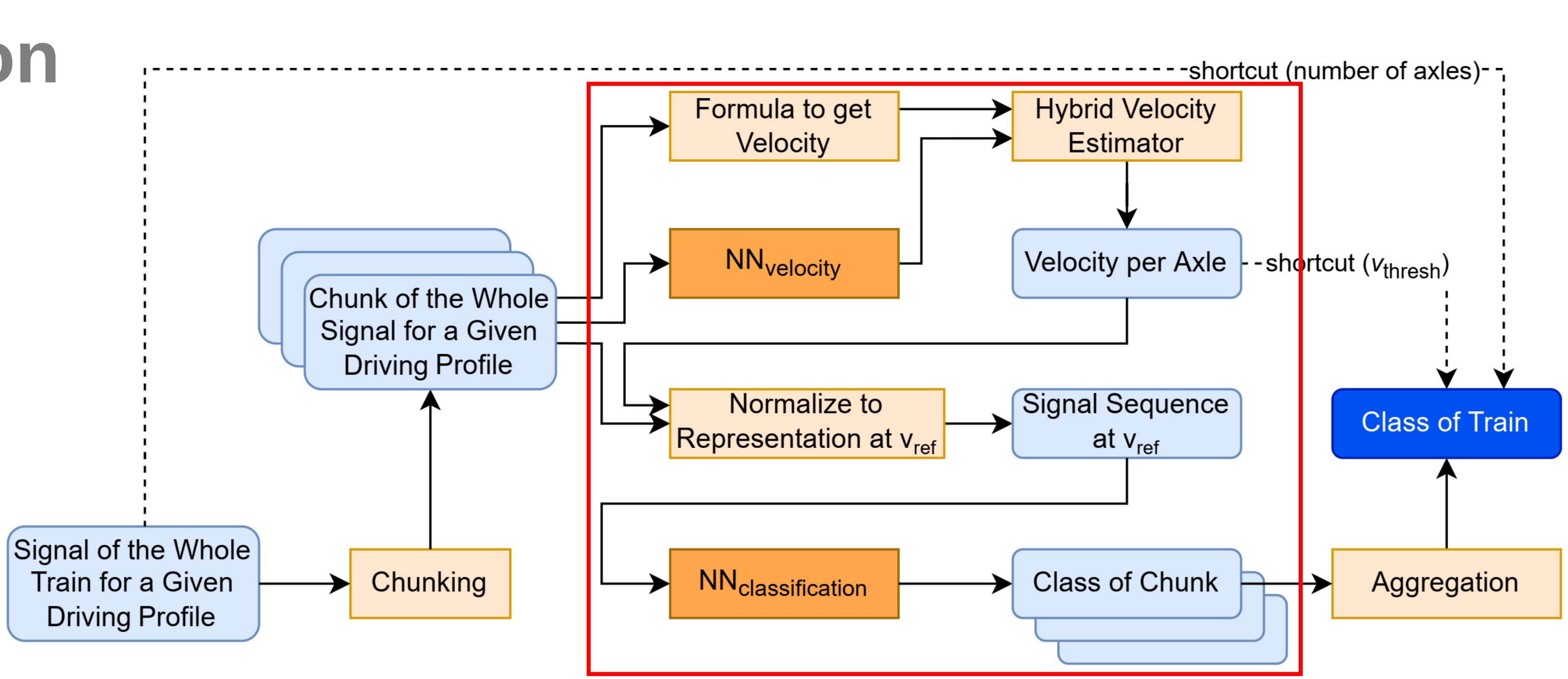


Table 2. Results of the robustness analysis for the different normalization methods and different sources of axle velocities.

	$norm_{linear}$			$norm_{actual}$		
	accuracy	precision	recall	accuracy	precision	recall
GT	0.9541	0.9591	0.9073	0.9667	0.9649	0.9389
formula	0.9538	0.9588	0.9066	0.9660	0.9644	0.9374
NN	0.9544	0.9583	0.9090	0.9657	0.9628	0.9382
hybrid	0.9539	0.9587	0.9072	0.9662	0.9643	0.9380

Evaluation

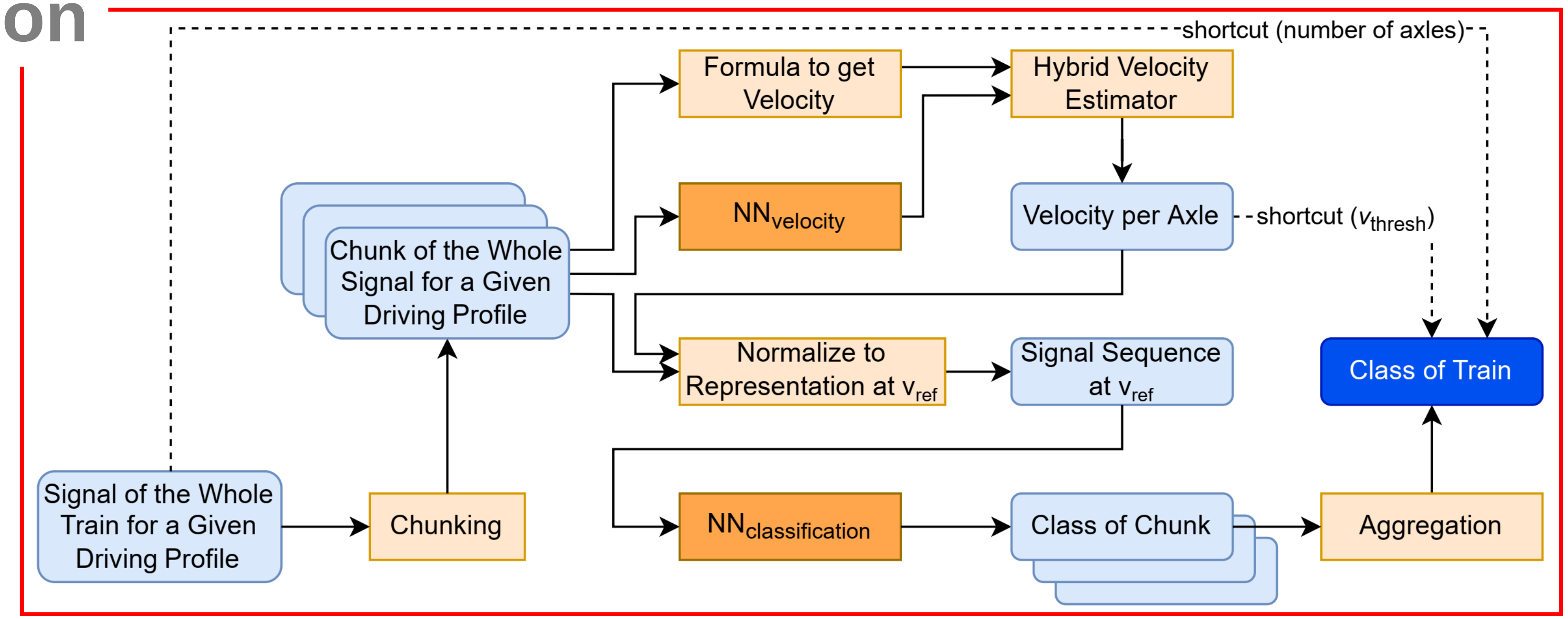
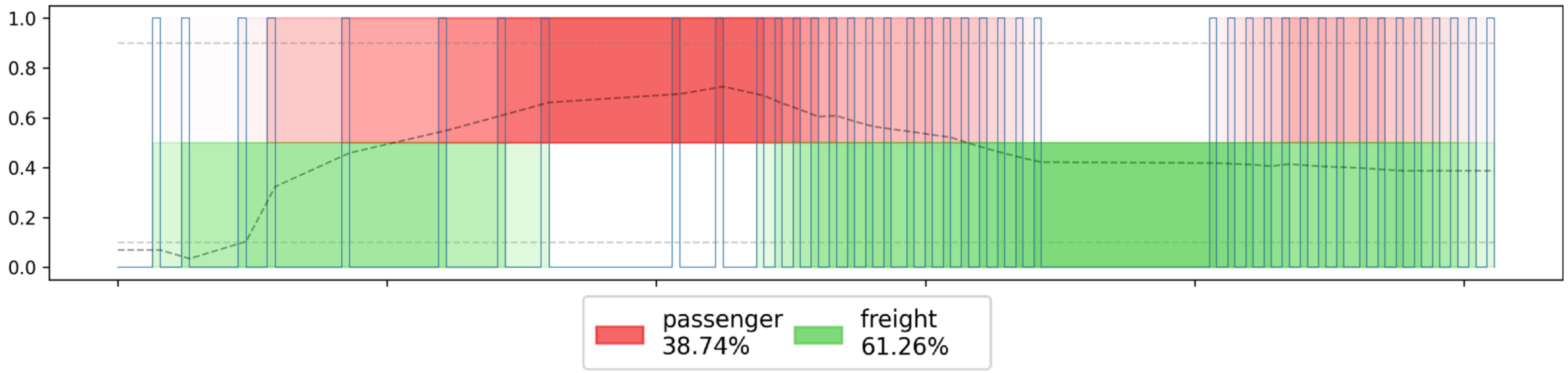


Table 3. Results of full train classifiers on $Test_{Trains}$

Method		accuracy
E2E	LSTM	0.8963
	FCN	0.9197
Ours	*none*	0.9378
	speed	0.9378
	axles	0.9984
	axles-speed	0.9984

Explanations

- Due to chunking, localization of errors is free
  - No need of approaches like LEAM or Grad-CAM to extract heatmaps
- Enables an attribution of errors for further investigation
  - Here: Detection of a labeling error in the train database ← Mixed train labeled as freight train



Conclusion

- Functional decomposition enables easily understandable architectures involving AI modules
- Components can be analyzed individually and in combination, leading to better insights into the system's behavior
- Further, post-hoc explanations for error localization can easily be derived
- In our case study, such a system can even outperform end-to-end architectures