DOI: 10.1111/mice.XXXXX

COMPUTER-AIDED CIVIL AND INFRASTRUCTURE ENGINEERING WILEY

RESEARCH ARTICLE

# An interpretable operational state classification framework for elevators through Convolutional Neural Networks

**J. Olaizola**[1] | **U. Izagirre**[1] | **O. Serradilla**[2] | **E. Zugasti**[1] | **M. Mendicute**[2] | **Jose I. Aizpurua**[3,4]

[1]Electronics and Computing Department, Mondragon University, Gipuzkoa, Spain

[2]Laboral Kutxa, Gipuzkoa, Spain

[3]Computer Science and Artificial Intelligence Department, University of the Basque Country (UPV/EHU), Donostia, Gipuzkoa, Spain

[4]Ikerbasque, Basque Foundation for Science, Gipuzkoa, Spain

**Correspondence**
Jose I. Aizpurua, Manuel de Lardizabal P. 1, Donostia-San Sebastian, 20018 Gipuzkoa (Spain).
Email: joxe.aizpurua@ehu.eus

**Funding information**
Education Department of the Basque Government through the Research Groups Support call (IT1451-22). Jose I. Aizpurua acknowledges financial support from the Spanish State Research Agency (grant No. RYC2022-037300-I).

## ABSTRACT

Ensuring the safe, reliable, and cost-efficient operation of transportation systems such as elevators is critical for the maintenance of civil infrastructures. The ability to monitor the health-state and classify different operational states (elevator moving up/down, stopped, doors opening/closing) may lead to the development of intelligent solutions, such as diagnostics and predictive maintenance. Accordingly, downtime and maintenance costs can be significantly reduced with an accurate monitoring of the operation parameters and dynamics. In this context, this paper presents a novel approach for the operational state classification of elevator systems based on a one-dimensional (1D) convolutional neural network (CNN), using exclusively a single axis (Z) of an accelerometer signal. The proposed model utilizes a single accelerometer and addresses the challenge of distinguishing overlapping signal patterns, such as those produced by vertical displacement and door movements. The approach includes an interpretability stage, which demonstrates the data processing involved in extracting features from the underlying physical phenomena captured in the acceleration signal. Obtained results have been validated with an on-site captured dataset which contains 250 elevator journeys and compared with three other classification methods that have been conventionally used: Generalized Likelihood Ratio Test (GLRT), barometer-assisted GLRT and three conventional ML models. It has been shown that the proposed approach is very accurate, with 96% of average F1 score, and importantly, includes the analytic relation of the classification model features.

## 1 | INTRODUCTION

Mechatronic transportation systems for humans and goods have been of great importance. Over the past decades, there has been a trend motivating the shift from rural to urban areas (Khanna, 2020). To maximize the use of available land, the vertical growth of cities has been accelerated, which requires optimized real estate property maintenance strategies, as claimed by (Taillandier et al., 2017). Simultaneously, the increasing height of buildings has led to a rise in the deployment of elevators in multi-story buildings (Khanna, 2020), whose maintenance must also be considered. In this context, significant efforts have been directed toward developing advanced technologies that integrate intelligence into the monitoring and maintenance of civil infrastructures and transportation systems (Adeli & Jiang, 2008). In the specific case of elevators new services have also been introduced to improve usability, accessibility, and condition monitoring (Awatramani et al., 2022), apart from enhancements in traditional control systems (Fernandez & Cortes, 2015).

Elevator maintenance costs are very high, and they can represent up to 48% of the total net sales (Kone, 2024; Otis, 2024). In this direction, condition monitoring is a key

activity for elevator maintenance planning to foresee operational deviations that may lead to failures and cause downtime and financial penalties. Condition monitoring can be used to decrease maintenance costs by anticipating downtime and develop energy- and money-saving strategies through elevator usage patterns (Adak et al., 2013; A. of German Engineers, 2007). Monitoring the operational state of the elevator can reveal useful information about the system condition, which may also show patterns that can evolve into more severe failures. (X.-Y. Jiang et al., 2022) describe seven different states during the nominal operation of an elevator, which can be reduced into three, if approximation operational states are neglected as displayed in **TABLE 1**.

**TABLE 1** Possible elevator status.

| ID | State | Description |
|---|---|---|
| $\mathcal{H}0$ | Stop | elevator stopped |
| $\mathcal{H}1$ | Elevator moving | elevator moving in its vertical displacement (up/down) |
| $\mathcal{H}2$ | Door moving | elevator door(s) moving (opening/closing) |

Understanding and accurately detecting these operational states is essential for improving elevator monitoring and predictive maintenance. Various approaches have been proposed in the literature, ranging from classical signal processing techniques to AI-driven methodologies. The following section reviews existing methods for operational state detection and identifies limitations that motivate the proposed methodology.

## 2 | LITERATURE REVIEW

This section reviews existing methods for detecting elevator operational states. Subsection 2.1 covers traditional approaches, Subsection 2.2 explores Artificial Intelligence (AI) -based techniques and Subsection 2.3 discusses their limitations and the contribution of the proposed framework.

### 2.1 | Traditional approaches

Alternative methods have been developed to measure and estimate not only the presented three operational states, but also phenomena that develop within them. Sensors have been widely used to measure physical magnitudes in the system, which allow engineers to devise analytic models to monitor physical processes related to these operational states.

In this regard, (Skog et al., 2010) proposed the Generalized Likelihood Ratio Test (GLRT)-based methodology to differentiate between the first two operational states of elevators, $\mathcal{H}_0$ and $\mathcal{H}_1$. This methodology uses the acceleration signal provided by an accelerometer in combination with a specifically formulated threshold to distinguish between the two operational states (this methodology is presented and analyzed in Section 5.1). (Guo et al., 2021) used a monitoring module comprised of four sensors: (i) an accelerometer to determine the position, speed, and direction of the cabin, (ii) a barometer to estimate the cabin height, (iii) three magnetic switches to detect the status of the doors and the position of the first floor, and (iv) a video camera to acquire the data inside the car. As a potential drawback, the cost of such setup should be considered, as there are 4 sensor systems installed in the elevator cabin. (Skog et al., 2017) also focused on monitoring the elevator operational states through sensor fusion. They proposed a framework based on multiple soft sensors to monitor car position, floor levels, vibrations, abnormal stop alarm, ride quality parameters and door status. The data used by the soft sensors is provided by an array of 32 inertial measurement units (IMUs) that include an accelerometer and a magnetometer, thus lowering sensors' intrinsic uncertainty. The number of employed IMUs increases the cost of the solution considerably. More recently, (Marinov et al., 2020) proposed a smart multi-sensor node to monitor the condition of elevators, merging barometer, accelerometer and magnetometer signals to obtain displacement and the speed of the elevator cabin. Accelerometer and barometer drift are reduced by merging both signals to obtain the actual position of the cabin. However, both acceleration and pressure patterns shown in the paper exhibit variability and drift even when the elevator is stopped, which leads to inconsistencies in the detection of the operational states when using conventional threshold-based approaches. Similarly, (Zhao et al., 2017) presented a hybrid floor identification algorithm using barometer and wireless access point data to locate individuals inside buildings, regardless of the vertical transportation system. However, it may not be suitable for elevator-specific scenarios due to the need for nearby access points.

### 2.2 | AI-based approaches

Artificial intelligence- (AI) based machine learning (ML) approaches, particularly deep learning (DL) models, offer a powerful alternative to traditional methods by learning complex patterns directly from raw data (Sheng et al., 2024). These models excel at identifying subtle and overlapping signal features that traditional approaches often fail to handle effectively. Built using large datasets and optimized through the minimization of an error-based loss function, ML and DL models achieve high accuracy.

Conventionally, one way to process sensor data for ML models has been through time- and frequency-domain feature extraction. These techniques rely on user-defined features, such as mean, variance, Fourier transform, signal energy, or kurtosis, which are computed iteratively using a rolling window. The extracted features are then used to design ML models for predictive tasks. While these methods allow engineers to approach prediction with some understanding of the physical phenomena, they lack the capability to autonomously learn and adapt to complex data patterns. Indeed, this is a limitation that some DL models are designed to overcome.

In this line, while not directly applied to elevator operational state detection, (Gu et al., 2024) proposed a DL model for urban risk assessment to evaluate city-scale earthquake-induced elevator passenger entrapment. The authors highlight the value of their DL model implementation, which processes a high volume of data, in contrast to their previous research, where simpler assumptions were made, resulting in findings that could be generalized. (Mishra & Huhtala, 2019b) presented a multi-layer perceptron (MLP) neural network that rely on features that are automatically extracted during training. The model they propose utilizes accelerometer vibration data to assess elevator ride quality. (Mishra & Huhtala, 2019a) also proposed a DL model that combines acceleration and magnetic signals to detect elevator faults, vertical displacement and stopped operational states and ride quality features. Other state-of-the-art ML algorithms, such as support vector machines (SVM) (Liono et al., 2018) and random forest (RF) (Mishra & Huhtala, 2019b) have also been applied to detect the vertical displacement of elevators. However, the models presented in these papers cannot be directly used to predict operational states comprehensively. This is either because they do not specifically focus on the signal patterns related to the three operational states or because they address only certain operational states while relying on additional sensors to enhance predictions.

Beyond the application field of elevators, scientific literature presents various ML- and DL-based approaches to detect and classify patterns within time-series signals, such as acceleration. In this line, (Zhang et al., 2019) presented a 1D CNN for the state identification of stiffness and mass changes in three different structures. The proposed model uses raw acceleration data from several accelerometers installed throughout the structures as input, processing all inputs in batches with as many dimensions as there are input channels in the convolutional layer. The vibration patterns processed and learned by their model rely solely on impact-shaped oscillations produced by hammer strokes. Among others, it is worth mentioning the paper published by (Li et al., 2022), that estimates machine failures projected in time-series through a WaveletKernelNet model and provides interpretability. For this purpose, they modify the first convolutional layer to process a Continuous Wavelet Transform (CWT) on input signal based on a mother wavelet. The CWT layer is a suitable method to extract features of the rapid decaying oscillations when the machine failures induce impact-shaped oscillations. However, the applicability is limited in other contexts. For instance, the operational states of elevator journeys induce not only rapid decaying oscillations in acceleration signals (see Section 3, door movement acceleration patterns) but also slower-evolving signal patterns (such as those from cabin's vertical displacement).

Similarly, Michau *et al.* also proposed a fast discrete wavelet transform (FDWT) -based deep NN that can extract a meaningful and sparse representation of high-frequency signals without labelled signals (Michau et al., 2022). The framework, named the denoising sparse wavelet network (DeSpaWN), mimics an L-level wavelet cascade, whose wavelet transforms are modeled as learnable CNN. Thus, authors prove wavelet transform operations can be replaced with convolutional layers that can be tuned by means of available data, achieving a signal denoising tool that can be used in further classification stages. Likewise, more recently, (Li et al., 2023) also posed graph wavelet networks for multiscale feature extraction and fault diagnosis of machines. In this approach authors present a spectral graph wavelet network that can deal with the diagnosis of faulty states. However, instead of processing sensors time-series data in the feature extraction layers of their model, they first create a graph from the different signal sources and the model performs a feature extraction based on the interrelations among graph parameters. This algorithm might be a good choice in environments where multiple data sources are available, and their correlations are not explicitly detectable, but the abstraction level from physical magnitudes is also higher than in approaches that do not employ graphs.

### 2.3 Limitations, contribution and organization

On the one hand, traditional methods for elevator operational state detection, such as the GLRT, have been widely used due to their simplicity and effectiveness in certain scenarios. However, these methods rely on predefined thresholds, which makes them highly sensitive to noise and the inherent variability of the operational conditions of the elevator. Additionally, GLRT and similar traditional approaches often struggle to handle overlapping and transient signals, such as those produced during the vertical displacement of elevators (see **FIGURE 1** and **FIGURE *3*A** patterns), which complicates the detection of the operational state. Sensor fusion-based methodologies, which combine data from multiple sensors, have been

proposed to address these limitations. However, they introduce additional complexity, increased cost, and greater computational demands, especially when dealing with noise reduction and signal processing.

On the other hand, despite the ability of AI-based models to learn patterns from signals and extract valuable information, they are often considered black boxes because their predictions rely on relationships automatically learned during training. This makes it difficult to understand how predictions are made. To address this limitation, researchers have increasingly focused on enhancing the interpretability and explainability of artificial intelligence (XAI) models. The limited transparency of many deep learning and machine learning models highlights the need for methods that connect accuracy with comprehensibility. This is especially important in contexts where actionable insights are required. Interpretable deep learning models allow practitioners to elaborate the connections between input signals and predictions, revealing underlying physical phenomena and fostering trust in the system. For example, interpretable DL models have been developed in various domains, such as structural health monitoring (Shabbir et al., 2024) and manufacturing processes (Li et al., 2022), highlighting their value in practical applications.

To tackle the detection of the operational states of interest (cf. **TABLE *1***), this paper presents an accurate and interpretable classification approach for elevators based on a 1D-CNN model. A 1D-CNN is well suited for this task, as convolutional layers effectively extract frequency-related features from single-axis acceleration data, a capability demonstrated in industrial monitoring and structural health applications (Li et al., 2022; Michau et al., 2022; Zhang et al., 2019).

The proposed method efficiently detects distinct acceleration patterns that emerge during elevator operations, even when patterns overlap, such as during vertical displacement of the elevator (see **FIGURE 1** and **FIGURE *3***A patterns). Computational novelty lies in the model's ability to handle overlapping and transient patterns, such as those caused by door movements and vertical displacement, which pose challenges for the traditional methods it has been compared to. Unlike predefined kernel-based models like WaveletKernelNet, the 1D-CNN dynamically trains its kernels to adapt to both short- and long-lasting signal patterns. Furthermore, the model achieves high precision and interpretability while relying solely on a single accelerometer signal, reducing complexity and cost, making it a scalable and practical solution for real-world applications.

The proposed approach leverages convolution operations within a 1D-CNN to obtain an interpretable representation during the feature extraction stage, which allows the engineers to acquire meaningful insights about the signal processing carried out to identify patterns associated with each elevator operational state. Efficiency is achieved by using a single sensor information, whilst the proposed methods to date use more sensors (Guo et al., 2021; Mishra & Huhtala, 2019a, 2019b; Skog et al., 2017). This monitoring system is universal and non-intrusive because it works as a standalone solution without the need to interact with other systems and can be smoothly integrated with mechanisms and control systems already working in the elevator.

The rest of the paper is structured as follows. Section 3 introduces the signal patterns that develop during the nominal operation of elevators. Section 4 describes the proposed approach. Section 5 gathers some of the most conventional approaches that have been used to date to tackle the detection of the operational states of elevators. Section 6 shows and discusses the model setup and experimental results. Finally, Section 7 gathers the conclusions and the future lines.

## 3 | ELEVATOR DYNAMICS & PROBLEM STATEMENT

The dynamic behavior of an elevator is captured using an accelerometer, which measures the acceleration of the elevator cabin's motion. **FIGURE 1** illustrates characteristic acceleration profiles during nominal elevator operation, highlighting three distinct dynamic patterns. Red rectangles indicate the acceleration patterns measured during the vertical displacement of the cabin, capturing its motion from the start of movement to the stopping point. Green rectangles represent acceleration patterns caused by vibrations from the cabin door mechanism as it opens and closes at each floor. Sections of the signal not tagged correspond to stationary or stop states, where both the cabin and doors remain motionless. These three patterns directly correlate with the operational states described in **TABLE 1**.

The proposed framework focuses on the classification of elevator operational states through a supervised learning process using a dataset, $S$, comprised on $K$ samples,

$$S = \{x_i, y_i\}_{i=1}^{K} \tag{1}$$

where the pair $\{x_i, y_i\}$ contains data related to the $i$-th observation, $x_i \in X$, $y_i \in Y$. The vector $X \in R^{K\times 1}$ contains information on the acceleration signal and $Y \in R^{K\times 1}$ contains information on the elevator operational state.

This is a multi-class classification problem, where each output $y_i$ can represent one of the following values (cf. **TABLE 1**): $y_i = \{\mathcal{H}_0, \mathcal{H}_1, \mathcal{H}_2\}$. The main challenge lies in designing a detection framework that (i) distinguishes transient (vertical displacement and door movement) and stationary (no-movement) acceleration patterns and (ii) accurately classifies the two transient patterns, which, despite

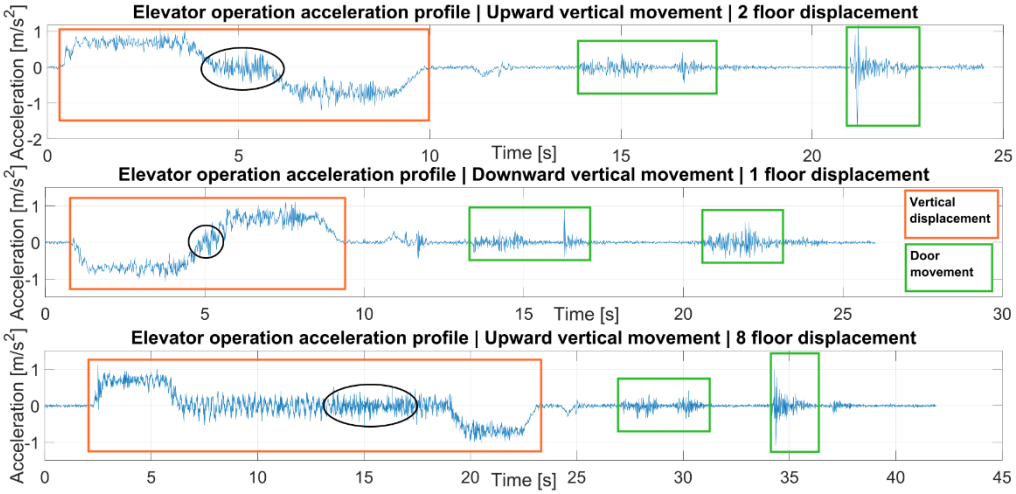


**FIGURE 1** Acceleration profiles of elevator under nominal operation.

their intrinsic differences in energy, frequency components, and duration, share similarities. Specifically, during the vertical motion of the cabin, occasional vibration patterns emerge, and they share temporal and frequency characteristics with those produced by door movements, as highlighted by the black ellipses in **FIGURE 1**. These characteristics of the acceleration signal make the prediction of operational states using computer-aided methods challenging for two reasons. First, the magnitude of vibrations generated during door movements is similar to those produced during the constant speed displacement of the cabin, making it difficult to differentiate these states based solely on signal amplitude. Second, the patterns marked in black ellipses share overlapping time and frequency components (c.f. **FIGURE *3***A approximately 28 Hz) with the patterns induced by door movements, further complicating the classification task. These two reasons pose significant challenges for accurate classification, particularly when relying on simpler signal processing methods such as FFT and CWT.

All the models in this research work have been assessed using the same training, testing and validation procedure. That is, an initial split of 70%-30% is performed to separate (a) training and testing from (b) the validation set. The training set is used to fit the model parameters, and the testing set is used to evaluate the model performance on the trained model, which is obtained by splitting further into 70%-30%. After selecting the model with the best performance, the results are calculated on the validation set, which is 30% of the dataset. For the training process and testing processes, the temporal sequence of the dataset was preserved. without random shuffling, to maintain the chronological order of the data.

The dataset was recorded and labeled by an expert from the elevator manufacturer which supports this research. It consists of approximately 1,000,000 rows (data samples captured at 100 Hz), representing 250 random elevator journeys. The data file contains two columns: one for the Z-coordinate acceleration values and another for the class of the operational state (0 for $\mathcal{H}_0$, 1 for $\mathcal{H}_1$, and 2 for $\mathcal{H}_2$). The journeys were randomly configured during the experiment, with varying starting and ending floors and different journey lengths to ensure variability. As part of the research in this paper, the preprocessed dataset is publicly released in DATASET.

## 4 | PROPOSED ELEVATOR STATUS CLASSIFICATION FRAMEWORK

### 4.1 | Proposed approach

To address the detection of the three operational states using a single acceleration signal a DL model has been designed. The design of the model starts with the analysis of the application requirements and available data. At the same time, the possibility to interpret and understand the internal data processing carried out throughout the DL model is of utmost relevance, as it can help find out the physical phenomena occurring during the elevator journey to assist engineers in future failure detection and maintenance tasks.

The acceleration dynamics analyzed in Section 3 show patterns of varying time duration that change either slowly (vertical displacement) or abruptly (door movements). These dynamics in the available data have an impact on the selection and design of the target algorithm, since a manner to capture both short- and long-lasting nature of patterns is required. In terms of the application requirements, an inference lag of 2 seconds is targeted, to provide near real-time classification.

Additionally, algorithm simplicity is another key consideration in the design of the model, defined as two complementary aspects: a shallow architecture to enhance interpretability and compact code to minimize the model's size. This simplicity makes the model not only easier to understand but also computationally efficient, ensuring its suitability for future deployment on real-time embedded systems.

Considering the requirements discussed in the previous paragraph, a 1D-CNN has been designed to setup an algorithm that can (i) provide fast inference, (ii) effectively capture the different patterns of the operational states in acceleration data and (iii) give rise to feature interpretability, in agreement with (Li et al., 2022)(Michau et al., 2022). To accomplish these objectives, the 1D-CNN approach developed in this work takes advantage of the ability of convolutional layers to mimic the CWT (Li et al., 2022; Mallat, 2016; Michau et al., 2022). Like the CWT, the 1D-CNN processes data at different resolutions using learned kernels. These kernels extract frequency features over a short time period at various scales, imitating wavelet functions. During training, the CNN adjusts these kernels to work like different types of custom wavelets, adapting to capture patterns of varying lengths and characteristics in the data. The proposed 1D-CNN model learns all kernel coefficients during training, offering great flexibility to capture both short-lived oscillations (door movements) and slower-evolving patterns (vertical displacement), making it suitable for the diverse signal characteristics encountered in elevator operations. Thus, the kernels of the convolutional layer of the proposed 1D-CNN can initially be defined as a generic wavelet function (see types in (Adeli & Jiang, 2006; X. Jiang & Adeli, 2005)) described in Equation (2), which is then tuned and refined during training to optimize feature extraction across multiple resolutions, scales and kernel functions.

$$\psi_k[\tau] \equiv |s|^{-p}\psi\left[\frac{\tau}{s}\right] \tag{2}$$

where $\psi_k[.]$ represents the $k$th generic wavelet function (initial $k$th kernel), $s$ is the scaling factor, $p \geq 0$ is a fixed constant and $\tau$ represents each sample of the generic wavelet function $\psi$.

The generic wavelet functions defined in Equation (2) act as the kernels of the convolutional layer, expressed as follows:

$$x[n] * \psi_k[\tau] = \sum_{\tau=0}^{T-1} \psi_k[\tau]x[n-\tau] \tag{3}$$

where $*$ denotes the convolution operation.

The equivalence between the CWT and the convolution operation performed by the convolutional layer allows to (i) extract the similarity between the input and a given $k$ kernel and (ii) tune each of the kernels of the convolutional layer as if they were independent wavelet functions that adapt their parameters ($s$, $p$ and $\psi$) conveniently as a whole to the patterns of varying duration and amplitude in the signal.

Different 1D-CNN configurations have been assessed, seeking a trade-off between the accuracy in the detection of the three operational states and the depth and complexity of the model in terms of its data processing interpretability. To achieve the highest possible interpretability, a single convolutional layer has been targeted, followed by MaxPooling1D (pool size = 2) for dimensionality reduction, a Dense layer with 100 nodes, and a final output layer with three nodes using softmax activation for classification.

The network configurations that have been tuned are the number of filters $K$ of the convolutional layer, the kernel size T of each filter and the block size $N$ of acceleration samples used as input. It has been assumed a sampling frequency of 100 Hz for the acquisition of the acceleration signal, which in turn has a great importance regarding the eligibility of the size of the input block. A sampling frequency of 100 Hz provides a sufficient trade-off between the data burden and the capability of the acquisition system to capture the dynamics of the elevator elements that participate in the operational states. In this regard, 50, 100, 150 and 200 sample blocks have been tested for the tuning of the hyperparameters, considering the block size of 200 samples as the maximum size as it defines the targeted maximum admissible time lag (2 seconds) for the inference time of the proposed 1D-CNN model.

Results of the evaluated configurations are shown in **TABLE 3** (Section 6.1), which demonstrates that hyperparameters that achieve the best performance are the combination of 100 filters or kernels ($K$=100), 10 coefficients for each kernel ($T$=10) and a block size of 150 samples ($N$=150).

To assess the performance of the different models evaluated in this research, a common assessment procedure has been performed to compare accuracy of the proposed model and those used as benchmark. The operational states identified by each methodology have been compared to actual operational states labelled experimentally. This comparison between the detected and actual operational states is presented by means of a qualitative assessment (see **FIGURE 9**) and the F1 score metric (Chicco & Jurman, 2020).

$$\mathrm{F1} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}} \tag{4}$$

where $TP$ denotes true positive, that is, the model correctly infers an outcome belongs to the analysed hypothesis. $FP$ denotes false positive, that is, the model incorrectly infers an outcome belongs to the analysed hypothesis. $FN$ denotes false negative, that is, the model incorrectly infers an outcome does not belong to the analysed hypothesis.

F1 score was selected as the primary evaluation metric due to the class imbalance present in our dataset. Specifically, samples representing door movement operational state are fewer than those for vertical displacement and stop operational states. The F1 score, which is the harmonic mean of precision and recall, is a better indicator of model performance in such cases, as it accounts for both $FP$ and $FN$. This metric is preferred over accuracy, which can be misleading in the presence of imbalanced classes (Aizpurua et al., 2021).

The proposed 1D-CNN framework and the steps carried

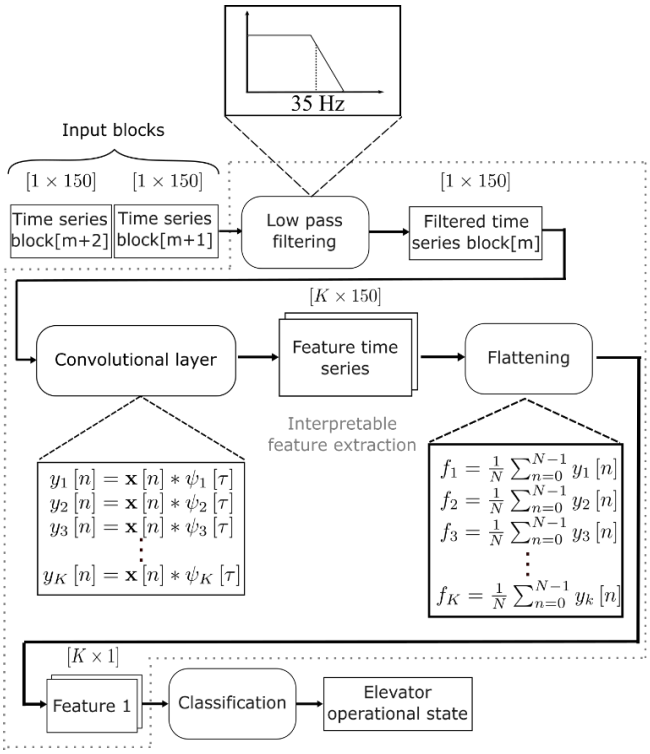


**FIGURE 2** Block diagram of the proposed approach.

out inside the algorithm are shown in **FIGURE 2**, in accordance with the obtained hyperparameters combination. The 1D-CNN processes fixed-length data blocks that are initially filtered using a low-pass filter with a cutoff frequency of 35 Hz. This filtering ensures that the patterns associated with operational states are preserved in the signal while removing unrelated frequency components. This cutoff frequency was selected based on the dominant frequency components observed in the acceleration patterns of the three operational states (**FIGURE 3**A). It is high enough to retain the relevant signal dynamics while effectively attenuating higher-frequency disturbances that do not contribute to classification. Besides, the operations delimited by the dot line are fully interpretable, which allows the designer to understand the underlaying acceleration magnitudes the classification step considers. The data processing performed by the convolution layer and the feature map obtained after the flattening operation will be described in more detail in Section 4.2. The final classification step is also described.

### 4.2 | Interpretability of features

To understand the feature extraction process in the 1D-CNN, the convolutional layer and the subsequent steps are analyzed using real acceleration data captured during an elevator journey. While the raw time-series data (see **FIGURE 1**) may seem visually interpretable, this paper focuses on how the 1D-CNN processes the input to identify patterns that can be overlapped at some operational states and adjust the signal processing framework steps (see **FIGURE 2**) accordingly. This interpretability improves the model's transparency and reliability, aiding decision-making, maintenance, and fault diagnosis for engineers and maintenance teams by providing clear links between input signals and model predictions.

The training step used to tune the 1D-CNN model plays a key role in fitting the 100 kernels/filters in the convolutional layer. To illustrate raw patterns that can be found using a traditional signal processing algorithm, **FIGURE 3**A provides a time vs. frequency plot obtained through a conventional CWT using Morlet wavelets, depicting how different scale wavelets or kernels capture energy concentrations of frequency components at various time instants. Energy concentrations corresponding to the beginning and end of vertical displacement are located around 0.7 Hz at the illustrated time instant, while those related to door movements appear around 28 Hz. It should be noted that although CWT effectively highlights the targeted operational states on the plot, it is difficult to distinguish other areas (marked by red ellipses) that also show energy concentrations around 28 Hz, which overlap with the operational state described by the vertical displacement. This limitation underscores the necessity of the subsequent steps in the 1D-CNN architecture for more accurate classification. **FIGURE 3**B complements this by showing the frequency responses of all 100 kernels of the proposed 1D-CNN after training, highlighting the passbands assigned to each kernel. These passbands correspond to the frequencies where most of the energy is concentrated in the acceleration signal, as indicated by the CWT plot in **FIGURE 5A**. This demonstrates that the tuning of the kernels aligns with the frequency components present

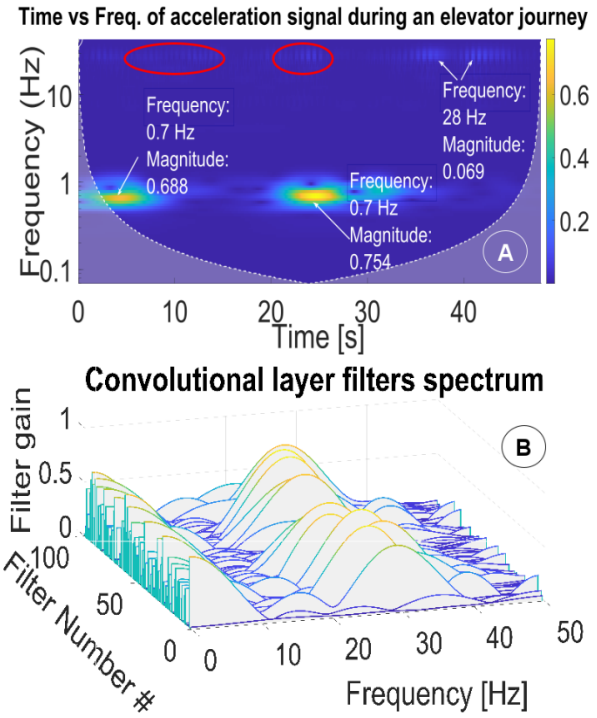


**FIGURE 3** CWT of the acceleration signal and the spectrum of the 100 kernels defined in the convolutional layer of the 1D-CNN.

in the acceleration signal, effectively adapting the kernels to the characteristics of the data.

The diagram shown in FIGURE 4 gathers all the steps carried out by the 1D-CNN model during the forward-propagation of the time series blocks. These four steps in FIGURE 4 are addressed in more detail in the following subsections where acceleration time series blocks of a complete elevator jouney are used to show the transformation they suffer before the final classification step is carried out.

### 4.2.1 | Time series data blocks

Acceleration time series data is introduced in batches or blocks of 150 samples into the convolutional layer. FIGURE 5.1 shows two different acceleration signal sections of a single elevator journey. The acceleration signal on image A is composed of 7 time series blocks (150x7 samples), whereas the signal on image B gathers 8 time series blocks (150x8 samples). Each block differentiated by colors is processed separately within the steps of the model.

### 4.2.2 | CNN layer output | 100 filters bank

Each block depicted in FIGURE 5.1 is convolved 100 times (once with each kernel in the convolutional layer) producing an output signal for each convolution as shown in FIGURE 5.2.

### 4.2.3 | Feature extraction for each time series block

The third step within the 1D-CNN model focuses on flattening of the filtered output blocks. Each feature in the feature map is created by calculating the mean value of an entire filtered output block, adding together 100 features. The feature map

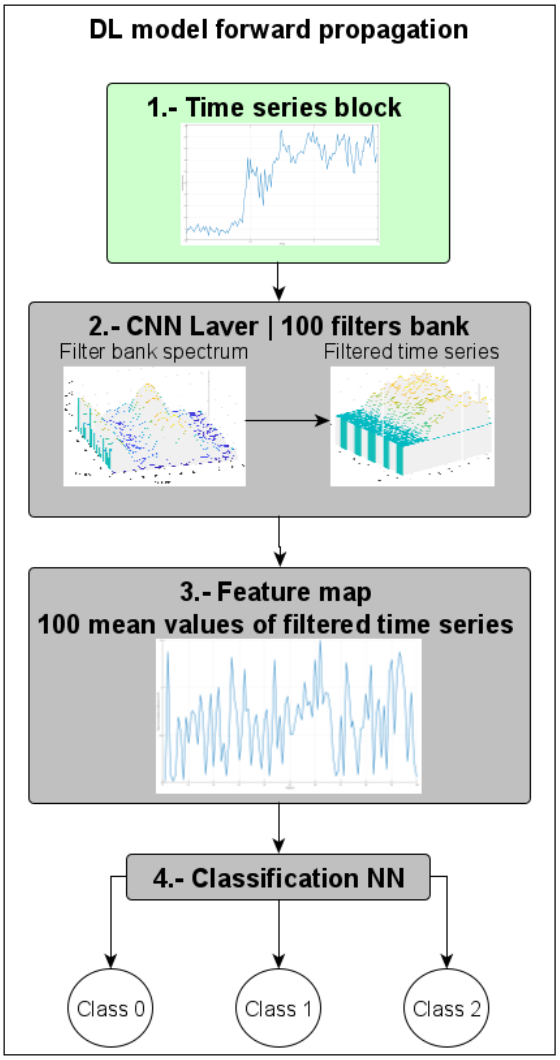


**FIGURE 2** 1D-CNN model operation for each time series block.

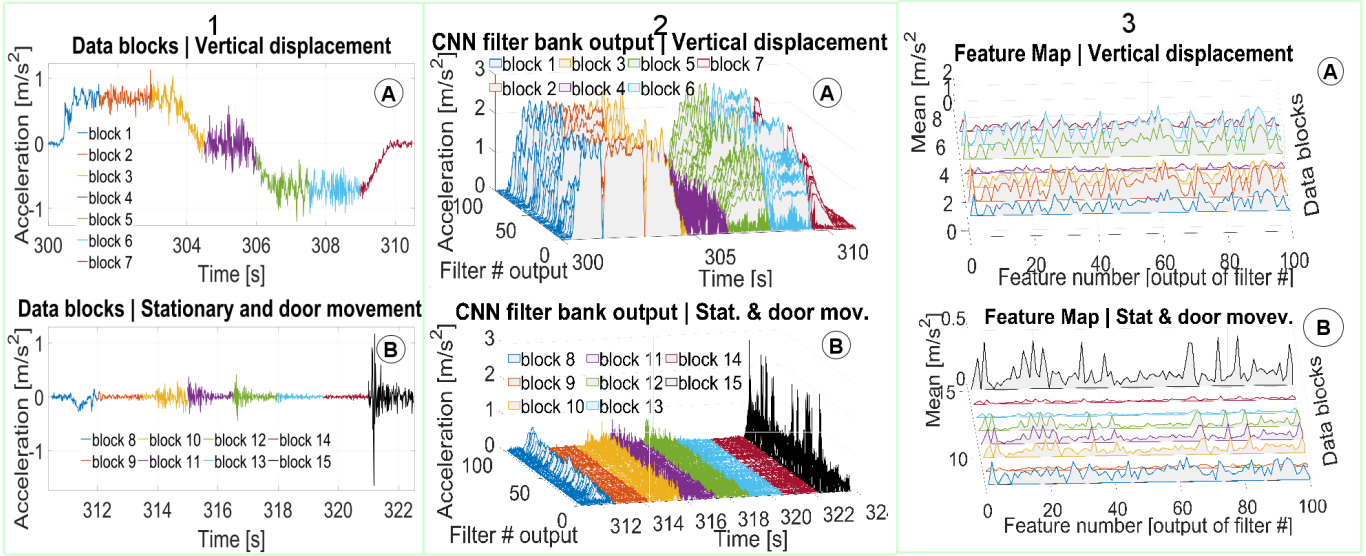


**FIGURE 3** Transformation of data blocks within 1D-CNN model forward propagation. Column 1 shows the blocks of data (colored differently) that enter the framework chronologically ordered. Column 2 depicts the output of the convolutional layer for each data block. Column 3 shows the features extracted after flattening is applied for each data block that entered the framework.

related to each time series block on FIGURE 5.1 is shown in FIGURE 5.3. Image A on FIGURE 5.3 shows the feature maps of all data blocks are similar in the relation between maximum and minimum values of the 100 features, which draw a similar profile. These feature maps belong to the vertical displacement acceleration pattern. Looking at image B, two different feature map patterns can be seen: blocks 10-11-12-15 draw a similar profile, whilst blocks 8-9-13-14 have different profiles. The former blocks correspond to acceleration patterns generated by door movement, while the latter blocks belong to the stationary state of the elevator. It should be noted that the profiles within the stationary state class are more heterogeneous in both magnitude and especially shape compared to profiles within the other two states. This variability is expected, as the stationary state can encompass a range of conditions that are not as tightly defined as the door movement or displacement states.

#### 4.2.4 | Classification of time series block

The final step of the 1D-CNN model classifies the time-series blocks based on their feature maps into three classes corresponding to the hypotheses $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$, as defined in the previous sections. This step is implemented using a fully connected neural network (NN) layer, which maps the feature maps to three neurons representing the possible outcomes. Each neuron outputs a value indicating the predicted probability of the input belonging to its respective class. The softmax activation function ensures that these output values are within the range [0,1] and sum to 1. The softmax function for class $i$ is defined as follows:

$$softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{3} e^{z_j}} \tag{5}$$

where $z_i$ is the raw score for class $i$. The final output represents the likelihood of the input belonging to each class, corresponding to the hypotheses $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$.

## 5 | IMPLEMENTATION OF STATE-OF-THE-ART METHODOLOGIES

This section describes the three conventional methodologies that have been used to classify the operational state of elevators, which will serve as benchmark. The focus of this paper is on the implementation of classification methods that have shown good performance in similar tasks, as shown in TABLE 2.

**TABLE 2** Benchmarking of different methods for elevator operational state classification.

| ID | Method | Refs |
|---|---|---|
| A | Generalized Likelihood Ratio Test (GLRT) | (Skog et al., 2010, 2017) |
| B | GLRT and sensor fusion | (Nikolov et al., 2020) |
| C | Time and frequency domain feature extraction and ML | (Liono et al., 2018; Mishra & Huhtala, 2019a, 2019b; Nguyen et al., 2019) |

The following subsections provide a detailed description of each methodology, outlining their fundamental principles and their suitability for comparison with the proposed 1D-CNN-based classification framework.

### 5.1 | Generalized Likelihood Ratio Test on acceleration

The Generalized Likelihood Ratio Test (GLRT) was originally employed to differentiate between stationary and non-stationary operational states in the acceleration signal (Skog et al., 2017). To our knowledge, the GLRT has not been applied to detect door movements, as the oscillations produced during this state (described in FIGURE 1) do not deviate significantly from the signal's mean. However, the GLRT methodology is still employed for the detection of the three operational states for a unified comparative framework.

The original GLRT operates within a composite hypothesis framework, where two hypotheses ($\mathcal{H}_0$ and $\mathcal{H}_1$) are tested based on the likelihoods derived from the acceleration signal (Kay, 1993; Skog et al., 2010). In the context of classifying different operational states from acceleration signal (Skog et al., 2010; Veltink et al., 1996) used the so-called acceleration-moving variance detector $T_v'(\boldsymbol{z}_n^a)$ within the GLRT framework, which yields the following mathematical expression:

$$T_v'(\boldsymbol{z}_n^a) = \frac{1}{N\sigma_a^2} \sum_{k\in\Omega_n} \|\boldsymbol{y}_k^a - \bar{\boldsymbol{y}}_n^a\|^2 < \gamma_v \tag{6}$$

where, both the upper-index $(.)^a$ and sub-index $(.)_a$ denote acceleration coordinate and the variables are defined as follows:

- $\boldsymbol{z}_n^a \triangleq \{\boldsymbol{y}_k^a\}_{k=n}^{n+N-1}$,
- $\bar{\boldsymbol{y}}_n^a = \frac{1}{N}\sum_{k\in\Omega_n} \boldsymbol{y}_k^a$,
- $\Omega_n = \{l \in N : n \le l < N-1\}$,
- $N$ is the total number of samples,
- $\sigma_a^2$ is the variance of acceleration,
- $\gamma_v$ is the threshold used to distinguish between $\mathcal{H}_0$ and $\mathcal{H}_1$,
- $\boldsymbol{y}_k^a$ represents every acceleration sample.

The GLRT was specifically designed to differentiate between stationary $\mathcal{H}_0$ and non-stationary $\mathcal{H}_1$ operational states of an accelerometer, depending on the value obtained compared to a threshold $\gamma_v$.

In this work, the objective is to determine the elevator operational state as defined in TABLE 1. Thus, two threshold levels $\gamma_{v_1}$ and $\gamma_{v_2}$ must be specified to test the set of hypotheses. Typically, these thresholds are set experimentally for each specific elevator system, although different signal processing techniques exist to define their values (Simão et al., 2017). In addition, these thresholds can also be concurrently (online) redefined throughout the runtime.

### 5.2 | GLRT and barometer sensor fusion

To enhance the classification of the operational states described by the three hypotheses $\mathcal{H}_0$, $\mathcal{H}_1$ and $\mathcal{H}_2$, the GLRT detector has been reinforced by the data provided by a barometer. (Nikolov et al., 2020) proved that the pressure variations measured by a barometer during the nominal operation of an elevator can improve the operational state detection. The pressure signal is an indicator that highlights vertical displacement movements of the elevator as shown in FIGURE 6. The linear variations (up/down slopes) of the pressure signal in the bottom image of FIGURE 6 (delimited by red rectangles) are perfectly synchronized with acceleration patterns produced during the vertical displacement of the cabin shown in the top image. Nevertheless, pressure signal does not seem to be effective in monitoring door movements (described by green rectangles in the top image), because it remains unchanged when vibration patterns are induced in the acceleration signal.

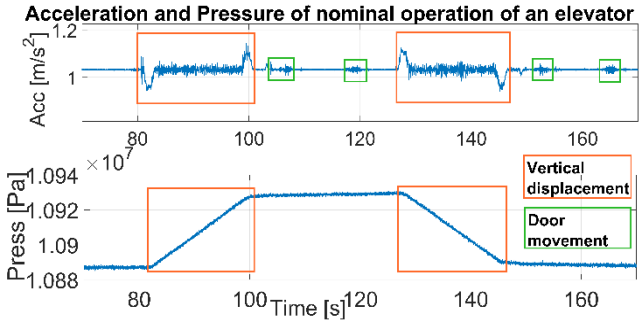


**FIGURE 4** Acceleration and pressure during the nominal operation of an elevator.

Knowing the strength (*i.e.* the ability to detect vertical displacement) and weakness (*i.e.* the inability to detect door movements) of the pressure signal, a combined detection framework has been developed. On the one hand, the derivation of the pressure signal has been used as an indicator of the vertical displacement of the elevator. On the other hand, the GLRT has been used as an indicator to detect the movement of the door, having to define, in this approach, only a single threshold $\gamma_v$ experimentally.

In scenarios where both indicators are triggered, the criterion of the pressure-based indicator prevails over the GLRT, since the pressure variation is more directly correlated to the operational state it is monitoring (vertical displacement).

### 5.3 | Feature extraction and traditional machine learning

There are different ML algorithms that have been used to address directly or indirectly the estimation of the elevator operational states targeted by this research work, e.g. (Liono et al., 2018; Mishra & Huhtala, 2019a, 2019b). To provide a general viewpoint of the different algorithms that can be used, this section presents three ML models and different configurations of them. The key point of this approach consists of extracting time- and frequency-domain features from acceleration data using the rolling window technique, and then applying the mentioned traditional ML models to classify the operational states.

The steps carried out within this methodology are shown in FIGURE 7. The approach starts with the *feature extraction* step, which extracts the main signal features. This has been done with the tsfresh library (Christ et al., 2018), with the default configuration for the *extract_features* function. By using the default configuration, a set of $K = 570$ features are computed, including some common features such as mean, variance, peaks, energy and kurtosis.

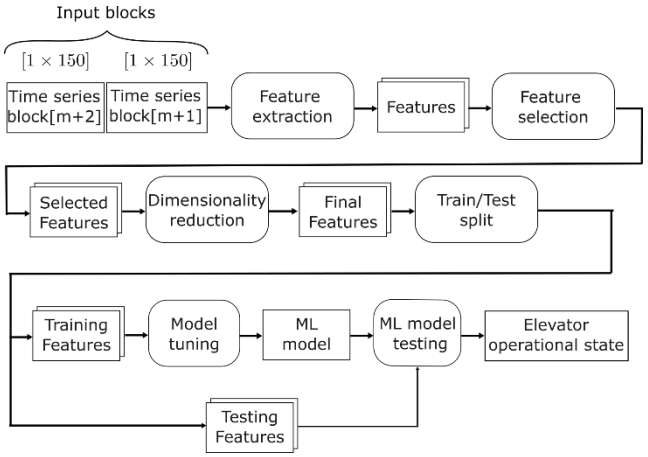


**FIGURE 5:** ML classification model design approach.

*Feature selection* step is implemented based on a features relevance table obtained training a RF in combination with the Gini index (Daniya et al., 2020). This is applied to select only the features that are representative of at least one class. As a result, from the initially computed $K = 570$ features, $K = 289$ relevant features are selected. With the selected features a *dimensionality reduction* is implemented through Principal Component Analysis (PCA). This reduces the input space, prevents multicollinearity and keeps 95% of variability, reducing the feature space into $K = 7$ features.

After data normalization, ensuring that all variables have the same weight for the distance-based models, the *hyperparameter tuning* of the ML models is carried out through a 10-fold cross-validation. To preserve the temporal nature of the dataset, the cross-validation process was conducted using a forward-chaining k-fold approach rather

than a traditional k=10 randomly sampled cross-validation. In this method, the first fold (k=1) trains the model on the initial 10% of the dataset and tests it on the subsequent 10%. For the second fold (k=2), the model is trained on the first 20% and tested on the next 10%, and this process continues iteratively until all folds have been evaluated. This approach ensures that the temporal sequence of the data is preserved, avoiding any leakage of future information into past predictions. Three ML classification algorithms have been tested including Support Vector Machines (SVM), Random Forests (RF) and MLP-based neural networks, using the *scikit-learn* library (Kramer & Kramer, 2016).

All machine learning models were tuned using a grid-search approach, analyzing their performance across a predefined grid of hyperparameters. For the SVM model, the main hyperparameters considered were the kernel type (*kernel*), the cost parameter ($C$) and the influence of training parameters on classification ($\gamma$). These were selected from the following values: $kernel = [rbf, polynomial]$, $C = [10,100,1000]$, $\gamma = [scale, auto]$. For the RF model, tuning was based on the maximum tree depth (*max_dept*), the minimum (*min_samples_split*) and maximum (*max_features*) number of features required to split a node. These parameters were evaluated with the following values: $max\_depth = [None, 5,10]$, $min\_sample\_split = [3,5]$ and $max\_features = [None, sqrt]$. Lastly, for the MLP model, the tuning process involved varying the number of hidden layers ($hidden\_layers$), the number of neurons per layer ($neurons$), the activation function ($activation$), and the training batch size ($batch\_size$). The following values were explored: $hidden\_layers = [10,50,100]$, $neurons = [10,50,100]$, $activation = [relu, logistic]$, $batch\_size = [32,64]$. The loss function used was a categorical-cross-entropy function, conducting 100 training epochs and optimizing the training process through Adam optimizer with a learning rate of 0.001.

Finally, the models were tested on the test-set to evaluate their performance and select optimal hyperparameters. The performance of the selected model is computed though the F1 score [cf. Equation (1)] and evaluated on the validation set.

# 6 | MODEL SETUP & EXPERIMENTAL RESULTS

## 6.1 | Model setup results

The training process was conducted using the same hyperparameters as those described for the conventional ML models in Section 5.3: 100 training epochs, the Adam optimizer with a learning rate of 0.001, ReLU activation for intermediate layers and the 'categorical-cross-entropy' loss function.

**TABLE 3** shows the different hyperparameter configurations and the F1 scores achieved by each of them in the testing set with unseen testing data. It shows that the configuration of 100 filters or kernels ($K$=100), 10 coefficients for each filter (T=10) and a block size of 150 samples ($N$=150) outperforms among the others. Hence, this configuration has been used to validate the model with the validation dataset.

**TABLE 3** 1D-CNN hyperparameters & F1 score of the testing set.

| **Filters [$K$]** | **Kernel size [$T$]** | **F1 score of Block size [$N$]** | | | |
|---|---|---|---|---|---|
| | | 50 | 100 | **150** | 200 |
| 50 | 10 | 77 | 84 | 86 | 86 |
| 50 | 20 | 74 | 87 | 89 | 89 |
| **100** | **10** | 82 | 91 | **97** | 92 |
| 100 | 20 | 83 | 89 | 92 | 92 |
| 150 | 10 | 82 | 87 | 90 | 91 |
| 150 | 20 | 80 | 88 | 91 | 89 |

Beyond classification accuracy, computational efficiency is a key factor for the pursued near real-time deployment. The proposed 1D-CNN model was designed to process input data with minimal computational overhead, ensuring responsiveness in embedded monitoring systems. With the selected configuration (K=100, T=10, N=150), the model processes each 150-sample input block within a maximum inference time of 56 milliseconds on a standard computing platform (e.g. Intel$^{R}$ Core$^{TM}$ i5-8265U CPU @ 1.60GHz). This inference time is well below the targeted 2-second lag, demonstrating that the model can be deployed effectively in practical elevator monitoring applications.

## 6.2 | Experimental setup

The experiments for the assessment of the analyzed methodologies have been carried out in the Onddi tower, a real elevator facility that is used as a test bench, courtesy of Orona. This elevator facility has 17 floors, and the elevator cabin has single door access in all the 17 stops. This elevator is a traction elevator, which is widely used in residential and office buildings. The elevator operated under normal conditions, with two people traveling inside the cabin. To simulate realistic usage, passenger entries and exits were mimicked at random stops.

The data acquisition system is TDK's DK-20789 board equipped with an IO1 Xplained Pro data logger expansion card that stores each data axis at 100 Hz sampling frequency (InvenSense, 2020). The data axes acquired from the DK-20789 are the Z axis of acceleration and the atmospheric pressure data of the barometer integrated in the board. The data acquisition system is fixed on the roof of the elevator cabin. **FIGURE 8** shows the Onddi tower (A) and the location

of the data acquisition system on cabin roof (B).

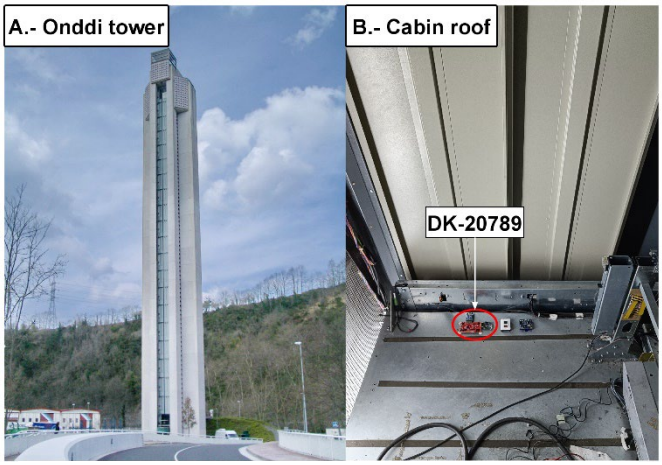


**FIGURE 6** Experiments test bench (A.- Onddi tower), courtesy of Orona and data acquisition setup (B.- Cabin roof).

The dataset used for training, testing, and validation consists of a single file containing data from 250 elevator journeys. The three operational states were labeled by an expert from the elevator manufacturer Orona, based on prior knowledge of the characteristic acceleration patterns induced during cabin rides and door movements. These patterns are well-characterized by the manufacturer, allowing the expert to manually segment the time slots corresponding to each operational state. In total, this file includes roughly 1 million data points.

To provide a clearer view of the dataset distribution, **TABLE 4** presents the percentage of samples corresponding to each operational state. In the entire dataset, 60.3% of the samples belong to $\mathcal{H}_0$, 25.3% to $\mathcal{H}_1$, and 20.8% to $\mathcal{H}_2$, ensuring that all states are sufficiently represented for training and evaluation

**TABLE 4** Label distribution in the dataset

| ID | Total samples | Percentage (%) |
|---|---|---|
| $\mathcal{H}0$ | 602,827 | 60.3 |
| $\mathcal{H}1$ | 253,280 | 25.3 |
| $\mathcal{H}2$ | 207,981 | 20.8 |

### 6.3 | Experimental results

The detection results of the three operational states obtained by means of the proposed 1D-CNN algorithm have been compared with the ones obtained through the conventional methodologies used in the field of elevators (cf. **TABLE 2**).

**TABLE 5** shows the average F1 scores for the tested algorithms on the validation set with operational states displayed in **TABLE 1**. As in the case of the proposed 1D-CNN, the optimal values of block size ($N$) for each algorithm were determined empirically by evaluating multiple sizes (50, 100, 150, and 200 samples) and selecting the size that maximized the F1 score across the test datasets. Hence, it should be noted that the results of the proposed 1D-CNN outperform the rest of the algorithms in both cases, on average and class by class.

**TABLE 5** Average F1 scores of the analyzed methodologies and algorithms.

| Model | Block size [$N$] | Av. F1 [%] | F1 scores per class [%] | | |
|---|---|---|---|---|---|
| | | | $\mathcal{H}_0$ | $\mathcal{H}_1$ | $\mathcal{H}_2$ |
| SVM | 150 | 70 | 80 | 69 | 63 |
| MLP Classifier | 150 | 73 | 81 | 71 | 68 |
| Random forest | 150 | 75 | 82 | 75 | 70 |
| GLRT | 150 | 75 | 87 | 79 | 58 |
| GLRT + baro | 100 | 81 | 89 | 86 | 68 |
| **1D-CNN** | **150** | **96** | **97** | **97** | **93** |

Furthermore, **FIGURE 9** shows the detection of the operational states by each algorithm along with the different acceleration patterns captured during several elevator journeys. For the sake of simplicity and clarity, only RF has been included from the three ML-based conventional algorithms that have been analyzed, as it is the one with the highest score. This figure includes the acceleration signal that has been processed by each of the algorithms, the detection made by each of the algorithms and the actual operational states that have been employed as benchmark. These actual operational states are represented by vertical bands. The darkest bands represent the actual vertical displacement section in the acceleration signal, whereas the brighter represent the acceleration section where door movement happens. White color bands symbolize sections where neither vertical displacement nor door movement occurs.

Regarding the detection results made by the conventional algorithms (GLRT, GLRT plus barometer and RF) **FIGURE 9** shows many discrepancies in the detection of the actual operational states. It is worth noting that conventional algorithms fail in inference consistency, since there are many transitions among their detected operational states under the same actual operational state. By contrast, the detection of the proposed 1D-CNN is more consistent and stable under any actual operational state, despite some lags or anticipations.

### 6.4 | Discussion

The quantitative (**TABLE 5**) and qualitative (**FIGURE 9**) experimental results demonstrate that the 1D-CNN model outperforms analyzed methodologies that have been conventionally applied, with an accuracy of 96% in detecting different elevator operational states. One of the key strengths is the ability to use a single data source (Z-axis acceleration), which simplifies the system and reduces costs while

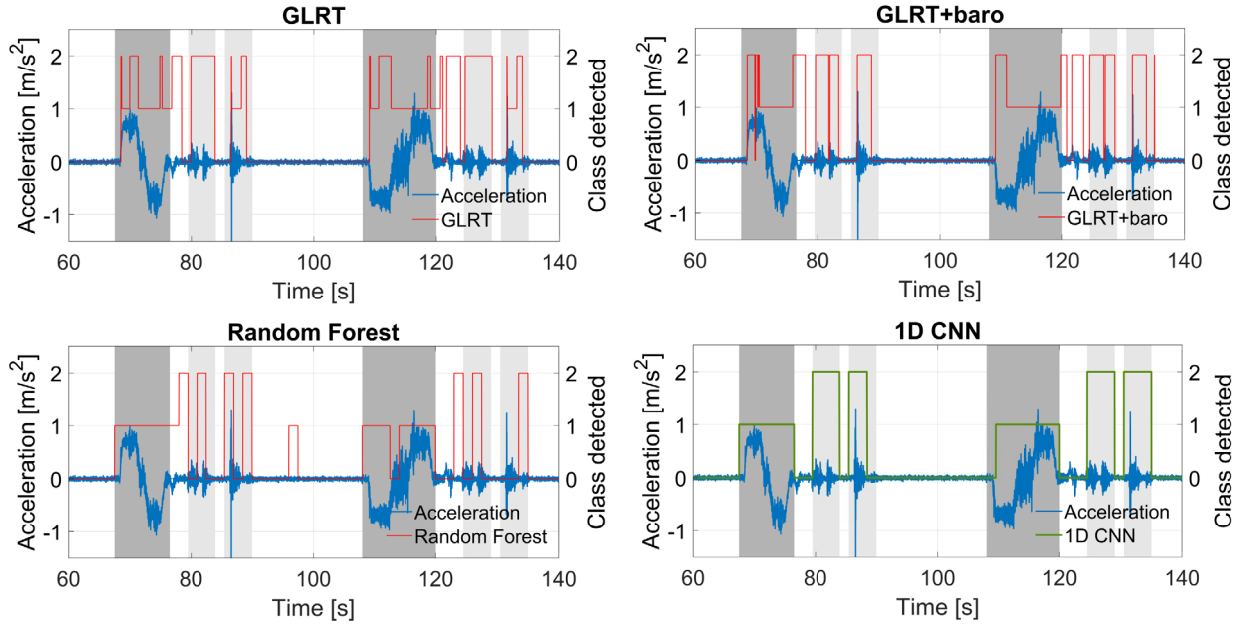


**FIGURE 7 Detection of the elevator's operational states by means of the analyzed models.**

maintaining strong performance. The model accuracy in distinguishing the three operational states highlights the robustness in a real elevator facility with a high variability in the length of elevator journeys. Besides, the simplicity of the 1D-CNN architecture ensures computational efficiency, making it suitable for real-time deployment on embedded systems (Garro et al., 2020).

When compared to the second most accurate model (GLRT with barometer), the proposed 1D-CNN demonstrates clear advantages in both performance and practicality. Despite the cost-effectiveness of modern accelerometer-barometer sensor chips, merging these signals does not improve accuracy, as shown in **TABLE 5**. This aligns with prior studies (Guo et al., 2021; Marinov et al., 2020; Zhao et al., 2017), which primarily use barometer data for height measurements and, in addition, require additional sensors to correct offset drift due to atmospheric changes. Moreover, **FIGURE 6** illustrates that barometer signals cannot detect door movements, a critical operational state, further limiting their utility. Incorporating barometer data also adds computational complexity, whereas our 1D-CNN, using only acceleration data, achieves superior results with simpler and more efficient implementation.

Despite the satisfactory performance, the discussion of the experimental results must also mention limitations. On the one hand, the experiments were conducted on a single elevator, and therefore, generalization of the model performance cannot be claimed. The model's performance largely depends on the training data, and while the current dataset provides useful insights, a larger and more diverse dataset would be necessary to improve the model's robustness and accuracy across a broader range of elevator types and usage conditions.

On the other hand, it should be noted that some of the deviations between the detected operational states and the labeled ones are not exclusively due to inaccuracies in the detection of the operational states, but manual labeling of the operational states has not been as accurate as possible by the expert in charge of the data preparation.

Finally, it is worth noting that the 1D-CNN model developed in this study offers greater flexibility in the convolutional layer compared to models like WaveletKernelNet (WKN) proposed by (Li et al., 2022). The WKN model, developed for predicting rapid decaying oscillations induced by impact-shaped vibration, employs predefined wavelet bases and requires only two parameters (s and u) to be tuned. In contrast, our 1D-CNN model does not restrict kernel training to predefined wavelet forms, allowing each kernel to be fully tuned during the training process. This flexibility enables the model to adapt to both long- and short-lasting patterns induced by elevator journeys, such as those related to vertical displacement and door movement, providing a more adaptable solution for operational state detection.

## 7 | CONCLUSIONS

This research has proposed an accurate and efficient monitoring system based on a 1D-CNN that is able to detect the operational states of an elevator with an accuracy of 96%, which improves the next most accurate algorithm, the GLRT with barometer fusion, by 18% in terms of average F1 score. The interpretability analysis provided in this paper allows to easily understand the feature extraction process carried out by the convolutional layer of the 1D-CNN, providing practitioners a meaningful way that allows to trace the feature extraction results with actual physical phenomena developing

within the elevator system.

Regarding further improvements of the proposed monitoring framework, several key areas need to be addressed to enhance the robustness and applicability of the proposed methodology.

First, while the 1D-CNN model has been validated on a traction elevator under normal operating conditions, it has yet to be tested on other types of elevators, such as hydraulic or guideless systems, where differing dynamics may significantly affect the acceleration signals. Additionally, experiments under varied operational conditions, such as heavier loads or emergency scenarios, are needed to further validate the model's adaptability and scalability.

Second, the model's long-term stability and the potential need for periodic retraining should be explored as more data becomes available over extended periods. This investigation would be critical to ensuring the robustness of the monitoring system against environmental changes and gradual wear and tear of elevator components.

Third, future research could explore the integration of unsupervised classification strategies to eliminate the need for manual labeling of operational states. By leveraging clustering algorithms or autoencoders, the monitoring framework could autonomously identify patterns in the acceleration signals and classify operational states without requiring expert intervention. This approach would not only reduce the time and effort involved in dataset preparation but also enhance the scalability of the framework for deployment in diverse elevator systems and operational conditions.

Fourth, the compact size of the acceleration blocks (150 samples) used, and the simplicity of the developed 1D-CNN model open opportunities for conceiving a real-time monitoring device. Such a device could be applied beyond operational state monitoring to include applications like anomaly detection, implementing reactive actions, improving elevator journey comfort through motor control, and supporting decision-making strategies for safe and maintenance mode operations. Moreover, the design process of such monitoring solutions must take into account the critical safety issues that must be satisfied when manufacturing the hardware and software that control systems involved in human transportation.

Lastly, the current framework could be extended to address damage scenarios in elevator systems. This would require developing methods to detect and monitor early signs of wear and tear, such as subtle changes in acceleration patterns or anomalies within specific operational states, which could serve as precursors to failure. Interpretability of the model would be crucial for understanding the physical phenomena behind the detected anomalies, building trust in the predictions, and supporting actionable maintenance decisions.

## ACKNOWLEDGMENTS

This research work has been funded by the Education Department of the Basque Government through the Research Groups Support call (IT1451-22). J.I. Aizpurua acknowledges financial support from the Spanish State Research Agency (grant No. IJC2019-039183-I).

## References

Adak, M. F., Duru, N., & Duru, H. T. (2013). Elevator simulator design and estimating energy consumption of an elevator system. *Energy and Buildings*, *65*, 272–280.

Adeli, H., & Jiang, X. (2006). Dynamic Fuzzy Wavelet Neural Network Model for Structural System Identification. *Journal of Structural Engineering*, *132*(1), 102–111. https://doi.org/https://doi.org/10.1061/(ASCE)0733-9445(2006)132:1(102

Adeli, H., & Jiang, X. (2008). *Intelligent infrastructure: neural networks, wavelets, and chaos theory for intelligent transportation systems and smart structures*. CRC press.

Aizpurua, J. I., Catterson, V. M., Stewart, B. G., McArthur, S. D. J., Lambert, B., & Cross, J. G. (2021). Uncertainty-Aware Fusion of Probabilistic Classifiers for Improved Transformer Diagnostics. *IEEE Trans. Systems, Man, and Cybernetics: Systems*, *51*(1), 621–633. https://doi.org/10.1109/TSMC.2018.2880930

Awatramani, J., Verma, G., Hasteer, N., & Sindhwani, R. (2022). Investigating strategies and parameters to predict maintenance of an elevator system. *Computational and Experimental Methods in Mechanical Engineering: Proceedings of ICCEMME 2021*, 323–332.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*, 1–13.

Christ, M., Braun, N., Neuffer, J., & Kempa-Liehr, A. W. (2018). Time series feature extraction on basis of scalable hypothesis tests (tsfresh–a python package). *Neurocomputing*, *307*, 72–77.

Daniya, T., Geetha, M., & Kumar, K. S. (2020). Classification and regression trees with gini index. *Advances in Mathematics: Scientific Journal*, *9*(10), 8237–8247.

Fernandez, J. R., & Cortes, P. (2015). A Survey of Elevator Group Control Systems for Vertical Transportation: A Look at Recent Literature. *IEEE Control Systems Magazine*, *35*(4), 38–55. https://doi.org/10.1109/MCS.2015.2427045

Garro, U., Muxika, E., Aizpurua, J. I., & Mendicute, M. (2020). FPGA-Based Stochastic Activity Networks for Online Reliability Monitoring. *IEEE Transactions on Industrial Electronics*, *67*(6), 5000–5011. https://doi.org/10.1109/TIE.2019.2928244

Gu, D., Zhang, N., Xu, Z., Wu, Y., & Tian, Y. (2024). Urban risk assessment model to quantify earthquake-induced elevator passenger entrapment with population heatmap. *Computer-Aided Civil and Infrastructure Engineering*, *39*(14), 2204–2222. https://doi.org/10.1111/mice.13287

Guo, Y., Liu, Y., Zhang, X., & Wang, G. (2021). The Real-Time Elevator Monitoring System Based on Multi-sensor Fusion. *Journal of Physics: Conference Series*, *2010*(1),

12182. https://doi.org/10.1088/1742-6596/2010/1/012182

InvenSense, T. D. K. (2020). DK-20789 data adquisition board. In *TDK InvenSense*. TDK InvenSense. https://invensense.tdk.com/products/dk-20789/

Jiang, X., & Adeli, H. (2005). Dynamic wavelet neural network for nonlinear identification of highrise buildings. *Computer-Aided Civil and Infrastructure Engineering*, *20*(5), 316–330. https://doi.org/https://doi.org/10.1111/j.1467-8667.2005.00399.x

Jiang, X.-Y., Huang, X.-C., Huang, J.-P., & Tong, Y.-F. (2022). Real-Time intelligent Elevator Monitoring and Diagnosis: Case Studies and Solutions with applications using Artificial Intelligence. *Computers and Electrical Engineering*, *100*, 107965. https://doi.org/10.1016/j.compeleceng.2022.107965

Kay, S. M. (1993). *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc.

Khanna, N. P. (2020). Urbanization and Urban Growth: Sustainable Cities for Safeguarding Our Future. In A. and B. L. and G. Ö. P. and W. T. Leal Filho Walter and Marisa Azul (Ed.), *Sustainable Cities and Communities* (pp. 953–965). Springer Int. https://doi.org/10.1007/978-3-319-95717-3_51

Kone. (2024). *Kone q3 2024*. https://www.kone.com/en/investors/reports-and-presentations/financial-reports/

Kramer, O., & Kramer, O. (2016). Scikit-learn. *Machine Learning for Evolution Strategies*, 45–53.

Li, T., Sun, C., Fink, O., Yang, Y., Chen, X., & Yan, R. (2023). Filter-Informed Spectral Graph Wavelet Networks for Multiscale Feature Extraction and Intelligent Fault Diagnosis. *IEEE Trans. Cybernetics*, 1–13. https://doi.org/10.1109/TCYB.2023.3256080

Li, T., Zhao, Z., Sun, C., Cheng, L., Chen, X., Yan, R., & Gao, R. X. (2022). WaveletKernelNet: An Interpretable Deep Neural Network for Industrial Intelligent Diagnosis. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *52*(4), 2302–2312. https://doi.org/10.1109/TSMC.2020.3048950

Liono, J., Abdallah, Z. S., Qin, A. K., & Salim, F. D. (2018). *Inferring Transportation Mode and Human Activity from Mobile Sensing in Daily Life*. 342–351. https://doi.org/10.1145/3286978.3287006

Mallat, S. (2016). Understanding deep convolutional networks. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *374*(2065), 20150203.

Marinov, M. B., Nikolov, D. N., Ganev, B. T., & Djamiykov, T. S. (2020). Smart Multisensor Node for Remote Elevator Condition Monitoring. *2020 21st International Symposium on Electrical Apparatus & Technologies (SIELA)*, 1–4. https://doi.org/10.1109/SIELA49118.2020.9167049

Michau, G., Frusque, G., & Fink, O. (2022). Fully learnable deep wavelet transform for unsupervised monitoring of high-frequency time series. *Proceedings of the National Academy of Sciences*, *119*(8), e2106598119. https://doi.org/10.1073/pnas.2106598119

Mishra, K. M., & Huhtala, K. (2019a). Elevator Fault Detection Using Profile Extraction and Deep Autoencoder Feature Extraction for Acceleration and Magnetic Signals. *Applied Sciences*, *9*(15). https://doi.org/10.3390/app9152990

Mishra, K. M., & Huhtala, K. J. (2019b). Fault Detection of Elevator Systems Using Multilayer Perceptron Neural Network. *IEEE Int. Conf. Emerging Tech. and Factory Automation*, 904–909. https://doi.org/10.1109/ETFA.2019.8869230

Nguyen, H. D., Tran, K. P., Zeng, X., Koehl, L., & Tartare, G. (2019). Wearable sensor data based human activity recognition using machine learning: a new approach. *ArXiv Preprint ArXiv:1905.03809*.

Nikolov, D. N., Marinov, M. B., Ganev, B. T., & Djamijkov, T. S. (2020). Nonintrusive Measurement of Elevator Velocity Based on Inertial and Barometric Sensors in Autonomous Node. *2020 43rd International Spring Seminar on Electronics Technology (ISSE)*, 1–5. https://doi.org/10.1109/ISSE49702.2020.9121077

of German Engineers, A. (2007). *VDI guideline Lift Energy efficiency*. http://info.wsisiz.edu.pl/~roksela/dzwigi/Energia/Energy_VDI_ENG.PDF

Olaizola, J. (2023). *Traction elevator dataset*. https://www.dropbox.com/scl/fo/hwi4k7zutmuekg6luovk8/ANBY7Cpgj-m8riKCECXYOws?rlkey=4bbo2q7iqxvdt0y0n2op69sop&st=ohc5v4y1&dl=0

Otis. (2024). *Otis Earnings Call Forward-Looking Statements q32024*. https://otisinvestors.com/events-and-presentations/default.aspx

Shabbir, K., Noureldin, M., & Sim, S. H. (2024). Data-driven model for seismic assessment, design, and retrofit of structures using explainable artificial intelligence. *Computer-Aided Civil and Infrastructure Engineering*, *August*, 1–20. https://doi.org/10.1111/mice.13338

Sheng, Z., Huang, Z., & Chen, S. (2024). Ego-Planning-Guided Multi-Graph Convolutional Network for Heterogeneous Agent Trajectory PredictionNo Title. *Computer-Aided Civil and Infrastructure Engineering*, *39*(22), 3357–3374.

Simão, M. A., Neto, P., & Gibaru, O. (2017). Unsupervised Gesture Segmentation by Motion Detection of a Real-Time Data Stream. *IEEE Transactions on Industrial Informatics*, *13*(2), 473–481. https://doi.org/10.1109/TII.2016.2613683

Skog, I., Handel, P., Nilsson, J.-O., & Rantakokko, J. (2010). Zero-velocity detection—An algorithm evaluation. *IEEE Transactions on Biomedical Engineering*, *57*(11), 2657–2666.

Skog, I., Karagiannis, I., Bergsten, A. B., Harden, J., Gustafsson, L., & Handel, P. (2017). A Smart Sensor Node for the Internet-of-Elevators - Non-Invasive Condition and Fault Monitoring. *IEEE Sensors Journal*, *17*(16), 5198–5208. https://doi.org/10.1109/JSEN.2017.2719630

Taillandier, F., Fernandez, C., & Ndiaye, A. (2017). Real Estate Property Maintenance Optimization Based on Multiobjective Multidimensional Knapsack Problem. *Computer-Aided Civil and Infrastructure Engineering*, *32*(3), 227–251. https://doi.org/https://doi.org/10.1111/mice.12246

Veltink, P. H., Bussmann, H. J., De Vries, W., Martens, W. J., & Van Lummel, R. C. (1996). Detection of static and dynamic activities using uniaxial accelerometers. *IEEE Transactions on Rehabilitation Engineering*, *4*(4), 375–385.

Zhang, Y., Miyamori, Y., Mikami, S., & Saito, T. (2019). Vibration-based structural state identification by a 1-dimensional convolutional neural network. *Computer-Aided Civil and Infrastructure Engineering*, *34*(9), 822–839. https://doi.org/https://doi.org/10.1111/mice.12447

Zhao, F., Luo, H., Zhao, X., Pang, Z., & Park, H. (2017). HYFI: Hybrid Floor Identification Based on Wireless Fingerprinting and Barometric Pressure. *IEEE Trans. Industrial Informatics*, *13*(1), 330–341. https://doi.org/10.1109/TII.2015.2491264