مجلة الدراسات التربوية والعلمية - كلية التربية - الجامعة العراقية
العدد السابع والعشرون - المجلد الحادي عشر - علوم الفيزياء - آيار 2026 م
doi.org/10.52866/esj.2026.11.27.15

Vehicle Classification by Audio Sound Signals Using a Scalogram

Ahlam M. Kadhim
Department of Physics, College of Science, Mustansiriyah University
Corresponding author: ahlammajead@uomustansiriyah.edu.iq

Abstract :

The rapid growth in the diversity and number of vehicles has produced increasing environmental pollution, creating a serious threat to human safety and health. These circumstances highlight the need for intelligent automated systems capable of accurately detecting vehicle types to aid traffic management, improve public safety, and decrease congestion. This work introduces an effective model classification built on Continuous Wavelet Transform CWT scalogram representations combined with a Convolutional Neural Network CNN. The enhanced time–frequency visual scalogram effectively describes non-stationary signals characterized by time and spectral changes, authorizing suitable deep feature extraction. A global sound dataset includes eight categories, and five of the most commonly used on the roads were selected for training (each class 250 signals, except cars, the data contains only 229 audio files). The other three, aircraft types, like a helicopter, were excluded. The data was divided 80 % training and 20 % testing. Image processing methods, like normalization, logarithmic transformation, and denoising, were applied to improve feature classification. The CNN model performs strongly and works with a balanced achievement, 97.31 % precision, 97.12% specificity, and an of 97.22 % F1-score. Class-wise analysis showed excellent performance across all vehicle types; buses achieved optimal results with 100 %, while motorcycles and trucks had near-perfect classification. The combination of enhanced CWT scalograms with CNN provides an accurate system for intelligent traffic monitoring and management.

Keywords: Vehicle classification, 1D Audio Signals, Enhanced Visual CWT, Deep Learning.

تصنيف المركبات باستخدام الإشارات الصوتية باستخدام مخطط سكالوكرام

د.احلام مجيد كاظم // الجامعة المستنصرية\ كلية العلوم \ قسم الفيزياء

مستخلص:

أدى النمو السريع في تنوع المركبات وعددها إلى تفاقم التلوث البيئي، مما يشكل تهديدًا خطيرًا لسلامة الإنسان وصحته. تُبرز هذه الظروف الحاجة إلى أنظمة آلية ذكية قادرة على الكشف بدقة عن أنواع المركبات للمساعدة في إدارة حركة المرور، وتحسين السلامة العامة، والحد من الازدحام. يقدم هذا العمل نموذج تصنيف فعالًا مبنيًا على تمثيلات المقياس الطيفي باستخدام تحويل المويجات المستمر (CWT) مع شبكة عصبية تلافيفية (CNN). يصف المقياس الطيفي المرئي المحسن للتردد الزمني الإشارات غير المستقرة التي تتميز بتغيرات زمنية وطيفية، مما يسمح باستخلاص ميزات عميقة مناسبة. تم استخدام مجموعة بيانات صوتية عالمية تضم ثماني فئات، وتم اختيار خمس فئات شائعة الاستخدام في الشوارع (250 إشارة لكل فئة عدا السيارات يوجد 229 تسجيل صوتي فقط)، وتم استبعاد الفئات الثلاث الاخرى مثل الهليكوبتر. قُسّمت البيانات بنسبة 80٪ للتدريب و20٪ للاختبار. طُبقت أساليب معالجة الصور، مثل التطبيع والتحويل اللوغاريتمي وإزالة التشويش، لتحسين تصنيف الميزات. يُظهر نموذج الشبكة العصبية التلافيفية أداءً قويًا ونتائج متوازنة، حيث بلغت الدقة 97.31٪، والنوعية 97.12٪، ودرجة F1 97.22٪. أظهر التحليل حسب الفئات أداءً ممتازًا لجميع أنواع المركبات؛ حيث حققت الحافلات نتائج مثالية بنسبة 100٪، بينما حظيت الدراجات النارية والشاحنات بتصنيف شبه مثالي. يوفر دمج مخططات CWT المحسّنة مع الشبكات العصبية التلافيفية نظامًا دقيقًا لرصد وإدارة حركة المرور بذكاء.

Introduction

In global society now, many problems are causing environmental changes and a significant increase in contributing to global warming and climate change [1]. The greenhouse gas emissions from vehicles during personal travel are a major problem [2]. Nowadays, the traffic scenario needs a reapproach to the vehicle classification technology. The recognition system's capacity to identify vehicle types enables the system to better manage the traffic, thereby decreasing road congestion and enhancing public safety for pedestrians. Classifying vehicles by type is a crucial and recent area of study in this sector for two main reasons. The primary reason is that the classification of vehicles allows experts to stop global warming, identify the sources, and assess and develop strategies to reduce pollution [3]. This helps law enforcement identify vehicles that violate environmental laws and supports the worldwide goal of mitigating environmental changes [4]. By guiding all types of vehicles to the right routes, vehicle identification aids in enhancing traffic flow and lowering total gas emissions in metropolitan centers [5]. By categorizing vehicle categories, cities can better design and improve a foundation that promotes sustainable transportation, even to the point of dedicated lanes for cars and bicycles in a smart city's systems [6,7]. If city traffic is only influenced by clean vehicles, such methods can identify non-compliant vehicles in low-emission areas. Additionally, the system's accuracy in recognizing vehicles improves traffic management in the smart city and reduces traffic congestion and hazardous gas emissions [8]. Numerous algorithms have been proposed for vehicle classification utilizing many data signals. The one effective framework for vehicle recognition using sound_signals, since each vehicle type has a distinctive sound [9, 10, 11]. Deep learning algorithms and machine learning techniques are the most used methodologies [4, 12]. One of the best techniques is the CNN [13, 14]. Time-frequency analysis is essential for precise feature extraction because the signals are naturally non-stationary. The Scalogram, which originat-

ed from the CWT, has been receiving more attention lately due to its superior performance among other techniques in the time–frequency domain. Recently, developments in deep learning, especially CNNs, have further enhanced vehicle audio recognition by allowing automated feature learning from 2D visual representations of CWT.

Review Related Studies

Improvements in deep learning methods, especially CNNs, have enhanced vehicle audio classification and enabled the automated learning of features from 2D or image representations like scalograms by CWT. This section shows some significant studies that use audio signals and a scalogram-located representation to classify vehicle categories. The earlier works, like Dalir et al. (2018), introduced a quadratic discriminant analysis approach to classify a vehicle type based on sound signals, achieving a mid-rate accuracy of ~ 90 % over a specific number of categories. The researchers extracted the features in the frequency and time domains, such as dominant frequency components, zero-crossing rate, and short-time energy. Although it's effective for correct classification, such approaches are constrained by their reliance on manually extracted features and lower-class diversity [15]. These approaches are still dependent on manual feature engineering and limited to four vehicle sound signal classes, which may restrict their scalability and adoption. The authors are working on vehicle sound audio transmitted via vehicle engine features linked to artificial neural networks to categorize vehicle types, including vans, cars, motorcycles, and trucks. Ashhad et al.'s approach performed categorization accuracies exceeding 90 [16]. The Scalogram, produce applying the CWT, specifies a time-frequency representation that adapts to both non-stationary transient components and low-frequency stationary components. These possessions create scalograms, especially acceptable for vehicle visual and audio entertainment transmitted signals. Scarpiniti et al. (2023) demonstrated that the scalogram-based likenesses significantly enhance categorization performance in audio recognition tasks, although they are used as a recommendation to

CNN models. Their judgments suggest that scalograms exhibit discriminatory temporal patterns that are often absent in typical spectrogram images [17]. Today, studies have met on combining scalogram likenesses accompanying deep learning architectures. Shabber, and others. (2024) suggest a fusion of spectrogram and scalogram analyses to use as a deep CNN VGG16 input layer to achieve an accuracy reach 96.8 % of four vehicle classes. The output projected a deep CNN scheme prepared on two together scalogram plus spectrogram fusion, showing that the mixed likeness advances classification veracity by occupying complementary time–frequency facts[18] .

New studies have joined the scalogram likenesses accompanying deep learning architectures. Despite the fusion technique improving discriminative ability, it causes an increase in computational and model complexity. In contrast, the results of the suggested algorithm support a single, strong time–frequency representation depending on the CWT to create a scalogram image, which is directly trained using a CNN without fusion mechanisms or additional features.

More recent studies, like Ashhad et al. [16], and Saleem et al. [19], employed modern handcrafted representations like LT-MFCC beside neural networks and the SVM optimized classifiers, achieving improved accuracies 91.98 % to 96.66 % ranging. However, these approaches are still dependent on manual feature engineering and limited to four vehicle sound signal classes, which may restrict their scalability and adoption. Shariff and so forth. discharged a CNN blueprint prepared on two composed scalograms with spectrogram merging, displaying the assorted reproduction advances categorization veracity by possessing completing time–frequency.

This work introduces a CNN–based framework for vehicle sound signal classification using audio wav files transformed into visual representations using a scalogram technique. Five vehicle categories are considered: buses, trucks, motorcycles, cars, and trains. The proposed approach achieves high accuracy across five vehicle categories using only enhanced CWT representations of the scalogram with a

CNN, demonstrating robust discriminative capability. Unlike prior works, it emphasizes scalability and efficient data representation, maintaining high performance accuracy while handling more classes and reducing pre-processing complexity.

Continuous Wavelet Transform (CWT)

The CWT technique is widely used because it evaluates how closely a sound signal matches an analysis wavelet function at different positions. This enables the detection of important characteristics in the signal over various frequency bands, supporting informative representations of the underlying signal characteristics. Scalograms were chosen as the basic time-frequency representation because of their significant advantages over traditional transformations in analyzing complex sound signals. CWT supports accurate time–frequency localization and efficient extraction of discriminative temporal features for each sound category [20, 21]. Scalogram multi-resolution analysis allows simultaneous description of high- and low-frequency components, which is demanded to detect random fast oscillations associated with sound patterns of various sounds. Scalograms also exhibit robustness to noise of real-world audio, while their intuitive visual CWT representation facilitates efficient CNN learning. Thus, CWT-based visual scalogram representations can be efficiently utilized as inputs to CNN layers, forming a strong approach for a correct and interpretable vehicle audio classification system [22].

The CWT plus CNN interconnected framework provides an efficient learnable technique for a classification system, enabling evaluators to delve deeper into the features of this dataset and improve direct learning of CNN by updating the layer's parameters. The sound waveform is depicted as a set of wavelet coefficients, thereby establishing automatic feature extraction [21]. Mathematically, the analysis is performed using the scalogram technique, which is a created visual representation of CWT. Given a stationary signal x(t), the CWT is defined as the product of x(t) with the following basis function family·

$$\Psi_{\tau,a}(t) = |a|^{-\frac{1}{2}}\Psi\left(\frac{t-\tau}{a}\right) \qquad (1)$$

Where: a≠0 is a scaling or dilation parameter, and τ is the time delay, i.e., $\Psi_{\tau,\,a}(t)$ is a scaled and translated version of the mother wavelet function Ψ(t). Hence, the CWT of signal x(t) is constructed as:

$$\Psi_x(\tau,a) = |a|^{-\frac{1}{2}} \int_{-\infty}^{\infty} x(t)\, \Psi^*\left(\frac{t-\tau}{a}\right) dt \qquad (2)$$

The expression (*) indicates the complex wave formula. The time-shift factor τ determines the temporal location of the wavelet function Ψ (τ, a), while the scale factor a orders its frequency components. When the factor |a|≪1, the wavelet function value represents a highly compressed and localized version of the mother wavelet Ψ(t), corresponding to high-frequency components with fine temporal resolution. Conversely, when |a|≫1, the wavelet becomes dilated in time, emphasizing low-frequency components with coarser temporal resolution. Scaling by a factor of $\frac{1}{a}$ enables proportional expansion of the wavelet, providing adaptive resolution across different frequency bands.

The similarity wavelet analysis between the wavelet $\Psi_{\tau,a}(t)$, and the signal x(t) is weighted as a and τ change. Dilation by a factor $\frac{1}{a}$ results in different signal expansions with distinct clarity. The features of the CWT time–frequency representations are outlined as follows. The Δτ varies inversely with the wavelet carrier frequency ω_0 of $\Psi_{\tau,a}(t)$; it can be made arbitrarily small for high frequencies. The frequency resolution Δω varies linearly with the wavelet carrier frequency ω0 of $\Psi_{\tau,a}(t)$; at low frequencies, it can be made arbitrarily small. Consequently, the CWT is suitable for the non-stationary signals analysis that contains high-frequency transients overlying low-frequency components, long-lasting [23].

So, the CWT is well-suited for analyzing non-stationary signals that contain both high-frequency and long-duration low-frequency components, such as audio [17, 23]. The squared magnitude of the CWT coefficients, known as the scalogram, represents the energy distribution of the signal in the time–frequency domain and is defined as:

$$S(\tau,a) \triangleq \Psi_x(\tau,a) = |W_x(\tau,a|^2 = \frac{1}{a}\left|\int_{-\infty}^{\infty} x(t)\, \Psi^*\left(\frac{t-\tau}{a}\right) dt\right|^2 \qquad (3)$$

Given scale a and time shift τ, the mother wavelet function Ψ(t) is typically a band function [24]. The Morlet wavelet, consisting of a complex sinusoid modulated by a Gaussian window, is particularly suitable for sound signal analysis due to its strong correspondence with human auditory perception. Owing to its proven effectiveness, representation, and sound classification [25, 26].

Methodology

The Vehicle Sounds dataset was acquired from Kaggle by the link[26] : https://www.kaggle.com/datasets/jan-boubiabderrahim/vehicle-sounds-dataset. The data contains 9448 audio recordings, which describe vehicle sound excerpts across eight diverse categories. This dataset has been applied in many recent research studies of sound-located vehicle classification systems. All audio recordings have a fixed duration of 5 seconds [28]. Audio signals were compressed to two channels. Each channel was stored separately as a numerical time series; they are saved in MATLAB. mat file. The program saved each of the two channels, with its corresponding time, in a separate file. According to eight vehicle categories, divided into class-specific folders, each folder contains the corresponding audio files. There are 250 audio recordings selected from each of the classes (Motorcycles, Trucks, Trains, Buses), while the data contains 229 audio recordings for the cars category. Using the proposed algorithm 1 and the preprocessing technique, each audio signal was transformed into a time–frequency visual representation. Table 1 displays the details of the rate splitting dataset (80% training, 20 testing). All the signals' categories create 1229 images for the experiments. The enhanced CWT scalogram enabled the extraction of discriminative spectral patterns while increasing the representational capacity of the dataset number for deep learning–based classification. Figure 1 illustrates spectral images of each vehicle category commonly found on the global roads.

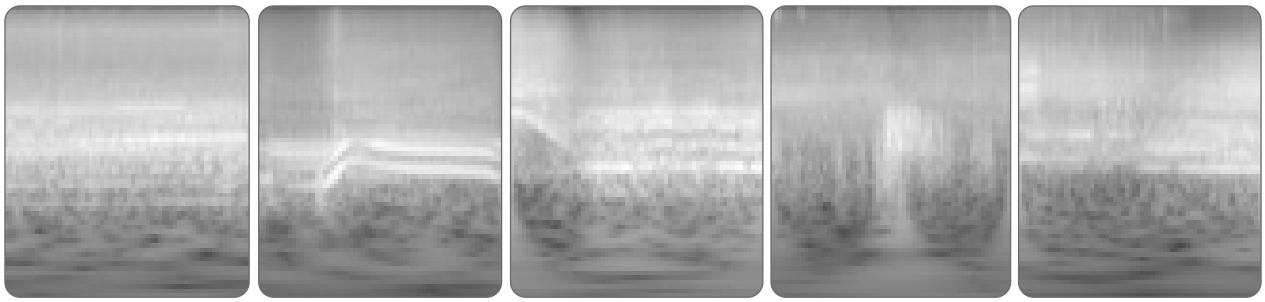

Figure 1: Spectral images for each visual audio vehicle's sounds.

Table 1. Dataset composition, rate 80 % for training and 20 % testing.

Class Name	Total Audio Signals	Audio Signals Selected	Scalogram Images	Training (80%)	Testing (20%)
Motorcycles	537	250	250	200	50
Trucks	265	250	250	200	50
Trains	2552	250	250	200	50
Buses	4221	250	250	200	50
Cars	229	229	229	183	46
Total	**12,638**	**1,229**	**1,229**	**983**	**246**

Figure 2 shows the general flowchart of this study. Each sound signal channel is extracted from audio recordings and then averaged. The sound signals were first pre-processed by removing the zero-centering signal (DC offset) through the subtraction of the signal mean, producing the zero-centered. The scalogram was created as the squared value for the CWT officiates presenting the signal sound energy in the time–frequency domain. To further improve feature clarity, a logarithmic transformation $\log_{10}(S+\varepsilon)$ was applied to enhance the visibility of low-energy segments within the CWT visual representation and allowed subtle yet spatial spectral features to be protected. Compressing the dynamic range, it prevents the model from being dominated by high-energy components, thereby reducing saturation effects during learn-

ing. This dynamic range compression also improves numerical stability and facilitates smoother gradient updates throughout training. Several scalogram energy values CWT may be equal to 0, where applying a logarithmic transformation directly results in $\log(0)=-\infty$, produce undefined values without altering the actual signal structure, and failure potential training. To avoid this issue, a small constant (ε) is added before the logarithmic operation.

The value of ε of precision, MATLAB is approximately 2.22×10^{-16}, which is the smallest positive value in double-precision floating-point representation.

As a result, the dynamic range focuses attention on energy components. Then the min–max normalization rescaled the scalogram values to the [0, 1] range, ensuring coordinated data scaling across samples, minimizing variability, and supporting stronger feature learning.

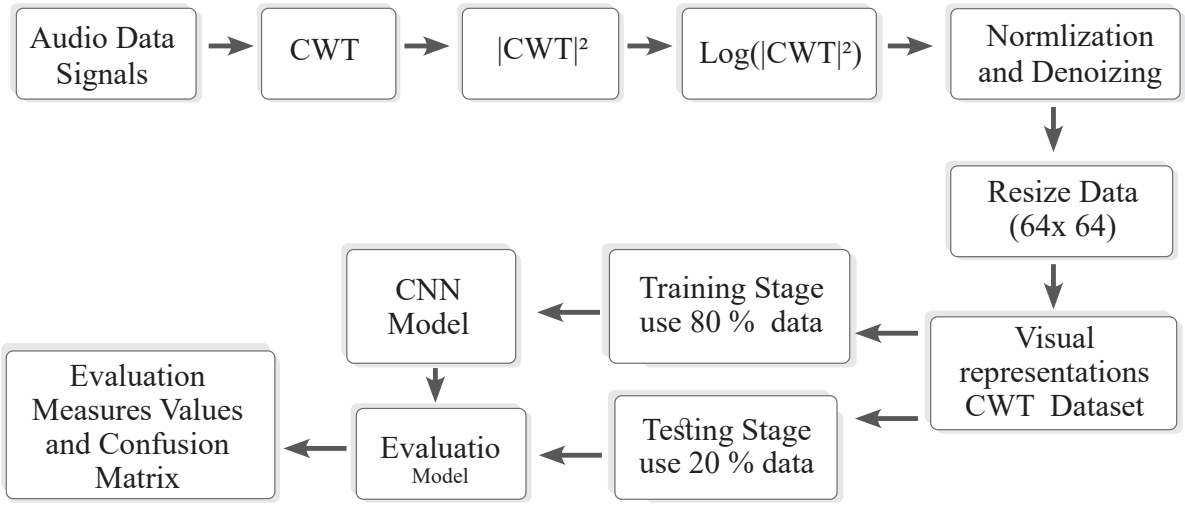

Figure 2. Workflow of the CNN proposed model based on the enhanced scalogram technique.

Algorithm 1: CWT Enhanced Representations

The proposed method vehicle sound audio, applies resampling, transforms the signals into time-frequency domain scalograms using CWT, normalizes and resizes the resulting representations, and then stores them in a

class folder located in the main folder. The scalogram images were resized to 64×64 pixels and saved in their original class directories to maintain the dataset structure. The output is a visual representation or a dataset image, which is used as an input to the CNN layers to train a classification model.

Input:

The main folder that contains the vehicle classes dataset, which is saved in audio format.

The interest Sample frequency f_s=44100 Hz

Determine image scalogram size 64x64.

Output: Mane Folder of Visual (images) dataset, divided into category folders.

Start:

1. Select the main folder that contains all category folders.

2. Create output folders to save the CWT visual dataset.

3. Read the subfolders that are named as the vehicle categories.

4. Select the folder path of each folder class.

5. Create an output folder path for each category within the main output folder.

6. Download sound files in wav format that are related to each category.

7. Read the sound signal and original sample frequency.

8. Resampling the sample frequency using the input frequency 44100 Hz.

9. Select one audio channel if there are two channels.

10. Apply the CWT of the channel signal.

11. The signal mean was subtracted to remove the DC component, producing zero-centered signals and enhancing the accuracy of time–frequency representations and feature extraction.

12. Calculate the spectral energy scalogram using the absolute value square for the CWT coefficients.

13. Apply a logarithmic transformation to reduce contrast and enhance key spectral patterns.

14. Enhancement and Normalized Scalogram.

15. Normalize the scalogram values to the [0, 1] domain to ensure training stability.

16. Resize the scalogram images to 64x64 to meet the network requirements.

17. Store the generated Image in PNG format.

18. Organize the images in subfolders according to name category.

19. Each image is automatically named based on the folder name.

End

The output of this algorithm is images unified names according to their category; the resulting images are ready to use in the training and testing stages of a CNN model.

Algorithm 2: Training Model

The training process was carried out using a MATLAB program code. The following steps explain in detail the training algorithm.

Inputs: Folders for the dataset image classes.

Outputs: The CNN Model and values for the accuracy evaluation measures.

Starting:

1. Define the directory path where the data is stored.

2. Download the data and prepare it for the training process.

3. Specify the width and height of the images.

4. Specify the number of categories for training the model.

5. Enter the class names that were entered into the training process.

6. Divide the data into two parts for training 80 % and for testing, 20 %.

7. Constructing the foundational CNN structure of multiple layers.

8. Set the Adem function as a classifier with a 0.0001 learning rate, 25 batch size, and 10 epochs.

9. Starting the training process.

10. Save the MATLAB matrix for the trained model.

11. Using a testing image set to evaluate the achievement of a trained model.

12. Calculate the performance metrics for the trained model and generate a Confusion Matrix.

13. In an Excel worksheet, save the evaluation metric results.

Ending

Training Progress and Loss Function

The exploratory results indicate that the proposed CWT combined with the CNN technique achieves high efficiency in vehicle sound signal classifica-

tion. The overall accuracy of 97.19 indicates the robustness and ability of the suggested approach, where the training curve in Figure 3 demonstrates stability and effective learning behavior. Clear from the figure that the training accuracy increases in the early epochs and then converges toward a high plateau; this behavior confirms the successful deep learning in the extraction of visual representation features. Simultaneously, the training loss function shows a smooth value decrease without rapid changes, which is a good convergence property and stable improvement. Overall, the training curves validate the strong and reliable performance of the proposed learning algorithm.

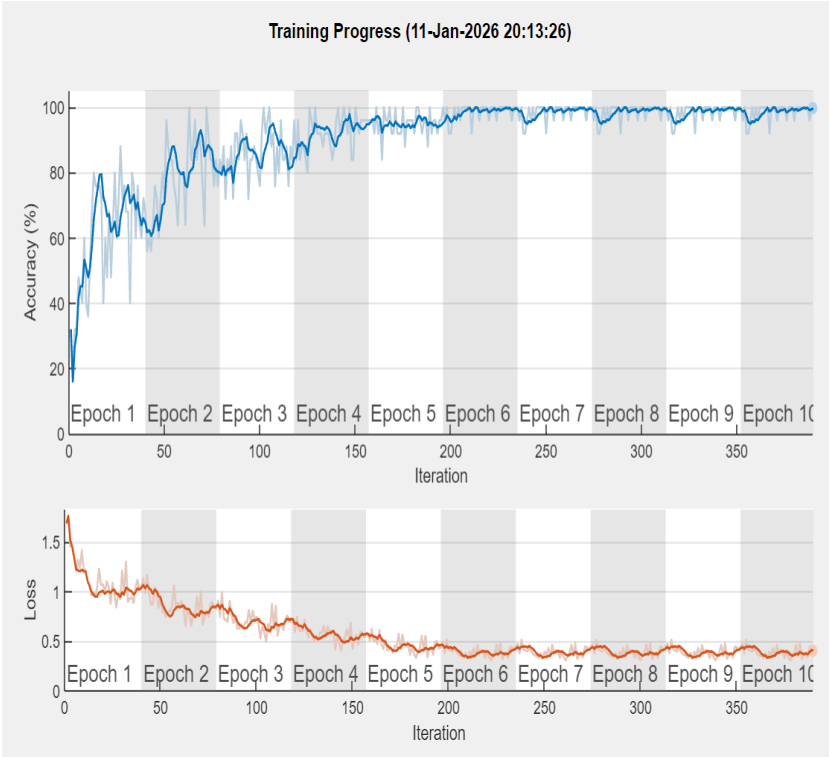

Figure 3. CNN model Training curves
of applying the scalogram-located vehicle sounds

Evaluation Measures

The measures in this work were calculated to calculate the difference between actual labels and predicted labels using a classification model. Four basic measures: accuracy, F1-score, precision, and recall. The accuracy measure is the proportion of correctly

classified cases of all scalogram images and is obtained by equation 5. The Precision measure evaluates the model's capability to correctly predict relative to all predictions made for that vehicle class and is calculated by equation 6. The recall assesses the model's ability to classify all samples that are truly properly to a given vehicle sound class and is obtained by equation 7. The f1-score is a harmonic of recall and precision, providing a stable measure of the model's performance, and is calculated using equation 8 [27, 28].

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive+False Positive}} \quad (5)$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive+False Negative}} \quad (6)$$

$$\text{F1_Score} = 2\frac{\text{Precision}\times\text{Recall}}{\text{Precision+Recall}} \quad (7)$$

$$\text{Accurcy} = \frac{\text{True Positive+True Negativ}}{\text{True Positive+False Positive+True Negativ+False Negativ}} \quad (8)$$

Results and Discussion

The results indicated that the suggested time-frequency dataset of a scalogram-based training CNN model is a stable classifier, achieving high specificity and sensitivity across all vehicle classes. The per-class precision and recall values are displayed in Table 2; there is consistently high accuracy for all the vehicle categories. For the buses class, the CNN model evaluation indicates that the ideal values for achieving perfect scores on the test set, demonstrating perfect separability of their acoustic pattern in visual CWT representation. Also, ideal accuracy performance for the trucks class highlights the effectiveness of the time–frequency representations in detecting distinctive acoustic properties. Similarly, trucks show a perfect balance between precision and recall (98.11%), reflecting the power of the extracted features, a scalogram that represents their acoustic patterns. The motorcycles category also shows robust and balanced evalu-

ation (96 %) precision and recall, while the trains category gets a high recall of (98 %), indicating active classification ability. Although the relatively low precision for the train category (92.45 %), the proposed acoustically similar vehicle sounds have irregular confusion. The Cars class achieves ideal precision, reaching to (100 %), supporting the no false-positive classifications. A slight reduction in recall (93.48 %) confirms the misclassification, likely due to the interface of acoustic features with other vehicle categories under specific operational conditions.

Table 2. Precision and Recall for each vehicle class.

class	**Precision (%)**	**Recall (%)**	Overall Precision **(%)**	Overall Specificity **(%)**	F1-score **(%)**
Motorcycles	96.00	96.00	**97.31**	**97.12**	**97.22**
Trains	92.45	98.00			
Trucks	98.11	98.11			
Buses	100.00	100.00			
Cars	100.00	93.48			

Table 1 summarizes the overall measure values. The model achieves a high overall precision value (97.31 %), specificity (97.12 %), and an F1-score (97.22 %). These results confirm that the suggested scalogram transformation with the CNN model keeps a well-balanced trade-off between specificity and sensitivity, ensuring reliable and strong classification achievement across varying vehicle sound categories.

Confusion Matrix

The confusion matrix shows robust diagonal dominance, indicating that the CNN model achieves high classification accuracy across all vehicle classes. Buses and cars show ideal classification achievement, with no noticeable misclassifications, reflecting highly distinctive acoustic temporal patterns and balanced time–frequency of their CWT representations' scalograms. Trucks also perform well, with

only the slightest confusion, indicating that the suggested CWT-based automatic feature extraction by CNN can well capture their spectral properties. Little misclassifications are noticeable for trains and motorcycles, as a result of similarities in their sound signatures, particularly under changing noise environment and speed conditions. Generally, the finite off-diagonal errors emphasize the strongest and most popular CNN model's ability, focusing attention on the effectiveness of combining CWT scalograms with CNN for vehicle sound classification. Generally, accuracy reaches 97.2 %, supporting the activity of the scalogram-located time-frequency similarity. Figure 3. Shows the confusion matrix of the CNN model for five-classes, vehicle audio sounds, utilizing the scalogram representations. The predominantly diagonal confusion matrix displayed in Table 3 confirms the robustness and accuracy of the suggested CNN classification model for all vehicle categories. Misclassification is slight and primarily happens among cars, trains, and motorcycles, which can be attributed to overlapping acoustic pattern features within specific operating conditions, like speed variations and the presence of environmental noise.

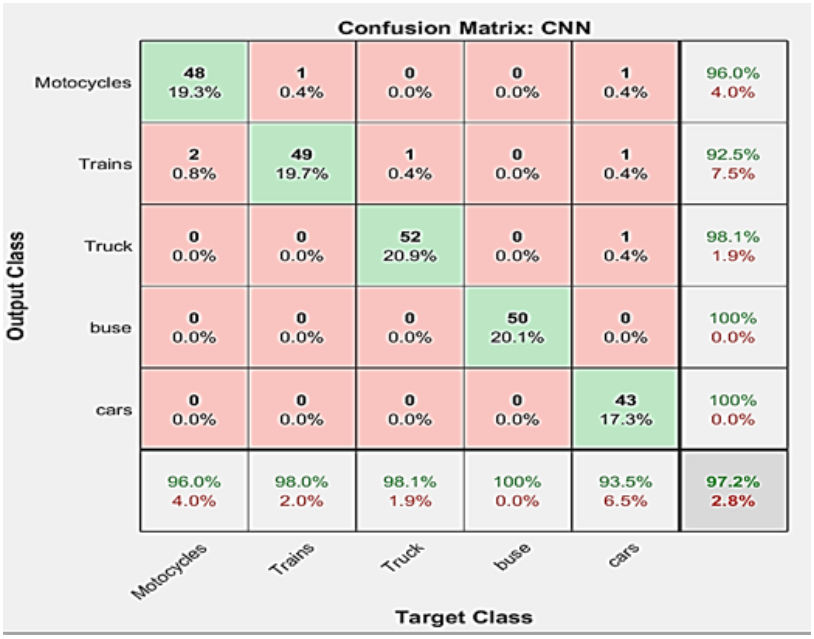

Figure 3. The confusion matrix of the CNN model classification.

Table 3. Confusion Matrix Analysis for Vehicle Audio Classification.

Vehicle Class	Correctly Classified (TP)	False Positives (FP)	False Negatives (FN)	Remarks
Motorcycles	48	2	2	Minor confusion with trains, cars
Trains	49	4	1	Slight overlap with motorcycles
Trucks	52	1	1	Highly distinctive acoustic patterns
Buses	50	0	0	Perfect classification (100%)
Cars	43	0	3	Some samples were confused with similar classes

The robustness-angled supremacy observed in the confusion origin shows that the CNN model successfully classifies sound patterns from the scalogram visual representations. Especially for the buses class, Precision and Recall reach 100 %, which can be attributed to their unique low-commonness power distribution and stable momentary architecture. Minor misclassifications were mainly noted between cars, motorcycles, and trains, that proper on account of partial flap in the sound signatures, especially under changes in speed and background buzz surrounding. Generally, the preformed extreme precision 97.12 % indicate that the model's powerful talent to correctly refuse non-aim categories is of critical importance for real-traffic world listening plans. Table 4 shows a comparison of the suggested approach with several study methods introduced in the literature review section.

Table 4. Comparison of the proposed algorithm with some other related works.

Study	Features / Representation	Classifier	No. of Classes	Accuracy (%)
Dalir et al. [15]	Time & frequency audio features	QDA	2–3	~90.0
Ashhad et al. [16]	Handcrafted audio features	Neural Network	4	91.98–96.66
Saleem, et al. [19]	LT-MFCC	SVM + PSO	4	94.5
Shabber, et al. [18]	Spectrogram + Scalogram fusion	CNN (VGG16)	4	96.8
Proposed Algorithm	Scalogram (CWT)	CNN	5	97.19

It is noted that the suggested approach achieves high classification accuracy, 97.19 %, with an increased number of classes, five vehicle types. Attention focuses on the scalability, efficiency, and computational effectiveness of the introduced approach, as well as its high capability to capture the discriminative acoustic patterns of diverse vehicle sound types.

Conclusion

The comparative analysis results using the other studies' approach further highlight the efficacy of the proposed CNN model, which demonstrates competitive performance while handling a larger number of vehicle categories, using a relatively pick-up feature-visualized enhanced scalogram representation. The training convergence analysis and confusion matrix emphasize the stability, reliability, and generalization capability of the proposed CNN model. The results validate the effectiveness of integrating a CWT-based scalogram with a CNN deep convolutional learning Model for vehicle sound signal classification. Future work will focus on extending the framework to greater and more varied datasets, collecting and recording them in real-world noise backgrounds, and using deep learning to hybridize the architectures to enhance scalability and robustness.

Acknowledgment

The author would like to express special thanks to the image processing laboratory staff in department of Physics, College of Science, Mustansiriya University.

References

1. Radziszewski, K., Anacka, H., & Weichbroth, P. (2021). Greencoin: A pro-environmental action-reward system.
2. Romero, Carlos Alberto, et al. "Strategies for reducing automobile fuel consumption." Applied Sciences 14.2 (2024): 910. https://doi.org/10.3390/app14020910.
3. Ahmed, Aser Alaa, et al. "Global warming potential, water footprint, and energy demand of shared autonomous electric vehicles incorporating circular economy practices." Sustainable Production and Consumption 36 (2023): 449-462.. https://doi.org/10.1016/j.spc.2023.02.001.
4. Luo, Yuxi, et al. "Sound-Convolutional Recurrent Neural Networks for Vehicle Classification Based on Vehicle Acoustic Signals." 2021 International Conference on Smart City and Green Energy (ICSCGE). IEEE, 2021.10.1109 / ICSCGE53744.2021.9654357.
5. Huang, Yi-Qi, et al. "Optimized YOLOv3 algorithm and its application in traffic flow detections." Applied Sciences 10.9 (2020): 3079.https://doi.org/10.3390/app10093079.
6. Mora, Luca, et al. "Smart city governance from an innovation management perspective: Theoretical framing, review of current practices, and future research agenda." Technovation 123 (2023): 102717. https://doi.org/10.1016/j.technovation.2023.102717.
7. Chen, Chen, et al. "An intelligent platooning algorithm for sustainable transportation systems in smart cities." IEEE Sensors Journal 21.14 (2020): 15437-15447.10.1109 / JSEN.2020.3019443.
8. Zhang, Linling, et al. "Potential for reducing carbon emissions from urban traffic based on the carbon emission satisfaction: Case study in Shanghai." Journal of Transport Geography 85 (2020): 102733. https://doi.org/10.1016/j.jtrangeo.2020.102733.
9. Kobiela, Dariusz, et al. "Vehicle type recognition: a case study of MobileNetV2 for an image Classification task." Procedia Computer Science 246 (2024): 3947-3956. https://doi.org/10.1016/j.procs.2024.09.169.
10. Suhao, Li, et al. "Vehicle type detection based on deep learning in traffic scene." Procedia computer science 131 (2018): 564-572. https://doi.org/10.1016/j.procs.2018.04.281.

11. Wu, Jian-Da, Wen-Jun Luo, and Kai-Chao Yao. "Acoustic signal classification using symmetrized dot pattern and convolutional neural network." Machines 10.2 (2022): https://doi.org/10.3390/machines10020090.
12. Wieczorkowska, Alicja, et al. "Spectral features for audio-based vehicle and engine classification." Journal of Intelligent Information Systems 50.2 (2018): 265-290. https://doi.org/10.1007/s10844-017-0459-2.
13. Kurowski, Adam, Szymon Zaporowski, and Andrzej Czyzewski. "1D convolutional context-aware architectures for acoustic sensing and recognition of passing vehicle type." 2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA). IEEE, 2020.10.23919 / SPA50552.2020.9241256.
14. Chen, Haoze, and Zhijie Zhang. "Hybrid neural network based on novel audio feature for vehicle type identification." Scientific reports 11.1 (2021): 7648. https://doi.org/10. 1109/ I2MTC43012.2020.9129183.
15. Dalir, Ali, Ali Asghar Beheshti, and Morteza Hoseini Masoom. "Classification of vehicles based on audio signals using quadratic discriminant analysis and high energy feature vectors." arXiv preprint arXiv:1804.01212 (2018). https://doi.org/10.5121/ijsc.2015.6105
16. Ashhad, Mohd, et al. "MVD: A novel methodology and dataset for acoustic vehicle type classification." arXiv preprint arXiv:2309.03544 (2023). https://doi.org/10.48550/arXiv.2309.03544.
17. Scarpiniti, Michele, Raffaele Parisi, and Yong-Cheol Lee. "A Scalogram-based CNN approach for audio classification in construction sites." Applied Sciences 14.1 (2023): 90.https://doi.org/10.3390/app14010090.
18. Shabber, Shaik Mulla, E. P. Sumesh, and Vidhya Lavanya Ramachandran. "Scalogram-based performance comparison of deep learning architectures for dysarthric speech detection." Artificial Intelligence Review 58.5 (2025): 128. https://doi.org/10.1007/s10462-024-11085-7.
19. Ullah, Saleem. "Vehicle type classification based on acoustic signals using LT-MFCC." Applied Acoustics 239 (2025): 110836.https://doi.org/10.1016/j.apacoust.2025.110836.
20. Shabber, Shaik Mulla, Mohan Ban-

sal, and Kodali Radha. "Machine learning-assisted diagnosis of speech disorders: a review of dysarthric speech." 2023 International Conference on electrical, electronics, communication and Computers (ELEXCOM). IEEE, 2023.10.1109/ELEXCOM58812.2023.10370116.

21. KK, M. S., Raju, R., Yassin, I., Eskandari, F., & MSA, M. A. (2024). Acoustic Vehicle Classification using Deep Learning Trained on a Spectrogram and Scalogram Fusion. https://doi.org/10.3390/app14010090.
22. Shanmugapriya, P., and V. Mohan. "Comparative analysis of deep learning models for dysarthric speech detection." Soft Computing 28.6 (2024): 5683-5698.
23. Gupta, Priyanka, Piyushkumar K. Chodingala, and Hemant A. Patil. "Morlet wavelet-based voice liveness detection using convolutional neural network." 2022 30th European Signal Processing Conference (EUSIPCO). IEEE, 2022.10.23919 /EUSIPCO55093.2022.9909835.
24. Lin, Jing. "Feature extraction of machine sound using wavelet and its application in fault diagnosis." NDT & e International 34.1 (2001): 25-30. https://doi.org/10.1016/S0963-8695(00)00025-6.
25. Darmawan, Yudhi, et al. "Reducing Complexity in Vehicle Recognition: A Comparison of Dimensionality Reduction Methods for Acoustic Features." ITEGAM-JETIA 11.55 (2025): 48-58. https://doi.org/10.5935/jetia.v11i55.1866.
26. [Rashed, Amr, et al. "Toward Inclusive Smart Cities: Sound-Based Vehicle Diagnostics, Emergency Signal Recognition, and Beyond." Machines 13.4 (2025): 258.https://doi.org/10.3390/machines13040258.
27. Kadhim, Ahlam M. "Optical recognition system and evaluation the transfer of Alex Net." 4th International Scientific Conference on Engineering Sciences and Advanced Technologies. Vol. 2830. No. 1. AIP Publishing LLC, 2023.https://doi.org/10.1063/5.0157750.
28. Ali, Sajad M., Hazim G. Daway, and Ahlam M. Kadhim. "Heart Diseases Classification by ECG Trace Images Using Deep Learning." International Conference on Digital Technologies and Applications. Cham: Springer Nature Switzerland, 2024.10.1007/978-52_3-031-68650-4.