CNN Architectures for Road Surface Wetness Classification from Acoustic Signals

Siavash Bahrami, Shyamala Doraisamy, Azreen Azman, Nurul Amelina Nasharuddin, and Shigang Yue

Abstract The classification of road surface wetness is important for both the development of future driverless vehicles and the development of existing vehicle active safety systems. Wetness on the road surface has an impact on road safety and is one of the leading causes of weather-related accidents. Although machine learning algorithms such as recurrent neural networks (RNN), support vector machines (SVM), artificial neural networks (ANN) and convolutional neural networks (CNN) have been studied for road surface wetness classification, the improvement of classification performances are still widely being investigated whilst keeping network and computational complexity low. In this paper, we propose new CNN architectures towards further improving classification results of road surface wetness detection from acoustic signals. Two CNN architectures with differing layouts for its dropout layers and max-pooling layers have been investigated. The positions and the number of the max-pooling layers were varied. To avoid overfitting, we used a 50% dropout layers before the final dense layers with both architectures. The acoustic signals of tyre to road interaction were recorded via mounted microphones on two distinct cars in an urban environment. Mel-frequency cepstral coefficients (MFCCs) features were extracted from the recordings as inputs to the models. Experimentation and comparative performance evaluations against several neural networks architectures were performed. Recorded acoustic signals were segmented into equal frames and thirteen MFCCs were extracted for each frame to train the CNNs. Results show that the proposed *CMCMDD*1 architecture achieved the highest accuracy of 96.36% with the shortest prediction time.

Keywords Deep learning · Convolutional neural networks · Road surface wetness detection · Audio classification

S. Bahrami · S. Doraisamy (✉) · A. Azman · N. A. Nasharuddin
Universiti Putra Malaysia, 43400 Serdang, Malaysia
e-mail: shyamala@upm.edu.my

S. Yue
University of Lincoln, Lincoln LN6 7TS, UK

R. Alfred and Y. Lim (eds.), *Proceedings of the 8th International Conference on Computational Science and Technology*, Lecture Notes in Electrical Engineering 835, https://doi.org/10.1007/978-981-16-8515-6_59

1 Introduction

Among the leading causes of weather related accidents is slippery road conditions. According to [1] there were 25,777 numbers of human fatality due to weather related accidents across Europe. The friction forces of the tyre on the road surface determine vehicle stability on the road, and this is an essential aspect in the development of vehicle safety systems [25]. An important aspect of developing an active safety feature for autonomous and semi-autonomous cars is the knowledge of road surface wetness condition. Using this knowledge, vehicles may automatically change their speed to maintain a safe distance from the vehicle in front of them, as well as providing higher manoeuvrability on slippery roads, allowing for more robust driver safety assistant systems [11].

Recently, new experiments detecting road surface wetness using mounted microphones and tyre sound recordings have been on the rise. Several studies have evaluated road surface conditions using acoustic signals by installing a microphone on the side of the road and collection of sound recordings from passing vehicles [14–16]. However with this method just a limited sections of the road acoustic signals are captured, these recording samples do not reflect the entire road condition. Recent investigations using data gathering on-board the vehicle have provided more comprehensive recording samples [2, 3, 6, 13, 19].

The use of acoustic signals can also help to overcome insufficient illumination, which can impact the effectiveness of computer vision systems. This research may be used in conjunction with systems like [8, 12], where one of the limitations is that these systems require external illumination and may function inadequately in low light. The majority of research on road surface wetness detection utilised octave-band features [3, 13] or auditory spectral features (ASF) from acoustic signals [2, 19]. One of the major goals of this study is to enhance the classification performance of the road surface wetness detection by using convolutional neural networks (CNN) with MFCCs features from acoustic signals.

For road wetness classification, several machine learning algorithms have been used, such as support vector machines (SVM) in research by [3, 13]. Recurrent neural networks (RNN), Long short-term memory (LSTM) and Bidirectional-LSTM(BLSTM) have been used by [2]. CNN was studied by [5, 19] and artificial neural networks (ANN) were investigated by [6]. Training a two-stream CNN model is computationally complex and require significant amount of resources as discussed in our previous work [5]. This complexity in the network architecture causes latency in the detection of wet road surface and compromises the vehicle safety system response time. Achieving the highest possible accuracy using less computationally complex models is critical for road safety systems.

CNN's superior classification performances have been demonstrated on other tasks of acoustic signal classification, such as environment sound categorization [20, 21]. A 4-layer CNN model with 2 convolution layers with max-pooling and 2 dense layers was trained by [20] where the feature set used was deltas of segmented spectrograms. The study by [21] used three convolutional layers which were interleaved

with two pooling layers followed by two fully connected layers with 64 and 10 hidden units, respectively. This study also also investigated experiments with data augmentation.

Based on the promising classification performances of these previous studies using CNNs on environment sound classification, we had proposed and evaluated several architectures in our previous study [5]. As for the acoustic features, MFCCs had shown superior performance in our recent work. We continue to investigate MFCCs with two differing CNN architectures that we propose in this paper for road wetness detection to improve classification performance. The first architecture layout begins with various depth of convolution layers stacked together followed by a max-pooling layer. We used a 50% drop out before the final fully connected layers to avoid overfitting. The convolution layer is interleaved with a pooling layer in the second architectural layout, which is followed by fully connected layers at the end. Similar to the first architecture, a 50% drop out was used. We validate the performance of these proposed CNN architectures.

The rest of the paper is structured as follows. Section 2 is the description of the proposed architectures. Section 3 presents the experimental setup and results. Section 4 discusses the conclusion and research directions for the future.

2 CNN Architectures for Road Wetness Detection

Two CNN architectures are investigated for road pavement wetness detection from recorded audio of tyre to road interaction. The first architecture is inspired by VGGNet [22] which instead of using a larger kernel size we stack multiple convolution layer with smaller kernel size with rectified linear activation units and the second architecture is inspired by Alexnet [17] and ZFNet [24]. These CNN models map a two-dimensional input X to a probability vector Z over two classes (wet and dry) via a series of layers. The number of hidden units and kernel sizes were chosen based on the preliminary experiments with 30 different combinations. Following this, the top three combination were selected for each of the proposed architectures as discussed in the following two subsections.

2.1 $C_n M D D_m$ Architecture

In this architecture, three levels of convolution depths have been studied. The sequence on how the layers are stacked up in this architecture, $C_n M D D_m$, are investigated. Convolution layers (C), Max-pooling (M), Dropout (D) and Dense layers (D) form this sequence. Varying levels of convolution (n) depths, $n = 1, 2, 4$ have been investigated. This is summarized as follows:

C_n is the convolution layer, where n is the number of layers.
M is max-pooling layer.
D is drop out layer.
D_m is the dense layer, where m indicates the number of layers.

Starting with only one layer of convolution with kernel size of (5,5), two layers with kernel sizes of (5,5) and (3,3) respectively, and lastly four layers of convolution with kernel sizes of (3,3). Smaller kernel sizes and stack multiple convolution layers were chosen as these has been discussed to be useful into obtaining higher accuracy [20–22]. Following the convolutional layers, a max-pooling layer of (2,2) size with a same size stride was selected. Default stride of 1 has been selected for convolution layers since max-pooling is used for down-sampling and selecting a bigger stride will cause more down-sampling of the feature maps. This process is followed by a drop out operation to avoid overfitting. Two or three fully connected layers were used for classification. Hyperparameters selected for the $C_n M D D_m$ architecture are shown in Table 1.

The last architecture $C_4 M D D_3$ is illustrated as an example in Fig. 1. This architecture has four convolution layers. The convolution layers start with 32 filters and increase exponentially to 256 filters. To reduce the computational complexity of the model,we kept the kernel sizes small and the same size across all four layers of convolution.

Table 1 $C_n M D D_n$ architecture and selected Hyperparameters.

Model	Num filters	Kernel size	Stride shape	Pool size	Dense layer size
$C_1 M D D_2$	32	(5,5)	(1,1)	(2,2)	128,2
$C_2 M D D_2$	64,128	(5,5)(3,3)	(1,1)	(2,2)	128,2
$C_4 M D D_3$	32,64,128,256	(3,3)(3,3)(3,3)(3,3)	(1,1)	(2,2)	256,128,2

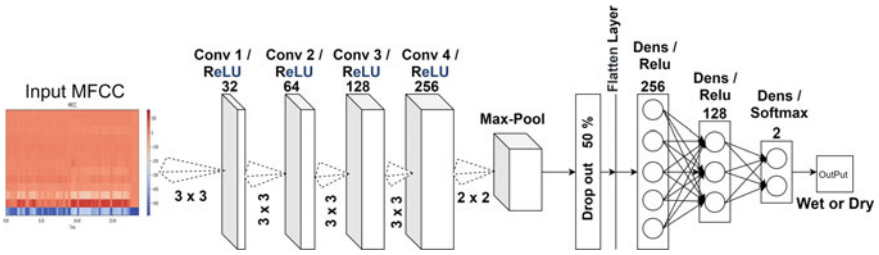

Fig. 1 $C_4 M D D_3$ architecture with MFCC as the input layer

2.2 *CMCMDD Architecture*

With an increase in the depth of convolution layers in the first architecture and possibly making the network more computationally complex, a second architecture is proposed. Pooling layers can be placed between convolution layers and lower kernel sizes can be used to improve network performance, as discussed in [24] and [22]. This has also been discussed as a better option for larger datasets in these studies. Two convolution layers are interleaved with a pooling layer in the second proposed network architecture. Convolution layers are followed by a drop out operation and deep fully connected layers for classification, similar to the first architecture. The sequence of how these layers are stacked up in this architecture is—convolution layer followed by max-pooling repeated twice and finishing with dropout and dense layers therefore we call it *CMCMDD*. Three layouts *CMCMDD*1, *CMCMDD*2 and *CMCMDD*3 are proposed. For this architecture, we kept the number of layers constant during the experiment and only changed the kernel sizes in convolution layers. Hyperparameters selected for the *CMCMDD* architecture are shown in Table 2:

Figure 2 illustrates the CMCMDD3 as an example, which has 2 convolution layers, the convolution layers start with 32 filters and then interleaved with a 2 by 2 max-pooling before going through the second convolution. The final dense layers and operations are similar with the first architecture.

Table 2 *CMCMDD* architecture and selected Hyperparameters

Model	CNN Layer Size	Kernel Shape	Stride Shape	Pool Size	Dense layer size
*CMCMDD*1	32,64	(5,5)(5,5)	(1,1)	(2,2)(2,2)	128,64,2
*CMCMDD*2	32,64	(5,5)(3,3)	(1,1)	(2,2)(2,2)	128,64,2
*CMCMDD*3	32,64	(3,3)(3,3)	(1,1)	(2,2)(2,2)	128,64,2

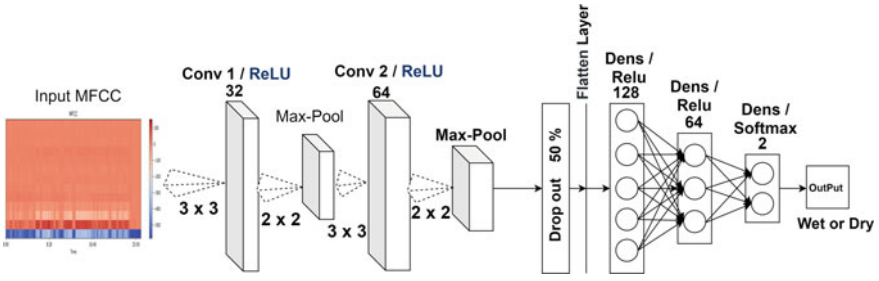

Fig. 2 *CMCMDD* architecture with two layers of convolution and kernel size of (3,3)

3 Experiments and Results

The accuracy measures of the proposed architectures are compared to the CNN architecture used in [19] and the BLSTM used in [2] which are the best performing state-of-the-art neural network approaches in road surface wetness detection. The input layer used to train the models for benchmarking is MFCCs of the acoustic signals. In addition, we also compared the proposed models to RNN-LSTM. The RNN-LSTM model that is used for benchmarking has two LSTM layer each with 128 units followed by a dropout layer. A single dense layer with 2 hidden units using $softmax$ activation function is used to classify the MFCCs of the signal.

3.1 Experimental Setup

This section discusses the experimental setup, the dataset and the acoustic features that were used for training and testing the models. Accuracy performance metric were used to compare our best performing CNN architecture with state-of-the-art models.

For training of the models, similar experimental setup to our previous study has been used. This includes the using the Adam optimizer with $beta1$ and $beta2$ set 0.9 and 0.999 respectively, learning rate of 0.001, batch size of 32 and a patience of 10 epochs with a maximum of 1000 epochs. For further details of the selection of these hyperparameters refer to [5].

Dataset Acoustic signals of tyre to road interaction needed for the experimentation in this study were recorded using microphones mounted on cars driven in urban environment.

To record the acoustic signals of tyre to road interaction, two Comica CVM-V02 microphones were placed near the tyre and the road surface. Previous studies [2, 3, 6, 13, 14] placed a single microphone near the rear tyres where there are less engine noise. In this study, acoustic signals were collected from two microphones mounted near both rear tyres. The same two cars used in our previous study [5] were used for data collection. Car 1 is a 2012 Toyota Aygo and Car 2 is a 2015 VW Tiguan. As shown in Fig. 3 a down facing camera and a dashcam were used to inspect the state of the road's surface. Both microphones were positioned on the front side of the tyres, which is less impacted by water splashing. The microphones were fitted with a deadcat furry windshield to reduce wind impact noises. Rest of the setup is similar to our previous study, for more details please refer to [5]. Figure 3b shows the setup of microphones and Fig. 3a shows in-cabin interface.

For data collection, two different routes were selected around city of Lincoln, UK, which consist of dual carriageways and urban roads. Data were collected for wet and dry condition during the daylight and night time. The wet condition was recorded either during the rain or right after rain has finished as shown in Fig. 3c, d. In total 8 recording session were conducted which resulted in 390 min of audio and video data.

Fig. 3 Data acquisition setup on Car 1 and sample images recorded using a down-facing camera

On each of the selected routes two session of dry recordings and two session of wet recording was conducted. This Complete data set has been made publicly available at https://sbahrami.com/dataset/ICCST21/.

Feature extraction For the training and evaluation of the models, MFCCs of the acoustic signals were used as feature inputs. MFCCs have been used for speech processing and were designed to retain phonetically significant acoustic information [10] but proven to be effective in other tasks such as music [18] and environment audio analysis [9]. To calculate MFCCs we must apply a mel-filter bank to the power spectrum of the signal and calculate the summation of the energy for each filter. To get the coefficients, logarithmic discrete cosine transform of each frequency component will be calculated. Unlike rain sound classification [7], audio recordings of tyre to pavement interaction is structured and has a constant sound that can be generally classified more accurately with MFCCs.

In total using all these data resulted in 63,334 mel-frequency ceptrum from the all the recording of the datasets six trips. The CNN models were trained using 13 MFCCs. By applying 26 filters to the signal, 26 coefficients will be obtained. To reduce feature vector complexity, we discard half of the coefficients as only lower frequencies are needed. Preliminary experiments were performed with higher number of coefficients. We found that by selecting a larger number of cepstral coefficients the model's complexity increased and it did not have a significant effect on the model's accuracy. Short-time fourier transform (STFT) was applied to extract MFCCs using a frame size of 30 milliseconds to obtain more samples from the data. For example, for a frame size of 30 milliseconds and a 16kHz signal, 480 samples are generated($0.030 * 16000 = 480$ samples).

Evaluation metrics On a similiar approach as [4], to evaluate the performance of the models $accuracy$, $recall$, $precision$ and $F - measure$ were used [23] and are defined as follow:

$$accuracy = \frac{tp + tn}{tp + fp + fn + tn} \tag{1}$$

$$recall = \frac{tp}{tp + fn} \tag{2}$$

$$precision = \frac{tp}{tp + fp} \tag{3}$$

$$F - measure = \frac{(\beta^2 + 1) * precision * recall}{\beta^2 * precision + recall} \tag{4}$$

where in this is case tp is when wet road detected correctly, fp dry road detected correctly, tn wet road detected incorrectly, fn dry road detected incorrectly. The $\beta = 1$ is used to evenly balance the $F - measure$, also refer to as F_1. If $\beta > 1$ it will favour precision and if $\beta < 1$ it will favour recall.

3.2 Results

The various hyperparameters discussed in Sect. 2 are used in the experimentation.To overcome the problem of overfitting, K-fold validations are commonly used. In order insure that we did not overfit the models we applied a 3-fold validation similar to our previous study. Details of these are available in [5]. The total of 162,221 samples were divided into 3 equal parts. For training and validation of the models 147,474 samples were used and the remaining of 14,747 were used for testing. The test data was only used for prediction and kept constant during all 3-folds. Based on the 3-fold iterations the average mean accuracy of the models were calculated. For each model we calculate, out of sample score base on the average of root mean square error (RMSE) of each iteration.

Tables 3 shows the overall performances of both models for $C_n MDD_m$ and *CMCMDD* architecture. All CNN models perform well in detecting road surface wetness but the *CMCMDD*1 architecture with 2 layers of convolution and 2 layers of dropout has achieved the highest overall average accuracy. The highest precision, recall and F_1 measures were achieved by *CMCMDD*1 as well. Higher precision and recall can be interpreted as, better prediction of the wet road. It can be concluded that using a larger kernel size with max-pooling overall performance is better. $C_4 MDD_3$ architecture was able to achieve 96.19% accuracy which is similar to *CMCMDD*1.

Following the best performing architecture from Table 3, further experiments were conducted to evaluate the effect of car and tyre differences on the classification performance of the proposed architectures. *CMCMDD*1 and $C_4 MDD_3$, were trained and tested using data recorded from Car 1 and Car 2 individually. Table 4 shows when the model is trained using Car 1 and tested using Car 2 and inversely, the classification performance of the models drop by 40%. This drop in classification performance can be attributed to the two vehicles' tyre types and sizes. Despite the fact that both cars were equipped with summer tyres at the time of recording, Car 1 had a smaller tyre than Car 2. As it can be seen in Table 4 when the models are trained and tested using data collected from the same car, better performance and the models can achieve accuracies above 99%. The study by [19] investigated the effect of tyre type, summer or winter tyre, on the classification performance. In our previous

Table 3 $C_n MDD_n$ and *CMCMDD* architecture and configuration with corresponding overall mean accuracy performance

Model	Accuracy	Precision	Recall	F_1	RMSE
$C_1 MDD_2$	0.9399	0.9400	0.9390	0.9394	0.2143
$C_2 MDD_2$	0.9612	0.9608	0.9612	0.9610	0.1718
$C_4 MDD_3$	0.9619	0.9611	0.9631	0.9618	0.1688
*CMCMDD*1	0.9636	0.9628	0.9644	0.9635	0.1686
*CMCMDD*2	0.9594	0.9587	0.9600	0.9592	0.1767
*CMCMDD*3	0.9607	0.9598	0.9617	0.9605	0.1734

Table 4 Classification performance of best performing $C_n MDD_m$ and *CMCMDD* architectures trained and tested using data collected from Car 1 and Car 2 separately

Trained	Tested	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
		$C_4 MDD_3$				*CMCMDD*1			
Car 1	Car 1	0.9946	0.9947	0.9945	0.9946	0.9959	0.9958	0.9959	0.9958
Car 1	Car 2	0.5197	0.5303	0.5297	0.5193	0.5441	0.5538	0.5532	0.5439
Car 2	Car 2	0.9911	0.9915	0.9906	0.9910	0.9945	0.9945	0.9944	0.9944
Car 2	Car 1	0.4326	0.3758	0.4157	0.3707	0.4587	0.4031	0.4403	0.3879

study [5] we investigated the effect of car type on classification performance using a smaller dataset. In this paper we investigate the effect using a larger dataset. Training the models with this larger dataset improved the overall classification performance. However, the problem still remains where the models only trained on a single car's acoustic data the classification performance drops significantly.

Achieving the best overall performance by *CMCMDD*1 in our experiments above, a benchmarking evaluation is performed against RNN-LSTM, CNN [19] and BLSTM [2]. The training was performed using the same dataset and MFCCs feature vectors. Table 5 presents the mean average accuracy over 3-fold and *CMCMDD*1 model achieved the best accuracy of 96.36% and F_1 score of 96.35%. Although the *CMCMDD*1 model improvement of accuracy performance over the BLSTM [2] and RNN-LSTM is relatively small, the training and predication time for BLSTM models are higher. The training time per epoch for *CMCMDD*1 using a Nvidia GeForce GTX 1080 GPU is 28 s while for the BLSTM is 149 s. All the models can classify MFCC of 1 s signal under 1 millisecond. The prediction time of 14,747 testing samples for the BLSTM is 4.52 s which is more than ten times of the *CMCMDD*1 model which only took 0.332 s. The training and prediction time of BLSTM was also discussed by [19] whereby the implementation of such models to the on-board vehicle safety system have to be considered, making the CNN models favourable over the BLSTM.

Table 5 Mean average accuracy over 3 folds

Model	Accuracy (%)	F_1 (%)	Model	Accuracy (%)	F_1 (%)
*CMCMDD*1	96.36	96.35	CNN [19]	93.29	93.26
RNN-LSTM	95.26	95.24	BLSTM [2]	95.05	95.02

4 Conclusion and Future Work

In this study, two differing CNN architectures $C_n MDD_m$, a seven-layer CNN with 1 to 4 layers of convolution followed by 2 or 3 dense layers and *CMCMDD*, a five-layers CNN with 2 layers of convolution and 3 dense layers were proposed and evaluated for wet road surface detection. These architectures use small kernel sizes to improve computation complexity. MFCCs have been proven to be an effective feature set as input layers. The proposed *CMCMDD*1 architecture with 2 layers of convolution and 5 by 5 kernel size achieved the best result of 96.36% mean average accuracy and *CMCMDD*1 would be useful as an overall road wetness classification approach. We showed that high level of classification performance is achievable using a simple network architecture that is less computationally complex and has a faster prediction time. Using CNN for road surface wetness detection has been shown comparable performance to using ANN, LSTM and BLSTM methods. Accuracy performance of the new CNN architectures proposed are promising for road wetness classification. We will continue to study the placement and number of the max pooling layers for prediction time optimization and higher classification performance. Further investigation on acoustic features or classification models that can better generalize the road surface acoustic data.

Future work will be to evaluate on a larger dataset and a more robust evaluation using data augmentation techniques to be performed for automated road surface wetness detection.

Acknowledgements This research has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No. 691154 STEP2DYNA and No. 778602 ULTRACEPT.

References

1. Reported road casualties in Great Britain: main results 2018. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/820562/Reported_road_casualties_-_Main_Results_2018.pdf
2. Abdic I, Fridman L, Brown DE, Angell W, Reimer B, Marchi E, Schuller B, (2016) Detecting road surface wetness from audio: a deep learning approach. In, (2016) 23rd International conference on pattern recognition (ICPR), pp 3458–3463, No 1. IEEE, Cancun. https://doi.org/10.1109/ICPR.2016.7900169

3. Alonso J, López JM, Pavón I, Recuero M, Asensio C, Arcas G, Bravo A (2014) On-board wet road surface identification using tyre/road noise and Support Vector Machines. Appl Acoust 76:407–415
4. Ambrosini L, Gabrielli L, Vesperini F, Squartini S, Cattani L (2018) Deep neural networks for road surface roughness classification from acoustic signals. Audio Eng Soc Convent, p 9934
5. Bahrami S, Doraisamy S, Azman A, Nasharuddin NA, Yue S (2020) Acoustic feature analysis for wet and dry road surface classification using two-stream CNN. In: 2020 4th International conference on computer science and artificial intelligence. ACM, New York, NY, USA, pp 194–200. https://doi.org/10.1145/3445815.3445847
6. Boyraz P (2014) Acoustic road-type estimation for intelligent vehicle safety applications. Int J Veh Safety 7(2):209
7. Brown A, Garg S, Montgomery J (2019) Automatic rain and cicada chorus filtering of bird acoustic data. Appl Soft Comput 81:105501. https://doi.org/10.1016/j.asoc.2019.105501
8. Cahyadi WA, Kim YH, Chung YH, Ghassemlooy Z (2015) Efficient road surface detection using visible light communication. In: International conference on ubiquitous and future networks, ICUFN, 2015-August, pp 61–63 (2015). https://doi.org/10.1109/ICUFN.2015.7182498
9. Chu S, Narayanan S, Kuo CCJ (2009) Environmental sound recognition with time-frequency audio features. IEEE Trans Audio Speech Language Process 17(6):1142–1158. https://doi.org/10.1109/TASL.2009.2017438
10. Davis S, Mermelstein P (1980) Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences. IEEE Trans Acoust Speech Sign Process 28(4):357–366. https://doi.org/10.1109/TASSP.1980.1163420
11. Ghandour R, Victorino A, Doumiati M, Charara A (2010) Tire/road friction coefficient estimation applied to road safety. In: 18th Mediterranean conference on control and automation, MED'10 - conference proceedings, pp 1485–1490. https://doi.org/10.1109/MED.2010.5547840
12. Jonsson P, Casselgren J, Thörnberg B (2015) Road surface status classification using spectral analysis of NIR camera images. IEEE Sens J 15(3):1641–1656. https://doi.org/10.1109/JSEN.2014.2364854
13. Kalliris M, Kanarachos S, Kotsakis R, Haas O, Blundell M (2019) Machine learning algorithms for wet road surface detection using acoustic measurements. In: Proceedings—2019 IEEE international conference on mechatronics, ICM 2019, vol 1, pp 265–270. IEEE, Ilmenau, Germany. https://doi.org/10.1109/ICMECH.2019.8722834
14. Kongrattanaprasert W, Kamakura T (2010) Detection of road surface states from tire noise using neural network analysis. IEEJ Trans Ind Appl 130(7):920–925
15. Kongrattanaprasert W, Nomura H, Kamakura T, Ueda K (2009) Application of neural network analysis to automatic detection of road surface conditions utilizing tire noise from vehicles. J Acoust Soc Am 125:2731. https://doi.org/10.1541/ieejias.129.761
16. Kongrattanaprasert W, Nomura H, Kamamura T, Ueda K (2010) Automatic detection of road surface states from tire noise using neural network analysis. In: Proceedings of 20th International Congress on acoustics (August), pp 1–4
17. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60(6):84–90. https://www.taylorfrancis.com/books/9781420010749, https://doi.org/10.1145/3065386
18. Maliki I, Sofiyanudin, (2018) Musical instrument recognition using mel-frequency cepstral coefficients and learning vector quantization. IOP Conf Ser: Mater Sci Eng 407(1):012118. DOIurlhttps://doi.org/10.1088/1757-899X/407/1/012118
19. Pepe G, Gabrielli L, Ambrosini L, Squartini S, Cattani L (2019) Detecting road surface wetness using microphones and convolutional neural networks. In: AES 146th International convention
20. Piczak KJ (2015) Environmental sound classification with convolutional neural networks. In: IEEE international workshop on machine learning for signal processing, MLSP, 2015-November, pp 1–6 (2015). https://doi.org/10.1109/MLSP.2015.7324337

21. Salamon J, Bello JP (2017) Deep convolutional neural networks and data augmentation for environmental sound classification. IEEE Sign Process Lett 24(3):279–283. https://doi.org/10.1109/LSP.2017.2657381
22. Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. In: International conference on learning representations, pp 1–14. San Diego (2014)
23. Sokolova M, Japkowicz N, Szpakowicz S (2006) Beyond accuracy, f-score and roc: a family of discriminant measures for performance evaluation. In: Sattar A, Kang BH (eds) AI 2006: advances in artificial intelligence. Springer, Berlin, pp 1015–1021
24. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: Fleet D, Pajdla T, Schiele B, Tuytelaars T (eds) 13th European conference on computer vision, ECCV 2014, vol 8689 LNCS, pp 818–833. Springer, Zurich. https://doi.org/10.1007/978-3-319-10590-1_53
25. Zhao YQ, Li HQ, Lin F, Wang J, Ji XW (2017) Estimation of road friction coefficient in different road conditions based on vehicle braking dynamics. Chin J Mech Eng (English Ed) 30(4):982–990. https://doi.org/10.1007/s10033-017-0143-z