Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Diagnostic on Car Internal Combustion Engine through Noise

William Surya Sjah[1], Ben Rahman[2)*], Djarot Hindarto[3], Alessandro Benito Putra Bayu Wedha[4]
[1,4] Bina Nusantara (BINUS ASO), Indonesia
[2,3] Fakultas Teknologi Komunikasi dan Informatika, Universitas Nasional, Indonesia
[1] William.sjah@binus.ac.id, [2] ben.rahman@civitas.unas.ac.id, [3] djarot.hindarto@civitas.unas.ac.id, [4] alessandro.wedha@binus.ac.id

Submitted : Apr 1, 2023 | **Accepted** : Apr 14, 2023 | **Published** : May 2, 2023

Abstract: Internal Combustion Engines are known for their unique sound characteristics. Through these sound characteristics, an experienced car mechanic will be able to diagnose the type of engine damage just by listening to the sound. This reduces the need to disassemble components to pinpoint machine faults which also contributes to a significant reduction in overall repair time. The main aim of this paper is to build a process to identify faulty machines through engine noise analysis with visual data to determine machine faults at an early stage. By capturing various types of engine sounds, data visualization uses healthy engine sounds and broken engine sounds obtained from cars as well as various types of broken engine sounds that are usually found in vehicles. This audio data will be used in audio signal processing combined with a linear regression classification algorithm. Visualization data can distinguish various types of sounds that are commonly found in damaged or damaged engines such as clicks, ticks, knocks and other types of sounds to determine the types of damage that are usually found in internal combustion engines. The data used comes from Kaggle, which is public data which is widely used as general data for data science activities. Visually, data from vehicle engines can be seen from the data on which car brand is the best in terms of sound. The results using linear regression show the R-squared score (R^2) or also called the coefficient of determination reaching 91.95%.

Keywords: Noise; Internal Combustion Engine; Diagnostics; Car Sounds

INTRODUCTION

Internal Combustion engines are a critical part of a modern car in the fact that it is still the most dominant and commonly found type of Power source for a car on the road according to World Energy Outlook 2022 by the International Energy Agency (IEA) (Li et al., 2022), (Alam et al., 2023), (Traivivatana et al., 2017). In fact, it's been the dominant source of power in automobiles for over a century, providing reliable and efficient power to a vehicle. However, just like any other machines, these engines will wear and age the more they are used and are subjected to varying driving conditions. Therefore, they can develop faults or malfunctions over time that can affect their performance and reliability. Because of this, it is important to detect malfunctions of an engine. Some of these faults or malfunctions can be diagnosed through the sound of the engine. A defective engine usually has a peculiar sound or noise when compared to its healthy state hence why it is possible to diagnose a car's engine from its sound. There are many types of engine fault that can be detected through sounds. Some of these common engine faults and their associated noises include:

*name of corresponding author

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Piston slap: a rapid slapping sound that can be heard from the engine. This noise is caused by an excessive amount of clearance between the piston skirt and cylinder wall. Connecting rod Noise: a knocking sound usually heard towards the bottom of the engine. Like piston slap, the noise is causing excessive amount of clearance between the crankshaft and connecting rod. Misfiring engine (Zheng et al., 2019), (Hindarto & Santoso, 2022), (Liu et al., 2011): a misfiring engine can produce a popping or knocking sound that occurs irregularly. This noise is often caused by a lack of combustion in one or more cylinders. Valve train noise: a clicking sound that usually quiets down as you raise the engine Revolutions Per Minute (RPM). Usually caused by a worn or sticking hydraulic lifters. Timing belt or chain noise: timing belt or chain noises that produce a high-pitched whining noise that is synchronized with the engine's RPMs. This noise is often caused by a loose or worn-out timing belt or chain. Through audio signal processing combined with classification algorithms (Carpinteiro et al., 2023), (Huang et al., 2023). This paper proposed a method to internal combustion engine diagnostic through noise, with a focus on its applications in cars. These noises can be diagnosed and classified to help diagnose an engine and determine which part of the engine is at fault without the need of taking apart the engine to find the source of a faulty engine which greatly reduce the time to repair a faulty engine since it is able to determine its fault at an early stage.

Data visualization (Ivanov et al., 2020), (Wedha et al., 2022) is the process of creating graphs, charts, or other visual illustrations to represent data or information. The goal is to make data easier for humans to understand and interpret. Data visualization can be used to describe patterns and trends in data, show comparisons between multiple data sets or categories, highlight anomalies or extreme values, and illustrate relationships and relationships between data. Examples of the types of data visualization include bar charts, pie charts, line charts, heat maps, and more. Visualization data (Hindarto & Handri Santoso, 2021) can be used in a variety of fields, including business, academic, healthcare, technology, and more. Effective data visualization should be easy to read, clear and easily understood by different audiences. This can be achieved by using the right colors, layouts and chart types for the data being displayed.

As a tool, data visualization (Dwyer et al., 2020), (Hindarto et al., 2021), (Wedha, 2022) can assist in diagnostics in several ways, depending on the type of data and the purpose of the analysis. In the health sector, data visualization can be used to assist doctors in diagnosing diseases and monitoring patient conditions. For example, a graph or chart showing changes in blood sugar or blood pressure levels over time can help doctors spot trends or anomalies and take appropriate action. In the business field, data visualization can assist in analyzing business performance and spotting problems or opportunities. For example, charts showing monthly sales or performance comparisons between business branches can assist managers in making strategic decisions and identifying areas for improvement. However, it is important to remember that data visualization is only a tool, and the results of the analysis must be considered in a wider context and with a deep understanding of the data used. In addition, any diagnosis or decision should be made by a suitably trained and qualified professional.

The following are some research questions that can help in diagnosing a car's internal combustion engine through sound and visualization data:

How can visualization data be used to assist in the diagnosis of car internal combustion engines through sound, and what types of visualization data are most effective? (RQ 1).

How to process and analyze sound data from a car's internal combustion engine, and convert it into visualization data that can be used in diagnosis? (RQ 2).

What kinds of algorithms or techniques are used in sound data processing and data visualization in car internal combustion engines, and what are their advantages and disadvantages? (RQ 3).

How to compare visualization data from a healthy car's internal combustion engine and a problem engine, and how to identify significant differences between the two? (RQ 4).

LITERATURE REVIEW

There have been many studies that discuss Linear Regression as a prediction. Linear regression model using Bayesian approach for energy performance of residential building (Permai & Tanty, 2018). In the statistics there are two types of points of view, Frequentist and Bayesian. The difference

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

between Frequentist and Bayesian is the point of view in terms of looking at a parameter. Bayesian views a parameter as a random variable, it means the value is not a single value. The modeling method that is most commonly used by researchers is the linear regression model. The Frequentist methods that are often used in linear regression are Ordinary Least Square (OLS) and Maximum Likelihood Estimation. An analysis of the impact of driving time on the driver's behavior using probe car data (Hyodo et al., 2017). Driver fatigue is an important factor in traffic accidents. Therefore, it is expected to develop a fatigue detection system that is less burdensome for the driver and encourages the driver to take a rest at an appropriate time after many hours of driving when driver fatigue is detected. In order to provide timely warnings, it is necessary to obtain information about the effect of hours of continuous driving on driver behavior. Identifying and modelling changes in chemical properties of engine oils by use of infrared spectroscopy (Wolak et al., 2021). The aim of this paper is to describe the interdependence of the FTIR surface characteristics spectral bands that reflect the number of chemical substances produced by the breakdown of oil in engines and car mileage depends on the oil used. The motor oil tested was used in 12 similar vehicles operated under the same conditions and attempts were made to describe the differences in kinetics changes in the group of oils studied and at the same time provide possible explanations of their causes to study the relationship between individual spectral bands. Mathematical model based on spectral peak area. Performance map of a LPG-diesel dual-fuel engine based on experimental and non-linear least squares determined Wiebe function (Vidal et al., 2021). Double engines that combine low pollution levels the properties of natural gas and diesel performance offer interesting perspectives in that case. Modern engines have to be experimentally designed and verified cards are created that pay a certain price. The model is validated with an error of 2% compared to the experimental data. Map made of primary fuel mass fraction z, equivalence ratio and the number of engine revolutions as a variable parameter shows a decrease in the specific cost of z and speed increase.

METHOD

Car diagnosis from engine sound is the process of identifying the source of a problem or the cause of an abnormal sound in a car engine through examining the sound produced by the engine. An abnormal engine sound can be an indication of a problem with the car' s engine and can indicate a variety of problems, such as damage to engine components, worn or loose engine parts, or problems with the fuel or cooling system.

Some of the steps that can be taken in car diagnostics from the sound of the engine include:

1. Listen carefully to the sound of the car engine and try to identify any abnormal sounds that appear, such as a rattling, rattling, or hissing sound.
2. Check the condition of the engine parts, such as engine oil, cooling water, fuel, and exhaust system, to make sure no problems are occurring.
3. Inspect engine components such as spark plugs, air filter, alternator, and transmission system to ensure that they are not damaged or worn.
4. Perform diagnostic tests such as checks of engine temperature, pressure, and other performance parameters to help identify engine problems.
5. In some cases, data processing or visualization can be used to assist technicians in interpreting and analyzing the sound data obtained.

After the problem or cause of abnormal sound is identified, the technician can repair or replace the problematic engine component to restore optimal performance from the car engine.

On-board diagnostics (OBD) is a vehicle self-diagnostics system designed to monitor and report real-time vehicle condition and performance. This system consists of sensors installed in the vehicle to gather information about various subsystems of the vehicle such as the engine, transmission, fuel system, emission system and other systems. The data collected is then analyzed and decoded by the vehicle's computer to determine whether the vehicle has a problem or not. The OBD system is also equipped with fault codes that allow repair technicians to identify problems quickly and accurately. When a problem occurs with the vehicle, the OBD system will generate an error code which can be

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

read using the OBD scanner. This error code will provide an indication of the problem with the vehicle and assist the repair technician in determining the required repair.

The linear regression algorithm is a method in machine learning that is used to predict the output value based on the given input value. This method is used to build a linear predictive model, namely a model that assumes a linear relationship between input and output. The linear regression algorithm tries to find the best-fit line that can represent the relationship between input and output. This best line is defined as the line that has the smallest prediction error between the output value predicted by the model and the actual output value. There are two types of linear regression:

Simple linear regression: Used when there is only one input variable and one output variable. Simple linear regression is a statistical method used to model a linear relationship between two variables, namely the independent variable (x) and the dependent variable (y). The purpose of simple linear regression is to find the best straight line that can describe the linear relationship between the two variables. This straight line can be used to predict the value of y based on the value of x.
Mathematically, simple linear regression can be expressed as follows:

$$y = \beta 0 + \beta 1x + \varepsilon \qquad (1)$$

Where:

y is the dependent variable (the variable you want to predict)

x is the independent variable (the variable used to predict)

β0 is the intercept or y value when x=0

β1 is the regression coefficient or the magnitude of the change in y when x increases by one unit.

ε is the error or difference between the actual y value and the y value predicted by the model.

To get the best straight line, we use the least squares regression method. This method looks for values of β0 and β1 that can minimize the sum of the squares of the differences between the actual y values and the y values predicted by the model. By using least squares regression, we can get the best straight-line equation as follows:

$$\hat{y} = \beta 0 + \beta 1x \qquad (2)$$

Where:

ŷ is the y value predicted by the model

β0 and β1 are intercept values and regression coefficients obtained from least squares regression.

To evaluate model performance, we can use several metrics such as R-squared (coefficient of determination) and Mean Squared Error (MSE). R-squared is used to measure how well the model can explain the variation of the data, while MSE is used to measure the difference between the actual y value and the y value predicted by the model.

Multiple linear regression: Used when there is more than one input variable and one output variable. Multiple linear regression is a regression analysis technique used to examine the relationship between two or more independent variables with one dependent variable. The main goal of multiple linear regression is to predict the value of the dependent variable based on the values of a given independent variable.
Mathematically, multiple linear regression can be written as:

$$Y = \beta 0 + \beta 1X1 + \beta 2X2 + \ldots + \beta nXn + \varepsilon \qquad (3)$$

Where:

Y is the dependent variable that you want to predict.

X1, X2, ..., Xn are independent variables.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

β0 is a constant or intercept.
β1, β2, ..., βn are regression coefficients which show how much influence the independent variables have on the dependent variable
ε is the error or random error

The purpose of multiple linear regression analysis is to determine the best β0, β1, β2, ..., βn values that can be used to predict the value of Y. This process is done by minimizing the sum of squares between the predicted value and the actual observed value (sum of squared residuals). There are several assumptions that must be met so that the results of multiple linear regression analysis can be considered valid, including:

1. The relationship between the independent and dependent variables is linear.
2. There is no multicollinearity between the independent variables.
3. The residuals have a normal distribution.
4. Homoscedasticity or uniformity of residuals across the range of independent variable values.

If these assumptions are not met, then the results of the multiple linear regression analysis cannot be considered valid and additional analysis must be carried out to improve the model.
The linear regression algorithm has the advantages of being easy to understand and fast in making predictions. However, the weakness of this method is the assumption that the relationship between input and output is linear, so it is not suitable for use in cases that have a non-linear relationship between input and output.

RESULT

Figure 1. Linear Regression Model (Linear Regression) is a statistical model used to model the relationship between one or more independent variables (X) and the dependent variable (Y) with the assumption that the relationship can be expressed linearly. The goal of the Linear Regression model is to study the straight-line equation that best describes the relationship between the X and Y variables. Mathematically, the Linear Regression model can be stated as follows:

$$Y = a + bX \quad (4)$$

Where:
Y is the dependent (or target) variable to be predicted.
X is the independent variable (or feature) used to predict Y.
a is the intercept or constant.
b is the regression coefficient, namely how much influence X has on Y.

To find the best a and b values, the Linear Regression model uses the Least Squares method. This method tries to minimize the sum of the squares of the difference between the predicted value and the observed value (actual value) of the dependent variable.

The Linear Regression model has several advantages and disadvantages. The advantage is that this model is easy to interpret and can be used to predict continuous values. In addition, this model can be used to identify independent variables that have a significant influence on the dependent variable. However, the Linear Regression model also has a weakness, namely this model can only produce linear predictions. In addition, this model is sensitive to outliers and basic assumptions must be met, such as the linear relationship between the independent and dependent variables and the assumptions of normality and residual homogeneity.

```
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score, accuracy_score
```

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

```
# Select the columns for the independent variables
X = df[['Engine Coolant Temperature(°C)', 'Engine RPM(rpm)',
    'Intake Air Temperature(°C)', 'Engine Load(%)', 'Throttle
Position(Manifold)(%)']]

# Select the column for the dependent variable
y = df['Mass Air Flow Rate(g/s)']
# Split the data into a training set and a test set
# I did a 80-20 split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=0)
# Fit a linear regression model to the data
model = LinearRegression()
model.fit(X_train, y_train)
# Make predictions on the test set
y_pred = model.predict(X_test)

# Evaluate the model
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)
print("Mean Squared Error:", mse)
print("R^2 Score:", r2, "\n")
# Create a DataFrame to store the actual and predicted values
results = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})

# Print the first 20 rows of the results DataFrame
print(results.head(20))

Output
Mean Squared Error: 13.472625318164965
R^2 Score: 0.919588666212247

        Actual      Predicted
838    2.700000     2.815943
632    3.680000     3.852193
997   28.790001    23.150110
913    2.320000     3.418839
795    2.650000     1.775174
581    4.100000     4.460151
665    4.000000     3.505610
241    2.750000     2.481916
1061  17.510000    17.170825
569    3.530000     4.991427
1070   3.760000     5.992472
322    2.510000     3.127746
299   31.930000    44.687881
438   33.369999    31.826858
```

Figure 1. Prediction using Linear Regression

Figure 2. Actual and predicted values plot is a visualization technique used to compare the actual and predicted values by the model. This helps in evaluating the performance of the model and seeing how close the predicted results are to the actual values. To plot the actual and predicted values, we must

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

first collect the actual and predicted values by the model for each data point. Then, we can plot the actual value on the x-axis and the predicted value on the y-axis and compare the two plots to see how close the points are. If the actual plot and the predicted values lie around the diagonal line, it means that our model has a good performance in predicting the values. However, if the plots are widely distributed and irregular, it indicates that our model is not performing well at predicting values. In plotting actual and predicted values, we can also add regression lines to see common patterns between the actual values and the values predicted by the model. These lines can help us see general patterns in the data and whether there is a linear relationship between the actual value and the predicted value.

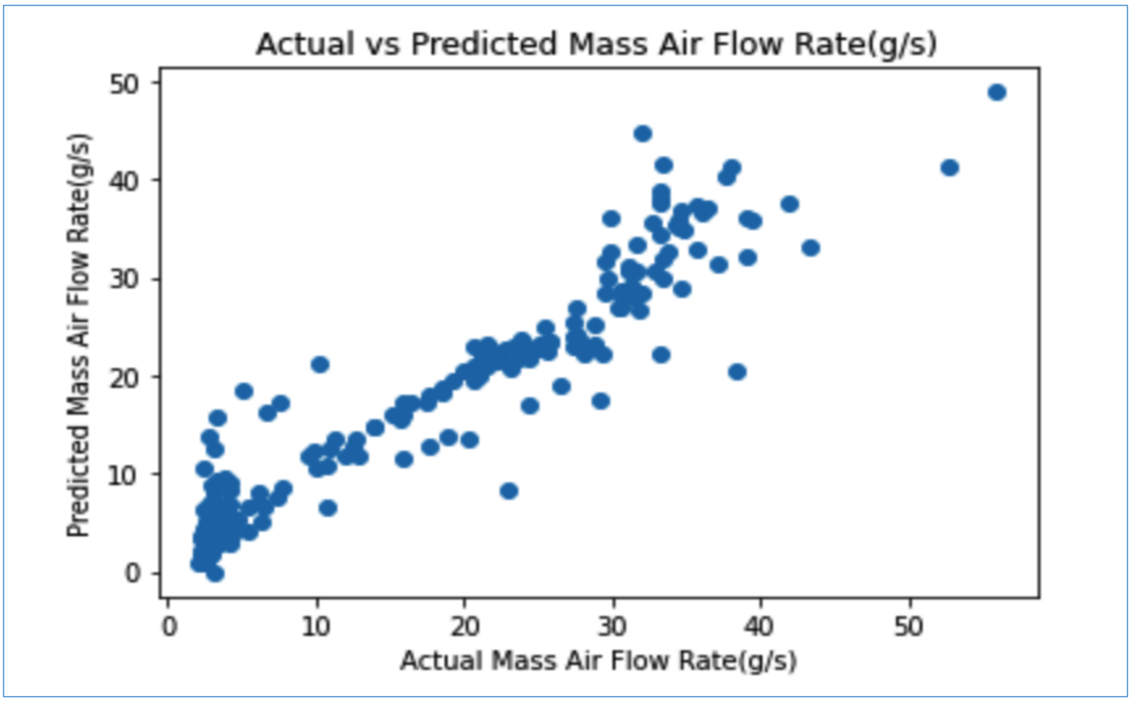

Figure 2. Plot model with Linear Regression

DISCUSSIONS

This section explains the questions from the Research Question contained in the introduction section. Here are some research questions and answers to research questions.

How can visualization data be used to assist in the diagnosis of car internal combustion engines through sound, and what types of visualization data are most effective? (RQ 1).

Visualization data can assist in the diagnosis of a car's internal combustion engine through sound by converting the sound data into a visual form that is easier to understand and analyze. By converting sound data into visuals, we can see certain patterns or tendencies in machines that cannot be seen by listening to sound alone. The type of data visualization that is most effective in assisting the diagnosis of a car's internal combustion engine by sound depends on the type of sound one wants to analyze. Several types of visualization data that can be used in the analysis of the sound of a car's internal combustion engine include:

1. Spectrogram: A spectrogram is a visual representation of the frequency spectrum of sound. The spectrogram shows the sound intensity at each frequency over time. In the diagnosis of a car's internal combustion engine, a spectrogram can be used to identify the sound frequency produced by the engine and compare it with normal data, to identify abnormal sounds. A spectrogram is a form of visualization of the sound frequency spectrum that shows the sound intensity at each frequency over time. The spectrogram is usually depicted as a 2D graph, with the x-axis showing time and the y-axis showing sound frequency. The brighter the color on the spectrogram indicates the higher the intensity of the sound at the appropriate frequency. In the diagnosis of a car's internal combustion engine, a spectrogram can help identify the sound frequencies produced by the engine and compare

*name of corresponding author

them with normal data to identify abnormal sounds. Every problem with the machine will produce sound with different frequency characteristics, so the spectrogram can help identify the type of problem that arises in the machine, such as an engine sound that is too noisy or humming. For example, sound caused by damaged bearings can be identified on a spectrogram by looking for abnormal or excessively high frequency patterns at certain times of the day. As such, a spectrogram can be a very effective tool in aiding the diagnosis of problems in a car's internal combustion engine through sound.

2. Waveform: Waveform is a visual representation of a sound waveform. The waveform shows the amplitude of the sound with respect to time. In the diagnosis of a car's internal combustion engine, the waveform can be used to identify noisy or abnormal sounds. Waveform is a form of visualization of a sound waveform that shows the amplitude of sound over time. Waveforms are usually depicted as 2D graphs, with the x-axis showing the time and the y-axis showing the sound amplitude. The amplitude of sound is shown as the height of the wave on the graph. In the diagnosis of a car's internal combustion engine, the waveform can help identify engine noise or abnormal sounds. When the machine is running normally, the waveform will show a relatively constant and smooth pattern. However, if there is a problem with the machine, the waveform will show patterns that are not constant and distorted. For example, a machine that is too noisy can be identified by a waveform that shows the amplitude of the sound that is too high and not constant at any given time. If there is a problem with an engine component, such as a faulty valve or piston, the waveform will show a distorted or inconsistent pattern at times. Thus, the waveform can be an effective tool in assisting the diagnosis of problems in the car's internal combustion engine through sound. However, to obtain more complete and accurate information, waveforms are often used together with other forms of visualization, such as spectrograms or polar diagrams.
3. Histogram: A histogram is a visual representation of the frequency distribution of sounds. The histogram shows the number of sounds at each frequency. In diagnosing a car's internal combustion engine, a histogram can be used to identify the dominant frequency in the engine and compare it with normal data, so as to identify abnormal sounds. A histogram is a form of visualization that describes the distribution of sound frequencies at a time. The histogram is usually depicted as a 2D graph, with the x-axis showing the frequency of a sound and the y-axis showing the number of sounds that occur at that frequency. In diagnosing a car's internal combustion engine, a histogram can be used to identify the dominant sound frequencies in the engine and compare them with normal data to identify abnormal sounds. When the engine is running normally, the histogram will show a relatively stable and evenly distributed sound frequency distribution. However, if there is a problem with the machine, the histogram will show a peak or dominance at certain frequencies that are not normal. For example, if there is a problem with the exhaust system, the histogram may show the frequency of the sound being too high or too low compared to the normal frequency of the engine. This can help identify problems with the exhaust system, such as a clogged or damaged exhaust pipe. As such, a histogram can be a very effective tool in aiding the diagnosis of problems with a car's internal combustion engine through sound. However, histograms should be used in conjunction with other forms of visualization, such as spectrograms or waveforms, to obtain more complete and accurate information about machine problems.
4. Mel-frequency cepstral coefficients (MFCC): MFCC is a visual representation of the sound signal processed in the cepstral domain. In the diagnosis of automobile internal combustion engines, MFCC can be used to identify the characteristic sound features of a healthy engine and compare it with the data of a problem engine. Mel-frequency cepstral coefficients (MFCC) are a very popular sound analysis technique in a variety of applications, including the diagnosis of automobile internal combustion engines. This technique converts the speech signal into a form that is easier to process and analyze, which can then be used to identify the distinctive sound features of a healthy engine and compare them to data on a problematic machine. In MFCC, the sound signal is converted from the time domain to the frequency domain through a Fourier transform, then a cepstral measurement (mel-frequency cepstral analysis) is performed to measure the strength of the sound frequency within a certain frequency range. The end result is a collection of vectors that represent sound, which can then be further processed to identify the distinctive features of the engine sound. In the

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

diagnosis of automobile internal combustion engines, MFCC can be used to identify the distinctive features of a healthy engine sound, such as the presence of certain frequency peaks over a certain frequency range, or the amplitude of sound over a certain frequency range. This can then be compared with the problematic machine data, to identify differences in the engine sound features and help determine what problems the machine may have. By using the MFCC technique, an engine sound recognition model can be developed that can distinguish between normal and abnormal sounds in a car's internal combustion engine. This model can process sound signals in real-time and provide useful information to help diagnose engine problems. In conclusion, the MFCC can be a very effective tool in aiding the diagnosis of problems in car internal combustion engines through sound, by identifying the distinctive sound features of a healthy engine and comparing them to the data of the problematic engine. However, this technique requires special knowledge and expertise in sound analysis, so it needs to be implemented by competent and experienced experts.

5. 3D visualization: 3D visualization can be used to spatially visualize sound patterns on machines, which can help identify the location of abnormal sounds and estimate the type of problem with the machine.

In practice, it is usually more effective to use more than one type of visualization data in the analysis of car internal combustion engine sound. This is because each type of visualization data provides different information and can complement each other in assisting the diagnosis.

How to process and analyze sound data from a car's internal combustion engine, and convert it into visualization data that can be used in diagnosis? (RQ 2).

To process and analyze the sound data from the internal combustion engine, the following steps can be taken:

1. Record the sound from the car's internal combustion engine using a suitable microphone or sound sensor.
2. Convert the voice signal to digital form using sound recording software or an analog to digital converter.
3. Speech signal analysis using signal processing software such as MATLAB, Python, or other signal processing software.
4. To produce useful visualization data, several techniques such as spectrograms, waveforms, histograms, and MFCC can be performed as previously described.
5. After the visualization data is generated, it can be compared with normal or reference data that has been collected before.
6. By comparing normal machine visualization data with problematic engine visualization data, problems with the machine can be identified and further diagnosis can be made.

At the analysis stage, several methods such as statistical analysis and machine learning can be used to identify patterns and trends in voice data. This can help identify abnormal sounds and determine possible engine problems.

What kinds of algorithms or techniques are used in sound data processing and data visualization in car internal combustion engines, and what are their advantages and disadvantages? (RQ 3).

There are several techniques and algorithms used in processing sound data and visualizing car internal combustion engine data, including:

1. Fast Fourier Transform (FFT): FFT is a technique used to convert a sound signal from the time domain to the frequency domain. FFT generates a sound frequency spectrum and can be used to identify the dominant frequency in engine sound. The advantage of FFT is that it is fast and easy to perform. However, the drawback of the FFT is the limited frequency resolution and it is not suitable for solving non-stationary problems in voice signals.
2. Principal Component Analysis (PCA) (Hindarto, 2022): PCA is a technique used to reduce the dimensions of voice data and identify important features of voice data. PCA can be used to reduce noise in voice data and improve accuracy in machine diagnosis. The advantage of PCA is its ability to reduce the dimensions of voice data and increase computational speed. However, the drawback of PCA is that it is difficult to interpret and can produce meaningless features.

*name of corresponding author

3. Convolutional Neural Network (CNN): CNN is a deep learning technique used to process voice data and generate accurate predictive models. CNN can be used to identify problems with a machine based on the sound it makes. The advantage of CNN is its ability to process complex voice data and generate accurate predictive models. However, the disadvantages of CNN are that it takes a long computation time and requires a large amount of data to train the model.

Each technique and algorithm have its advantages and disadvantages. Selection of the right technique depends on the purpose of data processing and the characteristics of the sound data to be processed.

How to compare visualization data from a healthy car's internal combustion engine and a problem engine, and how to identify significant differences between the two? (RQ 4).

To compare the visualization data from a healthy car's internal combustion engine and a problem engine, it can be done in several ways, including:

1. Comparing graphics or visualizations visually: Graphics or visualizations of the sounds of healthy machines and machines with problems can be visually compared. In general, differences in the sound of problematic machines will be seen on graphics or visualizations with different shapes or patterns from healthy machines. However, this method is sometimes not accurate enough to identify significant differences.
2. Comparing numeric values: Some types of visualization data, such as MFCC and histograms, generate numerical values that can be directly compared between healthy and troubled machines. A significant difference between the two numerical values may indicate a problem with the machine.
3. Use of machine learning methods: Machine learning methods can be used to compare sound data from healthy and troubled machines. In this case, the machine learning algorithm will be trained using voice data from healthy and problematic machines, and then it can be used to predict whether the sound from the newly heard machine is included in the category of healthy or problematic machines. This method can produce more accurate and reliable results.

In identifying significant differences between healthy and faulty engines, it should be noted that each engine has unique sound characteristics. Therefore, it is important to use sound data from the same engine and under the same conditions when comparing healthy and faulty engines. In addition, it is necessary to carry out careful evaluation and validation of the analysis results to ensure their accuracy.

CONCLUSION

In the automotive industry, the internal combustion engine (MPI) is known for its distinctive and unique sound characteristics. An experienced auto mechanic can diagnose the type of engine failure just by listening to the sound of the engine. However, this process still depends on the experience and skill of the mechanic concerned. Therefore, many automotive and technology companies are starting to develop technology that can assist mechanics in diagnosing engine damage more effectively and efficiently. One of the technologies proposed in this paper is the process of identifying engine failures through analysis of engine sounds and visual data. In this process, engine sound data is captured and analyzed using digital signal processing technology to identify abnormal or anomalous sound patterns. Additionally, visual data, such as infrared images or other sensor data, can also be captured to help more accurately identify engine failures. Using this technology, mechanics can identify engine failures at an early stage, which can significantly reduce repair time. In addition, this technology can also help increase the productivity of car mechanics and reduce repair costs incurred by vehicle owners. However, this technology is still in the development stage and must be tested before it can be widely implemented in the automotive industry. Also, keep in mind that replacement or repair of damaged components must be carried out by skilled and experienced mechanics, and this technology can only assist in identifying the fault at an early stage. Overall, this technology is promising and can help improve efficiency and accuracy in diagnosing engine damage in motorized vehicles. However, it is necessary to carry out further research and development of more advanced technology to ensure the successful implementation of this technology in the future.

REFERENCES

Alam, M. M., Aktar, M. A., Idris, N. D. M., & Al-Amin, A. Q. (2023). World energy economics and geopolitics amid COVID-19 and post-COVID-19 policy direction. *World Development*

SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023
DOI : https://doi.org/10.33395/sinkron.v8i2.12392
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Sustainability, *2*(July 2022), 100048. https://doi.org/10.1016/j.wds.2023.100048

Carpinteiro, C., Lopes, J., Abelha, A., & Santos, M. F. (2023). ScienceDirect ScienceDirect A Comparative Study of Classification Algorithms for Early A Comparative Study of Classification Algorithms for Early Detection of Diabetes Detection of Diabetes. *Procedia Computer Science*, *220*, 868–873. https://doi.org/10.1016/j.procs.2023.03.117

Dwyer, T., Cordeil, M., Czauderna, T., Delir Haghighi, P., Ens, B., Goodwin, S., Jenny, B., Marriott, K., & Wybrow, M. (2020). The Data Visualisation and Immersive Analytics Research Lab at Monash University. *Visual Informatics*, *4*(4), 41–49. https://doi.org/10.1016/j.visinf.2020.11.001

Hindarto, D. (2022). Perbandingan Kinerja Akurasi Klasifikasi K-NN, NB dan DT pada APK Android. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(1), 486–503. https://doi.org/10.35957/jatisi.v9i1.1542

Hindarto, D., & Handri Santoso. (2021). Android APK Identification using Non Neural Network and Neural Network Classifier. *Journal of Computer Science and Informatics Engineering (J-Cosine)*, *5*(2), 149–157. https://doi.org/10.29303/jcosine.v5i2.420

Hindarto, D., Indrajit, R. E., & Dazki, E. (2021). Sustainability of Implementing Enterprise Architecture in the Solar Power Generation Manufacturing Industry. *Sinkron*, *6*(1), 13–24. https://jurnal.polgan.ac.id/index.php/sinkron/article/view/11115

Hindarto, D., & Santoso, H. (2022). PERFORMANCE COMPARISON OF SUPERVISED LEARNING USING NON-NEURAL NETWORK AND NEURAL NETWORK. *Janapati*, *11*, 49–62.

Huang, A., Xu, R., Chen, Y., & Guo, M. (2023). Research on multi-label user classification of social media based on ML-KNN algorithm. *Technological Forecasting and Social Change*, *188*(August 2022), 122271. https://doi.org/10.1016/j.techfore.2022.122271

Hyodo, S., Yoshii, T., Satoshi, M., & Hirotoshi, S. (2017). An analysis of the impact of driving time on the driver's behavior using probe car data. *Transportation Research Procedia*, *21*, 169–179. https://doi.org/10.1016/j.trpro.2017.03.086

Ivanov, E., Khoroshavin, A., & Karsakov, A. (2020). Visual programming environment based on data visualization grammar specification. *Procedia Computer Science*, *178*(2019), 434–439. https://doi.org/10.1016/j.procs.2020.11.045

Li, L., Lin, J., Wu, N., Xie, S., Meng, C., Zheng, Y., Wang, X., & Zhao, Y. (2022). Review and outlook on the international renewable energy development. *Energy and Built Environment*, *3*(2), 139–157. https://doi.org/10.1016/j.enbenv.2020.12.002

Liu, J., Li, X., Zhang, X., Xu, S., & Dong, L. (2011). Misfire diagnosis of diesel engine based on rough set and Neural Network. *Procedia Engineering*, *16*, 224–229. https://doi.org/10.1016/j.proeng.2011.08.1076

Permai, S. D., & Tanty, H. (2018). Linear regression model using bayesian approach for energy performance of residential building. *Procedia Computer Science*, *135*, 671–677. https://doi.org/10.1016/j.procs.2018.08.219

Traivivatana, S., Wangjiraniran, W., Junlakarn, S., & Wansophark, N. (2017). Impact of Transportation Restructuring on Thailand Energy Outlook. *Energy Procedia*, *138*, 393–398. https://doi.org/10.1016/j.egypro.2017.10.178

Vidal, T. Y. G., Claude Valery, N. A., Emmanuel, A. N., Nelson, I. B. J., Moïse, L. N., Cyrille, M., & Ruben, M. (2021). Performance map of a LPG-diesel dual-fuel engine based on experimental and non-linear least squares determined wiebe function. *Scientific African*, *13*. https://doi.org/10.1016/j.sciaf.2021.e00900

Wedha, B. Y. (2022). Enterprise Architecture untuk Industri Truk Logistik di Indonesia. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, *9*(2), 1137–1150. https://doi.org/10.35957/jatisi.v9i2.1255

Wedha, B. Y., Karjadi, D. A., Wedha, A. E. P. B., & Santoso, H. (2022). Style Transfer Generator for Dataset Testing Classification. *SinkrOn*, *7*(2), 448–454. https://doi.org/10.33395/sinkron.v7i2.11375

Wolak, A., Molenda, J., Zając, G., & Janocha, P. (2021). Identifying and modelling changes in chemical properties of engine oils by use of infrared spectroscopy. *Measurement: Journal of the*

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 7, Number 2, April 2023 e-ISSN : 2541-2019
DOI : https://doi.org/10.33395/sinkron.v8i2.12392 p-ISSN : 2541-044X

International Measurement Confederation, *186*. https://doi.org/10.1016/j.measurement.2021.110141

Zheng, T., Zhang, Y., Li, Y., & Shi, L. (2019). Real-time combustion torque estimation and dynamic misfire fault diagnosis in gasoline engine. *Mechanical Systems and Signal Processing*, *126*, 521–535. https://doi.org/10.1016/j.ymssp.2019.02.048

*name of corresponding author