Neuro-Symbolic Workflow for Lightweight Detection of Traffic Anomalies in Real Time

Michiel Dhont[a,*] **and Elena Tsiporkova**[a]

[a]EluciDATA Lab of Sirris, Belgium
ORCID (Michiel Dhont): https://orcid.org/0000-0002-5679-0991, ORCID (Elena Tsiporkova): https://orcid.org/0009-0003-7202-3471

Abstract. Traffic plays a crucial role in modern life, influencing economy, public safety, environmental health, and overall quality of life. As cities grow and transportation networks become more complex, urban planning and monitoring become an even more difficult task. For this reason, extensive sensor networks (including smart cameras and inductive loops) are deployed across road infrastructures to collect valuable data on a continuous basis. Real-world traffic data is highly dynamic and complex, which makes accurate understanding, timely and meaningful forecasting, and actionable monitoring of traffic behaviour a significant challenge. This paper proposes a neuro-symbolic workflow for lightweight, real-time traffic anomaly detection, designed to handle diverse traffic conditions effectively. The approach consists of four key components: 1) a discretisation workflow allowing to partition historical multivariate time series into a limited set of interpretable traffic states; 2) a neural detection model allowing real-time identification (detection) of traffic states from newly incoming streaming traffic data; 3) a parametrised family of context-specific Markov chains allowing to capture state transition behaviour under specific conditions, e.g., location or time period of interest; 4) a decision logic allowing to identify anomalous traffic behaviour using the Jensen-Shannon divergence between the detected and predicted state probabilities. The proposed approach is validated in two distinct case studies: a dense urban traffic corridor in Brussels, Belgium, and a large-scale highway network in the San Francisco Bay Area, USA. While the Brussels dataset offers fine-grained temporal data over an extended period, the San Francisco dataset covers a vast number of monitored locations. The results demonstrate the effectiveness of our method in identifying anomalous traffic behaviour, providing valuable insights for traffic management and decision-making.

1 Introduction

Nowadays, a lot of streaming data is collected from complex real-world phenomena and machinery, allowing to monitor production processes, industrial assets operating remotely in the field or even complete infrastructures. Such streaming data is very dynamic in nature and complex in terms of multitude of parameters, number of assets, and time granularity. The evolution and dynamics of such complex data depends on a multitude of factors, often interdependent and influencing each other. Many of those factors, e.g., technical configurations, diverse operating contexts, varying environment conditions, or human activities, are difficult to monitor, model and predict. Therefore, it could be beneficial to determine for a context of interest (e.g., location or specific moment in time) whether the monitored phenomenon is behaving/operating as expected or whether there is an anomalous trend, which requires attention and mitigation.

Traffic serves as an ideal playground to study and explore such complex real-world phenomena. It represents a network of interconnected roads with highly varying temporal and spatial dynamics. For instance, while some areas exhibit similar traffic patterns, others differ substantially due to factors such as the type of traffic (e.g., leisure, shopping, commuting), the nature of the area (e.g., rural, urban), and the infrastructure (e.g., bending road, speed bump, low speed limit). These variations make it challenging to characterise 'normal' behaviour in an interpretable and scalable manner, making the identification of anomalies more challenging. The latter might be extremely relevant for the involved stakeholders, allowing them to better understand emerging trends and take informed actions.

This paper proposes a neuro-symbolic workflow for lightweight, real-time traffic anomaly detection. Designed for versatility, our approach adapts seamlessly to a wide range of scenarios, accommodating variations in traffic types and data availability (e.g., number of parameters, monitored locations, sampling frequency, and historical data volume). To demonstrate this flexibility, the approach is validated in two distinct traffic use cases. The first examines a hectic and dense urban traffic axis in Brussels, Belgium, near the city center. The second focuses on highways in the San Francisco Bay Area, USA. Note that while the American case covers a significantly larger number of monitored locations, the Brussels dataset provides more granular temporal data with richer parametrization over an extended observation period. Our validation results demonstrate that the neuro-symbolic workflow effectively identifies traffic anomalies across different contexts, providing valuable insights for traffic management and decision-making. Moreover, the lightweight nature of our approach makes it well-suited for real-time deployment, enabling efficient and scalable monitoring of complex traffic systems.

2 Related Work

2.1 Time Series Anomaly Detection

Time series data often has repetitive underlying patterns (e.g., daily behaviour). Consequently, these patterns can be discovered using appropriate decomposition techniques, such as non negative matrix factorisation, or tensor rank decomposition, and be interpreted as build-

* Corresponding Author. Email: michiel.dhont@outlook.com.

ing blocks to (approximately) reconstruct parts of the time series. In anomalous situations, the time series will not be able to properly be reconstructed using the known underlying patterns, so, such anomalies can be detected by a high reconstruction error [19, 11]. Li *et al.* [12] go even a step further by considering a coupled scalable Bayesian robust tensor factorisation model. Specifically, they couple the reconstruction error of multiple variables, using an N*th*-order tensor for each variable (instead of the typical two-way matrix). In their traffic use-case, they illustrate how to capture not only repetitive daily patterns, but also weekday and whole week patterns, and even patterns across different locations. However, best results are obtained using only a third-order tensor, ignoring the interactions across locations, which they attribute to the presence of off-ramps between locations, making the spatial relation not fixed.

Another common way to do anomaly detection on time series is to use symbolic time series analysis (STSA). STSA is a technique to simplify complex data into a sequence of symbols (using any possible partitioning technique such as uniform partitioning, maximum-entropy partitioning, and K-means), enabling easier detection of patterns, trends, and anomalies. Ghalyan *et al.* [8] propose an approach to detect anomalies using STSA by extracting state transition probabilities for optimally dimensioned time segments, and use these as features of each time window, allowing to classify anomalous vs. normal behaviour.

In this paper, we develop a lightweight anomaly detection mechanism relying on the transformation of multivariate time series into sequences of symbols. A finite set of humanly-interpretable traffic states (e.g., congestion, free-flow, traffic build-up) is extracted, in the absence of any ground truth, following a human-in-the-loop clustering strategy we proposed in [6].

2.2 Anomalies in Traffic

Anomaly detection is the approach of identifying anomalous behaviour. Although one might expect this definition to be satisfying at first sight, it is not unambiguous. Depending on the application domain, an anomaly can be interpreted different. In the use case of this paper, traffic, for example. Anomalies can be interpreted as traffic congestion, which is an unwanted situation in traffic that is preferably avoided. [17]. Alternatively, Li *et al.* [12] propose an approach to detect non-recurrent traffic congestion. With this approach, the daily (recurrent) traffic congestion during rush hours is considered expected behaviour. The latter allows to focus only on the situations that have the most impact on drivers: the unexpected congestion. A third alternative, which we adopt in this paper, is to detect all unexpected behaviour, regardless of the type of unexpected behaviour. In contrast to the previous two approaches, this method also aims to identify traffic events that lead to reduced traffic flow, such as holidays or road obstructions. This allows for a broader perspective on unexpected traffic behaviour, providing valuable insights for traffic stakeholders.

2.3 Neuro-Symbolic Modelling

Neuro-symbolic modelling integrates neural, logical, and probabilistic paradigms from artificial intelligence, creating unified frameworks/workflows that leverage the strengths of these diverse methodologies. By combining the principles of logic and neural networks, both of which inherently incorporate probability theory, neuro-symbolic systems enable high-level reasoning to coexist with low-level perception [10]. This synthesis overcomes the limitations of purely deep learning-based approaches, unlocking novel capabilities in AI [16].

The landscape of neuro-symbolic architectures is diverse, but two major categories stand out: architectures that use logic as a regularization mechanism within neural networks, and those that enhance logical reasoning methods with neural components. For instance, Hammoudeh *et al.* [9] demonstrate how neuro-symbolic modelling can integrate physics-based knowledge. They propose a method to predict the power of marine cargo vessels by combining a physics-informed symbolic module with a neural network, achieving performance that surpasses the state-of-the-art. Similarly, Capogrosso *et al.* [2] utilise a neuro-symbolic framework to enrich smart manufacturing. Their approach embeds structured, formal knowledge into industrial processes, by incorporating industrial ontologies within a diffusion model, showcasing the versatility of neuro-symbolic systems.

The workflow proposed in this paper leverages the potential of two fundamentally different modelling paradigms. A powerful, yet black-box, neural traffic state detection model is combined with a lightweight probabilistic model, composed of sequentially linked Markov chains, aiming at state transition prediction in a context-sensitive way. This decoupled architecture not only enhances interpretability but also enables efficient updates in response to concept drift, ensuring adaptability to evolving traffic patterns. By linking black-box neural models with structured probabilistic reasoning, our approach aligns with the core principles of neuro-symbolic AI, further demonstrating its potential in dynamic, real-world applications.

3 Methods

3.1 Modelling State Transitions: Parametrised Markov Chains

In the mobility domain, time series data often includes parameters such as average speed, vehicle flow, and road occupancy across multiple locations. This fine-grained, multivariate time series data, commonly found in industrial settings, can be challenging to interpret for humans and difficult to model accurately with algorithms. To address this, we proposed in [6] the extraction of a finite set of humanly-interpretable traffic states for the purpose of facilitating the annotation of mobility data with meaningful labels such as congestion, free-flow, traffic build-up, etc. These traffic states provide a symbolic representation for each time window at each location, effectively summarizing the traffic situation using a limited set of symbols with clear semantic meaning, thus enhancing human interpretability.

Transforming multivariate time series into sequences of symbols, analogous to a DNA sequence, unlocks a range of opportunities to integrate multiple complementary approaches for modelling state transition behaviour. For instance, representing a time series as a sequence of N semantically interpretable states $S_1, S_2, \ldots, S_N$, allows for modelling state transition probabilities using Markov chains. However, traffic dynamics, especially in urban areas, do not strictly follow Markovian logic. Traffic patterns are inherently periodic, influenced by daily, weekly, and seasonal human routines. As a result, state transition probabilities vary throughout the day and across different time periods. To account for this periodicity, we introduced in [4] the concept of a family of parametrised Markov chains ($\mathscr{M}$):

$$\mathscr{M} = \left\{ \mathbf{M}_{s,\kappa,t_{start},\lambda} | s, \kappa \in \mathbb{N}_{>0}; t_{start} \in [T_{min}; T_{max}); \lambda \in \Lambda \right\}, \tag{1}$$

where each transition probability matrix $\mathbf{M}_{s,\kappa,t_{start},\lambda}$ (of size $N \times N$) is uniquely defined by:

- **Granularity** s: The time step resolution for transition probability estimation.
- **Start time** t_{start}: The time reference point within the range $[T_{min}, T_{max}]$ (e.g., hours of the day or months of the year).
- **Time frame coverage** κ: The period $[t_{start}, t_{stop}]$ used to estimate transition probabilities, with $t_{stop} = t_{start} + s \cdot \kappa$.
- **Context** $\lambda \in \Lambda$: A unique temporal and/or spatial setting, such as a specific group of locations or time period.

The proposed family of Markov chains ($\mathscr{M}$) provides a flexible framework for analysing state transitions across varying temporal and spatial contexts. By associating each Markov chain with a specific context λ, we capture state transition behaviour unique to different locations, weekdays, or time periods. This enables a detailed analysis of periodic traffic dynamics and localized traffic bottlenecks. Moreover, the approach is computationally efficient, scaling linearly with data size ($\mathcal{O}(n)$), making it well-suited for large-scale applications. Moreover, Markov chains require relative small amounts of data (e.g., compared to deep models), allowing for fast retraining in case of concept drift. An important advantage of Markov chains is their inherent interpretability, making them highly explainable compared to complex deep learning models. Their transition probabilities offer clear insights into how traffic states evolve over time, enabling domain experts and traffic operators to analyse and understand mobility patterns more intuitively.

3.2 *Neuro-Symbolic Workflow for Anomaly Detection*

We proposed in [7], a novel hybrid modelling framework leveraging multiple data representations (temporal, time-frequency and symbolic) with the aim of forecasting traffic progression in terms of humanly-explicable state transitions. Three distinct modelling paradigms have been explored: neural, neural-to-symbolic, and symbolic-to-neural, and their potential to capture and forecast traffic dynamics has been convincingly demonstrated on real-world mobility data. The neural-to-symbolic approach is using in a sequential fashion a single deep learning state detection model, trained on the temporal representation and followed by a family of Markov chains managing the state prediction for different forecast horizons. This sequential neuro-symbolic workflow is being creatively adapted and enhanced in this work for the purpose of anomaly detection, as outlined in detail below and depicted in Figure 1.

The proposed neuro-symbolic approach consists of four key components applied in a sequential fashion: 1) *extraction of characteristic traffic states*: a discretisation workflow allowing to partition historical multivariate time series into a limited set of interpretable traffic states; 2) *neural traffic state detection*: a neural detection model allowing real-time identification (detection) of traffic states from newly incoming streaming traffic data; 3) *symbolic state transition prediction*: a parametrised family of context-specific Markov chains allowing to capture state transition behaviour under specific conditions, e.g., location or time period of interest; 4) *anomaly detection*: a decision logic allowing to identify anomalous traffic behaviour using the Jensen-Shannon divergence between the detected and predicted state probabilities.

Consider the availability of N traffic states $S_1, S_2, \ldots, S_N$, which are used to annotate/label the available historical time series data and subsequently train a *neural* state detection model $\mathscr{D}$. The model input is composed of time fragments extracted from the historical multivariate time series, while the output at any time t_i ($i \in \mathbb{N}_{>0}$) is a vector of soft labels $\hat{\boldsymbol{y}}_i$ of size N, expressing the confidence of being in one of the possible states at time t_i.

Next, the *symbolic* step of our workflow, based on the parametrised Markov chains, uses the detected vector at the previous time step $t_{i-1} = t_i - s$ to predict the expected state occurring at the present time step t_i, i.e., $\hat{\boldsymbol{y}}_{i-1}$, can be multiplied with the state transition matrix $\mathbf{M}_{s,\kappa,t_i-s,\lambda}$ of size $N \times N$ to generate a vector of expected (i.e., predicted) state probabilities $\hat{\boldsymbol{z}}_i$ at current time t_i. Subsequently, the distance between vectors $\hat{\boldsymbol{y}}_i$ and $\hat{\boldsymbol{z}}_i$ generated by the neural and the neuro-symbolic steps respectively, can be calculated using metrics such as Minkowski distance, cosine distance, or Jensen–Shannon divergence [13]. Such distance quantifies the deviation between the observed and expected behaviour for the given context at that moment in time, serving as an anomaly score.

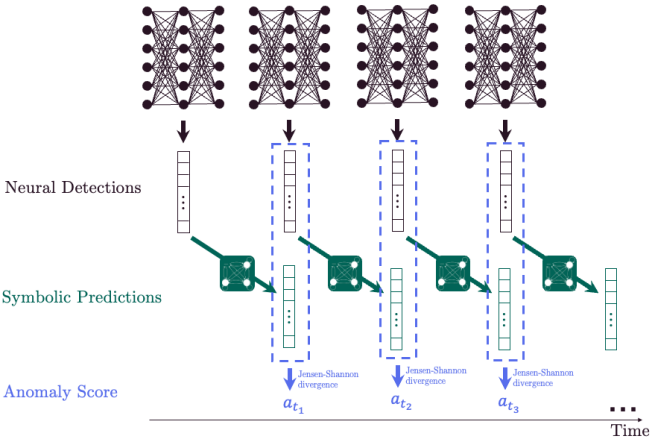


Figure 1: Overview of the anomaly detection workflow.

The so-generated anomaly scores serve as a continuous monitoring mechanism, enabling dynamic thresholding strategies to flag anomalies based on a streaming data in real-time. This approach allows for adaptive anomaly detection, offering traffic stakeholders greater flexibility in distinguishing between normal fluctuations and significant deviations requiring intervention.

3.3 *Dealing with High Spatial Variability*

Mobility is a complex system where road segments are spatially interconnected and influence each other. Data from road networks typically encompasses various types of spatial contexts (e.g., intersections vs. straight roads) and varying traffic usage (e.g., morning commuting vs. afternoon shopping traffic). Developing location-specific models (e.g., multiple families of parametrised Markov chains from Section 3.1) can potentially capture the unique characteristics of each spatial context. Unfortunately, in case of large number of spatial contexts with diverse traffic dynamics, this approach requires the construction and maintenance of a large number of models, each trained on a limited subset of data, which may hinder generalizability. Alternatively, one may opt for identifying groups of spatial contexts with similar traffic behaviour, enabling the development of tailored models that leverage shared traffic patterns while maintaining adaptability to local conditions.

Different spatial contexts, i.e., sections of the road network with available sensor data, can be grouped based on similar traffic dynamics. This requires suitable representations of these contexts that accurately capture the most discriminating traffic characteristics. Given the repetitive nature of traffic, such representations can benefit from comparison between recurring time periods as days, weeks, months, etc. which allows to identify trends and seasonality. We developed

in [5] a general visual analytics methodology that allows converting time series data into a series of circular heat maps covering recurring time periods. More concretely, the methodology allows to construct temporal fingerprints for each spatial context that capture both daily and weekly patterns. The time series are divided into sufficiently granular (e.g., 15-minute or 1-hour intervals) segments and subsequently, relevant traffic features, such as the median speed or average road density, are derived per time segment and aggregated across days and weeks, respectively. This results into two-dimensional (days vs. time segments) weekly fingerprints, which compactly capture the characteristic weekly behaviour for each spatial context. In addition, these fingerprints can be visualised using circular heatmaps, as illustrated in Figure 2, offering very insightful representation of traffic behaviour at a glance. For instance, as illustrated in Figure 2, different locations exhibit substantially different traffic patterns. The leftmost location shows a broad evening peak with lower median speed due to heavy traffic, while the middle location has a sharp, narrow morning peak, indicating a short spike in traffic density.

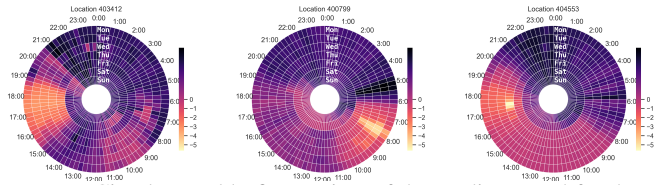


Figure 2: Circular weekly fingerprints of the median speed for three locations from the PeMS-Bay dataset.

Next to the possibility to visually examine and compare these fingerprints, the pairwise proximity between them can also be calculated. For instance, employing Pearson correlation[14] or other suitable distance metrics will allow to partition the different fingerprints, i.e., spatial contexts, into clusters of contexts exhibiting similar temporal behaviour. This approach substantially reduces spatial variability and creates more homogeneous groups.

4 Validation on Real-World Mobility Data

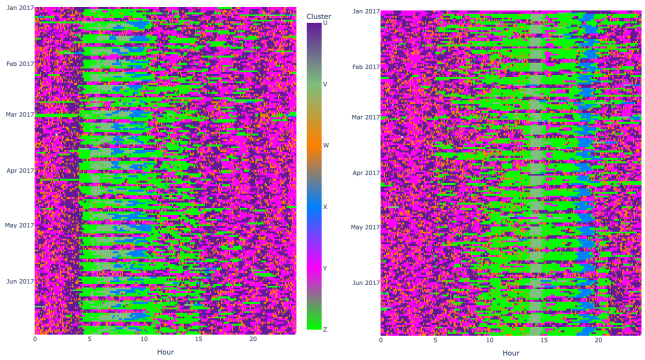
4.1 Use Cases

Our neuro-symbolic anomaly detection workflow is validated on two distinct use cases. The first use case, referred below as **Brussels use case**, is based on a dataset consisting of highly granular traffic measurements from a limited number of locations. Specifically, it covers 16 locations along the *Brussels small ring*, Belgium. The dataset is composed of minute-by-minute records of average velocity, vehicle flow, and road occupancy over more than two years (16/01/2020 - 04/05/2022). The second use case, referred as **PeMS-Bay use case**, utilises the widely known *PeMS-Bay* dataset, which includes average velocity data recorded every 5 minutes over a 6-month period (01/01/2017 - 30/06/2017) for 325 highway locations near the San Francisco Bay Area, USA. Note that, while the first use case dataset is concerned with a very dense urban mobility infrastructure and deals with challenges related to modelling highly granular (temporal) data with a rich feature set, the second use case dataset is collected from a wide highway road network and thus confronted with scalability issues due to its extensive spatial coverage. Moreover, the two use cases are based in two different continents, and may therefore exhibit fundamentally different traffic dynamics due to variations in infrastructure, driving behaviour, and regulatory policies.

4.2 Traffic States

For each of the two use cases 6 distinct traffic states are derived. For the Brussels use case, we build upon the validated traffic states established in [6]. These states were derived through an extensive human-in-the-loop clustering approach, ensuring interpretability and alignment with traffic theory, as confirmed by domain experts. Concretely, the available time series data is labelled with one of 6 traffic states (free-flow, traffic build-up, traffic intensity reduction, stable non-saturated traffic, variable non-saturated traffic, or congested traffic) at 15-minute intervals.

The PeMS-Bay dataset has a coarser time granularity of 5 minutes (compared to 1 minute for the Brussels dataset) and includes only a single measurement of average velocity (as opposed to average velocity, vehicle flow, and road occupancy for the Brussels dataset). Temporal features over 1-hour time windows are extracted, i.e., extracted from a parameter vector of 12 values, rather than 3 parameter vectors of 15 values used in the Brussels dataset. For the purpose of further enhancing the feature set granularity, the 1-hour windows are generated in a rolling-window fashion, providing new features every 5 minutes. After normalisation and scaling the features, majority voting using a diverse set of cluster validation measures (silhouette index [15], Calinski Harabasz index [1], Davies Bouldin index [3] and elbow index [18]) identified an optimal number of 6 clusters (traffic states) using k-means clustering. Although the number of clusters matches the one of the Brussels use case, their semantic meaning is not necessarily the same. The difference arises from the longer time period they represent (1 hour vs. 15 minutes) and the distinct geographic characteristics of the dataset. Unfortunately, for the PeMS-Bay dataset we cannot rely on support from traffic domain experts to assign a semantic meaning to each state. We can employ some visual analytics approaches instead to evaluate their potential to capture distinct traffic behaviour. For instance, the state transitions represented as label maps for two contrasting locations can be examined in Figure 3. Note that Figure 3a depicts a location with a clear morning rush hour (cluster V and X), while Figure 3b identifies peak traffic in the afternoon and early evening.



(a) Location 400001. (b) Location 400922.

Figure 3: Label map of traffic states at two locations in PeMS-Bay.

4.3 Models

Based on our neuro-symbolic workflow, two types of models must be developed, as outlined further in this section. For this, both datasets are divided into three subsets: 80% for training, 10% for validation, and 10% for testing. To prevent data leakage and ensure temporal consistency, all data points from a given day are assigned entirely to one of these three sets.

4.3.1 *Neural State Detection Models*

For both use cases, a state detection GRU model is trained using the training and validation data. Note that for each of the use case, 6 distinct traffic states are derived, i.e., the target variable for the state detection model is a probability vector of 6 values, representing the model's confidence in each possible traffic state.

For the Brussels use case a retrospective time window of 2 hours is considered, while PeMS-Bays dataset only considers a 1-hour time window as input. Despite being trained on a single parameter (vehicle speed), the PeMS-Bay model still achieves an accuracy of 73% and a categorical cross-entropy loss of 0.61 on the test set. This contrasts with the Brussels case study, where the model attained 86% accuracy and a loss of 0.33 thanks to the richer input of three parameters. The model architectures used can be consulted in Sections A and B.

4.3.2 *Parametrised Markov Chain Models*

The next building block of our anomaly detection approach are the (symbolic) families of Markov chains. For both use cases, the training and validation datasets are pooled together and used for training the families of Markov chains. For each use case, a family of time-aware models is constructed for each considered location. Practically, the days are segmented in time intervals and for each segment a separated model is trained.

In the Brussels case study, only 16 locations are considered, and the days are segmented into intervals of 15 min. Thus, for each location a family of 96 time-aware models is derived, where each model corresponds to a specific 15-minute time interval. To enable location-specific models while enhancing robustness, the historical traffic state transitions observed within that interval, along with data from the preceding and following hour, are incorporated from the training dataset.

Deriving location-specific families in the PeMS-Bay case would result in 325 families, which might suffer from low accuracies (since only 6 months of data is available) and contain a lot of redundancy (due to the high similarities across certain locations). For this reason, applying the approach discussed in Section 3.3, the total of 325 locations are partitioned into 18 groups exhibiting similar behaviour. The latter is done by first extracting weekly patterns for each location, as visualised in Figure 2. Subsequently, the pairwise Pearson correlation is computed between these fingerprints and the complete linkage algorithm is used to obtain the final partition. For each group, a family of 288 Markov chains is constructed to model traffic state transitions at every 5-minute interval throughout the day, based on the training data. To enhance robustness, each transition matrix is computed using a 65-minute time window centred around the target timestamp, incorporating data from 30 minutes before to 30 minutes after.

4.4 *From State Detection to Anomaly Detection*

Although the Brussels and PeMS-Bay use cases and associated datasets differ significantly, the neuro-symbolic anomaly detection workflow as proposed in Section 3.2 is highly adaptable and can be flexibly tailored to different traffic environments and data characteristics. The key aspects remain the same for both use cases. Concretely, for each location and each newly arriving time window of streaming data, the following entities are computed in real time:

- *Detected state probability distributions*: the GRU model takes as input a retrospective time window (3 x 15 matrix the for Brussels use case or vector of size 12 for the PeMS-Bay use case) and outputs a probability distribution over the 6 traffic states;
- *Expected state probability distributions*: the location-specific and time-aware Markov chain models estimate the state transition probabilities for the current time step, taking as input the state probability distributions produced by the GRU model for the previous time window;
- *Anomaly scores*: the Jensen–Shannon divergence between the detected and the expected state probability distributions is computed.

One advantage of using the Jensen–Shannon divergence, instead of for example the Minkowski distance, is a standardised value range between 0 and 1. The obtained anomaly scores for the Brussels training dataset are visualised in Figure 4. Note that the distribution of the anomaly scores appears to resemble a mixture of two normal distributions rather than a single one. This distinction arises from the very different traffic conditions during nighttime (which is easier to predict) and during daytime (which is more variable and challenging to predict). The red line in Figure 4 denotes the 99^{th} percentile, which is used in the examples below as the anomaly detection threshold. Other common statistical methods, as the six-sigma or the 1.5 IQR (interquartile range), can be also considered for this purpose.

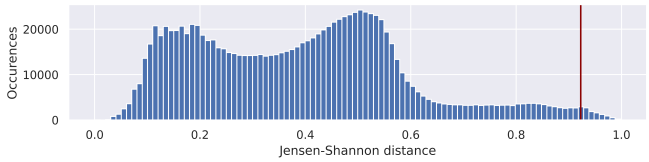


Figure 4: Histogram of the Jensen-Shannon divergence values on the test set of the Brussels use case. In red, the 99^{th} percentile is indicated.

In Figures 5 and 6, the anomaly scores for the 16 Brussels locations for two subsequent days, May 11th and 12th of 2021, are depicted. Depending on the desired sensitivity of the anomaly detection algorithm, anomalies can be flagged in several different ways. The most sensitive option would be to inspect the anomaly score at each timestamp and flag an anomaly whenever the detection *threshold is surpassed*. As it can be observed in the top plot of Figures 5 and 6, this approach can lead to the identification of numerous anomalies throughout the day for each location. This can lead to an excessive number of false positive alarms, which is not ideal for daily use by traffic operators. A more robust alternative will be to track *threshold exceeding with a moving average*, which smooths the anomaly score over a rolling window to prevent false positives caused by short-lived fluctuations or transient spikes in traffic behaviour. Applying a 1-hour moving average, as depicted at the middle plot of Figures 5 and 6, aids in eliminating short-term anomalies, thereby concentrating on the more severe cases. Note that on May 11th, the European Commission held a General Affairs Council in Brussels, while May 12th marked the end of the Ramadan period. It is expected that both events impacted substantially the *regular* traffic dynamics in Brussels, as observed and confirmed by the behaviour of the anomaly scores for these days. For May 11th, the daily evolution of the anomaly score across almost all locations exhibits several peaks (Figure 5), indicating that the European Commission event dramatically disturbed the traffic in Brussels on this day. For May 12th, only one clear anomalous moment is flagged (middle and bottom plot Figure 6), at Louisa tunnel (LOU_TD1) at 16:15. However, a second peak is observed at 20:30 for the same location, significantly exceeding the standard deviation range. This peak may be attributed to the

final day of Ramadan, occurring just one hour before sunset and the fact that there is a mosque in the immediate vicinity of Louisa tunnel.

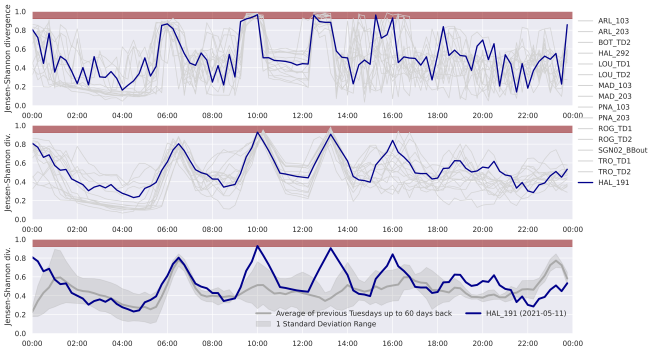


Figure 5: Evolution of the anomaly score on May 11th 2021, with the red line identifying the anomaly threshold. Top: anomaly score for each of the 16 Brussels locations (Hallepoort tunnel in blue); Middle: anomaly scores with a 1-hour moving average; Bottom: comparison of Hallepoort tunnel with the same weekdays of the past 2 months.

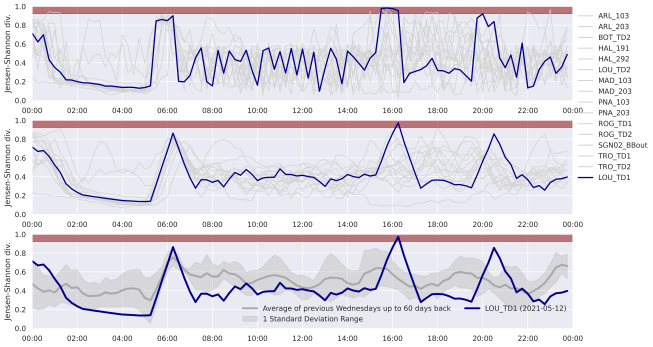


Figure 6: Evolution of the anomaly score on May 12th 2021, with the red line identifying the anomaly threshold. Top: anomaly score for each of the 16 Brussels locations (Louisa tunnel in blue); Middle: anomaly scores with a 1-hour moving average; Bottom: comparison of Louisa tunnel with the same weekdays of the past 2 months.

In Figure 7, the location and day with the most anomalies (i.e., 99% percentile threshold exceedings) for the PeMS-Bay use case is depicted (May 15th 2017). The anomaly scores of the anomalous location are compared against the average anomaly score for each location group (top and middle plot Figure 7). Considering the absence of any background knowledge for this use case, it is hard to validate the accuracy of the anomaly score behaviour. For the purpose of facilitating some interpretability, Figure 8 was generated alternatively. It depicts the number of anomalies on this specific day for all locations. In the purple zoomed-in square, three neighbouring locations with a high concentration of anomalies can be observed. The latter suggests the presence of an anomalous road segment, potentially caused by roadworks or a severe accident.

In Figure 9, the anomaly rate evolution when varying the anomaly score thresholds is depicted for both use cases. The difference of the anomaly rate evolution observed between the two use cases indicates that threshold selection should be tailored to each specific case, for example by considering a fixed percentile on a historical dataset. Moreover, the smoothing effect has a greater impact on the PeMS-Bay use case (Figure 9(B)), as evidenced by the larger deviation between the blue and grey curve compared to the Brussels use case (Figure 9(A)). This difference may be due to the higher frequency of anomaly calculations in the PeMS-Bay use case (every 5 minutes, compared to 15-minute intervals in the Brussels use case),

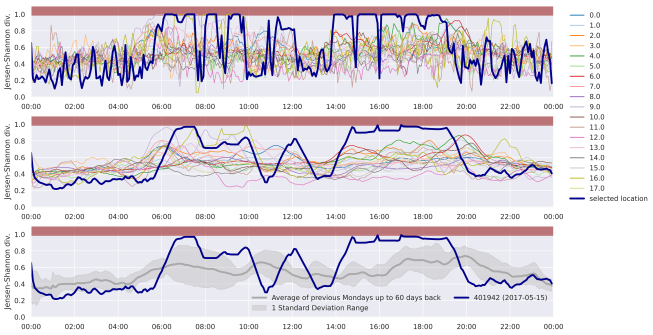


Figure 7: Evolution of the anomaly score on May 15th 2017, with the red line identifying the anomaly threshold. Top: average anomaly score for each of the 18 PeMS-Bay groups of locations; Middle: anomaly scores with a 1-hour moving average; Bottom: comparison of the most anomalous location (blue line) with the same weekdays

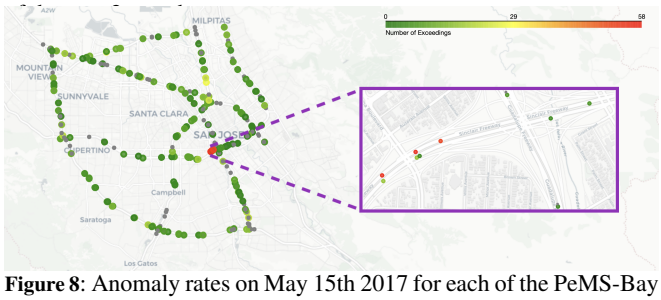


Figure 8: Anomaly rates on May 15th 2017 for each of the PeMS-Bay locations.

leading to a larger number of values being smoothed.

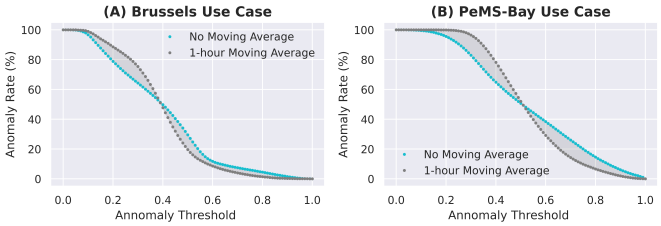


Figure 9: Percentage of detected anomalies in the test dataset across all possible threshold values.

5 Industrial Application

This work presents one of the applied outcomes from an industrial PhD research executed in close collaboration with an industrial partner in the mobility domain. They provided valuable domain expertise, offering insights into our results while also guiding the research toward practical applications with potential business impact. The technology readiness level of the proposed methodology is already quite advanced beyond 5 and its lightweight and interpretable nature makes it subject of ongoing plans of our industrial partner to embedded in its mobility platform.

6 Conclusion

In this paper, a neuro-symbolic workflow for real-time detection of non-recurrent traffic anomalies is presented. By decoupling detection and prediction through a combination of a neural model and symbolic Markov chains, our approach demonstrates flexibility across diverse traffic environments and data characteristics. The proposed framework was validated on two distinct case studies, urban traffic

in Brussels and highways in the San Francisco Bay Area, highlighting its adaptability to different spatial scales, temporal granularities, and feature sets.

Our results demonstrate that combining neural state detection with context-aware symbolic prediction is highly effective for anomaly detection. This approach captures short-term fluctuations while incorporating long-term traffic dynamics, enabling real-time monitoring with enhanced interpretability. Its flexibility and efficiency make it particularly well-suited for deployment in intelligent traffic management systems.

Acknowledgements

This research received funding from the Flemish Government (AI Research Program) and by the Brussels-Capital Region - Innoviris (TAPCOP project).

References

[1] T. Caliński and J. Harabasz. A dendrite method for cluster analysis. *Communications in Statistics-theory and Methods*, pages 1–27, 1974.

[2] L. Capogrosso, A. Mascolini, F. Girella, G. Skenderi, S. Gaiardelli, N. Dall'Ora, F. Ponzio, E. Fraccaroli, S. Di Cataldo, S. Vinco, et al. Neuro-symbolic empowered denoising diffusion probabilistic models for real-time anomaly detection in industry 4.0: Wild-and-crazy-idea paper. In *2023 Forum on Specification & Design Languages (FDL)*, pages 1–4, Turin, Italy, 2023. IEEE, IEEE.

[3] D. L. Davies and D. W. Bouldin. A cluster separation measure. *IEEE transactions on pattern analysis and machine intelligence*, 2:224–227, 1979.

[4] M. Dhont and E. Tsiporkova. Elucidating transition state behaviour from mobility data by cascades of markov chains. In *Proceedings of the Canadian Conference on Artificial Intelligence*, pages 1–12, Canada, 2023. Canadian AI.

[5] M. Dhont, E. Tsiporkova, T. Tourwé, and N. González-Deleito. Visual analytics for extracting trends from spatio-temporal data. In *Advanced Analytics and Learning on Temporal Data: 5th ECML PKDD Workshop, AALTD 2020, Ghent, Belgium, September 18, 2020, Revised Selected Papers 6*, pages 122–137. Springer, 2020.

[6] M. Dhont, E. Tsiporkova, and N. González-Deleito. Mining of spatiotemporal trajectory profiles derived from mobility data. In *2022 IEEE International Conference on Data Mining Workshops (ICDMW)*, pages 1–9. IEEE, 2022.

[7] M. Dhont, A. Munteanu, and E. Tsiporkova. Forecasting traffic progression in terms of semantically interpretable states by exploring multiple data representations. *Transactions on Intelligent Transportation Systems (T-ITS)*, accepted, waiting to be published.

[8] I. F. Ghalyan, N. F. Ghalyan, and A. Ray. Optimal window-symbolic time series analysis for pattern classification and anomaly detection. *IEEE Transactions on Industrial Informatics*, 18(4):2614–2621, 2021.

[9] A. Hammoudeh, I. Ghannam, H. Mubarak, E. Jean, V. Vandenbulcke, and S. Dupont. A neuro-symbolic approach for marine vessels power prediction under distribution shifts. In *2023 IEEE Jordan International Joint Conference on Electrical Engineering and Information Technology (JEEIT)*, pages 105–109, Amman, Jordan, 2023. IEEE, IEEE.

[10] P. Hitzler and M. K. Sarker. Neuro-symbolic artificial intelligence: The state of the art. *Neuro-Symbolic AI: The State of the Art*, 1(1):1–22, 2022.

[11] M. INOUE. Unsupervised anomaly detection in mixed processes using clusters. *Nippon Steel Technical Report*, 2024.

[12] Q. Li, H. Tan, Z. Jiang, Y. Wu, and L. Ye. Nonrecurrent traffic congestion detection with a coupled scalable bayesian robust tensor factorization model. *Neurocomputing*, 430:138–149, 2021.

[13] M. L. Menéndez, J. A. Pardo, L. Pardo, and M. d. C. Pardo. The jensen-shannon divergence. *Journal of the Franklin Institute*, 334(2):307–318, 1997.

[14] K. Pearson. Vii. note on regression and inheritance in the case of two parents. *proceedings of the royal society of London*, 58(347-352):240–242, 1895.

[15] P. J. Rousseeuw. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of computational and applied mathematics*, 20, 1987.

[16] M. K. Sarker, L. Zhou, A. Eberhart, and P. Hitzler. Neuro-symbolic artificial intelligence: Current trends. *Ai Communications*, 34(3):197–209, 2022.

[17] J. B. Slimane and M. B. Ammar. A novel approach of traffic congestion and anomaly detection with prediction using deep learning. *Journal of Electrical Systems*, 20(3s):2150–2159, 2024.

[18] R. L. Thorndike. Who belongs in the family. In *Psychometrika*. Citeseer, 1953.

[19] B. Yan, X. Ma, Q. Sun, and L. Shen. Physics-enhanced nmf toward anomaly detection in rotating mechanical systems. *IEEE Transactions on Reliability*, 2024.

A State Detection Model Architecture Brussels Use Case

Below the architecture of the GRU model for traffic state detection in the Brussels dataset is provided.

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 120, 3)]	0
gru (GRU)	(None, 120, 64)	13248
gru_1 (GRU)	(None, 64)	24960
dropout (Dropout)	(None, 64)	0
dense (Dense)	(None, 32)	2080
y1_output (Dense)	(None, 6)	198

Total params: 40,486
Trainable params: 40,486
Non-trainable params: 0

B State Detection Model Architecture PeMS-Bay Use Case

Below the architecture of the GRU model for traffic state detection in the PeMS-Bay dataset is provided.

Layer (type)	Output Shape	Param #	Connected to
input_3 (InputLayer)	[(None, 12, 1)]	0	[]
gru_4 (GRU)	[(None, 12, 64), (None, 64)]	12864	['input_3[0][0]']
dropout_3 (Dropout)	(None, 12, 64)	0	['gru_4[0][0]']
gru_5 (GRU)	[(None, 64), (None, 64)]	24960	['dropout_3[0][0]']
batch_normalization_2 (BatchNormalization)	(None, 64)	256	['gru_5[0][0]']
concatenate_1 (Concatenate)	(None, 192)	0	['batch_normalization_2[0][0]', 'gru_4[0][1]', 'gru_5[0][1]']
dense_2 (Dense)	(None, 32)	6176	['concatenate_1[0][0]']
y1_output (Dense)	(None, 6)	198	['dense_2[0][0]']

Total params: 44,454
Trainable params: 44,326
Non-trainable params: 128