inter.noise
21-24 AUGUST
SCOTTISH EVENT CAMPUS
GLASGOW 2022

IoT smart city framework using AI for urban sound classification

Simona Domazetovska[1]
Faculty of Mechanical engineering in Skopje
University Ss Cyril and Methodius in Skopje, North Macedonia

Damjan Pecioski[2]
Faculty of Mechanical engineering in Skopje
University Ss Cyril and Methodius in Skopje, North Macedonia

Viktor Gavriloski[3]
Faculty of Mechanical engineering in Skopje
University Ss Cyril and Methodius in Skopje, North Macedonia

Hristijan Mickoski[4]
Faculty of Mechanical engineering in Skopje
University Ss Cyril and Methodius in Skopje, North Macedonia

ABSTRACT

The advances of artificial intelligence approaches for automatically extracting and classifying disturbing sound events have great potential and application in the development of smart cities. In this paper, an urban sound event classification system based on deep learning technologies has been created by using the MEL frequency cepstral coefficients as feature extractors and the Convolutional Neural Networks as classifiers. The designed system was trained and tested using the UrbanSound8K dataset which resulted in high classification accuracy of 92.67% of the tested results. In addition to this, validation of unknown sound events was applied. Furthermore, the algorithm was implemented on a wireless sensor unit (WSU) capable of recording and classifying sounds within an urban environment and sending the classification data to the cloud. The implementation of such units could result in real-time sound classification which can be used in smart cities based on the Internet of Things technology. According to this, an IoT smart city framework using AI for urban sound classification was proposed. The created system could find its use in many applications by making contribution in creating a feasible and deployable real-time sound classification system.

[1] simona.domazetovska@mf.edu.mk

[2] damjan.pecioski@mf.edu.mk

[3] viktor.gavriloski@mf.edu.mk

[4] hristijan.mickoski@mf.edu.mk

1. INTRODUCTION

The Internet of Things (IoT) technology is a network of heterogeneous devices embedded with sensors, processing abilities and software that enables connecting and exchanging data with other devices and systems via communication networks (Wi-Fi, LoRaWAN, Bluetooth…) [1, 2]. The IoT technology enables intelligent, scalable IoT networks that process the data in real-time while reducing the operational cost providing more efficient, improved and autonomous system. The processed data is stored into online storage space services, that is further processed and shown on applications to the final users. Most applications use these storage servers to implement large scale algorithms based on the artificial intelligence (AI) technologies for natural language translation, image processing, and sound classification [3]. The modern urban areas tend to use the IoT technologies for development of smart cities that use information and communication technologies (ICT) to increase operational efficiency, share information and reduce the environmental pollution.

The environmental noise is one of the main pollutants and concerns due its negative impact on the people's wealth. According to the study in [4], the exposure to noise in urban areas caused by the heavy traffic and industries affect on 27.6 million people who live in urban areas in Europe, of which almost 45% suffer from anxiety caused by the noise pollution. Based on END [5], the European Union requires noise maps of big cities to check the noise level and create action plans for lowering and eliminating the noise exposure. In addition, the maps only quantitively describe the noise pollution, failing to recognize the class of the disturbing sound events. The classes of sound events are important as they have different subjective impact of the people's perception. For example, having loud noise in parks with a lot of people or street performers playing music are not considered as disruptive by the citizens as sound events from traffic or construction for the same noise level. This problem indicates using the sound event classification systems in the urban areas. The algorithms based on the AI technology could be a good candidate performing the task of sound event classification since the model can be trained to characterize sound noise through analyzing examples under various models and approaches.

The use of the AI-based algorithms could result in developing acoustic event detection and classification (AED/C) systems for urban sound classification, as shown in [6, 7, 8]. These technologies can be used for smart city development in the field of noise pollution by creating wireless acoustic sensor network (WASN) with low-budget sensors designed to classify the disturbing classes of sound events. Most of the WASN-based environmental noise monitoring systems are designed to continuously measure the sound levels in previously defined location, not defining the class of the sound event. However, the extraction of specific information about the sound sources presented in the acoustic environment is a key issue that must be done in order to meet the requirements of the legislation, besides allowing further detailed analyses based on the $L_{A,eq}$ computation. In [9], the researchers work on a project based on a low-cost, intelligent sensing platform capable of continuous, real-time, accurate, source-specific noise monitoring. This platform also provides an urban sound taxonomy, annotated datasets, and various cutting-edge methods for urban sound-source identification. Another application of the sound classification in the urban areas using the low-budget units is shown in [10, 11, 12], where the researchers develop systems that can measure the urban noise and recognize acoustic events with a high performance in real-life scenarios.

This paper proposes IoT smart city framework for urban sound event detection and classification through the development of deep learning algorithm that is trained and tested with labelled disturbing sounds using the UrbanSound8K dataset, and further validated by using unknown sounds. The created AED/C system to can be used for IoT smart cities development by implementing it on low-budget sensor unit that could help localize the disturbing classes of sound events.

This paper is structured as it follows: Section 2 introduces the design of the IoT smart city framework for sound event classification. Section 3 explains the applied architecture of the AED/C system by applying feature extraction, ML algorithms, testing and further validation, and cloud computing, while section 4 shows the results. Finally, section 5 presents the conclusions and the future work.

2. IOT SMART CITY FRAMEWORK FOR SOUND EVENT CLASSIFICATION

A smart city is a concept that uses ICT technologies to improve the operational efficiency of the city, share information with the public and provide better service quality. By using smart technologies and data analysis, the purpose of the smart city is to optimize the functions of the city and promote economic growth while improving the life quality of the citizens. While creating smart city, the IoT is a key technology which enables creating platform where all applications for better functioning are placed. Figure 1 shows the concept of a smart city by using various sensors for applications that help in gaining information about various events happening within the city, that can help in improvement of the environment, economy, mobility, and the life quality of the citizens.

Figure 1: Functionalities of a smart city

This paper focuses on the application of the AED/C systems for reducing the noise pollution which is known to be one of the most destructive pollutants in the urban environment that results in reducing the quality of life of the citizens. The classes of sound events that will be used in this paper are based on the research in [13], where researchers have developed a dataset and taxonomy for disturbing sounds in the urban environments, selecting 10 classes of sound events: car horn, engine idling, children playing, street music, dog barking, jack hammering, air conditioning, siren, drilling and gun shot.

The acoustic event detection and classification approaches is based on a two-main-stage process: the parameterization of the input audio—also known as feature extraction—and a machine learning approach, which is typically trained using supervised approach.

By implementing the AED/C algorithms into wireless acoustic sensor network, sound classification in large-scale can be used for smart city application [14]. This type of implementation deploys many end devices in a city framework that will be able to analyze the sounds within each area by collecting it and classifying in the belonged sound class. To design affordable and suitable WSU, according to the study in [15], the following requirements should be fulfilled:

- The device has to be built with low-cost components to create affordable sensor network consisted of several devices.
- The quality of the sensor must be provided in terms of long-term measurements.
- The processing unit has to be powerful enough for advanced calculation of the ML algorithms.
- The device should be connected to cloud for software analysis and sharing results.
- The network between the sensor units has to be able to communicate wireless.
- The final device has to be protected from outdoor conditions (e.g. water, wind).

3. ARCHITECTURE OF THE AED/C SYSTEM

The AED/C system should be able to recognize sound events through the application of AI, which relies on the idea that systems can learn from data, identify features, and make decisions with minimal human intervention. Figure 2 shows typical system for classification of sound events using supervised approach, focusing on the two main processes: feature extraction by using audio parametrization of the sound, and audio classification by applying machine learning algorithm.

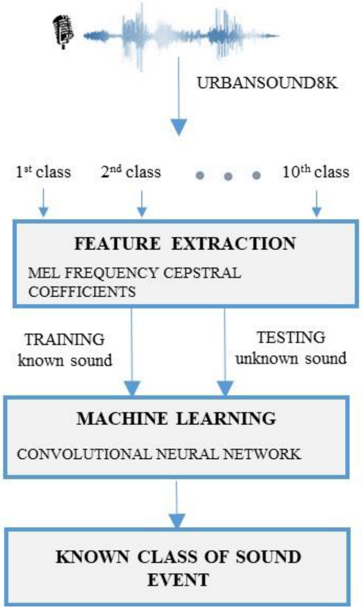

Figure 2: Architecture of acoustic event detection and classification system

When using these systems for supervised approach, first, the classes of sound events must be systemized and labeled according to sound taxonomy, which is provided by using UrbanSound8K dataset [13]. Afterwards, the chosen dataset is split into two parts for training and testing the model (90% for training, 10% for testing). Digital sound processing is applied on the audio signals by applying MEL frequency cepstral coefficients as feature extraction technique, so the audio signal of the sound event lowers its dimensionality and forms representative feature vector that is used as an input in the machine learning algorithm. The Convolutional Neural Network (CNN) is used as machine learning algorithm for training and testing the algorithm. The algorithm is trained with known sound events, and afterwards, tested with the unknown sound event. After applying this process, further validation can be applied in order to gain more information when using it in real application for urban sound classification.

3.1. Feature extraction

MEL Frequency Cepstral Coefficients (MFCCs) are used for the feature extraction process, as they are known to achieve high efficiency due to the double transformation that displays the signal in the cepstral domain, based on the MEL scale. The cepstral domain is obtained by applying two transformations, first is the Discrete Fourier transform, which serves to obtain a spectrum of frequencies, and the second, the Inverse Cosine Transform, which serves to process this spectrum into a cepstral domain. Parameters processed in the cepstral domain have high computational complexity, however, achieve high accuracy in sound event recognition and classification systems. Figure 3 shown the steps for the feature extraction process.

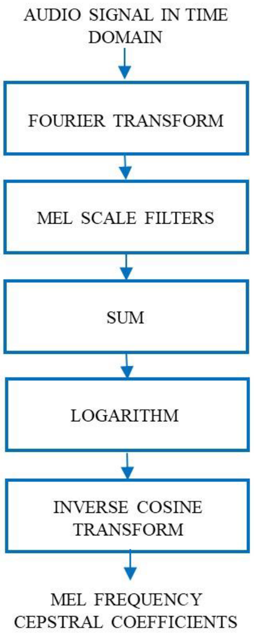

Figure 3: Steps for extracting MFCC

For the purpose of this study, 40 MFCC coefficients have been extracted, and zero padding on the signal was applied.

3.2. Classification

As the audio signals are converted into feature vectors that can be visualized as an image, a 2D convolutional neural network is used for the classification process. CNNs use relatively little pre-processing which means that the network learns to optimize the filters through automated learning, which is major advantage when applying the algorithm to WSU because of the independence from prior knowledge and human intervention. Figure 4 shows the architecture of the applied CNN model.

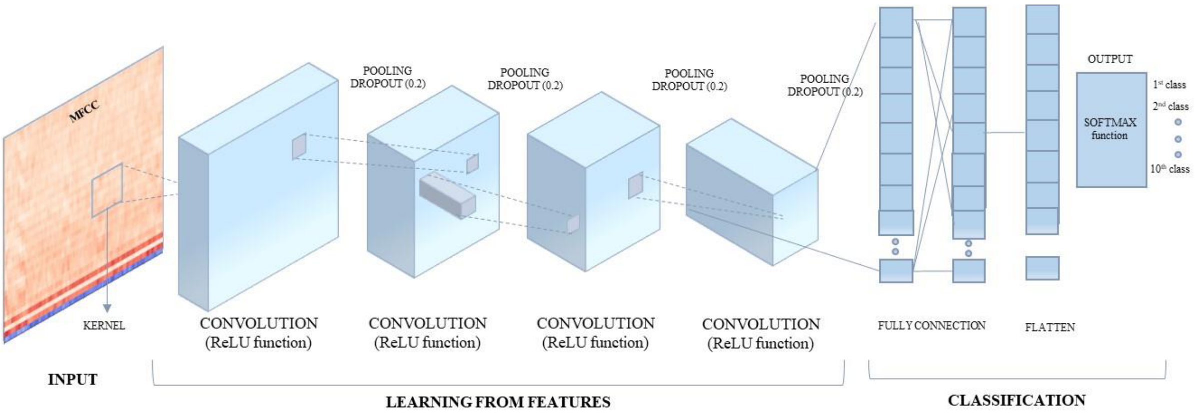

Figure 4: Architecture of the CNN

The number of filters in each layer vary from 16 to 128 filters, starting with a smaller number of filters for the first layer, up to 128 filters for the last convolution layer. It is known that the number of filters should increase when applying more convolutional layers. The kernel size is 2x2 and the ReLU activation function is used in all filters. Due to the large amount of data, the functions of pooling and dropout are applied in order to reduce the data. After the convolutional layers, a global average layer is applied, followed by the final classification layer using the SoftMax function.

3.3. Real-time recording and validation

After designing, training, and testing the AED/C system, a process of validation is applied using novel unknown sounds that are recorded in real time. Digital microphone is used for the sound recording, enabled by using the microphone wired through USB and a Python library called PySound.

The script specifies a constant recording of 4 seconds, which is enough to detect the class of the sound event. The recording is initialized by creating audio stream that uses the specifications of the digital microphone. Upon reaching the recording length, the library closes the stream, and a file handler stores the data into a WAV file. Afterwards the sound file is classified and sent to the cloud.

3.4. Cloud computing

According to [16], cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction through cloud storage technology. The cloud storage technology provides users with a means to store, retrieve, or back up their data through an online storage. The implementation of cloud storage allows instant reading of the data and eliminates the need of human on field intervention.

In this paper the cloud is only used as a storage device to work with the results taken from the system. After the initial recording of the sounds, they are processed and classified, resulting in a .csv file that shows the accuracy of the classification for the sounds. This file is then uploaded to the cloud. Using google clouds' own "Google Drive API", a connection is established between the system and a google drive storage. These files are uploaded every minute and can be accessed to see which sounds have the highest rate of appearance in the recorded time period.

4. EXPERIMENTAL RESULTS

4.1. Accuracy of the AED/C system

The accuracy of the constructed AED/C system is 97.95% while training the model that resulted in 92.68% accuracy while testing the algorithm. Figure 5 shows the confusion matrix, where comparison between the real and predicted class can be seen.

According to the results, it can be noticed that the biggest mistake occurs in the sound event children playing (84.53%) and dog barking (81.37%), where the biggest confusion for both classes is made with the sound event street music. The most precisely predicted sound event is the engine idling with an accuracy of 98.11%.

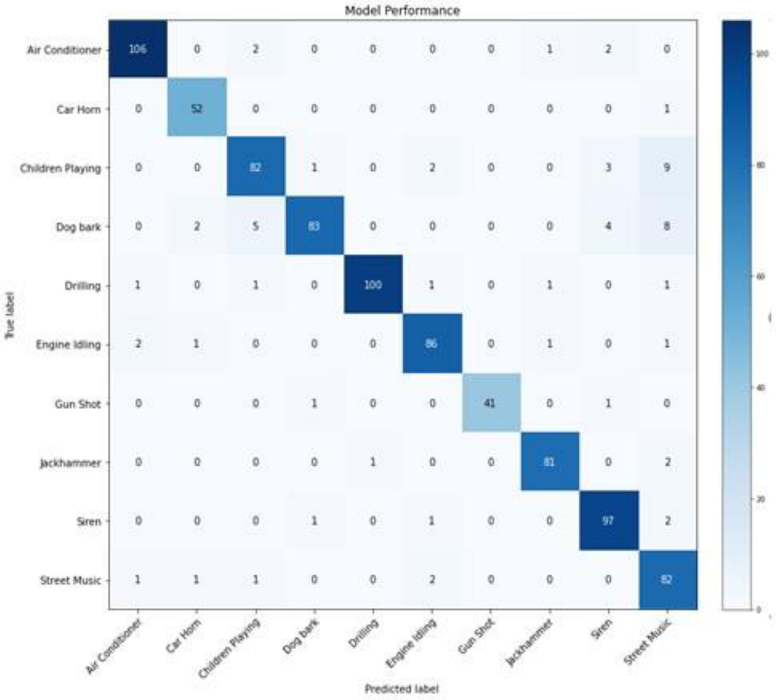

Figure 5: Confusion matrix of the tested results

4.2. Validation results

After training and testing the model, the next step is validation of the model using new real-time recorded sounds. The recording duration time is one minute, where each sound event lasts 4 seconds which leads in creating 15 sound events within one minute, and afterwards, the code is put into sleep mode for one minute. This algorithm is continuous, meaning that the process is repeating itself autonomously. After recording the sounds for one minute, the algorithm processes the results classifying them in the 10 chosen classes of sound events.

From the validation results, 10 minute recording was analyzed which provides 10 series for prediction of 15 sound events. From the analysis, it could be concluded that the accuracy for the 10 series of 15 sounds varies between 70 – 85%.

The validation accuracy of these new sounds recorded randomly within is lower than the testing results since there are times where nothing is recorded or sound which hasn't shown up in the training is recorded and the algorithm doesn't know how to process it. These results are then uploaded to the cloud where they can be analyzed and certain conclusions can be made such as streets with the most traffic, areas where most dogs are found, areas where construction is currently being done and so on.

4.3. Implementation of WSU for smart city application

The wireless sensor units are consisted of three main hardware components: microcontroller, sound recording device and external battery. Table 1 shows the parts that were used for prototyping of the wireless sensing units and the price per unit which indicates 100 € per unit to construct the low-budget sensor, while figure 6 shows the designed WSU.

Raspberry Pi is used as microcontroller, which is known to be one of the most popular and versatile board in IoT applications. With a Quad-core ARM Cortex A53 GPU as core, it works at 1.2 GHz, consuming an average of 350 mA and 1.9 W in its nominal conditions. Unlike many boards used for the same purposes, it embodies a considerably big memory solution: 1, 2 or 4 GB of SDRAM, with the possibility of attaching a microSD of any size. As for the cloud server, the design makes use of the Google Drive server. For the initial validation of the recorded sounds, a low-cost digital omnidirectional electret USB microphone is used.

Table 1: Components for construction one WSU

Part	Type	Price
Main board	Raspberry PI 3	50€
USB Microphone	OUT-AMLO-0872	10€
Box	Outdoor protection	10€
Battery	20000 mAh	20 €
Components	Cables	10 €
Total price per unit		**100€**

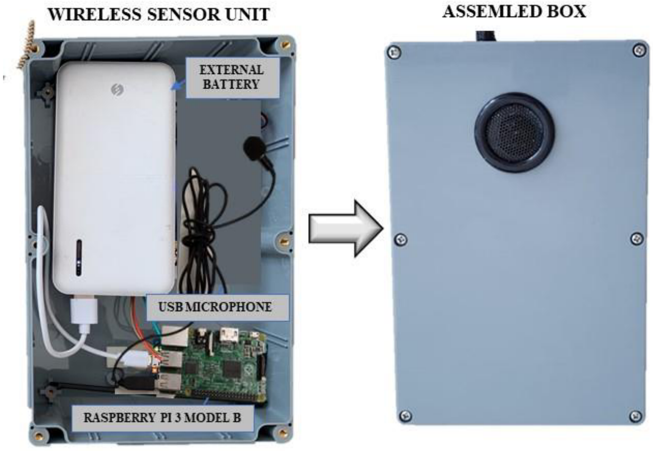

Figure 6: The designed wireless sensor unit and its components

The WSU is connected to the cloud and when the 1 minute of recording ends, the code runs giving its accuracy prediction for the recorded sounds. These results are uploaded to the cloud and later could be used to graph the most densely polluted areas and show which classification of noise pollution is detected in those areas. In this stage of the work, the code for the recording and the classification process has been transferred to the Raspberry Pi microcontroller as a proof of concept, but still needs to optimize.

5. CONCLUSIONS AND FUTURE WORK

This paper proposes a framework for implementation of acoustic event detection and classification systems in smart city applications. The use of MFCC as feature extractors and CNN as ML algorithm result in model with high accuracy, that can be further validated by recording and using unknown sound events. All the work has been transferred and tested on a wireless sensor unit in order to proof their use for sound classification.

Future work consists of testing the WSU and creating a network of several units' that would be put on chosen locations in heavy polluted areas. With the low price of the units, higher granularity can be achieved, and by continuous and autonomous data acquisition the concept of smart city and IoT can be deployed. The deployed WSUs could serve as a tool for increasing the noise awareness, allowing improvement of the city.

6. REFERENCES

1. R. Muoz, R. Vilalta, N. Yoshikane, R. Casellas, R. Martnez, T. Tsuritani, I. Morita, Integration of iot, transport sdn, and edge/- cloud computing for dynamic distribution of iot analytics and efficient use of network resources, Journal of Lightwave Technology 36 (7) (2018) 1420–1428. doi:10.1109/JLT.2018. 2800660.
2. A. Al-Fuqaha, M. Guizani, M. Mohammadi, M. Aledhari, M. Ayyash, Internet of things: A survey on enabling technologies, protocols, and applications, IEEE Communications Surveys Tutorials 17 (4) (2015) 2347–2376. doi:10.1109/COMST. 2015.2444095.
3. M. Shiraz, A. Gani, R. H. Khokhar, R. Buyya, A review on distributed application processing frameworks in smart mobile devices for mobile cloud computing, IEEE Communications Surveys Tutorials 15 (3) (2013) 1294–1313. doi:10.1109/ SURV.2012.111412.00045.
4. Noise in Europe 2017: updated assessment
5. Environmental Noise Directive END 2002/49/EC, 2002
6. Alsouda, Y., Pllana, S. and Kurti, A., 2018. A machine learning driven IoT solution for noise classification in smart cities. *arXiv preprint arXiv:1809.00238*.
7. Hernandez-Jayo, U. and Goñi, A., 2021. ZARATAMAP: Noise Characterization in the Scope of a Smart City through a Low Cost and Mobile Electronic Embedded System. *Sensors*, *21*(5), p.1707.
8. Das, J.K.; Ghosh, A.; Pal, A.K.; Dutta, S.; Chakrabarty, A. Urban Sound Classification Using Convolutional Neural Network and Long Short Term Memory Based on Multiple Features. In Proceedings of the 2020 Fourth International Conference On Intelligent Computing in Data Sciences (ICDS), Fez, Morocco, 21 October 2020; pp. 1–9
9. Bello, J.P.; Silva, C.; Nov, O.; Dubois, R.L.; Arora, A.; Salamon, J.; Doraiswamy, H. Sonyc: A System for Monitoring, Analyzing, and Mitigating Urban Noise Pollution. Commun. ACM 2019, 62, 68–77.
10. Vidaña-Vila, E., Navarro, J., Borda-Fortuny, C., Stowell, D. and Alsina-Pagès, R.M., 2020. Low-cost distributed acoustic sensor network for real-time urban sound monitoring. *Electronics*, *9*(12), p.2119.

11. Salvo, D., Piñero, G., Arce, P. and Gonzalez, A., 2020, November. A Low-cost Wireless Acoustic Sensor Network for the Classification of Urban Sounds. In *Proceedings of the 17th ACM Symposium on Performance Evaluation of Wireless Ad Hoc, Sensor, & Ubiquitous Networks* (pp. 49-55).
12. Luo, L., Qin, H., Song, X., Wang, M., Qiu, H. and Zhou, Z., 2020. Wireless Sensor Networks for Noise Measurement and Acoustic Event Recognitions in Urban Environments. *Sensors*, *20*(7), p.2093.
13. Salamon, J., Jacoby, C. and Bello, J.P., 2014, November. A dataset and taxonomy for urban sound research. In *Proceedings of the 22nd ACM international conference on Multimedia* (pp. 1041-1044).
14. Alías, F. and Alsina-Pagès, R.M., 2019. Review of wireless acoustic sensor networks for environmental noise monitoring in smart cities. *Journal of sensors*, *2019*.
15. Domazetovska, S., Anachkova, M., Gavriloski, V. and Petreski, Z., 2020, December. Wireless Acoustic Low-cost Sensor Network for Urban Noise Monitoring. In *Forum Acusticum* (pp. 677-682).
16. Peter Mell and Tim Grance, "The NIST Definition of Cloud Computing," Version 15, October 7, 2009, National Institute of Standards and Technology, Information Technology Laboratory, at http://csrc.nist.gov/ groups/SNS/cloud-computing/.