**D. Kantayeva, A. Zhalgas, Sh. Akhmetzhanova**

*Astana IT University,*

*Astana, Kazakhstan*

*d.kantayeva@astanait.edu.kz, aidana.zhalgas@astanait.edu.kz, sh.akhmetzhanova@astanait.edu.kz*

**AI-DRIVEN PREDICTION OF CONTAINER OVERFLOW FOR SMART WASTE COLLECTION IN URBAN ENVIRONMENTS**

**Abstract**

This paper presents the design and assessment of an artificial intelligence-enabled prediction model for smart waste collection in urban waste management. Employing Long Short-Term Memory (LSTM) neural networks and Extreme Gradient Boosting (XGBoost) regression, the proposed model predicts waste container fill percentages and identifies overflow risks based on historic trends and operational behavior. As data for Kazakhstan does not exist publicly, a simulated dataset that approximated real waste container behavior was established. The LSTM model achieved high overall accuracy but did not identify rare overflow episodes due to class imbalance, while XGBoost demonstrated excellent predictive capacity for frequent fill patterns with higher sensitivity to overflow limits. The results show that the hybrid predictive models can significantly improve proactive waste collection, lower resource usage, and facilitate the construction of smart city infrastructure in developing urban settings. The research results can be utilized to inform future policy initiatives and operational strategies for the implementation of the smart waste management system in Kazakhstan and elsewhere.

***Key words:*** *artificial intelligence, deep learning, LSTM, XGBoost, waste collection, prediction.*

**Introduction**

Effective municipal solid waste management maintains urban cleanliness and reduces environmental impact. Proper management minimizes the risks such as contamination of soil and water resources. Despite its importance, many cities worldwide continue to rely on traditional waste collection methods. This often leads to unnecessary fuel consumption and to overflow of waste containers. These issues highlight the need for more intelligent and responsive waste collection systems.

In recent years, the development of smart city technologies has introduced new possibilities for optimizing waste management through the use of real-time data, sensors, and artificial intelligence. One promising direction is the use of machine learning models for forecasting waste container fill levels.

The integration of machine learning techniques works well with forecasting waste container fill levels. Predictive models can anticipate peak load events allowing collection services to proactively schedule pickups before overflow occurs. We have identified the moment when the container reaches and exceeds 75% of its filling as the peak. Reinforcement learning algorithms can optimize the routing of collection trucks by considering environmental and operational factors such as traffic conditions and fuel consumption. The research includes forecasting container fill levels using utilizing time series forecasting and classification models to accurately predict when containers will reach critical fill thresholds.

This research is significant for Kazakhstan, where the adoption of sensor-based waste management systems remains limited and real-world data are scarce. By developing and validating a synthetic dataset alongside these machine learning models, this study aims to establish a foundational framework for future smart waste management solutions within the country’s urban infrastructure. After the conclusions, you can find our suggestions for improving the research results.

Furthermore, the outcomes of this research paper may be interesting for policymakers and technology developers about the benefits and practical considerations of implementing intelligent waste management systems in Kazakhstan. By demonstrating the effectiveness of machine learning-driven approaches, this work supports the transition toward smarter cities that improve citizen’s quality of life.

## Literature Review

The growing volume of urban waste and the limitations of static collection schedules have prompted municipalities to seek intelligent, data-driven solutions to improve waste management efficiency. In this context, artificial intelligence (AI) techniques, particularly machine learning (ML) and deep learning (DL), have emerged as powerful tools for forecasting waste container fill levels and enabling proactive collection strategies.

#### ****Smart Waste Management and IoT Integration****

Recent advances in **Internet of Things (IoT)** technologies have enabled the real-time monitoring of waste containers using embedded sensors that record fill levels, temperature, waste types, and geolocation [1]. These smart bins generate continuous data streams that can be used to develop predictive models for forecasting when a container is likely to overflow. Early efforts in this space focused primarily on rule-based systems or simple statistical models, which proved inadequate in capturing the nonlinear and dynamic patterns present in urban waste generation.

#### ****Machine Learning for Fill-Level Forecasting****

Machine learning algorithms such as **decision trees, support vector machines (SVM),** and **gradient boosting** have been widely applied to model container fill dynamics [2],[3]. Among these, **XGBoost** has shown promising results due to its ability to handle sparse features and temporal dependencies with minimal preprocessing. For instance, Zhang et al. [4] demonstrated the effectiveness of gradient boosting for forecasting waste generation at the district level by incorporating temporal features such as hour, day-of-week, and seasonal variations.

#### ****Deep Learning Approaches****

In recent years, **Recurrent Neural Networks (RNNs)** and their variants, especially **LSTM** networks, have been adopted for time-series prediction tasks in waste management due to their ability to learn from historical temporal dependencies. Studies such as that by Namoun et al. [5] implemented LSTM models to predict the fill level of containers up to 48 hours in advance, achieving higher accuracy compared to conventional models. LSTM’s memory cell architecture enables it to retain long-term contextual information, making it suitable for learning complex waste accumulation trends influenced by population density, time of day, and localized events.

#### ****Hybrid and Comparative Modeling****

Several studies have compared and combined ML and DL models to leverage their complementary strengths. For instance, Airaksinen et al. [6] compared LSTM, GRU, and XGBoost models on real-world sensor data and concluded that hybrid models provided better robustness and interpretability. These findings suggest that combining models like **LSTM (for sequential learning)** and **XGBoost (for structured feature learning)** may offer a reliable forecasting solution for smart bin systems.

#### ****Gaps and Research Motivation****

Despite the advances, many existing models overlook key operational constraints, such as collection intervals, traffic conditions, or truck routing integration. Moreover, most studies emphasize prediction accuracy without integrating the forecasts into actionable planning tools. There is a need for systems that not only **forecast container overflow**but also feed these predictions into **decision-making systems** for optimizing collection schedules and routes in real time.

This study addresses these gaps by developing an **AI-driven hybrid prediction framework**using both LSTM and XGBoost to forecast waste container overflow in urban environments. It further adopts a visual and evaluation-driven approach to support proactive, efficient waste collection management.

## *Public Waste Management Datasets*

Several international datasets support the analysis of waste management systems globally. Usually they are compiled by governmental agencies and international organizations.

Key datasets include:

* ***Eurostat Municipal Waste Data:*** This dataset offers detailed statistics on municipal waste generation and management across European Union member states. It facilitates comparative studies on waste treatment methods and policy effectiveness.
* ***OECD Municipal Waste Statistics:*** The Organisation for Economic Co-operation and Development (OECD) provides data on municipal waste generation, recycling rates, and disposal methods among its member countries, enabling cross-national performance assessments.
* ***World Bank “What a Waste” Database:*** This global dataset aggregates municipal solid waste data from numerous countries, providing a macro-level perspective on waste generation and management practices worldwide. It serves as a foundation for international benchmarking and policy formulation.
* ***Waste Atlas:*** Waste Atlas is an open-access, crowdsourced platform presenting municipal solid waste management data for over 160 countries and approximately 1,800 cities. It includes information on waste generation volumes, treatment infrastructure, and landfill sites, supporting both academic research and urban planning.

Several smart city deployments have generated real-time datasets through IoT-enabled waste management systems. Examples include:

* Utilization of NB-IoT and LoRaWAN technologies to monitor street waste containers in **Reykjavik (Iceland)**.
* Deployment of approximately 4,500 sensors for general waste monitoring in **Buenos Aires (Argentina)**.
* Real-time monitoring of 126 semi-underground waste bins **Patras (Greece)**.
* Implementation of smart dumpsters within a school district to optimize collection routes **Ohio** (**USA**).

The datasets are accessible through data.santander.es, opendata.cph.dk and other public data portals, either via REST APIs or as downloadable CSV/JSON files.

**Methodology**

This study aims to predict waste container fill levels and forecast critical thresholds to support timely waste collection. Two predictive modeling approaches were employed during the study Long Short-Term Memory (LSTM) neural networks and Extreme Gradient Boosting (XGBoost) regression, both evaluated for their ability to forecast fill levels over short horizons (next 48 hours) using historical data.

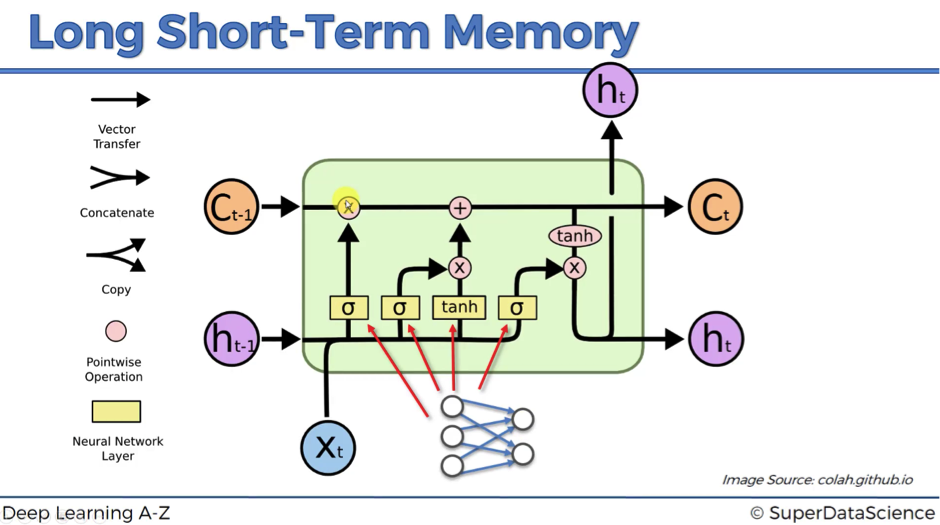
All numerical features were normalized using MinMaxScaler [7] to improve convergence and stability in training. It is a data normalization technique which transforms numerical features to fixed values (e.g. [0,1]). For each X value in the feature column it calculates scaled value:

where is a minimum value in the column, is a maximum value in the column.

***LSTM-Based Classification Approach***

To effectively model the temporal dynamics of container fill levels, this study employs a LSTM network [8] – a specialized type of recurrent neural network (RNN) designed to capture long-term dependencies in sequential data. Unlike traditional RNNs that struggle with vanishing gradients and limited memory capacity, LSTMs introduce a gated architecture that enables them to retain, update, or forget information across time steps.

The structure of an LSTM cell (Figure 1) consists of three primary gates: the forget gate, the input gate, and the output gate. These gates regulate the flow of information into and out of the cell's internal state, denoted as Ct, which serves as the long-term memory of the sequence. The output at each time step is denoted as ht​, referred to as the hidden state [10].



**Figure 1.** LSTM Cell Architecture [9]

1. Forget Gate:  
   The forget gate *ft* determines which information from the previous cell state *Ct−1​* should be discarded. It is computed as:

where *xt​* is the current input, *ht−1*​ is the previous hidden state, *Wf*​ are the learnable weights, and *σ* is the sigmoid activation function. The output of *ft*lies in the range [0, 1], where values closer to 0 mean “forget” and closer to 1 mean “retain”.

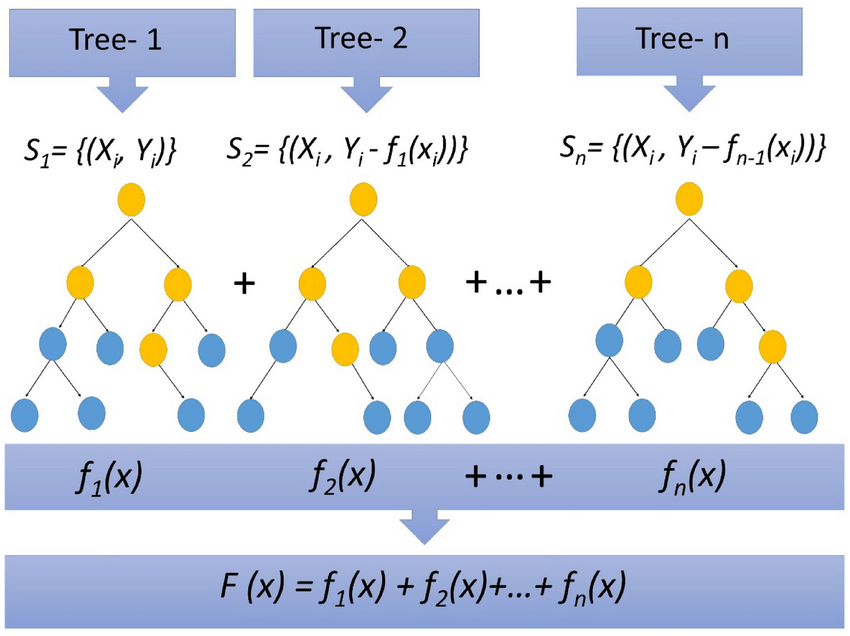
1. Input Gate and Candidate Memory**:**  
   The input gate*it* controls which new information should be added to the cell state. A candidate memory  is also computed using a tanh activation function:
2. Cell State Update**:**  
   The new cell state *Ct* is a combination of the retained previous state and the new candidate memory:
3. Output Gate and Hidden State**:**Finally, the output gate otot​ determines the next hidden state htht​, which is both used for predictions and passed to the next time step:

The gating mechanism ensures that relevant temporal information is retained across long sequences, making LSTMs highly suitable for tasks such as containerfill-level forecasting**,** where historical fill behavior significantly influences future trends. To capture temporal dependencies in fill patterns, an LSTM-based binary classifier was developed to predict whether a container would reach the critical fill threshold of 75% within the next 48 hours. As an input cosidered last 10 time steps (i.e., 2.5 hours of historical fill level data), whereas output is binary label, which shows 1 if fill level ≥ 75% in next 48 hours, else 0.

Each sequence-label pair was used to train an LSTM with two stacked recurrent layers (64 and 32 units) followed by a sigmoid-activated dense layer. The model was trained using binary cross-entropy loss and the Adam optimizer for 10 epochs. Performance was evaluated using classification metrics including accuracy, F1-score, and ROC-AUC.

***XGBoost-Based Regression Approach***

To complement the LSTM-based sequence modeling, we also employ XGBoost [11], a scalable and highly efficient tree-based ensemble learning method. XGBoost is particularly well-suited for structured tabular data and short-term forecasting tasks, making it a valuable component in our fill-level prediction system.



**Figure 2.** XGBoost Architecture [12]

As illustrated in Figure 2, XGBoost operates by building an ensemble of decision trees in a stage-wise, additive manner. The core idea is to improve the prediction accuracy iteratively by fitting new trees to the residual errors of the previous model [10].

* The training process begins with the first decision tree *f1(x),* trained on the original dataset *S1={(xi,yi)},* where *xi*​ represents the input features and *yi* the observed fill level.
* The second tree *f2(x)* is then trained on the residuals, i.e., the difference between the actual target *yi*​ and the output of the first model *f1(x)*. This forms a new dataset .
* This process continues for nn iterations, with each subsequent tree  minimizing the residual errors from the combined outputs of all preceding trees.

The final prediction is the sum of all individual trees:

This sequential learning mechanism enables the model to capture complex nonlinear relationships and subtle patterns in the feature space, including temporal and categorical effects relevant to container fill-level dynamics.

A parallel XGBoost regression model was trained to directly forecast future fill levels over a 48-hour horizon (192 time steps at 15-minute intervals). The model used lagged and contextual features (Hour, DayOfWeek, Minute, Time\_Step, and encoded Waste\_Type) to learn progression patterns in fill level dynamics. A final forecast curve was generated by recursively feeding predictions as inputs, updating the Time\_Step feature iteratively. The model was trained using squared error loss (reg:squarederror) with 100 estimators and default hyperparameters.

***Synthetic Dataset Generation***

In Kazakhstan, the implementation of sensor-based waste management systems remains limited, and as a result, there are currently no publicly available datasets relevant to this domain. To develop and test machine learning models, it was generated a synthetic dataset (Table 1) designed to closely repeat the behavior of real-world waste containers. The process was guided by logical rules and operational assumptions commonly adopted in smart city pilot projects.

Key Assumptions:

* The dataset simulates 10 waste containers and 10 collection trucks.
* Each container is monitored at 15-minute intervals over a period of one year, resulting in a total of more than 350,000 records.

**Table 1.** Information about dataset

| **Field Name** | **Description** |
| --- | --- |
| Container\_ID | Unique identifier for each waste container (C001- C010) |
| Timestamp | Exact datetime of the sensor reading (15-minute interval) |
| Fill\_Level | Current fill level of the container, as a percentage (0 - 100%) |
| Waste\_Type | Category of waste: liquid, paper, plastic, glass, or other |
| Latitude\_Container, Longitude\_Container | Geospatial coordinates of the container |
| Date | Date portion of the timestamp, for aggregation purposes |
| Truck\_ID | Unique identifier for the truck assigned to the container on that day |
| Total\_Distance\_km | Distance traveled by the truck to reach the container (recorded on collection) |
| Travel\_Time\_min | Estimated travel time for the truck to reach the container |
| Fuel\_Cost\_USD | Estimated cost of fuel consumption for the trip |

Operational Logic

* A collection truck is dispatched when the fill level of a container reaches or exceeds 75%. The system models a short delay to account for the truck’s travel time.
* After collection, the fill level is reset to 0%, and the fields for distance, travel time, and fuel cost are updated.
* Each container is assigned to one truck per day. For records where no collection occurs, the distance, travel time, and fuel cost fields remain zero.

***Visualization of Forecast Results***

To facilitate model interpretability and support operational decision-making, the forecasting results were visualized using a layout inspired by the Prophet framework. This visual approach integrates both historical observations and short-term forecasts within a unified timeline, enabling stakeholders to assess model behavior in context.

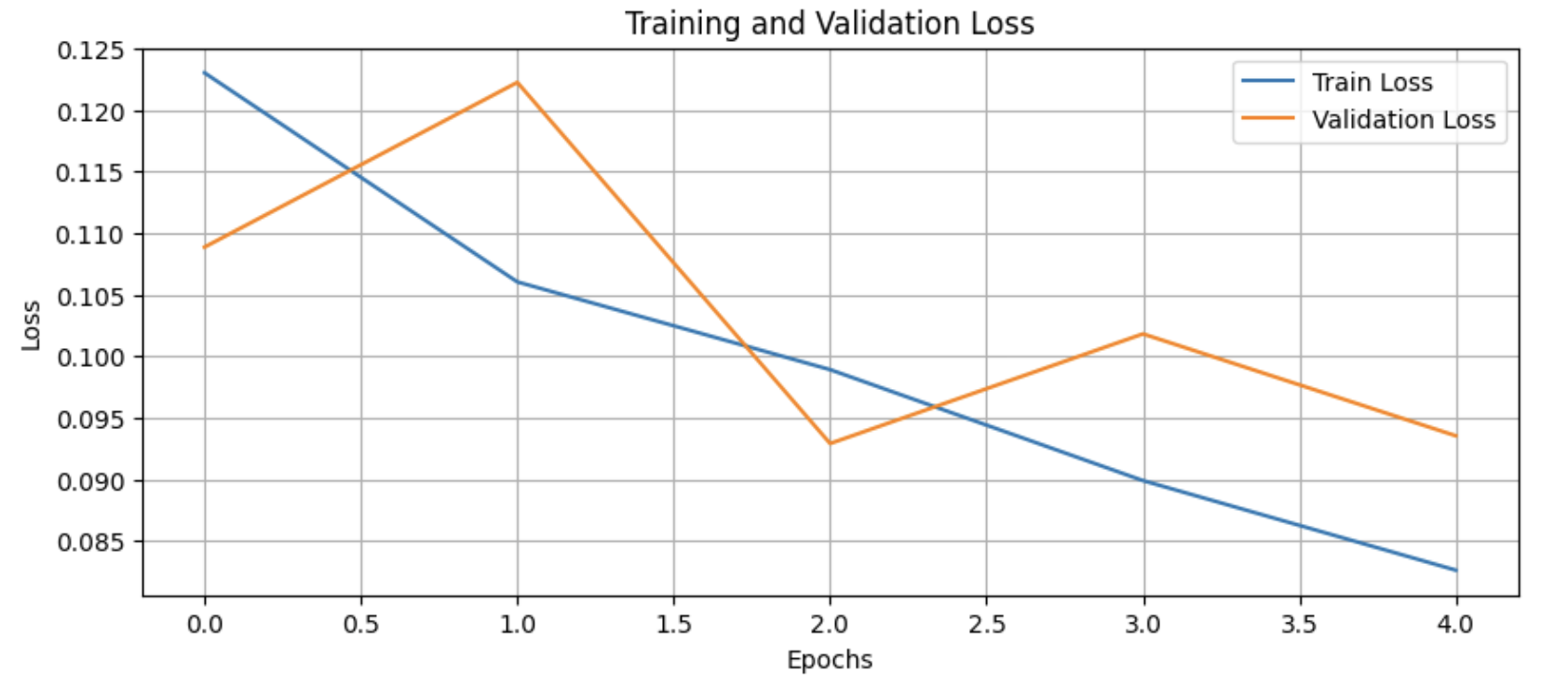
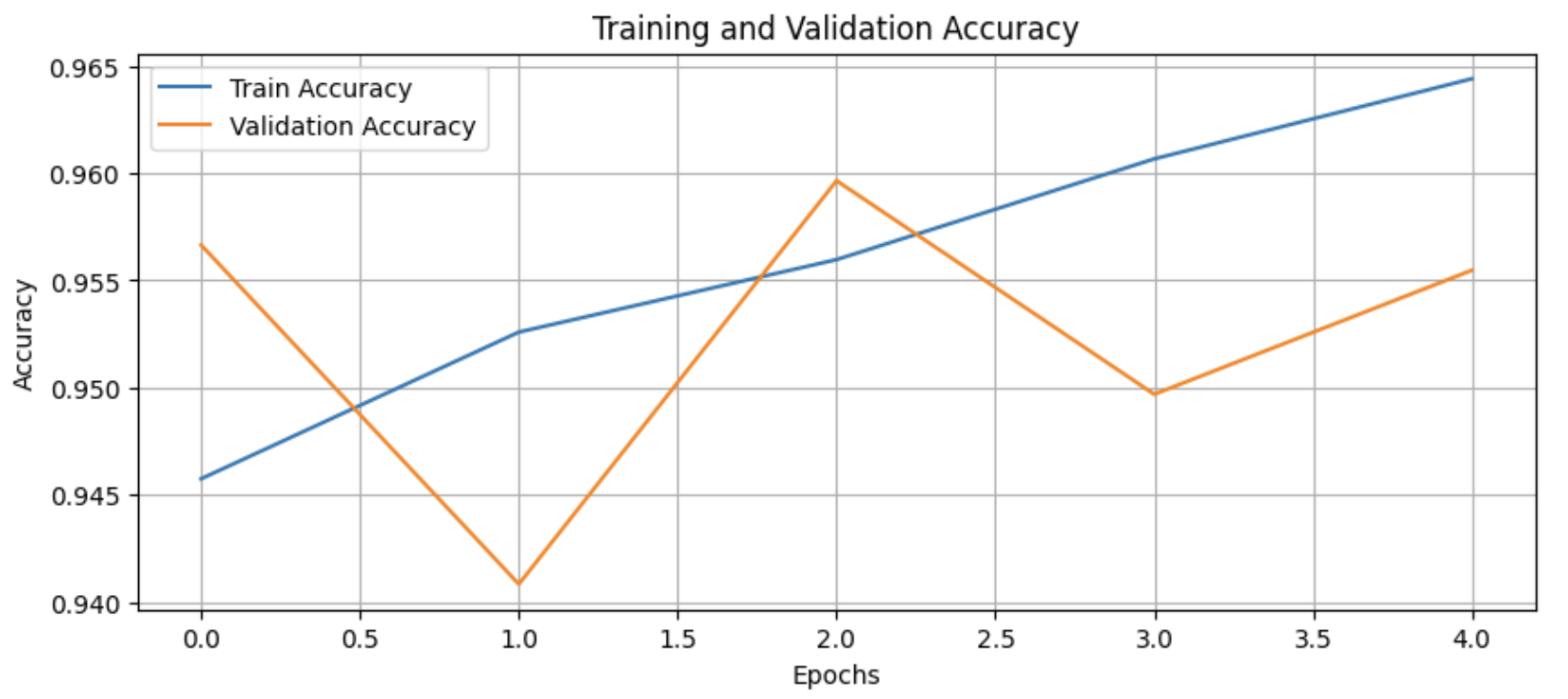
For each container, it was displayed:

* The last 7 days of observed data, shown as black scatter points, representing the actual fill level recorded at 15-minute intervals.
* The model's 48-hour (2-day) forecast, presented as a continuous blue line, indicating the expected fill level trajectory based on either the LSTM or XGBoost model.
* A horizontal red dashed line at the 75% threshold, marking the critical point at which waste collection is triggered according to the system design.
* A shaded region (±5%) around the forecast line, denoting uncertainty or potential variation in predicted values. While this is more naturally derived from probabilistic models, it is included for visual consistency.

This visualization strategy provides a clear temporal view of how fill levels are evolving and when containers are expected to exceed operational limits. By combining observed and predicted trends in a single figure, the plots support proactive planning, allow for anomaly detection (e.g., unexpected spikes), and help evaluate the effectiveness of the forecasting model in real-world scenarios. Figures generated for each container are stored and presented in the results section, with automated routines producing forecasts for all containers under monitoring.

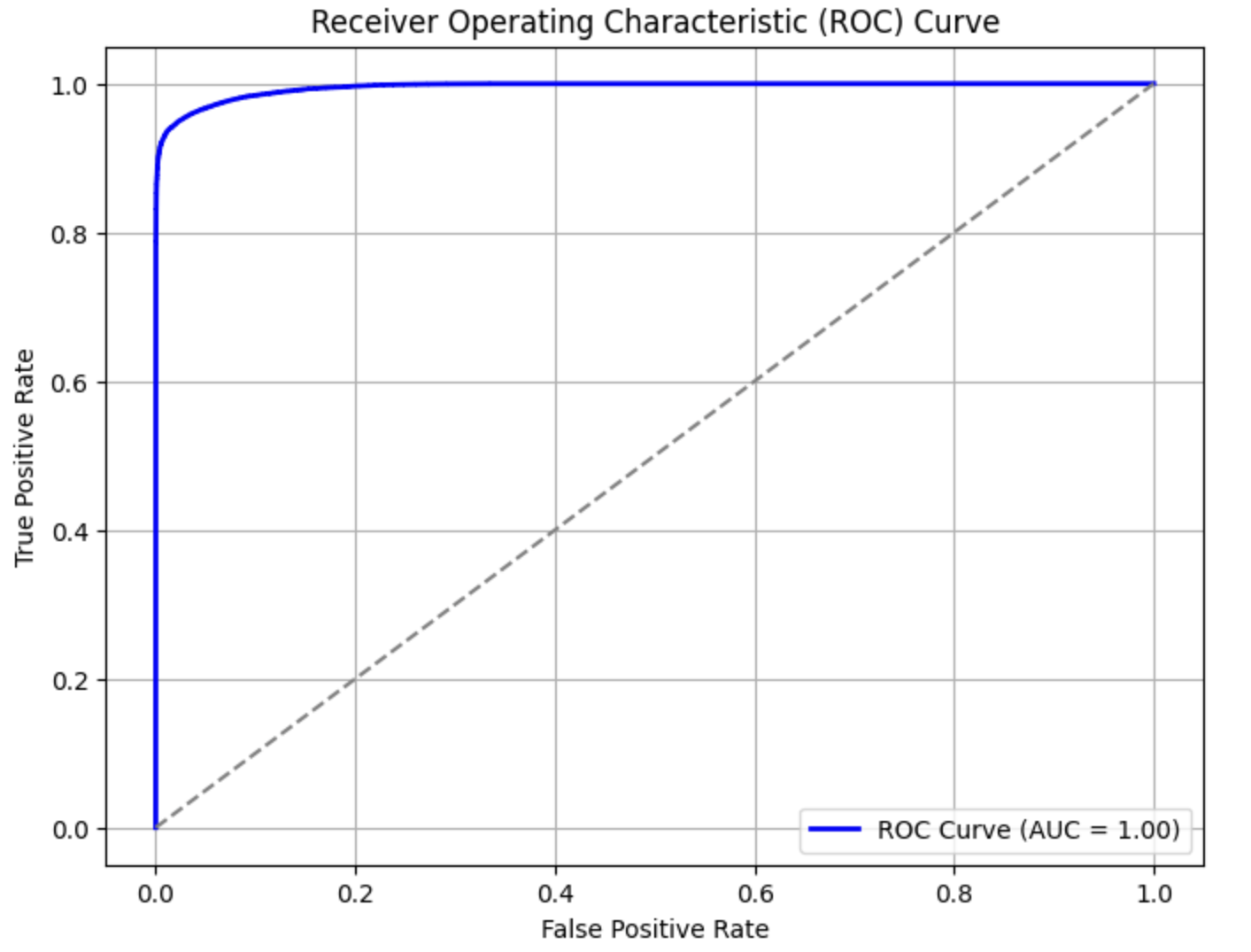
**Experimental Results**

The accuracy plot (Figure 3a) shows a steady increase in training accuracy across epochs, from ~0.946 to ~0.965. Validation accuracy exhibits more fluctuation but remains within a narrow band (~0.940–0.960), suggesting that the model is learning without significant overfitting. The training loss (Figure 3b) consistently decreases across epochs, while validation loss shows slight fluctuations. The convergence of training and validation loss values implies that the model generalizes well to unseen data, although the instability in validation loss suggests minor variance in performance depending on batch composition.

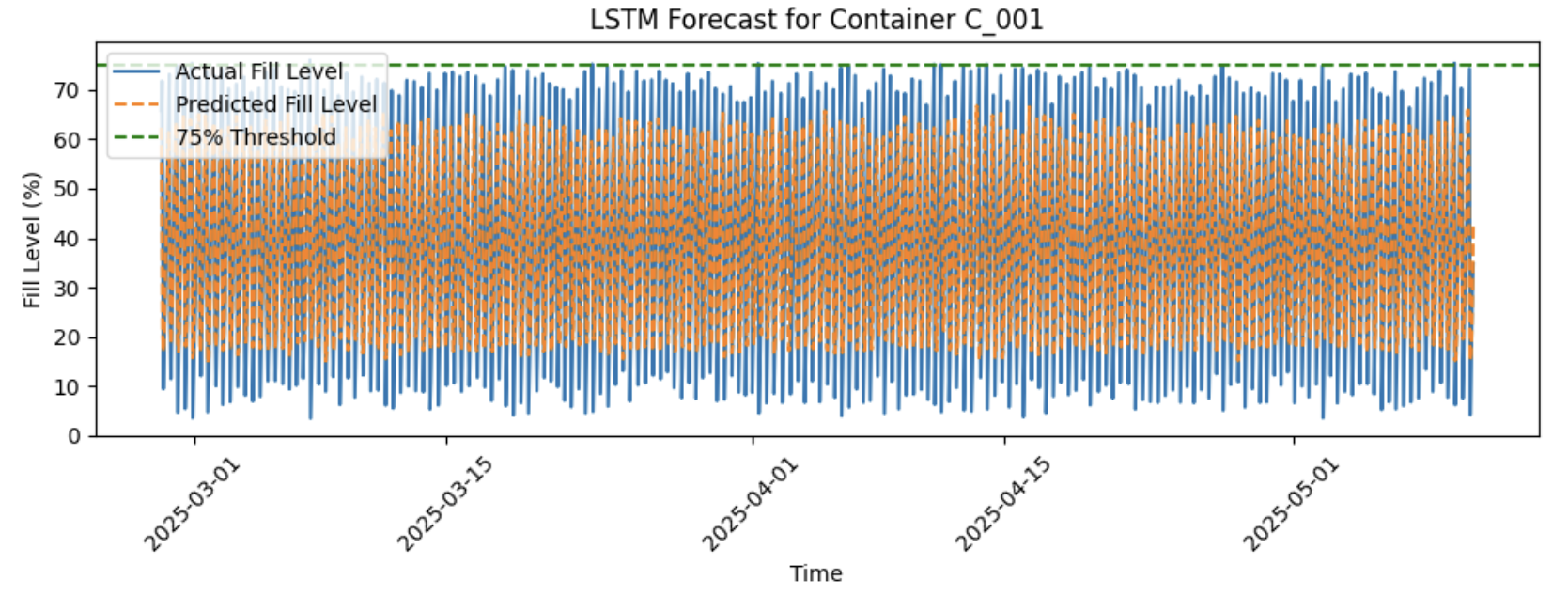


**Figure 3.** Training and Validation a) Accuracy, b) Loss

The ROC curve (Figure 4) yields an AUC of **1.00,** which appears overly optimistic given the model's failure to predict class 1. This discrepancy suggests the model perfectly separates class 0 instances but lacks resolution on the minority class. The high AUC may thus be misleading in imbalanced scenarios and should be interpreted cautiously.



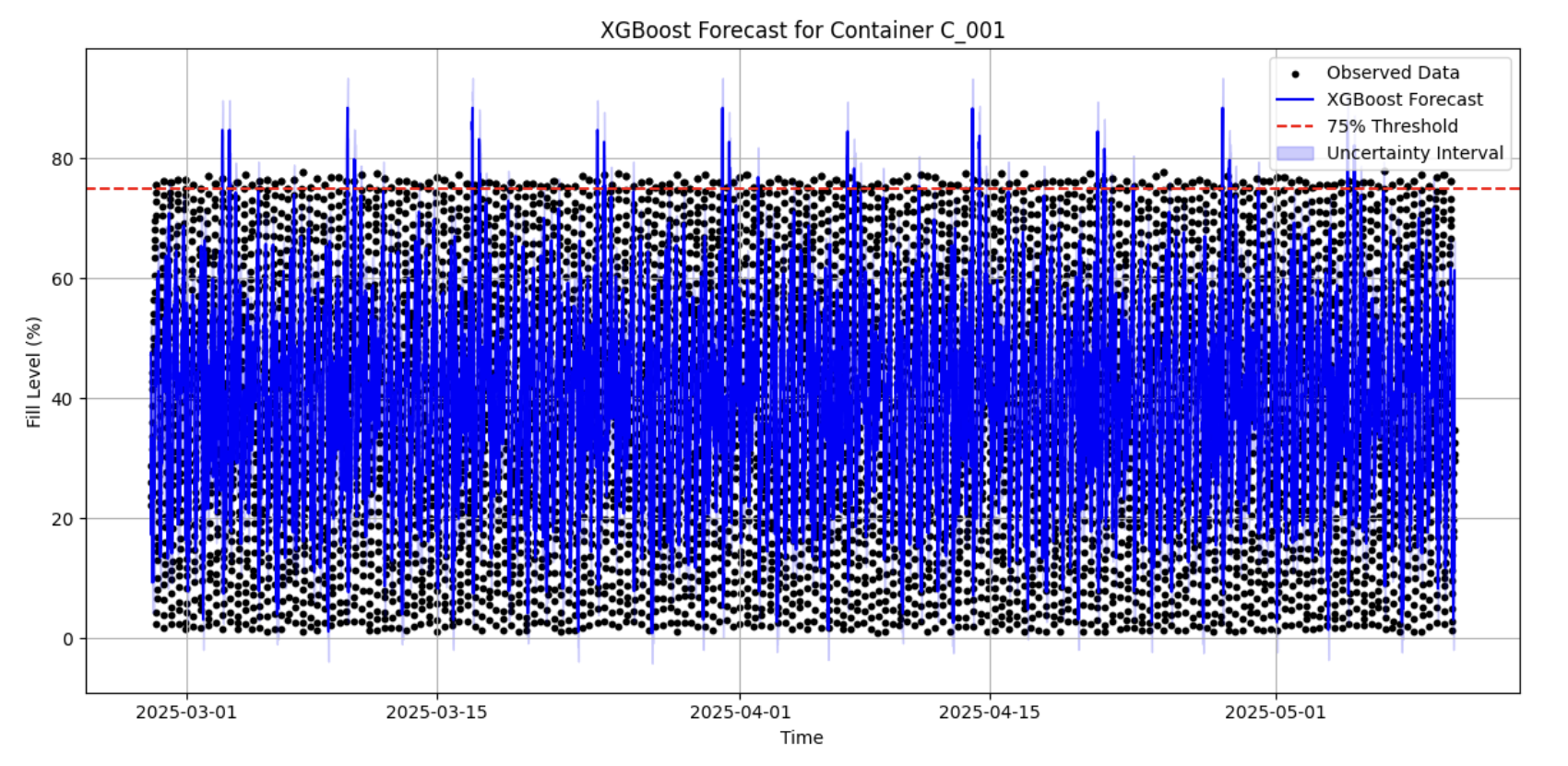
**Figure** **4.** The ROC curve



**Figure** **5.** LSTM Forecast

Figure 5 depicts the LSTM model's performance in forecasting the fill level of container C\_001 from March 1 to May 5, 2025. The solid blue line indicates the actual recorded fill levels, while the orange dashed line corresponds to the LSTM-predicted values. The green dashed horizontal line marks the 75% threshold—used as an indicator for scheduling waste collection. The LSTM model demonstrates a strong ability to replicate the general temporal dynamics of fill level changes. Visually, the forecasted values exhibit close alignment with the true fill levels, capturing periodic fluctuations and seasonal patterns across the two-month period. Quantitatively, the LSTM model achieved an overall accuracy of 99.5% (Table 2). These high scores indicate the model’s effectiveness in identifying normal operational states. However, performance on the minority class (1, overflow conditions) was limited, with precision, recall, and F1-score all equal to 0.000, and only 8 instances out of 1748 labeled as overflow in the dataset. This suggests that the model failed to detect any positive overflow events. While the weighted average F1-score was 0.993, the macro-averaged F1-score was only 0.499, revealing a lack of generalization to rare but critical cases.

The LSTM model demonstrates excellent accuracy in forecasting typical behavior but fails to recognize rare overflow conditions due to significant class imbalance. To improve detection of critical events, further strategies such as oversampling, anomaly-aware loss functions, or hybrid models may be required. Nevertheless, the model shows promise for regular fill level prediction and temporal trend learning.



**Figure** 6. XGBoost Forecast

Figure 6 illustrates the forecasted fill levels for container C\_001 using an XGBoost regression model, with the observed data, predicted values, and uncertainty intervals over a continuous time period from March 1 to May 5, 2025. The black dots represent the historical observed fill level data, while the solid blue line indicates the XGBoost model's forecast. The shaded blue region around the forecast line depicts the uncertainty interval, capturing potential variance in prediction outcomes. A red dashed horizontal line marks the 75% fill level threshold, indicating the critical level at which waste collection should ideally be triggered to prevent overflow. The forecast results demonstrate that the XGBoost model is capable of capturing daily fluctuations in fill level with high temporal granularity. Despite the presence of noise and spikes in the predicted values, the model trends align closely with the general pattern observed in the historical data. The frequent excursions of the forecasted fill levels beyond the 75% threshold suggest a high utilization rate for container C\_001, emphasizing the need for timely intervention in waste collection scheduling. Moreover, the narrow uncertainty bounds indicate a relatively stable and confident prediction performance by the model over the forecast horizon. Overall, the XGBoost model exhibits strong predictive ability for short-term fill level forecasting in smart waste containers, supporting the feasibility of data-driven optimization in urban waste collection systems.

**Table 2.** Details of Validation

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Precision** | | **Recall** | | **F1-score** | |
| **LSTM** | **XGBoost** | **LSTM** | **XGBoost** | **LSTM** | **XGBoost** |
| **0** | 0.995 | 0.949 | 1.000 | 0.990 | 0.998 | 0.969 |
| **1** | 0.000 | 0.080 | 0.000 | 0.017 | 0.000 | 0.028 |
| **Accuracy** |  |  |  |  | 0.995 | 0.940 |
| **Macro avg** | 0.498 | 0.515 | 0.500 | 0.503 | 0.499 | 0.498 |
| **Weighted avg** | 0.991 | 0.905 | 0.995 | 0.940 | 0.993 | 0.921 |

The model achieved an overall accuracy of 94.0%, primarily driven by its strong performance on the majority class (0, fill level below the threshold), which is shown in Table 2. These results indicate that the model is highly effective at identifying non-critical fill levels, minimizing false positives. However, performance on the minority class (1, fill level above threshold) was notably lower. This reflects the strong class imbalance in the dataset (6651 samples for class 0 vs. 358 for class 1), which limits the model’s ability to accurately predict rare overflow events. Consequently, the macro-averaged F1-score was 0.498, while the weighted average F1-score reached 0.921, demonstrating the dominance of the majority class in the overall performance metrics.

**Discussion**

The LSTM model demonstrates robust learning and predictive capabilities for the dominant class but suffers from low recall for overflow events. While training dynamics (loss and accuracy) indicate proper convergence and stability, the model's classification behavior reveals a need for strategies to better handle rare events—such as cost-sensitive training, synthetic data generation, or threshold optimization—to enhance performance on critical cases without sacrificing general accuracy.

The XGBoost model provides strong forecasting performance with high accuracy and reliability for typical fill levels. It learns daily usage patterns effectively and maintains consistent predictions with narrow uncertainty bounds. XGBoost does not train in epochs like neural networks, but the low uncertainty variance and smooth forecast trajectory suggest the model generalizes well on time-series data. Its tree-based ensemble structure captures non-linear patterns and seasonality, as evidenced by the model’s alignment with daily fill-level fluctuations. However, like the LSTM model, its inability to detect rare overflow conditions limits its operational robustness. Enhancing class sensitivity through imbalance-aware techniques will be crucial for improving the model’s practical utility in real-time smart waste collection systems.

**Conclusion and Future Work**

This study demonstrates the feasibility and promise of using AI-based predictive models, specifically LSTM and XGBoost, in forecasting waste container fill levels for enabling smart waste collection systems. Both models were able to learn general fill-level patterns; however, their abilities to detect overflow conditions were undermined by class imbalance in the data. To enhance model robustness, future work needs to incorporate additional variables such as seasonal influences, demographic profile, and event calendars, which influence waste generation. Moreover, the creation of an interactive dashboard with real-time geolocation of containers and route visualization is proposed to aid decision-making for operators. Other improvements such as anomaly-aware loss functions, oversampling methods, or ensemble models could be used to enhance detection of rare but critical overflow events. These innovations will enable the development of scalable, intelligent waste management systems that can adapt to the evolving demands of the urban environment.

**References**

1. Mitoi, M., Craciunescu, R., Vulpe, A., Fratu, O. (2015). Sensor-Based Environmental Monitoring for Ambient Assisted Living. In: Atanasovski, V., Leon-Garcia, A. (eds) Future Access Enablers for Ubiquitous and Intelligent Infrastructures. FABULOUS 2015. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering, vol 159. Springer, Cham. <https://doi.org/10.1007/978-3-319-27072-2_10>
2. Povetkin, K., Isaac, Sh. (2020). Identifying and addressing latent causes of construction waste in infrastructure projects, Journal of Cleaner Production, Volume 266, <https://doi.org/10.1016/j.jclepro.2020.122024>.
3. Ahmed, S., Mubarak, S., Du, J. T., & Wibowo, S. (2022). Forecasting the Status of Municipal Waste in Smart Bins Using Deep Learning. International Journal of Environmental Research and Public Health, 19(24), 16798. <https://doi.org/10.3390/ijerph192416798>
4. Zhang, H., Cao, H., Zhou, Y., Gu, Ch., Li, D. (2023). Hybrid deep learning model for accurate classification of solid waste in the society, Urban Climate, Volume 49, <https://doi.org/10.1016/j.uclim.2023.101485>.
5. Namoun, A., Hussein, B. R., Tufail, A., Alrehaili, A., Syed, T. A., & BenRhouma, O. (2022). An Ensemble Learning Based Classification Approach for the Prediction of Household Solid Waste Generation. Sensors, 22(9), 3506. <https://doi.org/10.3390/s22093506>
6. Airaksinen, M., Vanhatalo, S., & Räsänen, O. (2023). Comparison of End-to-End Neural Network Architectures and Data Augmentation Methods for Automatic Infant Motility Assessment Using Wearable Sensors. *Sensors*, *23*(7), 3773. <https://doi.org/10.3390/s23073773>
7. Ahsan, M. M., Mahmud, M. A. P., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect of Data Scaling Methods on Machine Learning Algorithms and Model Performance. *Technologies*, *9*(3), 52. <https://doi.org/10.3390/technologies9030052>
8. Benjamin Lindemann, Timo Müller, Hannes Vietz, Nasser Jazdi, Michael Weyrich. (2021). A survey on long short-term memory networks for time series prediction, Procedia CIRP, Volume 99, pages 650-655, <https://doi.org/10.1016/j.procir.2021.03.088>.
9. *SuperDataScience*. (n.d.). <https://www.superdatascience.com/blogs/recurrent-neural-networks-rnn-long-short-term-memory-lstm>
10. Garza-Ulloa, J. (2022). Chapter 6 - Deep Learning Models Evolution Applied to Biomedical Engineering, Applied Biomedical Engineering Using Artificial Intelligence and Cognitive Models, Academic Press, pages 509-607, <https://doi.org/10.1016/B978-0-12-820718-5.00012-X>.
11. Qi, Zh., Feng, Y., Wang, Sh., Li, Ch. (2025). Enhancing hydropower generation Predictions: A comprehensive study of XGBoost and Support Vector Regression models with advanced optimization techniques, Ain Shams Engineering Journal, Volume 16, Issue 1, <https://doi.org/10.1016/j.asej.2024.103206>.
12. Omarzai, F. (2025, March 25). XGBOOST Classification in depth - Fraidoon Omarzai - medium. *Medium*. <https://medium.com/@fraidoonomarzai99/xgboost-classification-in-depth-979f11ef4bf9>