JOURNAL OF
EMERGING INVESTIGATORS

Article

Predicting clogs in water pipelines using sound sensors and machine learning linear regression

Vivaan Rajawat[1], Gunjan Rajput[2]
[1] Pate's Grammar School, Cheltenham, Gloucestershire, United Kingdom
[2] Indian Institute of Technology, Indore, India

SUMMARY

A clog in a water pipe might not seem urgent, but it can escalate to leakages or even burst pipes, posing risks of structural damage. Furthermore, if the clog contains corrosive substances like household or industrial chemicals, it can accelerate pipe deterioration. Leaks resulting from blockages can lead to seepage problems and health hazards associated with living in damp conditions. To tackle these challenges, we aimed to develop a machine-learning algorithm capable of detecting and predicting water pipe blockages in distribution pipelines. We hypothesized that the varying intensity of sound in clean versus blocked pipelines could be utilized to devise a machine-learning algorithm for blockage detection and prediction. To test this hypothesis, we used acoustic sensors to collect data and employed linear regression classification techniques to categorize the data into two sets. We observed substantial variability in sensor readings between clean and blocked pipes. This variability formed the basis for training the machine learning model to identify and forecast pipeline blockages. In this work, the mean square error (MSE) is 5.28 and R^2 value of 0.98 for the model evaluated on the testing set. These metrics demonstrate the high accuracy of our model in predicting sensor values. Our results show that the intensity of sound varies in the clogged and clean pipe, and the variation can be used for clog detection.

INTRODUCTION

Occupants of damp or moldy buildings, caused by water leakages behind the walls, are at an increased risk of respiratory symptoms, respiratory infections, and exacerbation of asthma, with some evidence suggesting increased risks of allergic rhinitis and asthma (1). For instance, negative health impacts of leakages could be widespread considering the severity of leakage. Beyond health concerns, the wastage of this valuable resource due to leakage is another issue that needs to be addressed. The Environment Agency UK estimates that 3 billion liters, equivalent to the average daily use of 20 million people, are lost daily in the UK alone through leakages (2).

Water seepage and leakage issues not only pose health risks and waste of water but also cause structural damage to properties, which can be very expensive to repair (3). A typical network of water pipelines in a home or industrial building has tangled junctions of pipes, which are often located behind walls or underground, making them difficult to access **(Figure 1)**. Leakages in such inaccessible areas further exacerbate the problem. Underdeveloped countries are more susceptible to these issues due to a lack of infrastructure and economic challenges (2).

To address these challenges, researchers have been exploring various technology-based solutions to detect leakages, which are often closely linked to clogging in water pipelines. Clogging can increase internal pressure and stress within the pipes, eventually causing cracks or bursts that lead to leakages. Therefore, early detection and prevention of clogging are essential to minimizing leakage-related health and structural consequences. A previous study has investigated the use of the Internet of Things (IoT)—a network of interconnected physical devices—and machine learning algorithms to monitor water quality and detect issues in real-time (4). Additionally, acoustic sensors have been employed to identify leakages by analyzing sound variations within the pipes, while advanced imaging techniques have been used to detect clogs before they escalate (5). These solutions offer promising results but require further research and development to be effectively scaled and implemented (6,7).

Current methods primarily based on chemicals for detecting clogged pipes often have significant limitations that underscore the need for a new approach. Chemical methods for removing clogs typically use specialized solutions to dissolve organic materials and mineral deposits, while effective, involve expensive reagents and precise handling, which add to their operational costs (6). Another category of methods based on advanced imaging techniques, though accurate, require specialized equipment and expertise, making them cumbersome and expensive for routine use (6). The cost and logistic complexity make them impractical for widespread use in home settings (6).

In contrast, our proposed method uses a sensor-based approach, a more affordable and practical alternative. Sensors are relatively inexpensive and can be easily integrated into residential environments without requiring complex infrastructure or professional expertise (4). This method provides a cost-effective solution for detecting pipe clogs, addressing the limitations of existing methods by offering a simpler and more accessible option for everyday use **(Table 1)**. The proposed water pipeline clogging detection system works on a simple yet powerful concept that the intensity of sound varies depending on the degree of disturbance in the medium through which the object travels (8). These disturbances create pressure in the medium which is directly proportional to the intensity of the sound produced as established by **Equation 1**:

https://doi.org/10.59720/24-151

$$I = \frac{\Delta P^2}{2\,\rho\,V} \qquad \textbf{(Eqn 1)}$$

Where I is the intensity of sound in dB; ΔP is the change in pressure; ρ is the density of the material the sound is travelling through; V is the speed of the observed sound (8).

Based on this concept, we hypothesized that the intensity of sound in clean pipelines would vary significantly from blocked pipelines, and this characteristic could be utilized to devise a machine learning algorithm for blockage detection and prediction.

This research introduces a novel approach to pipe clog detection by incorporating linear regression to predict sound intensity levels. In addition to the other initial methodology utilizing a Random Forest classifier, this proposed method aims to provide a deeper quantitative understanding of the relationship between input features and actual sound intensity. Linear regression provides a deeper quantitative understanding of the relationship between input features – sound intensity and timestamp – and actual sound intensity by assigning specific weights to each input variable. These weights directly reflect how changes in each feature impact the output, offering a clear view of both the magnitude and direction of these relationships. In contrast to Random Forest classifiers, which focus on making accurate predictions but may lack transparency in feature contributions, linear regression offers interpretable insights into how individual input variables influence sound intensity (12,13). This makes it particularly useful for understanding and optimizing the factors that contribute to sound intensity. Such insights are particularly beneficial in applications like sound design or engineering, where knowing the precise contribution of each feature is crucial for optimization and decision-making.

We chose linear regression for its ability to predict a continuous variable—in our case, the intensity of sound—and the target value for the intensity of sound was modified accordingly. We adjusted the feature extraction which involves transforming raw data into a set of features that can be used to build a model to focus on predicting the precise sound intensity values rather than classifying states. The model is trained on a labeled dataset, and evaluation metrics such as mean squared error (MSE) and R^2 are employed to assess the regression model's accuracy in predicting continuous sound intensity levels The MSE indicates the average squared difference between predicted and actual values, and the R^2 value shows how closely a data set fits the linear regression trendline. We found a linear correlation between sound intensity variation and clog severity. This correlation in data is the key to training the machine learning algorithm. The designed machine learning model is capable of accurately predicting sensor values based on previous readings. Utilizing linear regression, the model achieves high accuracy on the testing set. By predicting clogs, this solution has the potential to significantly address the water leakage problem, prevent the wastage of this invaluable resource, and mitigate associated health risks.

Reference	Approach	Strength	Limitations
Jombo & Zhang (2023) (8)	Acoustic monitoring using microphones and signal processing for industrial machinery.	Non-invasive, capable of continuous monitoring, proven for industrial settings.	Requires complex filtering and feature engineering; not optimized for residential pipeline systems.
Müller et al. (2020) (9)	Image-based transfer learning for acoustic anomaly detection.	High accuracy using pretrained CNNs; applicable to varied sound data.	High computational cost; transfer learning may not generalize well to new environments.
Jisheng Bai et al. (2022) (10)	Self-supervised dual-path Transformer for anomalous sound detection.	Advanced deep learning model with excellent generalization; no manual labels required.	Resource-intensive; requires large datasets and tuning, limited interpretability.
Proposed Method (This Study)	Sound intensity modelled via linear regression using sensor data from pipelines.	Low-cost, interpretable, suitable for both industrial and household settings; enables prediction and maintenance.	Works best when there's a clear pattern in the data; careful selection of input features ensures accurate predictions.

Table 1. Literature Review. Comparative analysis of existing methods and contributions of the proposed method.

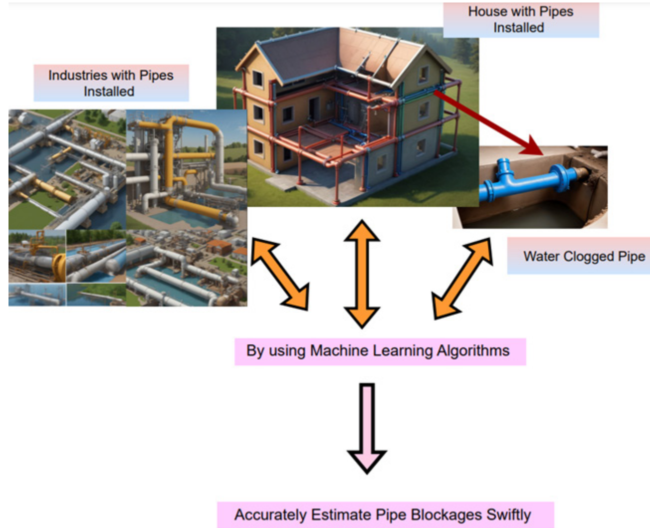

Figure 1. A typical mesh of pipelines in a home or industry. The vulnerability of water pipeline networks—both industrial and residential—to issues like clogs and leaks. It highlights how machine learning algorithms can be applied to predict such problems accurately and efficiently. By showing real-world examples and emphasizing the role of technology, the figure underscores the need for smart, cost-effective solutions, especially in regions with limited infrastructure.

RESULTS

We gathered sound sensor data in two scenarios: when the pipe is unclogged and when it is clogged. We observed that data followed the relationship between sound intensity and disturbance (clogs) behavior **(Equation 1)**. The sound sensor values recorded under clean and clogged pipe conditions show a significant variation – defined as the degree to which data points differ from one another. We observed that the sound sensor data exhibited greater fluctuations when the pipe was clogged; this increased variability is visually emphasized by the black curved on the graph (**Figure 2**). In contrast, when the pipe is clean, there is almost no variation. However, we observed noise during the initial phase in the data as depicted in the graph which was due to initial disturbance while setting up the hardware. We observed that the sensor value fluctuates over time, with some periodic patterns and occasional spikes. We also observed some missing values in the dataset, which we addressed using linear interpolation. This method estimates the missing values based on the known data points. We observed that the predicted values closely follow the actual values, with some minor deviations due to the noise in the data.

https://doi.org/10.59720/24-151

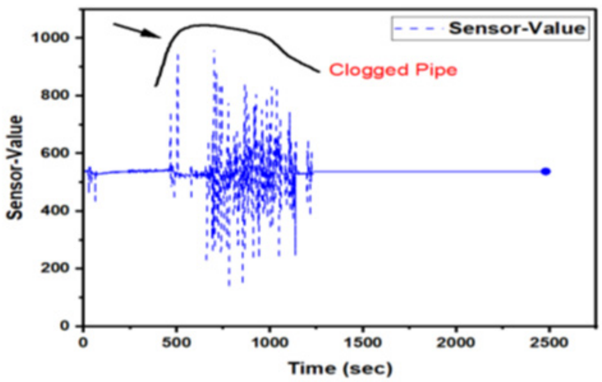

Figure 2. Variation in sound intensity in a clear and clogged pipe. Clogging started at around 600s and was resolved at around 1,200 s (black line). The density of data points indicates the level of clogging: A higher concentration of data points signifies more severe clogging. The dotted line represents the variable portion of the data, while the solid line indicates interpolated data points from the linear regression model. Variation in the sound intensity increases when the pipe starts clogging and gradually reduces when the pipe is clean.

The actual sensor values closely match the predicted values generated by the linear regression model for both the training and testing datasets, as illustrated in (**Figure 3**). The model demonstrated high predictive accuracy, as evidenced by the close alignment between the red and green lines representing predicted values for the training and testing sets, respectively; the blue and orange dots correspond to the actual data points use from training and testing datasets. We plotted the residual errors, which show the difference between the actual values and the predicted values. We observe that the errors are mostly small and random, which indicates that the model can capture the main patterns in the data (**Figure 4**).

We evaluated the performance of the model on the testing set by computing the MSE and the coefficient of determination (R^2). The MSE measures how close the predicted values are to the actual values, while the R^2 measures how much of the variance in the data is explained by the model. We calculated an MSE of 5.28 and an R^2 of 0.98, which indicates that the model can accurately predict the sensor values.

DISCUSSION

The machine learning algorithm, trained on historical sensor data, demonstrates exceptional accuracy in predicting sensor values, as evidenced by low MSE and high R^2 values. These results indicate the model's strong ability to anticipate sensor behavior, effectively capturing early signs of pipe clogging observed during our experimental setup. The machine learning algorithm is trained to learn the normal behavior of the sound sensor by analyzing historical data. This includes patterns in the sound signal when the pipe is clean and operating normally. Once trained, the model can predict what the sound signal should look like under normal conditions.

Similarly, the algorithm is trained to recognize patterns in the data that exhibit fluctuations—characteristic of the early stages of clogging in the pipe. This includes learning how the sound signal behaves when clogging begins to develop. By capturing these patterns during training, the model becomes capable of predicting what the sound signal should look like under both normal (unclogged) and abnormal (clogged) conditions.

Many researchers have worked over sensors to accurately record the data and drive insight out of it (14,15). Due to economic reasons, our research has been carried out by using wired sensors with a setup conducted at home. This limited the amount and accuracy of collected data. To address these constraints, we applied data augmentation to generate additional data offline and noise reduction to enhance data quality. We additionally used interpolation on our data. The interpolation of data points makes data robust by estimating unknown values within a dataset's range based on known data points, allowing for more comprehensive data without gaps in it, and is improving the performance and robustness of the model by training it with a complete dataset (14). To ensure the model's evaluation remains realistic, it is important

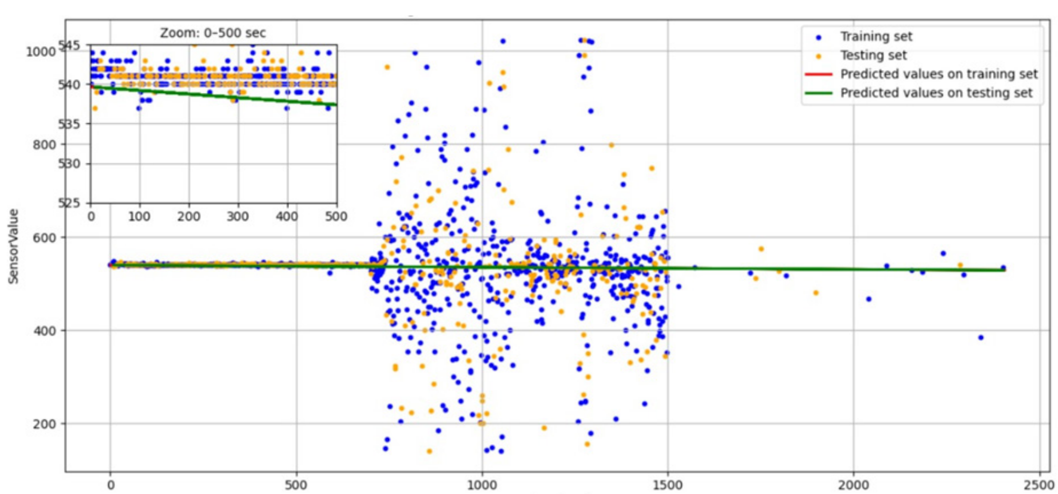

Figure 3. Training and testing data graph for clog detection. The sensor data is loaded into a Pandas Dataframe, visualized with Matplotlib, and missing values are filled by linear interpolation of eighteen data points. Splitting the data 70-30 for training and testing, a linear regression model is trained with sci-kit-learn to predict future sensor values, chosen for its simplicity in clog detection. The predictive values on training set (red line) and testing set (green line) are overlapping which signifies the higher accuracy of the model.

https://doi.org/10.59720/24-151

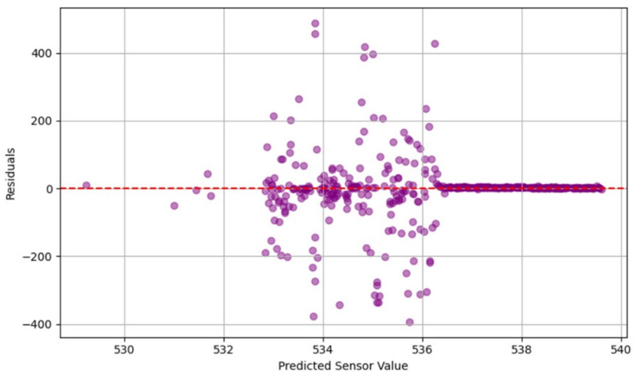

Figure 4. Residuals vs predicted values. A cluster of higher residuals appears between predicted sensor values 532–536. This suggests reduced model accuracy in this range. However, this issue is not entirely due to model limitations; it is largely attributed to the low resolution of the sensor used during data acquisition. The sensor's reduced precision in this operating band likely introduces noise and error that even a well-trained model cannot compensate for. In future work, the use of a higher-quality sensor could mitigate this problem and enhance prediction fidelity.

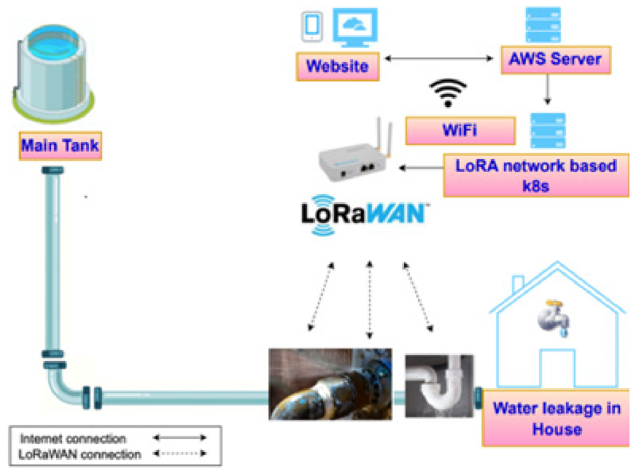

Figure 5. A hypothetical system design for predictive clog detection. A system where clog detection and prediction algorithms can be used. LoRa technology can be used to establish wireless connectivity between the server and sensors to collect the data in real-time and apply the algorithm to the collected data.

to consider the potential impact of augmented or interpolated data points on both the training and validation processes. While these techniques help compensate for limited or noisy data, they may not fully capture the complexity and variability of real-world signals. As a result, relying heavily on such data could lead to overly optimistic performance metrics during validation and reduce the model's ability to generalize effectively in practical applications (10).

The algorithm under study presents significant potential for various sectors, including water management bodies, facility management, and homeowners, by effectively addressing issues of water pipe clogging and leakages. Integrated into smart water-efficient home systems, our algorithm can predict pipe clogs, thereby mitigating the primary causes of pipe bursts and leakages. This proactive approach not only enhances water conservation efforts but also minimizes property damage and maintenance costs associated with water infrastructure.

Future solutions that make use of this algorithm can use a wireless sensor network (WSN) using a Long Range (LoRa) gateway (**Figure 5**). The LoRa technology wireless sensor network is thoroughly researched and provides a viable solution to tackle water leakage issues (14). A long-range gateway-based wireless sensor network (WSN) system is essential for effective water leakage detection, especially in large or hard-to-reach areas. It consists of wireless sensors that monitor conditions like pressure or sound along water pipes and a central gateway that collects this data. The long-range communication (such as LoRa) allows sensors to operate over wide distances with low power, making the system scalable and energy-efficient. This setup enables real-time monitoring, early leak detection, and predictive maintenance, helping to reduce water waste, prevent damage, and support smart infrastructure development.

The algorithm can be used to aid water management by implementing a WSN. A LoRa wireless sensor network consists of multiple sound sensors deployed at the appropriate locations around the mesh of the water pipeline. Leveraging long-distance, ultralow power LoRa technology, the system enables real-time data exchange between one gateway and three sensor nodes (**Figure 5**) The gateway acts as a link between the internet and the WSN and is responsible for data collection and aggregation. This aggregated data can be provided to the water clogging prediction machine learning algorithm to predict the issues. These issues can be directly reported to users via smartphone applications running on a mobile device or on an application server which is hosting a visual monitoring dashboard application. For transmission, we can use Kubernetes (k8s) which acts like a smart manager for your applications, ensuring they run smoothly, stay available, and can grow or shrink as needed without manual intervention. We can use virtual servers hosted on computer clouds provided by cloud service providers such as Amazon Web Services (AWS) for artificial intelligence/machine learning and other activities so that we can capture huge amounts of data.

Overall, the integration of sensors and machine learning algorithms for detecting and predicting clogs in the water pipelines water quality monitoring can have a significant impact on social welfare and contribute to achieving the United Nations Sustainable Development Goals, particularly to ensuring access to clean water and sanitation for all (15). However, it is worth researching and exploring further for other environmental factors such as temperature and pressure, and testing under varied conditions to capture more data to ensure the robustness of the experiments (8, 9). While our data analysis serves as a starting point, it requires further refinement to adapt to different scenarios(16). We also acknowledge our sensors' limitations in terms of the precision of their quality of recording minute details of the sound due to budget constraints. We are committed to improving their performance in future iterations. Moreover, we can consider other machine learning models, such as the Exponential Moving Average (EMA) model, to determine the fluctuations in the sound intensity to predict whether the pipe is clogged or unclogged.

An EMA model could enhance the current method by providing a more adaptive approach to capturing temporal dynamics in the data. Unlike the current methodology, which

https://doi.org/10.59720/24-151

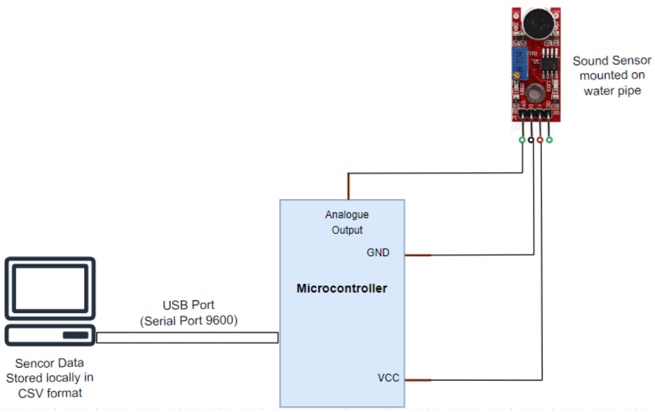

Figure 6. Data collection setup. The sensor was mounted on the pipe and connected to the microcontroller to receive and save the data locally. Initially, the pipe was kept clear and water flowed smoothly from it and gradually water flow was obstructed by introducing wool and cloth pieces to simulate a clog. The data was collected from the sensor and stored locally on a laptop in CSV format. The expected pattern in the data was observed and the ML algorithm was trained to predict the clogging.

might weigh all observations equally or rely on static weights, an EMA assigns exponentially decreasing weights to older data points, allowing more recent data to have a greater influence on the predictions (17). This approach could be particularly useful for capturing trends or patterns in time-series data where recent observations are more relevant to predicting future sound intensity levels. By continuously updating the weighted average as new data is received, the EMA model could lead to more accurate predictions, especially in dynamic environments where the input features or conditions are subject to frequent changes (17).

Additionally, the EMA's ability to smooth out short-term fluctuations while maintaining sensitivity to significant changes could provide more stability in predictions, potentially improving both the responsiveness and robustness of the model in real-world applications (17). A comparative analysis of the existing system which tackles this issue was not done as a part of this research, but it can be considered in the future.

MATERIALS AND METHODS

We used the wired sound sensor to carry out the experimentation for collecting the data. To ensure accurate data collection, a microcontroller is used to calibrate the sensors and record readings for subsequent analysis. This approach allows for precise measurement and tracking of multiple parameters, facilitating a comprehensive understanding of the environment being monitored.

Data collection

The data required to train and test the model is not available on authenticated resources on the internet, so we prepared the setup at home to collect the data (**Figure 6**). We used a small plastic drainpipe of approximately 20 inches in length and on top of it mounted the sound sensor. A shower hose is used as the water source in the bathtub. We used some pieces of cloth to clog the pipe. The sensor was connected to a microcontroller (Arduino) with the help of connecting wires and the microcontroller connected via the communication port of the laptop. We used Arduino Integrated Development Environment to develop code and push software images to the microcontroller. The serial monitor is kept at 9,600 baud rates, which is defined as the number of signal changes that occur in a communication channel per second, to display the readings and store the data in CSV format. The collected dataset contains the sensor value and the time elapsed since the start of the recording. To capture the data, five independent trials with the same plastic pipe were performed. At each trial, we kept the mounted sensor at a different location on the pipe to cover more realistic scenarios. Initially, the plastic pipe was kept unclogged, and later it was clogged by inserting some piece of cloth and wool. Later the pipe was unclogged by removing the cloth and wool from it. The experiment was repeated, and data were captured for augmentation and noise reduction methods.

Data augmentation and noise reduction

We used various techniques to enhance and clean the time series data. In our work, we used time warping to adjust the timing of data points slightly and added small random noise, known as jittering, to create more variability. Additionally, we used sliding windows to extract overlapping segments from the time series, which helps in generating multiple samples from the same data. We also employed synthetic data generation to create additional samples based on the existing data. We implemented smoothing techniques such as moving averages and Gaussian filters to reduce random fluctuations. We applied low-pass filters to remove high-frequency noise while preserving important trends. Before using the sensor data for training and testing the model, we used linear interpolation of 18 data points that were included in both the training and validation sets. The inclusion of interpolated points was intended to help fill gaps and provide a more continuous dataset, which can improve model performance by providing more consistent data for training and evaluation.

Training and testing machine learning algorithm

We implemented the algorithm by using the Python programming language, utilizing libraries such as Pandas for data manipulation, NumPy for numerical computations, Matplotlib for data visualization, and Scikit-learn for machine learning model development. The dataset was loaded into a Pandas DataFrame from a CSV file, and timestamps were transformed into seconds since the epoch to facilitate numerical processing. Missing values in the dataset were addressed using linear interpolation to ensure continuity in the data. The "Time" column was extracted as the input feature, while the "SensorValue" column was used as the target variable for the regression model. The dataset was then split into training and testing subsets in a 70-30 ratio using Scikit-learn's train_test_split method. A linear regression model was trained on the training set. The model takes as input the sensor value at the time, *t*, and tries to predict the sensor value at *t+1*, and predictions were generated for both the training and testing data. The results were visualized through scatter and line plots to compare actual and predicted values (**Figure 3**), while residuals were plotted to evaluate the model's performance and ensure its effectiveness in capturing patterns in the data (**Figure 4**).

https://doi.org/10.59720/24-151

ACKNOWLEDGMENTS

The first author would like to thank his physics teacher Mr. J Talbot and his father Manoj Kumar for their valuable insights and suggestions that were instrumental in shaping the direction of this project.

Received: June 3, 2024
Accepted: November 4, 2024
Published: October 11, 2025

REFERENCES

1. "WHO guidelines for indoor air quality: Dampness and mould." *World Health Organisation.* www.who.int/publications/i/item/9789289041683. Accessed 01-Mar-2023
2. "Water company performance report 2018: Leakage" *Environment Agency UK.* www.assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1005590/State_of_the_environment_water_resources_report.pdf. Accessed 22-Feb-2023.
3. Li, Jie et al. "Case study and back analysis of a residential building damaged by expansive soils" *Computers and Geotechnics*, vol. 56, March. 2014, pp. 89-99. https://doi.org/10.1016/j.compgeo.2013.11.005.
4. Zulkifli, C. et al. "IoT-based water monitoring systems: a systematic review" *Multidisciplinary Digital Publishing Institute*, vol. 14, no. 3621, 10 Nov. 2022, https://doi.org/10.3390/w14223621.
5. Rousso, B.Z. et al. "Smart water networks: A systematic review of applications using high-frequency pressure and acoustic sensors in real water distribution systems" *Journal of Cleaner Production*, vol. 410, no. 137193, 15 July 2023, https://doi.org/10.1016/j.jclepro.2023.137193.
6. Al Qahtani, T. et al. "A review on water leakage detection method in the water distribution network" *Journal of Advanced Research in Fluid Mechanics and Thermal Sciences*, vol. 68(2), no. 152-163, 31 March 2020, https://doi.org/10.37934/arfmts.68.2.152163.
7. Fereidooni, Z. et al. "A hybrid model-based method for leak detection in large scale water distribution networks. Journal of Ambient Intelligence and Humanized Computing" *Journal of Ambient Intelligence and Humanized Computing*, vol. 12, no. 1613-1629, 4 July 2021, https://doi.org/10.1007/s12652-020-02233-2.
8. Jombo G. and Zhang Y. "Acoustic-Based Machine Condition Monitoring—Methods and Challenges" *Eng*, vol. 4, 1 January 2023, https://doi.org/10.3390/eng4010004.
9. Müller et al. "Acoustic Anomaly Detection for Machine Sounds based on Image Transfer Learning" *ArXiv*, December 2020, https://doi.org/10.48550/arXiv.2006.03429.
10. Jisheng, B. et al. "Self-supervised dual-path transformer for anomalous sound detection" *Digital Signal Processing*, vol. 135, 30 April 2023, https://doi.org/10.1016/j.dsp.2023.103939.
11. Sojahrood, A. J. et al. "Probing the pressure dependence of sound speed and attenuation in bubbly media: Experimental observations, a theoretical model and numerical calculations" *Ultrasonics Sonochemistry*, vol. 95, May 2023, https://doi.org/10.1016/j.ultsonch.2023.106319.
12. Khorshidi, M.S. et al. "A Pressure sensor placement in water distribution networks for leak detection using a hybrid information-entropy approach" *Information Sciences*, vol. 516, no. 56-71, April 2021, https://doi.org/10.1016/j.ins.2019.12.043.
13. Paul, F. Smith et al. "A comparison of random forest regression and multiple linear regression for prediction in neuroscience" *Journal of Neuroscience Methods*, vol. 220, Issue 1, October 2013, https://doi.org/10.1016/j.jneumeth.2013.08.024.
14. Fan Bill Yu et al. "Machine learning model and strategy for fast and accurate detection of leaks in water supply network" *Journal of Infrastructure Preservation and Resilience*, vol. 2, no. 10, 15 April 2021. https://doi.org/10.1186/s43065-021-00021-6.
15. Tuyen Phong Truong et al. "Towards an IoT-Based System for Monitoring of Pipeline Leakage in Clean Water Distribution Networks" *EAI Endorsed Transactions on Smart Cities*, vol. 6, no. 17,p.e5, March 2022, https://doi.org/10.4108/eetsc.v6i17.230.
16. Milan, B.F. "Clean water and sanitation for all: interactions with other sustainable development goals", *Sustainable Water Resources Management ,* vol. 3, no.479-489, April 2017, https://doi.org/10.1007/s40899-017-0117-4.
17. Tai-hoon Kim et al. "A Smart, Sensible Agriculture System Using the Exponential Moving Average Model", *Symmetry*, vol. 12, no. 3, March 2020, https://doi.org/10.3390/sym12030457.

https://doi.org/10.59720/24-151

APPENDIX

```
######################################
Algorithm Code
####################################
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split

def run_linear_regression(file_path):
    try:
        data = pd.read_csv(file_path, header=None)
        data.columns = ['Timestamp', 'SensorValue']
        data['Time'] = pd.to_datetime(data['Timestamp']).astype(int) / 10**9
        data = data.interpolate(method='linear', limit_direction='forward', axis=0)
        print("Dataset loaded and preprocessed successfully!")
    except Exception as e:
        print(f"Error loading dataset: {e}")
        return

    X = data[['Time']].values
    y = data['SensorValue'].values
    X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
    model = LinearRegression()
    model.fit(X_train, y_train)
    y_train_pred = model.predict(X_train)
    y_test_pred = model.predict(X_test)

    plt.figure(figsize=(10, 6))
    plt.scatter(X_train, y_train, color='blue', label='Training set')
    plt.scatter(X_test, y_test, color='orange', label='Testing set')
    plt.plot(X_train, y_train_pred, color='red', label='Predicted values on training set')
    plt.plot(X_test, y_test_pred, color='green', label='Predicted values on testing set')
    plt.title('Linear Regression Model')
    plt.xlabel('Time (sec)')
    plt.ylabel('Sensor Value')
    plt.legend()
    plt.grid()
    plt.show()

if __name__ == "__main__":
    file_path = input("Enter the path to your dataset (CSV file): ")
    run_linear_regression(file_path)
#######################################
```