IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 11 December 2025, accepted 11 February 2026, date of publication 20 February 2026, date of current version 26 February 2026.

Digital Object Identifier 10.1109/ACCESS.2026.3666660

RESEARCH ARTICLE

Road Anomaly Classification Using Vibrational Signals With a Transfer Learning-Enhanced SepConv1D-LSTM Architecture

LORENZO MANONI, (Member, IEEE), SIMONE ORCIONI, (Senior Member, IEEE), AND MASSIMO CONTI, (Member, IEEE)

DII–Dipartimento di Ingegneria dell'Informazione, Università Politecnica delle Marche, I-60131 Ancona, Italy

Corresponding author: Massimo Conti (m.conti@univpm.it)

This work was supported by the PNRR Italian National Center for Sustainable Mobility (MOST), under SPOKE 7 "CCAM, Connected Networks and Smart Infrastructure"-WP4.

ABSTRACT Accurate road condition monitoring is critical for transportation safety and infrastructure management. Recent advances in Deep Learning and mobile sensing have enabled anomaly detection using smartphone-acquired vibrational signals. This paper proposes a hybrid SepConv1D-LSTM architecture that integrates separable convolutional layers and Long Short-Term Memory (LSTM) units, enhanced through a domain-specific transfer learning strategy. In the adopted approach, each backbone is pretrained on its optimal input data representation domain. We evaluated the proposed model on a publicly available smartphone-based dataset using a rigorous event-level stratified k-fold cross-validation protocol. Extensive experiments were conducted to compare the proposed approach with standalone models, such as SepConv1D, LSTM, GRU, and CNN-2D as well as several state-of-the-art Deep Learning methods across multiple signal representations, such as time, Discrete Wavelet Transform, FFT, and Continuous Wavelet Transform. Results demonstrate that the proposed hybrid SepConv1D-LSTM consistently outperforms existing methods in classification accuracy and robustness, highlighting the potential of hybrid architectures and domain-specific transfer learning for scalable road condition monitoring using low-cost mobile sensors.

INDEX TERMS Road condition monitoring, deep learning, convolutional neural networks, separable convolution, long short term memory, transfer learning, smartphone sensors.

I. INTRODUCTION

Road condition monitoring plays a critical role in ensuring transportation safety, infrastructure maintenance, and urban mobility planning. Traditional approaches to road surface assessment often rely on manual inspections or specialized equipment, which are costly, time-consuming, and difficult to scale. In recent years, the widespread use of low-cost sensors embedded in smartphones and vehicles has enabled the collection of vibrational signals that reflect road surface irregularities.

Processing these signals with modern Machine Learning (ML) and Deep Learning (DL) techniques provides a powerful means to develop automated, scalable and real-time road condition monitoring systems. DL models have demonstrated superior performance in extracting meaningful patterns from raw sensor data, outperforming classical ML methods in various signal classification tasks. In particular, convolutional and recurrent neural networks have been widely adopted to model spatial and temporal features in vibrational signals.

Recent state-of-the-art approaches for road condition monitoring from vibrational signals primarily rely on DL models built upon convolutional and recurrent layers. To further enhance classification performance, several studies have proposed hybrid models that combine convolutional and recurrent feature extractors. Despite these advances, significant challenges remain, including designing architectures

The associate editor coordinating the review of this manuscript and approving it for publication was Abedalrhman Alkhateeb.

that generalize well across diverse acquisition conditions, ensuring effective training when labeled data are limited, and adopting rigorous evaluation protocols that prevent data leakage by separating distinct road anomaly events between training and testing.

This paper addresses these challenges by introducing a novel hybrid DL architecture that combines separable convolutional layers (SepConv1D) with Long Short-Term Memory (LSTM) units, enhanced through a transfer learning strategy. The proposed model was evaluated on a publicly available smartphone-based dataset using a rigorous event-level k-fold cross-validation protocol, where distinct road anomaly events are assigned to training and test sets to prevent data leakage and ensure a realistic performance assessment.

The main contributions of this work are summarized as follows:

- We propose a hybrid SepConv1D-LSTM architecture that leverages the complementary strengths of convolutional and recurrent layers for road anomaly classification.
- We introduce an intra-domain transfer learning strategy, where each backbone is pretrained on its most suitable representation domain and jointly finetuned to improve convergence and generalization. The considered representation domains include the time-domain, Fast Fourier Transform, Discrete Wavelet Transform with Approximation Coefficients, Discrete Wavelet Transform with Detail Coefficients, and Continuous Wavelet Transform.
- We adopted a rigorous event-level stratified k-fold cross-validation protocol to prevent data leakage and ensure realistic evaluation.
- We provide a comprehensive experimental comparison against baseline and state-of-the-art DL models, demonstrating the superior performance of the proposed approach.

The remainder of the paper is organized as follows. Section II reviews existing literature on road condition monitoring using vibrational signals and transfer learning. Section III discusses the limitations of prior studies and motivates the approach proposed in this paper. Section IV describes the dataset, preprocessing pipeline, data representation domains and model architectures. Section V presents the experimental setup and results. Finally, Section VI concludes the paper and outlines directions for future research.

II. RELATED WORK

A. ROAD CONDITION MONITORING USING VIBRATIONAL SIGNALS

In recent years the application of ML and DL techniques has been extensively explored for the task of road condition monitoring from vibrational signals. A wide range of DL architectures and preprocessing methods have been applied to data collected with various types of sensors such as low-cost smartphone accelerometers and gyroscopes, inertial measurement units (IMUs), and high-precision inertial sensors.

Early approaches to road anomaly detection primarily relied on traditional ML techniques. For instance, Carlos et al. [1] found that SVM significantly outperformed threshold-based methods in classifying road types using synthetic data generated with the Pothole Lab simulator. Similarly, Basavaraju et al. [2] compared SVM, Decision Trees, and MLP classifiers for detecting road conditions (smooth, pothole, crack) using smartphone accelerometer data. Despite the computational cost of feature extraction, their framework supported real-time use, with MLP achieving the highest test accuracy of 92.12%.

In contrast, several subsequent works introduced more sophisticated approaches based on DL, demonstrating a clear superiority to traditional ML methods in classifying road anomalies from vibrational signals.

Menegazzo and von Wangenheim [3] used inertial data from MPU-9250 sensors such as 3-axis acceleration, 3-axis rotation and speed to classify road surfaces. The experiments demonstrated that DL models, including CNN-1D, LSTM, and CNN-LSTM significantly outperformed traditional ML methods, reaching 93.17% accuracy with a CNN-1D. Their framework was later extended to speed bump detection [4], where a CNN-LSTM achieved state-of-the-art results.

Singh et al. [5] proposed a custom binary feature layer for LSTM-based pothole detection, with their Improved-LSTM yielding the best performance.

Conversely, Sabor and Abdelraheem [6] introduced a CNN-2D with Low-High Frequency Feature Mixer blocks, based on residual connections and 2D convolutions to recognize asphalt bump, pothole or metal bump anomalies from 3-axis acceleration signals, interpreted as images. However, their study did not include a comparative evaluation against CNN-1D models trained on raw time-domain signals, leaving unresolved questions about the relative effectiveness of spatial versus temporal representations.

Ozoglu and Gökgöz [7] applied a CNN-2D to stacked 3-axis acceleration and 3-axis rotation data collected via smartphones, treating the combined signal as an image for pothole detection. Nevertheless, the work did not investigate the potential of 1D convolutional models on time-domain signals despite the simplicity of the binary classification task.

Recent years have witnessed notable advances in DL architectures for anomaly detection beyond vibrational signals. In particular, Generative Adversarial Networks (GANs) and blind-spot networks have been significantly advanced for anomaly detection in video and hyperspectral imagery. Wang et al. [8] proposed a novel global feature-injected blind-spot network within a self-supervised anomaly detection framework for hyperspectral images, incorporating patch-level shuffling and downsampling to disrupt spatial correlation among neighboring anomaly pixels. In a related study, Wang et al. [9] introduced a frequency-to-spectrum GAN to synthesize the background hyperspectral image (HSI) for anomaly detection. Singh et al. [10] investigated

anomaly detection in surveillance video and proposed a GAN framework featuring an attention-enhanced generator together with a dual-branch discriminator, enabling the model to discriminate between real and reconstructed frames, and between real and reconstructed anomalous content.

Several studies have explored crowdsourced approaches for road infrastructure condition monitoring, leveraging data collected from smartphones or low-cost sensors deployed from multiple users and vehicles. Tiwari et al. [11] proposed a real-time monitoring system based on a deep CNN trained on crowdsourced accelerometer data to detect road surface irregularities. Other works [12], [13], [14], [15] have focused on scalable frameworks for aggregating heterogeneous sensor data to enhance anomaly detection coverage and robustness in large-scale deployments.

Beyond architectural advancements, numerous studies have investigated alternative signal representation domains to enhance performance compared to raw time-domain inputs. To this end, Baldini et al. [16] demonstrated the superiority of 2D magnitude spectrograms over time-domain signals, Continuous Wavelet Transform (CWT), and Short-Time Fourier Transform (STFT) representations when used as inputs to CNN models. Their study employed acceleration and rotation data collected via a dashboard-mounted IMU for road anomaly and obstacle detection. Despite this, all anomaly types were grouped into a single class, without distinguishing between different road surface irregularities.

Likewise, Chen et al. [17] compared STFT and wavelet-based 2D representations of smartphone accelerometer data, using a CNN-2D to classify normal roads, transverse cracks, and manholes. However, the study did not explore the potential of 1D convolutional models applied to raw time-domain signals, nor did it assess whether simpler temporal representations could achieve comparable performance.

Alternatively, Martinelli et al. [18] extracted entropy and coefficient of variation from STFT sub-bands of z-axis acceleration to classify road surface conditions using SVM, DT, and KNN, distinguishing between no distress, short-term and long-term anomalies such as potholes and fatigue cracking. Despite this, the paper did not explore the benefits of DL models over traditional ML methods.

Recently, the effectiveness of hybrid CNN-LSTM models in classifying road anomalies from inertial sensor vibration data has been assessed. Raslan et al. [19], [20] proposed hybrid CNN-LSTM architectures for road anomaly classification using smartphone-based inertial and GPS data. Their models combined time and frequency-domain inputs and were evaluated using stratified k-fold cross-validation.

B. TRANSFER LEARNING

Transfer learning has emerged as a powerful strategy to improve DL model performance, particularly in scenarios with limited labeled data, by transferring knowledge from one domain to another [21].

In the context of road condition monitoring, transfer learning has been applied by using pretrained DL backbones on data from domains different from the target domain and derived from very large datasets.

Martinez-Ríos et al. [22] employed several state-of-the-art pretrained architectures such as GoogleNet [23], SqueezeNet [24], and ResNet [25] to detect transverse pavement cracks from Generalized Morse Wavelet images. These images were computed from denoised vertical acceleration data collected by a high-precision piezoelectric sensor.

To the best of our knowledge, existing studies applying transfer learning in the context of road condition monitoring have primarily adopted a cross-domain approach, leveraging models pretrained on large-scale datasets from different tasks or sensor modalities.

In contrast, inspired by the direction suggested by Otović et al. [26], this work explored the potential of intra-domain transfer learning, where both pretraining and finetuning are performed within the same application domain but across different signal representation domains.

III. MOTIVATION

Several studies that applied Deep Learning to vibrational signals for road condition monitoring, introduced novel hybrid multi-input architectures combining CNN and LSTM branches. The hybrid models were trained using data from different representation domains for CNN and LSTM backbones by carefully selecting the most appropriate domain for each input. Results consistently demonstrated superior classification performance compared to standalone CNN and LSTM models.

However, a closer examination of the existing literature reveals a number of methodological limitations that raise concerns regarding the generalizability, reproducibility, and practical applicability of the reported findings.

- Raslan et al. [19] proposed a hybrid SepConv1D-BiLSTM model for road anomaly classification using 3-axis accelerometer, gyroscope, and GPS data. The model was trained with stratified k-fold cross-validation using time-domain input for the CNN branch and frequency-domain input for the BiLSTM. Despite its novelty, the study suffers from significant limitations since the dataset comprised only 301 labeled sequences, with imbalanced class distributions and no detailed breakdown of category counts. Additionally, the absence of performance comparisons against a CNN-1D baseline restricts the evaluation of the hybrid model's effectiveness. These factors collectively undermine the generalizability and reproducibility of the reported results.
- In a subsequent study Raslan et al. [20] introduced a hybrid CNN-LSTM model for classifying road segments into four categories: normal road, potholes, speed bumps, and bad road. The performance of the proposed model was assessed by comparing it with other DL

architectures chosen as baselines. All the models were trained with a stratified k-fold cross-validation method (k=5) using several data representation domains such as time domain, FFT, Discrete Wavelet Transform, and Continuous Wavelet Transform.
To mitigate issues related to limited dataset size and class imbalance, the authors applied data augmentation by segmenting the signals of underrepresented classes with a fixed-length sliding window and class-specific overlap values.
Nevertheless, it was not specified in the paper whether data augmentation was performed before or after the generation of the k folds. Furthermore, the class consistency of the generated folds was not disclosed.
The potential presence of frames generated from the signal window event in the training and validation folds could lead to data leakage wherein the model is exposed during validation to signal patterns it has already seen in training. Such a scenario deviates from realistic deployment conditions, where input signals correspond to entirely distinct events that the model has not previously encountered during training.
- In both studies, data collection was performed with a 3-axis accelerometer and gyroscope placed in a fixed position in a single vehicle for all anomaly events. A limited diversity in data collection setups can significantly constrain the generalization ability of a hybrid architecture and hinder a reliable assessment of its performance.

To overcome the aforementioned limitations, this work proposed two key methodological innovations. First, a more rigorous evaluation protocol was adopted through an event-level k-fold cross-validation strategy, which ensures that training and validation sets contain entirely distinct signal events. This prevents any form of data leakage due to overlapping windows and better reflects realistic deployment scenarios.

Second, a novel hybrid SepConv1D-LSTM architecture was proposed and trained using a transfer learning strategy. In this approach, each backbone (SepConv1D and LSTM) was first independently pretrained on data from its most appropriate representation domain. The entire model was then fine-tuned jointly with a reduced learning rate to enhance convergence and performance.

This strategy leverages prior knowledge while preserving generalization capabilities of the hybrid architecture. The proposed approach is particularly beneficial in scenarios with limited data and diverse acquisition conditions, leading to superior classification performance compared to training the hybrid model from scratch.

IV. MATERIALS AND METHODS

A. DATASET DESCRIPTION

The dataset employed in this study was originally introduced by Gonzalez et al. [27] for the classification of road surface anomalies using data acquired from smartphone sensors. It comprises vibration signals collected with five Motorola Moto G smartphones, each equipped with an STMicroelectronics LIS3DH tri-axial accelerometer. Data acquisition was performed at a fixed sampling rate of 50 Hz, which reflects a realistic and practical configuration for commercial mobile devices. This setting ensures compatibility with low-cost, resource-constrained deployment scenarios.

To simulate real-world usage conditions, the smartphones were placed in various positions inside the vehicles, including the driver's door pocket, the central console, the driver's shirt and pants pockets, and a passenger's handbag located on the co-pilot seat. These heterogeneous placements introduce significant variability in the recorded signals, thereby increasing the complexity and realism of the dataset.

The recordings were collected using 12 different vehicles and categorized into five road condition classes: asphalt bumps, metal bumps, potholes, regular road, and worn out road. Each class is represented by 100 labeled samples, resulting in a balanced dataset with a total of 500 recordings.

Only the z-axis acceleration component was retained in the dataset, since road anomalies predominantly affect vertical motion, making this axis significantly more informative than the others.

The distribution of labeled events per vehicle is detailed in Table 1. Illustrative examples of z-axis acceleration signals corresponding to asphalt bump and worn out road events are shown in Figures 1 and 2, respectively.

This dataset was subsequently reused in a later study by [28], which introduced a comprehensive benchmark framework based on classical ML algorithms, including Decision Trees (DT), Support Vector Machines (SVM), and Multi-Layer Perceptrons (MLP). The study aimed to evaluate the effectiveness of various handcrafted feature extraction techniques for road anomaly classification. In this benchmark, each of the 500 vibration signal recordings was segmented into overlapping frames of fixed length w using a 50% overlap between consecutive windows. From each segmented frame, a feature vector of fixed size was extracted by computing a high number of statistical descriptors across three domains: time, frequency, and wavelet (using Daubechies-2 wavelets). Multiple combinations of these feature domains were tested to determine the most effective configuration for classification. The best overall performance was achieved by combining time-domain and wavelet-domain features. An MLP classifier trained with a window length of $w = 30$ attained the highest accuracy, reaching 52%.

Despite providing a valuable baseline, both the original dataset paper [27] and the classical ML benchmark by [28] present several limitations. The Bag-of-Words representation introduced in [27], for instance, is a relatively simplistic feature extraction technique tailored specifically to the dataset's structure. This narrow customization raises concerns regarding its generalizability to different data collection scenarios, sensor configurations, and deployment environments.

TABLE 1. Dataset consistency across vehicle types.

Vehicle	Pothole	Asphalt bump	Metal bump	Worn out	Regular road
Nissan Sentra 2013	3	4	6	8	7
Nissan Tsuru	1	13	33	0	0
Toyota Camry	15	1	1	17	11
Chevrolet	39	23	17	15	9
Chevrolet S10	1	9	2	8	3
Volkswagen Jetta	4	2	4	4	2
Nissan Sentra	4	5	2	8	14
Nissan Frontier	9	1	6	8	11
Volkswagen Beetle	6	6	9	6	9
Nissan Altima	7	3	2	7	1
Chevrolet Monza	0	12	6	0	20
Volkswagen Pointer	11	21	12	19	13
Total	100	100	100	100	100

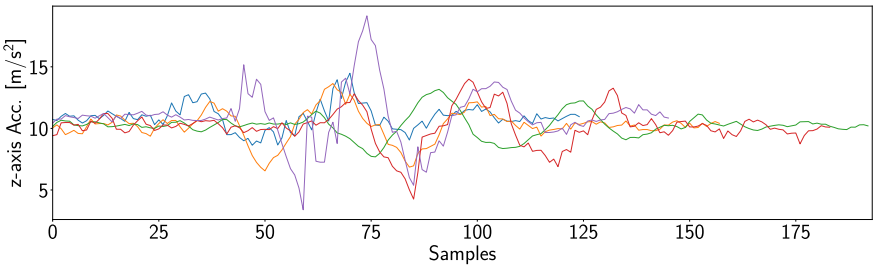


FIGURE 1. Examples of recorded signals for some asphalt bump events.

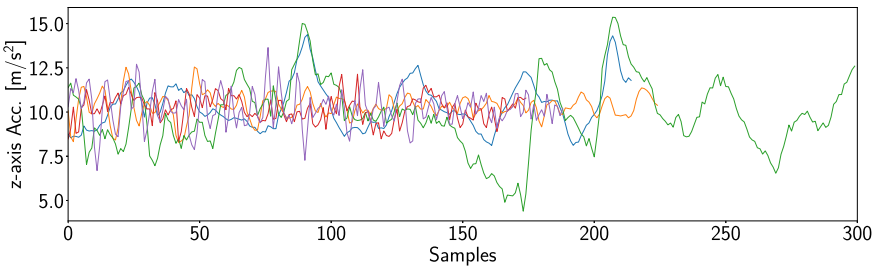


FIGURE 2. Examples of recorded signals for some worn out road events.

On the other hand, while the benchmark proposed in [28] provided a valuable investigation into the impact of handcrafted features extracted from various signal domains when applied to classical ML algorithms,it did not explore the capabilities of more advanced DL models. Such architectures have demonstrated superior performance in a wide range of sequence and signal classification tasks, owing to their ability to automatically learn hierarchical and discriminative feature representations directly from raw input data.

B. EVENT-LEVEL K-FOLD CROSS-VALIDATION

To ensure a realistic and unbiased estimation of model generalization capabilities, this study employed an event-level stratified k-fold cross-validation scheme. Unlike sample-level splitting, which may inadvertently introduce temporal correlation between training and test sets, an event-level splitting method ensures that no segments from the same anomaly instance appear in both training and test sets.

Formally, let $\mathcal{A}_c = \{a_{c,1}, a_{c,2}, \ldots, a_{c,N_c}\}$ denote the set of z-axis acceleration recordings for class c, where N_c is the number of recordings for class c. These are partitioned into k disjoint folds:

$$\mathcal{A}_c = \bigcup_{j=1}^{k} \mathcal{A}_c^{(j)}, \quad \mathcal{A}_c^{(j)} \cap \mathcal{A}_c^{(l)} = \emptyset \text{ for } j \neq l$$

ensuring $\left|\mathcal{A}_c^{(j)}\right| = N_c/k$. For each fold j, the training set comprises the segments generated from $\mathcal{A}_c \setminus \mathcal{A}_c^{(j)}$ for all c, and the test set is formed from $\bigcup_{c=1}^{C} \mathcal{A}_c^{(j)}$.

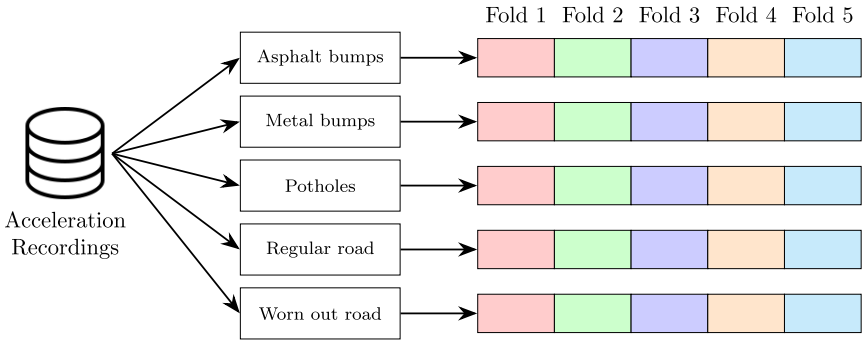


FIGURE 3. Event-level stratified 5-fold split technique adopted in this study.

This approach guarantees stratification across all classes and replicates realistic deployment conditions, where anomalies observed during testing are disjoint from those seen during training.

In this work, a k-fold event-level split strategy was adopted with a number of folds $k = 5$. The overall partitioning scheme is illustrated in Fig. 3.

C. DATA PIPELINE FOR TRAINING AND VALIDATION

The pipeline adopted in this study to generate training and validation data for the DL models involves three fundamental stages: segmentation using fixed-size overlapping windows, transformation into a specific representation domain, and normalization.

1) **Segmentation** Each time-domain z-axis acceleration signal $a_{c,n}$ was segmented using a sliding window of fixed length W. The segmentation served two purposes: first to standardize input length across samples, enabling uniform processing by DL models; second to augment the dataset by generating multiple partially overlapping windows from each signal, thus increasing the number of training samples.
To address class imbalance, a class-dependent overlap ratio O_c was applied. Underrepresented classes were assigned higher overlap values, thereby increasing their number of training samples. This class-specific augmentation approach led to a more balanced dataset across the five categories and improved the robustness of the supervised learning task.
2) **Domain transformation** After segmentation, each time-domain window was either retained in its raw form or transformed into an alternative representation domain to enhance feature representation. The following representations were considered: Fast Fourier Transform (FFT), Discrete Wavelet Transform (DWT) with Approximation Coefficients (DWT-CA), Discrete Wavelet Transform with Detail Coefficients (DWT-CD), and the Continuous Wavelet Transform (CWT). A detailed description of the data representation domains used in this study is provided in Subsection IV-D.
3) **Normalization** To ensure stable and unbiased model training, all input data were normalized prior to being fed into the DL models. For all representations except CWT, Min-Max normalization was applied to scale the values into (0,1) range, which is the most commonly used in Machine Learning and Deep Learning to prevent any single feature from dominating due to its magnitude.
Importantly, normalization parameters were computed exclusively on the training set to prevent information leakage into the validation set.
Let $X_d \in \mathbb{R}^{B\times W_d\times N_{\text{var}}}$ represent the training or validation data in a specific domain d, where B is the number of samples, W_d the sequence length, which depends on the segmentation window length W and the representation domain d. Since the input signal includes only the z-axis acceleration, $N_{\text{var}} = 1$. The normalization was computed as:

$$X^{\text{norm}}_{b,n,v} = \frac{X_{b,n,v} - X^{\min}_v}{X^{\max}_v - X^{\min}_v}, \quad \forall b \in \mathcal{T}\cup\mathcal{V},\ \forall n, v \quad (1)$$

where $X^{\min}_v$ and $X^{\max}_v$ are the minimum and maximum values computed over the training set $\mathcal{T}$.
For CWT representations, which produce 2D time-frequency matrices $X^{\text{cwt}} \in \mathbb{R}^{N\times S_L}$, where S_L is the number of CWT scales, normalization was applied independently to each sample:

$$X^{\text{cwt-norm}}_{b,n,s} = \frac{X^{\text{cwt}}_{b,n,s} - X^{\min}_b}{X^{\max}_b - X^{\min}_b}, \quad \forall b \in \mathcal{T}\cup\mathcal{V},\ \forall n, s \quad (2)$$

where $X^{\min}_b$, $X^{\max}_b$ were computed independently for each sample.

The overall framework employed for data preprocessing and DL model training andvalidation across the event-level stratified k-fold splits is illustrated in Fig. 4. The complete data preparation and training workflow, including segmentation, representation transformation, normalization, and model evaluation, was applied consistently across all DL

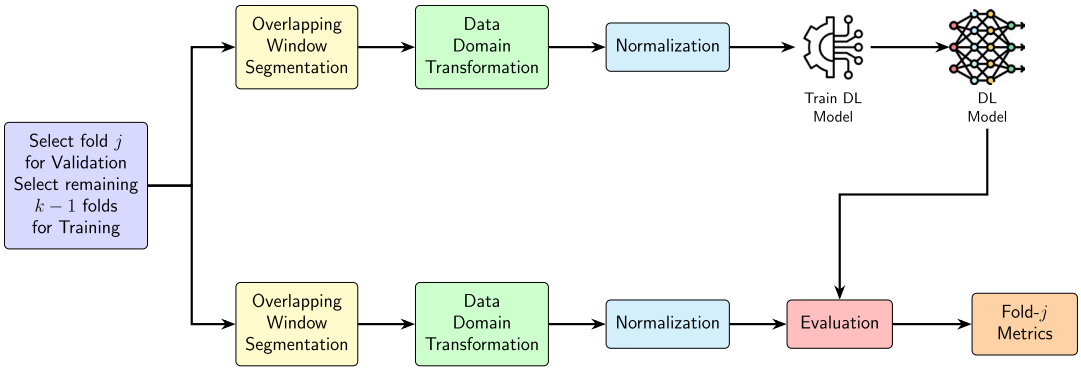


FIGURE 4. Structure of the overall pipeline adopted in this study for data preprocessing and DL model training/evaluation across folds.

TABLE 2. Samples consistency of each of the generated folds for each category.

Class	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	All folds
Asphalt Bumps	92	91	97	95	93	468
Metal Bumps	110	109	109	105	104	537
Potholes	140	138	139	139	144	700
Regular Road	112	114	115	112	118	571
Worn out	110	119	115	112	107	563
All Classes	564	571	575	563	566	2839

architectures and representation domains. This ensured fair comparisons and reproducibility of results.

A total of 2839 data samples were generated using the sliding window technique with a window size $W = 110$. The window size was selected by balancing a compromise: it was long enough so that the entire event dynamics were included in the segmented window for all categories while still generating a sufficient number of samples for all classes. In preliminary experiments window lengths from 100 to 130 yielded stable performance, with accuracy variations below 1%. This indicates that the proposed method and the reported results are robust to moderate variations of the window size. Tab. 2 reports the number of samples of all the five folds for each category. The use of class-specific overlap ratios O_c resulted in a well-balanced distribution of samples, enhancing the robustness of the supervised learning process.

D. DATA REPRESENTATION

In the preprocessing pipeline adopted in this study, the time-domain signals produced with the fixed-size sliding window were transformed into alternative representations to enhance their suitability for pattern-recognition tasks using DL methods.

The choice of signal representation plays a pivotal role in determining the performance of DL models, particularly when dealing with complex and noisy time-series data such as vehicle-induced vibration signals. While time-domain representations preserve the raw temporal structure of the signal, they may fail to highlight localized frequency patterns or longer-term oscillatory structures that are indicative of specific road anomalies.

To address this, several signal representation techniques were explored in this study to improve the discriminative power of the input data.

1) TIME-DOMAIN

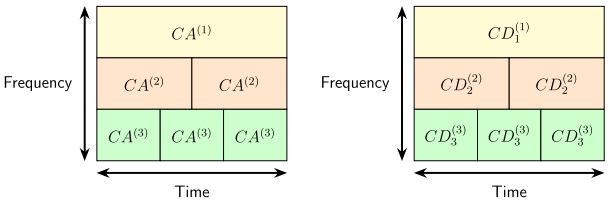
The raw z-axis acceleration signal, segmented into fixed-size windows using the sliding window procedure, was used as input in its original time-domain form. This representation retains the full temporal resolution and fine-grained variations of the signal, which is essential for capturing sequential patterns and transient events.

2) FAST FOURIER TRANSFORM (FFT)

The magnitude spectrum of each time-domain segmented window was computed using the Fast Fourier Transform. Only the first $W/2$ frequency bins were retained due to spectral symmetry. This representation emphasizes the dominant frequency components of the vibrational patterns induced by different road anomalies.

3) DISCRETE WAVELET TRANSFORM (DWT)

The DWT is a powerful signal processing technique that allows the decomposition of a signal into multiple frequency bands, each localized in time. Unlike the Fourier Transform, which provides only global frequency information,



(a) *CA* coefficients **(b)** *CD* coefficients

FIGURE 5. Structure of input data matrix to DL models for DWT representation.

DWT offers a joint time-frequency representation, enabling the analysis of how different frequency components evolve over time. At each level of decomposition, DWT operates by passing the signal through two types of filters.

- A low-pass filter, which extracts the slow, smooth variations in the signal, producing the approximation coefficients (CA)
- A set of band-pass and high-pass filters, which capture faster changes or finer details, resulting in the detail coefficients (CD).

After filtering, the signal is downsampled to reduce its resolution, and the process is repeated across multiple levels. This results in a hierarchical, multiscale representation that captures both coarse and fine-grained features of the signal.

In this study, DWT was applied using Daubechies-4 wavelets over three levels of decomposition. For each input signal window, both the approximation coefficients (DWT-CA) and the detail coefficients (DWT-CD) were extracted and evaluated as alternative input representations. This approach was motivated by the fact that different types of road anomalies manifest in distinct frequency ranges. For instance, bumps and potholes usually produce low-frequency structural changes, whereas surface roughness or worn out segments might introduce higher-frequency fluctuations.

By providing a multiresolution view of the signal, DWT allows DL models to access discriminative features, that may not be easily detectable in the raw time-domain data. Formally, applying a level-l DWT analysis to a signal vector $x \in \mathbb{R}^W$ yields:

- A single output vector of approximation coefficients $CA^{(l)}$
- A list of l vectors of detailed coefficients $CD_j^{(l)}$, $j = 1, \ldots, l$

To construct the input tensor for the learning models, either the approximation coefficients $CA^{(1)}, CA^{(2)}, CA^{(3)}$ or the detailed coefficients $CD_3^{(1)}, CD_3^{(2)}, CD_3^{(3)}$ were stacked along the frequency axis. Since the length of these vectors vary across decomposition levels, the coefficients of higher levels were repeated along time-axis to form a consistent matrix shape, as illustrated in Fig. 5.

4) CONTINUOUS WAVELET TRANSFORM (CWT)

The CWT provides a time–frequency representation of a signal by projecting it onto a set of wavelets continuously scaled and shifted along the time axis. Unlike the DWT, which analyzes the signal at discrete scales and involves downsampling, the CWT maintains the full temporal resolution. It produces a smooth and highly redundant representation that captures both fine and coarse features without loss of temporal focalization.

In this study, the CWT was computed using the complex Morlet wavelet, characterized by a bandwidth parameter $\beta = 1.5$ and center frequency $\gamma = 1.0$. The CWT coefficients were converted to magnitude power and log-scaled. A total of 60 scale values were used to compute the CWT. Cone-of-Influence masking was applied to mitigate boundary effects.

The result is a 2D time−frequency map that can be interpreted as a spectrogram, capturing both transient and periodic features of the signal. Figures 6, 7, 8, 9 and 10 show the computed representations used in this study, namely time domain, FFT, DWT (approximation coefficients) and CWT for each of the considered anomaly events: asphalt bump, metal bump, pothole, regular road and worn out road.

E. DL ARCHITECTURES

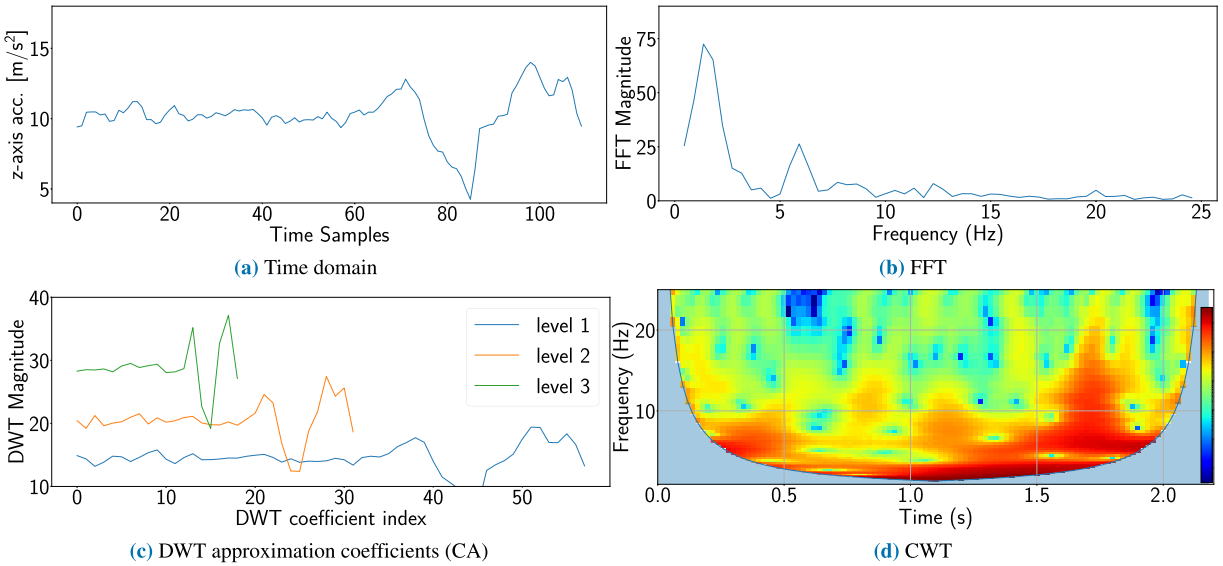
This section presents the DL architectures chosen as the baseline models for this study including a CNN-1D model based on 1D separable convolutions (named SepConv1D), two recurrent models based on LSTM and Gated Recurrent Units (GRU) and a CNN-2D, respectively designed for time−frequency inputs.

1) SepConv1D

The SepConv1D architecture is built upon the SeparableConv1D layer, a widely adopted technique in DL that decomposes a standard 1D convolution into two distinct operations: a depthwise convolution followed by a pointwise convolution. This decomposition significantly reduces the number of trainable parameters while preserving strong representational capacity, making it well-suited for deployment on resource-constrained edge devices.As shown in [29], [30], and [31] SepConv1D offers clear advantages in terms of model size (in both.h5 and.tflite formats) and total parameter count.

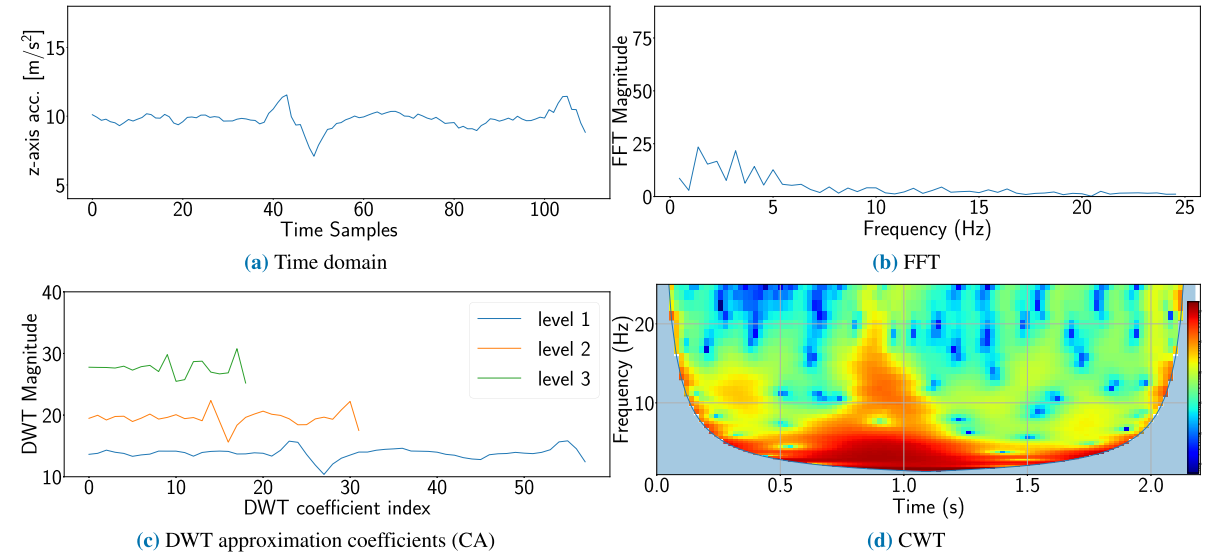
The structure of the SeparableConv1D layer used in our model is depicted in Fig. 11. The input tensor $X \in \mathbb{R}^{W \times C_{in}}$ is first processed by a depthwise convolutional layer, which applies a set of C_{in} independent 1D filters of length D, one per input channel. This operation captures local temporal patterns within each channel. The resulting feature maps are then passed through a pointwise convolution, which uses a number C_{out} of 1D filters with a length of 1 to combine information across channels, producing an output tensor of shape $W \times C_{out}$.

The full SepConv1D architecture, shown in Fig. 12, begins with an initial standard 1D convolutional layer with 64 filters



(a) Time domain (b) FFT (c) DWT approximation coefficients (CA) (d) CWT

FIGURE 6. Data representations of a "asphalt bump" event signal.



(a) Time domain (b) FFT (c) DWT approximation coefficients (CA) (d) CWT

FIGURE 7. Data representations of a "metal bump" event signal.

and a kernel size of 10. This is followed by a BatchNormalization, ReLU activation, and SpatialDropout1D layer with a rate of 0.2.

Next, four consecutive blocks are applied, each comprising a SeparableConv1D layer, BatchNormalization, ReLU activation and a Dropout of 0.2. To progressively reduce the temporal resolution, a MaxPooling1D with a stride of 2 is inserted before the Dropout in the first and third block. All SeparableConv1D layers have 128 output channels, a kernel size of 7 and "same" padding.

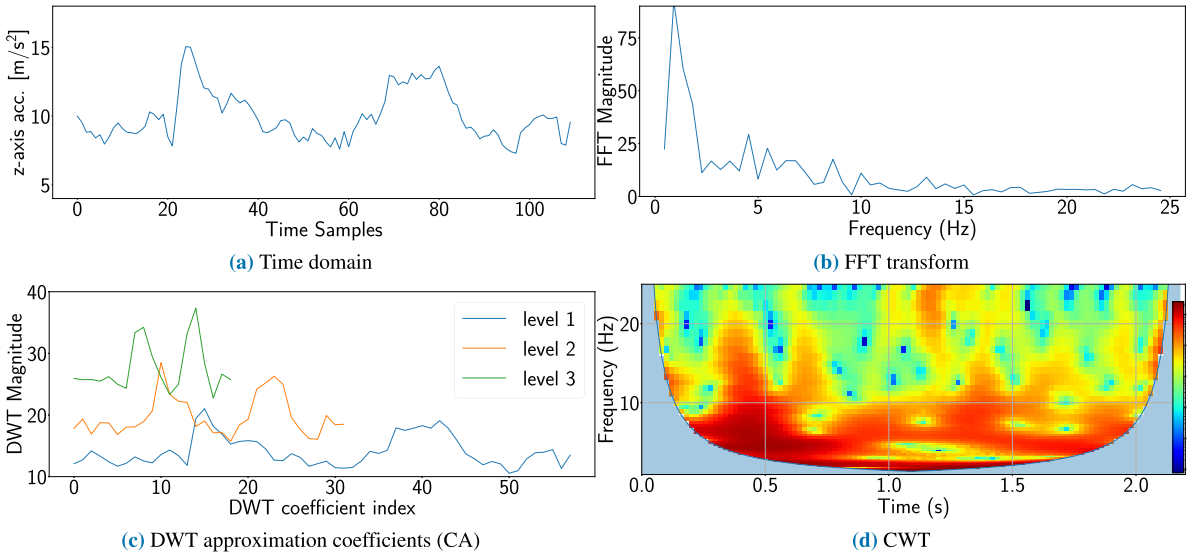
A global average pooling (GAP) layer is applied to the output of the fourth block. This is followed by a BatchNormalization and a Dropout of 0.2 to mitigate overfitting. Finally the output is provided to the final Dense layer with a Softmax activation, which produces the output class probabilities.

The portion of the network from the first Conv1D layer up to the input of the final Dense layer is referred to in this paper as the "SepConv1D backbone".

2) LONG SHORT-TERM MEMORY (LSTM)

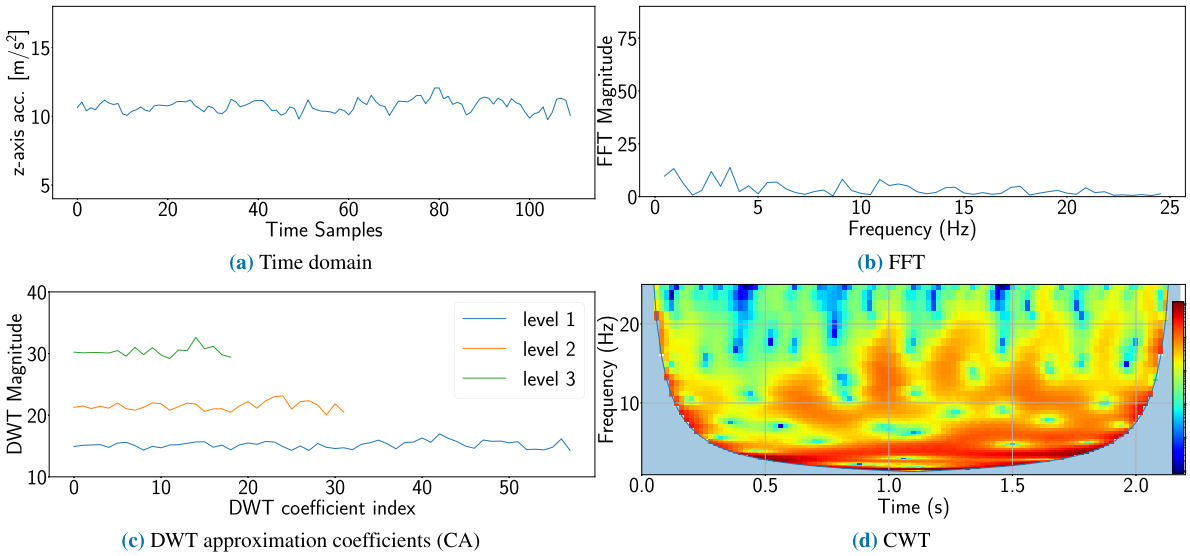
The LSTM architecture is a specialized form of recurrent neural network (RNN) designed to capture long-range dependencies in sequential data. This makes it particularly suitable for time-series modeling tasks such as accelerometer-based signal classification.

In this study, the LSTM model consists of two stacked LSTM layers, each with $U = 200$ units. The first layer is configured with `return_sequences = True`, allowing it to output the full temporal sequence of cell states with a shape $W \times U$. The second LSTM layer is configured with `return_sequences = False`, meaning it



(a) Time domain (b) FFT transform (c) DWT approximation coefficients (CA) (d) CWT

FIGURE 8. Data representations of a "pothole" event signal.



(a) Time domain (b) FFT (c) DWT approximation coefficients (CA) (d) CWT

FIGURE 9. Data representations of a "regular road" event signal.

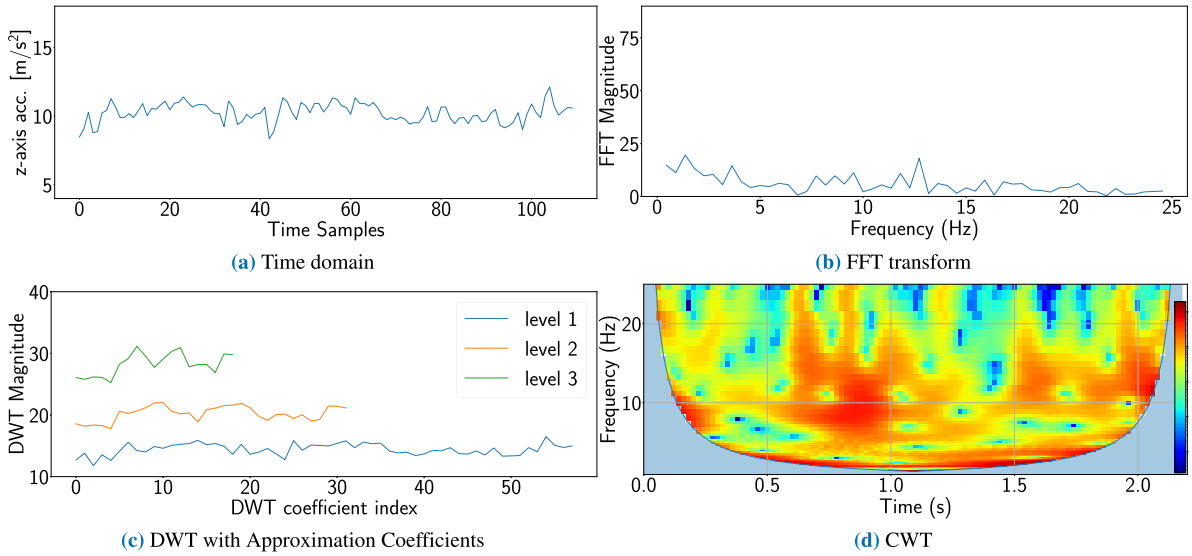
outputs only the final cell state, represented by a vector of size $U \times 1$.

To improve generalization and training stability, each LSTM layer is followed by a BatchNormalization and a Dropout layer with a rate of 0.5. This regularization strategy helps reduce overfitting and accelerates convergence. The LSTM architecture used in this study is shown in Fig. 13. The portion of the architecture from the input up to the input of the final Dense layer is referred to in this paper as "LSTM backbone".

3) GATED RECURRENT UNIT (GRU)

The GRU is a streamlined variant of the LSTM that retains the ability to model temporal dependencies while reducing computational complexity. In contrast to LSTM, GRU uses a simplified gating mechanism, with only two gates: an update gate and a reset gate, and does not maintain a separate cell state. This design allows GRUs to be trained faster and require fewer parameters, making them better suited for resource-constrained environments.

In this work, the GRU model consists of two stacked GRU layers, each with $U = 200$ cell units. Similarly to the LSTM model, the first GRU layer is configured with `return_sequences = True`, producing a full sequence of cell states with shape $W \times U$. Contrarily, the second layer uses `return_sequences = False`, outputting only the final state as a feature vector of shape $U \times 1$. Likewise, both GRU layers are followed



(a) Time domain (b) FFT transform (c) DWT with Approximation Coefficients (d) CWT

FIGURE 10. Data representations of a "worn out road" event signal.

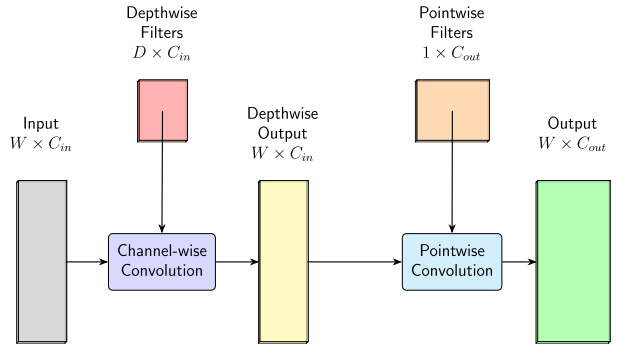


FIGURE 11. Structure of the SeparableConv1D layer.

by a BatchNormalization and a Dropout with rate 0.5 to mitigate overfitting and to improve training convergence. The architecture of the GRU model used in this work is shown in Fig. 14.

4) TWO-DIMENSIONAL CONVOLUTIONAL NEURAL NETWORK (CNN-2D)

A CNN-2D was employed in this study to process time-frequency representations of the input signals, specifically spectrograms derived from the CWT. This approach enables the model to exploit both temporal and spectral patterns of vibration signals by treating the spectrograms as 2D images.

The adopted CNN-2D architecture, shown in Fig. 15, is composed of four sequential convolutional blocks, each comprising a 2D convolutional layer with a kernel size of 7×7, followed by BatchNormalization, ReLU activation, MaxPooling2D with a pool size of 2×2 and a Dropout with a rate of 0.2. The number of filters increases progressively across the blocks: 32, 64, 128, and 256.

After the final convolutional block, the feature maps are flattened and passed through a Dense layer with 128 units and a ReLU activation. The final Dense layer with a Softmax activation, produces the class probabilities.

F. PROPOSED HYBRID SepConv1D-LSTM BASED ON TRANSFER LEARNING

To address the limitations identified in previous studies, this work introduced a novel improved hybrid architecture, which combines the strengths of two complementary components: a SepConv1D and LSTM backbone. The outputs of these branches are concatenated and passed through a Dense layer with 128 units and ReLU activation, producing a combined latent space representation of the learned features. This is followed by a final Dense layer with Softmax activation, to generate class probabilities. The architecture of the proposed hybrid SepConv1D-LSTM is illustrated in Fig. 16.

In this work, the input data domains for SepConv1D and LSTM branches were empirically chosen by selecting the best-performing configurations of their respective standalone models, as detailed in Section V. Specifically, the SepConv1D branch receives DWT-CA inputs, while the LSTM branch processes time-domain signals.

As opposed to prior hybrid models, the effectiveness of the proposed SepConv1D-LSTM architecture is further enhanced through a transfer learning strategy, which significantly improves training efficiency and generalization performance.

For each fold in the cross-validation process, the SepConv1D and LSTM branches are initialized with weights obtained from pretraining the respective standalone architectures, described in Subsection IV-E. The entire hybrid model, including both backbones and the classification head, is then fine-tuned using a reduced learning rate.

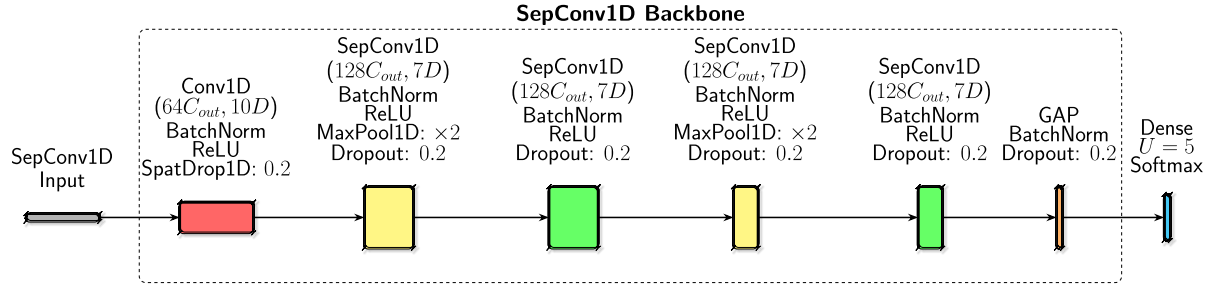


FIGURE 12. Architecture of SepConv1D.

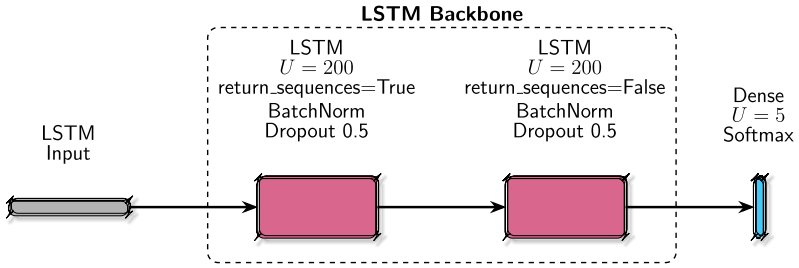


FIGURE 13. LSTM Architecture.

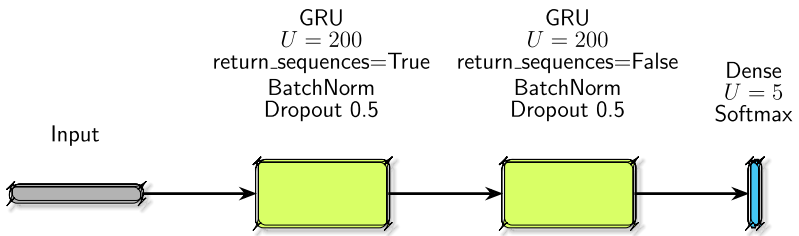


FIGURE 14. GRU Architecture.

This two-stage training approach offers several advantages. By leveraging high-quality features learned during pretraining, the poor convergence associated with random weight initialization is mitigated. In fact, by transferring prior knowledge from the pretrained backbones, the model starts training from a more informative initial state, enabling faster convergence and improved generalization across different anomaly categories. Moreover, this approach preserves the domain-specific strengths of each backbone, SepConv1D for spatial feature extraction, and LSTM for long-range temporal pattern modeling, without forcing them to adapt jointly to incompatible representations. Fig. 17 presents the transfer learning strategy employed in this study to improve the performance and the generalization capability of the proposed hybrid SepConv1D-LSTM model.

The strategy consists of transferring the learned weights of the SepConv1D and LSTM components to the hybrid SepConv1D-LSTM architecture, followed by a final fine-tuning stage.

TABLE 3. Details of environment used for experiments.

Library	Version
Python	3.11.13
TensorFlow	2.18.0
NumPy	1.26.4
PyWavelets	1.8.0

V. EXPERIMENTAL RESULTS

All experiments were conducted in a Google Colab environment equipped with a NVIDIA Tesla T4 GPU. The models were implemented using TensorFlow library with Keras backend. For signal transformation tasks, including the multi-level DWT analysis and CWT, PyWavelets library was employed. Basic array operations and numerical processing were performed using NumPy. The software environment details are summarized in Table 3.

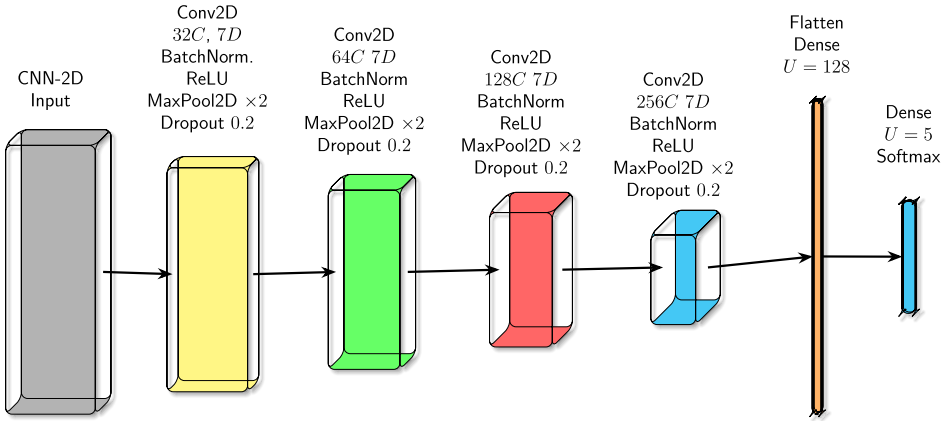


FIGURE 15. CNN-2D Architecture.

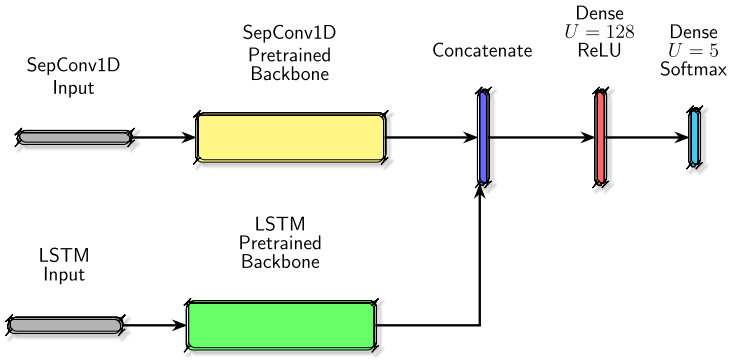


FIGURE 16. Hybrid SepConv1D-LSTM architecture.

All models were trained using the categorical crossentropy loss with Adam optimizer. An initial learning rate of 0.001 and a batch size of 24 were chosen. The learning rate was progressively decreased with a ''Reduce on Plateau'' scheduler, which multiplied the learning rate by a factor 0.75 every 100 epochs without improvement in validation accuracy. The training was terminated either after 400 epochs without improvement in validation accuracy or upon reaching a maximum of 1000 epochs.

For the proposed hybrid model, fine-tuning was performed after pretraining the individual SepConv1D and LSTM models. The same loss function and same optimizer were used but with an initial learning rate of 3×10^{-4}. The learning rate was decreased by a factor 0.75 every 50 epochs without validation accuracy improvements, and training was stopped after 200 stagnant epochs.

Tab. 4 summarizes the input representations and associated input shapes used for each of the baseline DL architectures tested in this study. For the hybrid SepConv1D-LSTM the input size with DWT-CA data for SepConv1D and Time data for LSTM are reported.

TABLE 4. Data types and input dimensions for each of the deep learning architectures.

Model	Data Type(s)	Input Shape(s)
SepConv1D	Time	(110, 1)
	FFT	(55, 1)
	DWT-CA	(58, 3)
	DWT-CD	(58, 3)
LSTM	Time	(110, 1)
	FFT	(55, 1)
	DWT-CA	(58, 3)
	DWT-CD	(58, 3)
GRU	Time	(110, 1)
	FFT	(55, 1)
	DWT-CA	(58, 3)
	DWT-CD	(58, 3)
CNN-2D	CWT	(110, 60, 1)
SepConv1D-LSTM	DWT-CA + Time	(58, 3) + (110, 1)

A. BASELINE DL MODELS

This section presents the performance evaluation of the baseline DL models: SepConv1D, LSTM, GRU and CNN-2D. Each model was trained using the event-level 5-fold cross-validation strategy described earlier, across all data representation domains, as summarized in Tab. 4.

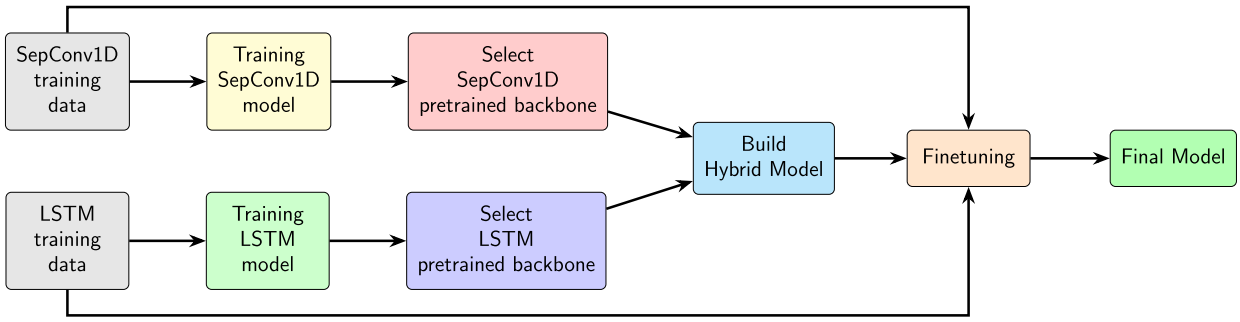


FIGURE 17. Flowchart of the adopted transfer learning strategy for the proposed hybrid SepConv1D-LSTM model.

1) SepConv1D

Tab. 5 summarizes the classification results obtained by SepConv1D architecture across all input data types, reporting 5-fold average Precision, Recall, F1-score for each class, along with overall metrics, including Accuracy, averaged across all classes and folds.

Among all input representations, DWT-CA yielded the best performance across all metrics with an average accuracy of 88.56%. Although the time-domain signals provided competitive results with an average accuracy of 86.78%, they were consistently outperformed by the DWT-CA representation across all metrics. Conversely, both FFT and DWT-CD representations achieved substantially lower classification effectiveness, with average accuracies of 75.05% and 69.79%, respectively.

Using DWT-CA inputs, the metal bumps class attained the best results, reaching an F1-score of 95.81%. Asphalt bumps and potholes were also classified with high accuracy, achieving F1-scores of 93.24% and 92.77%, respectively. Worn out road was the most challenging category, with an average F1-score of 78.33%. The regular road class showed intermediate performance, with an average F1-score of 82.26%.

2) LSTM

Tab. 6 presents the 5-fold average classification results for LSTM model. For the LSTM model, the time representation yielded the highest overall classification performance, achieving an average accuracy of 88.06% and F1-score of 88.02%. DWT-CA representation reached a comparable but slightly worse performance than time-domain signals with an accuracy of 87.36%. FFT and DWT-CD inputs performed markedly worse across all classes, demonstrating significantly lower classification quality compared to both DWT-CA and Time inputs.

As with SepConv1D, by using the best representation domain (time signals for LSTM), the worn out road category exhibited the weakest performance among all classes, with an F1-score of 75.51%. Conversely, the asphalt bumps, metal bumps, and potholes categories were all classified with high accuracy, each achieving F1-scores exceeding 92%. The regular road category showed intermediate performance, with a F1-score of 87.97%.

3) GRU

Tab. 7 reports the classification results obtained by GRU model for all data representations including per class and overall metrics averaged across folds. The GRU model showed comparable performance with time-domain and DWT-CA inputs, achieving accuracies of 87.71% and 87.11%, respectively. As observed with SepConv1D and LSTM architectures, the FFT and DWT-CD representations resulted in substantially reduced classification performance with average accuracies of 73.78% and 69.68%, respectively.

Concerning the results with time-domain inputs, the best class-specific performances were observed for metal bumps and asphalt bumps, achieving F1-scores of 97.02% and 93.01% respectively. The potholes category was also detected with high reliability, reaching an average F1-score of 92.20%. As with the other models, the worn out road category was the most challenging to classify, with an F1-score of 77.47% when using time-domain inputs. The classification quality for the regular road category was slightly above that of the worn out road, with an average F1-score of 78.90%.

4) CNN-2D

Table 8 shows the performance of the CNN-2D model trained on CWT inputs, with 5-fold average Precision, Recall, F1-score, and Accuracy.

The CNN-2D model, trained on CWT inputs, achieved an average accuracy of 82.53%. The highest class-specific performance was observed for the metal bumps category, reaching an average F1-score of 92.25%. High classification quality was also observed for the potholes and asphalt bumps categories, with F1-scores of 87.19% and 85.48%, respectively.

As with the other architectures, the worn out road class remained the most difficult to identify accurately, yielding a lower F1-score of 70.16%. Meanwhile, the regular road category showed intermediate classification performance, with an average F1-score of 77.17%.

TABLE 5. Classification performance of SepConv1D architecture for different input data representations in terms of 5-fold average Precision, Recall, F1-score, and Accuracy (%) for each class and macro-averaged across classes.

SepConv1D							
Input Data Type	**Metric**	**Asphalt Bumps**	**Metal Bumps**	**Potholes**	**Regular Road**	**Worn Out Road**	**Class Avg**
DWT-CA	Precision	93.12	97.19	92.25	82.02	78.99	**88.71**
	Recall	94.21	95.38	92.86	82.6	77.97	**88.61**
	F1-score	93.61	96.21	92.52	82.26	78.33	**88.59**
	Accuracy						**88.56**
Time	Precision	94.27	96.96	92.23	76.23	76.01	87.14
	Recall	90.63	95.32	92.70	81.34	73.33	86.66
	F1-score	92.34	96.03	92.44	78.66	74.43	86.78
	Accuracy						86.76
FFT	Precision	82.16	81.13	77.79	72.99	62.51	75.32
	Recall	79.98	75.01	85.41	80.48	52.40	74.66
	F1-score	81.01	77.37	81.27	76.49	56.69	74.56
	Accuracy						75.05
DWT-CD	Precision	58.63	79.14	79.57	63.04	68.21	69.72
	Recall	60.80	77.21	81.55	75.44	49.91	68.98
	F1-score	59.36	78.04	80.45	68.29	57.45	68.72
	Accuracy						69.79

TABLE 6. Classification performance of LSTM architecture for different input data representations in terms of 5-fold average Precision, Recall, F1-score, and Accuracy (%) for each class and macro-averaged across classes.

LSTM							
Input Data Type	**Metric**	**Asphalt Bumps**	**Metal Bumps**	**Potholes**	**Regular Road**	**Worn Out Road**	**Class Avg**
DWT-CA	Precision	93.61	95.05	92.20	79.51	77.05	87.48
	Recall	92.97	96.62	93.43	78.02	75.64	87.33
	F1-score	93.24	95.81	92.77	78.60	76.09	87.30
	Accuracy						87.36
Time	Precision	91.92	90.83	92.47	86.18	83.15	**88.14**
	Recall	97.85	95.19	93.75	89.83	69.16	**88.10**
	F1-score	94.79	92.96	93.10	87.97	75.51	**88.02**
	Accuracy						**88.06**
FFT	Precision	82.48	68.80	76.56	73.37	57.12	71.67
	Recall	74.60	64.69	82.44	75.10	56.31	70.63
	F1-score	78.23	65.77	79.37	74.17	56.30	70.77
	Accuracy						71.15
DWT-CD	Precision	61.15	80.43	78.10	65.73	61.77	69.44
	Recall	57.65	76.10	85.54	66.79	59.39	69.09
	F1-score	59.31	78.09	81.63	66.18	60.49	69.14
	Accuracy						70.26

5) SUMMARY OF BEST REPRESENTATIONS

Tab. 9 summarizes the performance of the baseline DL models in terms of average accuracy and F1-score with their respective best data representation domain. Some comments are reported below.

- For SepConv1D architecture DWT-CA yielded the best classification performances. This representation indeed preserves low-frequency information and coarse structures in the signal, which synergizes well with the convolutional filters applied at different resolutions.

TABLE 7. Classification performance of GRU architecture for different input data representations in terms of 5-fold average Precision, Recall, F1-score, and Accuracy (%) for each class and macro-averaged across classes.

GRU							
Input Data Type	**Metric**	**Asphalt Bumps**	**Metal Bumps**	**Potholes**	**Regular Road**	**Worn Out Road**	**Class Avg**
DWT-CA	Precision	92.67	96.7	91.62	78.69	79.28	**87.79**
	Recall	93.41	97.41	92.85	79.15	75.97	**87.76**
	F1-score	93.01	97.02	92.20	78.90	77.47	**87.72**
	Accuracy						**87.71**
Time	Precision	93.67	97.28	90.97	79.26	75.20	87.27
	Recall	93.87	96.07	92.27	78.07	75.53	87.16
	F1-score	93.70	96.58	91.61	78.61	75.29	87.16
	Accuracy						87.11
FFT	Precision	81.44	76.49	76.93	74.69	60.25	73.96
	Recall	79.96	75.30	81.69	76.82	54.00	73.55
	F1-score	80.43	75.68	79.15	75.63	56.34	73.45
	Accuracy						73.78
DWT-CD	Precision	61.10	80.29	80.43	63.15	60.37	69.07
	Recall	60.69	80.22	82.82	67.91	52.41	68.81
	F1-score	60.74	80.22	81.56	65.19	56.01	68.74
	Accuracy						69.68

TABLE 8. Classification performance of CNN-2D architecture trained on CWT representation, reported as 5-fold average Precision, Recall, F1-score, and Accuracy (%) for each class and averaged across classes.

CNN-2D							
Input Data Type	**Metric**	**Asphalt Bumps**	**Metal Bumps**	**Potholes**	**Regular Road**	**Worn Out Road**	**Class Avg**
CWT	Precision	86.73	94.97	84.80	73.78	74.87	**83.03**
	Recall	84.36	89.75	89.85	81.26	66.30	**82.30**
	F1-score	85.48	92.25	87.19	77.17	70.16	**82.45**
	Accuracy						**82.53**

- Concerning LSTM model, time-domain signals demonstrated a slight superiority compared to DWT-CA. This outcome could be attributed to the fact that LSTM is specifically designed to capture temporal dependencies and sequential patterns in the input data. Time-domain signals preserve the full temporal continuity and order of events, which is crucial for LSTM's memory-based learning dynamics.
- In contrast, for GRU model Time and DWT-CA provided comparable classification performance. This marginal difference suggests that the GRU model is robust to the type of temporal representation, likely due to its reduced gating complexity compared to LSTM, which enables effective modeling of sequential dependencies even in lower-resolution or transformed representations.
- The CWT representation combined with CNN-2D architecture provided inferior classification performance compared to SepConv1D and LSTM with the best input data types such as DWT-CA and time-domain. Such behavior could be explained by the fact that both LSTM and SepConv1D are much more effective in extracting meaningful features from their respective input data types than CNN is able to with CWT.

B. EVALUATION OF THE PROPOSED MODEL

Tab. 10 reports the classification results of the proposed hybrid model trained with the adopted transfer learning strategy. The input configuration was selected based on the best-performing data representation for each backbone: DWT-CA for the SepConv1D and time-domain signals for the LSTM branch. This configuration was chosen to leverage the complementary strengths of each architecture and ensure optimal compatibility with their respective input domains.

The proposed hybrid model demonstrated strong classification capabilities, attaining an average accuracy of 89.19% across the five folds. The best class-specific performance was observed for the metal bumps, asphalt bumps, and potholes

TABLE 9. Summary of classification performance for baseline DL models using their best-performing data representations, reported as 5-fold average Accuracy and F1-score (%).

Model	Best data domain	Accuracy	F1-score
SepConv1D	DWT-CA	**88.56**	**88.59**
	Time	86.76	86.58
LSTM	Time	**88.06**	**88.02**
	DWT-CA	87.36	87.30
GRU	DWT-CA	**87.71**	**88.72**
	Time	87.11	87.16
CNN-2D	CWT	82.53	82.45

categories, achieving average F1-scores of 97.17%, 93.55%, and 93.53%, respectively. These results indicate the model's ability to accurately recognize pronounced anomaly patterns in both frequency and time domains.

As consistently observed with all tested architectures, the worn out road category posed the greatest classification challenge with an average F1-score of 79.81%. The regular road class exhibited intermediate classification performance, with an average F1-score of 81.87%.

To assess the statistical robustness of the obtained results, in this study we computed both the standard deviation (SD) of accuracy values across all five folds and the one-tailed p-value derived from a paired t-test, comparing each of the baseline models against the proposed hybrid SepConv1-LSTM. Tab. 11 summarizes the comparison of the results obtained by the baseline DL models: SepConv1D, LSTM,GRU, and CNN-2D alongside the hybrid SepConv1D-LSTM trained from scratch. The results are reported in terms of 5-fold average metrics, SD and p-values with respect to the proposed model.

The proposed model consistently outperformed all baseline models, including the same hybrid architecture trained from scratch. These results underscore the effectiveness of the transfer learning strategy, which facilitates faster convergence and leads to a more optimal solution during the fine-tuning by initializing the model with pretrained weights on the most suitable data domain for each branch.

Among all baseline models, SepConv1D trained on the DWT-CA representation exhibited the closest performance to the proposed hybrid, while LSTM with time-domain inputs also performed competitively. However, both were statistically outperformed by the proposed hybrid model ($p < 0.05$). The CNN-2D model, trained on CWT inputs, showed the lowest performance across all metrics.

Interestingly, the hybrid SepConv1D-LSTM trained from scratch did not outperform the standalone SepConv1D and yielded only marginal improvements over LSTM. This outcome may be attributed to the challenging experimental conditions adopted in this study, including a rigorous cross-validation strategy that ensures disjoint signal events between training and validation sets. Additionally the dataset diversity, which spans multiple vehicles, varying speeds and heterogeneous sensor placements, further increased the complexity of the classification task, making it more difficult for the model trained from scratch to generalize effectively.

Tab. 12 further details the comparison between the proposed hybrid model and the baselines in terms of 5-fold average F1-scores for each class. The proposed architecture outperformed all other models on the most challenging and critical category, worn out road, achieving an average F1-score of 79.81%. Additionally, the hybrid model yielded the highest class-specific performance for metal bumps and potholes, with F1-scores of 97.17% and 93.53%, respectively. Although the GRU model achieved the highest F1-score for the asphalt bumps category, it underperformed compared to the proposed hybrid model on the remaining classes. The SepConv1D model slightly outperformed the hybrid model on the regular road and asphalt bumps classes; however, it achieved lower F1-scores for all other classes and exhibited an overall inferior classification performance. The proposed SepConv1D-LSTM enhanced with transfer learning, achieves the highest overall F1-score over all classes, as shown in Tab. 11. The number of samples used in training and test are similar for each class, as shown in Tab. 2.

A notable outcome of this study concerns the worn out road category, which consistently represents the most challenging type of anomaly across all architectures due to its subtle and heterogeneous signal patterns. Despite this intrinsic difficulty, the proposed hybrid SepConv1D–LSTM trained with the intra-domain transfer learning strategy achieved the highest F1-score for this class (79.81%), outperforming every baseline and state-of-the-art model. This improvement is particularly significant because worn out road segments are less distinctive and often overlap in appearance with regular road conditions, making them prone to misclassification. The superior performance of the proposed approach suggests that the complementary feature representations learned by the pretrained SepConv1D and LSTM backbones enable a more effective capture of the low-amplitude, long-term irregularities characteristic of worn surface degradation. Highlighting this result underscores that the benefit of the proposed method extends beyond the modest overall accuracy gain, providing a substantial advantage for the most ambiguous and practically important anomaly type.

C. COMPARISON WITH STATE-OF-THE ART ARCHITECTURES

To further evaluate the effectiveness of the proposed hybrid SepConv1D-LSTM architecture based on transfer learning, we conducted a comprehensive experimental comparison against several state-of-the-art DL models previously applied to road condition monitoring using vibrational signals.

- **Hybrid CNN-LSTM** (Raslan et al. [20])
  This model was proposed in the aforementioned study [20] to classify road abnormalities from vibrational signals collected with an MPU-9250 sensor. This

TABLE 10. Classification results of the proposed hybrid SepConv1D-LSTM model based on transfer learning in terms of 5-fold average Precision, Recall, F1-score and Accuracy (%) for each class and averaged across classes.

Proposed hybrid SepConv1D-LSTM							
Input Data Type	**Metric**	**Asphalt Bumps**	**Metal Bumps**	**Potholes**	**Regular Road**	**Worn Out Road**	**Class Avg**
DWT-CA + Time	Precision	93.10	97.37	93.1	82.06	80.67	89.26
	Recall	94.04	97.00	94.00	81.77	79.18	89.20
	F1-score	93.55	97.17	93.53	81.87	79.81	89.19
	Accuracy						89.19

TABLE 11. Comparison of classification performance of the proposed hybrid SepConv1D-LSTM model with SepConv1D, LSTM, GRU and CNN-2D in terms of 5-fold average metrics (%), SD and *p*-values computed on accuracies across folds.

Model	**Input data**	**Prec.**	**Rec.**	**F1**	**Acc.**	**SD**	**P-val***
SepConv1D-LSTM (transfer)	DWT-CA + Time	**89.26**	**89.20**	**89.19**	**89.19**	0.0158	-
SepConv1D-LSTM (scratch)	DWT-CA + Time	88.29	88.25	88.20	88.28	0.0155	0.0094
SepConv1D	DWT-CA	88.71	88.61	88.59	88.56	0.0130	0.0365
LSTM	Time	88.14	88.10	88.02	88.06	0.0156	0.0251
GRU	Time	87.27	87.16	87.16	87.11	0.0181	0.0056
CNN-2D	CWT	83.03	82.30	82.45	82.53	0.0056	0.0001

TABLE 12. Comparison of 5-fold average F1-score for each class, obtained by the proposed hybrid SepConv1D-LSTM architecture to those obtained by other baseline models.

Model	**Input data**	**Asphalt Bumps**	**Metal Bumps**	**Potholes**	**Regular Road**	**Worn Out Road**
SepConv1D-LSTM (transfer)	DWT-CA+ Time	93.55	**97.17**	**93.53**	81.87	**79.81**
SepConv1D-LSTM (scratch)	DWT-CA+ Time	92.35	97.07	92.80	81.50	77.26
SepConv1D	DWT-CA	93.61	96.21	92.52	**82.26**	78.33
LSTM	Time	93.11	96.67	92.23	79.99	78.09
GRU	Time	**93.7**	96.58	91.61	78.61	75.29
CNN-2D	CWT	85.48	92.25	87.19	77.17	70.16

TABLE 13. Comparison of classification performance of the proposed hybrid SepConv1D-LSTM model with the other DL state-of-the-art architectures in terms of 5-fold average metrics (%), SD and *p*-values computed on accuracies across folds.

Model	**Input data**	**Prec.**	**Rec.**	**F1**	**Acc.**	**SD**	**P-val.***
SepConv1D-LSTM (transfer)	DWT-CA + Time	**89.26**	**89.20**	**89.19**	**89.19**	0.0158	-
SepConv1D-LSTM (scratch)	DWT-CA + Time	88.29	88.25	88.20	88.28	0.0155	0.0094
CNN-LSTM [20]	DWT-CA + Time	81.82	81.71	81.71	81.72	0.0195	0.0001
Transformer [32]	DWT-CA	85.68	85.62	85.46	85.53	0.0167	0.001
CNN-1D [3]	DWT-CA	84.36	83.49	83.67	83.66	0.0141	0.0007
CNN-1D [11]	DWT-CA	84.67	84.70	84.64	84.68	0.0087	0.0003
CNN-2D [7]	CWT	82.73	82.15	82.18	82.18	0.0057	0.0003

architecture consists of two parallel branches. The first branch is a CNN-1D subnetwork comprising two 1D convolutional layers with ReLU activations and same-padding: the first with 128 filters and a kernel size of 10, and the second with 64 filters and kernel size of 3. These are followed by a Flatten layer and a Dropout layer with a rate of 0.2. The second branch is an LSTM layer with 128 units followed by a Dropout of 0.2. The outputs of both branches are concatenated and passed through a Dense layer with 128 units and ReLU activation, followed by the final classification Dense layer.

- **Transformer** (Shavit and Klein [32])
  Originally introduced for natural language processing tasks, Transformer architectures have been

TABLE 14. Comparison of 5-fold average F1-score for each class, obtained by the proposed hybrid SepConv1D-LSTM architecture to those obtained by other state-of- the art DL architectures.

Model	Input data	Asphalt Bumps	Metal Bumps	Potholes	Regular Road	Worn Out Road
SepConv1D-LSTM (transfer)	DWT-CA + Time	**93.55**	**97.17**	**93.53**	**81.87**	**79.81**
SepConv1D-LSTM (scratch)	DWT-CA + Time	92.35	97.07	92.80	81.50	77.26
CNN-LSTM [20]	DWT-CA + Time	89.56	87.00	88.66	74.50	68.82
Transformer [32]	DWT-CA	92.65	95.54	91.31	77.53	70.25
CNN-1D [3]	DWT-CA	89.12	93.43	80.79	75.30	71.45
CNN-1D [11]	DWT-CA	87.11	88.71	86.61	74.05	66.98
CNN-2D [7]	CWT	85.84	89.90	86.30	77.75	71.09

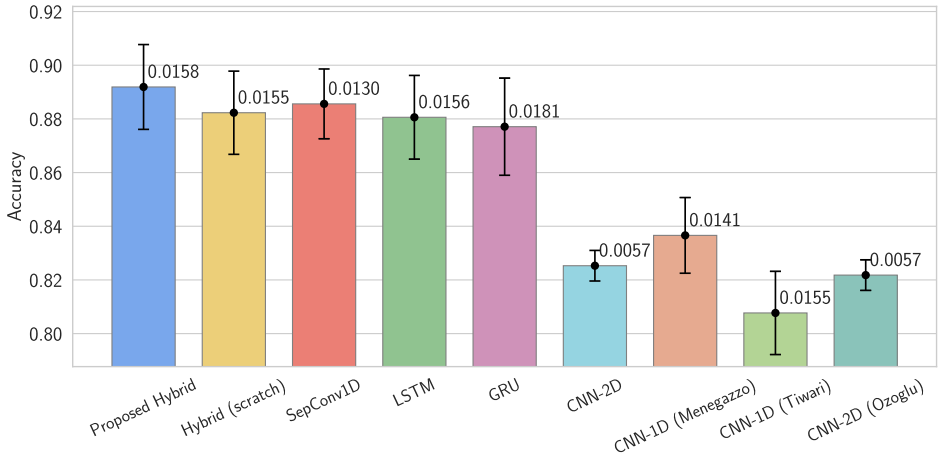


FIGURE 18. Performance comparison in terms of 5-fold average accuracy of the proposed model with the DL baseline models and the other state-of-the art architectures.

increasingly adopted for time series classification due to their ability to model long-range dependencies using self-attention mechanisms, without recurrence. The adapted architecture used in this study includes a CNN-1D feature extractor consisting of four Conv1D layers with GeLU activations, followed by a stack of six Transformer encoder blocks. The final encoder output is passed through a Dense layer with LogSoftmax activation to generate class probabilities.

- **CNN-1D** (Menegazzo and von Wangenheim [3])
  This model was identified as the best-performing architecture in [3] for vibrational signal-based road type classification in a broad comparison involving CNN-1D, LSTM, and CNN-LSTM. The architecture consists of two initial blocks, each formed by a sequence of Conv1D layer, ReLU activation, a BatchNormalization and a Dropout of 0.15. The Conv1D layers have a kernel size of 3 and a filters number of 64 and 32 respectively. The two blocks are followed by a third block made up by a Conv1D with 32 filters, a ReLU activation, a GlobalAveragePooling1D, a BatchNormalization and a Dropout of 0.2. Finally, the output is provided to a couple of consecutive Dense layers, the first with a ReLU activation and the second with Softmax activation producing the class probabilities.
- **CNN-1D** (Tiwari et al. [11])
  This CNN-1D was developed for a crowdsourced, real-time road surface monitoring system from 3-axis accelerometer signals and speed. The model includes two initial Conv1D layers with 24 and 64 filters respectively, each followed by MaxPooling1D. These are followed by three additional Conv1D layers with 64 filters each. The resulting feature maps are then passed through three Dense layers with 1024 units and finally to the output layer for classification.
- **CNN-2D** (Ozoglu and Gökgöz [7])
  This architecture was designed to process multi-channel vibrational data (accelerometer and gyroscope) as 2D images for pothole detection. The image height corresponds to sensor dimensions and the width to time steps. The tested configuration consists of four Conv2D layers with 5×5 kernels and l_2 regularization (0.0001), each followed by MaxPooling and Dropout of rate 0.3. The first Conv2D layer has 32 filters; subsequent layers use 64 filters. The output tensor is flattened and processed by a Dense layer with 64 units and ReLU activation, followed by a Dropout of 0.5.

All the considered DL models in this study including the baseline models, the novel proposed SepConv1D-LSTM based on transfer learning and the other state-of-art architectures were trained and evaluated under identical preprocessing procedures, data generation steps and the same event-level k-fold evaluation protocol. Moreover, all models were configured with the same training hyperparameters. This unified experimental setup ensures methodological consistency and enables a fair and reliable comparison of classification performance across architectures.

For the CNN-1D models by Menegazzo and Gökgöz [3] and Tiwari et al. [11], input data in the DWT-CA domain were used. This choice was guided by our earlier findings, which showed that DWT-CA consistently yields superior performance in convolutional architectures. Similarly, DWT-CA was also used for the Transformer-based model by Shavit and Klein [32], given its initial convolutional backbone. Likewise, the hybrid model proposed by Raslan et al. [20] was configured with DWT-CA as input for CNN-1D backbone and with time-signals for the LSTM branch. In contrast, the CNN-2D model proposed by Ozoglu and Gökgöz [7] was trained using input data represented in the CWT domain due to the structural compatibility between two-dimensional convolutional filters and the spectrogram-like time-frequency matrices produced by CWT.

Tab. 13 presents the comparative evaluation of the proposed hybrid SepConv1D–LSTM model, trained via a transfer learning strategy, against the same hybrid model trained from scratch and several state-of-the-art architectures, in terms of 5-fold average classification metrics.

The proposed hybrid SepConv1D-LSTM model significantly outperformed all the other architectures with strong statistical confidence as evidenced by the p-values well below the conventional threshold of 0.05 for all comparisons.

Among the existing approaches, the Transformer-based model achieved the highest classification performance, with an average accuracy of 85.53% and an average F1-score of 85.46%. Nevertheless, the proposed transfer learning-based hybrid model yielded a substantial performance improvement over this architecture. Interestingly, the hybrid CNN–LSTM model introduced in [20] exhibited the lowest overall performance among the compared state-of-the-art models. These findings are further supported by the per-class F1-score comparison across five folds, as reported in Table 14.

Fig. 18 visually summarizes the classification accuracy and standard deviation of all evaluated models. The proposed hybrid SepConv1D-LSTM architecture, trained with transfer learning, clearly outperformed all other models, achieving the highest accuracy while maintaining low variability across folds. Notably, even compared to its non-pretrained counterpart, the transfer learning approach yielded a measurable improvement in both performance and consistency. The state-of-the-art models used for road condition monitoring including, CNN-1D, CNN-2D, Transformer and the hybrid alternatives, significantly underperformed against the proposed model, highlighting the effectiveness of combining domain-specific representations with a carefully designed hybrid architecture and transfer learning strategy.

VI. CONCLUSION

This work introduced a hybrid SepConv1D-LSTM architecture for road anomaly classification based on vibrational signals collected via smartphone sensors. By combining SepConv1D layers for extracting spatial feature with LSTM units for modeling temporal dependencies, the proposed model effectively captures the complex patterns associated with road surface irregularities.

A key contribution of this study is the adoption of an intra-domain transfer learning strategy, wherein each backbone was pretrained on its optimal input representation, DWT-CA for SepConv1D and time-domain signals for LSTM, and subsequently finetuned in a unified framework. This approach enhanced convergence stability and generalization, particularly under conditions of limited data and heterogeneous sensor setups.

To ensure robust and realistic evaluation, an event-level stratified 5-fold cross-validation protocol was employed, mitigating data leakage and simulating deployment scenarios more accurately.

The proposed model was systematically compared against a range of baseline architectures, including its standalone SepConv1D and LSTM components, as well as GRU and CNN-2D models, never used for this type of application. Additionally, several state-of-the-art DL architectures, already used for road condition monitoring, were evaluated under consistent experimental conditions. A total of 22 architectures were compared using a dataset of about 90,000 samples properly subdivided in training and test. Different transformations of the time-domain data were considered: time-domain, Fast Fourier Transform, Discrete Wavelet Transform with Approximation Coefficients, Discrete Wavelet Transform with Detail Coefficients, and Continuous Wavelet Transform. The hybrid model consistently achieved superior performance for all classification metrics compared to the baseline models and the other state-of-the-art architectures, with statistical significance across all folds.

Future research directions may include more the exploration of more advanced fine-tuning techniques to further enhance generalization capability, such as a layer-wise training scheme or an approach based on an independent finetuning of the pretrained SepConv1D and LSTM branches. Additionally, future work could explore the performance of more sophisticated feature fusion mechanisms, incorporating attention-based modules and regularization techniques to enhance the combined feature representation. The methodology of Deep Learning applied to several data representation domains such as time-domain, FFT, DWT, and CWT will be extended by the authors to different applications.

REFERENCES

[1] M. R. Carlos, M. E. Aragon, L. C. Gonzalez, H. J. Escalante, and F. Mariínez, "Evaluation of detection approaches for road anomalies based on accelerometer readings—Addressing Who's who," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 10, pp. 3334–3343, Oct. 2018.

[2] A. Basavaraju, J. Du, F. Zhou, and J. Ji, "A machine learning approach to road surface anomaly assessment using smartphone sensors," *IEEE Sensors J.*, vol. 20, no. 5, pp. 2635–2647, Mar. 2020.

[3] J. Menegazzo and A. von Wangenheim, "Road surface type classification based on inertial sensors and machine learning," *Computing*, vol. 103, no. 10, pp. 2143–2170, Oct. 2021.

[4] J. Menegazzo and A. von Wangenheim, "Speed bump detection through inertial sensors and deep learning in a multi-contextual analysis," *Social Netw. Comput. Sci.*, vol. 4, no. 1, p. 18, Oct. 2022.

[5] P. Singh, A. E. Kamal, A. Bansal, and S. Kumar, "Road pothole detection from smartphone sensor data using improved LSTM," *Multimedia Tools Appl.*, vol. 83, no. 9, pp. 26009–26030, Aug. 2023.

[6] N. Sabor and M. AbdelRaheem, "CMNN-RADC: A crowdsensing convolutional-based mixer neural network road anomalies detector and classifier," *Internet Things*, vol. 22, Jul. 2023, Art. no. 100771.

[7] F. Ozoglu and T. Gökgöz, "Detection of road potholes by applying convolutional neural network method based on road vibration data," *Sensors*, vol. 23, no. 22, p. 9023, Nov. 2023.

[8] D. Wang, L. Zhuang, L. Gao, X. Sun, and X. Zhao, "Global feature-injected blind-spot network for hyperspectral anomaly detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 21, pp. 1–5, 2024.

[9] D. Wang, L. Gao, Y. Qu, X. Sun, and W. Liao, "Frequency-to-spectrum mapping GAN for semisupervised hyperspectral anomaly detection," *CAAI Trans. Intell. Technol.*, vol. 8, no. 4, pp. 1258–1273, Dec. 2023.

[10] R. Singh, A. Sethi, K. Saini, S. Saurav, A. Tiwari, and S. Singh, "Attention-guided generator with dual discriminator GAN for real-time video anomaly detection," *Eng. Appl. Artif. Intell.*, vol. 131, May 2024, Art. no. 107830. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0952197623020146

[11] S. Tiwari, R. Bhandari, and B. Raman, "RoadCare: A deep-learning based approach to quantifying road surface quality," in *Proc. 3rd ACM SIGCAS Conf. Comput. Sustain. Societies*, New York, NY, USA, Jun. 2020, pp. 231–242.

[12] A. K. Pandey, R. Iqbal, T. Maniak, C. Karyotis, S. Akuma, and V. Palade, "Convolution neural networks for pothole detection of critical road infrastructure," *Comput. Electr. Eng.*, vol. 99, Apr. 2022, Art. no. 107725.

[13] M. Hijji, R. Iqbal, A. Kumar Pandey, F. Doctor, C. Karyotis, W. Rajeh, A. Alshehri, and F. Aradah, "6G connected vehicle framework to support intelligent road maintenance using deep learning data fusion," *IEEE Trans. Intell. Transp. Syst.*, vol. 24, no. 7, pp. 7726–7735, Jul. 2023.

[14] H. Xin, Y. Ye, X. Na, H. Hu, G. Wang, C. Wu, and S. Hu, "Sustainable road pothole detection: A crowdsourcing based multi-sensors fusion approach," *Sustainability*, vol. 15, no. 8, p. 6610, Apr. 2023.

[15] K. Bansal, K. Mittal, G. Ahuja, A. Singh, and S. S. Gill, "DeepBus: Machine learning based real time pothole detection system for smart transportation using IoT," *Internet Technol. Lett.*, vol. 3, no. 3, p. 156, May 2020.

[16] G. Baldini, R. Giuliani, and F. Geib, "On the application of time frequency convolutional neural networks to road Anomalies' identification with accelerometers and gyroscopes," *Sensors*, vol. 20, no. 22, p. 6425, Nov. 2020.

[17] C. Chen, H. Seo, and Y. Zhao, "A novel pavement transverse cracks detection model using WT-CNN and STFT-CNN for smartphone data analysis," *Int. J. Pavement Eng.*, vol. 23, no. 12, pp. 4372–4384, Oct. 2022.

[18] A. Martinelli, M. Meocci, M. Dolfi, V. Branzi, S. Morosi, F. Argenti, L. Berzi, and T. Consumi, "Road surface anomaly assessment using low-cost accelerometers: A machine learning approach," *Sensors*, vol. 22, no. 10, p. 3788, May 2022.

[19] E. Raslan, M. F. Alrahmawy, Y. A. Mohammed, and A. S. Tolba, "IoT for measuring road network quality index," *Neural Comput. Appl.*, vol. 35, no. 3, pp. 2927–2944, Jan. 2023.

[20] E. Raslan, M. F. Alrahmawy, Y. A. Mohammed, and A. S. Tolba, "Evaluation of data representation techniques for vibration based road surface condition classification," *Sci. Rep.*, vol. 14, no. 1, pp. 1–14, May 2024.

[21] Z. Khademi, F. Ebrahimi, and H. M. Kordy, "A transfer learning-based CNN and LSTM hybrid deep learning model to classify motor imagery EEG signals," *Comput. Biol. Med.*, vol. 143, Apr. 2022, Art. no. 105288.

[22] E. A. Martinez-Ríos, R. Bustamante-Bello, and S. A. Navarro-Tuch, "Generalized Morse wavelets parameter selection and transfer learning for pavement transverse cracking detection," *Eng. Appl. Artif. Intell.*, vol. 123, Aug. 2023, Art. no. 106355.

[23] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 1–9.

[24] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "Squeezenet: Alexnet-level accuracy with 50x fewer parameters and <0.5MB model size," 2016, *arXiv:1602.07360*.

[25] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[26] E. Otovic, M. Njirjak, D. Jozinovic, G. Mausa, A. Michelini, and I. SStajduhar, "Intra-domain and cross-domain transfer learning for time series data—How transferable are the features?" *Knowl.-Based Syst.*, vol. 239, Mar. 2022, Art. no. 107976.

[27] L. C. González, R. Moreno, H. J. Escalante, F. Martínez, and M. R. Carlos, "Learning roadway surface disruption patterns using the bag of words representation," *IEEE Trans. Intell. Transp. Syst.*, vol. 18, no. 11, pp. 2916–2928, Nov. 2017.

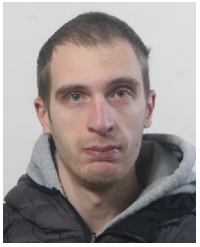
[28] I. Ferjani and S. Ali Alsaif, "How to get best predictions for road monitoring using machine learning techniques," *PeerJ Comput. Sci.*, vol. 8, p. e941, Apr. 2022.

[29] J. Sanchez, N. Soltani, R. Chamarthi, A. Sawant, and H. Tabkhi, "A novel 1D-convolution accelerator for low-power real-time CNN processing on the edge," in *Proc. IEEE High Perform. extreme Comput. Conf. (HPEC)*, Sep. 2018, pp. 1–8.

[30] X. Yang, L. Shu, K. Li, Z. Huo, and Y. Zhang, "SA1D-CNN: A separable and attention based lightweight sensor fault diagnosis method for solar insecticidal lamp Internet of Things," *IEEE Open J. Ind. Electron. Soc.*, vol. 3, pp. 291–303, 2022.

[31] T. Yang, J. Chen, H. Deng, and B. He, "A lightweight intrusion detection algorithm for IoT based on data purification and a separable convolution improved CNN," *Knowl.-Based Syst.*, vol. 304, Nov. 2024, Art. no. 112473.

[32] Y. Shavit and I. Klein, "Boosting inertial-based human activity recognition with transformers," *IEEE Access*, vol. 9, pp. 53540–53547, 2021.



LORENZO MANONI (Member, IEEE) received the B.Sc. and M.Sc. degrees in electronics engineering and the Ph.D. degree in information engineering from the Università Politecnica delle Marche, Ancona, Italy, in 2015, 2018, and, 2022, respectively. He was a Research Fellow with the Department of Information Engineering (DII), in 2022. He is currently a Researcher with DII. His current research interests include embedded systems, machine learning, deep learning, artificial intelligence applications, convolutional neural networks, signal processing, algorithms analysis and design, and bio-signal analysis.

SIMONE ORCIONI (Senior Member, IEEE) received the Laurea and Ph.D. degrees in electronics engineering from the Università Politecnica delle Marche, Ancona, Italy, in 1992 and 1995, respectively. He was appointed as an Assistant Professor, in 2000, teaching courses in analog and digital electronics and authored a text book. In 2017, he was an Adjunct Professor with the Ubiquitous Computing Laboratory (UC-Lab), HTWG Konstanz, Germany, where he is currently a Guest Researcher. Since 2021, he has been held the position of an Associate Professor with the Department of Information Engineering, Università Politecnica delle Marche, where he is also the President of the Board of the Degree Programs in Electronic Engineering. He has authored more than 50 journal articles and over 100 contributions to international conference proceedings and edited book chapters. His research has addressed statistical device modeling and simulation, analog circuit design, cyber-physical system simulation, and both linear and nonlinear system identification. His current research interests include nonlinear digital signal processing, li-ion battery impedance spectroscopy, and electronics for renewable energy applications. He has acted as a reviewer for 20 international journals, a program committee member for seven international conferences, the program chair for three international conferences, an editor for four international volumes, and an inventor for two patents. He has been included in Stanford University's "World's Top 2% Scientists" in 2023 and 2024 edition of "Annual Influence Ranking" and the "Lifetime Scientific Influence Ranking." He has served as a Guest Editor for *EURASIP Journal on Embedded Systems, Frontiers in Energy Research, and Sensors* (MDPI).

MASSIMO CONTI (Member, IEEE) graduated in electronics engineering from the University of Ancona, Italy, in 1987. He is an Associate Professor with the Dipartimento di Ingegneria dell'Informazione (DII), Università Politecnica delle Marche (UNIVPM), Ancona, Italy, from 1998. In 1990, he is a Researcher in electronics at UNIVPM. His research activity in the field of microelectronics is mainly devoted to system level design of low power integrated circuits, electronic smart systems for ambient assisted living, design of energy harvesting systems, battery management system, V2G vehicle to grid connection, smart grid, state of health estimation, analysis of cell mismatch on battery pack performances, design of BMS and battery life tracing for reuse, recycle and end-of-life, smart mobility, NFC for food traceability, coauthor of more than 255 papers on Int. Books, Journals or Conferences. Scopus: 195 pubs, 1588 citations, h-index: 21. He is a Coordinator of European and National research projects. He is an Editor of 11 International Books, a Lead Guest Editor of special issue of *International Journals*, and a General Chair of ten International Conferences. In 2024 and 2025 in the top 2% most cited scientists, ranking elaborated by Stanford University/Elsevier. In 2025-26 Elected Chair of the of Consumer Power and Energy (CPE) Technical Committee of IEEE Consumer Technology Society (CTSoC). For more information visit the link www.univpm.it/massimo.conti.