Metodi di deep learning acustico per il riconoscimento dei dissesti della pavimentazione stradale

Alessandro Monticelli

Dipartimento di Fisica E. Fermi,
Università di Pisa,
Largo Bruno Pontecorvo, 3, 56127 Pisa
a.monticelli1@studenti.unipi.it

Ricevuto: 28/2/2023
Accettato: 17/5/2023

DOI: 10.3280/ria2-2023oa15509
ISSNe: 2385-2615

Nel seguente lavoro è stata proposta una metodologia basata su tecniche di deep learning per la valutazione delle condizioni della superficie stradale a partire da segnali acustici misurati all'interno della cavità dello pneumatico. Il progetto è stato svolto in collaborazione con Ipool srl., nel contesto del progetto SURFAce, finanziato dalla regione Toscana. Sono state proposte tre architetture di classificazione: una LSTM *(Long short-term memory network)* basata sull'andamento temporale di un insieme di descrittori spettrali e due CNN *(Convolutional neural network),* una incentrata sugli spettrogrammi dei segnali, l'altra sui Mel-frequency cepstral coefficients (MFCC). Il dataset di ground truth è stato acquisito tramite un laboratorio mobile e classificato mediante strumenti di analisi appositamente sviluppati. Due delle tre architetture proposte hanno fornito risultati incoraggianti. L'implementazione di tali strumenti su dispositivi mobili potrebbe rendere possibile la classificazione dello stato della pavimentazione in tempo reale con ridotti costi economici e temporali.
Parole chiave: tire cavity noise, deep learning, valutazione delle condizioni stradali

Acoustical deep learning methods for road pavement distress recognition
In the following work, a deep learning-based methodology was proposed to evaluate road surface conditions starting from acoustic signals measured inside the car tire cavity. The project was carried out in collaboration with Ipool srl., in the context of the SURFAce project, funded by the Tuscany region. Three different classification architectures were proposed: An LSTM, based on the time series of a set of spectral descriptors, and two CNNs, the first focused on the signals' spectrograms and the second on their Mel-frequency cepstral coefficients (MFCCs). The ground truth data set was acquired through a mobile laboratory and classified through aptly developed analysis tools. Two of the three proposed architectures have provided encouraging results, and their implementation on portable systems could lead to real-time pavement classification in a cost and time-efficient way.
Keywords: tire cavity noise, deep learning, road condition assessment

1 | Introduzione

Il rumore da traffico veicolare rappresenta uno dei principali agenti di disturbo acustico in ambiente urbano. Le due sorgenti principali che contribuiscono a tale emissione acustica sono il rumore del motore e il rumore da rotolamento associato agli pneumatici. Ricerche sull'argomento (ad esempio quella condotta da Ipool) hanno permesso di evidenziare come il rumore da rotolamento sia dominate a velocità superiori a 40km/h [1]. La progressiva introduzione dei veicoli elettrici, inoltre, ha ridotto ulteriormente la componente rumorosa associata al funzionamento del motore dell'autoveicolo [2]. Per questo motivo, il monitoraggio del rumore da rotolamento costituisce uno dei campi di ricerca più attivi nell'ambito dell'acustica ambientale.

Il metodo più utilizzato per la valutazione dell'impatto acustico del rumore da rotolamento è costituito dal metodo standardizzato CPX [3] ("*Close proximity method*") che consiste nella misurazione dei livelli acustici nell'immediata vicinanza della zona di contatto tra lo pneumatico e la pavimentazione stradale.

In tempi recenti, è stato proposto un nuovo metodo per l'esame del rumore da rotolamento, basato sull'analisi dei segnali di rumore acustico misurati all'interno della cavità dello pneumatico (metodo TCN [4], "*Tire cavity noise*"). Alcuni studi recenti [5] hanno evidenziato una forte correlazione tra i livelli CPX a basse frequenze (dove la componente di rumore influenzata dalle vibrazioni dello pneumatico causate dall'interazione con la pavimentazione risulta dominante) e livelli TCN a basse frequenze. Ricerche analoghe hanno permesso di evidenziare una simile correlazione tra lo spettro TCN e lo spettro della tessitura stradale [6], valutato considerando la limitata deformabilità dello pneumatico, facendo riferimento alla descrizione di tale fenomeno fornita dalla formula di enveloping di Clapp [7] riportata in equazione (1).

$$\frac{\pi E u(x)}{2(1-\nu^2)} + c_0 = -\int_a^b p(\epsilon) \ln|\epsilon - x| \, d\epsilon \qquad (1)$$

Nell'equazione E indica il modulo di Young dello pneumatico, ν il suo coefficiente di Poisson, $u(x)$ la deformazione dello pneumatico nella zona di contatto, valutata lungo la direzione longitudinale; $p(\varepsilon)$ il campo di pressione in prossimità della zona di contatto, valutato lungo la direzione longitudinale; a e b indicano gli estremi di tale area.

Tali indagini sembrano suggerire la possibilità di ricavare informazioni rilevanti sullo stato della tessitura stradale a partire da segnali TCN, valutati nell'intervallo delle basse frequenze (100-1000 Hz), dove la correlazione tra le osservabili risulta significativa [8].

La correlazione così evidenziata tra rumore TCN e tessitura stradale non risulta facilmente modellizzabile da un punto di vista matematico, sebbene un modello correlante la macrotessitura stradale e il rumore TCN sia stato fornito da J. Pinay et al. [9], i quali hanno ricavato una relazione che lega il L_{eq}(TCN) ad un indice sintetico per la valutazione delle condizioni stradali (MPD, "*Mean profile depth*"). In Equazione (2) è riportata la relazione tra il livello di pressione interna e le altre grandezze fisiche rilevanti per la sua valutazione.

$$L_{TCN} = a + b * \log(v) + cMPD + dF_{load} + ep \qquad (2)$$

Nell'equazione precedente v indica la velocità del veicolo, MPD la Mean profile depth della pavimentazione, F_{load} il carico verticale associato alla ruota in esame e p la pressione di gonfiaggio dello pneumatico. I parametri a, b, c, d ed e sono coefficienti di regressione.

Nel 2022, tuttavia, uno studio condotto da Schiaffino et al. [10] ha dimostrato come metodi di classificazione basati sul machine learning possano essere utilizzati ai fini della classificazione delle condizioni del manto stradale, ottenendo risultati incoraggianti.

Nel presente lavoro, è stata valutata la possibilità di sviluppare sistemi di riconoscimento dello stato della pavimentazione stradale a partire da segnali TCN basati sulle recenti tecniche di deep learning. Sono state quindi sviluppate tre diverse architetture, le quali sono state applicate ad un problema di classificazione a tre classi.

2 | Metodi

2.1 | Apparato sperimentale

Le misurazioni sono state effettuate attraverso un laboratorio mobile equipaggiato con la seguente strumentazione:

I. una videocamera GoPro, per l'acquisizione di rilievi video;
II. un encoder rotativo, connesso alla ruota posteriore sinistra, capace di inviare 20 segnali impulsivi per giro, utilizzato per la valutazione della distanza percorsa e della velocità del veicolo;
III. un accelerometro, connesso alla medesima scheda di acquisizione per l'encoder, capace di misurare le vibrazioni trasmesse dai dissesti stradali al veicolo;
IV. un sensore TCN, costituito da un microfono a condensatore connesso ad un microcontroller Raspberry Pi Zero. Tale strumento permette la misurazione dei segnali di rumore all'interno dello pneumatico e l'invio dei file wav al computer di bordo mediante una porta Wi-Fi.

I dati relativi alla posizione sono stati inoltre monitorati tramite i sensori GPS presenti nei telefoni cellulari, in maniera tale da permettere una geolocalizzazione dei risultati di misura. I segnali di pressione acustica TCN sono stati acquisiti ad una frequenza di campionamento di 16kHz.

I segnali sono stati sincronizzati con l'ausilio di un dosso artificiale, rilevabile distintamente da tutti gli strumenti coinvolti nel processo di misura. In questo modo è stato possibile rapportare tutti i segnali acquisiti alla medesima base temporale, in modo tale da associare a ciascun segnale audio un'informazione sulla condizione della relativa pavimentazione, ottenuta a partire dal rilievo video.

I siti di misura sono stati selezionati valutando la possibilità di acquisire dati relativi a sezioni omogenee e estese di pavimentazioni in diverse condizioni. Le condizioni di traffico sono state inoltre valutate, poiché, affinché i dati potessero essere considerati rappresentativi, le misure effettuate a velocità minori di 30 km/h o in presenza di forti accelerazioni o decelerazioni sono state scartate. Le misure, infine, sono state effettuate in condizioni meteorologiche favorevoli durante la fascia serale (19-22) del periodo estivo (luglio-agosto 2022), su pavimentazioni asciutte.

2.2 | Creazione del ground truth data set

I dati acquisiti durante le campagne di misura sono stati in seguito suddivisi in segmenti da 1s ciascuno, i quali sono stati utilizzati per la creazione di un dataset di ground truth, contenente segnali classificati utilizzabili per l'allenamento di un classificatore. In Figura 1 è riportata una rappresentazione schematica della metodologia sperimentale

I dati così ottenuti sono stati suddivisi in tre classi distinte: una relativa a pavimentazioni in buone condizioni ("*good*"), una relativa a superfici danneggiate ("*bad*") e una relativa a tessiture contenenti uno specifico tipo di ammaloramento ("*pothole*"), definita come una buca di forma approssimativamente circolare di diametro superiore ai 100 mm [11]. Attraverso tale processo di classificazione, è stato possibile selezionare un data set di 2850 campioni (950 per ciascuna classe), su cui sono state effettuate le operazioni di allenamento e validazione delle reti neurali.

A partire da questi dati, sono stati isolati due diversi data set: un set di allenamento, costituito da 800 campioni per ciascuna classe, e un set di validazione, costituito da 150 campioni per ciascuna classe.

In un secondo momento, è stato realizzato un ulteriore set di dati, utilizzato per testare l'efficienza finale della rete. Tale data set (detto data set di test) è stato realizzato a partire da 150 campioni per ciascuna classe.

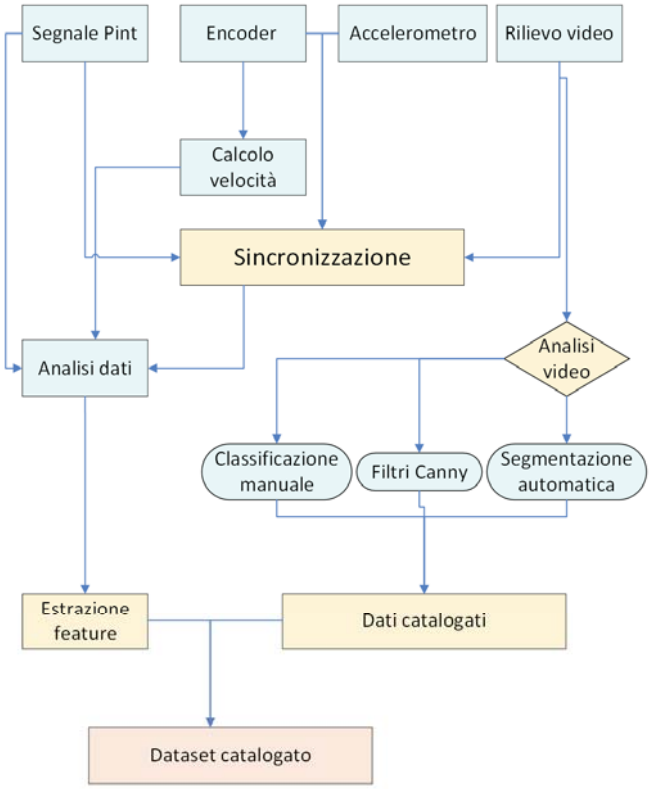

Fig. 1 – Rappresentazione schematica della metodologia sperimentale

Schematic of the experimental methodology

2.3 | Architetture testate

Nel presente lavoro sono state testate tre distinte architetture per la classificazione dei dati audio TCN: una rete LSTM [12] (*Long short term memory network*) e due reti CNN [13, 14] (*Convolutional neural network*). La scelta di queste architetture dipende dai diversi modi possibili in cui i segnali acustici possono essere caratterizzati: le reti LSTM permettono di classificare dati di input rappresentati tramite serie temporali, mentre le reti CNN permettono una classificazione di un segnale a partire da una sua rappresentazione grafica (ad esempio spettrogrammi, spettrogrammi in scala di frequenze Mel e MFCC [15]).

La prima rete CNN è stata progettata per permettere una classificazione a partire dagli spettrogrammi dei segnali ("*SpectNet*") in ingresso, la seconda è stata invece sviluppata per permettere una classificazione a partire da una caratterizzazione basata sugli MFCC del segnale ("*MFCCNet*").

3 | Risultati

I dati acquisiti durante la campagna di misura sono stati elaborati in modo tale da permettere una caratterizzazione preliminare dei segnali, evidenziandone le proprietà utilizzabili ai fini della classificazione. In particolare, sono state valutate 7 feature spettrali, due tipologie di rappresentazione in tempo frequenza del segnale (spettrogrammi e spettrogrammi Mel) e l'andamento temporale dei primi 13 coefficienti del *Cepstrum* in scala Mel (MFCC).

3.1 | Risultati per l'approccio LSTM

I segnali sono stati descritti tramite delle specifiche caratteristiche relative alla loro rappresentazione nel dominio delle frequenze (feature spettrali). Sono state considerate le seguenti feature [16] nel processo di caratterizzazione del segnale:

I. i centroidi spettrali;
II. la deviazione spettrale;
III. l'acutezza spettrale;
IV. la curtosi spettrale;
V. la pendenza spettrale [17];
VI. il punto di roll-off dello spettro [18];
VII. l'energia sonora.

Il numero delle feature è stato ulteriormente ridotto mediante l'applicazione della PCA *(Principal Component Analysis)* [19, 20], attraverso la quale sono state definite 3 nuove componenti che sono risultate sufficienti a spiegare più del 90% (il 93.1%) della varianza del data set (Figura 2).

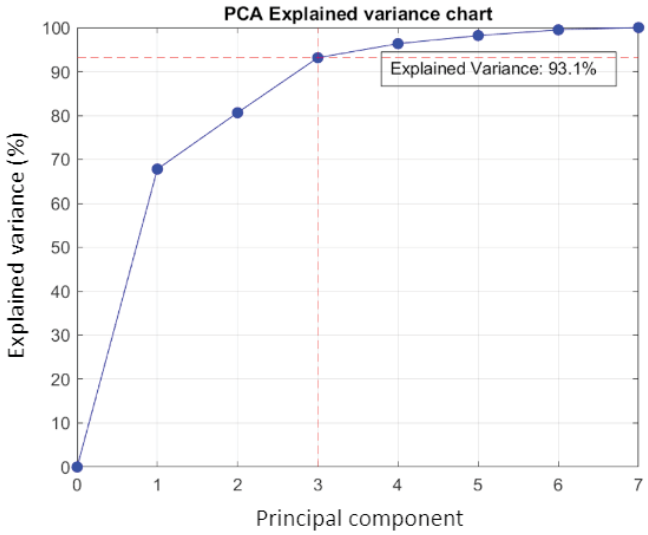

Fig. 2 – Varianza cumulativa descritta dalle componenti principali

Description of the cumulative variance for the principal components

Le feature spettrali possono essere definite su intervalli costituiti da un numero arbitrario di campioni: è quindi possibile ricavare un valore medio della feature in esame sul segnale nella sua interezza (valutato cioè su 16.000 campioni, si ricorda che i segnali esaminati hanno durata pari a 1s e sono stati acquisiti con una frequenza di campionamento di 16kHz). I valori medi sono stati utilizzati per condurre l'analisi preliminare e per ricavare le direzioni di massima varianza tramite PCA; tuttavia, è possibile costruire sequenze temporali relative ai descrittori spettrali riducendo l'intervallo su cui essi vengono valutati. Proiettando questi dati lungo le direzioni di massima varianza definite dalla PCA, è possibile ottenere

tre sequenze temporali che possono essere utilizzate come input per una rete LSTM. La lunghezza dell'intervallo su cui viene valutato il valore delle feature è stato posto uguale a 1024 campioni, con un overlap del 50% (512 campioni) tra un intervallo e il successivo. Su ciascun campione è stata effettuata una finestratura di tipo Hamming [21].

In conclusione, il dato di input per il classificatore basato su un'architettura LSTM era costituito da 3 sequenze temporali composte da 30 campioni ciascuna, relative all'andamento temporale delle tre feature ottenute tramite PCA a partire dall'andamento temporale di diversi descrittori spettrali.

La rete LSTM utilizzata è stata addestrata sul data set di allenamento con un learning rate pari a 10^{-4}, su un numero di epoch pari a 100 e un batch size pari a 64, tramite l'algoritmo di ottimizzazione "*adam*" *(Adaptive moment estimator)* [22, 23]. In questo modo, è stato possibile ottenere un'accuratezza sul data set di validazione pari al 92.2%. Risultati simili sono stati ottenuti con un numero di epochs pari a 30, un batch size pari a 64 e un learning rate pari a 10^{-3} (accuratezza sul data set di validazione: 90.7%).

I risultati complessivi della rete sono stati infine valutati sul data set di test. L'accuratezza complessiva della rete LSTM sul data set di test è risultata pari a 80.9% (Figura 3). La performance complessiva della rete è stata valutata calcolando i parametri di recall, precisione e F-1 score per ciascuna classe (Tabella 1).

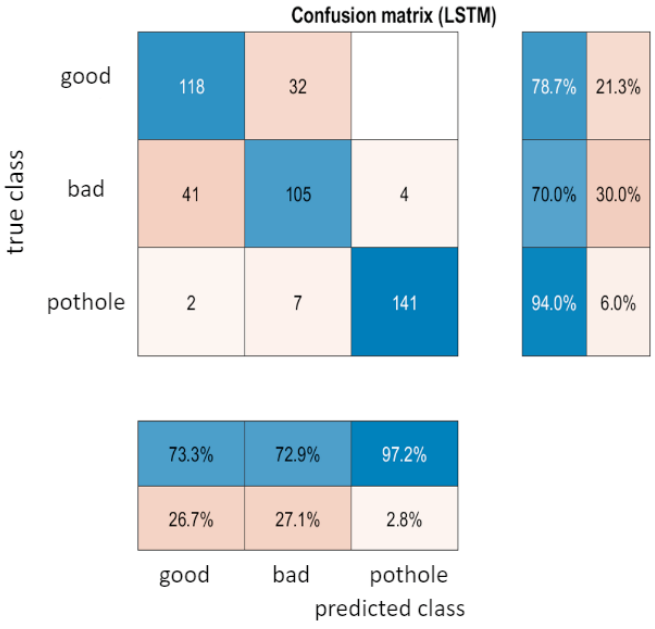

Fig. 3 – Risultati per una rete LSTM su dati appartenenti al set di test
Results of a LSTM network on the test datasets

Tab. 1 – Indici di qualità per la rete LSTM
Quality indices for the LSTM network

	Integra	*Dissestata*	*"Pothole"*
Precisione	73.3%	72.9%	97.2%
Recall	78.7%	70.0%	94.0%
F1-Score	75.9%	71.4%	95.6%

La precisione di una rete neurale è definita come la frazione degli elementi effettivamente appartenenti ad una determinata classe sul totale degli elementi classificati dalla rete come appartenenti a tale classe; il parametro di recall indica invece la frazione degli elementi appartenenti ad una determinata classe che sono stati correttamente classificati dalla rete neurale. Il parametro F-1 score rappresenta la media armonica tra i parametri di precisione e di recall.

3.2 | Risultati per un approccio CNN

3.2.1 | Risultati per SpectNet

Una prima CNN è stata proposta per consentire una classificazione della pavimentazione a partire dagli spettrogrammi in frequenze Mel del segnale. Sono state definite 32 bande in frequenza Mel nell'intervallo compreso tra i 100 e i 1000 Hz. I dati audio sono stati rappresentati in tempo-frequenza tramite una matrice numerica di 230dimensioni 32×59.

Questi dati sono stati utilizzati come input per *SpectNet*: la rete è stata addestrata su 100 epochs, con un learning rate di 0.01 e un barch size di 64. In questo caso, è stato utilizzato un algoritmo di ottimizzazione di tipo "*sgdm*" ed è stato possibile ottenere una accuratezza massima sul set di validazione pari al 71.56% (Figura 4). A causa di questa accuratezza significativamente più bassa di quella riportata per la rete LSTM, è stato ricercato un approccio alternativo alla classificazione tramite reti CNN.

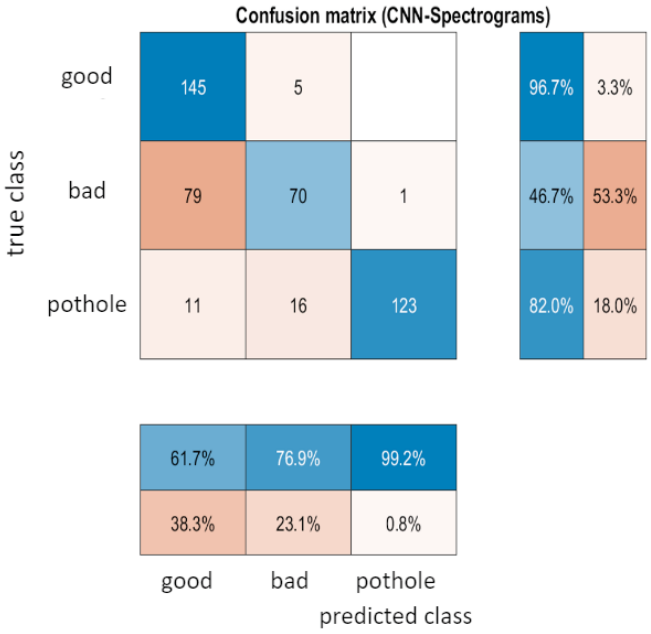

Fig. 4 – Risultati di spectNet sul data set di validazione
Results of SpecNET on the validation dataset

In Tabella 2 vengono riportati gli indici di qualità per la rete SpectNet.

Tab. 2 – Indici di qualità per la rete SpectNet (valutati su data set di validazione)
Quality indices for the SpecNet network (evaluated on the validation dataset)

	Integra	*Dissestata*	*"Pothole"*
Precisione	61.7%	76.9%	99.2%
Recall	96.7%	46.7%	82.0%
F1-Score	75.1%	58.1%	89.8%

3.2.2 | Risultati per MFCCNet

Una seconda architettura di tipo CNN è stata proposta. I segnali sono stati caratterizzati attraverso l'andamento temporale dei loro primi 13 MFCC (è stato aggiunto anche l'andamento temporale dell'energia del segnale). Ai fini della classificazione, sono stati utilizzati dati di input in formato matriciale con dimensioni 14×59.

La rete neurale è stata addestrata su un totale di 30 epochs con algoritmo di ottimizzazione di tipo "sgdm", un batch size di 64 ed un learning rate pari a 10^{-3}. In questo modo è stato possibile ottenere un'accuratezza del 94% sul validation set.

L'accuratezza di *MFCCNet* sul data set di test è invece risultata essere pari a 81.8% (Figura 5). Anche in questo caso, la performance complessiva del classificatore è stata valutata attraverso i parametri di precisione, recall e F1-score per ciascuna classe (Tabella 3).

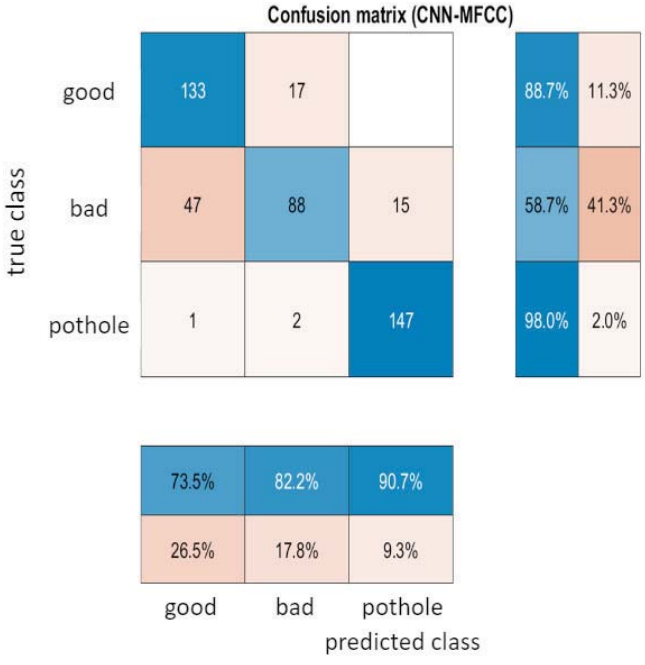

Fig. 5 – Risultati di MFCCNet sul data set di test
Results of MFCCNet on the test dataset

I risultati ottenuti utilizzando *MFCCNet* risultano simili a quelli ottenuti con il classificatore di tipo LSTM. Si noti, in particolare, l'elevata performance nel riconoscimento di segnali associati a pavimentazioni che presentano un ammaloramento di tipo "pothole".

Tab. 3 – Indici di qualità per la rete MFCCNet
Quality indices for the MFCCNet

	Integra	*Dissestata*	*"Pothole"*
Precisione	73.5%	82.2%	90.7%
Recall	88.7%	58.7%	98.0%
F1-Score	80.4%	68.5%	94.2%

4 | Conclusioni

Due dei tre classificatori proposti, la rete LSTM e *MFCCNet*, hanno fornito risultati promettenti su un data set di test (rispettivamente l'80.9% e l'81.8% di accuratezza). In particolare, è stata raggiunta un'ottima precisione nel riconoscimento dei potholes. Al momento, non sono state ancora raggiunte elevate accuratezze per le pavimentazioni in cattivo stato: è importante sottolineare che la classe relativa alle pavimentazioni in cattivo stato comprende segnali relativi a pavimentazioni che presentano diverse tipologie di ammaloramento. Si può ipotizzare che una classe così eterogenea non possa essere descritta adeguatamente mediante i processi di caratterizzazione proposti. Inoltre, è possibile che migliori performance possano essere ottenute aumentando il numero di classi, ovvero dividendo i dati in un numero maggiore di classi caratterizzate da una maggiore omogeneità (data set più specifici), e, dunque, facilmente caratterizzabili. Test su strada potrebbero essere condotti per testare l'effettiva capacità del sistema di classificare il manto stradale in tempo reale.

L'implementazione di sistemi di classificazione basati sui segnali TCN potrebbe portare in futuro alla realizzazione di protocolli efficienti ed economici per il monitoraggio e la manutenzione delle infrastrutture stradali. Inoltre, attraverso la georeferenziazione delle misure, potrebbe essere possibile sviluppare piani d'azione circoscritti a determinate aree di interesse, permettendo interventi tempestivi e riducendo inoltre i costi associati alle operazioni di bonifica stradale.

Conclusions

Two of the three classifiers proposed in this paper, the LSTM network and *MFCCNet* provided promising results over a test data set (80.9% and 81.8% accuracy, respectively). In particular, a very high accuracy has been reached in pothole recognition tasks. At the moment, high precision classification performances for degraded pavements have yet to be achieved: it is important to underline that the class associated to degraded road pavements includes signals that are linked to different kinds of road asperities. It is possible to speculate that such a differentiated class could not be adequately described through the proposed methodologies. Moreover, it might be possible to achieve better performances by increasing the number of classes, dividing the data in a greater number of classes, characterized by an increased homogeneity (more specific data sets), and, therefore, more easily describable. Road tests might be conducted in order to assess the effective classification capability of the system in a real time road condition classification task.

The implementation of such TCN-based systems could lead to the development of efficient and cost-effective protocols for the monitoring and maintenance of road infrastructures. Moreover, georeferencing the results might allow for the development of localized action plans, that could lead to timely interventions, all while decreasing the overall associated costs.

Ringraziamenti

L'autore desidera ringraziare le seguenti persone: i dottori Francesco Bianco, Stefano Carpita, Simon Kanka per il loro puntuale e preciso lavoro di revisione e i professori Gaetano Licitra e Francesco Fidecaro per la loro attenta supervisione. L'autore desidera inoltra ringraziare Ipool srl. per il supporto fornito durante le attività di ricerca e di elaborazione dei dati.

Glossario

MPD: Mean profile depth.
STM: Long short-term memory network.
CNN: Convolutional neural network.
PCA: Principal component analysis.
Adam: Adaptive moment estimator.
SGDM: Stochastic gradient descent with momentum.
MFCC: Mel frequency cepstral coefficients.
MFCCNet: Rete neurale convoluzionale che permette di classificare suoni sulla base di una rappresentazione bidimensionale dell'andamento degli MFCC.
SpectNet: Rete neurale convoluzionale che permette di classificare suoni mediante i loro spettrogrammi.

Bibliografia

[1] G. Bitelli, A. Simone, F. Girardi, C. Lantieri, Laser scanning on road pavements: A new approach for characterizing surface texture, Sensors 12 (2012) 9110-9128. https://doi.org/10.3390/s120709110.

[2] M.A. Pallas, M. Bérengier, R. Chatagnon, M. Czuka, M. Conter, M. Muirhead, Towards a model for electric vehicle noise emission in the European prediction method CNOSSOS-EU, Appl. Acoust. 113 (2016) 89-101. https://doi.org/10.1016/j.apacoust. 2016.06.012.

[3] ISO 11819-2:2017 Acoustics - Measurement of the influence of road surfaces on traffic noise - Part 2: The close-proximity method, International Organization for Standardization, Geneva, Switzerland, 2017.

[4] J. Masino, B. Daubner, M. Frey, F. Gauterin, Development of a tire cavity sound measurement system for the application of field operational tests, in: 10th Annual International Systems Conference, SysCon 2016 - Proceedings, Institute of Electrical and Electronics Engineers Inc., Orlando, FL, 2016: 7490624. https://doi.org/10.1109/SYSCON.2016.7490624.

[5] L.G. Del Pizzo, F. Bianco, A. Moro, G. Schiaffino, G. Licitra, Relationship between tyre cavity noise and road surface characteristics on low-noise pavements, Transport. Res. D-Tr. E. 98 (2021) 102971. https://doi.org/10.1016/j.trd.2021.102971.

[6] ISO 13473-3:2002 Acoustics - Characterization of pavement texture by use of surface profiles, International Organization for Standardization, Geneva, Switzerland, 2002.

[7] P. Klein, J.F. Hamet, Road texture and rolling noise: an envelopment procedure for tire-road contact, 2004, 17p. hal-00546120.

[8] A. Del Pizzo, Analysis of Tyre Rolling Noise on Low Noise Pavements, PhD Thesis, University of Pisa, Italy, 2021.

[9] J. Pinay, H. J. Unrau, F. Gauterin, Prediction of close-proximity tire-road noise from tire cavity noise measurements using a statistical approach, Appl. Acoust. 141 (2018) 293-300. https://doi.org/10.1016/j.apacoust.2018.07.023.

[10] G. Schiaffino, L. G. Del Pizzo, S. Silvestri, F. Bianco, G. Licitra, F.G. Pratico, Machine learning techniques applied to road health status recognition through tyre cavity noise analysis, J. Phys. Conf. Ser. 2162 (2022) 012011. https://doi.org/10.1088/1742-6596/2162/1/012011.

[11] Bollettino ufficiale della regione Lombardia - 1° supplemento straordinario. Allegato B, D.g.r. 25 gennaio 2006 - n. 8/1790 (in Italian).

[12] S. Hochreiter, J. Schmidhuber, Long short-term memory, Neural Comp. 9 (1997) 1735-1780. https://doi.org/ 10.1162/neco.1997.9.8.1735.

[13] Z. Li, F. Liu, W. Yang, S. Peng, J. Zhou, A survey of convolutional neural networks: analysis, applications, and prospects, IEEE T. Neur. Net. Lear. 33 (2021) 6999-7019. https://doi.org/10.1109/TNNLS.2021.3084827.

[14] K. Fukushima, S. Miyake, Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition, in: Amari, Si., Arbib, M.A. (eds) Competition and Cooperation in Neural Nets. Lecture Notes in Biomathematics, vol 45. Springer, Berlin, Heidelberg: pp. 267-285. https://doi.org/10.1007/978-3-642-46466-9_18.

[15] S. Davis, P. Mermelstein, Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences, IEEE T. Acoust. Speech. 28 (1980) 357-366. https://doi.org/10.1109/TASSP.1980.1163420.

[16] G. Peeters, A large set of audio features for sound description (similarity and classification) in the CUIDADO project, CUIDADO Project Report, 2004.

[17] A. Lerch, An introduction to audio content analysis: Applications in signal processing and music informatics, Wiley-IEEE Press, Hoboken, NJ, 2012.

[18] E. Scheirer, M. Slaney, Construction and evaluation of a robust multifeature speech/music discriminator, in: Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP. Part 1 (of 5), Institute of Electrical and Electronics Engineers, 1997: pp. 1331-1334.

[19] D.S. Wilks, Statistical methods in the atmospheric sciences, 3rd ed., Academic Press, Oxford, UK; Amsterdam, The Netherlands; Waltham, MA; San Diego, CA, 2011.

[20] S. Raschka, Python machine learning, 2nd ed., Packt Publishing Ltd, Birmingham, UK, 2015.

[21] Y. Wang, S. Ji, H. Xu, Non-stationary signals processing based on STFT, in: 2007 8th International Conference on Electronic Measurement and Instruments, ICEMI, Xi'an, China, 2007: pp. 3301-3304. https://doi.org/10.1109/ICEMI.2007.4350914.

[22] J. Pomerat, A. Segev, R. Datta, On neural network activation functions and optimizers in relation to polynomial regression,

in: Proceedings - 2019 IEEE International Conference on Big Data, Big Data 2019, Institute of Electrical and Electronics Engineers Inc., Los Angeles, CA, 2019: pp. 6183-6185. https://doi.org/10.1109/BigData47090.2019.9005674.

[23] Z. Zhang, Improved Adam optimizer for deep neural networks, in: 2018 IEEE/ACM 26th International Symposium on Quality of Service, IWQoS 2018, Institute of Electrical and Electronics Engineers Inc., Banff, Canada, 2018: pp. 1-2. https://doi.org/10.1109/IWQoS.2018.8624183.