

# Deep Learning Approach for Traffic Congestion Sound Classification using Circular Neural Networks

Muhammad Ariq Muthi<sup>1)</sup>, Putu Harry Gunawan<sup>2)\*</sup>

<sup>1,2)</sup> CoE Hemic, School of Computing, Universitas Telkom, Bandung, Indonesia  
<sup>1)</sup>[ariqjabriq@student.telkomuniversity.ac.id](mailto:ariqjabriq@student.telkomuniversity.ac.id), <sup>2)</sup>[phgunawan@telkomuniversity.ac.id](mailto:phgunawan@telkomuniversity.ac.id)

**Submitted** : Jun 26, 2024 | **Accepted** : Jul 4, 2024 | **Published** : Jul 15, 2024

**Abstract:** Traffic congestion has become one of the main problems that occur in big cities around the world. Traffic congestion also has a negative impact if not handled seriously. Traffic congestion occurs because there is a buildup of vehicle volume that exceeds the capacity of the road. The efficiency and quality of living in cities can be negatively impacted by traffic congestion, which can also result in higher fuel consumption, pollution, and delays. There needs to be a method that can overcome and identify this. Therefore, by classifying sounds, this research aims to reduce traffic congestion. The author uses deep learning with the Convolutional Neural Network (CNN) method as the algorithm model. The model employs Mel-Frequency Cepstral Coefficients (MFCC) as the primary feature extraction technique to capture the essential characteristics of the audio signals. This research is expected to be able to classify traffic congestion sounds with good accuracy, so it can be used as a solution to overcome traffic congestion. Experiments were conducted using a training dataset, and for testing, the road sound dataset has been collected at traffic light intersections. To evaluate the proposed method, the implementation showed promising results, achieving an accuracy of 97.62% on the training data and 88.19% on the test data in classifying traffic congestion sounds.

**Keywords:** Traffic congestion, CNN, MFCC, Deep Learning, Classification

## INTRODUCTION

Traffic congestion is always one of most common problems in major cities around the world. If not addressed seriously, traffic congestion can also have negative impacts such as transportation delays, high fuel consumption, especially for four-wheeled vehicles, and greenhouse gas emissions produced by exhaust fumes. Therefore, there needs to be an effective method to monitor and detect traffic congestion by using sound classification.

Traffic congestion occurs when the number of vehicles expands beyond the maximum capacity of the road, causing traffic congestion by increasing density and disturbing the traffic flow. According to (Lu et al., 2021), many factors contribute to traffic congestion, including minimal use of public transportation, which increases the number of private vehicles. When cars are forced to stop or move very slowly, total congestion happens. Examples of traffic congestion problems can be seen in Bandung. According to (Akhmad Hermawan & Haryatiningsih, 2022), the population of Bandung increases every year due to birth rates and population migration for various purposes (education, trade, and family economic improvement). Bandung is the most congested city in Indonesia according to the Asian Development Bank (ADB), even surpassing other major cities like Surabaya and Jakarta. This is caused

\*name of corresponding author



by the large number of motor vehicles in Bandung, which increases almost every year. Traffic congestion in both urban and nonurban areas leads to wasted time and energy, increased pollution and stress, reduced productivity, increase the demand for costs on society as well as the country for spending. This problem poses an ongoing challenge for metropolitan cities, as well as medium and small cities (Kumar et al.,2021).

A method to identify traffic congestion involves classifying congestion sounds. One method that can be used is sound signal processing. The process of capturing congestion sounds by recording with a microphone. The sounds of congestion are then classified by applying deep learning to the audio samples in the (.wav) format. A branch of machine learning called deep learning is capable of learning independently. In order to process sound signals and categorize them into categories related to congestion, the author applies Convolutional Neural Network (CNN) methods as algorithm models. Multi-Layer Perceptron (MLP) is the base of Convolutional Neural Network, a machine learning technique implied for processing two-dimensional data (Tania Emanuella et al.,2021).

### LITERATURE REVIEW

In the research (Wahyuningtias, 2021), on the application of feature extraction for urban segmentation by deep learning with CNN, the objective was to demonstrate the ability of convolutional neural network (CNN) to extract noise data from environmental and urban acoustics using the Mel-Frequency Cepstral so Coefficients (MFCC) method. The results of this study demonstrated that this model performs well and can accurately predict acoustic signals. The highest audio scores included hearing air conditioning, drilling audio, hearing music in the streets, and hearing a dog, with an accuracy above 0.8.

In a study by (Ihsan & Fauzan, 2022), researchers explained how to detect audio threats using Convolutional Neural Network (CNN) method. Social media has become a favorite place to share both positive and negative content. Negative comments are often expressed as hate speech. Thus, the aim of this study was to develop a system that can detect hazards in audio using Convolutional Neural Network (CNN). The authors collected data from Twitter's website to obtain information for action. The confusion matrix technique was used as the model's evaluation tool to calculate precision, recall, and F1 scores. The audio was then converted from audio to text using Google Text-To-Speech, which produced a prediction of whether or not the audio contained a threat. The model using CNN proved to be more effective in processing information. Based on the test results, an accuracy of 70% was achieved in cases of data imbalance. However, balancing the data, shifting the denser layers, and increasing the epochs resulted in an accuracy of 82%.

In the scientific paper (Yohannes & Wijaya, 2021), the research focused on classifying the meaning of baby cry characteristics through CNN based on a combination of Mel-Frequency Cepstral Coefficients (MFCC) and Discrete Wavelet Transform (DWT) components. The researchers explained how to combine MFCC and DWT components. This study used baby cry sounds retrieved from the Donateacry dataset. In the following sections, the researchers implement the proposed model in a programming language. The MFCC and DWT feature vectors were resized to match the input layer size of the CNN. The experiment was performed with an input size corresponding to the 13 CNN layers. Following the activation process, a tendency of babies was tested in both training and test sounds, with 67% performing training and 33% for testing separation. The training and testing data covered 61 training samples and 30 test samples. Baby crying sounds were extracted with MFCC, and the audio signal was decomposed into 4 units with DWT. The aim of CNN classification was to distinguish between 4 types of baby cries characteristics based on feature data obtained from MFCC and DWT extraction processes Performance results of MFCC and DWT feature combination to detect baby cry type achieved 32.76% average accuracy , 32.63% recall, and 68.33% accuracy.

### Convolutional Neural Network

Convolutional Neural Network (CNN) is one type of algorithm in deep learning that has been used to solve problems such as voice recognition, object detection, and visual object detection (Zatarain-Cabada et al., 2018). CNN is a variation of multilayer perceptron inspired by the human neural network.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

CNN is a variation of multilayer perceptron inspired by human neural networks. CNN has three layers of a 3D structure of neurons (width, height, and depth). In the CNN model, the input  $x$  for each layer is organized into three dimensions: height, width, and depth, or  $m \times m \times r$ , where the height ( $m$ ) is equal to the width (Alzubaidi et al., 2021). Generally, in Fig. 1 the types of layers in CNN are divided into two.

**Feature learning:** The Feature learning process, commonly referred to as the extraction layer, is responsible for encoding images into features that transform the image into numerical values representing the image itself. Feature learning consists of three main layers : Convolution Layer ReLU activation, and Pooling Layer.

**Classification:** The Classification process consists of several layers responsible for classifying each neuron extracted in the feature learning process. This process generates class scores for classification. The main components of the Classification process are : Flatten Layer, Fully Connected Layer, and Softmax Layer.

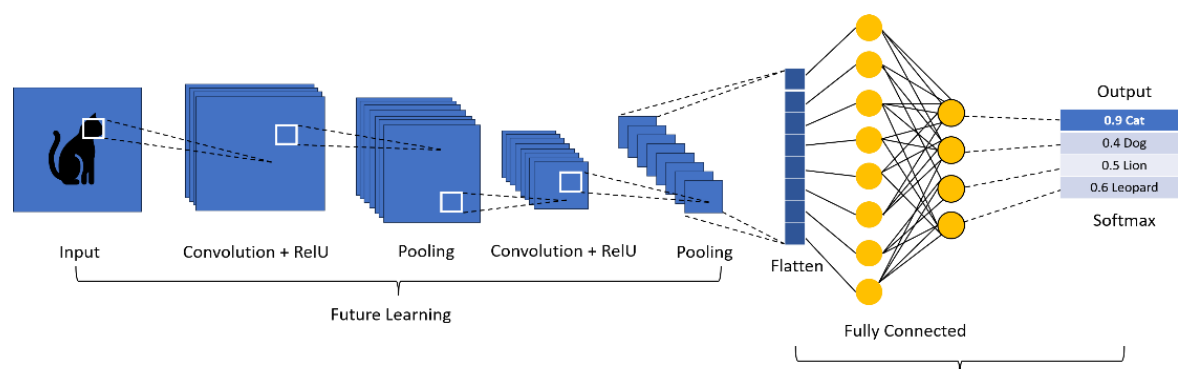


Fig. 1 Convolutional Neural Network Layer

### Deep Learning

The concept of "deep learning" describes machine learning methods that automatically learn hierarchical representations in deep architectures for classification through the use of supervised and/or unsupervised strategies. Deep learning is a class of machine learning methods that use many layers of information processing stages in hierarchical architectures to classify patterns and learn unsupervised features (Mishra & Gupta, 2017). The system learns patterns from the input data and improves also responses over time at the output. It becomes smart, smarter, and eventually smartest with continuous learning, all without human involvement (Sharma et al., 2021). There are three types of layers in deep learning, the input layer, hidden layer, and output layer. Deep learning algorithms are divided into several types, but algorithms The most frequently used is Convolutional Neural Network (CNN), Recurrent Neural Network (RNN), and Long Short-Term Memory Network (LSTM) (Ramba, 2020).

### Mel-Frequency Cepstral Coefficients

For this research, the Mel-Frequency Cepstral Coefficients (MFCC) method is being used for feature extraction because it effectively processes audio signals, enabling the extraction of characteristics present in the audio signal. MFCC is one of the most frequently implemented acoustic characteristics in speech end detection and speech recognition (Ma & Zou, 2020). According to (Limantoro et al., 2016), MFCC has several advantages: it can capture sound characteristics to recognize patterns in specific sounds, it provides output in the form of vectors with small data sizes without losing the characteristics of the extracted data, and its algorithm works similarly to how humans perceive audio signals.

\*name of corresponding author



## METHOD

In this stages, this flowchart system below will classify sounds based on traffic congestion noise on the road by comparing noisy and quiet sounds using a CNN algorithm model. Subsequently, predictions will be made based on the classification results from this model.

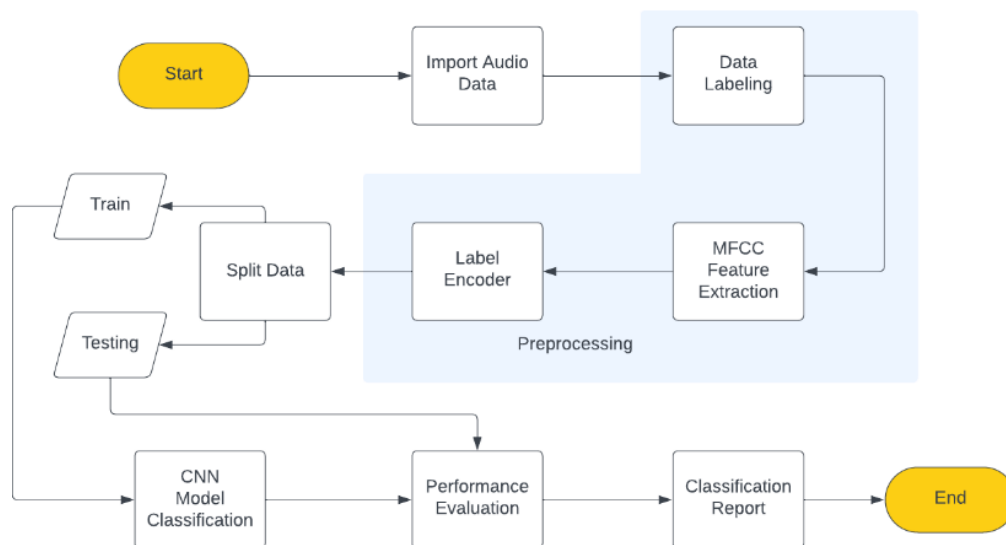


Fig. 2 System Model

## Data Collection

The data collection in this study is conducted by recording audio directly at a traffic congestion location, in Fig. 3 specifically at the traffic light intersection on Jl. Soekarno Hatta, Bandung, which is known to have the longest traffic light duration in Indonesia. The data is captured using a smartphone microphone with recordings lasting 10 to 15 minutes, which will later be segmented into 5 to 10 seconds.



Fig. 3 In Location Jl. Soekarno Hatta

## Data Labelling

Data labeling is a preprocessing stage using FFT (Fast Fourier Transform) feature extraction to measure frequencies and automatically label data by comparing the amplitude of the dataset. Fast Fourier Transform (FFT) is an improved version of the Discrete Fourier Transform (DFT) algorithm, which is used to convert a digital signal from the time domain to the frequency domain. This is a signal analysis technique which works to extract and compress specific signal features without dropping any important information (Ali et al., 2021). Below is a visualization of the obtained data set by plotting the audio signals into two categories The X-axis is amplitude, and the Y-axis is time in seconds. In Table 1 this dataset is divided into two categories: "congested" and "smooth".

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Table 1. Labeling Results

File	Label	Total FFT
siang54.wav	congested	241125
siang71.wav	congested	247116
malam101.wav	smooth	222847
malam103.wav	congested	220413
malam105.wav	smooth	220413

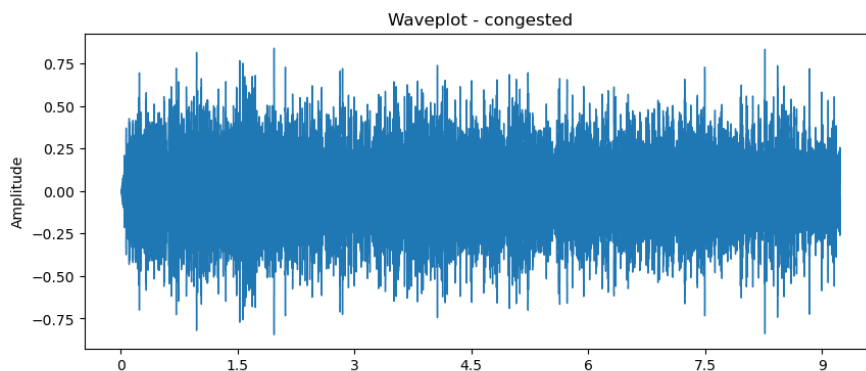


Fig. 4 congested Label

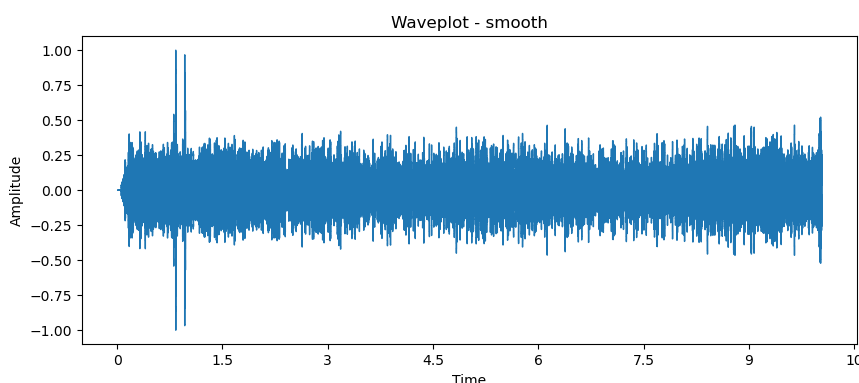


Fig. 5 smooth Label

Fig. 4 and Fig. 5 help to understand the differences in audio characteristics between congested and smooth traffic conditions. The plot for congested traffic displays higher amplitude values, indicating louder noise levels over time, and the plot for smooth traffic shows lower amplitude values, reflecting quieter noise levels over time. The amplitude variations over time provide a clear distinction that can be used for training the classification model.

Table 2 Label Description

Label	Total Sample
congested	275
smooth	205

### MFCC Feature Extraction

MFCC is an audio feature extraction commonly used in speech recognition systems. In the calculation of the MFCC, will do the audio signal in short frames. Then take power spectrum which helps in identifying the different frequencies present in audio (Lakshmi et al., 2019). Technically, MFCC can be applied using the librosa library, as you can see in Fig. 6 how MFCC visual is represented.

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

According to (Abdul et al.,2022), the MFCC can be calculated through five consecutive processes: signal framing, computing the power spectrum, applying a Mel filter bank to the power spectra, calculating the logarithm values of all filter banks, and finally applying the Discrete Cosine Transforms (DCT). The Mel scale can be calculated using the following equation –

$$mel = 2595 \times \log_{10}\left(1 + \frac{f_{hz}}{700}\right) \quad (1)$$

Description :

Equation (1) frequency measured in Hertz (f)

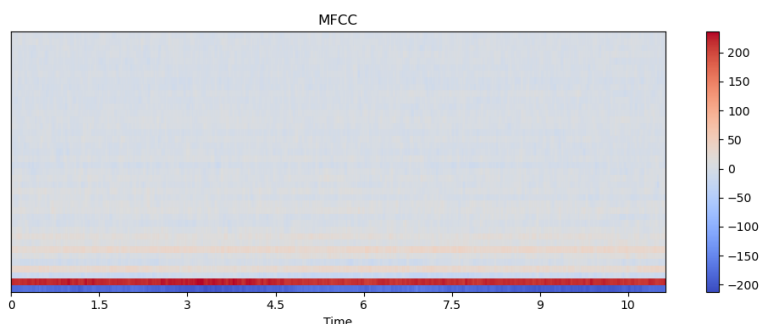


Fig. 6 Example MFCC Visual Representation

### Label Encoding

An important preprocessing step in machine learning is label encoding, label encoding is simply assigning an integer value to every possible value of a categorical variable. Label encoding is a right method for getting that list of values. Label encoding is the most simple determined technique, but it also has its uses (Hancock & Khoshgoftaar, 2020). Label encoding is used in audio data analysis to convert categorical labels of audio samples into numerical values so as to convert them into the machine-readable form which are then filtered by features from audio files such as MFCC.

### Classification Model

The data obtained during the feature extraction phase will be used for training and testing. For the training model, the dataset will be divided into 70% for training data and 30% for testing data. In Fig. 7 The classification stage will evaluate the results of the feature extraction using the CNN method to determine the accuracy score and the feasibility of the extracted features for classifying traffic congestion noise. In table 3 below, experiment parameter are used to optimize the performance of the model to be trained.

Table 3 Experiment Parameter

Label	Value
Optimizer	Adam
Epoch	20
Loss	sparse_categorical_crossentropy
Metrics	accuracy

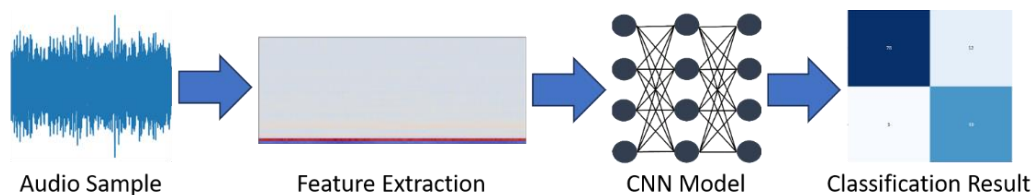


Fig. 7 Main Process

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

Table 4 CNN Model

CNN Model
<pre> model = Sequential() model.add(Conv2D(32, kernel_size=(3, 3), activation='relu', input_shape=(40, 480, 1))) model.add(MaxPooling2D(pool_size=(2, 2))) model.add(Flatten()) model.add(Dense(128, activation='relu')) model.add(Dense(64, activation='relu')) model.add(Dense(32, activation='relu')) model.add(Dense(16, activation='relu')) model.add(Dropout(0.4)) num_classes = len(np.unique(y_train)) model.add(Dense(num_classes, activation='softmax'))                     </pre>

For explanation, in Table 4, the CNN model is used to classify sound with an input size of 40x480x1, indicating that each input sample is a 40x480 matrix with a single channel and reflecting the results that have been extracted by MFCC from the audio signal. In the first layer, there is a 2D convolutional layer with 32 filter, each of size 3x3, using the ReLU activation. convolution layer operation will extract spatial features from the input matrix. Followed by the use of a max pooling layer with a 2x2 pool size used to reduce dimensions, and increase computational efficiency.

Then we move on to the stage of using the flatten layer. This layer is used to transform the 2D matrix into a 1D vector, which will be suitable for use on dense layers. After that, the dense layer, or fully connected layer, consists of 128, 64, 32, and 16 neurons, using the ReLU activation to prevent non-linear results. To prevent overfitting, a dropout layer of size 0.4 is used, which will deactivate 40% of neurons.

The final layer is dense output using Softmax activation, which will produce probabilities from both categories. This architectural model combines convolutional layers, feature extraction, and dense layers, which are suitable for handling the representation of MFCC features and for classifying audio.

### Performance Evaluation

Several metrics and methods were used to evaluate the performance of the CNN model for classifying audio recordings into congested and smooth categories. The primary evaluation metric was accuracy, which measured the ratio the correct number of predictions from the training and test datasets to those categories. After accuracy, a detailed classification report was included, providing precision, recall, and F1-score for each category. Precision refers to the ratio of true positive predictions compared to the overall positive predicted results, recall measures the ratio of true positive predictions compared to the total true positive data, and the F1-score is the reconciled precision and recall, providing a balanced measure of performance.

Furthermore, a confusion matrix in Table 5 was utilized to visualize the performance and make a classification report of the model in distinguishing between the two categories, showing the number of true positive, true negative, false positive, and false negative predictions. The confusion matrix is a measurement tool that can be used to estimate performance or accuracy in the in the classification process (Luque et al., 2019).

Table 5 Confusion Matrix

		True Class	
		Positive	Negative
Predicted Class	Positive	True Positive (TP)	True Negative (TN)
	Negative	False Positive (FP)	False Negative (FN)

\*name of corresponding author



$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (5)$$

## RESULT

### Model Performance Results

The CNN model that has been created in table 4 will be trained using the parameters in table 3 with 20 epochs using the “Adam” optimizer. After various training and testing CNN model for classifying audio recordings into "congested" and "smooth" categories demonstrated promising results. In Table 6 below, the model achieved an accuracy of 97.62% on the training data, indicating effective learning from the dataset. On the test data, the model attained an accuracy of 88.19%, showcasing its ability to generalize to unseen data.

Table 6 Model Results

Model	Training Data	Test Data	Train Accuracy	Test Accuracy
CNN	336	144	97.62%	88.18%

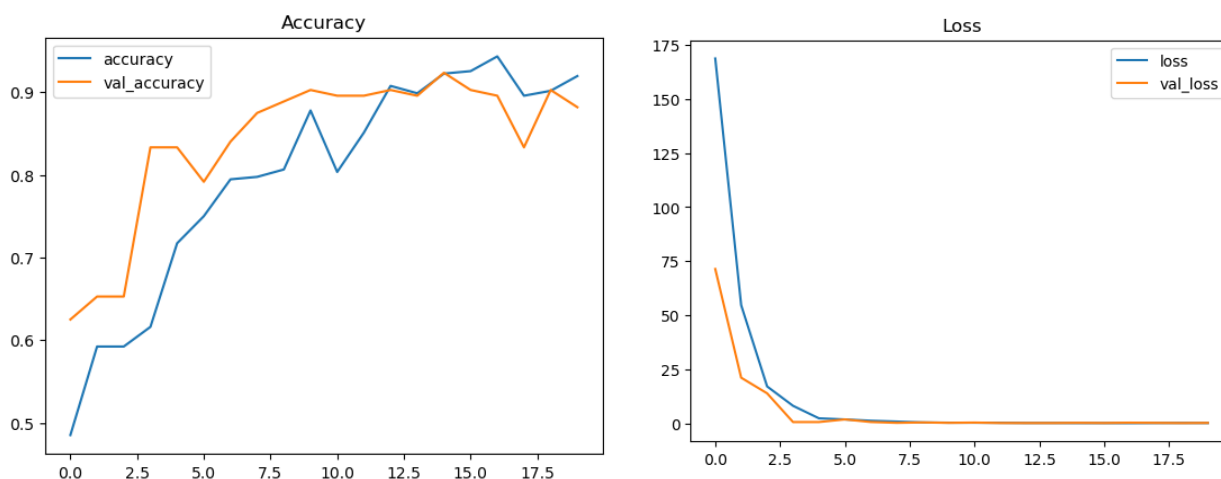


Fig. 8 Model Accuracy & Loss Result

Fig. 8 provides visualization of the results of model accuracy and loss result, which, as you can see, showed that the performance of model accuracy and val accuracy decreased slightly in the 17th epoch but still showed good results after the training. The loss results show a drastic decrease in the 2nd epoch.

### Performance Evaluation Results

In sound classification, performance evaluation like accuracy, precision, recall, and F1-score will be use as a metric to measure and determine how optimal the performance of the model. The most important



metric in this evaluation is the F1-score. The F1-score will combine precision and recall to find the balance between these two metrics.

### F1-Score Table

The performance evaluation of the CNN model in classifying traffic conditions is summarized in the F1-score table in Table 7. The model exhibits an accuracy of 0.95 and a recall of 0.86 for the "congested" category, resulting in a high F1 score of 0.90. For the "smooth" category, the model achieves a precision of 0.79 and an impressive recall of 0.93, yielding an F1 score of 0.85. Overall accuracy of the model is 88%, calculated from 144 test samples.

Table 7 F1-Score Table

	precision	recall	f1-score	support
congested	0.95	0.86	0.90	90
smooth	0.79	0.93	0.85	54
accuracy			0.88	144
macro avg	0.87	0.89	0.88	144
weighted avg	0.89	0.88	0.88	144

### Confusion Matrix

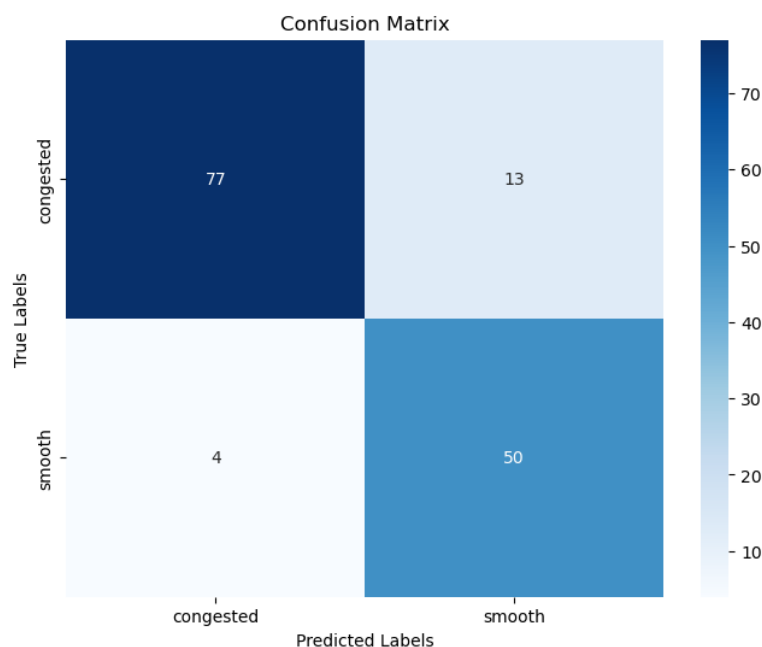


Fig. 9 Classification Report using Confusion Matrix

The confusion matrix visualizes how CNN is successful in classifying the categories of traffic conditions, which can be seen in Fig. 10 in classifying traffic conditions as either "congested" or "smooth" categories in 144 test samples. The confusion matrix classifies 77 out of 90 as congested labels. resulted in 13 false negatives in which the congested condition was misclassified as smooth. Then, in the smooth category, the confusion matrix classifies 50 out of 54 of the smooth labels. with only 4 false positives where the smooth conditions were incorrect and identified as congested.

\*name of corresponding author



## DISCUSSION

The results from the testing phase indicate that the CNN model performs well in classifying audio recordings into "congested" and "smooth" categories. The model achieved a high accuracy of 88.19% on the test data and 97.62% on the training data. Both categories have good accuracy, recall, and F1-score, with the "congested" category having particularly high accuracy 0.95 and F1-score 0.90, and the "smooth" category having strong recall 0.93. This indicates that the model is highly accurate in predicting congestion and effectively identifies smooth conditions. The confusion matrix shows that the model makes few errors, with only 13 false negatives being "congested" and only 4 false negatives being "smooth".

Accurate detection of congestion by classifying sound can provide real-time data to traffic management systems, potentially reducing transportation delays and fuel consumption, and subsequently decreasing gas emissions. These findings are particularly relevant for all other city. Especially in Bandung, a city known for heavy traffic congestion especially at traffic light intersections In Fig. 3 Jl. Soekarno Hatta, where the dataset was collected. Accurate identification and classification of traffic conditions in such critical areas can help authorities take appropriate measures to reduce traffic congestion. Despite the fact that the data used contains only 480 samples, which is relatively small for a CNN model, the model performed very well. The slight imbalance in accuracy and recall between the two groups could be addressed in future work.

## CONCLUSION

The study developed and analyzed a CNN model that efficiently classifies traffic conditions based on audio recordings. The results show that the model is effective in distinguishing between "congested" and "smooth" conditions, achieving even higher accuracy and performance metrics despite a relatively small dataset of 480 samples. A classification model can be a valuable tool for real-time traffic monitoring and management systems, aiding in the reduction of congestion and its associated risks. Future work should consider using larger datasets, advanced data augmentation techniques, and more comprehensive and scalable solutions for traffic condition monitoring based on audio data.

## REFERENCES

- Lu, J., Li, B., Li, H., & AL-Barakani, A. (2021). Expansion of city scale, traffic modes, traffic congestion, and air pollution. *Cities*, 108, 102974.
- Tania Emanuella, C., Musfita, M., & Lawi, A. (2021). Klasifikasi Suara Kucing dan Anjing Menggunakan Convolutional Neural Network. *Proceeding KONIK (Konferensi Nasional Ilmu Komputer)*, 5(1), pp. 321– 327.
- Sharma, N., Sharma, R., & Jindal, N. (2021). Machine Learning and Deep Learning Applications-A Vision. *Global Transitions Proceedings*, 2(1), 24–28. <https://doi.org/10.1016/j.gltip.2021.01.004>
- Ramba, L. S. (2020). Design Of A Voice Controlled Home Automation System Using Deep Learning Convolutional Neural Network (DL-CNN). *Telekontran : Jurnal Ilmiah Telekomunikasi, Kendali Dan Elektronika Terapan*, 8(1), 57–73. <https://doi.org/10.34010/telekontran.v8i1.3078>
- Zatarain-Cabada, R., Barron-Estrada, M. L., González-Hernández, F., & Rodriguez-Rangel, H. (2018). Emotion recognition using a convolutional neural network. In *Advances in Computational Intelligence: 16th Mexican International Conference on Artificial Intelligence, MICAI 2017, Ensenada, Mexico, October 23-28, 2017, Proceedings, Part II 16* (pp. 208-219). Springer International Publishing.
- Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., Santamaría, J., Fadhel, M. A., Al-Amidie, M., & Farhan, L. (2021). Review of deep learning: concepts, CNN

\*name of corresponding author



This is anCreative Commons License This work is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License.

- architectures, challenges, applications, future directions. *Journal of Big Data*, 8(1). <https://doi.org/10.1186/s40537-021-00444-8>
- Wahyuningtyas, V. (2021). Implementasi Ekstraksi Fitur untuk Klasifikasi Suara Urban Menggunakan Deep Learning. *Sains, Aplikasi, Komputasi dan Teknologi Informasi*, Vol 3, No 1, pp. 10-17, April 2021.
- Ihsan, F. M., & Fauzan, M. N. (2022). Identifikasi Audio Ancaman Menggunakan Metode Convolutional Neural Network. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, 10(4), 446. <https://doi.org/10.26418/justin.v10i4.52433>
- Yohannes, Y., & Wijaya, R. (2021). Klasifikasi Makna Tangisan Bayi Menggunakan CNN Berdasarkan Kombinasi Fitur MFCC dan DWT. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 8(2), 599–610. <https://doi.org/10.35957/jatisi.v8i2.470>
- Ma, Q., & Zou, Z. (2020). Traffic State Evaluation Using Traffic Noise. *IEEE Access*, 8, 120627–120646. <https://doi.org/10.1109/ACCESS.2020.3006332>
- Limantoro, W. S., Faticah, C., Umi, D., & Yuhana, L. (2016). Rancang Bangun Aplikasi Pendeteksi Suara Tangisan Bayi. *JURNAL TEKNIK ITS*, 5(2). doi:10.12962/j23373539.v5i2.17817.
- Akhmad Hermawan, R., & Haryatiningsih, R. (2022). Dampak Kemacetan di Kota Bandung bagi Pengguna Jalan. *Bandung Conference Series: Economics Studies*, 2(1). <https://doi.org/10.29313/bcses.v2i1.680>
- Lakshmi C, C., Deeksha, A. B., & Deeksha, S. (2019). *Predicting the Reason for the Baby Cry Using Machine Learning*. <https://doi.org/10.5281/zenodo.2656353>
- Abdul, Z. K., & Al-Talabani, A. K. (2022). Mel Frequency Cepstral Coefficient and its Applications: A Review. In *IEEE Access* (Vol. 10, pp. 122136–122158). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2022.3223444>
- Ali, S., Tanweer, S., Khalid, S.S., & Rao, N. (2021). Mel Frequency Cepstral Coefficient: A Review. *Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development, ICIDSSD 2020, 27-28 February 2020, Jamia Hamdard, New Delhi, India*.
- Mishra, C., & Gupta, D. L. (2017). Deep Machine Learning and Neural Networks: An Overview. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 6(2), 66. <https://doi.org/10.11591/ijai.v6.i2.pp66-73>
- Kumar, M., Kumar, K., & Das, P. (2021). Study on road traffic congestion: A review. In *Recent Trends in Communication and Electronics* (pp. 230–240). CRC Press. <https://doi.org/10.1201/9781003193838-43>
- Hancock, J. T., & Khoshgoftaar, T. M. (2020). Survey on categorical data for neural networks. *Journal of Big Data*, 7(1). <https://doi.org/10.1186/s40537-020-00305-w>
- Luque, A., Carrasco, A., Martín, A., & de las Heras, A. (2019). The impact of class imbalance in classification performance metrics based on the binary confusion matrix. *Pattern Recognition*, 91, 216–231. <https://doi.org/10.1016/j.patcog.2019.02.023>