Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.Doi Number

A Spectrogram-Based Approach for Tire Type Classification Using Pass-By Noise

Mustafa Demetgul[1], Lukas Schwemer[1], and Sanja Lazarova-Molnar[1]

[1]Institute of Applied Informatics and Formal Description Methods (AIFB) KIT - Karlsruhe Institute of Technology

Corresponding author: Mustafa Demetgül (e-mail: mustafa.demetguel2@kit.edu).

ABSTRACT This study presents a non-intrusive approach for tire type classification using pass-by noise recorded from a single roadside microphone. Unlike camera-based or sensor-based systems, the proposed method enables contactless monitoring under real traffic conditions without requiring any modifications to the vehicle. The approach is based on the assumption that tire–road interaction produces characteristic acoustic patterns that can be used to distinguish between different tire types. The recorded signals are first filtered using a 400–5000 Hz bandpass filter to focus on the most informative frequency range. They are then segmented using fixed window sizes with overlapping intervals and transformed into Log-Mel spectrograms to represent their time–frequency structure. The effects of window size, overlap, and preprocessing parameters are systematically investigated. Data augmentation is applied using a polyphase-based downsampling strategy to increase variability. The framework is evaluated on the TyRoN Tyre Road Noise dataset using a Leave-One-Group-Out validation strategy to evaluate performance across different operational conditions. The best results are obtained with a hybrid approach combining ResNet-based feature extraction and an Extra Trees classifier, achieving up to 96.88% accuracy across six tire classes.

INDEX TERMS: *Pass-by noise, tire classification, spectrogram, machine learning, deep learning, hybrid ML architecture*

I. INTRODUCTION

Road safety remains a critical challenge in modern transportation systems. Despite continuous advancements in vehicle automation and advanced driver assistance systems, accident rates associated with loss of traction, reduced braking performance, and adverse weather conditions remain significant [1], [2], [3].

In this context, the condition and type of tires represent a fundamental yet often under-monitored safety factor. Incorrect tire selection (e.g., summer tires used under winter conditions), excessive tread wear, or the use of non-compliant tire types can significantly increase braking distance, reduce traction, and compromise vehicle stability [4], [5]. Previous studies have shown that tread depth and seasonal tire mismatch substantially influence wet braking performance and accident probability [6], [7], [8].

In many countries, including Germany, seasonal tire regulations are enforced to improve safety during adverse weather conditions. However, current enforcement mechanisms rely primarily on manual inspection during roadside checks. This approach is time-consuming, subjective, and not scalable for large traffic volumes [9], [10].

Several studies have employed computer vision techniques to analyze tire tread patterns and estimate wear conditions using convolutional neural networks. While these camera-based approaches can achieve high classification accuracy under controlled conditions, they are often sensitive to lighting variations, occlusion, dirt contamination, and camera angle. In addition, reliable detection typically requires low vehicle speeds or stationary inspection setups, limiting their applicability in real-world traffic scenarios [11], [12], [13], [14].

Beyond vision-based systems, alternative approaches such as Tire Pressure Monitoring Systems (TPMS) [15], vibration-based tire condition monitoring [16], and embedded smart tire sensors [17] have been investigated. However, these methods generally require direct instrumentation, vehicle integration, or sensor installation, which restricts scalability for large-scale traffic monitoring applications.

Tire-road interaction produces characteristic acoustic signatures. Underlying noise generation mechanisms include tread block impact excitation, air pumping effects in tread gaps, adhesion-slip events at the contact surface, and structural vibrations of the tire carcass [18], [19]. Collectively, these mechanisms produce distinct spectral and temporal features in the emitted noise signal. When captured using roadside crossing measurement devices, these acoustic emissions naturally encode distinctive information about tire structure and surface characteristics without the need for direct visual access or vehicle instrumentation [20]. Unlike camera-based systems or embedded smart sensors, acoustic monitoring enables non-intrusive and contactless data collection under realistic traffic conditions. Consequently, these acoustic fingerprints offer a scalable and infrastructure-based opportunity for automated tire classification in dynamic traffic environments [21], [22]. These measurable acoustic characteristics form the basis for data-driven learning models capable of distinguishing between different tire types under real-world operating conditions. Integrating AI-based tire detection into intelligent transportation systems (ITS) could enable large-scale, automated compliance monitoring, early risk detection, and data-driven traffic safety strategies [23], [24], [25], [26].

While there is growing interest in tire assessment, most of the existing techniques rely on computer vision or other sensor-based inspection setups. These require the car to be stationary or to be measured in a controlled environment. This limits their usefulness in real-world traffic monitoring situations. A pass-by audio tire classification setup, where a car's tires can be classified only using a roadside microphone, is a non-intrusive, low-cost, and deployable framework. It not only enables tire classification but also avoids complex sensors, vehicle-mounted hardware, or standstill inspection procedures, supporting future scalable roadside tire monitoring applications.

In sound-based classification approaches, some preprocessing steps are used to improve the quality of the signal, emphasize discriminative time-frequency structures and increase stable performance of learned features under varying conditions. Bandpass filters are frequently used to remove unwanted frequencies while preserving the band with the most discriminative features. This initial conditioning step is often used to filter out low and high-frequency artefacts while preserving the mid-band where discriminative noise is most informative [27], [28], [29].

Numerous studies use spectrogram representation in acoustic scene and environmental sound classification using spectrograms either as a direct machine learning inputs or to enable CNN-based architectures to learn discriminative time-frequency patterns. Their results establish spectrogram-derived features as a robust and widely viable preprocessing step in audio recognition across diverse conditions [30], [31], [32], [33], [34], [35].

To increase data size, polyphase-style down-sampling augmentations can be used to generate additional samples with complementary phase components. The additional training views around the same underlying signal preserve the original samples' core while expanding the database without creating exact duplicates. This is supported by formal polyphase signal-processing and empirical augmentation studies showing the improved performance of time-series datasets expanded using structured transformations [36], [37].

In the area of audio classification, traditional machine learning as well as convolutional neural networks are frequently used to learn discriminative patterns either directly from the data or from engineered features and spectrogram representation. When pipelines produce fixed handcrafted feature vectors, classical machine learning methods are often employed as they offer strong performance and fast training/inference. Particularly, tree-based ensemble methods such as LightGBM, XGBoost, Random Forest, and Extra Trees have found wide adoption due to their promising generalization under limited data [38], [39], [40], [41].

On the CNN side, spectrograms are often used as an image-like input for acoustic event and sound classification, spanning applications using pass-by noise for roadside monitoring to speech processing. This enables established architectures such as LeNet or AlexNet, but also modern feature extraction backbones like ConvNext, EfficientNet, and ResNet to learn discriminative patterns in the spectrogram on a local and global scale [32], [42], [43], [44], [45]. Recently, transformer-based audio models such as AST and BEATs have shown strong performance in acoustic classification tasks by capturing long-range temporal dependencies. However, these architectures typically require larger datasets and higher computational resources. Therefore, this study focuses on lightweight CNN and hybrid approaches suitable for limited-data and roadside deployment scenarios[46], [47].

To leverage the strengths of CNNs' ability to extract meaningful patterns from spectrogram time-frequency representations and the efficiency of classical machine learning algorithms when working with handcrafted feature vectors, many sound classification studies use hybrid approaches. These combine the deep feature extraction with lightweight decision models to improve accuracy and promising generalization, as demonstrated in studies using a mix of CNNs and ensemble learners or SVMs [48], [49], [50].

Table I summarizes the closest related work. Kongrattanaprasert [24] uses pass-by noise from vehicles traveling at 30 – 60 km/h on dry or wet roads. He distinguishes between winter and summer tires using a simple autocorrelation-based thresholding method rather than deep learning in a binary classification problem. The author reports 100% accuracy in a one-hour experimental setup. Demetgul et al. [49] shared their initial results in earlier studies, which laid the foundation for the present work. In this study, a classification accuracy of 90% was achieved using only machine learning algorithms; however, this performance was found to be insufficient [51]. Meng et al. [48] utilize pass-by noise to classify vehicle type, speed range, and movement direction. By using MFCC spectrograms in a hybrid ResNet50 combined with a

IEEE Access
Multidisciplinary ⁝ Rapid Review ⁝ Open Access Journal

LightGBM setup, they report accuracy up to 99% on the IDMT-Traffic dataset, demonstrating the feasibility of pass-by noise-based classification. Ashhad et al. [25] also use pass-by noise from the IDMT-Traffic dataset to classify vehicle sub-types such as car, truck, motorcycle, and no vehicle. The authors employ a CNN to classify MFCC, GFCC, and mel-spectrogram-based samples, reporting accuracies of up to 98.95%, extending the research done on the IDMT-Traffic dataset. Taking together, these prior works show the feasibility of using pass-by noise for classification. While they report high accuracies, [25], [48] are not tire-focused and instead perform related vehicle classification. Kongrattanapraserts [24] work is restricted to a simple binary classification setup, using a simple autocorrelation approach.

TABLE I: COMPARES RELATED WORK TO OUR CONTRIBUTION

Ref.	Method	Contribution	Acc.
[24]	Pass-by acoustic tire-noise analysis; autocorrelation-based threshold classifier	Acoustic tire-type classification of winter and summer tire. Same vehicle, road, and no acceleration	100%
[48]	Classification of MFCCs using hybrid ResNet50 + LightGBM setup	Classification of vehicle type, speed range, and movement direction from pass-by audio using the IDMT-Traffic dataset	99%
[51]	Used Log-Mel spectrogram-based classification using ML algorithms	Classified 6 tire types and 2 pressures solely using pass-by vehicle noise on the TyRoN dataset	90%
[52]	Signal-processing-based passenger-car detector, Log-Mel-spectrogram features, and SVM / MLP classifiers	Classified studded vs. non-studded passenger-car tires from in-road acoustic recordings	96.5%
Ours	Log-Mel spectrogram-based classification using ML, CNN, and hybrid architectures with strict LOGO validation	Classified 6 tire types solely using pass-by vehicle noise recorded from a single roadside microphone under multiple vehicles, road surfaces, speed ranges, and acceleration conditions	96.88%

Unlike many existing studies that focus on vehicle classification, binary tire categorization, or controlled sensor-based inspection setups, this work investigates a more challenging multi-class tire classification problem using only pass-by noise recorded from a single roadside microphone. Rather than proposing a fundamentally novel neural network architecture, the contribution of this study lies in the systematic integration and evaluation of signal processing, handcrafted feature engineering, CNN-based feature learning, and hybrid feature fusion within a realistic roadside monitoring framework.

To fill this gap, this work makes the following contributions. First, it extends the existing literature by addressing a more challenging six-class tire classification problem using only pass-by noise recorded from a single roadside microphone, enabling a non-intrusive and potentially scalable monitoring framework. Second, a comprehensive framework is proposed that combines signal processing, handcrafted feature extraction (Log-Mel spectrograms and PCA), and deep learning-based representations, allowing a systematic comparison between classical machine learning, convolutional neural networks, and hybrid approaches.

Third, we demonstrate that a hybrid architecture combining deep feature extraction with ensemble-based machine learning significantly improves classification performance under the evaluated operational conditions.

To improve robustness and provide a more realistic assessment of generalization, the dataset was collected under diverse real-world conditions, including multiple vehicles, road surfaces, varying speed ranges, and different acceleration profiles.

In addition, a strict Leave-One-Group-Out (LOGO) validation strategy was employed, where each test split consists of entirely unseen operational conditions. This setup prevents data leakage and provides a more realistic assessment of generalization across evaluated operational conditions.

The dataset includes measurements from four different vehicles (Volkswagen ID.4, Porsche Taycan, Volkswagen e-Golf, and Audi Q8) under multiple speed ranges (45–80 km/h, 80–100 km/h, and 100–120 km/h) and acceleration levels (1, 2, and 3 m/s² at 50 km/h).

Furthermore, six different tire types are considered, enabling a realistic and diverse evaluation scenario. Recent work has demonstrated the feasibility of tire-road analysis when implemented as an end-to-end pipeline, covering data acquisition, signal processing, feature engineering, and model inference [53].

Building on this foundation, the proposed framework leverages pass-by acoustic measurements as a non-intrusive alternative to vision-based or sensor-based approaches. Unlike close-proximity methods requiring vehicle-mounted sensors, this approach enables scalable data acquisition under real traffic conditions while preserving discriminative tire–road acoustic signatures [48].

The remainder of this paper is organized as follows. Section 2 describes the data collection and preprocessing pipeline. Section 3 presents the experimental methodology, including pre-experiments and model configurations. Section 4 discusses the results, and Section 5 concludes the paper.

II. DATA PREPARATION

The following chapter details the dataset and preprocessing pipeline used for our main line of experiments. First, we detail how our dataset was recorded, including vehicles, driving scenarios, and classes. The chapter then explains the entire preparation pipeline covering window extraction, band-pass filtering, Log-Mel spectrogram generation, augmentations, and PCA for the classical machine learning models.

A. DATA COLLECTION

The data used in this project were recorded in the TyRoN Tyre Road Noise research project, which investigates tire-road noise under controlled conditions. The data were recorded on two different roads: one ISO 10844-compliant test surface and one non-ISO asphalt surface, enabling evaluation under both standardized and real-world road conditions. Four different cars were used to record the data: a Volkswagen ID.4, a Porsche Taycan, an Audi Q8, and a Volkswagen e-Golf. Six different tires were used across these four vehicles to record the data, including Continental PremiumContact 6 AO (Pre), Continental EcoContact 6 Q (Eco), Hankook Ventus S1 evo3 (Ven), Pirelli P Zero R (Pze), BF Goodrich Summer RTT (Sum), and Uniroyal Rainsport 5 (Rain). Figure 1 shows the vehicle, tire, and location setup for our data collection. The measurements were taken at different speeds and accelerations.

FIGURE 1: The vehicle, tire, and location setup used for recording the data.

From constant speed driving at 45 km/h, 50 km/h, 60 km/h, 80 km/h, and 100 km/h, to dynamic speeds starting at 50 km/h with accelerations of 1 m/s², 2 m/s², and 3 m/s². This recording setup ensures capturing both stationary and transient tire-road interaction effects.

While the broader TyRoN Tyre Road Noise research data includes several sensing modalities, such as tire cavity microphones, roadside arrays, and thermal measurements, this work focuses solely on the pass-by noise recorded by a stationary microphone. This enables the investigation of a low-cost and non-intrusive tire classification framework under realistic traffic conditions.

B. SAMPLE PREPARATION

Figure 2 shows a 3x3 matrix of a few raw noise recordings. These signals make up the raw basis of our data used to train different models in our main line of experiments. After a data cleaning process including the removal of invalid files and duplicate files and a harmonization of channel names, a total of 207 recordings were ready for preprocessing. To increase the sample count, a windowing strategy was adopted. By using shifted 2s windows, multiple samples could be extracted per file. An overlap of 60% was used for our main line of experiments; a pre-experimentation study, detailed later in the Methodology chapter, determined these optimal parameters. This strategy increases the number of usable samples while preserving temporal continuity. Since tire-road noise signals are nonstationary and continuous, and their spectral characteristics may vary within one pass-by recording, this strategy was particularly beneficial.

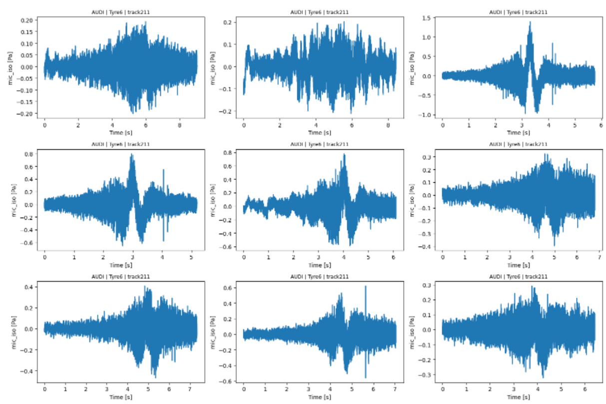


FIGURE 2: The raw waveforms of our tire-pavement interaction noise.

By extracting multiple overlapping segments, the dataset is enriched through the generation of partially overlapping signal segments. This results in a total of 1,082 samples. After window extraction, each signal segment was filtered by applying a fourth-order Butterworth band-pass filter. The cutoff frequencies were set between 400 Hz and 5000 Hz,

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3714728

suppressing very low- and high-frequency components, emphasizing the information-rich mid-section for the targeted classification task. While the signal is not completely cut off at these values, the signal surrounding them is gradually weakened. This step further contributes to the spectrogram generation, detailed in a later section, by reducing spectral contributions outside the region of interest.

The selected 400–5000 Hz range was motivated by prior tire-road noise studies and preliminary spectral investigations on the TyRoN dataset [54]. Previous studies have shown that discriminative tire-road interaction patterns are predominantly concentrated within the mid-frequency region, while low-frequency components are often dominated by vehicle body vibrations and environmental noise. Preliminary spectrogram analyses further indicated that the selected range contains the most discriminative tire-dependent acoustic patterns across the investigated tire classes [55], [56], [57].

C. POLYPHASE AUGMENTATIONS

Since CNNs usually require a large amount of data for training, we already applied a windowing approach, increasing our sample count from 207 files to 1,082 data examples. To further increase the diversity of the training data, we employ an optional polyphase down-sampling augmentation in some lines of our main experiments. The intuition is to artificially reduce a sensor sampling rate by 50%, preserving the general frequency patterns while losing some of its fine-grained information. Figure 3 illustrates a representative example of the data collection setup.

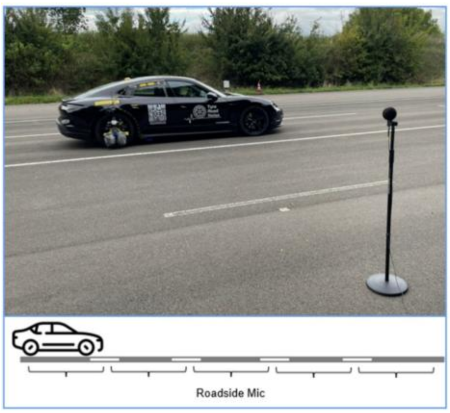


FIGURE 3: Roadside pass-by noise data collection experimental setup.

For each window, we split the signal into even- and odd-indexed samples. This produces two additional versions of the same segment. While their general structural pattern remains the same, we create additional highly similar samples without directly copying the original. Accordingly, the augmented samples are not intended to represent statistically independent observations but rather to improve the robustness of the models against small phase variations while preserving the physical characteristics of the original tyre–road noise signals. This is specifically attractive in this low-file setting, since it does not require synthetic spectral manipulations and remains closely tied to the underlying waveform. Starting from 1,082 base datapoints, we increase the total number of training samples to 3,246.

During the subsequent LOGO evaluation, the augmented samples retain the operating-condition label of their corresponding original recording. Consequently, all augmented samples belonging to the held-out speed or acceleration condition remain exclusively within the corresponding test set, while augmented samples from the remaining operating conditions are used only for training. This ensures that no augmented samples originating from the held-out operating condition are included in the training data.

D. SPECTROGRAM GENERATION

After the band-pass filtering step and the creation of optional augmentations, we transform each window into a Log-Mel spectrogram representation. This step is directly beneficial for our main line of experiments in two different ways. By compressing the raw waveform into a structured time-frequency description, the resulting representation can be simultaneously used for our ML and CNN lines of experiments. On the ML side, the extracted frequency dimensions can be used directly as handcrafted vectors for further processing and later classification. On the CNN side, the 2D image representations are used as the input, so a convolutional architecture can learn discriminative patterns in the image-like data. Figure 4 shows a time-frequency representation of an EcoContact 6Q sample.

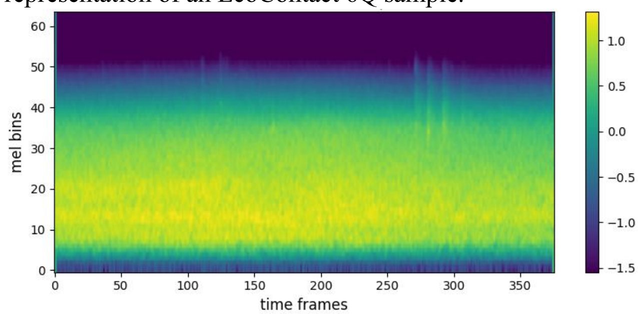


FIGURE 4: An EcoContact 6Q sample converted to a Log-Mel spectrogram representation.

E. DIMENSIONALITY REDUCTION

For the classical machine learning models, principal component analysis (PCA) is applied after Log-Mel spectrogram generation. Each flattened Log-Mel representation contains 12,032 features.

The reduction of this high-dimensional representation serves two purposes: It suppresses redundancy of feature dimensions and reduces the number of input features to improve the tractability of the tree-based model architecture. After applying PCA, the representation is reduced to 50 components. In contrast, PCA is not used for the CNNs, since they can directly operate on the 2D spectrogram tensor. The preprocessing pipeline, therefore, branches after Log-Mel spectrogram generation, depending on the model family later used.

To summarize, the data preparation pipeline transformed the raw TyRoN pass-by noise recordings into a structured data set for our tire-classification problem. Starting from 207 valid HDF5 files, which were acquired on 4 vehicles and 2 different roads, the audio signals were windowed into 2-s samples with 60% overlap, resulting in 1,082 band-pass filtered base samples for our non-augmented lines of experiments. By using augmentations, we increased this number to 3,246 data points. The base and the augmented samples were then transformed into Log-Mel spectrogram representations. For the classical machine learning models, PCA was applied to reduce dimensionality, whereas the CNNs used the image-like representations directly. The resulting dataset was used in our main lines of experimentation, as detailed in the following methodology section.

III. Methodology

The following chapter describes the experimental methodology, starting with an overview of our overall workflow, our decision to use a strict Leave-One-Group-Out (LOGO) approach, the pre-experiments that defined key parameters, and finally, our three main lines of experiments. The central objective was to directly compare classical machine learning techniques, CNN-based architectures, and a hybrid classification setup on the same data to explore the feasibility of tire-classification based solely on pass-by noise. The overall workflow is illustrated in Figure 5. All experiments were conducted in Python using VS Code and PyTorch/TensorFlow on a workstation equipped with two NVIDIA RTX 6000 Ada Generation GPUs (48 GB VRAM each). To improve reproducibility, fixed random seeds were used throughout the experiments.

The raw audio recordings are first preprocessed using Log-Mel spectrogram conversion, as detailed in the data preparation chapter. Afterwards, they are split into train/test LOGO setups, which will be discussed in the following section. Finally, our three model families are trained and evaluated under identical conditions, which allows consistent and controlled comparison of their performance.

B. LOGO(Leave-One-Group-Out)

To obtain a reliable evaluation, a strict LOGO strategy was adopted. This choice is essential because window-based sampling can otherwise introduce data leakage, as neighboring or overlapping windows from the same recording may contain highly similar information. For this reason, the split was performed strictly at the condition level, ensuring that no file contributed samples to both the training and test sets. In addition, recordings collected under the same operating conditions, such as identical speed or acceleration settings, can also be very similar across files. To prevent this second form of leakage, complete condition groups were held out during testing. As a result, the model is always evaluated on entirely unseen speed or acceleration groups, which provides a more realistic assessment of generalization across operating conditions.

Each operating condition consists of multiple independent pass-by recordings rather than a single recording. Depending on the operating condition, the number of recordings varies, with at least 10 independent recordings available for every condition.

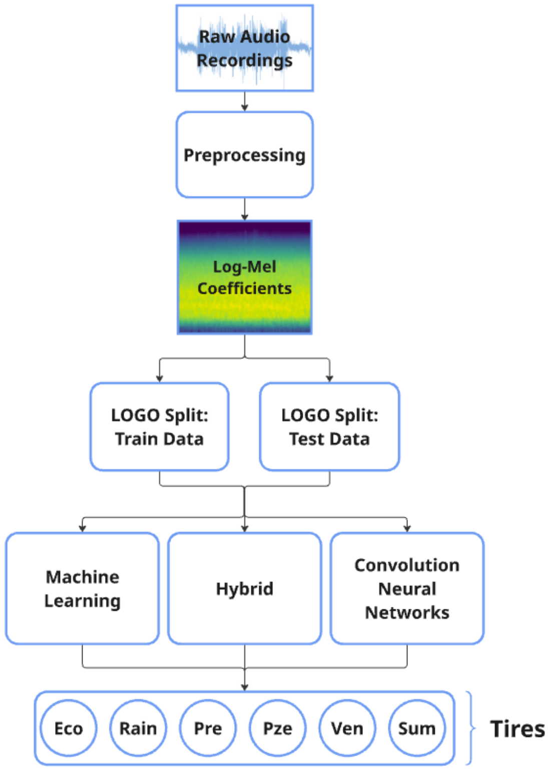


Figure 5: The pipeline of our main line of experiments.

During each LOGO evaluation, all recordings belonging to the selected speed or acceleration group are assigned exclusively to the test set, while recordings from all remaining operating conditions are used for training.

For overall comparison, evenly weighted averages across LOGO groups are reported in Table II. However, these aggregate values must be interpreted with care, since some groups are inherently easier than others when more similar conditions are represented in the training set. Therefore, detailed per-group results are additionally reported in the results section and discussed further in the limitations section.

Eight distinct LOGO setups were created from the data, as detailed in the data preparation chapter: 45 km/h, 50 km/h, 60 km/h, 80 km/h, and 100 km/h, and dynamic speeds starting at 50 km/h with accelerations of 1 m/s², 2 m/s², and 3 m/s². In each setup, one of those groups is chosen as the test data, and the rest provide the training data for the models. This is beneficial as in most setups, the model is evaluated on previously unseen operational conditions rather than only unseen windows. Metrics are first calculated for each of the eight LOGO classes individually.

TABLE II: TABLE TRAIN AND TEST SAMPLE DISTRIBUTIONS FOR EACH INDEPENDENT LOGO EVALUATION

LOGO Setup	Train Samples	Test Samples	Train Ratio	Test Ratio
1 m/s²	965	117	89.19%	10.81%
2 m/s²	964	118	89.09%	10.91%
3 m/s²	957	125	88.45%	11.55%
45 km/h	771	311	71.26%	28.74%
50 km/h	662	420	61.18%	38.82%
60 km/h	1014	68	93.72%	6.28%
80 km/h	927	155	85.67%	14.33%
100 km/h	954	128	88.17%	11.83%

Each row represents an independent LOGO evaluation rather than a single partition of the complete dataset. For example, in the 45 km/h LOGO setup, all 311 samples acquired at 45 km/h constitute the test set, whereas the remaining 771 samples collected under all other operating conditions are used for training. The same procedure is repeated independently for every speed and acceleration condition. Consequently, the train/test ratios differ across LOGO setups because the number of available samples varies between operating conditions.
They are evenly weighted across classes in the corresponding test split. For easier comparison, the resulting 8 LOGO macro metrics are then averaged across all setups again. Both the per-LOGO split and the combined numbers are reported in the Results chapter. To evaluate robustness across operational conditions, all models were evaluated independently on eight different LOGO setups, and accuracy, precision, recall, F1-score, and confusion matrices were analyzed across the different test conditions. For compact presentation, averaged metrics across the LOGO conditions are reported for most models. In addition, detailed condition-level results are explicitly presented for representative architectures in Figures 8, 10, and 12, illustrating performance variability across different operational conditions.

C. Pre-Experiments

Before starting the main line of experiments, several key design choices had to be established. Two main preliminary studies will be presented in the following section. These were implemented to reduce arbitrariness in parameter choice and to provide a heuristic for optimal hyperparameter setup. Since the segmentation strategy, feature representation, overlap ratio, and window length affect the number and mutual similarity of extracted samples, we investigated the optimal overlap ratio and window length before continuing with our final configuration. To investigate the influence of acoustic feature representation and temporal segmentation on the separability of tire classes, several pre-experiments were conducted before defining the final configuration. Different feature representations, including raw waveform signals, Log-Mel spectrograms, and MFCC-based features with delta and delta-delta coefficients, were evaluated using PCA-, UMAP-, and HDBSCAN-based embedding analysis. As shown in Fig. 6(a), raw waveform representations produced highly mixed embedding structures with limited class separation. Although the Log-Mel representation combined with HDBSCAN-based clustering improved the compactness of several tire-dependent acoustic manifolds, noticeable overlap between classes still remained, as illustrated in Fig. 6(b). Furthermore, temporal window lengths of 4 and 6 seconds were investigated to assess their effects on cluster compactness and inter-class separation. The comparison between Fig. 6(c) and Fig. 6(d) demonstrates that shorter temporal windows yield less compact clusters, whereas excessively long windows produce fragmented embedding structures. Among the investigated configurations, the 6-second window provided the best balance between compact cluster formation and inter-class separation. Therefore, additional window-length experiments were later conducted within the supervised ML and CNN pipelines to determine the final configuration used in the main experiments.

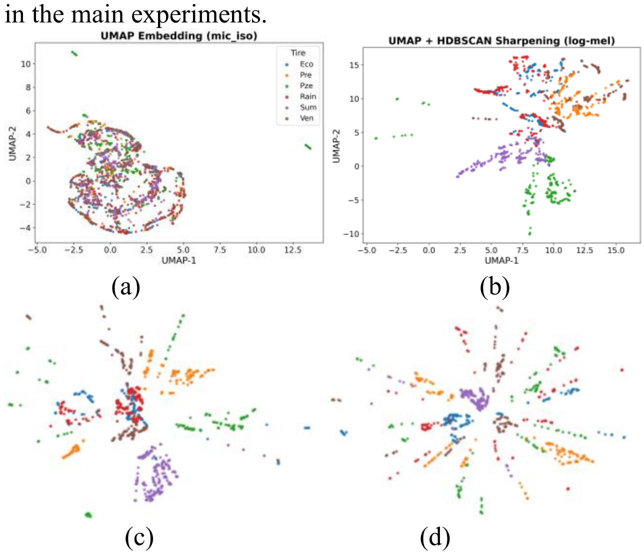


Figure 6: The Acoustic embedding analysis for different feature representations and temporal window lengths.

These observations demonstrate the complexity of tire-dependent acoustic manifolds and motivate the use of supervised machine learning and deep learning approaches for effective classification.
Second, an overlap study was conducted to find the optimal overlap ratio for window-based sample extraction. We tried to balance sample density and information redundancy. We trained LightGBM on our base dataset without augmenting the data samples. Window length was fixed to 4 s, the evaluation was done using the full LOGO procedure

described above. The model was configured with a learning rate of 0.05, 63 leaves, unrestricted tree depth, subsample, and column-sampling ratios of 0.9 in a multiclass classification setup. A random seed of 42 was used to ensure reproducibility. Training was done over 5000 boosting rounds, with early stopping. The overlap ratios explored were 50%, 60%, 70%, 80%, 90%. Since we were operating on limited amounts of data, the goal was to generate as many samples as possible while finding an optimal trade-off between sample count and avoidance of information redundancy. As visible in Figure 7, the best accuracy was achieved with a 90% overlap ratio with 90.02%, closely followed by the 60% overlap ratio with 88.01%.

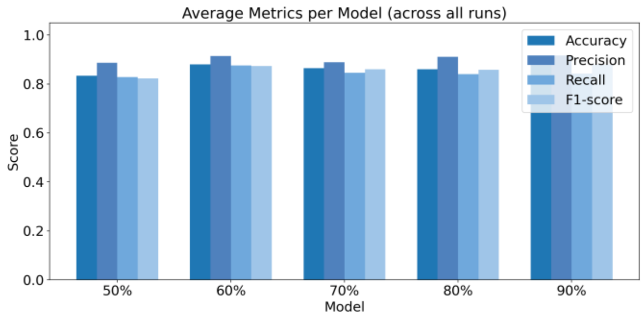
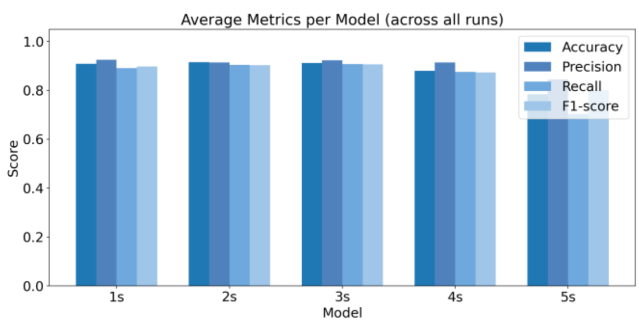
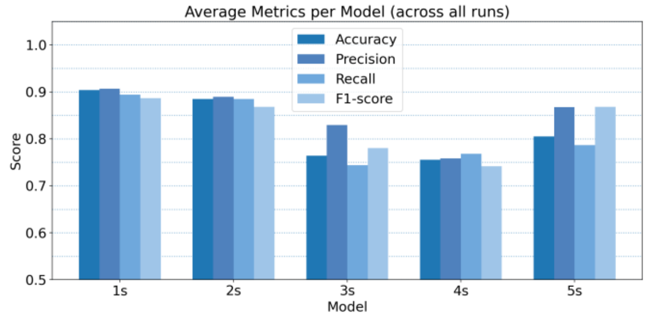


Figure 7: The Results from our overlap comparison

Since 90% overlap results in a strong overlap among samples and 60% overlap still achieved strong results, we continued with the second setting, since it yielded the best trade-off between sample count and avoidance of in-sample information redundancy. After establishing a fixed overlap ratio of 60% moving forward, another round of pre-experiment was conducted to determine the optimal temporal window length. The objective was to identify a segment length that preserves sufficient discriminative information while remaining small enough to yield a high sample count. This was especially important for our CNN line of experiments, as they are known to require large amounts of input data. Therefore, this study was conducted using LightGBM and EfficientNet-B0. In both cases, polyphase augmentation was disabled, and the full LOGO procedure was applied. On the EfficientNet side, the training was performed over 30 epochs with a batch size of 32 and a learning rate of 1e-4. The five window sizes compared were 1 s, 2 s, 3 s, 4 s, and 5 s, as visible in Figure 8a and b.



(a)



(b)

Figure 8: The Results from our window size comparison

For LightGBM, the strongest results were obtained for 2s windows with an accuracy of 91.48%. EfficientNet's best performance was with 1 s windows with 90.36%, closely followed by 2 s with 88.47%. In terms of F1-score, the 3s window performed best for LightGBM and the 1s window for EfficientNet. Considering both families jointly, 2 s windows were selected as the final configuration for the main experiments, as they showed strong performance across both cases. Furthermore, a single window size allows us to make direct comparisons of a unified training setup. Together with the previously established 60% overlap, this concludes our pre-experimentation phase, which established key data metrics for our main lines of experiments as detailed in the next section.

D. Main Experiment I: Classical Machine Learning

This section investigates how classical machine learning algorithms perform on a tire classification task using solely pass-by noise as an input feature. It's a six-class classification problem, with the data being split into a LOGO setup as detailed earlier. This experiment serves as a baseline reference against the CNN and hybrid approaches, which will be detailed later. The inputs consist of 50 handcrafted feature representations, a result of the Log-Mel spectrogram and PCA preprocessing steps. Four classifiers are compared, namely LightGBM, XGBoost, RandomForest, and ExtraTrees; a random seed of 42 was set wherever possible. LightGBM was configured with 5000 estimators, 0.05 learning rate, 63 leaves, unrestricted depth, subsample, and column sample ratios of 0.9. The settings for XGBoost were, namely, maximum depth of 8, learning rate of 0.05, subsample and column sample ratios of 0.9, mlogloss evaluation metric, and histogram-based tree construction. RandomForest used 1000 trees, unrestricted depth, minimum split size of 2, minimum leaf size of 1, square-root feature sampling, and bootstrap sampling. The same setup was used for ExtraTrees, but without bootstrap sampling.

Performance was computed for each LOGO setup individually, then averaged evenly weighted. We trained the classical models in two main setups: First, using the normal database and secondly by adding the additional augmented samples from the polyphase augmentation as described in the Data Preparation chapter. Tracked metrics for all three experiments included accuracy, precision, recall, and f1-score. This experimental block was designed to quantify the performance achieved by classical machine learning models in a tire classification problem, and serves as a baseline for later comparison with our CNN-based and hybrid approaches.

E. Main Experiment II: Convolutional Neural Networks

The second main experiment block investigates how trained CNNs perform in the given tire classification task. The networks learn directly from the Log-Mel spectrogram representation by studying discriminative spatial patterns directly from the two-dimensional time-frequency structure. The purpose was to evaluate whether CNN-based approaches can provide an advantage over the classical machine learning algorithms. All CNNs directly received the structured input representations, allowing the CNNs to exploit correlations that are not explicitly modeled in the flat representations used by the machine learning algorithms. Four architectures were compared in this experiment, namely LeNet, ResNet18, ConvNext-Tiny, and EfficientNet-B0. These networks were selected to cover a range from simple convolutional designs to advanced architectures. Since the dataset is relatively compact, the appropriate smaller architectures from within each family were chosen, wherever possible. To ensure a fair comparison, all four models were trained under the same setup: learning rate of 1e-4, batch size of 32, and a total of 30 epochs. The loss function, cross-entropy loss, was used in all four cases. The data split was performed according to the LOGO setup detailed earlier. As in the machine learning setup, metrics were first calculated per LOGO class and then averaged for easier comparison. This parallel approach ensures comparability between CNN- and ML-based results. This also applied to the dataset variants that the models were trained and evaluated on. In the base configuration 1,082 samples were present, which were increased to 3,246 samples in the augmented setup. This experiment was performed to investigate whether direct convolutional feature learning on image-like representations yields an advantage over the classical machine learning algorithms. It forms the second major comparison line of the study.

F. Main Experiment III: Hybrid Network

The final main experiment explores a hybrid architecture. It combines deep feature extraction with a classical machine learning model. The aim was to test whether the complementary strengths of both families can improve performance beyond what they can do individually. On the one hand, convolutional neural networks are well-suited to learn structured and discriminative patterns from the spectrogram representations; on the other hand, tree-based classifiers bring robustness and a training behavior that can work well with smaller numbers of samples. The motivation, therefore, follows directly from the two preceding experiments. First, the CNN extracts patterns from the full spectrogram tensors, and the classical machine learning algorithm then uses this additional information and combines it with the handcrafted representations from the traditional approach. This hybrid approach enables combining both information sources within a single model pipeline.

The architecture was implemented using a ResNet18 backbone for deep feature extraction and ExtraTrees as the final decision model. First, the ResNet model was pretrained directly on the spectrograms. The learning rate was set to 1e-4, the batch size to 32, and the training was done for 30 epochs. Afterwards, its network weights were frozen and the final classification layer removed. This pretraining stage enabled the network to extract discriminative features from samples.

In parallel to this feature pathway, another representation was built for the machine learning approach from the Log-Mel spectrograms along the preprocessing steps discussed before, and the input was reduced to 50 principal components.

As a result, the hybrid model received the same input through two complementary pathways. The features extracted from the CNN backbone and the handcrafted feature vector were concatenated before classification. These inputs were combined and used to train the ExtraTrees classifier, which performed the final classification.

Two setups were evaluated in this study: a normal unmodified dataset consisting of 1,082 samples and the augmented counterpart with 3,246 samples. They were again split along the LOGO setup. Results were averaged as previously reported. Overall, this experiment was designed to determine whether it is possible to increase the performance of a classification model using pass-by noise by processing the same data sample twice through a deep learning and a handcrafted representation approach. It presents the final comparative component of the main line of experiments, exploring whether a fused representation strategy is advantageous for tire classification from pass-by noise.

This complementary design was motivated by the observation that CNN-based representations can capture local and global spectrogram structures, while tree-based ensemble methods often provide stronger robustness under limited-data conditions and partially overlapping feature distributions.

The comparative results further indicate that augmentation affects different model families differently. While CNN-based architectures, particularly ResNet and the hybrid framework, benefited substantially from the increased sample diversity introduced by the augmentation pipeline, several classical machine learning models showed reduced performance under the augmented setup.

This behavior may be related to the high correlation between overlapping and polyphase-generated samples, which can affect tree-based ensemble decision boundaries differently than representation-learning-based CNN architectures. These observations suggest that increased sample count alone does not necessarily improve performance uniformly across all model families under limited-data conditions.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3714728

IV. Results

A. Results: Classical Machine Learning

The preliminary embedding analysis indicated that tire-dependent acoustic patterns are only partially separable in the unsupervised feature space, which further motivates the use of supervised learning approaches. On the base dataset, all ML models achieved strong classification accuracies across the board (>0.90) as visible in Table III. XGBoost achieved the best performance with an accuracy, precision, recall, and F1-Score of 0.9172, 0.9229, 0.8999, 0.9020, respectively, showing that classical machine learning algorithms are suited to discriminate between handcrafted features derived from the pass-by noise. Furthermore, this shows that even with a relatively low number of training samples, the model was able to learn sufficient patterns to generalize to the previously unseen test set.

TABLE III: EXPERIMENTAL RESULTS USING DATA WITHOUT AUGMENTATIONS

Model	Accuracy	Precision	Recall	F1-Score
LightGBM	0.9148	0.9142	0.9046	0.9029
XGBoost	**0.9172**	0.9229	0.8999	0.9020
RandomForest	0.9009	0.9221	0.8810	0.8851
ExtraTrees	0.9145	0.9385	0.9164	0.9178
LeNet	0.5816	0.6589	0.5035	0.5861
ResNet	0.5717	0.7885	0.5712	0.6274
ConvNext	**0.9112**	0.9212	0.8957	0.8936
EfficientNet	0.8847	0.8894	0.8848	0.8678
Hybrid	**0.9553**	0.9613	0.9430	0.9390

On the augmented dataset, all models still delivered decent accuracies (>0.85) as visible in Table IV, but average performance visibly dropped compared to the non-augmented dataset. This might be due to the augmentation increasing data size, without increasing informational diversity sufficiently.

TABLE IV: EXPERIMENTAL RESULTS USING DATA WITH AUGMENTATIONS

Model	Accuracy	Precision	Recall	F1-Score
LightGBM	0.8517	0.8645	0.8354	0.8363
XGBoost	0.8540	0.8692	0.8336	0.8382
RandomForest	0.8629	0.8918	0.8413	0.8485
ExtraTrees	**0.8854**	0.9164	0.8856	0.8852
LeNet	0.3524	0.4404	0.2886	0.3158
ResNet	**0.9447**	0.9509	0.9382	0.9381
ConvNext	0.9062	0.9016	0.8946	0.8919
EfficientNet	0.8690	0.8609	0.8495	0.8499
Hybrid	**0.9688**	0.9677	0.9650	0.9646

The models may have been unable to deal with the heavily correlated new samples introduced by the polyphase augmentation. Nonetheless, the best-performing model, ExtraTrees, still came close to the non-augmented performance with accuracy, precision, recall, and F1-score of 0.8854, 0.9164, 0.8856, and 0.8852. To get a better understanding of this model's specific performance, it's important to view it in the context of the LOGO setup, described in the methodology chapter. Since one group was excluded from serving as a test set in each setup, some scenarios may have been easier than others, due to more similar data being present in the train set. This is not visible when only considering the averaged metrics. Looking at the detailed overview in Figure 9 helps to truly evaluate the performance and how well this model would generalize in a deployment setup. These observations further suggest that increasing sample count through highly correlated augmentations does not necessarily improve classical machine learning performance, whereas CNN-based models may benefit more strongly from increased sample diversity under limited-data conditions.

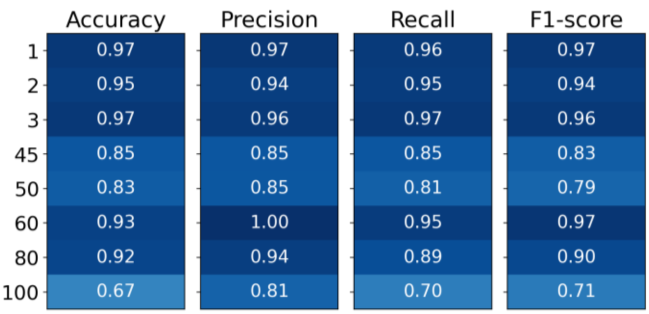


Figure 9: **Detailed ExtraTrees results under different conditions**

While the acceleration runs 1 m/s², 2 m/s², and 3 m/s² all showed strong performance with accuracies above 0.95, the setups with constant acceleration are mixed. Here, the edge cases of lower (e.g., 45 km/h) and higher speeds (e.g., 100 km/h) generally perform worse than the ones in the center. This may be due to the center speeds in the data distribution being more supported by the surrounding data when being left out as a test set than the edge cases, where fewer data points are available. Similar considerations had to be made for the average confusion matrix in Figure 10.

Since the held-out test set may not always include all present classes, and sample counts differ, the averaged confusion matrix is presented in normalized percentage form. The strongest performance was reached in the Sum class with a class recall of 0.98.

The most confusion happened at the Rain class, with 0.16 being wrongly predicted as Eco and 0.06 as Pre, showing partial overlap with two other tires.

Overall, the classical machine learning models showed that it is possible to classify tire types based solely on pass-by noise. While performance was solid across both datasets, it

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and
content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3714728

did not benefit from introducing additional augmentations; instead, performance suffered significantly.

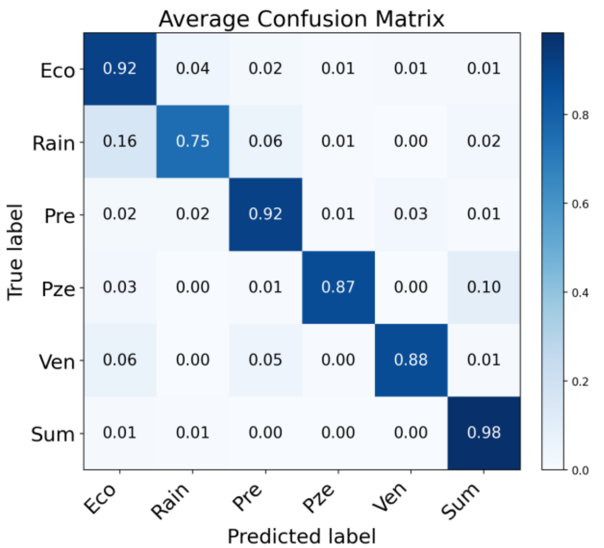


Figure 10: **The Confusion matrix for ExtraTrees tire classification**

B. Results: Convolutional Neural Networks

The convolutional neural networks showed a markedly different behavior than the classical machine learning approach. On the basic dataset, performance varied strongly between models, ranging from ResNet’s accuracy of 0.5717 to ConvNexts with 0.9112, as visible in Table III. The trend of larger models showing a better performance was also visible, with LeNet being strongly outperformed by EfficientNet. Unlike with the classical machine learning approaches, the models partially seemed to suffer from the small data size of only 1,082 total samples, proving insufficient to learn the relevant discriminative features to consistently classify correctly.

On the augmented dataset, the performance behavior changed significantly under the augmented setup. ResNet benefited substantially from the additional training data available, with a performance increase of 0.373 in terms of accuracy, as visible in Table IV. Precision, recall, and F1-score also showed strong results with 0.9509, 0.9382, and 0.9381, respectively. The previous frontrunners in ConvNext and EfficientNet maintained a solid performance but with slight decreases in accuracy (<0.02), while LeNet's performance further deteriorated to 0.3524.

To further investigate ResNet performance on the augmented dataset, we provide Figure 11, which shows a heatstrip plot of detailed per LOGO setup performances.

The 1 m/s², 2 m/s², and 3 m/s² all showed strong performances with accuracies between 0.97 – 1.00. This shows that ResNet showed strong generalization across the accelerated pass-by scenarios. The setups using constant speed showed slightly more variation. As with the ExtraTrees model, the edge cases of 45 km/h and 50 km/h still performed well with accuracies of 0.96 and 0.94, but were weaker than the central speed configurations, with 60 km/h reaching a precision of 1.00 and an F1-score of 0.99, and 80 km/h showing 0.98 across all metrics.

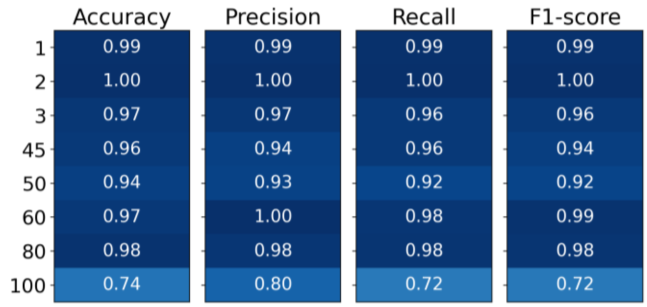


Figure 11: **Detailed Resnet results under different conditions**

The most difficult setup again was 100 km/h, where accuracy, precision, recall, and F1-score were 0.74, 0.80, 0.72, and 0.72, respectively. As previously discussed, this may be due to center cases being better supported by neighboring operating conditions. Figure 12 shows the average confusion matrix with specific information regarding the per-tire performances.

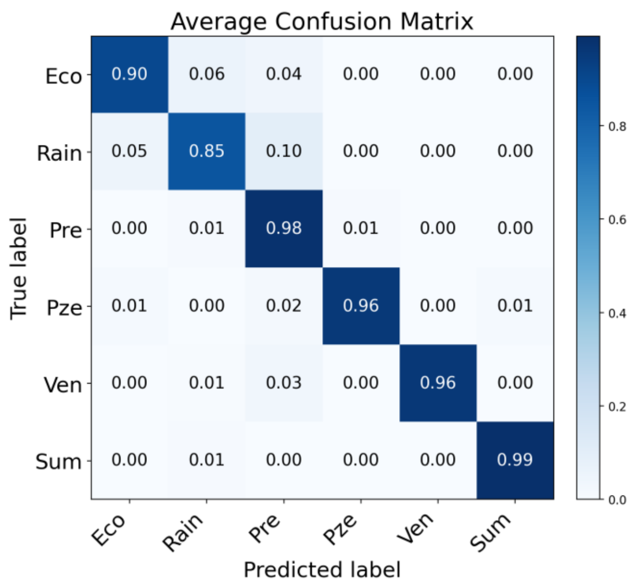


Figure 12: **The Confusion matrix for Resnet tire classification**

The strongest class-wise performance was delivered by Sum with a class recall of 0.99. In contrast, the weakest tire was Rain, which was confused in 0.1 cases with Pre and in 0.05 cases with Eco. This shows the same pattern as with ExtraTrees model performance, where most confusion happened between separating Rain from the Pre and Eco tires, but in noticeably fewer cases on the CNN side. Overall, the best CNNs were able to differentiate between the different classes by leveraging their pattern recognition capabilities to find discriminative patterns in the time-frequency spectrograms without other major preprocessing. Their performance varied noticeably, though displaying

clear differences between models able to converge in the latent space and those unable to reliably separate classes.
The introduction of additional data in the form of augmented samples substantially improved ResNet's performance, increasing it from the worst to the best model across both ML and CNN setups. This highlights the importance of a sufficiently large dataset and the feasibility of increasing sample count through polyphase augmentations.

C. Results: Hybrid Network

In the Hybrid network setup, the question was whether it's possible to improve performance by combining classical machine learning methods in the form of ExtraTrees, with deep learning-based feature extraction provided by a ResNet architecture. Table IV shows that the Hybrid model already achieved the best overall performance among all evaluated approaches on the base dataset, without additional augmentations. It reached an accuracy of 0.9553, a precision of 0.9613, a recall of 0.9430, and an F1-score of 0.9390. This indicates that combining learned deep representations with additional complementary information allowed the model to discriminate very effectively between tire types even without augmented data. On the augmented dataset, the Hybrid model further improved its performance and achieved the strongest results of this study. With an accuracy of 0.9688, a precision of 0.9677, a recall of 0.9650, and an F1-score of 0.9646 it shows promising generalization performance and provides evidence that the Hybrid approach benefited clearly from the additional variability introduced by augmentation. Figure 13 (Spiderweb Graph) gives an overview across test groups. It shows that the hybrid model performed consistently across almost all LOGO setups. While the general trends from previous observations still show that 100 km/h was the most difficult setup with an accuracy of 0.89, the Hybrid network significantly improved this performance from the standalone models. Other accuracies in the range of 0.96 – 1.00 further consolidate the strong results of the Hybrid model on augmented data. Additionally, no pronounced breakdown was observed for any of the edge-case groups. This indicates that the Hybrid model generalized not only well on the central regions of the data distribution but also remained comparatively stable when lower- or higher-support operating conditions were excluded entirely during training. The confusion matrix shown in Figure 14 (Confusion Matrix) further supports this trend. The only major source of confusion occurred between the Rain and Eco classes, with the Pre class being classified correctly much more consistently than with the standalone models. Otherwise, it demonstrates a very high degree of separability between the tire classes, with particularly good results in the Pze and Sum classes, with per-class recalls of 0.99. Overall, the off-diagonal entries are extremely small, showing that the hybrid model maintained a highly stable and balanced class separation across all tire categories. When taken together, the Hybrid network results represent the strongest findings of this study. On the non-augmented data, the model was highly effective with an accuracy of above 95%. This improved by another 1.35% when introducing the polyphase augmentation. In contrast to the ML and CNN standalone approaches, the per-setup and per-class performances remained much more stable, with the confusion matrix showing only minor confusion between some classes, and almost completely eliminating them between others.

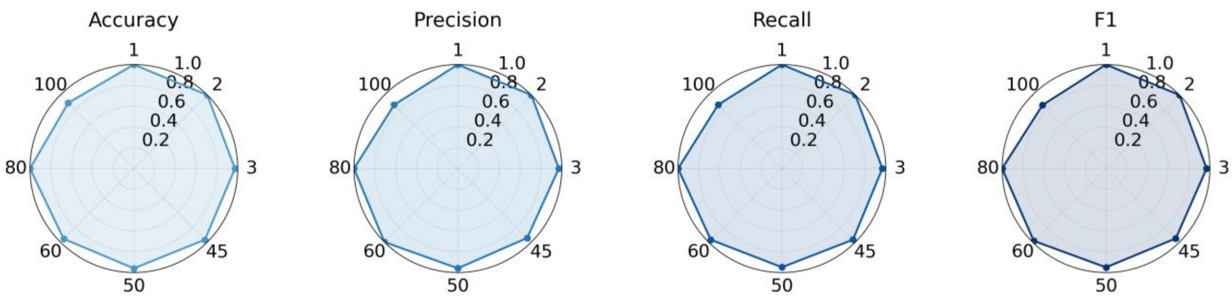


Figure 13: Hybrid ML structure results

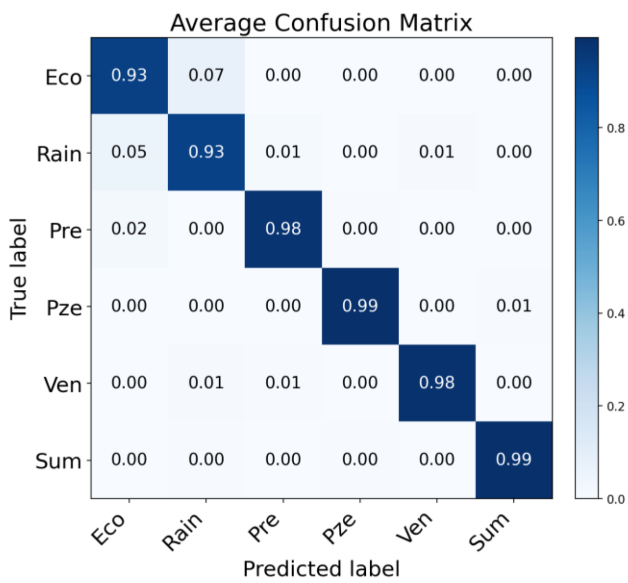


Figure 14: The Confusion matrix for hybrid ML structure tire classification

This suggests that this setup was particularly well-suited to exploit the enlarged training distribution while still maintaining stable class separation. It is also important to highlight that the Hybrid model used the same classification setup as the ML models, with handcrafted features, produced using a dimensionality reduction on the Log-Mel coefficients, being used to classify the tires solely using pass-by noise. The difference was the additional path this raw data took. By additionally using a deep learning backbone architecture to extract features, the ExtraTrees model was provided with broad, high-quality input features, which helped achieve this significant increase in classification performance. These results indicate that the Hybrid architecture was able to combine strong feature extraction with robust generalization to accurately classify tires using pass-by noise.

V. Conclusion and Discussion

This study presented a pass-by noise-based approach for tire type classification using a single roadside microphone, enabling a fully non-intrusive and scalable monitoring framework. The results demonstrated that tire–road interaction produces distinctive acoustic patterns that can be effectively exploited to differentiate between tire types without requiring any vehicle-mounted sensors or camera systems. By focusing on the 400–5000 Hz frequency range and employing a 2-second window with 60% overlap, stable and informative signal representations were obtained. The transformation into Log-Mel spectrograms enabled both classical machine learning and deep learning models to extract discriminative time–frequency features. Preliminary embedding analyses further demonstrated that tire-dependent acoustic manifolds exhibit partial overlap and complex nonlinear structures, indicating that reliable classification cannot be achieved through simple feature-space separation alone.

The experimental results showed that classical machine learning models achieve strong performance even under limited data conditions, while deep learning models require larger datasets but benefit significantly from augmentation. The best performance was obtained using a hybrid framework combining CNN-based feature extraction (ResNet) with a classical machine learning classifier (Extra Trees), achieving an accuracy of 96.88% and providing more stable performance across varying operating conditions. The use of a strict Leave-One-Group-Out (LOGO) validation strategy demonstrated promising generalization across unseen speed and driving conditions. However, performance degradation observed at higher speeds (e.g., 100 km/h) indicates that further improvements in dataset balance and diversity are necessary. In addition, while augmentation increased the dataset size, its limited contribution to performance highlights the need for more informative augmentation strategies.

Overall, the results confirm that pass-by noise demonstrates the feasibility of a practical and cost-effective solution for tire classification in real-world traffic environments. Compared to vision-based or sensor-based systems, the proposed approach enables contactless monitoring under realistic conditions, making it shows potential for future large-scale deployment. Although the proposed framework demonstrated promising performance under the evaluated operational conditions, further validation under unseen road surfaces, environmental conditions, and tire aging scenarios is necessary to fully assess large-scale real-world generalization.

Future work will focus on extending the dataset to include varying tire wear levels and aging conditions to improve generalization and real-world applicability. Furthermore, integrating acoustic-based classification with complementary sensing modalities such as computer vision may provide a more comprehensive understanding of vehicle and tire characteristics. Although leave-one-vehicle-out and leave-one-road-surface-out evaluations represent important future research directions, the current dataset structure includes limited tire configurations for some vehicles and only two road surface categories, limiting the statistical meaningfulness of such splits at the present stage. Future work may additionally investigate transformer-based acoustic architectures and self-supervised audio representation learning approaches for improved generalization under large-scale datasets. Compared to large transformer-based acoustic architectures, the proposed lightweight hybrid framework may provide advantages for future roadside and edge-oriented deployment scenarios due to its lower computational complexity and reduced hardware requirements. Additional research directions include tire pressure estimation from acoustic signals and further refinement of classification performance. Finally, the proposed system has the potential to be developed into a roadside monitoring solution for intelligent transportation systems, supporting automated traffic safety assessment and large-scale smart city applications.

ACKNOWLEDGMENT

The authors extend their thanks to the Tyre Road Noise project, which is funded by the German Federal Ministry of Digital and Transport (BMDV) as part of the mFUND innovation initiative, for providing the data used in this study. However, no funding was provided by the project to cover publication fees. The authors gratefully acknowledge financial support for open access publication by the KIT Open Access Publication Fund.

REFERENCES

[1] WHO, “Global status report on road safety 2023: country and territory profiles.” Accessed: Mar. 24, 2026. [Online]. Available: https://www.who.int/publications/i/item/9789240087712
[2] N. Gavanas, K. Anastasiadou, E. Nathanail, and S. Basbas, “Transport Policy Pathways for Autonomous Road Vehicles to Promote Sustainable Urban Development in the European Union: A Multicriteria Analysis,” *Land*, vol. 13, no. 11, Oct. 2024, doi: 10.3390/land13111807.
[3] National Highway Traffic Safety Administration, “Traffic safety facts annual report,” US Department of Transportation, 2020.
[4] O. Orynycz, K. Tucki, A. Wasiak, R. Sobótka, and A. Gola, “Evaluation of the Brake’s Performance Dependence Upon Technical Condition of Car Tires as a Factor of Road Safety

Management," *Energies*, vol. 13, no. 1, Dec. 2019, doi: 10.3390/en13010009.

[5] I. Edunyah, "Causes of tyre failure on road traffic accident; a case study of Takoradi Township," *Int. J. Sci. Res. Publ*.

[6] W. Blythe and T. D. Day, "Single Vehicle Wet Road Loss of Control; Effects of Tire Tread Depth and Placement," presented at the SAE 2002 World Congress & Exhibition, Detroit, Michigan, United States, Mar. 2002, pp. 2002-01–0553. doi: 10.4271/2002-01-0553.

[7] D. Uguru-Okorie, "The role of tire in car crash, its causes, and prevention," Aug. 2018.

[8] P. Stabile, G. Mastinu, M. Gobbi, and P. Bardini, "Tire aging: A state-of-the-art review," *Engineering Science and Technology, an International Journal*, vol. 69, p. 102101, Sep. 2025, doi: 10.1016/j.jestch.2025.102101.

[9] M. Schulthoff, M. Kaltschmitt, C. Balzer, K. Wilbrand, and M. Pomrehn, "European road transport policy assessment: a case study for Germany," *Environ Sci Eur*, vol. 34, no. 1, p. 92, Sep. 2022, doi: 10.1186/s12302-022-00663-7.

[10] S. Jansen, A. Schmeitz, and L. Akkermans, "Tire use and road safety – background to policy recommendations for new EU measures," in *6th International Munich Chassis Symposium 2015*, P. Pfeffer, Ed., Wiesbaden: Springer Fachmedien, 2015, pp. 129–147. doi: 10.1007/978-3-658-09711-0_12.

[11] R. Wang, Q. Guo, S. Lu, and C. Zhang, "Tire Defect Detection Using Fully Convolutional Network," *IEEE Access*, vol. 7, pp. 43502–43510, 2019, doi: 10.1109/ACCESS.2019.2908483.

[12] T. Mignot, F. Ponchon, A. Derville, S. Duffner, and C. Garcia, "An automatic inspection system for the detection of tire surface defects and their severity classification through a two-stage multimodal deep learning approach," *J Intell Manuf*, vol. 36, no. 5, pp. 3427–3445, Jun. 2025, doi: 10.1007/s10845-024-02378-3.

[13] Z. Tong *et al.*, "Machine learning-driven intelligent tire wear detection system," *Measurement*, vol. 242, p. 115848, Jan. 2025, doi: 10.1016/j.measurement.2024.115848.

[14] W. Kazmi, I. Nabney, G. Vogiatzis, P. Rose, and A. Codd, "Vehicle tire (tyre) detection and text recognition using deep learning," in *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE)*, Aug. 2019, pp. 1074–1079. doi: 10.1109/COASE.2019.8842962.

[15] R. Matsuzaki and A. Todoroki, "Wireless Monitoring of Automobile Tires for Intelligent Tires," *Sensors*, vol. 8, no. 12, pp. 8123–8138, Dec. 2008, doi: 10.3390/s8128123.

[16] S. Arora, S. Naveen Venkatesh, V. Sugumaran, A. Prabhakaranpillai Sreelatha, and V. S. Mahamuni, "Enhancing Tire Condition Monitoring through Weightless Neural Networks Using MEMS-Based Vibration Signals," *Journal of Engineering*, vol. 2024, no. 1, p. 1321775, 2024, doi: 10.1155/2024/1321775.

[17] H. Askari, A. Khajepour, M. B. Khamesee, and Z. L. Wang, "Embedded self-powered sensing systems for smart vehicles and intelligent transportation," *Nano Energy*, vol. 66, p. 104103, Dec. 2019, doi: 10.1016/j.nanoen.2019.104103.

[18] K. Iwao and I. Yamazaki, "A study on the mechanism of tire/road noise," *JSAE Review*, vol. 17, no. 2, pp. 139–144, Apr. 1996, doi: 10.1016/0389-4304(95)00004-6.

[19] M. Heckl, "Tyre noise generation," *Wear*, vol. 113, no. 1, pp. 157–170, Dec. 1986, doi: 10.1016/0043-1648(86)90065-7.

[20] Y. Azizi, "Generation mechanisms of tire/road noise," in *Automotive Tire Noise and Vibrations*, Elsevier, 2020, pp. 91–114. doi: 10.1016/B978-0-12-818409-7.00006-4.

[21] E. Ascari, M. Cerchiai, L. Fredianelli, and G. Licitra, "Statistical Pass-By for Unattended Road Traffic Noise Measurement in an Urban Environment," *Sensors*, vol. 22, no. 22, Nov. 2022, doi: 10.3390/s22228767.

[22] M. Zappatore, A. Longo, and M. A. Bochicchio, "Crowd-sensing our Smart Cities: a Platform for Noise Monitoring and Acoustic Urban Planning," *Journal of Communications Software and Systems*, vol. 13, no. 2, pp. 53–67, Jun. 2017, doi: 10.24138/jcomss.v13i2.373.

[23] A. Bansal and N. K. Garg, "Environmental Sound Classification: A descriptive review of the literature," *Intelligent Systems with Applications*, vol. 16, p. 200115, Nov. 2022, doi: 10.1016/j.iswa.2022.200115.

[24] W. Kongrattanaprasert, "Detection of tire types using tire noise from passing vehicles," in *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, Dec. 2014, pp. 1–4. doi: 10.1109/APSIPA.2014.7041753.

[25] M. Ashhad, U. Goenka, A. Jagetia, P. Akhtari, S. K. Ambat, and M. Samuel, "Improved Vehicle Sub-type Classification for Acoustic Traffic Monitoring," Feb. 06, 2023, *arXiv*: arXiv:2302.02945. doi: 10.48550/arXiv.2302.02945.

[26] H. Parineh, M. Sarvi, and S. A. Bagloee, "A Deep Wavelet-Fourier Method for Monaural Vehicle Speed Estimation in Real-Life Settings," *IEEE Sensors Journal*, vol. 24, no. 9, pp. 15337–15346, May 2024, doi: 10.1109/JSEN.2024.3381276.

[27] A. Dawn and W. Ansar, "Studying the Effect of Audio Filters in Pre-Trained Models for Environmental Sound Classification," Aug. 24, 2024, *arXiv*: arXiv:2408.13644. doi: 10.48550/arXiv.2408.13644.

[28] Y. R. Veeranki, R. McNaboe, and H. F. Posada-Quintero, "EEG-Based Seizure Detection Using Variable-Frequency Complex Demodulation and Convolutional Neural Networks," *Signals*, vol. 4, no. 4, pp. 816–835, Nov. 2023, doi: 10.3390/signals4040045.

[29] J. Ludeña-Choez and A. Gallardo-Antolín, "Feature extraction based on the high-pass filtering of audio signals for Acoustic Event Classification," *Computer Speech & Language*, vol. 30, no. 1, pp. 32–42, Mar. 2015, doi: 10.1016/j.csl.2014.04.001.

[30] W. Mu, B. Yin, X. Huang, J. Xu, and Z. Du, "Environmental sound classification using temporal-frequency attention based convolutional neural network," *Sci Rep*, vol. 11, no. 1, p. 21552, Nov. 2021, doi: 10.1038/s41598-021-01045-4.

[31] K. J. Piczak, "Environmental sound classification with convolutional neural networks," in *2015 IEEE 25th International Workshop on Machine Learning for Signal Processing (MLSP)*, Boston, MA, USA: IEEE, Sep. 2015, pp. 1–6. doi: 10.1109/MLSP.2015.7324337.

[32] Q. Kong, Y. Cao, T. Iqbal, Y. Wang, W. Wang, and M. D. Plumbley, "PANNs: Large-Scale Pretrained Audio Neural Networks for Audio Pattern Recognition," *IEEE/ACM Trans. Audio Speech Lang. Process.*, vol. 28, pp. 2880–2894, 2020, doi: 10.1109/TASLP.2020.3030497.

[33] A. Mesaros, T. Heittola, and T. Virtanen, "Acoustic Scene Classification: An Overview of Dcase 2017 Challenge Entries," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, IEEE, Sep. 2018, pp. 411–415. doi: 10.1109/IWAENC.2018.8521242.

[34] G. S. Mafra, N. Q. K. Duong, A. Ozerov, and P. Perez, "ACOUSTIC SCENE CLASSIFICATION: AN EVALUATION OF AN EXTREMELY COMPACT FEATURE REPRESENTATION," 2016.

[35] L. M. Schwemer, M. Demetgül, and S. Lazarova-Molnar, "RoadAceVision: Deep Multimodal Learning for Road Surface Classification," 2026.

[36] H. Bolcskei and F. Hlawatsch, "Discrete Zak transforms, polyphase transforms, and applications," *IEEE Trans. Signal Process.*, vol. 45, no. 4, pp. 851–866, Apr. 1997, doi: 10.1109/78.564174.

[37] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *PLoS One*, vol. 16, no. 7, p. e0254841, Jul. 2021, doi: 10.1371/journal.pone.0254841.

[38] X. Wang, G. Zhang, S. Lou, S. Liang, and X. Sheng, "Two-round feature selection combining with LightGBM classifier for disturbance event recognition in phase-sensitive OTDR system," *Infrared Physics & Technology*, vol. 123, p. 104191, Jun. 2022, doi: 10.1016/j.infrared.2022.104191.

[39] S. Koç and R. İlenç, "Tool-Wear Classification Using Sound-Based Ensemble Learning in the Turning of AISI 1050 Steels," *J. of Materi Eng and Perform*, Feb. 2026, doi: 10.1007/s11665-026-13434-9.

[40] A. Kaçar, D. Avci, and İ. Türkoğlu, "Enhancing Environmental Sound Classification Performance through Data Fusion: A Comparative Machine Learning Analysis".

[41] M. Badawy *et al.*, "From Sensors to Insights: Interpretable Audio-Based Machine Learning for Real-Time Vehicle Fault and

Emergency Sound Classification," *Machines*, vol. 13, no. 10, Sep. 2025, doi: 10.3390/machines13100888.

[42] T. Arias-Vergara, P. Klumpp, J. C. Vasquez-Correa, E. Nöth, J. R. Orozco-Arroyave, and M. Schuster, "Multi-channel spectrograms for speech processing applications using deep learning methods," *Pattern Anal Applic*, vol. 24, no. 2, pp. 423–431, May 2021, doi: 10.1007/s10044-020-00921-5.

[43] K. W. Cheng *et al.*, "Spectrogram-based classification on vehicles with modified loud exhausts via convolutional neural networks," *Applied Acoustics*, vol. 205, p. 109254, Mar. 2023, doi: 10.1016/j.apacoust.2023.109254.

[44] T. Pellegrini, I. Khalfaoui-Hassani, E. Labbé, and T. Masquelier, "Adapting a ConvNeXt Model to Audio Classification on AudioSet," in *INTERSPEECH 2023*, ISCA, Aug. 2023, pp. 4169–4173. doi: 10.21437/Interspeech.2023-1564.

[45] K. M. Lim, C. P. Lee, Z. Y. Lee, J. Y. Lim, and J. N. Mogan, "Acoustic Event Classification with Enhanced EfficientNet," in *2024 IEEE 12th Conference on Systems, Process & Control (ICSPC)*, Dec. 2024, pp. 13–17. doi: 10.1109/ICSPC63060.2024.10862919.

[46] K. Zaman, M. Sah, C. Direkoglu, and M. Unoki, "A Survey of Audio Classification Using Deep Learning," *IEEE Access*, vol. 11, pp. 106620–106649, 2023, doi: 10.1109/ACCESS.2023.3318015.

[47] S. Chen *et al.*, "BEATs: Audio Pre-Training with Acoustic Tokenizers," 2022, *arXiv*. doi: 10.48550/ARXIV.2212.09058.

[48] J. Meng, M. Demetgül, S. Lazarova-Molnar, F. Gauterin, and A. Vinel, "Acoustic Traffic Attribute Classification for ITS: A Comparative Study of Machine Learning and CNN Approaches," in *2025 IEEE 102nd Vehicular Technology Conference (VTC2025-Fall)*, Oct. 2025, pp. 1–7. doi: 10.1109/VTC2025-Fall65116.2025.11310562.

[49] J. S. Luz, M. C. Oliveira, F. H. D. Araújo, and D. M. V. Magalhães, "Ensemble of handcrafted and deep features for urban sound classification," *Applied Acoustics*, vol. 175, p. 107819, Apr. 2021, doi: 10.1016/j.apacoust.2020.107819.

[50] W. Shen *et al.*, "Fusion of acoustic and deep features for pig cough sound recognition," *Computers and Electronics in Agriculture*, vol. 197, p. 106994, Jun. 2022, doi: 10.1016/j.compag.2022.106994.

[51] M. Demetgul, Lukas Michael Schwemer, and Lazarova-Molnar, Sanja, "Classification of Tyre Type and Pressure from Pass-by Noise," presented at the 5th International Tyre Colloquium (IAVSD), Fraunhofer ITWM, Kaiserslautern, Germany, Mar. 2026.

[52] A. Hakala, "SOUND BASED CLASSIFICATION OF STUDDED TIRES," Tampere University, 2020.

[53] M. Demetgul and S. Lazarova-Molnar, "Data-driven artificial intelligence applications for tyre-road-noise prediction and road condition monitoring: A review and future directions," *Measurement*, vol. 264, p. 120203, Mar. 2026, doi: 10.1016/j.measurement.2025.120203.

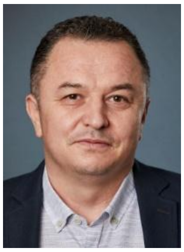
[54] M. Demetgul, Heinzelmann, Marc, and Lazarova-Molnar, Sanja, "Tyre–Road Noise Prediction Using Multi-Task Learning and LOGO Cross-Validation," presented at the DAGA2026 Dresden, Dresden, pp. 542–545. doi: 10.71568/daga2026.393.

[55] Z. Li, "Spectral Comparison of Pass-By Traffic Noise," in *ASME 2018 Noise Control and Acoustics Division Session presented at INTERNOISE 2018*, Chicago, Illinois, USA: American Society of Mechanical Engineers, Aug. 2018, p. V001T01A014. doi: 10.1115/NCAD2018-6149.

[56] D. Yang, F. Wang, H. Liu, T. Ma, H. Zhang, and W. Li, "Prediction of mid-high frequency noise in new energy electric vehicles," *Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering*, vol. 238, no. 13, pp. 3879–3891, Nov. 2024, doi: 10.1177/09544070231196271.

[57] T. Li, "Literature review of tire-pavement interaction noise and reduction approaches," *J. vibroeng.*, vol. 20, no. 6, pp. 2424–2452, Sep. 2018, doi: 10.21595/jve.2018.19935.

This article has been accepted for publication in IEEE Access. This is the author's version which has not been fully edited and content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2026.3714728



DR. MUSTAFA DEMETGÜL is a postdoctoral researcher at the KIT AIFB Institute. He is a member of the SYDSEN research group and works on the TyRoN Tyre Road Noise project. His research interests include artificial intelligence, anomaly detection, computer vision, predictive maintenance, signal processing, and structural health monitoring. He is on the editorial boards of several international journals and has published extensively in international journals on anomaly and fault detection. Before joining AIFB, he worked as a solution developer for computer vision at the ZEISS Innovation Hub KIT. Here he carried out projects on computer vision and GANs. With the support of a scholarship from the Alexander Humboldt Foundation, he also did postdoctoral research at the KIT WBK Institute, focusing on sensorless monitoring in machines and anomaly detection using computer vision and time series data. Before Germany, he gained experience in Turkey and the USA. From 2007 to 2009, he was a postdoctoral fellow in the Department of Mechanical Engineering at Florida International University (FIU), USA, where he worked on structural health monitoring and energy harvesting. He then served as Assistant Professor and later Associate Professor in the Department of Mechatronics Engineering at Marmara University, Turkey. During his tenure, he established and managed a laboratory dedicated to fault diagnosis and structural health monitoring.

PROF. DR.-ING. SANJA LAZAROVA-MOLNAR is a Professor at the Institute of Applied Informatics and Formal Description Methods, Karlsruhe Institute of Technology in Germany. In addition, she is a Professor with the Faculty of Engineering at the University of Southern Denmark in Denmark. Her research focuses on data-driven simulation, digital twins, and cyber-physical systems modeling for reliability and energy efficiency enhancement. Actively engaged in developing advanced methodologies, she leverages her expertise to optimize complex systems through data-driven simulation. She leads activities focused on digital twins and data-driven simulation in several European and national projects. Furthermore, Professor Lazarova-Molnar assumes leadership roles in IEEE and The Society for Modeling & Simulation International (SCS), currently serving as Representative of SCS to the Winter Simulation Conference (WSC) Board of Directors. She was also one of the Proceedings Editors for the Winter Simulation Conference in 2019 and 2020 and is an associate editor of SIMULATION: Transactions of The Society for Modeling and Simulation International.

LUKAS MICHAEL SCHWEMER is an assistant researcher at the Karlsruhe Institute of Technology (KIT), where he works at the Institute of Applied Informatics and Formal Description Methods (AIFB) as part of the TyRoN Tyre Road Noise project. His research focuses on deep neural networks and artificial intelligence for autonomous systems. A particular emphasis is on multimodal learning, computer vision, time-series analysis, and robust model design. His work has achieved significant performance improvements in road surface classification, contributing to increased safety in autonomous driving applications. He has authored and co-authored international conference publications in the field of machine learning.

He received his M.Sc. degree in industrial engineering from the Karlsruhe Institute of Technology (KIT), where his thesis focused on deep multimodal learning for road surface classification. Prior to his current position, he gained experience during an internship as an assistant researcher at the A*STAR Institute of High-Performance Computing (IHPC), Singapore. Here he worked on novel continual learning approaches to mitigate catastrophic forgetting in deep neural networks. His research interests include artificial intelligence, deep multimodal learning, continual learning, anomaly detection, computer vision, and time-series modeling, with a strong focus on real-world deployment and robust AI systems.