OPEN ACCESS

International Journal of Advances in Artificial Intelligence and Machine Learning Vol 1 No 1 (2024) https://doi.org/10.58723/ijaaiml.v1i1.285

Check for updates

Research Article

AI-Enhanced Gross Pollutant Traps: A Smart Approach to River Health and Pollution Control

Received: Oct 04, 2024	Revised: Oct 14, 2024	Accepted: Nov 05, 2024	Publish: Nov 06, 2024

Chang Shi Ying, Bong Peak May, Soo Ting Fang, Lee Wai Yi, Misinem*

Abstract:

Flooding and river pollution pose significant challenges in Malaysia, exacerbated by the inefficiencies of Gross Pollutant Traps (GPTs), which rely on manual monthly cleaning processes. These conventional methods are inadequate for addressing the dynamic influx of pollutants, particularly during adverse weather conditions. This research proposes an innovative AI-powered framework that integrates logistic regression for weather prediction and Convolutional Neural Networks (CNNs) for real-time garbage classification. By predicting weather patterns and classifying pollutants, this system optimizes GPT maintenance, enhancing its effectiveness and efficiency. The proposed system leverages real-time data from sensors, cameras, and weather forecasts, enabling authorities to implement proactive maintenance strategies based on accurate weather predictions and pollutant types. Logistic regression models forecast adverse weather conditions, while CNNs accurately classify garbage types, allowing targeted GPT cleaning during periods of increased pollutant buildup. The logistic regression model achieved an accuracy of 86.41%, and the CNN model attained a classification accuracy of 79.37%, showcasing strong performance in predicting weather conditions and categorizing pollutants. The integration of AI technologies in GPT maintenance significantly enhances environmental planning, mitigates flooding risks, and improves the accuracy of pollution monitoring. This solution provides valuable insights for decision-makers, helping them allocate resources effectively and maintain sustainable water management practices. In conclusion, the AI-driven system offers a robust and efficient approach to optimizing GPT operations, contributing to better environmental protection and urban sustainability.

Keywords: Convolution Neural Network, Gross Pollutant Traps, River Health and Pollutant Control

1. INTRODUCTION

In Malaysia, the Department of Irrigation and Drainage has strategically placed Gross Pollutant Traps (GPTs) to combat flooding and river pollution (Mohd Ahmed Hafez, 2019). Despite their importance, these GPTs face inefficiencies and inconsistencies due to the manual cleaning procedures conducted monthly since 2015, using nets and portable tools. Research has primarily focused on assessing GPT efficacy, leaving a crucial knowledge gap regarding the composition and volume of pollutants, which hampers effective environmental planning (Md Khalid, R., 2018).

To address these challenges, we propose a novel paradigm that revolutionizes GPT maintenance by integrating cutting-edge technology. Our framework optimizes GPT processes and overcomes the limitations of traditional cleaning techniques by leveraging artificial intelligence. Specifically, we employ logistic regression to predict weather patterns and Convolutional Neural Networks (CNNs) to classify pollutants (Masood & Ahmad, 2021). This approach enables a proactive and efficient GPT maintenance program while enhancing pollution identification accuracy.

Our system's foundation lies in the optimization and integration of these predictive models. Logistic regression forecasts adverse weather conditions that could increase pollutant influx, allowing for anticipatory GPT maintenance (Ma et al., 2020). Concurrently, CNNs transform garbage categorization, providing decision-makers with real-time insights into pollutant types and weather-related concerns. This comprehensive solution aims to transform environmental planning, strengthen river pollutant control methods, and usher in a new era of sustainable water management.

The use of artificial intelligence (AI) in environmental management has been gaining traction in recent years. Studies have shown that AI can significantly enhance the efficiency and effectiveness of environmental monitoring and maintenance systems. For instance, logistic regression models have been successfully used to predict weather

Publisher Note:
CV Media Inti Teknologi stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

patterns and their impact on environmental systems (Ma et al., 2020). These models help in anticipating adverse weather conditions, allowing for proactive measures to be taken to mitigate potential environmental damage.

Convolutional Neural Networks (CNNs) have also been widely used in environmental applications, particularly in the classification of pollutants. CNNs are capable of accurately identifying and categorizing various types of waste, which is crucial for effective waste management (Masood & Ahmad, 2021). This technology has been applied in various fields, including marine pollution monitoring and urban waste management, demonstrating its versatility and effectiveness.

The integration of AI in GPT maintenance is a relatively new concept, but it has shown promising results. A study by Masood and Ahmad (2021) highlighted the potential of using CNNs for real-time garbage classification, which can significantly improve the efficiency of GPT operations. Similarly, logistic regression models have been used to predict weather patterns and their impact on pollutant influx, allowing for more effective planning and maintenance of GPTs (Ma et al., 2020).

The implementation of this system enables continuous, real-time monitoring and collection of data concerning the types, quantities, and compositions of pollutants entering the Gross Pollutant Traps (GPTs) and the river system. This dynamic approach enhances the efficiency of GPTs by providing instant insights into the evolving pollution scenario (Fuertes et al., 2019). It ensures that authorities are alerted promptly to any pollution events, allowing for rapid response and minimizing environmental damage (Musthaq et al., 2024).

Traditional GPT maintenance involves manual cleaning once a month, which often results in challenges during periods of heavy rainfall and increased rubbish. The AI-powered system determines the most efficient cleaning schedule based on real-time data by using Garbage Classification. By doing so, it prevents unnecessary or delayed maintenance that can lead to flooding during adverse weather conditions (Meseguer & Quevedo, 2017). This optimization reduces the risk of flooding and ensures that GPTs are cleaned precisely when needed (da Costa et al., 2023).

The AI system not only aids in monitoring and controlling pollution but also supports long-term environmental planning for the sustainability of Malaysia's rivers. Reducing pollution and minimizing flood risks contribute to the overall well-being of aquatic ecosystems and the communities that rely on them. It aligns with the intention to maintain the ecological balance and preserve water resources for future generations.

This system employs continuous real-time monitoring and weather-predictive modeling to forecast and anticipate potential impacts. This involves predicting specific weather events, such as rain, drizzle, snow, or fog, that have been proven to affect pollutant buildup in Gross Pollutant Traps (GPTs). Using weather prediction models integrated with logistic regression, decision-makers can proactively plan maintenance activities. With such a proactive approach, authorities can prevent overflows, mitigate environmental damage, and improve the overall efficiency of GPT operations ahead of adverse weather events. Weather predictive modeling allows the system to ensure that GPT system performance is aligned with expected weather patterns, contributing significantly to environmental protection.

The Gross Pollutant Trap (GPT) initiative's Garbage Classification module seeks to reduce operational risks by accurately and promptly identifying different forms of waste. The system utilizes the predictive powers of the Convolutional Neural Network (CNN) to mitigate future issues like overflow and inefficiency in GPT operations. Targeted and preventive maintenance procedures are made possible by CNN's precise classification of various waste elements. During times of increased pollution influx, prompt intervention based on known waste kinds lowers the chance of GPT overflow. This method of risk mitigation ensures that the filtering infrastructure operates at its best, minimizes environmental harm, and supports the smooth operation of GPT systems (Nnamoko et al., 2022). By using intelligent waste classification, the goal is to improve the GPT system's dependability and effectively control pollution and the longevity of environmental conservation efforts.

By using logistic regression's prediction power, the Weather Prediction module is essential to reducing the risks connected to GPT operations. The goal of this proactive strategy is to deal with possible problems like overload or inefficiency in GPT systems. By accurately predicting unfavorable weather conditions, such as rain, drizzle, snow, or fog, the logistic regression model helps GPT maintenance decision-makers make timely decisions. The technology facilitates the smooth operation of GPT systems by resolving possible problems before they become more serious, hence lowering the possibility of environmental harm and guaranteeing the best possible performance from filtering systems. The goal is to make the GPT infrastructure more resilient to weather-related problems, resulting in a strong and effective system that supports efficient pollution control and environmental preservation.

The Smart River System for environmental planning and pollutant control serves a diverse range of users, each playing a vital role in river management and protection. Federal government departments responsible for environmental protection gain real-time insights and enhanced decision-making capabilities, enabling efficient resource allocation for tasks like pollution control and river cleanups. The system's advanced data analysis and integrated water quality sensors streamline compliance with regulations, making it an essential tool for sustainable river management.

2. MATERIAL AND METHOD

The establishment of a Smart River System demands a holistic and meticulous design approach, encompassing software, hardware, and artificial intelligence (AI) components. This interdisciplinary strategy is crucial for the systematic development and deployment of advanced functionalities within the system. The initial phase involves extensive research to inform data acquisition, employing techniques such as data cleansing and analysis for comprehensive preprocessing. This sets the foundation for subsequent stages, where machine learning and deep learning algorithms play pivotal roles. The chosen algorithms, linear regression for machine learning and Convolutional Neural Network for deep learning, are tailored to extract meaningful insights from the gathered data.

Real-time data processing is a cornerstone of the Smart River System, achieved through the integration of cameras for garbage classification and weather sensors utilizing linear regression models. The incorporation of machine learning models enhances the system's ability to make informed decisions based on the analyzed data. The culmination of the system's functionalities is realized in the visualization phase. This includes the implementation of a messaging and notification system, a dashboard for comprehensive monitoring, and remote management capabilities. The integration of water quality sensors further augments results accuracy and overall system efficacy, ensuring that the Smart River System operates intelligently and adaptively to dynamic environmental conditions.

In crafting this sophisticated system, the intricate interplay of software, hardware, and AI elements is a key consideration. The seamless integration of linear regression and Convolutional Neural Network algorithms, coupled with real-time data processing and advanced visualization, forms the foundation for a Smart River System that not only monitors and analyzes but also actively contributes to the preservation and enhancement of river ecosystems.

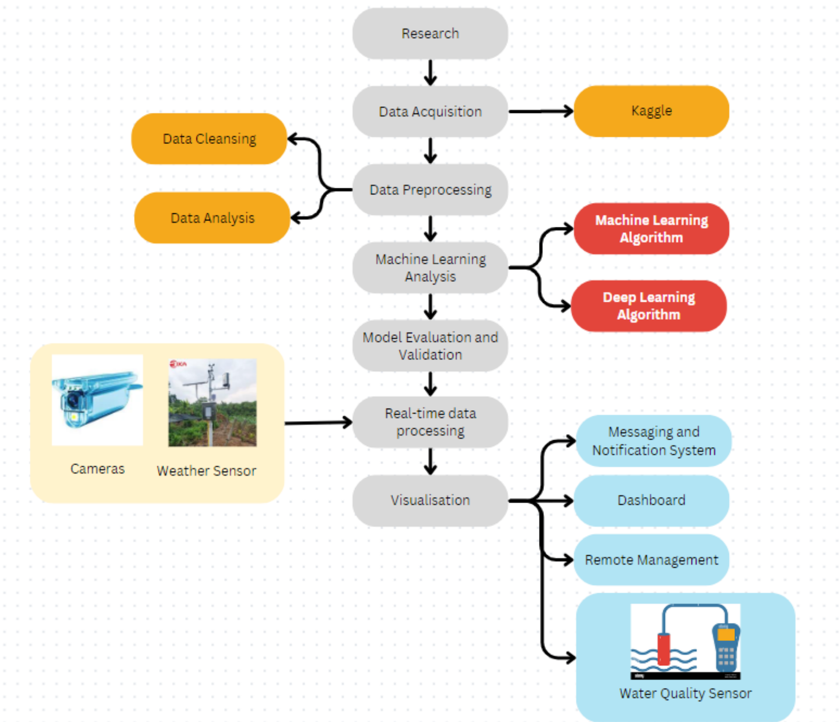


Figure 1. Overview of the operational framework of the application

This diagram represents the flow of an artificial intelligence-driven system designed for environmental monitoring and water quality management, particularly through the application of machine learning and deep learning techniques. The process begins with research, where problems such as water pollution and environmental concerns are explored in depth to understand data needs and the approach to managing water quality. Following this, the system collects data through data acquisition from multiple sources, including sensors, cameras, weather stations, and external databases like Kaggle, which help provide both real-time and historical data for analysis.

After data collection, the system moves into data preprocessing, where it undergoes two essential steps: data cleansing and data analysis. Data cleansing ensures that inaccurate or irrelevant data is removed, improving the quality of the dataset. In the data analysis phase, patterns are examined to prepare the data for the next step, which is machine learning analysis. At this point, the preprocessed data is analyzed through machine learning and deep learning algorithms to create predictive models that can detect pollution patterns, forecast future conditions, and identify anomalies in water quality.

Once the models are built, they go through a phase of model evaluation and validation. This step ensures that the machine learning models are reliable, accurate, and capable of making sound predictions based on real-time data. After validation, the system processes data in real time by receiving continuous inputs from various sources, including cameras, weather sensors, and water quality sensors, allowing for constant monitoring of environmental factors affecting the river.

The system then presents the processed data through visualization, offering users graphical insights into river conditions, pollution levels, and trends. Visualization is often displayed on a dashboard, which acts as a central hub where users, such as government authorities, environmental agencies, or the public, can monitor and interpret water quality metrics easily. The system is further equipped with a messaging and notification system, providing real-time alerts about pollution incidents or other significant environmental changes that require immediate attention.

Another crucial feature is remote management, enabling users to manage and control the system from a distance, ensuring prompt responses without the need for physical intervention. Finally, the integration of water quality sensors ensures that the system has continuous data input to monitor water conditions, detect pollutants, and maintain water bodies in alignment with environmental standards.

This AI-driven system integrates machine learning, real-time data collection, and advanced monitoring technologies to create an intelligent solution for environmental management and river pollution control. Continuous monitoring and data-driven insights facilitate efficient water management, timely interventions, and improved decision-making for all stakeholders.

3. RESULT AND DISCUSSION

The results of this study provide significant insights into the effectiveness of the Smart River System in managing Gross Pollutant Trap (GPT) operations and improving overall water quality monitoring. The system's integration of artificial intelligence (AI) with real-time data collection and machine learning algorithms has demonstrated enhanced efficiency in detecting pollutants, predicting pollution trends, and optimizing resource allocation for environmental interventions. This section discusses the outcomes of the implemented system, including its performance metrics, the accuracy of pollution detection, and the effectiveness of its predictive algorithms in various environmental scenarios.

Moreover, this section highlights the key findings from real-time data processing, which involved multiple sources such as water quality sensors, cameras, and weather sensors. The system's capacity to provide continuous monitoring and instant feedback through dashboards and notification systems has been evaluated to understand its impact on decision-making by local authorities, environmental agencies, and other stakeholders.

The discussion further explores the implications of these results in practical settings, comparing them with existing methods of water management and pollution control. The advantages, limitations, and areas for improvement in the Smart River System will also be examined, focusing on its potential scalability and adaptability to other regions or environmental contexts. Through this analysis, the study aims to provide a comprehensive understanding of how AI-driven solutions can enhance GPT operations and contribute to more sustainable river management.

Logistic Regression (Weather Prediction)

Logistic regression, despite its name, is primarily a classification method used to predict the likelihood of an input belonging to a specific category. It transforms the linear combination of input features into a probability value between 0 and 1 using the logistic or sigmoid function. This probability helps establish a decision boundary, classifying inputs based on whether the probability exceeds a defined threshold, typically 0.5.

In the context of the Gross Pollutant Trap (GPT) initiative, logistic regression is crucial for predicting weather patterns that affect the efficiency of environmental filtration systems. Classifying weather data indicating the need for GPT maintenance helps anticipate conditions like rain or fog that lead to increased pollutant buildup. This predictive capability allows for proactive maintenance planning, optimizing cleaning schedules, and mitigating risks of overflow or inefficiency. Logistic regression's interpretability also provides insights into how different weather features influence the need for GPT cleaning, supporting a data-driven, cost-efficient approach to environmental protection.

In this research, the dataset from Kaggle, which is linked at https://www.kaggle.com/ Datasets/ananthr1/weather-prediction is used. The Weather Prediction dataset on Kaggle, provided by user Ananthr1, is designed for teaching and exploring machine learning and deep learning techniques in the context of weather forecasting. The dataset contains multiple features such as temperature, humidity, wind speed, and precipitation, making it suitable for developing predictive models to forecast weather conditions based on historical data (Ananth R., 2021).

The dataset is structured to allow users to perform data preprocessing tasks such as data cleaning, normalization, and feature engineering, which are essential for building accurate models. This dataset's real-world applicability is significant, especially for projects involving environmental monitoring, climate analysis, or disaster preparedness. It also serves as a foundational dataset for experimentation with various algorithms, including decision trees, random forests, and neural networks, to predict outcomes like rain or other weather events. The use of this dataset can lead to a deeper understanding of how weather patterns evolve and the potential to create systems that offer precise predictions.

To build our Logistic Regression model, we acquired the dataset and initial code from Kaggle.com. The dataset comprises various columns, as outlined in Table 1 and Figure 2. For our model, we specifically chose to focus on key features, namely precipitation, temp_max, temp_min, and wind. These selected features were utilized as the foundation for our model training process.

Table 1. The column name of the dataset

Feature	**Description**
date	The date of data collection for weather records
precipitation	Various types of precipitation occurring on land surfaces and in open water bodies, including rainfall, sleet, snow, hail, or drizzle
temp_max	Maximum Temperature
temp_min	Minimum Temperature
wind	Wind speed
weather	Labeled output

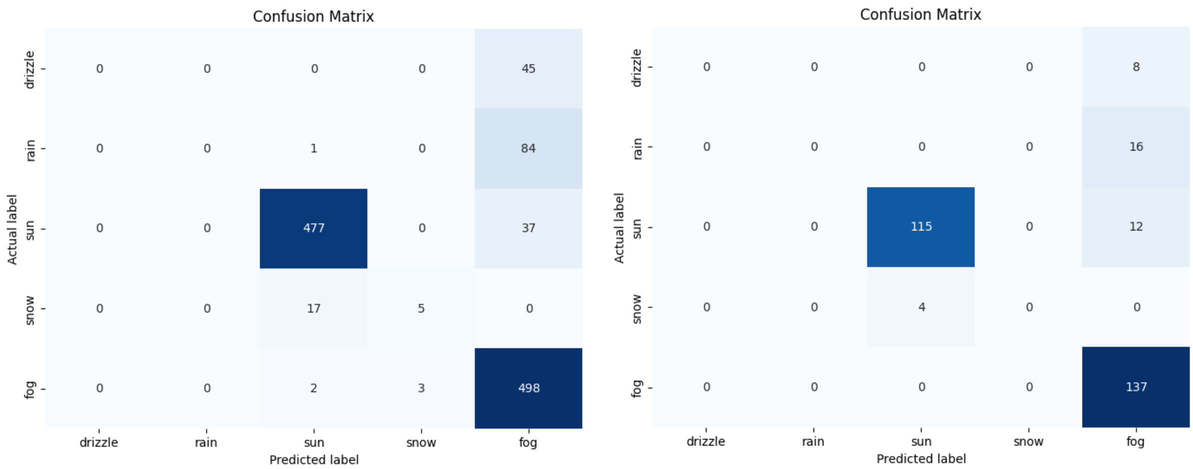
	date	precipitation	temp_max	temp_min	wind	weather
0	2012-01-01	0.0	12.8	5.0	4.7	drizzle
1	2012-01-02	10.9	10.6	2.8	4.5	rain
2	2012-01-03	0.8	11.7	7.2	2.3	rain
3	2012-01-04	20.3	12.2	5.6	4.7	rain
4	2012-01-05	1.3	8.9	2.8	6.1	rain
...	...	...	...	...	...	...
1456	2015-12-27	8.6	4.4	1.7	2.9	rain
1457	2015-12-28	1.5	5.0	1.7	1.3	rain
1458	2015-12-29	0.0	7.2	0.6	2.6	fog
1459	2015-12-30	0.0	5.6	-1.0	3.4	sun
1460	2015-12-31	0.0	5.6	-2.1	3.5	sun

1461 rows × 6 columns

Figure 2. The structure of the dataset

Figure 3 shows the dataset consists of 6 columns, including the last column, the label of the dataset, the weather, and 1461 rows.

The initial code, sourced online, commences by loading a weather dataset and segregating it into predictor features and target variables. Following this, the data is divided, with 80% assigned to model training and 20% for testing purposes. A logistic regression model is initialized, trained on the training dataset, and evaluated for accuracy on both the training and test sets. It displays a training accuracy of 86.30% and a slightly lower test accuracy of 83.83%, as shown in Figure 3.



(a) Training confusion matrix (b) Testing confusion matrix

Classification Report:

	precision	recall	f1-score	support
drizzle	0.00	0.00	0.00	45
rain	0.00	0.00	0.00	85
sun	0.96	0.93	0.94	514
snow	0.62	0.23	0.33	22
fog	0.75	0.99	0.85	503
accuracy			0.84	1169
macro avg	0.47	0.43	0.43	1169
weighted avg	0.76	0.84	0.79	1169

Classification Report:

	precision	recall	f1-score	support
drizzle	0.00	0.00	0.00	8
rain	0.00	0.00	0.00	16
sun	0.97	0.91	0.93	127
snow	0.00	0.00	0.00	4
fog	0.79	1.00	0.88	137
accuracy			0.86	292
macro avg	0.35	0.38	0.36	292
weighted avg	0.79	0.86	0.82	292

(c) Training classification report (d) Testing classification report

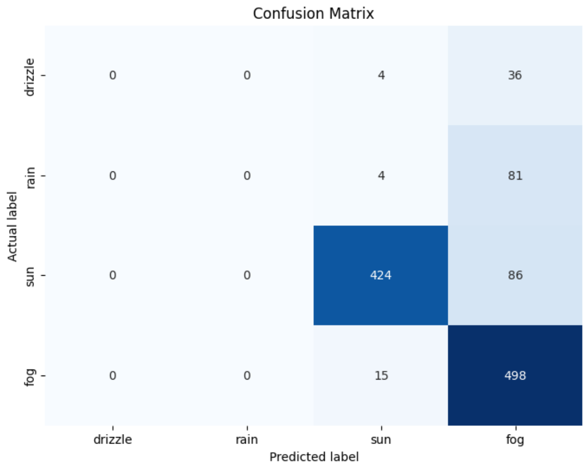
Figure 3. Confusion matrix and classification report

Logistic Regression Optimization

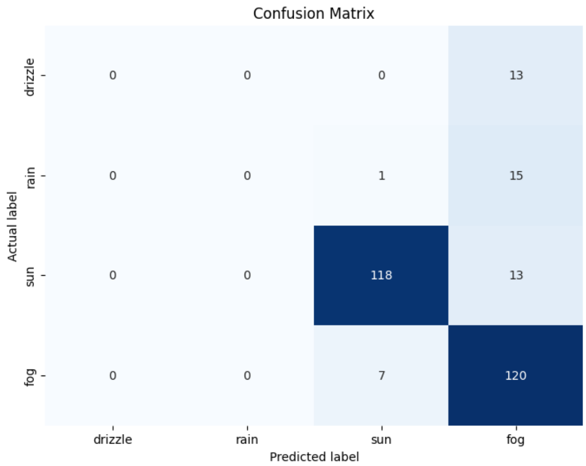
In our analysis of Malaysia's weather patterns, we excluded rows where the weather column indicated 'snow,' as snow is not typical in Malaysia. This ensures our machine-learning model is trained and tested on data that accurately represents the local climate, enhancing its ability to predict GPT maintenance needs specific to Malaysia's weather conditions.

The updated code improves the logistic regression model for the GPT project by including 'StandardScaler()' for feature standardization. This ensures uniform scaling across all variables, preventing biases from differing data ranges. Addressing varied feature scales is crucial for model performance, enhancing its ability to identify patterns and improving accuracy and reliability in forecasting weather patterns related to GPT maintenance.

The adjustment in model fitting emphasizes training the model exclusively on the training data, a vital precaution to avert overfitting and guarantee an impartial assessment on the test dataset. This refinement notably bolsters the model's ability to generalize, resulting in more reliable and precise forecasts for unexpected weather scenarios linked to GPT maintenance planning. After these adaptations, the accuracy of both the training and test models shows enhancement, reaching 85.37% and 86.41%, albeit with a marginal increase compared to the initial model in the test set, as shown in Figure 4.



(a) Training confusion matrix



(b) Testing confusion matrix

Classification Report:

	precision	recall	f1-score	support
drizzle	0.00	0.00	0.00	40
rain	0.00	0.00	0.00	85
sun	0.95	0.83	0.89	510
fog	0.71	0.97	0.82	513
accuracy			0.80	1148
macro avg	0.41	0.45	0.43	1148
weighted avg	0.74	0.80	0.76	1148

(c) Training classification report

Classification Report:

	precision	recall	f1-score	support
drizzle	0.00	0.00	0.00	13
rain	0.00	0.00	0.00	16
sun	0.94	0.90	0.92	131
fog	0.75	0.94	0.83	127
accuracy			0.83	287
macro avg	0.42	0.46	0.44	287
weighted avg	0.76	0.83	0.79	287

(d) Testing classification report

Figure 4. GridSearchCV for Logistic Regression Model

Incorporating ‘GridSearchCV’ stands as a critical step, systematically delving into various hyperparameters to pinpoint the most effective combination for the logistic regression model. Adjusting parameters like ‘C’ for regularization strength, ‘solver,’ and ‘max_iter’ is pivotal for refining the model’s efficacy. This meticulous process elevates the model's performance by identifying the most fitting configurations, ultimately aligning it more precisely with the distinct demands of the GPT project, as shown in Figure 5.

```
param_grid = {
    'C': [0.001, 0.01, 0.1, 1, 10, 100],  # Regularization parameter
    'solver': ['liblinear', 'lbfgs', 'sag', 'newton-cg'],  # Solver types
    'max_iter': [100, 200, 300]  # Maximum iterations
}
model = LogisticRegression()

grid = GridSearchCV(model, param_grid, cv=5)
grid.fit(x_train, y_train)
```

```
print("Best Parameters:", grid.best_params_)
print("Best Score:", grid.best_score_)
```

```
Best Parameters: {'C': 100, 'max_iter': 100, 'solver': 'lbfgs'}
Best Score: 0.843254688445251
```

Figure 5. GridSearchCV of Logistic Regression Model for the best parameter and score

The implementation of a predictive feature for GPT maintenance based on forecasted weather significantly enhances the system by making it both predictive and prescriptive. This feature allows for proactive decision-making by recognizing weather patterns historically linked to increased pollutant accumulation in GPTs. It triggers timely alerts to operators, indicating potential cleaning needs based on anticipated weather conditions, thus enabling more effective and proactive maintenance plans. This proactive strategy is crucial for the GPT project, as it helps manage the impact of various weather conditions on pollutant flow. For instance, rainfall and drizzle often necessitate immediate or more frequent cleaning, while fog can indirectly introduce pollutants. Even sunny weather can lead to the gradual accumulation of debris, requiring long-term cleaning. Understanding these diverse weather impacts allows for tailored and prompt maintenance strategies, aligning the GPT project with proactive environmental stewardship.

Convolutional Neural Network (CNN) for Garbage Classification

Convolutional neural networks (CNNs) are artificial neural networks (ANN) that are well-suited for image recognition and analysis. In the Gross Pollutant Trap (GPT) project context, a Convolutional Neural Network (CNN) is utilized for garbage classification. The CNN model is designed to analyze and categorize various types of waste and pollutants. The model operates by employing a series of convolutional layers, pooling layers, and fully connected layers. These layers work together to extract features from input images and classify them into different garbage types, as shown in Figure 6.

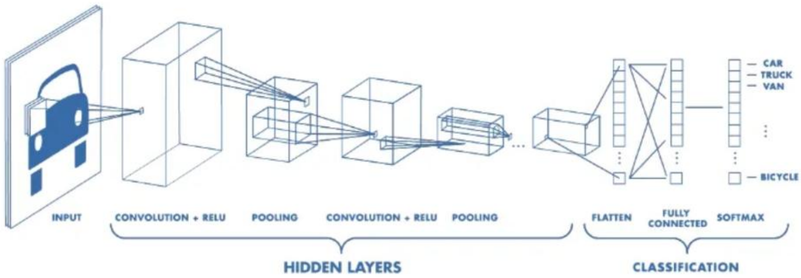


Figure 6. Architecture of a CNN (Mayank Mishra, 2020)

Convolutional layers in a CNN detect patterns and features in input images by applying filters and performing convolution operations, resulting in feature maps that highlight important image details. Pooling layers, such as max pooling or average pooling, reduce the spatial dimensions of these feature maps while retaining key information, making the model more robust and computationally efficient. Fully connected layers then classify the images based on the extracted features by connecting every neuron in one layer to every neuron in the next.

During training, the CNN model optimizes its weights and biases through backpropagation, minimizing a loss function based on the error between predicted outputs and ground truth labels. Once trained, the model can classify new images by passing them through its layers and producing a probability distribution over different garbage types, assigning the image to the type with the highest probability.

In this research, the dataset from Kaggle, which is linked at https://www.kaggle.com/datasets/asdasdasasdas/garbage-classification is used. The Garbage Classification dataset on Kaggle, provided by user asdasdasasdas, is designed for machine learning tasks related to waste classification. This dataset primarily focuses on classifying various types of waste into different categories, which is crucial for projects aimed at promoting recycling and effective waste management. The dataset includes images of real-life garbage and contains six distinct classes of waste: cardboard, glass, metal, paper, plastic, and trash (CChange CS., 2018)) This dataset is ideal for deep learning models, particularly for image classification tasks like smart recycling systems or garbage sorting facilities. It includes six classes (Cardboard, Glass, Metal, Paper, Plastic, Trash) and is structured for convolutional neural networks (CNNs) or transfer learning models. The initial CNN model processes image data, sets up data generators, constructs a neural network architecture, and trains the model, achieving an accuracy of 76.786%. Data augmentation techniques and feature standardization are used to enhance model performance.

The CNN model is built using the Sequential API in Keras, with several convolutional blocks followed by classification layers. It includes Conv2D layers with max pooling, a flattened layer, Dense layers with ReLU activation, and dropout layers to prevent overfitting. The model is compiled with ‘categorical_crossentropy’ loss and ‘adam’ optimizer and trained for 100 epochs using data generators for training and validation. The training process includes steps for parallel processing and model checkpointing to save the best model based on validation accuracy. This structured approach ensures robust training and accurate image classification.

In the final training epoch (Epoch 100), the model achieved an approximate training accuracy of 79.37% ('acc') with a corresponding loss of 0.5887. Notably, the validation accuracy ('val_acc') exhibited improvement from around 76.79% to 77.68%. This progress prompted the model to save its state, optimize performance, and showcase its ability to generalize well to unseen data, as shown in Figure 7.

```
model.compile(loss='categorical_crossentropy',
              optimizer='adam',
              metrics=['acc']) # RMS PROP - No accuracy

#es=EarlyStopping(monitor='val_loss', mode='min', verbose=1, patience=50)
```

```
history = model.fit(train_generator,
                    epochs=100,
                    steps_per_epoch=2276//32,
                    validation_data=test_generator,
                    validation_steps=251//32,
                    workers = 4,
                    callbacks=callbacks_list)
```

```
71/71 [==============================] - ETA: 0s - loss: 0.5805 - acc: 0.7897
Epoch 96: val_acc did not improve from 0.76786
71/71 [==============================] - 129s 2s/step - loss: 0.5805 - acc: 0.7897 - val_loss: 0.8134 - val_acc: 0.7545
Epoch 97/100
71/71 [==============================] - ETA: 0s - loss: 0.5496 - acc: 0.8035
Epoch 97: val_acc did not improve from 0.76786
71/71 [==============================] - 124s 2s/step - loss: 0.5496 - acc: 0.8035 - val_loss: 0.9625 - val_acc: 0.7232
Epoch 98/100
71/71 [==============================] - ETA: 0s - loss: 0.5586 - acc: 0.8057
Epoch 98: val_acc did not improve from 0.76786
71/71 [==============================] - 125s 2s/step - loss: 0.5586 - acc: 0.8057 - val_loss: 0.8748 - val_acc: 0.7589
Epoch 99/100
71/71 [==============================] - ETA: 0s - loss: 0.5555 - acc: 0.7968
Epoch 99: val_acc did not improve from 0.76786
71/71 [==============================] - 124s 2s/step - loss: 0.5555 - acc: 0.7968 - val_loss: 0.7927 - val_acc: 0.7455
Epoch 100/100
71/71 [==============================] - ETA: 0s - loss: 0.5887 - acc: 0.7937
Epoch 100: val_acc improved from 0.76786 to 0.77679, saving model to trained_model.h5
71/71 [==============================] - 126s 2s/step - loss: 0.5887 - acc: 0.7937 - val_loss: 0.7465 - val_acc: 0.7768
```

Figure 7. 100 epochs of the proposed CNN model

CNN Optimization

The optimized model enhances the original architecture by integrating BatchNormalization after each Conv2D layer, which standardizes input data and promotes stable, faster convergence during training. The model also features increased depth and complexity with larger filter sizes in convolutional layers, allowing it to detect more intricate data patterns. Additionally, the inclusion of a higher-dimensional Dense (512) layer before the final classification layer helps extract sophisticated features, improving the model's ability to handle complex datasets and generalize to new data.

The original architecture used Conv2D layers with MaxPooling2D layers to extract and down-sample features, followed by dense layers for classification. The optimized model, however, includes BatchNormalization and Dropout layers after each Conv2D layer, with increasing filter sizes (32, 64, 128, 256). This design culminates in a Dense (512) layer with BatchNormalization and a higher dropout rate before the final Dense (6) layer with softmax activation. During training over 100 epochs, the model achieved a training accuracy of 80.05% but showed a validation accuracy of 46.61%, indicating a performance gap between training and validation datasets, as shown in Figure 8.

```
history = model.fit(train_generator,
                    epochs=100,
                    steps_per_epoch=len(train_generator),
                    validation_data=test_generator,
                    validation_steps=len(test_generator),
                    workers=4,
                    callbacks=callbacks_list)
```

```
72/72 [==============================] - ETA: 0s - loss: 0.2464 - acc: 0.9121
Epoch 96: val_acc did not improve from 0.78486
72/72 [==============================] - 591s 8s/step - loss: 0.2464 - acc: 0.9121 - val_loss: 4.1966 - val_acc: 0.3904
Epoch 97/100
72/72 [==============================] - ETA: 0s - loss: 0.5456 - acc: 0.8304
Epoch 97: val_acc did not improve from 0.78486
72/72 [==============================] - 635s 9s/step - loss: 0.5456 - acc: 0.8304 - val_loss: 12.7271 - val_acc: 0.1873
Epoch 98/100
72/72 [==============================] - ETA: 0s - loss: 1.0768 - acc: 0.6656
Epoch 98: val_acc did not improve from 0.78486
72/72 [==============================] - 620s 9s/step - loss: 1.0768 - acc: 0.6656 - val_loss: 8.1055 - val_acc: 0.2311
Epoch 99/100
72/72 [==============================] - ETA: 0s - loss: 0.6997 - acc: 0.7491
Epoch 99: val_acc did not improve from 0.78486
72/72 [==============================] - 625s 9s/step - loss: 0.6997 - acc: 0.7491 - val_loss: 3.4709 - val_acc: 0.3426
Epoch 100/100
72/72 [==============================] - ETA: 0s - loss: 0.5238 - acc: 0.8005
Epoch 100: val_acc did not improve from 0.78486
72/72 [==============================] - 610s 8s/step - loss: 0.5238 - acc: 0.8005 - val_loss: 2.3760 - val_acc: 0.4661
```

Figure 8. 100 epochs of the proposed CNN Optimization model

Integration of Both Models (Weather Prediction and Garbage Classification)

The code snippet integrates weather predictions and garbage classification to provide actionable insights for GPT maintenance. It uses a placeholder for weather prediction, currently set to ‘rain,’ to recommend immediate cleaning during adverse weather. The CNN model’s output for garbage classification, represented by hypothetical probabilities, determines the detected garbage type based on a threshold. The combined output advises on maintenance actions based on weather conditions and identified garbage types, promoting proactive waste management.

The integration of these predictions allows for context-specific guidance, suggesting immediate GPT cleaning during adverse weather and identifying specific garbage types for tailored handling. This approach emphasizes the practical use of AI models in optimizing waste management strategies by considering dynamic environmental conditions and waste characteristics, ensuring effective and timely maintenance.

4. CONCLUSION

The integration of weather prediction and garbage classification models significantly enhances decision-making for GPT maintenance. By identifying adverse weather conditions like rain or drizzle, the system prompts immediate cleaning, ensuring GPTs remain effective during conditions that increase debris. The precise garbage classification model further aids in tailored disposal procedures, improving waste management strategies.

This system’s impact extends to environmental protection by enabling pollution monitoring and optimizing GPT maintenance schedules. It helps mitigate ecological damage and reduces flooding risks during adverse weather. Additionally, it supports city planning by providing insights into pollutant control and resource allocation, benefiting urban development and sustainability.

Various user groups, including federal departments, local authorities, environmental NGOs, research institutions, the general public, and emergency services, benefit from this AI-powered system. It aids in informed decision-making, compliance with environmental standards, innovative conservation solutions, pollution reporting, and swift response to environmental emergencies. The development methodology emphasizes flexibility, data-driven decision-making, and iterative progress. It involves thorough research, meticulous data collection, and tailored machine-learning models like logistic regression for weather prediction and CNNs for garbage classification. Continuous testing and optimization ensure the models’ accuracy and performance, making this system a testament to innovation and efficiency in environmental sustainability and urban planning.

AUTHOR INFORMATION

Corresponding Authors

Misinem, Universitas Bina Darma, Palembang, Indonesia.

Email: misinem@binadarma.ac.id

Authors

Chang Shi Ying, INTI International University, Malaysia.

Bong Peak May, INTI International University, Malaysia.

Soo Ting Fang, INTI International University, Malaysia.

Lee Wai Yi, INTI International University, Malaysia.

REFFERENCE

Ananth R. (2021). Retrieved October 1, 2023. *Weather Prediction*. https://www.kaggle.com/ datasets/ananthr1/weather-prediction

AP Dr. Mohd Ahmed Hafez, D. Prof. Ir. Dr. L. M. S. Ir. H. B. D. Ir. M. A. I. A. D. R. C. O. (2019). *Gross Pollutant Traps to Enhance Water Quality in Malaysia*. Partridge Publishing.

CChange CS. (2018). *Garbage Classification*. https://www.kaggle.com/datasets/asdasdasasdas

/garbage-classification

da Costa, T. P., Gillespie, J., Cama-Moncunill, X., Ward, S., Condell, J., Ramanathan, R., & Murphy, F. (2023). A Systematic Review of Real-Time Monitoring Technologies and Its Potential Application to Reduce Food Loss and Waste: Key Elements of Food Supply Chains and IoT Technologies. *Sustainability*, *15*(1). https://doi.org/10.3390/su15010614

Fuertes, W., Cadena, A., Torres, J., Benítez, D., Tapia, F., & Toulkeridis, T. (2019). Data Analytics on Real-Time Air Pollution

Monitoring System Derived from a Wireless Sensor Network. In Á. Rocha, C. Ferrás, & M. Paredes (Eds.), *Information Technology and Systems* (pp. 57–67). Springer International Publishing.

Ma, Y., Ma, B., Jiao, H., Zhang, Y., Xin, J., & Yu, Z. (2020). An analysis of the effects of weather and air pollution on tropospheric ozone using a generalized additive model in Western China: Lanzhou, Gansu. *Atmospheric Environment*, *224*, 117342. https://doi.org/https://doi.org/10.1016/j.atmosenv.2020.117342

Masood, A., & Ahmad, K. (2021). A review on emerging artificial intelligence (AI) techniques for air pollution forecasting: Fundamentals, application and performance. *Journal of Cleaner Production*, *322*, 129072. https://doi.org/https://doi.org/10.1016/j.jclepro.2021.129072

Mayank Mishra, 2020, Medium: Towards Data Science: Convolutional Neural Networks, Explained, Available at: https://towardsdatascience.com/convolutional-neural-networks-explained-9cc5188c4939

Md Khalid, R. (2018). REVIEW OF THE WATER SUPPLY MANAGEMENT AND REFORMS NEEDED TO ENSURE WATER SECURITY IN MALAYSIA. In *International Journal of Business and Society* (Vol. 19, Issue 3).

Meseguer, J., & Quevedo, J. (2017). Real-Time Monitoring and Control in Water Systems. In V. Puig, C. Ocampo-Martínez, R. Pérez, G. Cembrano, J. Quevedo, & T. Escobet (Eds.), *Real-time Monitoring and Operational Control of Drinking-Water Systems* (pp. 1–19). Springer International Publishing. https://doi.org/10.1007/978-3-319-50751-4_1

Nnamoko, N., Barrowclough, J., & Procter, J. (2022). Solid Waste Image Classification Using Deep Convolutional Neural Network. *Infrastructures*, *7*(4). https://doi.org/10.3390/infrastructures7040047