Enhancing Continuous Risk Assessment: The Role of Safety Engineers in Early Hazard Identification

Anil Ranjitbhai Patel
RPTU Kaiserslautern-Landau
Germany
apatel@rptu.de

Peter Liggesmeyer
Fraunhofer IESE
Germany
peter.liggesmeyer@iese.fraunhofer.de

***Abstract*—As Automated Driving Systems (ADS) revolutionize the intelligent transportation landscape, ensuring unparalleled safety is increasingly essential. Traditional risk assessment methodologies, primarily designed for human-driven vehicles, struggle to adapt to the complex, ever-changing environment of ADS. This paper introduces a cyclic process aimed at augmenting continuous risk assessment for ADS, addressing the limitations of existing standards, which focuses on the functional safety of road vehicles but assumes static risk and the presence of a human driver for control and responsible for safety. Our proposed process transcends these limitations by integrating learning-based risk assessment, aiding safety engineers in early hazard detection for ADS development. The cornerstone of this approach is the Plan-Do-Train-Adjust-Assess cyclic process, which facilitates continuous improvement in risk assessment under diverse driving conditions. This method leverages advanced learning algorithms and integrates risk-specific contextual information, thus bridging traditional gaps in risk assessment. Critically, the process allows for the evaluation of severity and controllability to vary across different dynamic environment. This variability is determined by factors such as the operational domain, system complexity, and the evolving risk knowledge obtained through an iterative process. The insights gained from assessing severity and controllability aid in creating and refining essential safety mechanisms.**

***Index Terms*—Automated Driving System, Dynamic Risk Assessment, Severity, Controllability, Safety Engineer**

I. Introduction

Current automotive safety standards are inadequate for Automated Driving Systems (ADS) because they lack explicit definitions of unacceptable risk levels for these advanced systems. These standards, including ISO 26262 [1] and ISO 21448 [2], rely on implicit risk knowledge and traditional risk management methods, which are not sufficiently aligned with the unique risks and dynamic interactions of ADS operational environment. Additionally, the traditional focus on the driver as the primary responsible entity for vehicle safety does not apply to ADS, and the assignment of Automotive Safety Integrity Level (ASIL) ratings is complicated by this shift.

The Hazard Analysis and Risk Assessment (HARA) process in automotive safety often leads to inconsistent results, mainly due to challenges faced by safety engineers when assigning ASIL ratings to identified hazards. The subjective judgment of different safety engineers, personal biases, shaped by an engineer's background, experiences, or preferences, introduces a subjective element into these assessments [3]. This subjectivity, along with the lack of standardized criteria for evaluating hazard severity, controllability, and exposure, leads to varying interpretations and inconsistencies in risk assessments across different engineers [4].

Enhancing continuous risk assessment and early hazard identification for ADS presents multiple challenges [5, 6]. Firstly, establishing comprehensive scenario coverage is critical, as current standards may not fully capture the vast array of potential real-world scenarios an ADS might encounter, often focusing instead on a limited set of predefined conditions. Secondly, the integration of complex system interactions poses a challenge. Traditional approaches might struggle to holistically integrate the complex interactions between various subsystems of ADS, failing to analyze the risk introduced by these systems effectively. Thirdly, integrating and analyzing diverse data streams is essential yet challenging. Current frameworks may not effectively address this aspect of data integration and analysis. Another significant challenge is the adaptation of risk thresholds. Current standards often rely on rule-based or static risk thresholds defined by safety engineer, which may not be effective under all operational conditions. Lastly, validation under various operational conditions and continuous model improvement is essential. Standard validation processes may not adequately account for the wide variability in operational conditions faced by ADS and fail to specify measures, particularly in an open traffic.

Therefore, this paper introduces a cyclic Plan-Do-Train-Adjust-Assess (PDTAA) process inspired from [7] and shown in Fig. 1, which addresses those challenges to enhance continuous risk assessment and early hazard identification for ADS. The PDTAA process, with its iterative planning and simulation phases, offers extensive and evolving scenario coverage, adapting to new data and insights. It facilitates the integration of complex system interactions by continuously evaluating and adjusting these interactions during simulations. Moreover, the model's focus on robust data processing and supervised Machine Learning (ML) training allows for more effective integration and analysis of multifaceted data, to overcome rule-based for each and every situation, thereby enhancing the accuracy of risk assessment. This approach not only harnesses the predictive power of ML but also tailors it to the dynamic and varied requirements of runtime risk

2325-6664/24/$31.00 ©2024 IEEE
DOI 10.1109/DSN-W60302.2024.00027

assessment.

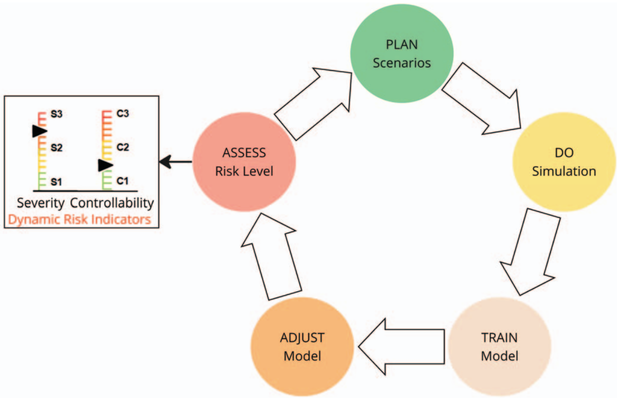

Fig. 1. Concept of Enhancing Continuous Risk Assessment

Unlike static risk assessment, this process adapts based on evolving data and contextual insights, leading to more nuanced and situation-appropriate risk assessments. The ADJUST phase of the PDTAA process allows for the dynamic recalibration of risk features and thresholds, based on continuous learning, to ensure their relevance and accuracy under changing conditions. The safety engineer plays a crucial role in ensuring continuous assessment and model improvement by selecting appropriate risk features and their thresholds to enhance adaptability. The prediction of risk assessment ratings can be utilized to signal to ADS developers the level of attention or investment required to mitigate the risk associated with a particular hazard through the development or enhancement of safety mechanisms.

II. Related Work

This section explores various studies, each contributing uniquely to the domain of risk assessment for ADS using diverse methodologies. These works collectively highlight the range of approaches for enhancing risk prediction, moving beyond traditional methods and can be grouped into four categories based on their implementation.

Firstly, scenario-based approaches focus on detailed scenarios, facilitating the identification of a wide range of risks, including those that might not be apparent in less nuanced methodologies [8]–[10]. These methods may struggle to adapt quickly to new information or changes in conditions that were not initially considered in the scenario design.

Secondly, simulation-based methods create realistic and controlled environments where various driving conditions and scenarios can be tested without the risks associated with real-world experimentation. These approaches reduce the need for expensive physical prototypes and real-world testing, significantly lowering development costs. Simulations allow for quick iterations over design changes, enabling faster refinement of ADS technologies and risk assessment models [11]–[13]. The accuracy of simulation-based techniques is contingent upon the models' fidelity. Inaccuracies in modeling can lead to misleading risk assessments.

Thirdly, iterative methods support continuous learning and improvement, ensuring that risk assessment models evolve in response to new data and insights. They offer the ability to dynamically adapt to changing conditions and incorporate feedback from real-world operations, enhancing the reliability of risk assessment process [14]–[16]. However, the need for continuous refinement and adaptation can introduce delays, as each iteration requires time to collect and analyze data before implementing changes.

Lastly, ML models excel at identifying complex patterns and dependencies in large datasets, which might be difficult or impossible for safety engineers to discern. These approaches automate the risk assessment process, significantly reducing the time and resources needed for manual analysis [17]–[21]. While these models are good at adapting to new scenarios, there is always a risk of overfitting to the training data, which can reduce the model's ability to generalize to new, unseen scenarios.

Integrating these approaches into a cohesive PDTAA process could enhance the efficiency of risk assessments in autonomous driving. The amalgamation of scenario-based, simulation-based, iterative-based, and learning-based approaches could forge an ADS risk assessment process that is not only more adept at identifying and mitigating risks but also proficient in learning from and adapting to the complexities of real-world driving scenarios. By harnessing the strengths of each method, such an integrated process would not only anticipate unforeseen risks but also quantify the risk.

III. Methodology

This paper introduces the PDTAA cyclic process, a method for enhancing continuous risk assessment and early hazard identification for ADS. It utilizes risk-specific context information from Highway Lane Following scenarios with Adaptive Cruise Control (ACC), which serves as the use case for this study.

The PDTAA process uniquely blends runtime and design time activities within its cyclic methodology, providing a thorough and comprehensive analysis. The runtime phases of the PDTAA process, namely the PLAN and DO phases dynamically simulate ADS behavior across varied operational contexts, while the ASSESS phase translates simulation outcomes into dynamic risk indicators, such as severity and controllability.

On the other hand, the design time phases, TRAIN and ADJUST, allow for in-depth analysis beyond the constraints of runtime processing. In these phases, ML models are rigorously trained and tested, allowing for an in-depth exploration and fine-tuning. Specifically, during the ADJUST phase, the model is fine-tuned with insights gleaned from simulation data. This step integrates qualitative assessments and expert opinions, thereby enriching the risk assessment process. Furthermore, this incorporation of expert evaluation into the ML models not only complements ML-based approaches that solely depend

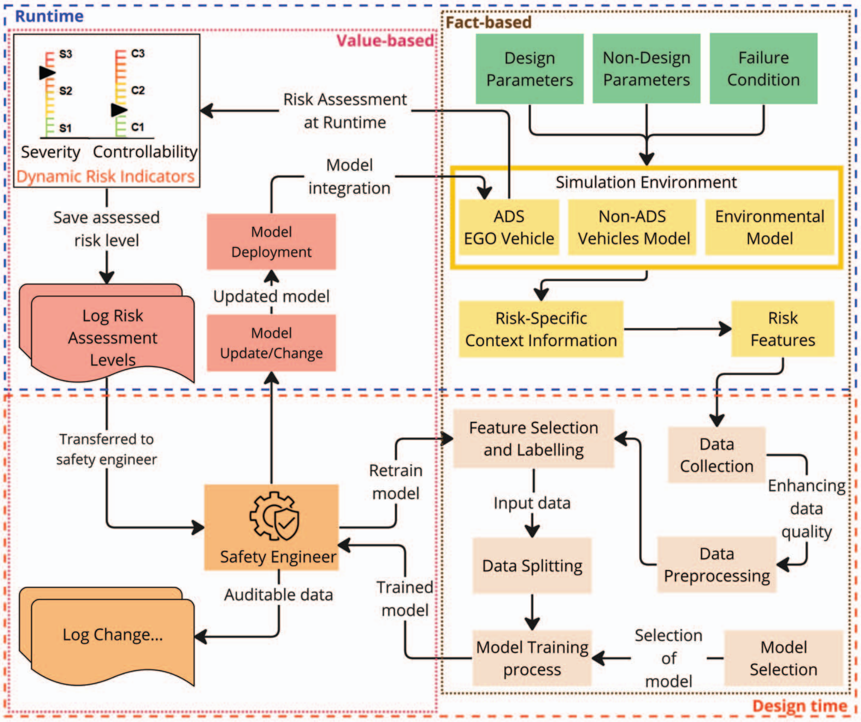

Fig. 2. Overview of the steps involved in the Plan-Do-Train-Adjust-Assess (PDTAA) Process in Risk Assessment

on data-driven insights but also significantly enhances the robustness of the models. Through this symbiotic relationship, the models benefit from a blend of empirical data and human expertise, ensuring a more comprehensive and resilient risk assessment framework.

As illustrated in Fig. 2, the PDTAA process combines runtime and design time activities, fact-based simulations, and value-based adjustments. This blend enables a balance between the immediacy of runtime data and the depth of expert analysis. The continuous assessment and model improvement mechanism ensure the system's efficacy and adaptability in autonomous driving's dynamic landscape. Delving deeper, each phase of the PDTAA process, starting with the PLAN-Scenarios, plays a crucial role in this risk assessment process.

A. *PLAN-Scenarios*

The PLAN-Scenario phase plays a crucial role in preparing for the DO-Simulation stage by identifying and categorizing essential parameters and conditions. This preparation is foundational and revolves around three core pillars: Design Parameters, Non-Design Parameters, and Failure Conditions. Each of these elements is vital for creating detailed and comprehensive simulation scenarios that accurately reflect potential real-world situations.

Design Parameters represent the controllable aspects of the system that directly influence both the dynamic behavior of the vehicle and the decision-making processes of the ADS. These parameters include, but are not limited to, sensor specifications, software algorithms, hardware configurations, and vehicle dynamics. For example, acceleration and deceleration capabilities of ADS, the positioning of sensors and the configuration of ACC controllers are considered design parameters. By adjusting these parameters during simulations, such as changing acceleration or deceleration capabilities of ADS, the impact of various design choices on the system's overall performance can be evaluated. This evaluation helps in understanding how changes in design can affect safety.

Non-Design Parameters, in contrast, include elements that are outside the direct control of ADS developers but still have a significant influence on the behavior of ADS. These parameters cover a wide range of factors such as environmental conditions, the behavior of other traffic participants (i.e. non-ADS), vehicle mass, the coefficient of friction, and road conditions. Although these parameters cannot be controlled, their effects on the ADS must be thoroughly understood and accounted for in simulation scenarios to ensure comprehensive risk assessment.

Lastly, Failure Conditions focus on potential system failures that could jeopardize safety. These are identified through meticulous analysis, often using HAZOP guiding words to pinpoint issues like the omission of sensor signals or wrong value being considered by actuators. By intentionally introducing these failure conditions into simulations, the response of the ADS to potential failures can be assessed. This approach enables the creation of a diverse set of conditions, including those that are rare or extreme, to ensure that the system is tested against all foreseeable hazardous events. This way, we can create a wide range of conditions, even rare and extreme ones, ensuring coverage of all foreseeable hazardous events, including low probability but high impact situations.

B. DO-Simulation

The DO-Simulation phase integrates the ADS-equipped ego vehicle model, non-ADS vehicle model, and environmental model. The ego vehicle model is comprehensive, incorporating an ACC system, braking mechanisms, sensors such as radars, cameras, LiDAR, and the necessary computational algorithms for operation. Simultaneously, the non-ADS vehicle model simulates the surrounding vehicles (e.g. lead and side vehicles) for various acceleration and deceleration capabilities, and unpredictable traffic behavior. Complementing these is the environmental model that replicates external conditions such as weather, road types, and traffic patterns.

Utilizing inputs from the previously identified parameters and failure conditions, simulations are conducted to replicate ADS interactions with the traffic ecosystem. The behavioral algorithms and dynamic environmental factors, such as rain, snow, and fog, enrich the complexity of the driving scenarios. This phase tests the ego vehicle's adaptability and decision-making processes against such unpredictable conditions, from weather variation to aggressive maneuvers by other vehicles. By exposing the ego vehicle to a range of hazards and operational challenges, the simulation provides valuable risk-specific context information and data, capturing at every 0.1-second interval.

The risk-specific context information includes static and dynamic variables and is used to derive quantifiable attributes called risk features. Static variables include attributes like maximum acceleration and deceleration of both the ego and lead vehicles, lane width, vehicle mass, road friction, and response times, whereas dynamic variables capture runtime changes such as the velocity of the ego, velocity of the lead, relative velocity, relative distance, relative deceleration, relative lateral speed, and relative lateral distance.

These variables are instrumental in calculating risk features such as Time To Collision (TTC) for immediacy, Time To Escape (TTE) for avoidance, Time To Stop (TTS) for responsiveness, Minimum Distance to Avoid Crash (MDAC) for proximity, Safe Distance (SFD) for buffering, Stopping Distance (STD) for halting, Deceleration Rate to Avoid Crash (DRAC) for mitigation, and Kinetic Energy (KE) for force. Such risk features, rooted in established studies [22]–[24], and it can be fed into ML models to identify, assess, and predict risks.

C. TRAIN Model

In this phase, the focus is on utilizing risk features generated during the DO phase to predict dynamic risk indicators through the training and testing of supervised ML models. Initially, data labeling is conducted to enable the ML model to distinguish between various risk levels accurately. This process involves tagging each data point with an output label that represents different levels of severity and controllability [1], which the model aims to predict based on the inputs (as shown in Table. I). Such supervised learning ensures that ML models can learn from input-output pairs and make informed predictions.

TABLE I
THRESHOLDS FOR CLASSIFICATION RULE-SET

Level	**TTC**	**MDAC**	**DRAC**
1	>15s	>30m	$\leq$ 1 m/s^2
2	<10s	<20m	>1 & $\leq$ 3 m/s^2
3	<5s	<10m	>3 & $\leq$ 5 m/s^2
4	<1s	<5m	>5 m/s^2

Severity Ratings				
Feature	TTC1	TTC2	TTC3	TTC4
MDAC 1	S0	S1	S2	S3
MDAC 2	S1	S1	S2	S3
MDAC 3	S2	S2	S2	S3
MDAC 4	S3	S3	S3	S3

Controllability Ratings				
Feature	DRAC1	DRAC2	DRAC3	DRAC4
MDAC 1	C0	C1	C2	C3
MDAC 2	C1	C1	C2	C3
MDAC 3	C2	C2	C2	C3
MDAC 4	C3	C3	C3	C3

Following data labeling, the dataset undergoes normalization and balancing as shown in Fig. 3. Normalization addresses datasets with risk features on vastly different scales, which can hinder the training process. By uniformly scaling the feature space, normalization facilitates smoother and faster convergence to optimal model parameters, thereby mitigating numerical instability and accelerating the learning process. On the other hand, balancing the dataset is crucial for ensuring that the model learns equally from all risk classes, regardless of their occurrence frequency. This balanced learning approach enhances the model's ability to generalize to unseen data, improving its predictive accuracy across various risk categories.

The upper plot in Fig. 3 illustrates a complex and interwoven landscape, underscoring the challenge of discerning clear, linear relationships for risk assessment. This complexity accentuates the necessity for ML algorithms, which are adept at uncovering subtle correlations within multifaceted data sets, correlations that might elude traditional rule-based risk assessments. With the data preprocessed through labeling, balancing, and normalization, the ML model proceeds to training. The training and testing process is adapted from a framework established in [19]. During this stage, the model learns the patterns and correlations within the prepared dataset that indicate different levels of severity and controllability.

The model selection assesses the effectiveness of several ML algorithms, including Support Vector Machines (SVM),

[1] In this paper, the research focuses solely on severity and controllability for the specified hazardous events, with the exposure rating considered as E4 (highly probable). The classification of controllability and severity as system attributes underscores their inherent relation to the system's design, operation, and potential failure modes [4].

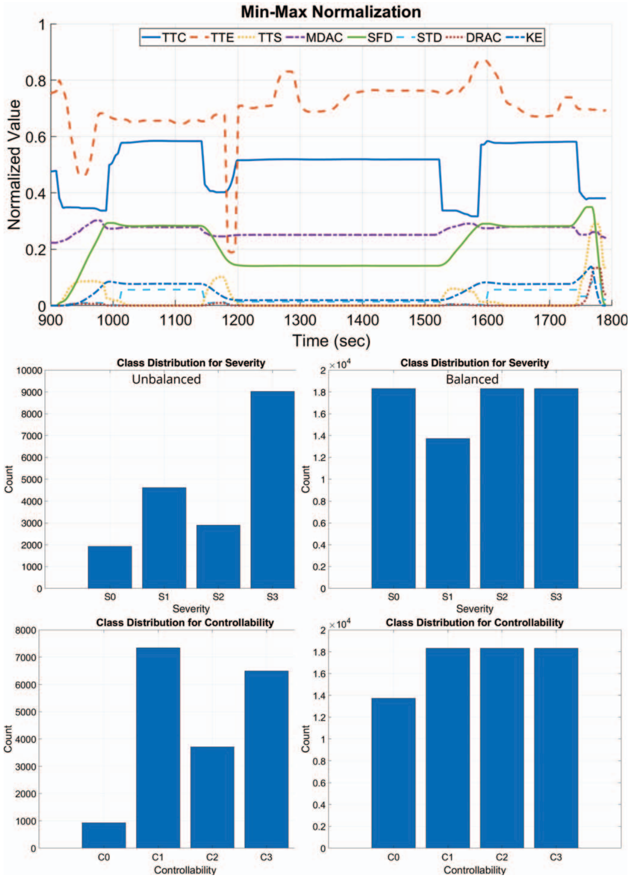

Fig. 3. Data Preprocessing: Normalization and Balancing

Random Forests (RF), Artificial Neural Networks (ANN), Gradient Boosting Decision Trees (GBDT) and others [2]. The decision to focus on these four models is based on their respective characteristics. While numerous models could be considered for comparative evaluation under different conditions, this study opts for simplicity by concentrating on SVM, RF, GBDT, and ANN.

SVMs demonstrate effectiveness in high-dimensional spaces with clear class separation but may falter in settings with substantial noise. In contrast, RFs are proficient in handling complex, non-linear scenarios through their decision tree ensemble, albeit with a lack of interpretability. ANNs are noted for their adaptability, efficiently identifying intricate patterns in large datasets, though they require significant data and computational resources. GBDT, on the other hand, excel in optimizing predictive accuracy by sequentially correcting errors from previous trees, making them highly effective in both linear and non-linear problem spaces.

The choice of model depends on the specific requirements of the driving scenario. Ultimately, the safety engineer, with the ability to assess the parallel performance of the models, is responsible for making the final decision regarding model performance. They can evaluate the models in parallel and make an informed judgment.

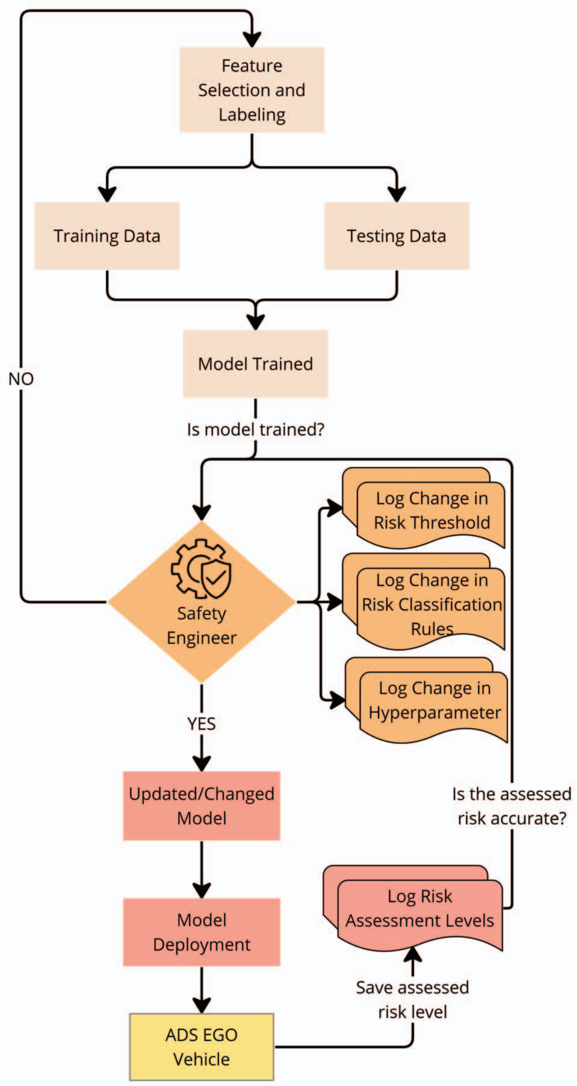
Feature Selection and Labeling
Training Data
Testing Data
Model Trained
Is model trained?
NO
Safety Engineer
Log Change in Risk Threshold
Log Change in Risk Classification Rules
Log Change in Hyperparameter
YES
Updated/Changed Model
Is the assessed risk accurate?
Log Risk Assessment Levels
Model Deployment
Save assessed risk level
ADS EGO Vehicle

Fig. 4. Focused View of the Safety Engineer's Integral Role

D. ADJUST Model

During this phase, the safety engineer rigorously evaluates the ML model's performance using metrics like accuracy, precision, recall, specificity, and F1 score, focusing on its runtime risk indicator prediction capabilities. If the model's predictions do not meet the predefined accuracy criteria ($>=$ 99%), it undergoes adjustments or retraining process, as shown in Fig. 4. This process may involve sessions with modified risk thresholds, changes in classification rules, or hyperparameter tuning to ensure the model meets performance criteria. The risk features utilized by the model may be refined

[2] Sufficient research exists on how these models operate within the scientific domain. Therefore, to maintain focus on the title, the methods are intentionally not discussed in detail and are used off-the-shelf in this research.

or adjusted for relevance and comprehensive representation of operational complexities. Modifications are made to the classification rules or thresholds for classifying risk labels, aiming to improve the accuracy of predictions and prevent biases in the model. Hyperparameters are also fine-tuned to achieve an optimal balance between bias and variance, aiming for a model that generalizes effectively to new unseen and unlabeled data. Continuous post-deployment monitoring and periodic reassessment based on feedback loops ensure the model's sustained performance and adaptability to evolving risk landscapes. Safety engineers also maintain comprehensive documentation of all changes, updates, and performance logs related to the ML model for future reference.

E. ASSESS Risk Level

In this phase, the refined ML model is integrated into the ADS for runtime risk assessment, focusing on detecting severity and controllability indicators in scenarios not previously encountered. These indicators allow safety engineers to establish safety goals for the necessary enhancements or modifications to ADS, including improvements to braking systems or alterations to sensor configurations. Monitoring the ML model's performance during runtime risk assessments introduces unique challenges, particularly when it encounters new, unlabeled data, and ground truths are unknown to the safety engineer.

To navigate these challenges, safety engineers employ data visualization tools to scrutinize the model's predictions against new data, facilitating the discovery of emerging patterns, trends, or anomalies indicative of previously unrecognized risk indicators. This understanding aids safety engineers in pinpointing unseen risk indicators and grasping the model's decision-making logic.

IV. Experimental Results

In our MATLAB-based implementation of the PDTAA process, our objective is to simulate and enhance the risk assessment process, taking the Highway Lane Following example as a reference [25].

TABLE II
Model Performance Metrics

Severity					
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Specificity**
GBDT	0.9992	0.9982	0.9993	0.9988	0.9998
ANN	0.9102	0.8679	0.9241	0.8951	0.9718
RF	0.9992	0.9984	0.9993	0.9989	0.9997
SVM	0.9937	0.9905	0.9938	0.9922	0.9976

Controllability					
Model	**Accuracy**	**Precision**	**Recall**	**F1-Score**	**Specificity**
GBDT	0.9989	0.9982	0.9993	0.9988	0.9996
ANN	0.8690	0.8162	0.8979	0.8551	0.9549
RF	0.9989	0.9981	0.9993	0.9987	0.9997
SVM	0.9907	0.9834	0.9940	0.9887	0.9969

We simulated multiple scenarios, ranging from 12 m/s to 35 m/s speeds for the ego vehicle, along with varying speeds of the lead vehicle to address the dynamic driving tasks of acceleration, deceleration, and braking. The acceleration rate for both the lead and ego vehicles was set at 2 m/s², and the deceleration rate for both was set at -2 m/s². In the real world, these rates can vary, but for the sake of simplifying our analysis, we chose to keep them the same in our study.

Our dataset did not include any failure conditions during training and testing. Our goal was to teach ML models about dynamic driving conditions by varying speeds and behaviors from surrounding traffic. The training dataset comprised 80%, and the testing dataset accounted for 20%. After the models were trained, they were tested against the dataset, and the performance metrics were calculated.

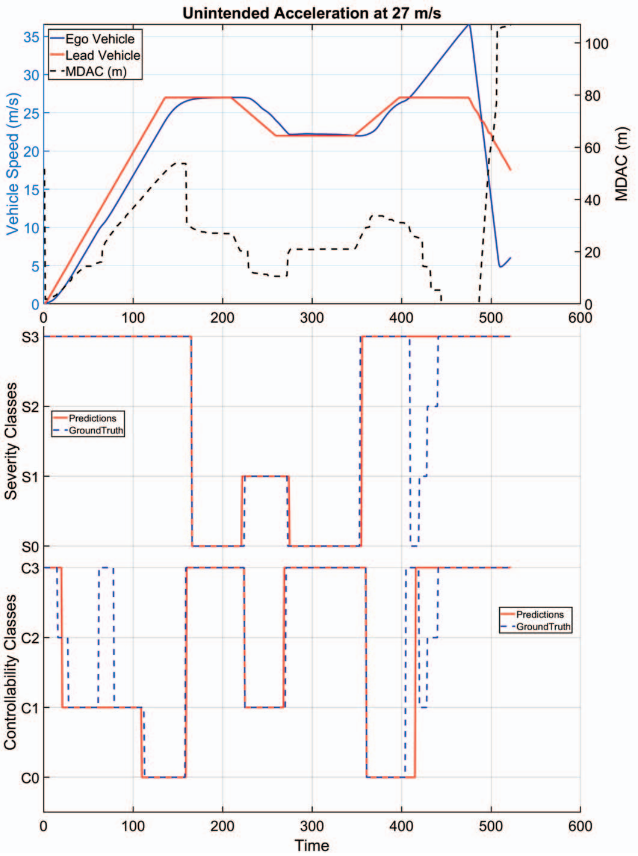

Fig. 5. Model Validation with Unintended Acceleration at 27 m/s

As shown in Table. II, following tuning, adjustments, and hyperparameter optimization by the safety engineer, the GBDT model outperformed the others across all five metrics, leading to its deployment in an ADS-equipped ego vehicle to predict dynamic risk indicators for scenarios not used in training and testing.

Fig. 5 illustrates the behavior of both ego and lead vehicles during acceleration, deceleration, and braking operations. Initially, the severity was already high as the ego vehicle was following the lead vehicle with a very short distance, posing a risk of crash. In non-failure scenarios, the GBDT model predicts the severity and controllability based on the driving behavior of the lead vehicle. Notably, around the 400-second mark, a “wrong value” failure was injected, leading to an immediate spike in controllability ratings, as predicted by the GBDT model. This injection further exacerbated the situation, with the minimum distance required to avoid a crash increasing. In this scenario, while the lead vehicle was decelerating, the ego vehicle experienced unintended acceleration at a speed of 27 m/s. Consequently, this failure led to a front-end collision, as indicated by the elevated severity and controllability indicators. While numerous such scenarios can be simulated and evaluated to understand the ADS’s behavior comprehensively, due to page limits, only one scenario is shown here.

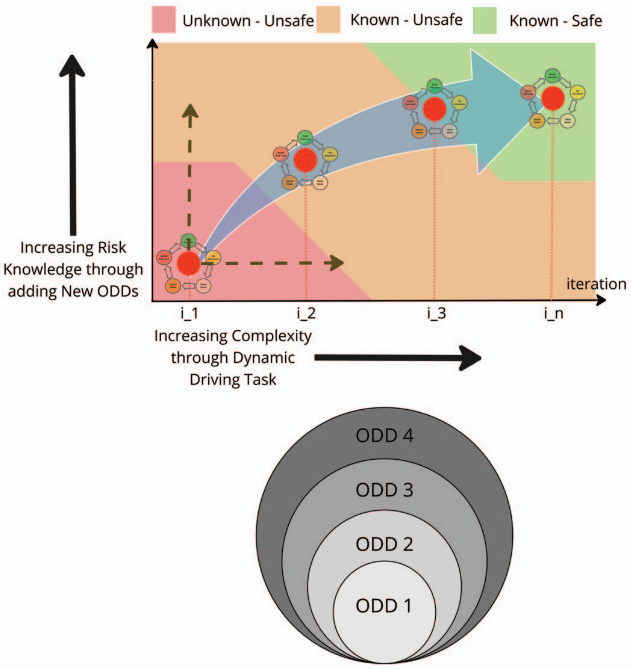
Unknown - Unsafe
Known - Unsafe
Known - Safe
Increasing Risk Knowledge through adding New ODDs
iteration
i_1
i_2
i_3
i_n
Increasing Complexity through Dynamic Driving Task
ODD 4
ODD 3
ODD 2
ODD 1

Fig. 6. Iterative Enhancement of Risk Knowledge Through PDTAA process

The process repeats cyclically once safety updates are made, with each iteration building on the accumulated risk knowledge and risk indicators from previous cycles. The cyclic process continues until all unknown-unsafe situations within a particular ODD are explored. As depicted in Fig. 6, the process begins with ODD 1 in the first iteration, encountering the first failure, followed by subsequent iterations for each addressing additional failures. Once all failures are attempted, a new ODD included. For instance, if highway lane following straight road driving is considered in ODD 1, then ODD 2 might be a curved road, ODD 3 an uphill scenario, ODD 4 a downhill scenario, and so forth. The iterations on the X-axis correspond to increasing complexity in DDT; for example, the first iteration might cover basic acceleration/deceleration and braking operations, the second iteration could introduce steering maneuvers, the third might involve overtaking maneuvers, and the fourth could include exiting and entering the construction zone.

This way runtime risk indicators can assist safety engineers in augmenting ADS design to increase redundancy or incorporate fault-tolerant mechanisms, relying on empirical evidence and objective ratings rather than subjective and biased assessments that do not fully consider system capabilities and environmental factors.

V. Limitation

The PDTAA process encounters several challenges within continuous risk assessment, primarily due to the complexity and evolving nature of ADS.

First, the intricate interconnectedness of systems within ADS presents a significant hurdle. These systems include a variety of components, such as advanced algorithms and runtime data processing modules. The PDTAA process may struggle to accurately model and predict the outcomes of these complex interactions, potentially leading to unforeseen dependencies and emergent behaviors. Second, fully understanding both the functional and non-functional requirements in an ever-changing environment is difficult. The PDTAA process might not entirely capture these requirements, especially in unique or rare scenarios. Third, the integration of various sensors and subsystems, crucial for precise risk modeling, introduces additional complexity. Handling data from diverse sources, each with its own set of nuances and potential inaccuracies, poses a challenge. Fourth, the extensive computational demands required to model a broad range of driving conditions and runtime interactions can impede the effectiveness and efficiency of the PDTAA process. This may result in delays in risk assessment and decision-making.

Despite these challenges, it is important to recognize the strengths and advantages of the PDTAA process. While there may be an over-reliance on ML algorithms, these algorithms offer significant benefits in terms of automation, scalability, and the ability to rapidly process vast quantities of data. They are capable of identifying complex patterns and dependencies that might elude traditional rule-based systems. Furthermore, well-trained and validated ML models can deliver valuable insights and predictions in runtime, enhancing risk assessment for ADS. Their adaptability and continuous learning capabilities render them highly suitable for dynamic environments. Efforts to improve the transparency and interpretability of “black-box” models are ongoing in the field of ML model interpretability, potentially enhancing this aspect of the PDTAA process. Although gaps in training data present challenges, advancements in data collection and simulation techniques are helping to close these gaps and enhance model robustness. The continual refinement and optimization of ML models contribute to their growing effectiveness in the PDTAA process over time.

VI. Conclusion

The PDTAA process offers an approach to improve HARA for ADS. This process overcomes the shortcomings of traditional HARA, which often depends on subjective judgments and falls short in providing uniform criteria for hazard evaluation.

The PDTAA process incorporates several key enhancements, including comprehensive scenario coverage, data integration, computing risk features, and continuous improvement. One of its strengths is the ability to encompass a broad spectrum of real-world scenarios, surpassing the capabilities of traditional methods. In terms of data integration and risk features, the process effectively leverages diverse data, with a focus on the TRAIN phase that employs ML model to predict risk indicators under various operational conditions using risk features and risk thresholds.

Furthermore, the process places significant emphasis on validation and ongoing refinement, especially in the ASSESS phase. This commitment ensures that the ML model remains effective based on runtime data, contributing to its overall reliability and performance in continuous risk assessment and early hazard detection. The predicted severity and controllability ratings can be utilized to trigger safety mechanisms or guide the allocation of resources for creating new safety mechanisms. For instance, in a highway lane-following scenario with ACC, the PDTAA process can dynamically evaluate the risk of front-end collisions during unintended acceleration.

These ratings can then be employed to mitigate risk through runtime safety mechanisms, as presented in [26, 27]. This proactive approach goes beyond reacting to the immediate driving environment; it adapts to changing conditions. This adaptability enables the vehicle to make risk informed decisions, ensuring consistent safe behavior at all times.

ACKNOWLEDGMENT

This paper discusses research conducted within the framework of a broader project sponsored by the Performance Center Simulation and Software-based Innovation (Leistungszentrum Simulations- und Software-basierte Innovation) based in Kaiserslautern, Germany.

REFERENCES

[1] ISO 26262-1:2018, ”Road vehicles - Functional safety - Part 1: Vocabulary,” 2018.

[2] ISO 21448:2022, ”Road vehicles — Safety of the intended functionality,” International Organization for Standardization, 2022.

[3] T. Aven, ”On how to define, understand and describe risk,” in Reliability Engineering & System Safety, vol. 95, no. 6, pp. 623-631, 2010.

[4] S. Khastgir, S. Birrell, G. Dhadyalla, H. Sivencrona, and P. Jennings, ”Towards increased reliability by objectification of HARA of automated automotive systems,” Safety Science, vol. 99, pp. 166-177, 2017.

[5] F. Beringhoff, J. Greenyer, C. Roesener, and M. Tichy, ”Thirty-one challenges in testing automated vehicles: Interviews with experts from industry and research,” in 2022 IEEE Intelligent Vehicles Symposium, pp. 360-366, IEEE, 2022.

[6] W. M. D. Chia, S. L. Keoh, C. Goh, and C. Johnson, ”Risk assessment methodologies for autonomous driving: A survey,” IEEE Transactions on Intelligent Transportation Systems, vol. 23, no. 10, pp. 16923-16939, 2022.

[7] ISO 9001: 2015 Quality Management Systems-Requirements. Guidance Document.

[8] J. Krampe and M. Junge, ”Injury severity for hazard & risk analyses: calculation of ISO 26262 S-parameter values from real-world crash data,” Accident Analysis & Prevention, vol. 138, p. 105321, 2020.

[9] E. De Gelder, H. Elrofai, A. Khabbaz Saberi, J. P. Paardekooper, O. Op Den Camp, and B. De Schutter, ”Risk quantification for automated driving systems in real-world driving scenarios,” IEEE Access, vol. 9, pp. 168953-168970, 2021.

[10] J. Reich and M. Trapp, ”SINADRA: towards a framework for assurable situation-aware dynamic risk assessment of autonomous vehicles,” in 2020 16th European Dependable Computing Conference, pp. 47-50, IEEE, 2020.

[11] B. Wu, Y. Yan, D. Ni, and L. Li, ”A longitudinal car-following risk assessment model based on risk field theory for autonomous vehicles,” International Journal of Transportation Science and Technology, vol. 10, no. 1, pp. 60-68, 2021.

[12] A. Duracz, A. Aljarbouh, F. A. Bartha, J. Masood, R. Philippsen, H. Eriksson, J. Duracz, F. Xu, Y. Zeng, and C. Grante, ”Advanced hazard analysis and risk assessment in the ISO 26262 functional safety standard using rigorous simulation,” In: Cyber Physical Systems, pp. 108-126, Springer International Publishing, 2020.

[13] S. Hallerbach, Y. Xia, U. Eberle, and F. Koester, ”Simulation-based identification of critical scenarios for cooperative and automated vehicles,” SAE International Journal of Connected and Automated Vehicles, vol. 1, no. 2018-01-1066, pp. 93-106, 2018.

[14] F. Warg, M. Gassilewski, J. Tryggvesson, V. Izosimov, A. Werneman, and R. Johansson, ”Defining autonomous functions using iterative hazard analysis and requirements refinement,” in Computer Safety, Reliability, and Security: SAFECOMP Springer International Publishing, 2016.

[15] C. Fayollas, H. Bonnin, and O. Flebus, ”SafeOps: A concept of continuous safety,” in 2020 16th European Dependable Computing Conference, pp. 65-68, IEEE, 2020.

[16] B. Kramer, C. Neurohr, M. Büker, E. Böde, M. Fränzle, and W. Damm, ”Identification and quantification of hazardous scenarios for automated driving,” in International Symposium on Model-Based Safety and Assessment, pp. 163-178, Cham: Springer International Publishing, 2020.

[17] P. Feth, M. N. Akram, R. Schuster, and O. Wasenmüller, ”Dynamic risk assessment for vehicles of higher automation levels by deep learning,” in Computer Safety, Reliability, and Security: SAFECOMP Springer International Publishing, 2018.

[18] C. Katrakazas, M. Quddus, and W.-H. Chen, ”A new integrated collision risk assessment methodology for autonomous vehicles,” Accident Analysis & Prevention, vol. 127, pp. 61-79, 2019.

[19] A. R. Patel and P. Liggesmeyer, ”Machine learning-based dynamic risk assessment for autonomous vehicles,” in 2021 International Symposium on Computer Science and Intelligent Controls, pp. 73-77, IEEE, 2021.

[20] P. Mundt, I. Kumara, W. Van Den Heuvel, D. A. Tamburri, and A. S. Andreou, ”Knowgo: An adaptive learning-based multi-model framework for dynamic automotive risk assessment,” in International Symposium on Business Modeling and Software Design, pp. 268-278, Cham: Springer International Publishing, 2022.

[21] M. Khatun, R. Jung, and M. Glaß, ”Scenario-based collision detection using machine learning for highly automated driving systems,” Systems Science & Control Engineering, vol. 11, no. 1, p. 2169384, 2023.

[22] J. Weast, M. Elli, and I. Alvarez, ”To err is human: The role of human-derived safety metrics in an age of automated vehicles,” SAE Technical Paper 2021-01-0875, 2021.

[23] J. Wishart, S. Como, M. Elli, B. Russo, J. Weast, N. Altekar, E. James, and Y. Chen, ”Driving safety performance assessment metrics for ADS-equipped vehicles,” SAE International Journal of Advances and Current Practices in Mobility, vol. 2, no. 2020-01-1206, pp. 2881-2899, 2020.

[24] L. Westhofen, C. Neurohr, T. Koopmann, M. Butz, B. Schütt, F. Utesch, B. Neurohr, C. Gutenkunst, and E. Böde, ”Criticality metrics for automated driving: A review and suitability analysis of the state of the art,” Archives of Computational Methods in Engineering, vol. 30, no. 1, pp. 1-35, 2023.

[25] The MathWorks Inc., ”MATLAB version: 9.13.0 (R2022b),” Natick, Massachusetts: The MathWorks Inc., 2022. [Online]. Available: https://www.mathworks.com

[26] M. Trapp, D. Schneider, and G. Weiss, ”Towards safety-awareness and dynamic safety management,” in 2018 14th European Dependable Computing Conference, pp. 107-111, IEEE, 2018.

[27] A. R. Patel, N. B. Haupt, and P. Liggesmeyer, ”A Conceptual Framework of Dynamic Risk Management for Autonomous Vehicles,” in New Trends in Intelligent Software Methodologies, Tools and Techniques, pp. 475-486, IOS Press, 2022.