



OPEN ProWaste for proactive urban waste management using IoT and machine learning

Thompson Stephan¹, S. M. Hari Krishna², Chia-Chen Lin³✉, Ujjwal Sumesh², Saurabh Agarwal⁴✉ & Hyunsung Kim⁵✉

Urban waste-collection centres (WCCs) routinely overflow because maintenance routes are scheduled reactively rather than on data-driven forecasts. Overspill, odour, and leachate therefore threaten public health and sustainability targets in rapidly growing smart cities. We introduce ProWaste, an end-to-end Internet-of-Things and machine-learning platform that proactively prioritises WCC servicing. Fifteen automated and manual indicators, including population density, weather, maintenance history, and weekly waste build-up, are streamed from low-cost sensors, public APIs, and a mobile app to a cloud database. Twenty-five off-the-shelf classifiers were benchmarked under repeated stratified cross-validation; a Decision Tree Classifier offered the best balance of interpretability and near-top accuracy. Binary Particle Swarm Optimisation (BPSO) removed 80% of the inputs, revealing that three features alone predict criticality with >99% accuracy on a hold-out test set. SHAP analysis confirms the interpretability of the three-feature model. The predicted class and confidence score are pushed to a Sustainable Smart Waste Management (SSWM) app that alerts field teams and dynamically reorders maintenance queues. Compared with current practice, ProWaste can eliminate missed pickups while reducing on-road inspections and data bandwidth. The proposed architecture is readily transferable to other cities and can be extended to recycling or composting streams.

Keywords Smart waste management, Internet of things, Classification, Wrapper-based optimization, Feature selection, Urban waste collection

The growth of urban populations has significantly increased city and municipal garbage generation, making efficient waste management a critical concern. Urban Waste Collection Centers (WCCs) can utilize smart waste management systems that track garbage levels and enable prompt collection, preventing health hazards¹ caused by waste buildup. Integrated systems that include small data centers, alarm systems, and mobile devices used by waste collectors and monitoring officers can significantly improve waste management. Proper waste management can reduce environmental degradation, climate change problems, and public health issues. The COVID-19 outbreak has also raised awareness of the need to properly dispose of household garbage, which is often dumped in urban waste collection centers. Urban waste collection centers are typically smaller in size than municipal landfills, which are designed to handle waste from larger areas, including commercial and industrial waste. Bengaluru, like many other urban areas in India, has struggled with waste management and faced challenges related to waste collection centers. These centers can cause significant environmental and health problems if waste is disposed of without proper treatment or management. WCCs located near residential areas can release toxic fumes and pollutants into the air, causing respiratory disorders and other medical conditions affecting the local population. Waste disposal centers can also pollute groundwater and soil, causing long-term harm to the environment.

To address these issues, many urban waste collection centers are now adopting Internet of Things (IoT)-based waste management systems. These systems leverage IoT technology to improve garbage disposal by making garbage collection routes more efficient and by offering real-time data on waste levels. The capability of the IoT to improve human living globally enables it to be a promising technology for the waste management

¹Thumbay College of Management and AI in Healthcare, Gulf Medical University, Ajman, UAE. ²Department of Computer Science and Engineering, M. S. Ramaiah University of Applied Sciences, Bangalore 560054, India.

³Department of Computer Science and Information Engineering, National Chin-Yi University of Technology, Taichung 411, Taiwan. ⁴School of Computer Science and Engineering, Yeungnam University, Gyeongsan 38541, Republic of Korea. ⁵Department of Smart Security, Kyungil University, Gyeongsan 38428, Republic of Korea.

✉email: ally.cclin@ncut.edu.tw; saurabh@yu.ac.kr; kim@kiu.ac.kr

industry. IoT facilitates communication and connectivity between low-power devices and online activities. Numerous applications worldwide have been incorporating various IoT-based activities to deliver cutting-edge services for smart cities². After reaching USD 2,733.1 million in 2024 with 15.6% year-on-year growth, the global smart waste management market is forecast to hit USD 3,170.5 million in 2025 and expand at a 16% CAGR to USD 13,986.4 million by 2035³. The smart waste management industry is driven by global concern over solid waste and urbanization, particularly in the context of urban waste collection centers (WCCs). By utilizing IoT, smart devices, and sensors, these centers can optimize waste collection schedules and reduce costs caused by operational inefficiencies. Waste creation varies by location and season, making it impractical to remove garbage cans regularly. The World Bank estimates that 2.24 billion tonnes of solid garbage were generated worldwide in 2020. The yearly output of garbage is anticipated to rise to 3.88 billion tonnes by 2050, up 73% from 2020. As IoT technology advances in this industry, it becomes possible to reduce costs, manage resources more effectively, and improve the sustainability of waste services, particularly in cities where trash management has become more prevalent and increasingly challenging.

In India, the waste management market is growing due to industrial activity and population density. However, the lack of appropriate infrastructure and rules for collection, disposal, and recycling is causing waste management to be in a terrible state, particularly in urban areas. Every year, India generates 62 million tonnes of garbage, of which 70% is gathered, 12 million is managed, and 31 million is disposed of in landfills⁴. By 2030, urban municipal solid waste output is anticipated to reach 165 million tonnes. Additionally, the waste recycling and management industry is projected to be valued at US\$ 13.62 billion by 2025, experiencing a yearly growth rate of 7.17%⁴. The problem of trash in India is severe, and the growing urban population is adding to the challenge of developing efficient urban waste management techniques. To tackle this issue, many Indian towns are utilizing real-time data to monitor rubbish collection and enhance street sanitation to clean up their streets. By combining all of this real-time data, quick judgments can be made to optimize garbage collection routes, collection times, fuel usage, and response times, for example, by overlaying data into a GIS base map⁵.

Many governments have recognized the need for waste management and have taken measures to implement it as part of their Smart Communities and Clean Development Mission programs. WCCs serve as central locations for the collection, segregation, and processing of waste, and they allow for the efficient transport of waste to landfills or recycling facilities. The fast rise in urban populations due to migration and unplanned urbanization has increased the amount of urban waste and changed the characteristics of urban waste. To design a long-term waste management strategy that meets the goals of government efforts, it is essential to figure out the types, volumes, and current waste management techniques⁶. Most household waste ends up in municipal solid waste landfills, which have strict safety and monitoring regulations⁷. However, the uneven dispatch of maintenance teams to different urban WCCs, as shown in Fig. 1 for the city of Bangalore, can lead to some areas being neglected and tons of waste going uncollected. The implementation of WCCs can help address this issue by allowing for the efficient collection and transport of waste to the appropriate disposal or recycling facility. Figure 1a shows the comparative analysis of total waste collected and wet waste ratio across Bangalore zones. Blue lines represent total waste (solid for April, dashed for May), and red dashed lines represent the wet waste ratio. Figure 1b shows the capacity of various waste collection sources. Auto Tippers lead with the highest capacity, followed by Compactors, Wet Waste Processing Plants, Bio-Methanation Plants, and Sanitary Landfills. Figure 1c depicts the total waste collected in Bangalore zones for April and May 2024. Blue bars for April and orange bars for May show a general increase in waste collection from April to May. Figure 1d illustrates the average chemical composition of municipal solid waste for 2024.

The Proactive Waste Management project is inspired by the necessity for efficient, data-driven, and optimal solid waste management strategies in smart cities. Proactive maintenance of WCCs is essential to attaining this goal, and their implementation is necessary to create sustainable and intelligent cities.

Smart cities face the challenge of implementing a proactive waste management program that connects various sectors, including Urban WCCs. Smart city consultants can evaluate and develop waste management strategies, while IoT and associated technologies can address some problems. Machine learning is used in this work to analyze data from IoT sensors for proactive waste management solutions, including data collected from WCCs. Regression and classification are two efficient ML-based solutions for IoT-based applications⁸. This is due to three key factors: First, all of the devices in IoT applications are linked, and large volumes of data are collected regularly. Moreover, activities could be preprogrammed to trigger based on intriguing responses from collected data or specific conditions. Secondly, computers can learn and perform tasks such as categorization. These systems are trained using various algorithms that analyze test data. Thirdly, sample data is characterized by measurable attributes (known as features), and machine learning algorithms aim to predict output values by identifying relationships between these features (referred to as labels). After training, the knowledge acquired is applied to identify trends or make judgements based on fresh data. By installing sensors in WCCs, data on the type and quantity of waste can be collected, helping to inform waste management strategies. For example, if data shows that a particular WCC is becoming more critical than others, additional maintenance teams can be dispatched to that WCC to ensure timely waste collection. Machine learning (ML) schemes utilize this information to make predictions about waste generation patterns, enabling proactive waste management measures to be taken. Furthermore, by integrating IoT sensors and ML schemes, WCCs may be optimized to be more efficient and reduce costs associated with waste management.

The structure of the manuscript is as follows: Section “[Related work](#)” surveys prior work and positions the present research, closing with the specific gaps that ProWaste addresses. Section “Proposed waste management system: ProWaste”³ presents the ProWaste framework—sensor stack, data-flow architecture, mobile app, dataset composition, and preprocessing steps. Section “[Feature selection](#)” explains the wrapper-based feature-selection strategy, defining the BPSO objective, particle encoding, and search procedure. Section “[Performance evaluation](#)” describes the experimental design and results, including the 25-model benchmark, convergence curves, three-

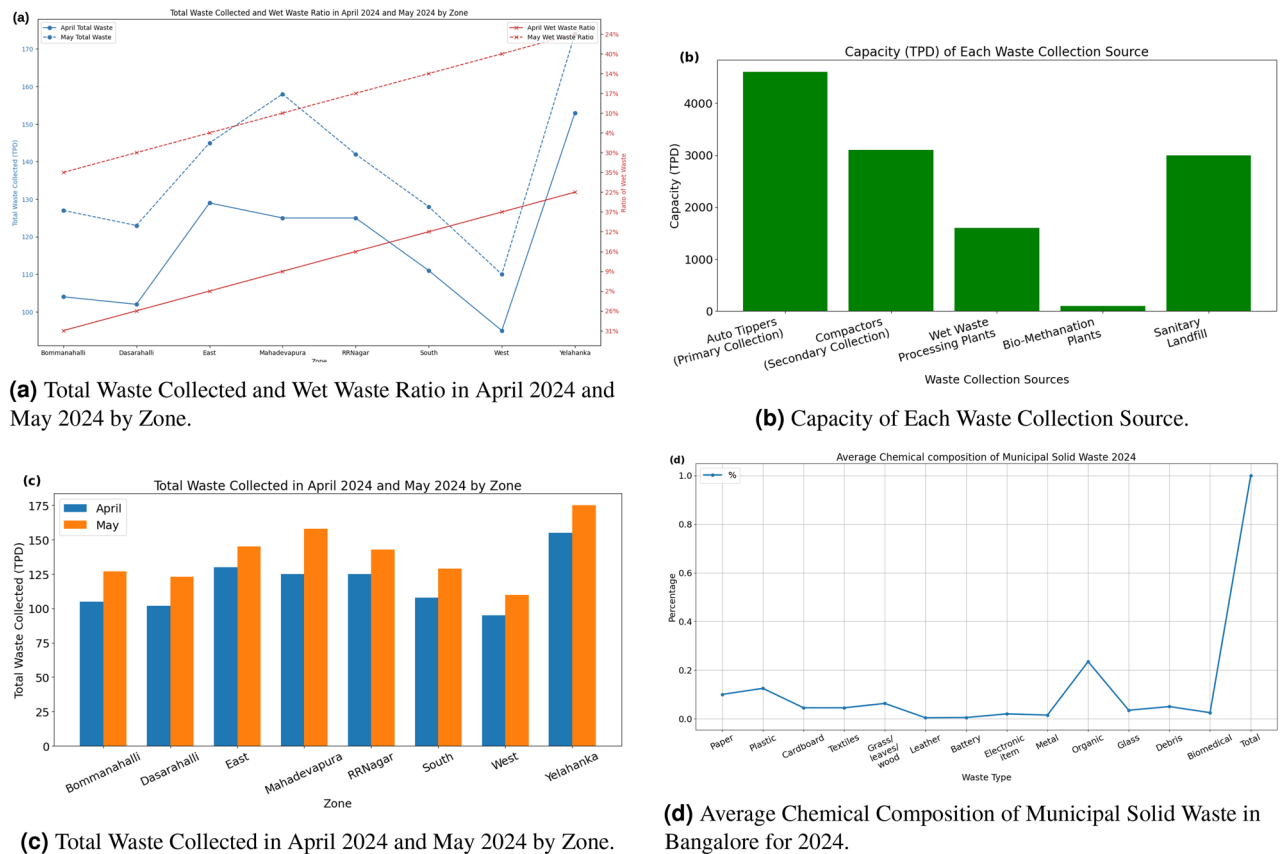


Fig. 1. Visual summary of waste data and characteristics across different zones.

feature decisive model, and SHAP interpretability analysis. Section “**Conclusion**” summarises the findings, discusses practical implications for municipal operations, and outlines directions for extending ProWaste to additional waste streams and city contexts.

Related work

Urban Waste Collection Centers (WCCs) play a critical role in managing urban waste, yet proactive solutions for overspill and gas emissions are limited; this study introduces ProWaste, an IoT and machine learning system for WCC management, building on prior works utilizing IoT, ultrasonic sensors, and automation techniques in waste management. Patil et al. offered an intelligent e-waste management system with an ultrasonic sensor that measures trash volume and sends a notification to a city web server for sanitation⁹. An Android app is used to receive notifications, reducing the need for human monitoring. The challenge of energy-efficient data collection in Smart Cities is also addressed in the article. Ijmaru et al.¹⁰ proposed an Internet of Vehicles-based data collection technique for waste management in Smart Cities, with energy consumption analysis and preference for LoRaWAN protocol. Vishnu et al.¹¹ conducted a study on automation approaches in waste management using RFID, WSN, and IoT-based systems, and concluded that IoT-based systems with LoRaWAN protocol outperform other approaches. Shanthini et al.¹² proposed an IoT-based smart trash can connected via an IoT system for waste management in urban areas, using a Node MCU microcontroller, waterproof sensors, and Wi-Fi connectivity. Prastyabudi et al.¹³ proposed a conceptual design for a smart trash can connected via an IoT system, with data classification, concept identification, and conceptual model building. Limitations of the study include the lack of evidence of the concept’s versatility and verification.

Municipal solid waste management (MSWM) is an essential aspect of modern society’s supply chain. However, the cost associated with MSWM is a significant problem that needs to be reduced. To tackle this issue, Navid et al.¹⁴ proposed a method that employs two sub-models to reduce operational costs. The first sub-model utilizes the Vehicle Routing Problem (VRP) to route fleets between waste generation and separation facilities. The second sub-model aims to distribute resources from a separation facility to various treatment plants or landfills. To minimize overall transportation costs and maximize recycling revenues, this study uses random constraint programming to work with stochastic optimization models. This research indicates that determining the ideal number of vehicles can support managers and policymakers in tough circumstances. Another solution proposed for managing waste in smart cities is iSmartWMS¹⁵. In this paper, the authors provide a comprehensive description of the architecture and components of an intelligent waste management system. They also detail the software tools, sensors, and techniques utilized in iSmartWMS. A proof-of-concept prototype of the iSmartWMS solution has been developed and tested to a limited extent. If implemented on a

large scale, this solution could effectively serve the purpose of waste management, given the involvement of a large number of stakeholders according to the established architecture. Supervision plays an essential role in waste management by addressing and resolving various related issues. The deployment of AI and digital technology in intelligent waste management has been suggested in¹⁶ and¹⁷. The former presents the necessary technologies for achieving intelligent waste management, while the latter describes how digital technology and AI were used to reform the process of municipal solid waste management in Russia. IoT is also identified as a key technology for waste management in smart cities in¹⁸. Lastly,¹⁹ discusses waste management planning and the actualization of Angkuts in Indonesia for designing waste-handling management systems.

The study presented in²⁰ proposes involving Internet service providers (ISPs) in interactions with clients and the city council to improve smart waste management (SWM) services. The concept aims to provide clients with a more dynamic and adaptable service while benefiting all parties. The study suggests employing flexible and individualized SLAs and presents a proof-of-concept model to optimize costs for customers engaging with the SWM scheme. In India, rapid industrialization and population growth have made municipal solid waste (MSW) a significant environmental concern. This paper²¹ proposes a smart waste collection method to optimize MSW processing. The Sa'diyah et al.²² proposed a study on waste management initiatives in Bogor City, where IT-based initiatives such as the Trash Bank and 3R Trash Program were implemented. However, the City of Bogor encountered barriers to public engagement and knowledge during its application. The municipal production waste (MPW) in urban residential areas is also a significant problem. Three studies propose solutions for smart waste management in smart cities. Dzhuguryan et al.²³ suggest a design for smart supply chain scenarios for the transportation of municipal production waste (MPW) from a city multi-floor manufacturing cluster using smart supply chain management (SSCM) technologies. Abuga et al.²⁴ propose a real-time smart garbage bin system strategically placed across smart cities using fuzzy logic, while Vishnu et al.²⁵ suggest an IoT-enabled solid waste management system for smart cities that tracks the location and unfilled level of bins using sensor nodes and sends this information to a central monitoring station via a graphical user interface (GUI) for garbage collection authorities. These solutions aim to overcome the drawbacks of traditional garbage collection and management systems and enhance waste management in smart cities. Meegan et al. suggested employing sensors to monitor the volume of trash within bins²⁶. The bin's GPS pinpoints its location, and this information is transmitted to a designated phone number via a GSM module. Zhang et al. suggested a strategy for the garbage detection system for small objects, using multi-branch dilated convolutions and deformable convolution networks in place of the feature extraction networks in the original Cascade RCNN²⁷. Gondal et al. presented a real-time smart garbage classification model that classifies waste using a hybrid technique based on pictures taken by a camera and pushed into the appropriately designated bucket using an automatic hand hammer²⁸. The proposed model's accuracy during training, testing, and validation was 0.99%. The proposed work uses cloud applications and long-range (LoRa) connections to monitor the bins in real-time using a customized sensor node and gateway node²⁹. The study³⁰ collected socioeconomic indicators from Tehran between 1991 and 2013 and identified significant factors using Pearson correlation analysis and three machine learning methods. Faisal et al.³¹ proposed a novel approach to waste management using the IoT and an advertising solution built on network-attached storage technology to generate income for waste management organizations.

Jacob et al.³² propose a smart prediction and monitoring IoT system for garbage disposal that uses commercially available components and a GPS module to monitor bins. LSTM is used to collect data and predict forthcoming waste. Linyuan et al.³³ propose an accurate MSW prediction model using DDMs and the ITD algorithm. The study highlights the lack of research on ITD in environmental systems. Agrawal et al.³⁴ propose a system that uses the MQTT protocol to transmit and receive data between devices and provides real-time updates on trash levels on a web-based monitoring tool. To ensure a clean environment, precise trash collection and disposal are essential. Humaun et al. presented an intelligent solution for garbage collection and disposal, which involves real-time monitoring of smart bins powered by solar energy³⁵. IoT technologies and their potential applications in infrastructure projects were discussed in a survey³⁶. The improper management and monitoring of e-waste may have negative effects on ecosystems and human health. Several papers propose IoT-based solutions for waste management. Liaqat et al.³⁷ suggest a system to measure garbage level and bin temperature in real-time, while Zaki et al.³⁸ focus on drainage overflow management. These solutions aim to improve waste collection, drainage management, and overall quality of life.

Various studies have proposed IoT-based waste management solutions in smart cities. A low-cost and effective smart compost bin was developed using environmental sensors and the LoRaWAN protocol³⁹. A review paper on existing IoT-based waste level management solutions in smart cities⁴⁰ showed the widespread use of Arduino Uno in earlier research. Pardini et al.⁴¹ proposed monitoring the bin's contents by sensors that provide real-time information on the compartments' degree of filling. Bano et al.⁴² suggested a smart bin mechanism based on AI that monitors trash cans in real-time and prevents overloading. An investigative case⁴³ carried out at Ton Duc Thang University in Vietnam showed the benefits of the suggested method, which is affordable, simple to use, and replaceable. Mishra et al.⁴⁴ developed a system to optimize solid waste collection in Bhubaneswar City by reducing the number of waste collection vehicles (WCVs) and their travel distance, which can lower capital and operational expenditures. The suggested route optimization technique showed a 30.28% reduction in overall WCVs' route distance, reducing OpEx by 29.07% and CapEx by 26.83% through fleet size reduction. Melakessou et al.⁴⁵ explored a case study on a corporation providing a commercial garbage collection service in Luxembourg and investigated how to use various waste data sources to develop meaningful indicators for collection processes.

Several researchers have proposed smart waste management systems based on the Internet of Things (IoT). Tariq et al.⁴⁶ proposed a system for monitoring waste bins and managing municipal solid waste in smart cities, including garbage collection, fire detection, and waste generation predictions. Srinivas et al.⁴⁷ created a 3D smart bin model that was tested at the IIIT-Kottayam IoT Cloud Research Lab in India. Sharma et al.⁴⁸ proposed an architecture for waste management in Indian smart cities, integrating intelligent bins with the IoT to improve

garbage collection and sorting. Kellow et al.⁴⁹ presented an IoT-based reference model and comparison analysis of waste management solutions currently on the market, with a focus on improving collection time and lowering costs while promoting citizenship. Chen et al.⁵⁰ suggested a smart trash system that detects not only the volume of trash but also its foul scent, allowing for effective waste collection and scheduling. Maksimovic⁵¹ discussed the advantages and potential for the adoption of IoT-powered waste management systems, which can address environmental and health issues caused by increasing waste volumes.

Smart cities face inefficiencies in garbage collection systems, which can be addressed through the use of IoT-based smart bins. The proposed prototype by Harith et al.⁵² provides frequent updates on bin status, enabling efficient route planning and decreasing collection times and expenses. Waste disposal and recycling are also addressed through reward-based strategies that incentivize recycling and door-to-door rubbish collection, as proposed by Pelonero et al.⁵³. To monitor waste levels, an IoT-based SC bin⁵⁴ embedded with sensors sends SMS alerts to municipal corporation members when the bin fills up with trash. Ultrasonic sensors can also be used to collect data on waste levels in each bin, which can be used to develop dynamic routes for garbage trucks, as proposed by Mohit et al.⁵⁵. In a survey, Vladyslav et al.⁵⁶ explore cutting-edge technologies to address conventional garbage management issues. Their solution has three phases, including regularly monitoring trash levels, utilizing all available data, and providing users with updated information through a web-based application. In another study, Harith et al.⁵² propose an IoT-based smart bin prototype as a solution to inefficiencies in garbage collection systems, aiming to achieve smart cities. Their system provides frequent bin status updates, enables accurate route planning, reduces collection times and expenses, and conserves gasoline. Several studies have proposed IoT-based solutions for solid waste management. A smart system for smart cities was proposed⁵⁷, which uses a NodeMCU microcontroller and sensor hardware system along with an Android application to track the data sent by the dustbin module. Abdullah et al.⁵⁸ recommended an improved waste management system that uses different truck sizes based on the kind of waste and integrates Internet of Things devices to allow communication between waste management centers, smart bins, waste source regions, and waste collection trucks. Bakhshi et al.⁵⁹ proposed an IoT-based waste monitoring system that makes use of Raspberry Pi, ultrasonic sensors, and machine learning analytics to monitor waste capacity and detect garbage collection scheduling. The results showed that fuel efficiency was increased by up to 46% and collection times reduced by up to 18% over a ten-day testing and validation period.

The present study responds to three persistent limitations in municipal solid-waste research and practice:

- Prevailing systems intervene only after overflow events, a reactive stance that proves inadequate as urban waste volumes accelerate;
- Many models ignore influential factors such as proximity to activity hubs, recent maintenance history, and weather dynamics, thereby restricting predictive power;
- Few investigations report end-to-end field deployments, leaving real-world viability largely untested.

To address these gaps, the work makes the following contributions:

- Introduces ProWaste, an IoT-enabled, machine-learning framework that fuses 15 automated and manual variables, predicts daily WCC criticality, and pushes mobile alerts to maintenance crews;
- Conducts a 25-model benchmark and wraps the selected Decision-Tree classifier with Binary PSO feature selection, yielding a compact three-feature model that exceeds 99.8% macro-F₁ while remaining interpretable;
- Validates the system on 6,954 daily records from 57 Bengaluru WCCs over four months, demonstrating reliable real-time operation and measurable reductions in unnecessary collections.

Recent work has advanced autonomous cleaning robots for coastal litter⁶⁰, vision-based solid-waste recognition that couples an attention-enhanced YOLO detector with DeepSORT tracking for real-time bin sorting⁶¹, and embedded “Internet-of-Video-Things” platforms that classify plastics on conveyor belts with quantised deep networks⁶². While these systems excel at on-site object detection or collection, they focus on single domains (beach sand, camera-visible items, or plastic streams) and require high-resolution imaging hardware, substantial computational power, or specialised robotics. ProWaste differs fundamentally by shifting the optimisation point upstream: instead of identifying individual waste items, it fuses fifteen low-cost heterogeneous signals (environmental, demographic, operational) to forecast the criticality of each urban Waste-Collection Centre in advance. An optimized classifier distils these inputs to three interpretable drivers, enabling municipality crews to pre-empt overspill with minimal sensing infrastructure and without GPU-class edge devices. Hence, ProWaste complements vision-centric sorting solutions by orchestrating city-scale routing and maintenance. It offers a scalable, explainable, and resource-efficient layer that can later ingest richer data from advanced sensors or camera systems as they become economically viable.

Proposed waste management system: ProWaste

The proposed proactive waste management system, named ProWaste, considers both automated and non-automated input parameters, with varying impact factors in the ML Module, and the respective criticality score for the WCCs is generated. As shown in Fig. 2, the ProWaste framework uses machine learning and IoT as its building blocks. The important components of the ProWaste conceptual framework, which gives an overall picture of the possible courses of action, are described below.

- *Physical sensors*: A low-cost sensor stack was selected. All modules contain no imaging or personal-data capabilities and operate passively; therefore, no additional permits or privacy approvals are required. During deployment, each node self-checks via a 30-s zeroing cycle at start-up and pushes hourly QC flags (range,

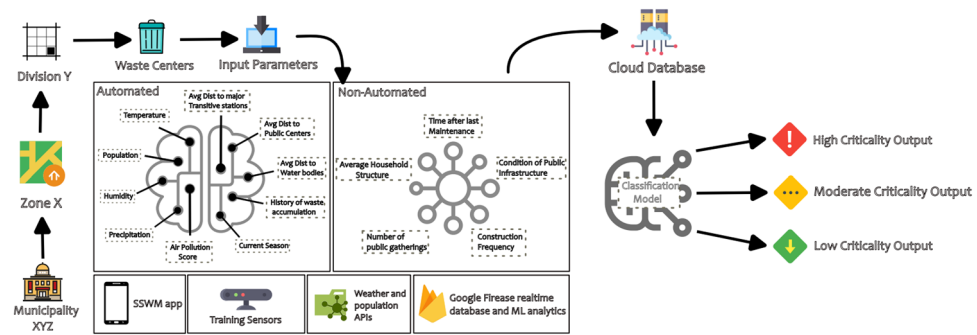


Fig. 2. ProWaste conceptual framework.

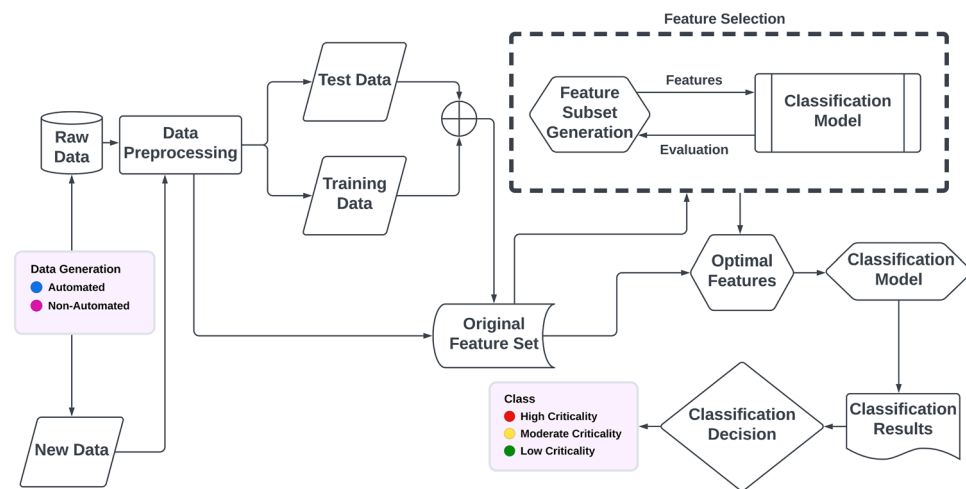


Fig. 3. Data classification process for criticality classes.

rate-of-change, and inter-sensor consistency) to the cloud; any flag failure triggers an app alert for field re-calibration.

- **APIs:** Multiple APIs are used for various functionalities, including but not limited to communication between the cloud, mobile application, and ML module, transfer of data between the three, receiving real-time data from weather applications, etc.
- **Google Firebase:** This database provided by Google Firebase is used for the storage of data such as municipal employee details, WCC details, etc.
- **Mobile application:** The Sustainable Smart Waste Management (SSWM) mobile application built on the Flutter framework is used as the key entity for allowing municipality users to view the data concerning WCCs and their current criticality score and input parameters, as well as to receive timely notification alerts for ensuring optimal waste management.

General model description

Figure 3 illustrates the classification process for new data into three categories: high, moderate, and low criticality. The process begins with raw data, which is subjected to a series of preprocessing steps to clean, normalize, and convert it into a model-compatible format. After preprocessing, the data is split into testing and training subsets. The training subset is used to train the model, whereas the testing subset is utilized to evaluate the model's performance. The preprocessed data is used to create an original feature set, consisting of all the features that can be used for classification. A wrapper method is then used to select the best features from the extracted set, generating a subset of features that are most pertinent to the classification task. The subset of features generated by the wrapper method is further refined to obtain the optimal features for the classification task. Finally, a classification algorithm is trained on the optimal set of features and used to classify new data points based on their feature values, outputting a decision for each data point and indicating the class to which it belongs.

Data description & generation

As described below, the input data are generated from fifteen different sources, including automated and non-automated sources:

Population

This input data pertains to the population count in the 10 km radius area surrounding the respective WCC. The data will be the categorization of the population count based on approximate estimates among the following three:

- High population (>1,00,000 people)
- Moderate population (50,000 – 1,00,000 people)
- Low population (<50,000 people)

The population data is taken from the API provided by <https://geoq.io/>, which is called geoIQ's data API. With this API, the population data of any selected area can be extracted. This data, however, is not real-time and is updated every 2–3 years. Since the scope of the research work is constrained to smart cities, where population variance is at its minimum in that timespan, the data is relevant. Using the API, the population count of the respective area surrounding the WCC for residential waste is taken and accordingly classified into one of the three categories above. This category is then sent to the ML model as the input feature. This feature is static and non-variable for up to 2–3 years. As and when the APIs update their databases with fresh data, so will the input feature and the refreshed category. This input feature holds a high significance for the output variable, as the population and the amount of waste produced are closely correlated. Waste production increases with an area's population.

Average distance to public centers

This continuous feature records the road-network mean distance (in km) between a WCC and all public-facing facilities located within a 10 km buffer, including educational institutes (schools, colleges), recreational venues (malls, play arenas, cinemas, city centres), large apartment complexes, permanent gathering spots (clubs, pubs), hotels and restaurants, and any registered community facility. Pairwise distances are retrieved via the Google Maps Distance API; the arithmetic mean is then computed

$$A = \frac{1}{n} \sum_{i=1}^n a_i, \quad (1)$$

where n is the number of detected facilities and a_i is the individual road distance in kilometres. The raw value (float) is stored under `Avg_Dist_Public_Centers_km` and fed directly to the model; its directional influence on waste accumulation is learned automatically during training.

Average distance to water bodies

Analogously, this feature measures the mean road distance (km) from the WCC to all perennial water bodies—rivers, lakes, ponds, and open drains—located within a 10 km radius. Distances are harvested with the Google Maps Distance API and averaged with Eq. (1). The resulting float is stored as `Avg_Dist_WaterBodies_km`. Whether shorter or longer proximity drives higher criticality is not hard-coded; the model infers the relationship from data.

Average distance to major transit stations

This continuous variable represents the road-network mean distance, in kilometres, from each WCC to every major transit hub contained within a 10 km buffer. Transit hubs comprise train stations, primary bus terminals, airports, harbours, taxi depots, and public parking facilities (surface lots and multi-level garages). Pairwise road distances are obtained through the Google Maps Distance API exactly as in Eq. (1); the arithmetic mean of the resulting list is stored as `Avg_Dist_Transit_km`. Two counteracting phenomena motivate the inclusion of this feature. First, it is identified that WCCs situated farther from transit hubs often accumulate more waste because they receive less casual surveillance and experience longer service intervals. Conversely, it is also expected that WCCs positioned very close to high-traffic hubs may accumulate excess waste owing to constant footfall and opportunistic disposal by commuters. By supplying the continuous distance value rather than a preset category, the learning algorithm can infer which effect dominates under local conditions and adjust predictions accordingly.

Temperature

This input feature is concerned with the temperature of the area where the respective waste center is located. The area is a 20 km radius circle centered on the waste center in question. Elevated temperatures speed up waste decomposition and gas production, leading to an increase in leachate⁶³.

Air-pollution score

For each WCC day-snapshot, the nearest OpenAQ monitoring station supplies the *daily mean Air-Quality Index (AQI)* over a 20 km radius. The integer value is recorded in the dataset column `Air_Pollution_AQI`.

- Low : $0 \leq \text{AQI} \leq 70$
- Moderate : $71 \leq \text{AQI} \leq 110$
- High : $\text{AQI} \geq 111$

The raw numeric AQI is streamed in real time to the ML pipeline; the above categories are employed solely for visual alerts in the SSWM mobile app.

Humidity

The Open-Meteo API provides the *daily mean relative humidity* (%) for the 20 km circle around each WCC. The integer value is stored as `Humidity_%` and consumed directly by the model.

- Low : 0–40 %
- Moderate : 41–70 %
- High : 71–100 %

As with AQI, these bands are used only for colour-coded indicators; the ML engine operates on the continuous humidity percentage, which is refreshed continuously via the live data stream.

Precipitation

This input feature captures the *daily total precipitation depth* (in mm) recorded within a 20 km radius of the waste-collection centre. Values are obtained from the same weather APIs used for *Temperature*. The levels are discretised as:

- Low (< 3.8 mm)
- Moderate (3.8–7.6 mm)
- High (> 7.6 mm)

Because rainfall can fluctuate sharply from one day to the next, the corresponding data stream is ingested in real time and the feature is refreshed continuously for the ML model.

Current season

This input feature defines the climatic conditions of the smart city (since the current season is not localized, but generalized to, at the very least, an entire state) and is categorized as one of the following:

- Normal—Spring, Autumn, and Winter
- Negative—Summer and Rainy

The data is gathered through the same weather API mentioned in the previous features. This data will define the seasons, such as summer, winter, etc. “Negative” denotes seasons that accelerate decomposition and odour generation. Based on these data, the category of the input feature will be selected and sent to the ML model. This data feature will vary according to the seasons, so the period would be 2–4 months for each data before they are updated by the API.

Time since the last maintenance date

This input feature represents the duration in days between the present time and the last maintenance date of the WCC. The data for this feature is obtained from the municipality’s database, which is updated regularly through a mobile application. The cloud calculates the feature, and the value is sent to the ML model. The longer the duration between maintenance, the higher the likelihood of waste accumulation and toxicity at the respective WCC. This feature is important as it serves as a criticality score reset parameter after maintenance is performed.

Average house structure

This input feature accounts for the different house types present in a smart city and is a subdivision of public centers, but can be taken as a separate factor on its own. It defines the average house type present in the 10 km radius, whether it’s dominated by apartments or single/double-story buildings, etc. The category of the area will fall within the following:

- High number of apartments (meaning dominated by multi-story apartments)
- Moderate number of apartments (average mixture of apartments and single homes)
- Low number of apartments (meaning dominated by single houses)

This data can also be acquired from the Google Maps API, though a thorough data cleaning process must be in place. The data taken here are from apartments within a 10 km radius. If there is a presence of more than 10 such apartments, then the area surrounding the waste center can be categorized as having a high number of apartments. If between 5 and 10, then it will fall under an average number of apartments, and anything lower is categorized as a low number of apartments. This selected category is sent as the input feature. This data feature is semi-variable; updating of the feature data can be done when apartment complex constructions are done and updated in Google Maps as operational. This input feature also holds moderate significance to the output variable. If apartment complexes dominate the area, then it indirectly shows more population concentration in that area, and hence more waste accumulation

Prior information as to the weekly aggregation of trash in a certain zone

Information about the weekly accumulation of waste in a particular area is an essential input for the model. This feature comprises data related to the average weekly accumulation of waste in the vicinity of the WCC. The data is classified into three categories based on the level of waste accumulation: high, moderate, and low. Data gathering for this feature is done through the output factors of the ML model. The output factor of the model is the criticality score of the WCC, which is stored in the database. This data is used as an input feature and represents the historical data of the WCC. The period of this data is semi-variable and is regularly updated, along

with the output factors being sent to the municipality. This input feature holds high significance for the output variable. The historical waste accumulation rate of an area covered by a WCC can be used to predict the future accumulation rate of that area, considering that certain factors remain constant.

Approximate public gatherings per month

The input feature of the approximate number of temporary public gatherings that occur in the 10km radius area surrounding the WCC per month is essential for the model. Data for this feature is collected through manual inputs from the municipality of the concerned city or area. At the end of every month, they input the data for this parameter through the mobile application. The data is obtained from their sources and is updated routinely every month.

Condition of public infrastructure

The input feature of the condition of public infrastructure in the 10km radius area surrounding the respective WCCs is essential for the model. This feature includes data on the average condition of the public infrastructure in the area. It is categorized into one of the following three categories: well-maintained, sufficiently maintained, or poorly maintained. The data for this feature is gathered through manual inputs from the municipality via the mobile application. The municipality can update the data as and when they see fit based on the observed conditions of the public infrastructure. Once the inputs are given in the form of the three categories, the average category is calculated and sent as an input feature.

Construction works or any other social works and their frequency

This categorical feature describes the average frequency of construction and other social works within a 10 km radius of each waste center, reflecting the intensity and occurrence of such activities in the area and classified into one of three levels: high frequency, moderate frequency, or low frequency.

The data is gathered once again by manual inputs from the municipality side through the mobile SSWM application, which is similar to the previous feature. This data has to be updated by the municipality at least once every 3 months, with multiple changes in the data being possible if such changes occur in the frequency of the construction works in the respective area. This feature has a periodic variability, at least as it must be updated every 3 months. If the increased frequency of construction and other such social works is observed, waste accumulation also proportionally increases. A complete description of all input variables and their update schedules is provided in Table 1.

Waste-Collection Centre locations were selected primarily for public convenience and logistical accessibility; as such, their spatial distribution may introduce measurement bias. Because the present study aims to demonstrate a scalable analytical framework rather than produce a definitive geostatistical survey, this trade-off was accepted.

The functional flow layout diagram of the ProWaste model is displayed in Fig. 4. In block (A), the automated input parameters are taken for a certain waste center using various data-gathering techniques. This data is then sent two ways, one to the cloud database for storage and the other to the SSWM app for monitoring. The parameters are sent with the help of the respective APIs and updated regularly whenever there is a change in the data gathered. In block (B), the input data values from the parameters are routinely checked for any disparity or observable outliers. These may be caused by certain errors in the data-gathering technique, such as API server downtime, data corruption, etc. If any such changes to the parameter values are required, they are made by municipal software personnel and updated through the SSWM app to the cloud database. In block (C), the input parameters from the cloud database are continuously updated with the most recent data in all scenarios. The ML Module accesses these input parameter values to generate a criticality score as the output variable. This criticality score is updated in the database. In block (D), once the criticality score is updated, it is sent to the decision-making algorithm that decides the priority of maintenance required based on the said criticality score. A higher criticality score puts the respective waste center in its corresponding position in the maintenance priority queue. The SSWM app is used to update and notify the relevant municipality. In block (E), the municipality, on receiving the notification, can handle maintenance at their discretion, based on the priority queue displayed in the SSWM app, after which post-maintenance values are updated to both the database and the app. In block (F), the non-automated input parameters are observed by municipality personnel and manually entered into the database through the SSWM app. These parameters may or may not be periodic, based on conditions observed by the municipality. Whenever there is a significant change in these parameters, they are updated by the municipal software personnel and uploaded through the SSWM app to the cloud database.

In Fig. 5, (A) shows the registration and logging in of municipal personnel with their respective IDs. There will be two types of users with different access levels in the application: software personnel and maintenance personnel. The software personnel users will have access to all input parameter details for every waste center, both automated and manual. They will be able to report any errors or disparities observed in the automated input parameters. Along with this, it is their responsibility from the side of the municipality to manually enter the non-automated input parameters periodically or whenever the application notifies them of such. The criticality score of each waste center in real-time can also be viewed. Maintenance personnel users will have access to the waste center locations and an overview of their details, along with their current criticality score and maintenance status. They will also have a “Priority Queue Alerts” feature, which notifies them and their respective teams of new waste centers that have gone critical (the criticality score is high) and need maintenance. They can also keep track of the maintenance of different waste centers of different criticalities, which are arranged in a priority queue for optimal waste management.

In (B), the dashboard acts as a quick notification to alert the current top waste centers that need maintenance (in the case of maintenance personnel) or the current input parameters that need to be updated for respective waste centers (in the case of software personnel). It can be used to provide quick access to maintenance updates

Variable	Role	Type / Unit	Description	Collection source	Update freq.
Date	–	–	Calendar day of snapshot (00:00–23:59).	Edge-gateway system clock	Daily
WCC_ID	–	–	Unique Waste-Collection-Centre identifier (wcc001–057)		–
Population_Category	X-var	Cat. {High, Moderate, Low}	Residential density within 10 km radius of the WCC.	GeoIQ population raster (100 m)	2–3 yr
Avg_Dist_Public_Centers_km	X-var	Float (km)	Road-network mean distance to schools, malls, hotels, clubs (10 km buffer).	Google Maps Distance API	Monthly
Avg_Dist_Transit_km	X-var	Float (km)	Mean distance to major transit hubs (rail/bus depots, airport, parking).	Google Maps Distance API	Monthly
Avg_Dist_WaterBodies_km	X-var	Float (km)	Mean distance to perennial water bodies (rivers, lakes, open drains) within 10 km (road-network distance).	Google Maps Distance API	Monthly
Temperature_C	X-var	Float (°C)	Daily mean ambient temperature at the WCC centroid.	Open-Meteo API (ERA5 down-scale)	Daily
Air_Pollution_AQI	X-var	Integer (AQI)	Daily mean Air-Quality Index (CPCB scale).	OpenAQ API (nearest station)	Daily
Humidity_%	X-var	Integer (%)	Daily mean relative humidity.	Open-Meteo API	Daily
Precipitation_mm	X-var	Float (mm)	Daily total rainfall proxy.	Open-Meteo API	Daily
Current_Season	X-var	Cat. {Normal, Negative}	Macro-climatic regime: <i>Normal</i> = Spring/Autumn/ Winter, <i>Negative</i> = Summer/Rainy.	Open-Meteo seasonal calendar	2–4 months
TimeSinceLastMaintenance_days	X-var	Integer (days)	Days elapsed since the last logged maintenance visit.	SSWM mobile-app event log	Daily (auto)
Avg_House_Structure	X-var	Cat. {HighAp., ModerateAp., LowAp.}	Dominant housing pattern (apartments vs. single homes) in 10 km buffer.	Google Places POI count	Annual
Weekly_Waste_Accumulation	X-var	Cat. {High, Moderate, Low}	Seven-day moving-window waste build-up category.	Derived from historic criticality logs	Weekly
Infrastructure_Condition	X-var	Cat. {WellMaintained, SufficientlyMaintained, PoorlyMaintained}	Subjective rating of roads, drains, and lighting around WCC.	Manual entry by municipal engineer	Monthly
Approx_Public_Gatherings	X-var	Integer (events)	Number of temporary public events (festivals, fairs, rallies) within 10 km during the preceding month.	Manual entry by municipal culture office	Monthly
Construction_Frequency	X-var	Cat. {High, Moderate, Low}	Intensity of construction or other sanctioned social works within 10 km, derived from active permits and field audit.	Municipal permit database + engineer survey	Quarterly
CriticalityClass	Target	Cat. {Low, Moderate, High}	Priority label used for route scheduling.	Classifier	Daily

Table 1. Variable Description (6 954 daily snapshots; 57 WCCs; 1 March–30 June 2024).

or parameter update functionality. In (C), the zone list shows all the different zones under the municipal scope; this can be viewed in either tabular format or map format. All details are stored in the cloud and delivered in real-time to the mobile interface. In (D), the waste center list shows all the different waste centers under each zone, which can also be viewed in tabular format or map format. All details are stored in the cloud and received in real-time by the mobile interface. In (E), the waste center details show the user all data concerning the selected waste center, including all input parameters, the current criticality score, and the waste center’s location. There are also options to report data errors and edit non-automated input parameters. The option for reporting maintenance is available to maintenance personnel, who can tag the waste center as having undergone maintenance or still in the process of maintenance. The reason for this is to have optimal waste management among many maintenance teams to avoid overlap and increase efficiency. In (F), priority alerts are sent to users as and when new waste centers need to undergo maintenance. Their position in the priority queue is based on their criticality score relative to the other waste centers. The algorithm that determines the queue considers the criticality score as well as other factors such as the closest distance to maintenance headquarters, the proximity of waste centers to one another, and so on. This feature also indicates which centers are undergoing maintenance currently and which centers are still pending maintenance. When the centers are done with maintenance and are reported as such by the maintenance team, they are removed from the queue.

The map view displayed in Fig. 6 is an illustration of the User Layer Interface that the SSWM application provides to municipality personnel. It presents a detailed pictorial overview of all WCCs within a certain zone, along with their real-time criticality score. The location markers indicate nine urban WCCs, and the surrounding circular color indicates the criticality of each WCC. Overlapping criticality areas indicate higher criticality, and locations that fall within such areas correspond to a higher criticality score. Similar to the Google Map interface, users can toggle different surrounding structures, such as the nearest public centers, weather indicators, etc., to provide more detailed clarity to the municipality’s users and personnel. Figure 6 is for illustration only and does not represent any actual location or municipality; any such correlation is purely incidental.

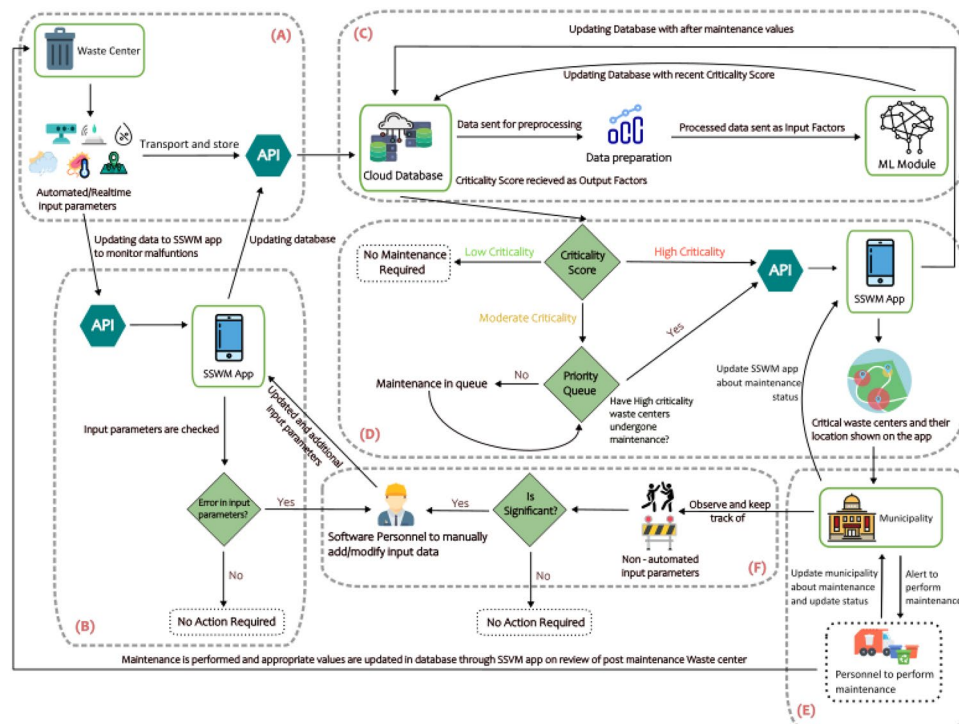


Fig. 4. Functional flow block diagram of ProWaste model.

Feature selection

The feature extraction method involves the two crucial processes of feature construction and feature selection. The term “feature construction” (also known as “data preprocessing”)⁶⁵ encompasses any techniques that modify the original feature in some way, such as data standardization, normalization, and noise filtering. Making it simpler to recognize the underlying information in data is the goal of this key preprocessing phase. By eliminating the characteristics with the lowest potential for information, feature selection merely looks for the ideal feature subset. The benefit of feature selection is that it allows us to execute dimensionality reduction on our data by reducing the number of features. This can protect from the curse of dimensionality and speed up the classification process. Although there are many different feature selection methods available, there are three things that essentially tell them apart⁶⁵: the search strategy for feature subset creation, the prediction performance, and the assessment method. The first refers to the search method that was used to analyze the potential solutions in the space of feature combinations. The remaining two relate to the assessment criteria, or the procedure and metrics employed to rate the excellence of each feature subset. The feature selection algorithms are categorized into three classes: wrapper methods, embedding methods, and filter techniques, based on the subset assessment approach⁶⁶. Wrappers investigate the range of prospective features utilizing a search algorithm, assessing every subset by employing a model. This approach risks over-fitting the model and can be computationally expensive. In the search plan, filters function similarly to wrappers. Still, the features are chosen by analyzing a performance metric that does not necessitate developing a model, as opposed to comparing against a model. Similar to wrapper approaches, embedded feature selection methods relate feature selection to the categorization stage. Better search space coverage is the key benefit of the wrapper approach over embedded methods. The wrapper method’s key benefit over filter techniques is that the final subset’s prediction performance is linked to the relevant measure of choice, in this instance, the classifier. The wrapper technique is an attractive option when obtaining the most accurate model, which is the main goal, and speed is not a concern.

Wrapper methods

Using the predictor itself during subset evaluation is the defining feature of wrapper approaches. Candidate subsets are scored by a learning algorithm according to their predictive performance⁶⁵. Although wrappers often yield superior accuracy for the chosen base learner, they require training a model for *every* feature mask examined; hence, an explicit search strategy is essential because enumerating all combinations is infeasible. Popular heuristics include greedy hill climbing, which iteratively modifies a current subset and keeps a change only if it improves the score. Since exhaustive search is impossible, the best subset observed so far is returned when a stopping criterion is met—for example, no further improvement, a target score, or a maximum run-time.

Model selection for wrapper feature selection

Each candidate mask is evaluated by fitting a *base learner* and measuring its validation performance. After benchmarking 25 classifiers (Table 2). We selected a **DecisionTreeClassifier** because it

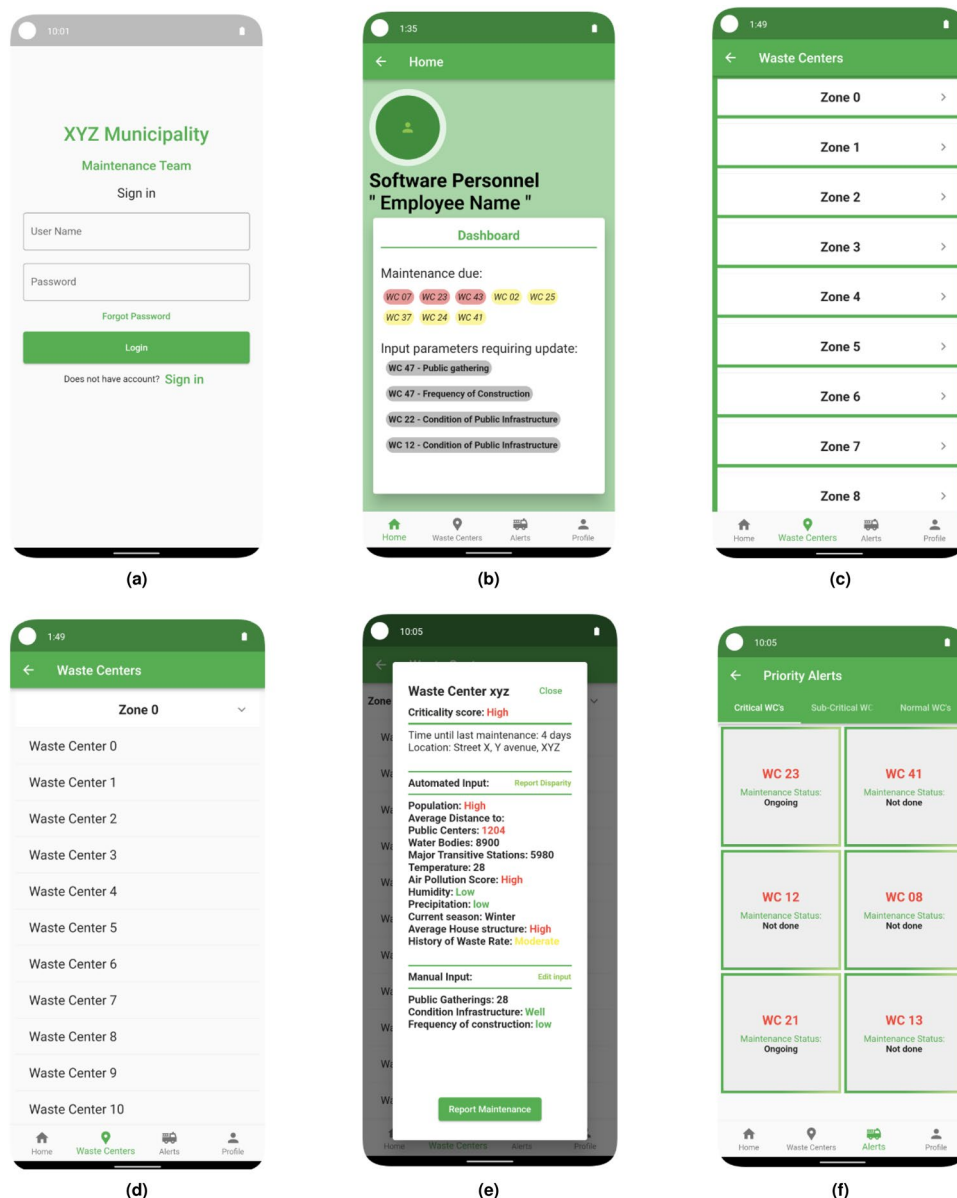


Fig. 5. Interface of municipal waste management application with user access levels and priority queue alerts.

- offers high single-model accuracy, yet trains orders of magnitude faster than large ensembles—critical when thousands of masks are tested;
- produces transparent if-then rules, allowing municipal engineers to audit why a WCC is flagged *High*;
- accepts both numerical and one-hot categorical inputs without additional engineering inside the swarm loop.

Objective function

Let the binary vector $h_x \in \{0, 1\}^{N_T}$ activate $N_F = \|h_x\|_1$ of the N_T encoded predictors. To balance predictive power and parsimony, we minimise

$$f(h_x) = \alpha(1 - \text{Acc}) + (1 - \alpha) \frac{N_F}{N_T}, \quad \alpha = 0.99, \quad (2)$$

where *Acc* is the *multi-class accuracy* of the decision-tree based learner on the validation fold, and N_F/N_T is the fraction of features retained. With $\alpha = 0.99$ the search places 99% weight on reducing classification error and 1% on shrinking the subset.

Given a $K \times K$ confusion matrix C whose element C_{ij} counts instances of true class i predicted as j , multi-class accuracy is

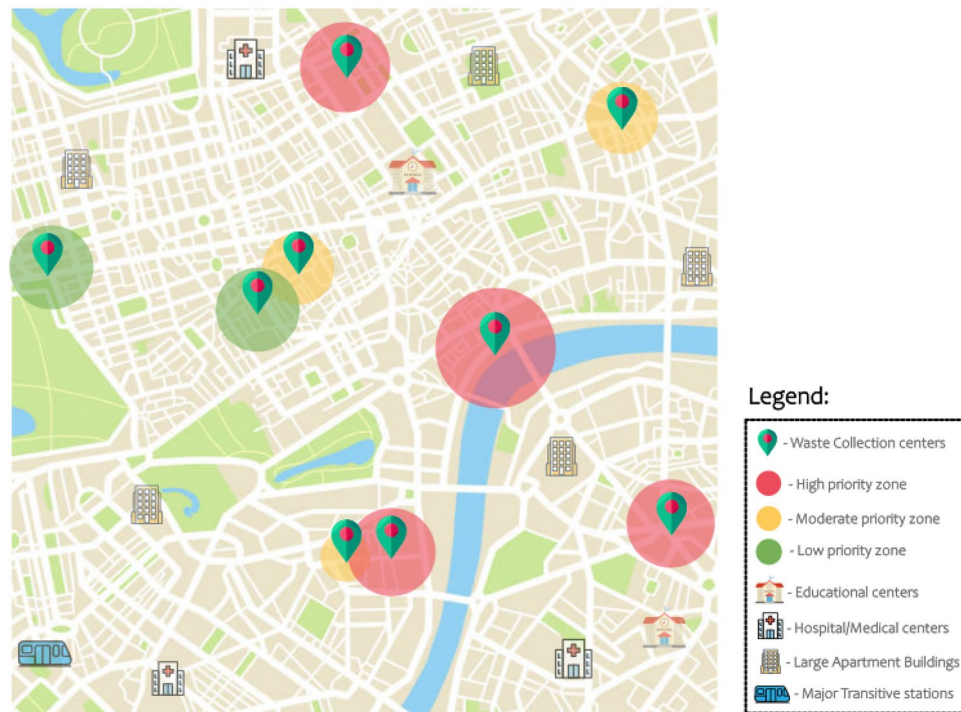


Fig. 6. Map view of SSWM Application's Waste Center Criticality Scores for Municipalities ⁶⁴.

$$Acc = \frac{\sum_{k=1}^K C_{kk}}{\sum_{i=1}^K \sum_{j=1}^K C_{ij}}, \quad (3)$$

i.e. the proportion of correctly classified samples. This generalises the binary definition and avoids class-specific terms such as true/false positives.

Because enumerating the 2^{N_T} possible masks is computationally prohibitive, we optimise $f(h_x)$ with Binary Particle Swarm Optimisation (BPSO).

Application of BPSO in feature selection

BPSO represents each particle as a binary mask of length N_T ; a '1' keeps the corresponding column, a '0' drops it. Positions and velocities are initialised at random, and each iteration is updated via cognitive and social components. Our settings—30 particles, 100 iterations, 20 independent runs—balance exploration with run-time (see Algorithm 1).

Encoding

Let $m_{xy} \in \{0, 1\}$ be the position of particle x on dimension y and $v_{xy} \in \mathbb{R}$ its velocity. Initialisation uses

$$m_{xy} \leftarrow \begin{cases} 1 & \text{if } r \geq 0.5, \\ 0 & \text{otherwise,} \end{cases} \quad r \sim \mathcal{U}(0, 1), \quad (4)$$

$$v_{xy} \leftarrow -v_{\max} + 2k v_{\max}, \quad k \sim \mathcal{U}(0, 1). \quad (5)$$

Search procedure

All features are normalised to zero mean and unit variance before the swarm search begins. Thirty particles, each testing on average $\approx 15\%$ of the N_T columns, and 100 iterations per run were sufficient to reach stable minima in every one of the 20 runs.

Model	Accuracy	BalancedAcc.	F1_macro	BestParams	Acc_Rank	BalAcc_Rank	F1_Rank	AvgRank	References
AdaBoostClassifier	0.998897	0.998086	0.998587	{'n_estimators': 50, 'learning_rate': 1.0}	1	1	1	1.00	67
BaggingClassifier	0.998466	0.997954	0.998039	{'n_estimators': 10, 'max_features': 1.0}	2	2	2	2.00	68
LGBMClassifier	0.998370	0.997873	0.997917	{'n_estimators': 100, 'learning_rate': 0.1, 'num_leaves': 31}	3	3	3	3.00	69
DecisionTreeClassifier	0.998226	0.997839	0.997733	{'criterion': 'gini', 'max_depth': None, 'min_samples_split': 2}	4	4	4	4.00	70
RandomForestClassifier	0.997651	0.996667	0.996993	{'n_estimators': 200, 'max_features': 'sqrt'}	5	5	5	5.00	71
ExtraTreesClassifier	0.985236	0.976512	0.981340	{'n_estimators': 100, 'max_features': 'sqrt'}	6	6	6	6.00	72
SVC	0.968266	0.957637	0.960816	{'C': 1.0, 'kernel': 'rbf'}	7	7	7	7.00	73
NuSVC	0.964144	0.945881	0.954204	{'nu': 0.5, 'kernel': 'rbf'}	8	8	8	8.00	74
LinearDiscriminantAnalysis	0.949957	0.941619	0.938551	{'solver': 'svd'}	9	9	9	9.00	75
CalibratedClassifierCV	0.944205	0.919877	0.931539	{'method': 'sigmoid'}	10	10	10	10.00	76
LogisticRegression	0.932988	0.904546	0.914007	{'C': 1.0, 'penalty': 'l2'}	11	12	11	11.33	77
GaussianNB	0.922587	0.908919	0.903256	{'var_smoothing': 1e-09}	12	11	12	11.67	78
LinearSVC	0.895697	0.855221	0.866455	{'C': 1.0}	13	13	13	13.00	79
KNeighborsClassifier	0.865543	0.829040	0.841961	{'n_neighbors': 11, 'weights': 'distance'}	14	15	14	14.33	80
NearestCentroid	0.862429	0.849923	0.834539	{'metric': 'euclidean'}	15	14	15	14.67	81
ExtraTreeClassifier	0.842161	0.825114	0.824790	{'criterion': 'gini', 'max_depth': None, 'min_samples_split': 2}	16	16	16	16.00	82
LabelSpreading	0.813489	0.794286	0.798710	{'kernel': 'rbf', 'gamma': 20, 'n_neighbors': 7}	17	17	17	17.00	83
LabelPropagation	0.813345	0.794040	0.798565	{'kernel': 'rbf', 'gamma': 20, 'n_neighbors': 7}	18	18	18	18.00	84
Perceptron	0.795134	0.755192	0.745572	{'alpha': 0.0001, 'max_iter': 200}	19	19	19	19.00	85
PassiveAggressiveClassifier	0.771931	0.731353	0.724623	{'C': 1.0, 'loss': 'hinge', 'max_iter': 200}	20	20	20	20.00	86
RidgeClassifier	0.747341	0.652367	0.643795	{'alpha': 1.0, 'tol': 0.0001}	21	21	21	21.00	87
Continued									

Model	Accuracy	BalancedAcc.	F1_macro	BestParams	Acc_Rank	BalAcc_Rank	F1_Rank	AvgRank	References
RidgeClassifierCV	0.747293	0.652155	0.643434	{'alphas': [0.1, 1.0, 10.0], 'store_cv_values': 'deprecated'}	22	22	22	22.00	88
BernoulliNB	0.702760	0.625630	0.617116	{'alpha': 1.0, 'binarize': 0.0, 'fit_prior': True}	23	23	23	23.00	89
QuadraticDiscriminantAnalysis	0.416018	0.421123	0.365857	{'reg_param': 0.0, 'tol': 0.0001}	24	24	24	24.00	90
DummyClassifier	0.408398	0.333333	0.193316	{'strategy': 'most_frequent'}	25	25	25	25.00	91

Table 2. Performance-based ranking of the classification models.

```

1: Input: training set  $(\mathbf{X}, \mathbf{y})$ ; particle count  $N$ ; feature dimension  $N_T$ ; inertia  $w$ ; cognitive/social coefficients  $c_1, c_2$ ;  $\alpha = 0.99$ .
2: for run = 1 to 20 do
3:   Randomly initialise  $\mathbf{m}_x, \mathbf{v}_x$  ( $x = 1:N$ )
4:   Evaluate each mask; set personal bests  $m_x^{PB}$ ; set global best  $m^{SB}$ 
5:   for iteration = 1 to 100 do
6:     for each particle  $x$  do
7:        $\mathbf{v}_{xy} \leftarrow w\mathbf{v}_{xy} + c_1u(\mathbf{m}_{xy}^{PB} - \mathbf{m}_{xy}) + c_2r(\mathbf{m}_y^{SB} - \mathbf{m}_{xy})$ 
8:        $\mathbf{m}_{xy} \leftarrow \begin{cases} 1 & \text{if rand} < \sigma(\mathbf{v}_{xy}), \\ 0 & \text{otherwise} \end{cases}, \sigma(z) = [1 + \exp(-z)]^{-1}$ 
9:       Compute  $f(\mathbf{m}_x)$  via Eq. 2
10:      Update  $m_x^{PB}$  and  $m^{SB}$  if fitness improved
11:    end for
12:    Record best fitness of this iteration
13:  end for
14:  Store run-level record: best mask, fitness, test metrics
15: end for
16: Output: mask achieving the lowest validation fitness (decisive run); corresponding test metrics and selected feature names.
```

Algorithm 1. Binary PSO wrapping a Decision-Tree base learner

Performance evaluation

As shown in Table 2, the twenty-five models were ranked based on performance metrics, namely accuracy, balanced accuracy, and F1-score. The twenty-five models were chosen to give balanced, practical coverage of tabular-data classifiers while keeping the 10×3 repeated-fold benchmark feasible on a municipal workstation. They span all principal paradigms—linear (Logistic, Ridge, Perceptron), distance-based (k-NN, Nearest Centroid), probabilistic (Gaussian NB, Bernoulli NB, LDA, QDA), kernel (SVC, NuSVC, Linear SVC), tree and ensemble methods (Decision-, Random-, Extra-Trees, Bagging, AdaBoost, LightGBM), semi-supervised graph learners (LabelPropagation, LabelSpreading), and online margin algorithms (Passive-Aggressive). Every candidate natively supports multi-class targets and works with the scaled + one-hot matrix produced by the ColumnTransformer, needs only moderate CPU time (so deep nets and very large GBDTs were excluded), and offers built-in or SHAP-compatible explainability—an adoption requirement for municipal engineers. Key settings for the repeated-stratified K-fold benchmark include loading the ProWaste dataset, applying 10-fold splits repeated three times with a fixed random seed of 42, and preprocessing numeric features via StandardScaler and categorical features via OneHotEncoder(handle_unknown='ignore') within a ColumnTransformer inside a Pipeline. Twenty-five classifiers are evaluated in parallel using accuracy, balanced accuracy, and macro-F1 metrics, and their mean metric ranks determine the final ordering.

In the experiment, we applied a 10×3 repeated-stratified cross-validation (i.e., a 10-fold stratified CV repeated three times, for a total of 30 validation folds) to obtain point estimates, 95 % confidence intervals (CIs), a Friedman omnibus test, and Wilcoxon signed-rank post-hoc comparisons with Holm correction (Fig. 7), thereby providing three complementary layers of statistical evidence.

The AdaBoostClassifier tops the table with a mean accuracy of 0.998897; Balanced Accuracy and macro-F1 display equally low variance, whereas mid-tier models such as LinearSVC exhibit markedly broader intervals (accuracy 0.895697), signalling weaker generalisation. The Friedman p -value ($\approx 7 \times 10^{-134}$) rejects the null hypothesis of equal performance, justifying pair-wise Wilcoxon tests: only BaggingClassifier,

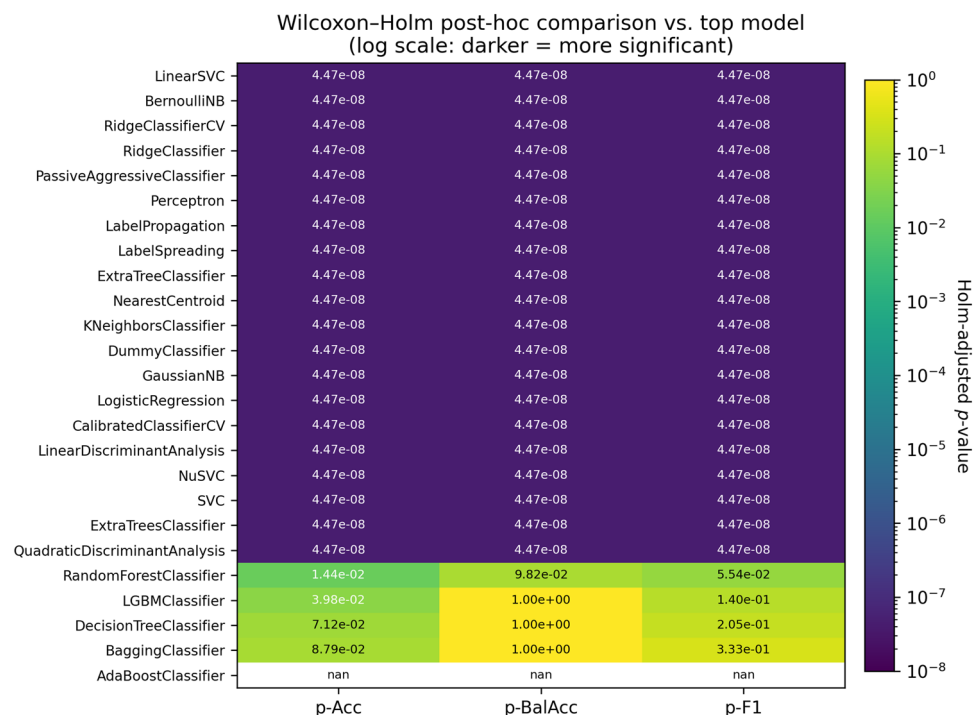


Fig. 7. Heat-map of Holm-adjusted Wilcoxon p -values for Accuracy, Balanced Accuracy, and macro-F1, comparing each model to the top performer.

LGBMClassifier, and DecisionTreeClassifier return adjusted p -values > 0.05 against AdaBoostClassifier, so their differences are not statistically significant—a pattern mirrored across all three metrics. Rank aggregates corroborate these findings, cleanly separating the leading trio from the remaining twenty-two learners. Because AdaBoostClassifier, Bagging, and LGBM offer no statistically reliable gain over a single tree, we favour the DecisionTreeClassifier for the wrapper-based feature-selection and XAI pipeline: it is natively interpretable (explicit if-then rules auditable by municipal engineers), orders of magnitude faster to retrain during the thousands of BPSO evaluations, supplies deterministic feature-importance vectors that integrate seamlessly with SHAP, and—given its CI width of ≈ 0.002 —shows no greater over-fitting risk than the more complex ensembles. Hence, the DecisionTreeClassifier combines statistically indistinguishable predictive power with maximal transparency and computational efficiency, making it the most suitable base learner for ProWaste. Based on the ranking, the DecisionTreeClassifier is selected as the base model for the wrapper feature selection process because it offers clear, actionable explanations through its transparent, rule-based structure. Its single-tree architecture provides direct feature importance measures and unambiguous decision paths, which can be further analyzed at both global and local levels using SHAP values, making it an ideal choice for explainable AI applications. The BPSO algorithm identified three optimal features that yielded the best results for the classification process. The experimental results of the wrapper feature selection process are displayed in Fig. 8. Figure 8 offers an integrated performance snapshot of the feature-optimised Decision-Tree model generated by twenty 100-iteration BPSO runs. Sub-plot (a) displays the confusion matrix for the hold-out test set (1,391 instances), where the diagonal dominance and a single off-diagonal entry confirm that the decisive model misclassifies only one record. Sub-plot (b) traces the mean best-fitness trajectory averaged over the twenty runs; the curve drops sharply in the first 15–20 iterations and stabilises around iteration 35, illustrating the swarm's rapid convergence toward an optimal three-feature subset. Sub-plot (c) summarises the resulting predictive quality: overall Accuracy, Balanced Accuracy, and macro-F1 all exceed 0.999, proving that drastic dimensionality reduction does not degrade performance. Finally, sub-plot (d) lists the retained predictors—*Weekly_Waste_Accumulation*, *TimeSinceLastMaintenance_days*, and *Air_Pollution_AQI*—emphasising that these three variables alone are sufficient to discriminate Low, Moderate, and High criticality with near-perfect precision.

Figure 9 elucidates how the decisive BPSO-Decision-Tree model arrives at its predictions by means of SHAP analysis. Sub-plot (a) reports mean absolute SHAP values per class, showing that the three variables retained by BPSO dominate the explanation for Low, Moderate, and High criticality alike. Sub-plot (b) ranks the global importance of every input: *TimeSinceLastMaintenance_days* contributes approximately an order of magnitude more than any other feature, followed by *Air_Pollution_AQI* and *Weekly_Waste_Accumulation*, whereas the remaining predictors are essentially inert. Sub-plot (c) presents the SHAP summary (beeswarm) plot; points to the right of the zero line increase the probability of the High class, while those to the left favour the Low class, again underlining the leverage exerted by the three selected variables. Subplots (d)–(f) zoom in on a single test instance (first sample) whose raw values are *Population_Category* = High, *Avg_Dist_Public_Centers_km* = 1.15,

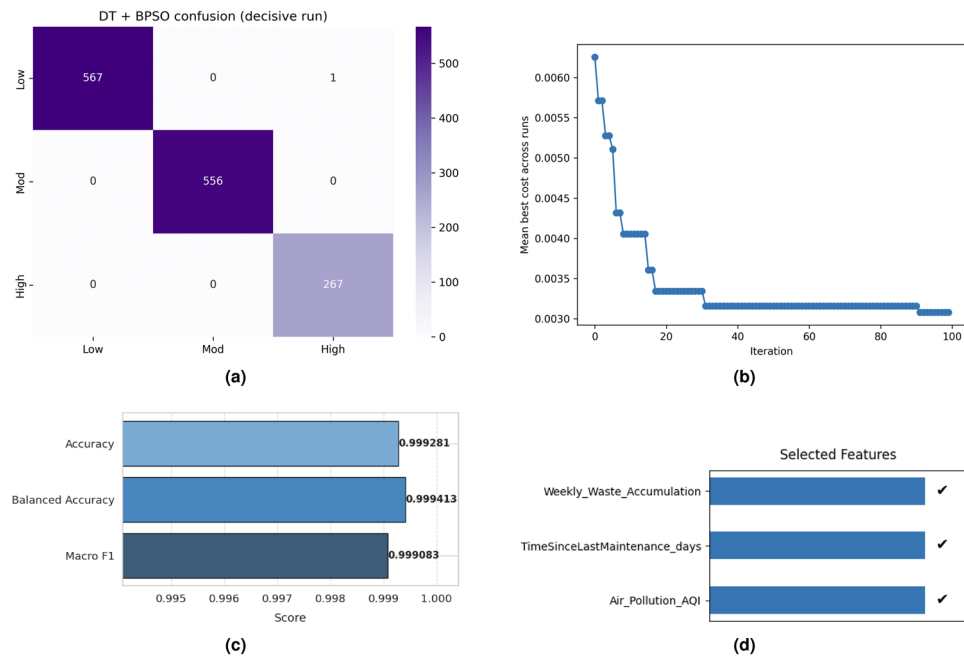


Fig. 8. Performance of the decisive BPSO-Decision-Tree model. **(a)** Confusion matrix on the hold-out test set. **(b)** Mean best-fitness trajectory (average of the twenty 100-iteration runs). **(c)** Bar chart of Accuracy, Balanced Accuracy, and macro-F1 obtained with the decisive subset. **(d)** Final feature subset selected by BPSO.

Avg_Dist_Transit_km = 2.02, *Temperature_C* = 28.7, *Air_Pollution_AQI* = 67, *Humidity_%* = 55, *Precipitation_mm* = 0.07, *TimeSinceLastMaintenance_days* = 3, *Avg_House_Structure* = *HighApartments*, *Weekly_Waste_Accumulation* = *High*, *Infrastructure_Condition* = *PoorlyMaintained* (true label: *Low*). Sub-plot (d) decomposes the log-odds for the *Low* class: the recent maintenance interval (3 days) and moderate AQI drive the prediction strongly toward *Low*. Sub-plot (e) performs the same decomposition for the *High* class, illustrating how the identical features push the log-odds upward yet remain insufficient to overturn the *Low* decision. Sub-plot (f) shows the middle ground for the *Moderate* class. Taken together, the experiments demonstrate that the BPSO-selected subset not only preserves classification accuracy but also yields an interpretable mechanism in which just three variables govern both global behaviour and individual-case reasoning.

Conclusion

This study introduced ProWaste, an end-to-end IoT and machine-learning platform that streams fifteen heterogeneous indicators—from population density and weather to maintenance history and weekly waste build-up—from 57 Bengaluru WCCs over 6,954 daily snapshots into a cloud database. Twenty-five off-the-shelf classifiers were benchmarked under repeated stratified cross-validation, and a DecisionTreeClassifier wrapped by Binary Particle Swarm Optimization was selected, removing 80% of inputs while preserving near-perfect performance (>99% accuracy on a hold-out test set). SHAP analyses confirm that the three retained features (Weekly Waste Accumulation, Time-Since-Last-Maintenance, Air Pollution AQI) remain highly interpretable for municipal engineers, and a Flutter-based mobile app completes the data-to-alert pipeline for field teams. Compared with current reactive scheduling, ProWaste can virtually eliminate missed pickups and cut unnecessary inspections. Future work will extend this pilot to multi-city and seasonal deployments to evaluate transferability, integrate more advanced licensed sensors (e.g., infrared gas analyzers, ultrasonic depth meters, and leachate conductivity probes) for higher-fidelity data capture, broaden coverage to additional waste streams (recycling, compost, e-waste) and multi-label criticality, and deploy adaptive learning and drift-detection mechanisms to self-calibrate as input distributions shift. We also plan to link ProWaste with vehicle-routing solvers, life-cycle cost analyses, and blockchain-based audit trails, transforming it from a high-accuracy predictor into a comprehensive decision-support platform for sustainable smart-city waste management.

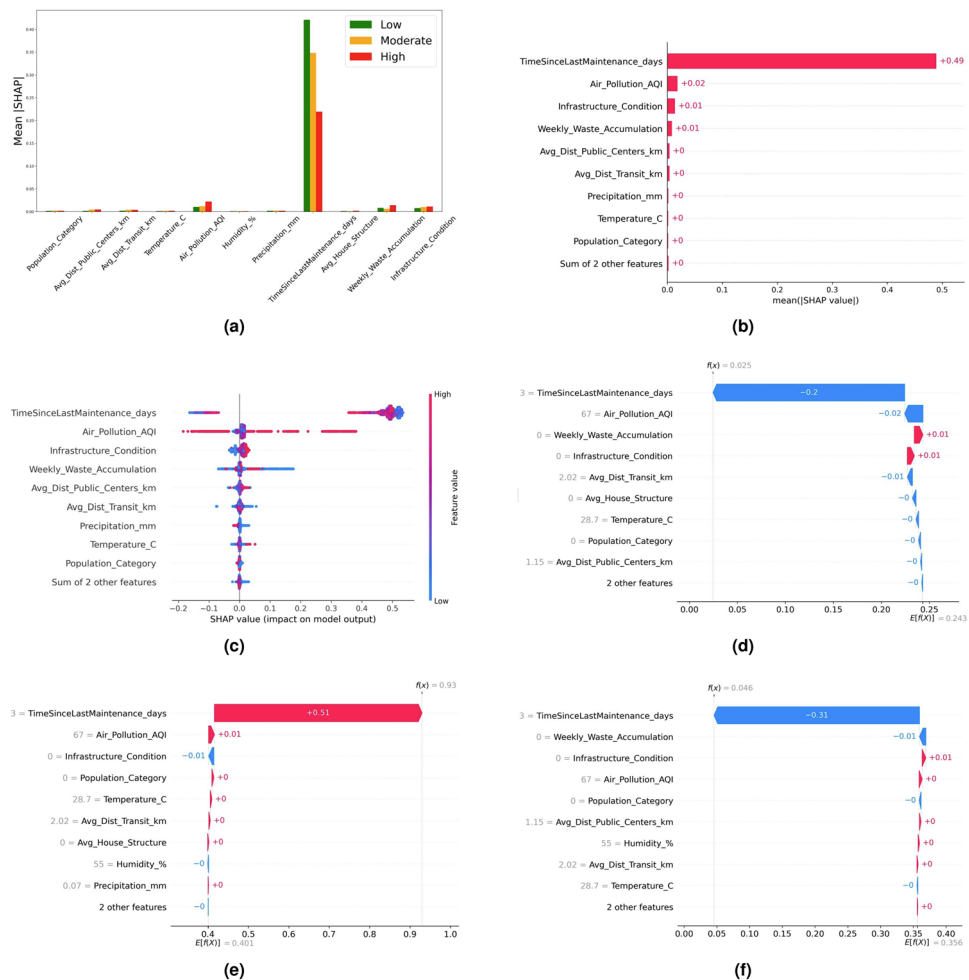


Fig. 9. SHAP-based feature-importance analysis of the decisive model. **(a)** Mean absolute SHAP values per class, confirming dominance of the three selected features. **(b)** Global mean SHAP bar chart: ‘TimeSinceLastMaintenance_days’ is by far the most influential predictor. **(c)** SHAP summary-dot plot showing the distribution of impacts; positive values shift predictions toward the *High* class. **(d)–(f)** Waterfall plots, illustrating how the key features increase or decrease the predicted criticality score of a particular instance (first sample).

Data availability

The datasets analyzed during the current study are available from the corresponding author upon reasonable request.

Received: 20 June 2024; Accepted: 20 June 2025

Published online: 30 July 2025

References

- Wire, B. Smart waste management systems global market report 2022: Growing volume of e-waste driving adoption worldwide—<https://www.researchandmarkets.com/> (2022).
- Silva, B. N., Khan, M. & Han, K. Towards sustainable smart cities: A review of trends, architectures, components, and open challenges in smart cities. *Sustain. Cities Soc.* **38**, 697–713 (2018).
- Future Market Insights. Smart waste management market growth—trends & forecast 2025 to 2035 (2025). Accessed Jan 20, 2025.
- Paul, A. Corpbiz has emerged as the most trusted consulting firm for waste management licensing in India (2022).
- Ming, N. Y. Exclusive: How India’s smartest city is tackling waste (2020).
- Cheela, V. S., Ranjan, V. P., Goel, S., John, M. & Dubey, B. Pathways to sustainable waste management in Indian smart cities. *J. Urban Manag.* **10**, 419–429 (2021).
- Behm, J. How well do you know your landfills?.
- Bzai, J. et al. Machine learning-enabled internet of things (IoT): Data, applications, and industry perspective. *Electronics* **11**, 2676 (2022).
- Patil, J. S. & Sankpal, S. V. Iot enabled e-waste management system for smart city. *Int. J. Sci. Dev. Res.* **6**(3), 1555–1560 (2021).
- Ijmaru, G. K., Ang, L. M. & Seng, K. P. Transformation from IoT to IoV for waste management in smart cities. *J. Netw. Comput. Appl.* **204**, 103393 (2022).

11. Vishnu, S., Ramson, S. J., Rukmini, M. S. S. & Abu-Mahfouz, A. M. Sensor-based solid waste handling systems: A survey. *Sensors* **22**, 2340 (2022).
12. Shanthini, E. et al. Iot based smart city garbage bin for waste management. In *2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 105–110 (IEEE, 2022).
13. Prastyabudi, W. A. & Permata, O. A. A conceptual design of waste management: Smart bin deployment in Surabaya smart city. *Acta Mechanica Malaysia (AMM)* **4**, 5–9 (2021).
14. Akbarpour, N., Salehi-Amiri, A., Hajiaghahi-Keshteli, M. & Oliva, D. An innovative waste management system in a smart city under stochastic optimization using vehicle routing problem. *Soft. Comput.* **25**, 6707–6727 (2021).
15. Aithal, P. et al. Smart city waste management through ICT and IoT driven solution. *Int. J. Appl. Eng. Manag. Lett. (IJAEML)* **5**, 51–65 (2021).
16. Fayomi, G. et al. Smart waste management for smart city: Impact on industrialization. In *IOP Conference Series: Earth and Environmental Science*, vol. 655, 012040 (IOP Publishing, 2021).
17. Wang, Q. & Li, S. Experience of urban solid waste management in Russia under the concept of smart city and its enlightenment to Shenyang. In *IOP Conference Series: Earth and Environmental Science*, vol. 719, 042021 (IOP Publishing, 2021).
18. Nasreen Banu, M. I. & Metilda Florence, S. Convergence of artificial intelligence in IoT network for the smart city—waste management system. In *Expert Clouds and Applications*, 237–246 (Springer, 2022).
19. Elyta, E., Wiko, G., Fahrana, Y., Rahman, I. & Zhan, F. F. Model-based waste management handling angkuts application to smart city: The perspective of human security. *Turkish J. Comput. Math. Educ.* **12**(4), 1555–1560 (2021).
20. Peoples, C. et al. A smart city economy supported by service level agreements: A conceptual study into the waste management domain. *Smart Cities* **4**, 952–970 (2021).
21. Shivran, P. S. Smart city waste management system. *Int. J. Comput. Appl.* **174**, 30–33. <https://doi.org/10.5120/ijca2021921059> (2021).
22. Sadiyiah, A. F., Purnomo, E. P. & Kasiwi, A. N. Waste management in the implementation of smart city in Bogor city. *Jurnal Ilmu Pemerintahan Widya Praja* **46**, 271–279 (2020).
23. Dzhuguryan, T. & Deja, A. Sustainable waste management for a city multifloor manufacturing cluster: A framework for designing a smart supply chain. *Sustainability* **13**, 1540 (2021).
24. Abuga, D. & Raghava, N. Real-time smart garbage bin mechanism for solid waste management in smart cities. *Sustain. Cities Soc.* **75**, 103347 (2021).
25. Vishnu, S. et al. IoT-enabled solid waste management in smart cities. *Smart Cities* **4**, 1004–1017 (2021).
26. Scariah, M. et al. Automated waste monitoring and alert system using IoT. In *Emerging Technologies in Data Mining and Information Security*, 349–357 (Springer, 2021).
27. Zhang, C., Zhang, X., Tu, D. & Wang, Y. Small object detection using deep convolutional networks: Applied to garbage detection system. *J. Electron. Imaging* **30**, 043013 (2021).
28. Gondal, A. U. et al. Real time multipurpose smart waste classification model for efficient recycling in smart cities using multilayer convolutional neural network and perceptron. *Sensors* **21**, 4916 (2021).
29. Akram, S., Singh, R., Gehlot, A. & Thakur, A. Design and implementation of a wide area network based waste management system using blynk and cayenne application. *Iran. J. Electr. Electron. Eng.* **17**, 1941–1941 (2021).
30. Ghanbari, F., Kamalan, H. & Sarraf, A. An evolutionary machine learning approach for municipal solid waste generation estimation utilizing socioeconomic components. *Arab. J. Geosci.* **14**, 1–16 (2021).
31. Faisal, T., Awawdeh, M. & Bashir, A. Design and development of intelligent waste bin system with advertisement solution. *Bull. Electr. Eng. Inf.* **10**, 940–949 (2021).
32. John, J. et al. Smart prediction and monitoring of waste disposal system using IoT and cloud for IoT based smart cities. *Wirel. Pers. Commun.* **122**, 243–275 (2022).
33. Fan, L. et al. Introducing an evolutionary-decomposition model for prediction of municipal solid waste flow: Application of intrinsic time-scale decomposition algorithm. *Eng. Appl. Comput. Fluid Mech.* **15**, 1159–1175 (2021).
34. Agrawal, R. & Sharma, A. Smart bin management system with IoT-enabled technology. In *Soft Computing: Theories and Applications*, 145–153 (Springer, 2021).
35. Kabir, M. H., Roy, S., Ahmed, M. T. & Alam, M. Iot based solar powered smart waste management system with real time monitoring—an advancement for smart city planning. *Glob. J. Comput. Sci. Technol.* **20**, 11–20 (2020).
36. Hassan, R. J. et al. State of art survey for IoT effects on smart city technology: Challenges, opportunities, and solutions. *Asian J. Res. Comput. Sci.* **22**, 32–48 (2021).
37. Ali, N. A. L., Ramly, R., Sajak, A. A. B. & Alrawashdeh, R. IoT e-waste monitoring system to support smart city initiatives. *Int. J. Integr. Eng.* **13**, 1–9 (2021).
38. Zaki, T., Jahan, I. T., Hossain, M. S. & Narman, H. S. An IoT-based complete smart drainage system for a smart city. In *2021 IEEE 12th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON)*, 0553–0558 (IEEE, 2021).
39. Ziouzos, D., Baras, N., Dasygenis, M. & Tsanaktisidis, C. Envisioning IoT applications in a smart city to underpin an effective municipal strategy: The smartbin project. In *SHS Web of Conferences*, vol. 102, 04020 (EDP Sciences, 2021).
40. Shah, A. et al. A review of IoT-based smart waste level monitoring system for smart cities. *Indones. J. Electr. Eng. Comput. Sci.* **21**, 450–456 (2021).
41. Pardini, K. et al. A smart waste management solution geared towards citizens. *Sensors* **20**, 2380 (2020).
42. Bano, A., Ud Din, I. & Al-Huqail, A. A. AIoT-based smart bin for real-time monitoring and management of solid waste. *Sci. Progr.* **2020**, 6613263 (2020).
43. Anh Khoa, T. et al. Waste management system using IoT-based machine learning in university. *Wirel. Commun. Mob. Comput.* **2020**, 6138637 (2020).
44. Mishra, A. & Kumar Ray, A. Iot cloud-based cyber-physical system for efficient solid waste management in smart cities: A novel cost function based route optimisation technique for waste collection vehicles using dustbin sensors and real-time road traffic informatics. *IET Cyber-Phys. Syst. Theory Appl.* **5**, 330–341 (2020).
45. Melakessou, F., Kugener, P., Alnaffakh, N., Faye, S. & Khadraoui, D. Heterogeneous sensing data analysis for commercial waste collection. *Sensors* **20**, 978 (2020).
46. Ali, T., Irfan, M., Alwadi, A. S. & Glowacz, A. Iot-based smart waste bin monitoring and municipal solid waste management system for smart cities. *Arab. J. Sci. Eng.* **45**, 10185–10198 (2020).
47. Srinivas, M., Benedict, S. & Sunny, B. C. Iot cloud based smart bin for connected smart cities—a product design approach. In *2019 10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 1–5 (IEEE, 2019).
48. Sharma, A. & Battula, R. B. Architecture for waste management in Indian smart cities (AWMINS). In *2019 International Conference on Information and Communication Technology Convergence (ICTC)*, 76–83 (IEEE, 2019).
49. Pardini, K., Rodrigues, J. J., Kozlov, S. A., Kumar, N. & Furtado, V. Iot-based solid waste management solutions: A survey. *J. Sens. Actuator Netw.* **8**, 5 (2019).
50. Chen, W.-E., Wang, Y.-H., Huang, P.-C., Huang, Y.-Y. & Tsai, M.-Y. A smart IoT system for waste management. In *2018 1st International Cognitive Cities Conference (IC3)*, 202–203 (IEEE, 2018).
51. Maksimovic, M. Leveraging internet of things to revolutionize waste management. *Int. J. Agric. Environ. Inf. Syst. (IJAIEIS)* **9**, 1–13 (2018).

52. Harith, M. Z. et al. Prototype development of IoT based smart waste management system for smart city. In *IOP Conference Series: Materials Science and Engineering*, vol. 884, 012051 (IOP Publishing, 2020).
53. Pelonero, L., Fornaia, A. & Tramontana, E. From smart city to smart citizen: rewarding waste recycle by designing a data-centric IoT based garbage collection service. In *2020 IEEE International Conference on Smart Computing (SMARTCOMP)*, 380–385 (IEEE, 2020).
54. Bhandari, R., Rathod, S., Singh, N., Desai, D. & Kotadiya, H. Iot based smart city bin. *Int. J. Comput. Appl.* **176**, 26–29. <https://doi.org/10.5120/ijca2020920103> (2020).
55. Badve, M., Chaudhari, A., Davda, P., Bagaria, V. & Kalbande, D. Garbage collection system using IoT for smart city. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 138–143 (IEEE, 2020).
56. Kalyuzhnyy, V., Costa, M., Silva, P. & Santos, J. Smart city IoT system-collectmywaste. In *2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT)*, 1–5 (IEEE, 2020).
57. Lokhande, A., Kshirsagar, A. & Deokar, S. Smart city waste management system for Swachh Bharat with IoT. *Int. J. Res. Appl. Sci. Eng. Technol.* **7**, 1523–1529 (2019).
58. Abdullah, N., Alwesabi, O. A. & Abdullah, R. Iot-based smart waste management system in a smart city. In *International Conference of Reliable Information and Communication Technology*, 364–371 (Springer, 2018).
59. Bakhshi, T. & Ahmed, M. Iot-enabled smart city waste management using machine learning analytics. In *2018 2nd International Conference on Energy Conservation and Efficiency (ICECE)*, 66–71 (IEEE, 2018).
60. Cicceri, G. et al. An intelligent hierarchical cyber-physical system for beach waste management: The BIOBLU case study. *IEEE Access* **11**, 134421–134445 (2023).
61. Ma, W. et al. DSYOLO-trash: An attention mechanism-integrated and object tracking algorithm for solid waste detection. *Waste Manage.* **178**, 46–56 (2024).
62. Shukhratov, I. et al. Optical detection of plastic waste through computer vision. *Intell. Syst. Appl.* **22**, 200341 (2024).
63. Wang, Y., Pelkonen, M. & Kaila, J. Effects of temperature on the long-term behaviour of waste degradation, emissions and post-closure management based on landfill simulators. *Open Waste Manag. J.* **5**, 19–27 (2012).
64. Inkscape Project. *Inkscape: Vector Graphics Editor*. Inkscape Project (2024). Version 1.3.2.
65. Guyon, I., Gunn, S., Nikravesh, M. & Zadeh, L. A. *Feature Extraction: Foundations and Applications* Vol. 207 (Springer, Berlin, 2008).
66. Kuncheva, L. I. *Combining Pattern Classifiers: Methods and Algorithms* (John Wiley & Sons, Hoboken, 2014).
67. Sahu, P. K. Artificial intelligence system for verification of schizophrenia via theta-EEG rhythm. *Biomed. Signal Process. Control* **81**, 104485 (2023).
68. Latif, S. et al. Analysis of birth data using ensemble modeling techniques. *Appl. Artif. Intell.* **37**, 2158273 (2023).
69. Jung, A. Recognizing mental states when diagnosing psychiatric patients via bci and machine learning. In *Proceedings of the Future Technologies Conference (FTC) 2022, Volume 1*, 644–655 (Springer, 2022).
70. Lipsa, S. & Dash, R. K. A novel dimensionality reduction strategy based on linear regression with a fine-pruned decision tree classifier for detecting DDoS attacks in cloud computing environments. In *Artificial Intelligence: First International Symposium, ISAI 2022, Haldia, India, February 17–22, 2022, Revised Selected Papers*, 15–25 (Springer, 2023).
71. Zhai, X. et al. Classification of arctic sea ice type in CFOSAT scatterometer measurements using a random forest classifier. *Remote Sens.* **15**, 1310 (2023).
72. Liao, C. Employee turnover prediction using machine learning models. In *International Conference on Mechatronics Engineering and Artificial Intelligence (MEAI 2022)*, vol. 12596, 227–231 (SPIE, 2023).
73. Huseinbasic, E. Angle instability and oscillations control using svc: A deep reinforcement learning enhanced local controller. *Eur. J. Electr. Eng. Comput. Sci.* **7**, 74–78 (2023).
74. Tian, W. et al. Structural analysis and classification of low-molecular-weight hyaluronic acid by near-infrared spectroscopy: A comparison between traditional machine learning and deep learning. *Molecules* **28**, 809 (2023).
75. Li, S. et al. Linear discriminant analysis with generalized kernel constraint for robust image classification. *Pattern Recogn.* **136**, 109196 (2023).
76. Kuschel, L. et al. Robust methylation-based classification of brain tumours using nanopore sequencing. *Neuropathol. Appl. Neurobiol.* **49**, e12856 (2023).
77. Cornish, R. P., Bartlett, J. W., Macleod, J. & Tilling, K. Complete case logistic regression with a dichotomised continuous outcome led to biased estimates. *J. Clin. Epidemiol.* **154**, 33–41 (2023).
78. Akhtar, M. S. & Feng, T. Evaluation of machine learning algorithms for malware detection. *Sensors* **23**, 946 (2023).
79. Merrin, P. N. et al. Machine learning based automated disaster message classification system using linear svc algorithm. In *Intelligent Cyber Physical Systems and Internet of Things: ICoICI 2022*, 869–879 (Springer, 2023).
80. Moon, N. N. et al. Machine learning approach to predict the depression in job sectors in Bangladesh. *Curr. Res. Behav. Sci.* **2**, 100058 (2021).
81. Tran, T.-N., Nguyen, T.-L., Hoang, V. T. & Voznak, M. Sensor clustering using a k-means algorithm in combination with optimized unmanned aerial vehicle trajectory in wireless sensor networks. *Sensors* **23**, 2345 (2023).
82. Malik, A. et al. Landslide susceptibility prediction based on decision tree and feature selection methods. *J. Indian Soc. Remote Sens.* **51**, 771–786 (2023).
83. Li, J. & Fan, S. GRNN: Graph-retraining neural network for semi-supervised node classification. *Algorithms* **16**, 126 (2023).
84. Yue, Y., Wang, G., Hu, J. & Li, Y. An improved label propagation algorithm based on community core node and label importance for community detection in sparse network. *Appl. Intell.* **53**, 17935–17951 (2023).
85. Ahmed, S. A software framework for predicting the maize yield using modified multi-layer perceptron. *Sustainability* **15**, 3017 (2023).
86. Pal, A. et al. Survey of fake news detection using machine intelligence approach. *Data Knowl. Eng.* **144**, 102118 (2023).
87. Frizzarin, M. et al. Classification of cow diet based on milk mid infrared spectra: A data analysis competition at the “international workshop on spectroscopy and chemometrics 2022”. *Chemom. Intell. Lab. Syst.* **234**, 104755 (2023).
88. Bouchiha, D., Bouziane, A. & Doumi, N. Machine learning for Arabic text classification: A comparative study. *Malays. J. Sci. Adv. Technol.* **4**, 163–173 (2022).
89. Elbagir, S. & Yang, J. Sentiment analysis of twitter data using machine learning techniques and scikit-learn. In *Proceedings of the 2018 International Conference on Algorithms, Computing and Artificial Intelligence*, 1–5 (2018).
90. de Andrade Silva, H. et al. Detection of terbufos in cases of intoxication by means of entomotoxicological analysis using ATR-FTIR spectroscopy combined with chemometrics. *Acta Trop.* **238**, 106779 (2023).
91. Aguirre-Munizaga, M., Del Cioppo Morstadt, J. & Samaniego-Cobo, T. Analysis of classification algorithms for the prediction of purchase intention in electronic commerce. In *Technologies and Innovation: 8th International Conference, CITI 2022, Guayaquil, Ecuador, November 14–17, 2022, Proceedings*, 30–42 (Springer, 2022).

Author contributions

Conceptualization, T.S., H.K.S.M., C.L., U.S., S.A., and H.K.; writing—original draft preparation, T.S., H.K.S.M., C.L., U.S., S.A., and H.K.; writing—review and editing, T.S., H.K.S.M., C.L., U.S., S.A., and H.K.; All authors have read and agreed to the published version of the manuscript.

Funding

This work was supported in part by the National Science and Technology Council under Grant NSC113-2410-H-167-012-MY3 and NSC113-2634-F-005-001-MBK.

Declarations

Competing interests

The corresponding author is responsible for submitting on behalf of all authors of the paper.

Additional information

Correspondence and requests for materials should be addressed to C.-C.L., S.A. or H.K.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

© The Author(s) 2025