COMPUTER DESIGN SYSTEMS. THEORY AND PRACTICE

Vol. 4, No. 1, 2022

https://doi.org....

UDK 004.9+621.3.049.75

M. Melnyk[1], K. Pytel[2], M. Orynchak[2] , V. Tomyuk[1], V. Havran[1]
[1] Lviv Polytechnic National University
[2]AGH University of Science and Technology in Kraków

ANALYSIS OF ARTIFICIAL INTELLIGENCE METHODS FOR RAIL TRANSPORT TRAFFIC NOISE DETECTION

Nowadays, many cities all over the world suffer from noise pollution. Noise is an invisible danger that can cause health problems for both people and wildlife. Therefore, it is essential to estimate the environmental noise level and implement corrective measures. There are a number of noise identification techniques, and the choice of the most appropriate technique depends upon the information required and its application. Analyzing audio data requires three key aspects to be considered such as time period, amplitude, and frequency. Based on the above parameters, the source of noise can be identified.

This research paper suggests the utilization of artificial intelligence and machine learning algorithms for the traffic noise detection process. Computational methods are the fastest and most innovative way to analyze raw data sets and predict results. Identifying patterns in these methods requires a large amount of data and computing power. Machine learning models can be trained using three types of data: experimental sound libraries, audio datasets purchased from data providers, and data collected by domain experts. In the scope of the study, an experimental dataset was used to train a model that predicts the correct outcomes based on the inputs, using supervised learning. Developing an accurate model requires high-quality data input. However, incorrect data collection can cause noise in feature sets, as can human error or instrument error. Traffic sound events in the real environment do not usually occur in isolation but tend to have a significant overlap with other sound events. A part of this paper is dedicated to the problems that may arise during traffic noise detection, like incorrect data processing and data collection. It also discusses the ways to improve the quality of the input data. The study also states that the field of transport noise detection would greatly benefit from the development of a centralized railway database based on constructive railroad data, and from a centralized database with railway-specific datasets. Based on preliminary results of traffic noise analysis, modernization of the tram lines was proposed to reduce the environmental noise.

Key words: sound, AI, IoT, python, rail transport, machine learning.

INTRODUCTION

Most major cities around the world suffer from noise pollution caused by vehicular traffic. Noise is defined as a disturbing, unwanted or unpleasant sound that has a significant impact on the quality of life. For a healthy noise-free environment, it is therefore essential to have an accurate and reliable method of estimating vehicular traffic noise.

Today, we have AI and machine learning to extract insights, inaudible to human ears, from speech, voices, snoring, music, industrial and traffic noise, and other types of acoustic signals. Audio analysis is transforming, exploring, and interpreting acoustic signals recorded by digital devices. It uses a variety of technologies, including state-of-the-art deep learning algorithms, in order to understand sound data. The application of audio analysis has already gained widespread acceptance in a variety of industries, including entertainment, healthcare, and manufacturing.

Environmental sound recognition focuses on the identification of noises around us, promising a bunch of advantages for the automotive and manufacturing industries. It's vital for understanding the context of IoT applications. Audio data represents analog sounds in digital form, preserving the main properties of the original. As we know from school lessons in physics, sound is a wave of vibrations travelling through a medium like air or water and finally reaching our ears. Analyzing audio data requires three key aspects to be considered such as time period, amplitude, and frequency.

The time period is how long a certain sound lasts or, in other words, how many seconds it takes to complete one cycle of vibrations.

Amplitude is defined as the amount of sound intensity (measured in decibels) that we perceive as loudness. Rhythmic vibrations per second are referred to as frequency and are measured in Hertz (Hz). The frequency of people's behavior is expressed as a low or high pitch.

Pitch is a subjective criterion, whereas frequency is an objective criterion. The hearing range of a human is between 20 and 20.000 Hz.

Scientists claim that most people perceive as low pitch all sounds below 500 Hz — like a plane engine roar. Keep in mind that pitch is considered by humans at 2,000 Hz and higher.

Audio data file formats in general are unstructured data, like texts and images, which means they are not arranged in rows and columns. Alternatively, audio can be stored in the following formats:

- WAV or WAVE (Waveform Audio File Format) was developed by Microsoft and IBM. Particular formats leave the original sound recording uncompressed;
- AIFF (Audio Interchange File Format) format which allows to proceed uncompressed audio;
- FLAC (Free Lossless Audio Codec) – format that allow to be compressed without losing sound quality. Was created by the Xiph.Org Foundation [2];
- MP3 (MPEG-1 audio layer 3) was developed by the Fraunhofer Society in Germany and is supported globally. It's the most popular file format since it allows music to be transferred back and forth between portable devices and computers. Though MP3 compresses audio, it still offers acceptable sound quality [2].

Audio data transformation:

● waveform is a visual representation of an audio signal that reflects an amplitude and its changes over a defined period of time. The horizontal (X) axis reflects the time and vertical (Y) - the amplitude. However, it doesn't indicate audio wave frequencies [1, 2];

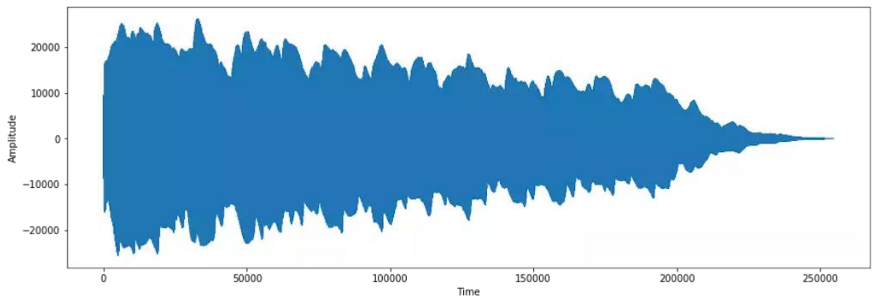

Fig 1. Simple waveform

● spectrum or spectral plot, on the other hand, represents frequencies of the sound wave on the X-axis and its amplitude on the Y-axis. Though this type of sound data visualization enables us to analyze frequency content, but misses the time component [1, 11];

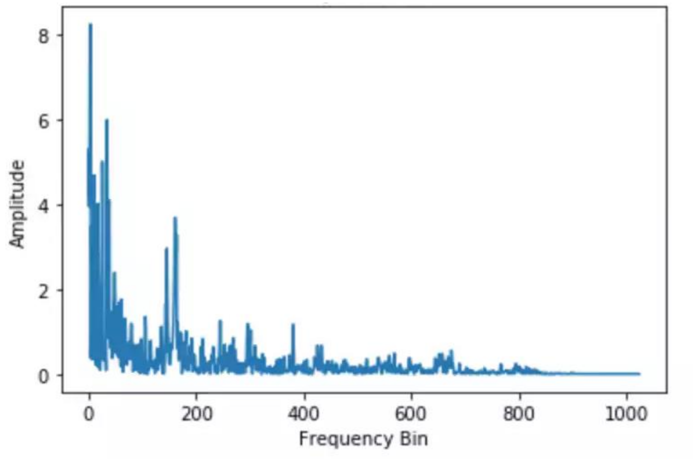

Fig 2. Spectrum plot

● spectrogram is a visual representation of a signal that covers all three characteristics of sound. The time is shown on the X-axis, frequencies on the Y-axis, and amplitude can be differed by color. The louder the event the brighter the color. When audio wave is silent, it is colored in black. Having three dimensions of the audio wave on one graph is very convenient as we may analyze the frequencies changes over time, amplitude, and spot various patterns and problem areas (like noises) on the visual graph [1, 14];

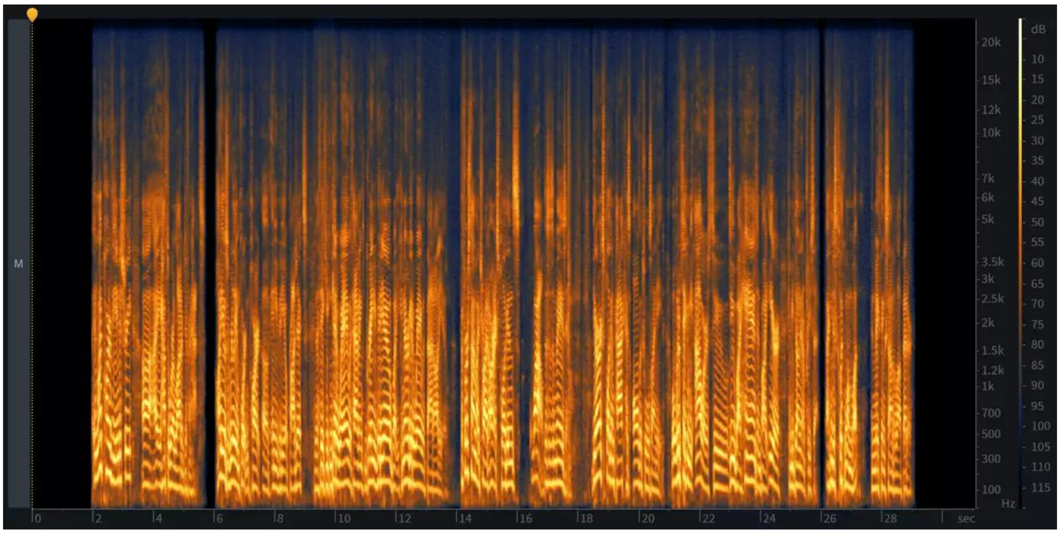

Fig 3. Spectrogram

● mel spectrogram is defined as a type of spectrogram based on the mel scale that describes how people perceive sound characteristics. On this spectrogram values in Hertz are converted into the mel scale, incorporates the unique feature of human hearing. Human ear can distinguish low frequencies better than high frequencies. Try to play tones from 500 to 1000 Hz and then from 10,000 to 10,500 Hz. Though the frequency rages are the same, the former frequency range would seem much broader than the latter. This spectrogram is widely used in

- music, for genre classification, instrument detection in songs, and human speech emotion recognition. [1, 12];

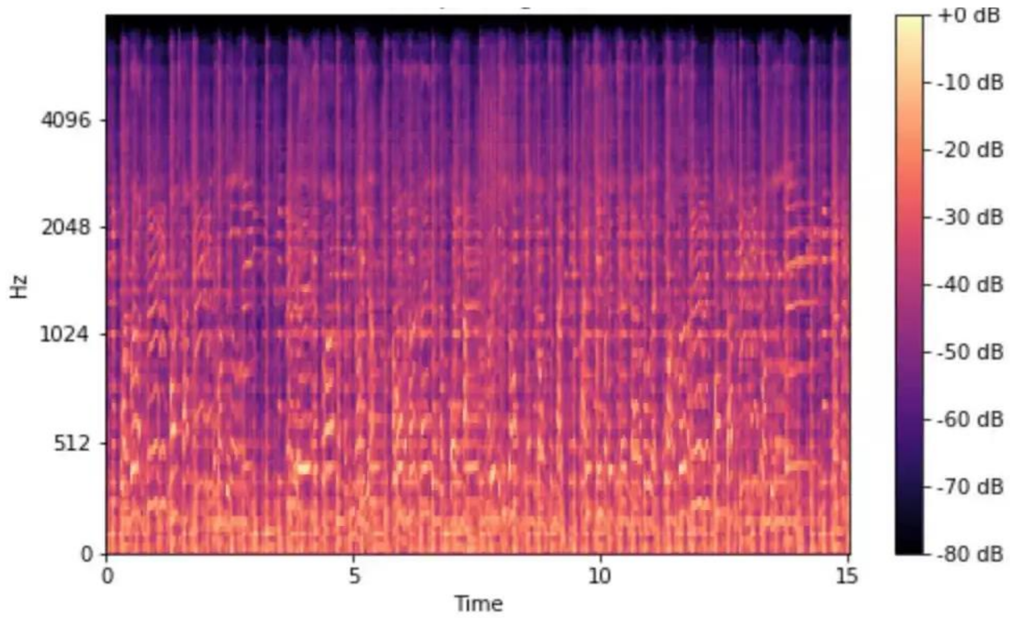

Fig 4. Mel Spectrogram

●Fourier transform (FT) is a mathematical function that decomposes a signal into frequency components, in other words, into spikes of different amplitudes and frequencies. FT is used to convert waveforms into spectrum plots, which allows to check the same signal from a different angle and perform frequency analysis. Fourier transform can be used to understand signals and troubleshoot errors in them [1,15].

●Fast Fourier Transform (FFT) is the algorithm developed for computing the described above Fourier transform [1,13].

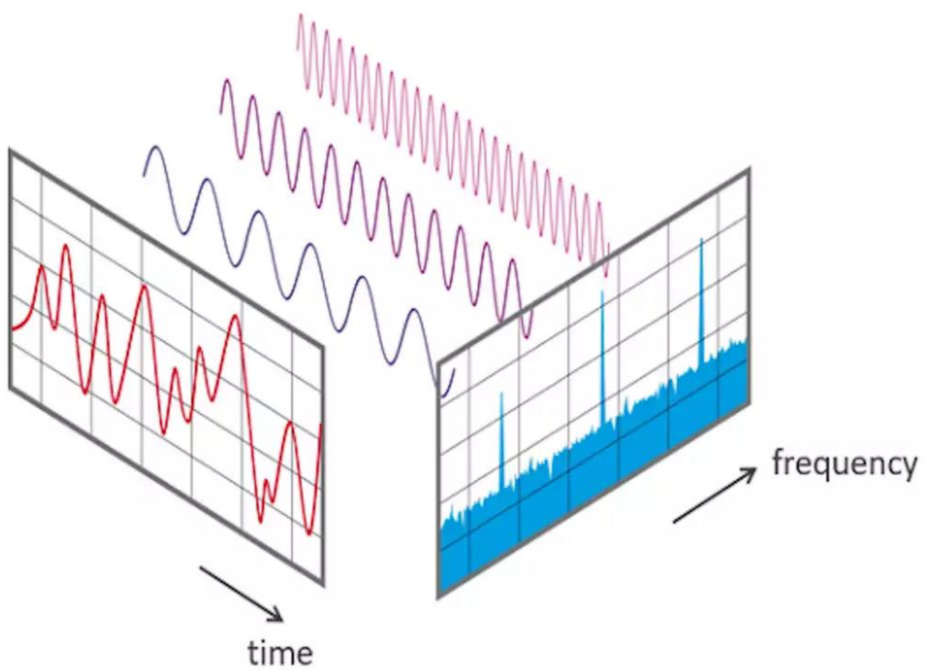

Fig 5. View of a signal in the time and frequency domain

MACHINE LEARNING NOISE DETECTION MODELING

Machine learning (ML) is a computational tool trained to automatically solve a problem instead of explicitly programming the rules. Mathematically formalized statistical models attempt to approximate the behavior of phenomena. There is a dataset available, and inputs and outputs are known in supervised learning.

ML's main goal is supervised learning with input-output data sets. The task is to use this dataset to train a model that predicts the correct outcomes based on the inputs. The image below presents the workflow to train a model using supervised learning [3].

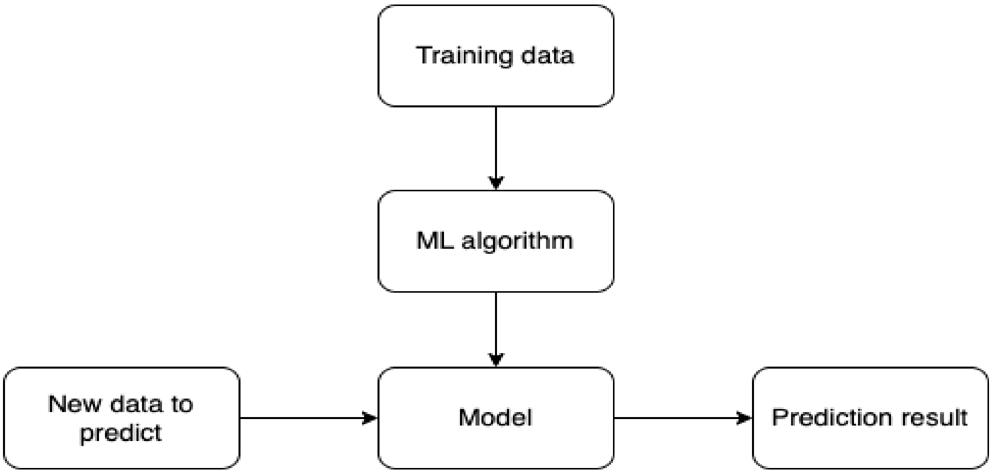

Fig 6. Generic ML methodology model

Figure 6 illustrates a generic methodology model, using which we can be guided through the noise detection process. Artificial intelligence algorithms may provide a model based on newly collected data as a starting point for ML algorithms. This model is then used as a starting point for ML algorithms that are then applied to the data. Afterwards, the algorithm finalizes the prediction result. Being engaged by endless columns of incoming data may allow us to see the calculated results. Computational methods are the fastest and most innovative way to analyze raw data sets and predict results.

Incorrect data collection can cause noise in feature sets, as can human error or instrument error. Examples of anomalies in training ML with AI here are missing values, outliers, and wrong/inconsistent formats. This is not a biconditional relationship though. Even though these effects may be caused by incorrect collection and therefore be interpreted as noise, they may also be representative of real behavior-likely exceptions-and thus be viewed as signals.

When performing exploratory data analysis, it is imperative to keep this in mind. Both interpretations could have major implications for the model's effectiveness, so you shouldn't rush to accept either.

Incorrect processing also contributes to feature set noise. The problem can either result from too much filtering, which alters the true distribution of data, or from not enough filtering, which produces irrelevant signals that confuse model learning.

A last source of noise comes from attacks, where ill-intentioned actors add intentional noise to data so they can manipulate what the model learns and, in turn, bias predictions.

SOLUTION FOR DEVELOPMENT

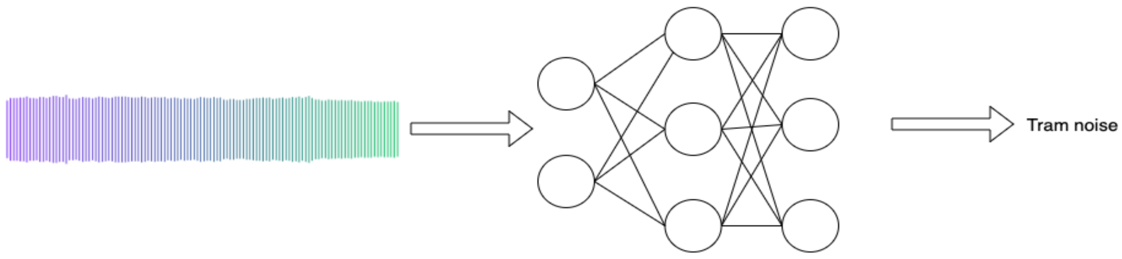

Fig. 7. Tram noise processing by AI

Sound analysis software automates all of these tasks, in most cases, by supporting the following operations:

- Import audio data
- Add annotations (labels).
- Edit recordings and split them into pieces.
- Remove noise.
- Signal conversion to visual representation (waveforms, spectrum plots, spectrograms, mel spectrograms).
- Do preprocessing operations.
- Analyze time and frequency content.
- Extract sound features and more.

Now that we have a basic understanding of sound data, let’s take a closer look at the key stages of an end-to-end audio analysis project.

1. Access standard file formats containing audio data specific to a project.
2. Utilize software tools to prepare data for your machine learning project
3. The extraction of audio features from a visual representation of a sound file.
4. Training the machine learning model on audio features.

Machine learning models can be trained using three types of data: free sound libraries, audio datasets purchased from data providers, and data collected by domain experts.

On the Net, there is an abundance of datasets, but what we cannot control is the quality and quantity of data, and how it is recorded.

Voice recordings, environment sounds, noises, and honestly everything you can think of can be found at sources like Freesound and BigSoundBank. There is the soundscape of applause, and the set with skateboard sounds, for example.

Specifically, sound libraries aren't designed for machine learning projects. In this particular case, we put additional pressure on completion, labelling, and quality control. In this particular case, we additionally put pressure by setting completion, labelling, and quality control.

Audio datasets, however, are created with particular machine-learning tasks in mind. During bioacoustics monitoring projects, the Machine Listening Lab collected more than 7,000 excerpts for its Bird Audio Detection dataset. This dataset contains 2,000 tagged audio recordings, including the ESC-50: Environmental Sound Classification dataset. There are fifty semantic classes categorized into five categories, each of which is 5 seconds long.

A large collection of audio data is available through Google's AudioSet service. This collection consists of over 2 million human-labelled 10-second sound clips from YouTube videos. Six hundred and thirty-two classes are covered in the dataset, ranging from music and speech to splinters and toothbrush noises.

COMMERCIAL AND EXPERT DATASETS

The data integrity of commercial audio sets for machine learning is better than that of free ones. Our knowledge of ProSoundEffects selling datasets will help us train models for speech recognition, environmental sound classification, audio source separation, and other applications. In total, the company has 357,000 files recorded by experts in film sound and classified into 500+ categories.

But what if the sound data you're looking for seems too specific or rare. What if you need full control of recording and labelling. Well, then better do it in partnership with reliable specialists from the same industry as your machine learning project.

A dataset such as this was obtained by taking experimental measurements from around the world and combining them with measurements from our side. This is because scientists monitor the noise tram to identify healthy noise vibration patterns. As a result, we prepared a labelled data set with about 12,000 samples of tram sound. This data set would help us in our upcoming paper to define sound prediction with the help of AI.

Besides enriching data with meaningful tags, we have to preprocess sound data to achieve better prediction accuracy. Here are the most basic steps for speech recognition and sound classification projects.

Sound is cut into short segments (frames) of the same length (usually 20-40 ms) for further segment-wise processing.

By using spectral leakage minimization, audio processing can minimize frequency smearing and amplitude distortion. There are several window functions (Hamming, Hanning, Flat Top, etc) applied to different types of signals, though the Hanning variation works well in 95 % of cases.

A window preserves the average value by reducing or smoothing its amplitude at the beginning and end of each frame.

Preprocessed audio data is visualized to produce audio features or descriptors, which may be categorized into one of three major domains:

- The time domain is represented by waveforms
- The frequency domain is represented by spectrum plots
- Spectrograms represent the time and frequency domain.

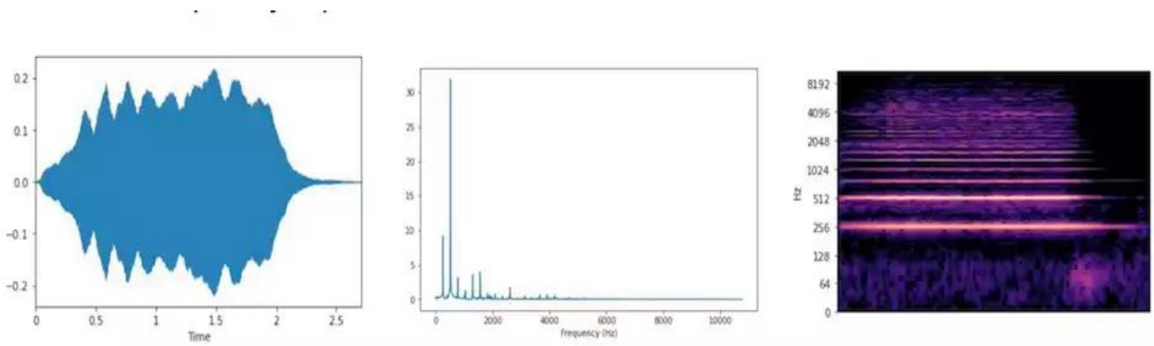

Fig. 8. Audio data visualization: waveform for time domain, spectrum for frequency domain, and spectrogram for time-and-frequency domain

The spectrogram is a visual representation of sound that combines both time and frequency components. You can get a spectrogram from a waveform by applying the short-time Fourier transform [1].

Many features can be found in the time-frequency domain, including the mel-frequency spectral coefficients (MFCCs). Due to their sensitivity to human hearing, they are based on the mel scale and the mel spectrograms.

Speech and voice recognition systems were the first applications of MFCCs, and it's not surprising that they are still used today. As well as music processing, they can also be used for acoustic diagnostics, including detecting snoring.

Based on the solution generated by ML plus AI mix, we can obtain the following results. By analyzing raw data sets, we can verify the noise plot by spectral and amplitude analyses. Then we must implement the Overlap-Add (OLA) method which prevents the loss of vital information due to windowing [4]. OLA allows 30-50 % overlap between adjacent frames, allowing them to be modified without distortion. Thus, the affected subject takes the form of a random sound pressure pattern. In the case of vibration propagation, it is a very similar issue, where mechanical vibration from the source of vibration moves through the geological environment. This process of vibration propagation is referred to as technical seismicity and is caused by traffic. As a result, the vehicle fleet is being modernized, and thus unfavorable noise and vibration emissions are being reduced. However, this process is slow and needs to be accelerated to increase efficiency over time. Replacing all vehicles is very expensive and environmentally unfriendly. Therefore, it is necessary to address the second influencing factor, and that is the technical condition of the tram line [4]. Modernizing the tram line involves not only reducing noise and vibration emissions but also degrading the materials of the tram components.

Conclusion

The main conclusion that can be drawn is that noise recognition, as a vital part of a healthy noise-free environment, can be measured using machine learning and artificial intelligence technologies. In the scope of this study the AI tram noise processing model was developed, and key stages of end-to-end audio analysis were described. Artificial intelligence algorithms provide a model based on newly collected data as a starting point for ML algorithms. One of the main concerns that was highlighted in this study is the quality of input datasets. Machine learning models can be trained using three types of data: free sound libraries, audio datasets purchased from data providers, and data collected by domain experts. A centralized unified database with railway-specific datasets would greatly contribute to the conducting of research in this area. Additionally, further investigation of the existing datasets would be required to understand their size, quality, and applicability in additional detail.

References

1. Roberts L. Understanding the Mel Spectrogram [Електронний ресурс] / Leland Roberts // Medium. – 2020. – Режим доступу до ресурсу: https://medium.com/analytics-vidhya/understanding-the-mel-spectrogram-fca2afa2ce53.
2. Audio Analysis With Machine Learning: Building AI-Fueled Sound Detection App [Електронний ресурс] // Altexsoft. – 2022. – Режим доступу до ресурсу: https://www.altexsoft.com/blog/audio-analysis/.
3. Déborah M. Python AI: How to Build a Neural Network & Make Predictions [Електронний ресурс] / Mesquita Déborah // Realpython. – 2020. – Режим доступу до ресурсу: https://realpython.com/python-ai-neural-network/#artificial-intelligence-overview.
4. Overlap-Add (OLA) STFT Processing [Електронний ресурс] // Center for Computer Research in Music and Acoustics (CCRMA). – 2022. – Режим доступу до ресурсу: https://ccrma.stanford.edu/~jos/sasp/Overlap_Add_OLA_STFT_Processing.html.
5. Zuzana Papánová, Daniel Papán, Libor Ižvolt, Peter Dobeš, Modernization of Heavy Loaded Tram Radial Effect on Noise and Vibration, Appl. Sci. 2022, 12(14), 6947; https://doi.org/10.3390/app12146947// Received: 21 March 2022 / Revised: 27 June 2022 / Accepted: 5 July 2022 / Published: 8 July 2022
6. Special Issue "Design of Track System and Railway Vehicle Dynamics Analysis" [Електронний ресурс] // MDPI. – 2022. – Режим доступу до ресурсу: https://www.mdpi.com/journal/applsci/special_issues/Track_System_Railway_Vehicle_Dynamics.
7. Morihara, T.; Yokoshima, S.; Matsumoto, Y. Effects of Noise and Vibration Due to the Hokuriku Shinkansen Railway on the Living Environment: A Socio-Acoustic Survey One Year after the Opening. Int. J. Environ. Res. Public Health 2021, 18, 7794. [Google Scholar] [CrossRef] [PubMed]
8. Maclachlan, L.; Ögren, M.; Van Kempen, E.; Hussain-Alkhateeb, L.; Persson Waye, K. Annoyance in Response to Vibrations from Railways. Int. J. Environ. Res. Public Health 2018, 15, 1887. [Google Scholar] [CrossRef] [PubMed][Green Version]

9. Lercher, P.; De Coensel, B.; Dekonink, L.; Botteldooren, D. Community Response to Multiple Sound Sources: Integrating Acoustic and Contextual Approaches in the Analysis. Int. J. Environ. Res. Public Health 2017, 14, 663. [Google Scholar] [CrossRef] [PubMed][Green Version]

10. Bao, Y.; Li, Y.; Ding, J. A Case Study of Dynamic Response Analysis and Safety Assessment for a Suspended Monorail System. Int. J. Environ. Res. Public Health 2016, 13, 1121. [Google Scholar] [CrossRef] [PubMed]

11. A Systematic Review of Artificial Intelligence Public Datasets for Railway Applications, September 2021, DOI:10.3390/infrastructures6100136, RAILS - Roadmaps for A.I. integration in the rail Sector (EU Horizon 2020 - Shift2Rail JU)

12. Understanding Spectrograms [Електронний ресурс] // iZotope. – 2020. – Режим доступу до ресурсу: https://www.izotope.com/en/learn/understanding-spectrograms.html.

13. Fast Fourier Transformation FFT - Basics [Електронний ресурс] // NTi-audio. – 2020. – Режим доступу до ресурсу: https://www.nti-audio.com/en/support/know-how/fast-fourier-transform-fft.

14. Audio Feature Extraction [Електронний ресурс] // Devopedia. – 2021. – Режим доступу до ресурсу: https://devopedia.org/audio-feature-extraction.

15. Schutz, Michael, and Jonathan M. Vaisberg. 2012. "Surveying the temporal structure of sounds used in Music Perception." Music Perception: An Interdisciplinary Journal, vol. 31, no. 3, pp. 288-296. doi: 10.1525/mp.2014.31.3.288. Accessed 2021-05-23.

М. Мельник[1], К. Питель[2], М. Оринчак[1], В. Том’юк[1], В. Гавран[1]
[1]Національний університет “Львівська політехніка”,
[2]Гірничо-металургійна академія імені Станіслава Сташиця в Кракові, Польща

АНАЛІЗ МЕТОДІВ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ВИЯВЛЕННЯ ШУМУ ВІД РУХУ РЕЙКОВОГО ТРАНСПОРТУ

Сьогодні багато міст у всьому світі страждають від шумового забруднення. Шум - це невидима небезпека, яка може спричинити проблеми зі здоров'ям як людей, так і дикої природи. Тому важливо оцінити рівень шуму в навколишньому середовищі та запровадити коригувальні заходи. Існує кілька методів ідентифікації шуму, і вибір найбільш відповідного методу залежить від необхідної інформації та її застосування. Аналіз аудіоданих вимагає врахування трьох ключових аспектів, таких як період часу, амплітуда та частота. На підставі зазначених параметрів можна визначити джерело шуму.

У цій дослідницькій статті пропонується використовувати штучний інтелект і алгоритм машинного навчання для процесу виявлення шуму що видає транспорт. Обчислювальні методи є найшвидшим і найінноваційнішим способом аналізу необроблених наборів даних і прогнозування результатів. У цих методах потрібен великий обсяг даних і обчислювальна потужність для ідентифікації шаблонів. Моделі машинного навчання можна навчити, використовуючи три типи даних: безкоштовні звукові бібліотеки, набори аудіоданих придбані в постачальників даних, і дані, зібрані експертами домену. У рамках дослідження експериментальний набір звукових даних використовувався для навчання моделі, яка передбачає правильні результати на основі вхідних даних, використовуючи контрольоване навчання. Розроблена модель вимагає якісних даних для отримання точного результату. Однак неправильний збір даних може спричинити шум у наборах даних, як і людська помилка чи помилка приладу. Звукові події дорожнього руху в реальному середовищі зазвичай не відбуваються ізольовано, а мають тенденцію накладатися на інші події. Частина цієї статті присвячена проблемам, які можуть виникнути під час виявлення шуму транспорту, наприклад неправильній обробці та збору даних, а також способам покращення якості вхідних даних. У

дослідженні також стверджується, що виявленню транспортного шуму значно посприяє розробка централізованої бази даних залізниць на основі конструктивних даних про залізницю та централізованої бази даних із спеціальними наборами даних залізниці. На основі отриманих результатів аналізу шуму транспортного руху було виявлено необхідність в модернізації трамвайних шляхів для зменшення шуму.

Ключові слова: звук, AI, IoT, python, рейковий транспорт, машинне навчання.