**Bulletin of Electrical Engineering and Informatics**
Vol. 15, No. 6, December 2026, pp. 5216~5227
ISSN: 2302-9285, DOI: 10.11591/eei.v15i6.12025 

# Modelling sump pump energy usage patterns using artificial intelligence

**Joshua B. Ancheta[1,2,3], Aaron Don M. Africa[1], Luigi Carlo M. De Jesus[1,3], Stanley Glenn E. Brucal[1,3]**
[1]Department of Electronics, Computer, and Electrical Engineering, De La Salle University, Manila, Philippines
[2]College of Engineering and Architecture, Lyceum of the Philippines-Cavite, Cavite, Philippines
[3]School of Engineering, Asia Pacific College, Makati City, Philippines

**Article Info**

***Article history:***

Received Dec 2, 2025
Revised Sep 13, 2026
Accepted Oct 4, 2026

***Keywords:***

Energy forecasting
Long short-term memory
Machine learning
Random forest
Smart buildings
Sump pumps

**ABSTRACT**

Water typically seeps into the basements of buildings that are close to lakes or situated in low-lying areas, and hence these places require continuous pumping to avoid flooding. Such situations very often lead to high energy consumption which is not well-understood. This research models and analyzes the energy usage patterns of sump pumps installed at Basement 3 of Asia Pacific College (APC) through artificial intelligence techniques. High-resolution (5-minute) and daily aggregated datasets were recorded by SCT013 current sensors and then were converted to energy consumption values. Four prediction models—linear regression, random forest, support vector machine (SVM), and long short-term memory (LSTM)— were trained and tested using root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination ($R^2$) metrics. For the 5-minute dataset, random forest achieved RMSE=20.78, MAE=2.05, and $R^2$=0.226, whereas LSTM obtained RMSE=18.22, MAE=1.56, and negative $R^2$. For the daily dataset, random forest achieved RMSE=299.43, MAE=222.98, $R^2$=0.091, while LSTM reached RMSE=289.80, MAE=208.04, and $R^2$=0.014. The models indicate strong diurnal patterns and show the effect of temporal resolution on predictive performance. All reported LSTM results correspond to direct one-step-ahead forecasting and therefore characterize short-horizon predictive accuracy rather than multi-step deployment performance. The developed framework serves as a steppingstone towards future intelligent pump control and energy-efficient operation.

*This is an open access article under the CC BY-SA license.*

***Corresponding Author:***

Aaron Don M. Africa
Department of Electronics, Computer, and Electrical Engineering, De La Salle University
Manila, Philippines
Email: aaron.africa@dlsu.edu.ph

## 1. INTRODUCTION

The energy usage of building amenities has been a major topic of debate over the last few years when it comes to sustainability and resilience of the infrastructure. Among the various energy-consuming subsystems, pumping systems-in particular, those operating in basements that are subject to groundwater infiltration-get into a somewhat different position as their operation depends on environmental conditions rather than being scheduled routinely [1], [2]. In a lot of buildings, sump pumps are used which automatically turn on in situations when water levels go up and this frequently leads to the occurrence of irregular energy patterns that are still a mystery. This problem is getting bigger with the rise of the number of houses that are built nearby water bodies and, thus, the ingress of groundwater is becoming a common cause of the pumping mechanisms' malfunction.

Asia Pacific College (APC) basement is always under the threat of groundwater infiltration since the school is located close to a natural lake and urban stormwater channels. Consequently, Basement 3's sump pumps are in constant operation as water accumulation in the facility is prevented. Even though the pumps help in avoiding flood, the institution is unaware of any formal documentation that records the time when the pumps run at full capacity, which pump loads change the most, or how energy consumption varies across the seasons. Such a situation arises when the building systems are continuously running for safety reasons rather than efficiency, which, in the long run, results in energy wastage and unmonitored mechanical strain of the systems [3], [4]. Without a system in place to analyse the situation, it is a mystery as to how often the pumps are used and hence it is very difficult to find ways of improvement.

Existing literature on building systems sheds light on the significant benefits of understanding load behavior through computational modeling techniques that can unfold hidden temporal structures and operational irregularities [5]–[9]. These publications demonstrate that predictive modeling frameworks are capable of revealing the patterns which cannot be detected in raw logs. Moreover, several researchers emphasize that while dealing with pumping systems it is equally important to comprehend the diurnal cycles, equipment aging behavior, water-inflow patterns, and environmental factors [10], [11]. Despite the growing body of work on intelligent building energy forecasting and pump-related energy analysis, two important gaps remain in the literature. First, most existing studies rely on hourly, daily, or otherwise aggregated datasets, whereas high-frequency sump-pump energy datasets capable of capturing short pumping cycles, rapid switching events, and transient surges are rarely documented, particularly for basement drainage applications. Second, prior studies commonly evaluate model performance at only one temporal scale, making it difficult to determine how predictive behavior changes when the same physical system is represented at both fine-grained and aggregated resolutions. In response to these gaps, this study develops a real-world sump-pump energy dataset from Basement 3 of APC using 5-minute measurements collected from July 19, 2023 to July 31, 2025, and compares forecasting performance on both the original high-frequency series and its daily aggregated counterpart. This design directly addresses the limited availability of high-frequency sump-pump datasets and the lack of comparative model evaluation across multiple temporal resolutions. Unlike scheduled loads such as HVAC or lighting, sump-pump operation is event-driven and environmentally triggered: the pumps activate in response to groundwater infiltration rather than occupancy or time-of-day programming, producing a highly intermittent, zero-inflated signal punctuated by abrupt high-amplitude surges. This behavior makes sump-pump energy prediction a distinct technical barrier relative to conventional building-energy forecasting, because most variance is concentrated in rare transient events that aggregated datasets tend to smooth out. Consequently, this study advances machine-learning-based building-energy analytics by contributing a rarely available high-frequency sump-pump benchmark and a controlled cross-resolution evaluation that quantifies how the same physical system is represented at fine-grained and aggregated scales.

The research deciphers the setup of the sump pump at APC by means of a structured modeling approach which is mainly concentrated on grasping the energy aspects at various temporal scales. The high-resolution data corresponding to 5-minute intervals enable the identification of activity bursts, operational anomalies, and fluctuations within a single day. On the other hand, the daily aggregates depict the broader seasonal tendencies as well as the longer duration shifts influenced by the factors that come from outside such as rainfall or water-table variations [12]–[15]. The creation of two temporal views helps in revealing the relationships that otherwise may be overlooked in a single-resolution dataset.

The paper justifies the choice of four predictive modeling techniques, namely, linear regression, random forest, support vector machine (SVM), and long short-term memory (LSTM), which come from different methodological perspectives and aim at revealing the consumption data patterns in which irregularities are present [16]–[18]. Among them, linear regression represents a basic simple model; random forest detects non-linear features; SVM learns multifaceted boundaries in the noisy input space; and LSTM deals with sequential information. Examining these models simultaneously points to the answer to which approach is the most stable one under the practical and performance limitations imposed by the quality of the data and the pump behavior at APC. This research does not aim at providing a fully automatic control system as its main goal is to help create a solid analytical basis that leads to the clarification of the real energy patterns of the sump pumps.

## 2. METHOD

### 2.1. Data acquisition

Energy data were collected from the sump-pump installation at APC Basement 3. The monitored system consisted of four (4) units of 5-hp, 3-phase submersible pumps, with two units located on the Humabon side and two units on the SLEX side. The operational dataset used in this study covers the period from July 19, 2023 to July 31, 2025, providing a long duration record suitable for analyzing both regular pump activity and periods of elevated energy demand.

Current measurements were acquired using the YHDC SCT-013-000 split-core AC current transformer, rated for 0-100 A AC with a 100 A:50 mA output ratio and an approximate 1:2000 turns ratio. The sensor typically operates within the 50-60 Hz frequency range and is commonly reported to provide approximately ±1% accuracy under Grade B conditions, with typical linearity ranging from ±1% to ±3% depending on the specific unit and installation condition. Because the monitored equipment consists of high-power sump pumps, the selected sensor was considered appropriate for the intended monitoring application.

In the sensing circuit, a 33 Ω burden resistor was used for the SCT-013-000 on a 5 V measurement system. This configuration is suitable because it keeps the sensor output within a safe and measurable voltage range for analog-to-digital conversion while helping prevent ADC saturation under high-current conditions. The selected burden resistor therefore supported stable signal acquisition for the monitored pump loads.

To improve data reliability, the sensing setup was calibrated using a 1000 W heater as a reference load. During calibration, the measured response of the SCT-013-000-based circuit was compared against the expected electrical behavior of the heater load, and the resulting correction factor was applied to improve agreement between observed and expected values. This procedure helped validate the combined performance of the current transformer, burden resistor, and analog acquisition stage prior to long-term deployment. Since the monitored pumps operate within the practical high-current range of the selected sensor, the instrumentation setup was considered adequate for sustained monitoring. Calibration was performed against a resistive reference load and cross-checked across the operating current range spanned by the monitored pumps (from near-idle draw up to full three-phase load), rather than at a single operating point, so that the derived correction factor remained valid over the range encountered in deployment. Uncertainty was also propagated from the sensing stage to the reported energy values: the transformer and acquisition chain contribute an approximate plus-or-minus 1 to 3% measurement tolerance, which propagates to a comparable relative bound on the per-interval and daily energy estimates. This bound is small relative to the pump-activation dynamics being modeled and does not affect the comparative ranking of the forecasting models.

The raw acquisition stream was recorded at 5-minute intervals in order to preserve short pumping cycles, rapid switching events, and abrupt increases in electrical demand associated with basement water inflow conditions. From this source record, two datasets were constructed. The first retained the original 5-minute temporal resolution for high-frequency forecasting and short-interval pattern analysis. The second aggregated the same records into daily totals to support lower-frequency trend analysis and model comparison across temporal resolutions. This dual-resolution design is central to the study because it allows direct evaluation of how predictive models behave when the same physical system is represented in both fine-grained and aggregated forms [19]–[21]. No substantial sensor drift was observed during the monitoring period; however, periodic recalibration remains recommended for long-term deployment.

### 2.2. Data preprocessing

The unprocessed data logs had some missing entries, sudden dips, and inconsistent timestamps which needed rectification before modeling. To keep the chronological order intact, all the timestamps were standardized to one time zone reference. The missing short intervals were taken care of by linear interpolation, and the missing parts that were more than a few minutes were dealt with using conservative forward-fill methods that keep the structure of the surrounding data intact [22]. These changes guaranteed that the depiction of the pumps' performance was in a continuous and dependable sequence.

Besides that, the sensor values were improved by applying a Savitzky-Golay filter which lessens the noise part of the signal while the total shape of the signal remains the same. It was very important because the pumps in the sumps generally have very abrupt changes when they are turned on or off, and the filter helped in making sure that the modeling approaches reflect real pump operations and not just electrical noise [23]. Upon completion of the smoothing and cleaning processes, the current measurements were transformed into energy units through the use of calibration factors derived from the reference clamp-meter readings. Other time-related variables such as hour-of-day, day-of-week, rolling averages, and lagged variables were also added. These variables have shown great potential in revealing the deeply rooted structured dependencies in energy systems where repetitive cycles are the norm [24], [25].

### 2.3. Model development

The performance of the models was compared by developing four different predictive models: linear regression, random forest, SVM, and LSTM. Linear regression is the simplest one and is used to set up a benchmark which reflects the linearly dependent performance. With the help of ensemble decision structures, random forest allows the model to capture abnormal non-linear behavior, thus it is very effective for noisy energy data. The SVM uses flexible mathematical boundaries that try to separate the most intricate relationships between features and target variables, especially when the patterns are faint and interwoven [26], [27].

The LSTM has memory feature through sequential processing, and this feature is very useful in cases where the past significantly influences the future.

Every model was subjected to hyperparameter search through grid-based tuning along with cross-validation. The evaluation stage referred to the same set of features and the same splits of data for the different models to make the comparison between them fairer. This controlled setting permitted the research to pinpoint which model was most responsive under the limitations of APC's dataset and which model was less capable of handling the temporal variability or small sample size. The integrated modeling approach did not aim to give a universal forecasting solution but to uncover structural insights into the behavior of the sump pumps. To support reproducibility, hyperparameters were selected by grid search over documented ranges with a fixed random seed of 42. Random forest was tuned over number of estimators (40 to 200) and maximum depth (8 to unrestricted); the SVM over the regularization parameter C (0.1, 1, and 10) with a radial-basis kernel and scaled gamma; and the LSTM over hidden units (32 and 64) with the Adam optimizer at a learning rate of 0.001. The LSTM sliding window was set to 12 steps, corresponding to one hour of 5-minute observations, which was long enough to capture short activation cycles while limiting sequence length on the intermittent signal. Model selection used blocked time-series cross-validation with expanding training windows so that every validation fold contained only observations chronologically later than its training data.

For the LSTM implementation, the forecasting task was formulated as a direct one-step-ahead sequence prediction problem. A fixed-length sliding window of past observations was used as input, and the model predicted the next value in the sequence. This design was selected to provide a controlled and comparable evaluation setting across models while preserving chronological order in the time-series data. Compared with recursive rolling prediction, the direct one-step-ahead approach avoids the accumulation of prediction error across multiple future steps and is therefore more appropriate for analyzing the short-horizon predictive capability of the models under the present dataset conditions. Since the study primarily aims to characterize sump-pump energy patterns and compare model behavior across temporal resolutions, this setup was considered more suitable than autoregressive multi-step inference.

### 2.4. Evaluation metrics

Model performance was evaluated using root mean square error (RMSE), mean absolute error (MAE), and coefficient of determination ($R^2$). RMSE was used to emphasize larger prediction errors, particularly during abrupt increases in sump-pump energy consumption. MAE was used to measure the average magnitude of prediction error in a more uniform manner. $R^2$ was used to determine how well each model explained the variance of the observed energy-consumption series relative to a mean-value baseline [28]–[30]. For all models, these metrics were computed on the chronologically held-out test set in order to preserve temporal order and ensure fair comparison across methods.

For the LSTM model, evaluation was performed using a direct one-step-ahead sliding-window forecasting approach. Each test prediction was generated from a fixed sequence of previously observed values from the dataset, and the model's own predictions were not recursively reused as inputs for subsequent forecasting steps. Thus, the reported LSTM results reflect direct sequence-based regression rather than teacher forcing during inference or rolling autoregressive prediction. This evaluation strategy was selected because it provides a stable and interpretable basis for comparing LSTM against the non-sequential machine learning models used in the study, while avoiding artificial error accumulation that may arise in recursive multi-step forecasting.

The occurrence of a negative $R^2$ value for the LSTM on the high-frequency dataset indicates that, although the model achieved relatively low RMSE and MAE, it explained the variance of the observed series less effectively than a simple baseline prediction based on the mean of the test data. This can occur when predictions remain numerically close to many target values but fail to reproduce the timing and magnitude of high-variance events, such as abrupt surges in pump activity. Accordingly, the negative $R^2$ suggests that the LSTM was able to approximate short-interval values but did not fully capture the broader variability structure of the sump-pump energy series.

### 2.5. Experimental setup

All the work was done in a controlled workstation environment. The processing and modeling were implemented by Python-based libraries. The working environment was always performing at its best regardless of how complicated the model was, and it was also allowing the reproduction of each step of the experiment. The datasets were split up chronologically by 80/20. Thus, the readings' temporal progression was preserved. This stopped the information from leaking, which very often leads to evaluation results being distorted when forecasting time series [31]. The experimental design was mainly about the real deployment conditions, where the future values have to be predicted from the past data always and no retrospective adjustment is allowed. To prevent data leakage, all preprocessing operations that learn parameters from the

data, including gap interpolation, Savitzky-Golay smoothing, feature scaling, and the construction of lag and rolling features, were fitted on the training partition only and then applied unchanged to the held-out test partition; the same discipline was preserved within each cross-validation fold, and the random seed was fixed at 42 across all experiments.

## 3. RESULTS AND DISCUSSION

### 3.1. Diurnal energy consumption patterns

Figure 1 illustrates the diurnal energy behavior of the monitored sump pumps across different phases. The data show that pump activity is unevenly distributed throughout the day, with certain periods exhibiting more frequent and pronounced energy-use events than others. This indicates that the sump-pump system does not operate under a uniform constant load, but instead responds dynamically to varying inflow conditions and drainage demands within the basement environment [32].

The use of a 5-minute monitoring interval made it possible to capture short-duration fluctuations, abrupt activation bursts, and irregular demand changes that would likely be obscured in hourly or daily aggregates. These short-interval variations are important because they reflect the actual operating behavior of the pumps under changing hydraulic conditions. Consequently, the diurnal patterns observed in Figure 1 reinforce the value of high-frequency monitoring for identifying transient load behavior, detecting operational irregularities, and supporting more reliable predictive modeling of sump-pump energy use.

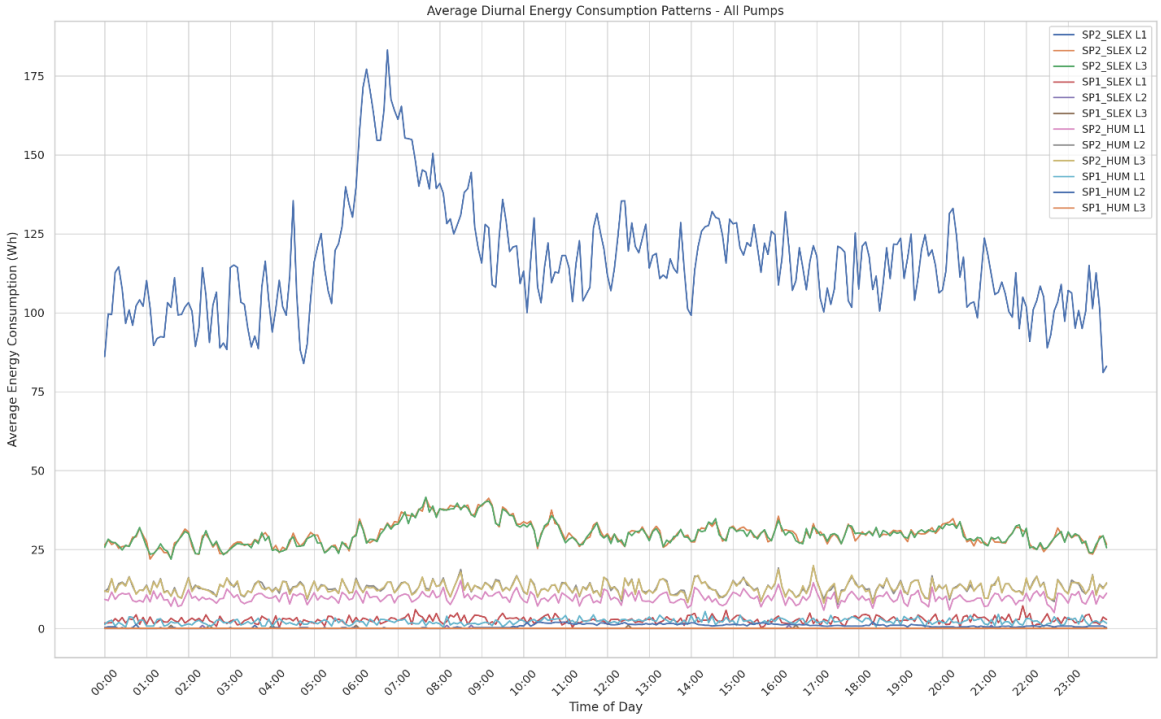



Figure 1. Diurnal energy consumption pattern of sump pumps

### 3.2. Daily trends and variability

Figure 2 presents the daily energy consumption series together with a 7-day rolling mean and standard deviation. The daily aggregated record reveals substantial variability over the monitoring period, with consumption ranging from near-zero values during low-activity periods to pronounced peaks during high-demand intervals [32]–[34]. The monitored record spans July 19, 2023 to July 31, 2025, and the highest daily consumption in the analyzed series occurred on March 17, 2025, reaching approximately 237.285 kWh. The overall pattern suggests relatively moderate daily consumption during much of 2023 and 2024, followed by a marked increase in early 2025.

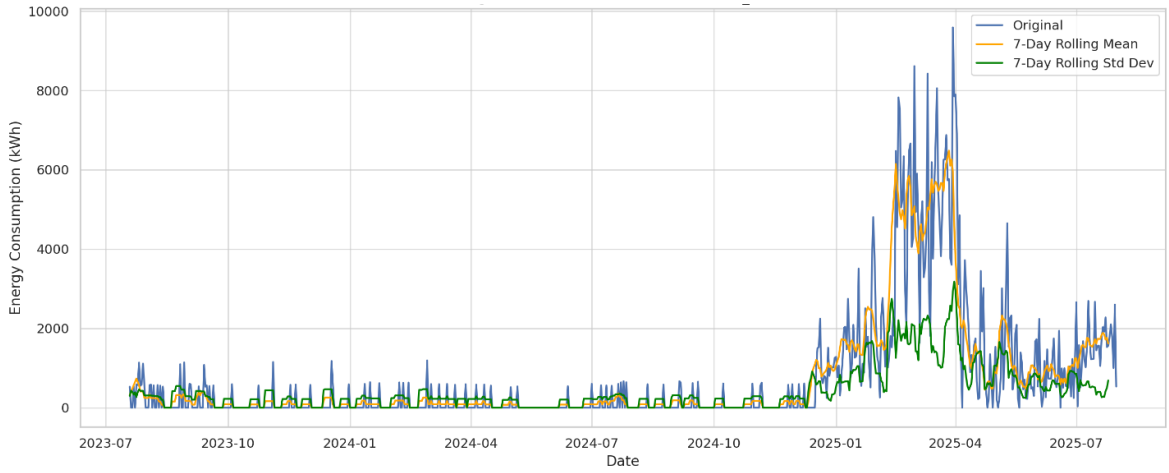



Figure 2. Daily energy consumption trend with 7-day rolling mean and standard deviation

At the seasonal level, the dataset shows stronger monthly variation than weekly seasonality. Figure 3 further compares daily sump-pump energy consumption with Makati daily rainfall. Daily energy use does not exhibit a clear or consistent weekday-weekend pattern, indicating that weekly operational scheduling is not the dominant driver of sump-pump activity. In contrast, monthly averages increase substantially from late 2024 into early 2025. The mean daily energy consumption rose from approximately 41.93 kWh in November 2024 to 60.27 kWh in December 2024, 89.73 kWh in January 2025, 136.08 kWh in February 2025, and 197.60 kWh in March 2025, before decreasing again in the succeeding months. This pattern suggests that the observed variability is more consistent with longer-duration environmental or hydraulic influences than with recurring weekly operational routines. To examine whether rainfall contributed to these fluctuations, the daily sump-pump energy series was compared with historical daily rainfall data for Makati, Philippines. The comparison showed very weak correlation values for both same-day rainfall (lag 0, r=-0.041) and one-day delayed rainfall response (lag 1, r=-0.028). These results indicate that rainfall, when represented as daily precipitation totals, did not exhibit a meaningful linear relationship with the observed daily sump-pump energy consumption during the monitoring period. Therefore, the early-2025 increase in pump energy demand cannot be directly attributed to rainfall based on the present correlation analysis alone.

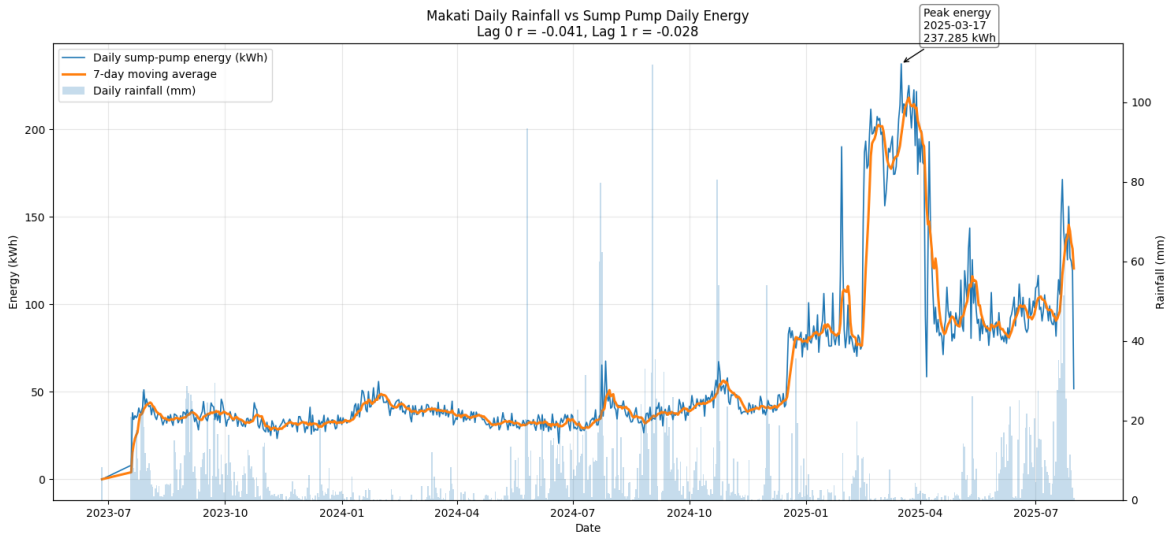



Figure 3. Comparison of Makati daily rainfall and sump-pump daily energy consumption

Although rainfall remains a plausible hydrological influence, the weak correlation suggests that other factors may have contributed more strongly to the observed consumption patterns. These may include groundwater accumulation, delayed subsurface seepage, drainage-system behavior, pump scheduling, maintenance condition, or site-specific hydraulic factors not captured by daily rainfall totals alone. Accordingly, the early-2025 consumption surge should be interpreted as a real operational anomaly in the dataset, but one whose cause remains unresolved under the currently available variables. The negligible rainfall correlations further suggest that daily precipitation alone is insufficient as a predictor of sump-pump energy demand in the studied basement environment. Future analyses may therefore benefit from incorporating additional explanatory variables such as water-level measurements, groundwater indicators, hourly rainfall, soil saturation proxies, and maintenance or operational logs, which may better capture the physical processes driving sump-pump activation.

### 3.3. Model evaluation on high-frequency data

Figure 4 demonstrates the effectiveness of the models on the high-frequency dataset, and the numerical results are summarized in Table 1. Among the classical models, random forest achieved the highest $R^2$ value of 0.226, indicating the strongest ability to capture nonlinear relationships in the 5-minute energy data. Linear regression and SVM obtained similar $R^2$ values of 0.169, suggesting only limited capability in representing more complex short-interval fluctuations. The LSTM obtained the lowest RMSE and MAE, which indicates favorable short-horizon point-wise prediction accuracy. However, its negative $R^2$ value (-0.010) shows that it did not explain the overall variance of the signal as effectively as the other models [35]–[37]. This divergence between error and variance metrics reflects how each model treats the signal. Random forest recursively partitions the feature space and fits nonlinear conditional means, so it reproduces a larger share of the observed variance and attains the highest R-squared. The LSTM instead minimizes point-wise error and tends to predict values close to the recent local level, which yields the lowest RMSE and MAE but systematically underestimates the sparse, high-amplitude activation surges that dominate the variance term of R-squared; the result is strong point accuracy alongside weak variance explanation. A blocked time-series cross-validation with expanding windows reproduced this ordering, but the R-squared values varied substantially from fold to fold, confirming that variance explanation on this intermittent signal is limited and unstable and that the results should be interpreted as short-horizon estimation rather than deployment-ready forecasting.

This result should be interpreted in light of the adopted evaluation procedure. The LSTM was tested using a direct one-step-ahead sliding-window approach, where each prediction was generated from a fixed sequence of observed past values rather than through recursive autoregressive forecasting. Under this setup, the model can still achieve relatively low average prediction error while failing to reproduce the full amplitude and timing of abrupt energy fluctuations. Therefore, the LSTM's performance suggests that direct one-step sequence learning is useful for short-interval estimation, but not yet sufficient for variance-sensitive modeling of the sump-pump system under the current feature set and architecture.

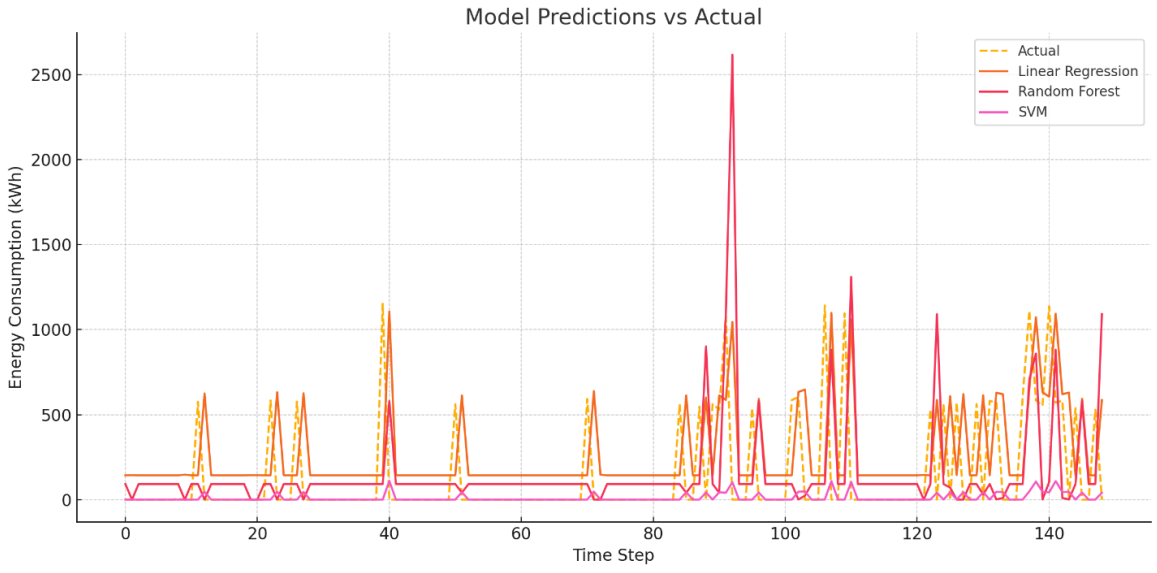



Figure 4. Actual vs predicted energy values using machine learning models (5-minute dataset)

Table 1. Performance metrics for classical and LSTM models (5-minute dataset)

| Model | RMSE | MAE | $R^2$ |
|---|---|---|---|
| Linear regression | 21.598 | 2.301 | 0.169 |
| Random forest | 20.781 | 2.048 | 0.226 |
| SVM | 21.605 | 2.263 | 0.169 |
| LSTM | 18.217 | 1.564 | -0.01 |

### 3.4. Model evaluation on daily data

Information on daily levels in Table 2 shows lower capacities of all methods due to fewer cases and greater differences between days. The highest $R^2$ of 0.091 was scored by the random forest while the lowest RMSE and MAE were LSTM's achievement. Although the error values are lowered, the low $R^2$ that is found in all models is an indication of the fact that other contextual variables such as rainfall, pump maintenance logs, or groundwater sensors may have considerable potential for improving the prediction accuracy [38]–[40]. The daily dataset is devoid of many brief-interval signals that are available in high-frequency readings and this most likely leads to weaker model representation of total daily variance.

Table 2. Performance metrics for classical and LSTM models (daily dataset)

| Model | RMSE | MAE | $R^2$ |
|---|---|---|---|
| Linear regression | 308.359 | 235.507 | 0.045 |
| Random forest | 299.426 | 222.982 | 0.091 |
| SVM | 312.721 | 236.105 | 0.030 |
| LSTM | 289.800 | 208.042 | 0.014 |

### 3.5. Long short-term memory performance and short-interval forecasting potential

The LSTM model's favorable RMSE and MAE on the 5-minute dataset indicate that it was able to estimate short-interval values with relatively low average error. However, its negative $R^2$ value shows that the model did not adequately represent the broader variance structure of the observed energy-consumption series, particularly during abrupt inflow-related surges and other high-amplitude events [41]–[45]. This means that while the model could follow the general magnitude of many observations, it was less successful in capturing the timing and intensity of strong fluctuations. This outcome is also consistent with the forecasting design adopted in the study. The LSTM was evaluated using a direct one-step-ahead sliding-window strategy, in which predictions were generated from observed historical windows and were not recursively reused for future inputs. This approach was intentionally selected because it provides a controlled basis for comparison with the other models and avoids recursive error propagation. However, it also means that the present LSTM experiment primarily evaluates short-horizon predictive capability, rather than long-horizon autonomous forecasting performance. The results therefore suggest that LSTM has promising potential for real-time monitoring and short-step estimation of sump-pump energy use, but that additional improvements are needed before it can serve as a robust stand-alone forecasting model. Such improvements may include longer input windows, richer contextual variables, deeper architectures, and future testing under multi-step autoregressive forecasting conditions to assess performance in more operationally realistic deployment scenarios.

## 4. CONCLUSION

This research focused on the energy consumption of sump pumps operating in Basement 3 of APC using two temporal resolutions: 5-minute readings and daily aggregates. The analysis revealed pronounced diurnal cycles and substantial variation in energy demand over time, reflecting the influence of environmental and operational factors on pump activity. Among the evaluated models, random forest showed the strongest ability to represent overall variance across both temporal resolutions. In contrast, the LSTM achieved the lowest point-wise errors in some cases but showed weaker variance representation, particularly on the high-frequency dataset where the $R^2$ value became negative. These results indicate that different forecasting methods offer different strengths depending on the temporal resolution and the statistical characteristics of the data. The LSTM in this study was evaluated using a direct one-step-ahead sliding-window forecasting approach, which was chosen to provide a controlled and interpretable comparison with the other machine learning models while preserving chronological structure in the data. This setup was appropriate for assessing short-horizon predictive behavior, but it does not yet represent recursive multi-step forecasting conditions. As such, the reported LSTM results should be interpreted as evidence of short-interval sequence modeling capability rather than full long-horizon deployment readiness. The results also showed

that the early-2025 increase in daily pump energy consumption was a clear feature of the monitored record. However, comparison with Makati daily rainfall revealed negligible same-day and one-day lag correlations, indicating that daily rainfall alone was not a strong explanatory variable for the observed energy-use variability. This suggests that sump-pump demand in the studied building is influenced by more complex hydraulic and operational factors than precipitation totals alone can capture. Future work should incorporate additional environmental and operational variables, including rainfall, water-level measurements, groundwater-related indicators, temperature, humidity, and maintenance logs, to better explain the drivers of pump activity. It should also extend the present direct forecasting setup toward multi-step and autoregressive forecasting, explore automated control or decision-support strategies, and validate the methodology in other buildings with different hydraulic and structural conditions. The scope of these conclusions is bounded by the evaluation design. The study analyzes a single institutional building, so generalization to other hydraulic or infrastructure settings remains to be established; the forecasting evaluation is one-step-ahead and therefore should not be read as evidence of multi-step or autonomous deployment readiness; and the omission of explanatory variables such as water level, groundwater indicators, and sub-daily rainfall constrains causal interpretation, since daily rainfall alone did not explain the observed consumption surge. Within these limits, the practical engineering significance of the work is that high-frequency monitoring exposes transient activation behavior that daily aggregation conceals, providing an analytical basis for intelligent pump management. Future work should incorporate explainable-AI techniques such as feature-attribution analysis, multi-source environmental sensing, and multi-step forecasting evaluated under realistic operational deployment conditions.

## ACKNOWLEDGMENTS

The authors would like to express their gratitude toward God Almighty, who enabled the completion of this research. Without His grace, this project would never have been completed. Our appreciation also goes to the APC AI and Robotics Innovation Hub (AIRHUB), which provided resources for the study. Special thanks go to De La Salle University (DLSU) for its immeasurable contributions, with which a partnership has been formed in fostering an environment for innovation and excellence. These institutions have supported the authors in every aspect of conducting this research from start to finish, for which they are very grateful.

## FUNDING INFORMATION

Authors state no funding involved.

## AUTHOR CONTRIBUTIONS STATEMENT

This journal uses the Contributor Roles Taxonomy (CRediT) to recognize individual author contributions, reduce authorship disputes, and facilitate collaboration.

| Name of Author | C | M | So | Va | Fo | I | R | D | O | E | Vi | Su | P | Fu |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Joshua B. Ancheta | ✓ | ✓ | ✓ | | ✓ | ✓ | ✓ | ✓ | ✓ | | ✓ | | ✓ | |
| Aaron Don M. Africa | ✓ | ✓ | | ✓ | ✓ | ✓ | | | | ✓ | | ✓ | | ✓ |
| Luigi Carlo M. De Jesus | ✓ | | ✓ | | | ✓ | ✓ | | | ✓ | ✓ | ✓ | ✓ | |
| Stanley Glenn E. Brucal | ✓ | | | ✓ | | | | | | ✓ | | ✓ | | |

C : **C**onceptualization
M : **M**ethodology
So : **So**ftware
Va : **Va**lidation
Fo : **Fo**rmal analysis
I : **I**nvestigation
R : **R**esources
D : **D**ata Curation
O : Writing - **O**riginal Draft
E : Writing - Review & **E**diting
Vi : **Vi**sualization
Su : **Su**pervision
P : **P**roject administration
Fu : **Fu**nding acquisition

## CONFLICT OF INTEREST STATEMENT

Authors state no conflict of interest.

## DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author, A.D.M.A., upon reasonable request.

## REFERENCES


[1] A. Mosavi and A. Bahmani, "Energy consumption prediction using machine learning; a review," *Preprints*, Mar. 2019, doi: 10.20944/preprints201903.0131.v1.

[2] X. Gan, X. Gong, J. Pei, G. Pavesi, and S. Yuan, "Machine Learning Approaches for Enhancing Energy Efficiency and Stability in Parallel Pumping Systems," *Water Resources Management*, vol. 39, no. 4, pp. 1719–1746, Nov. 2024, doi: 10.1007/s11269-024-04043-z.

[3] R. Jin *et al.*, "Artificial Intelligence Enabled Energy-Saving Drive Unit With Speed and Displacement Variable Pumps for Electro-Hydraulic Systems," *IEEE Transactions on Automation Science and Engineering*, vol. 21, no. 3, pp. 3193–3204, May 2023, doi: 10.1109/TASE.2023.3276766.

[4] T. K. Wijaya, Sholahudin, M. I. Alhamid, K. Saito, and N. Nasruddin, "Dynamic optimization of chilled water pump operation to reduce HVAC energy consumption," *Thermal Science and Engineering Progress*, vol. 36, p. 101512, Oct. 2022, doi: 10.1016/j.tsep.2022.101512.

[5] Y. Guo, Y. Liu, Y. Wang, C. Du, H. Li, and Z. Zhang, "A deep belief network-based energy consumption prediction model for water source heat pump system," *Applied Thermal Engineering*, vol. 255, p. 124000, Jul. 2024, doi: 10.1016/j.applthermaleng.2024.124000.

[6] N. G. Jolfaei, B. Jin, C. Chow, F. Bressan, and N. Gorjian, "An optimised energy saving model for pump scheduling in wastewater networks," in *Lecture Notes in Mechanical Engineering*, 2018, pp. 197–208, doi: 10.1007/978-3-319-95711-1_20.

[7] G. Seo, S. Yoon, M. Kim, C. Mun, and E. Hwang, "Deep Reinforcement Learning-Based Smart Joint Control Scheme for On/Off Pumping Systems in Wastewater Treatment Plants," *IEEE Access*, vol. 9, pp. 95360–95371, 2021, doi: 10.1109/ACCESS.2021.3094466.

[8] C. Cao, Q. Pan, S. Wang, and C. Wen, "Prediction of Energy Consumption Characteristics of Electro-Proportional Piston Pumps Based on GA-SVR," in *2023 9th International Conference on Fluid Power and Mechatronics (FPM)*, Lanzhou, China: IEEE, Aug. 2023, pp. 1–7, doi: 10.1109/FPM57590.2023.10565398.

[9] Y. Ruan *et al.*, "Predictive model of pump energy consumption based on image feature extraction and instance-based transfer learning," *Journal of Building Engineering*, vol. 79, p. 107770, Nov. 2023, doi: 10.1016/j.jobe.2023.107770.

[10] K. K. Ye, G. Demirezen, A. S. Fung, and E. Janssen, "The use of artificial neural networks (ANN) in the prediction of energy consumption of air-source heat pump in retrofit residential housing," in *IOP Conference Series: Earth and Environmental Science*, vol. 463, no. 1, p. 012165, Mar. 2020, doi: 10.1088/1755-1315/463/1/012165.

[11] F. S. Fadnes, R. Banihabib, and M. Assadi, "Using Artificial Neural Networks to Gather Intelligence on a Fully Operational Heat Pump System in an Existing Building Cluster," *Energies*, vol. 16, no. 9, pp. 1–33, May 2023, doi: 10.3390/en16093875.

[12] A. Domingues, A. Barbosa, and P. C. Bartolomeu, "Heat Pump Energy Consumption Optimization Using Machine Learning," in *IECON 2023- 49th Annual Conference of the IEEE Industrial Electronics Society*, Singapore: IEEE, Oct. 2023, pp. 1–7, doi: 10.1109/IECON51785.2023.10311953.

[13] A. Shaqour and A. Hagishima, "Systematic Review on Deep Reinforcement Learning-Based Energy Management for Different Building Types," *Energies*, vol. 15, no. 22, pp. 1–27, Nov. 2022, doi: 10.3390/en15228663.

[14] S. U. Khan, N. Khan, F. U. M. Ullah, M. J. Kim, M. Y. Lee, and S. W. Baik, "Towards intelligent building energy management: AI-based framework for power consumption and generation forecasting," *Energy and Buildings*, vol. 279, p. 112705, Jan. 2023, doi: 10.1016/j.enbuild.2022.112705.

[15] P. P. Phyo, Y. C. Byun, and N. Park, "Short-Term Energy Forecasting Using Machine-Learning-Based Ensemble Voting Regression," *Symmetry*, vol. 14, no. 1, p. 160, Jan. 2022, doi: 10.3390/sym14010160.

[16] S. A. Bagloee, M. Asadi, and M. Patriksson, "Minimization of water pumps' electricity usage: A hybrid approach of regression models with optimization," *Expert Systems with Applications*, vol. 107, pp. 222–242, Oct. 2018, doi: 10.1016/j.eswa.2018.04.027.

[17] C. Wang *et al.*, "Residential water and energy consumption prediction at hourly resolution based on a hybrid machine learning approach," *Water Research*, vol. 246, p. 120733, Nov. 2023, doi: 10.1016/j.watres.2023.120733.

[18] T. Ahmad and H. Chen, "Utility companies strategy for short-term energy demand forecasting using machine learning based models," *Sustainable Cities and Society*, vol. 39, pp. 401–417, May 2018, doi: 10.1016/j.scs.2018.03.002.

[19] E. Sarmas, E. Spiliotis, V. Marinakis, G. Tzanes, J. K. Kaldellis, and H. Doukas, "ML-based energy management of water pumping systems for the application of peak shaving in small-scale islands," *Sustainable Cities and Society*, vol. 82, p. 103873, Jul. 2022, doi: 10.1016/j.scs.2022.103873.

[20] Z. Eddaoudi, Z. Aarab, K. Boudmen, A. Elghazi, and M. D. Rahmani, "A Brief Review of Energy Consumption Forecasting Using Machine Learning Models," *Procedia Computer Science*, vol. 236, pp. 33–40, 2024, doi: 10.1016/j.procs.2024.05.001.

[21] C. Jamunadevi, N. Rajasekaran, S. Pandikumar, R. Raja, and K. P. N. Prasanth, "Energy Performance Forecasting in Buildings Using Machine Learning," in *2024 International Conference on Communication, Computing and Internet of Things (IC3IoT)*, Chennai, India: IEEE, Apr. 2024, pp. 1–6, doi: 10.1109/IC3IoT60841.2024.10550233.

[22] N. Kushwaha and A. A. Waoo, "Energy Consumption Prediction by using Machine Learning," *International Journal For Multidisciplinary Research*, vol. 5, no. 6, Dec. 2023, doi: 10.36948/ijfmr.2023.v05i06.9988.

[23] S. Seyedzadeh and F. P. Rahimian, "Machine learning for building energy forecasting," in *Green Energy and Technology*, 2021, pp. 41–76, doi: 10.1007/978-3-030-64751-3_4.

[24] M. Bahij, M. Cherkaoui, and M. Labbadi, "Energy Consumption Forecasting in Industrial Sector Using Machine Learning Approaches," in *Learning and Analytics in Intelligent Systems*, vol. 7, 2020, pp. 155–164, doi: 10.1007/978-3-030-36778-7_17.

[25] N. V. U. Reddy, Y. P. Rangaiah, A. Nagpal, J. Reddy, S. Ziara, and Y. K. Bargavi, "Transfer Learning for Energy Consumption Forecasting in Smart Buildings," in *2024 7th International Conference on Contemporary Computing and Informatics (IC3I)*, Greater Noida, India: IEEE, Sep. 2024, pp. 1549–1555, doi: 10.1109/IC3I61595.2024.10828603.

[26] V. D. Novikov and R. M. Khamitov, "Forecasting Consumer Activity Using Machine Learning Methods," *International Journal of Advanced Studies: Transport and Information Technologies*, vol. 14, no. 1, pp. 205–214, Mar. 2024, doi: 10.12731/2227-930x-2024-14-1-290.

[27] K. Danilov and S. Maltseva, "Energy consumption forecasting as a service," in *2021 IEEE 23rd Conference on Business Informatics (CBI)*, Bolzano, Italy: IEEE, Sep. 2021, pp. 167–172, doi: 10.1109/CBI52690.2021.10068.

[28] D. Koroteev, T. Koroteeva, and J. Huang, "Forecasting energy consumption in home energy management systems using machine learning method," *Construction and Architecture*, vol. 11, no. 2, p. 6, Jun. 2023, doi: 10.29039/2308-0191-2023-11-2-6-6.
[29] P. Jurek and A. Plichta, "Forecasting Energy Consumption in Energy Clusters using Machine Learning Methods," in *Proceedings - European Council for Modelling and Simulation, ECMS*, ECMS, Jun. 2024, pp. 500–506, doi: 10.7148/2024-0500.
[30] A. V. Alekseev, "Study of the forecasting problem of energy consumption of water pumping station," in *E3S Web of Conferences*, vol. 102, p. 03001, Jun. 2019, doi: 10.1051/e3sconf/201910203001.
[31] T. X. Dan and P. N. K. Phuc, "Application of Machine Learning in Forecasting Energy Usage of Building Design," in *2018 4th International Conference on Green Technology and Sustainable Development (GTSD)*, Ho Chi Minh City, Vietnam: IEEE, Nov. 2018, pp. 53–59, doi: 10.1109/GTSD.2018.8595595.
[32] F. El Husseini, H. N. Noura, O. Salman, and K. Chahine, "Machine Learning in Smart Buildings: A Review of Methods, Challenges, and Future Trends," *Applied Sciences*, vol. 15, no. 14, pp. 1–40, Jul. 2025, doi: 10.3390/app15147682.
[33] B. U. Islam, M. Rasheed, and S. F. Ahmed, "Review of Short-Term Load Forecasting for Smart Grids Using Deep Neural Networks and Metaheuristic Methods," *Mathematical Problems in Engineering*, pp. 1–14, Sep. 2022, doi: 10.1155/2022/4049685.
[34] H. Wang, Z. Lei, X. Zhang, B. Zhou, and J. Peng, "A review of deep learning for renewable energy forecasting," *Energy Conversion and Management*, vol. 198, p. 111799, Oct. 2019, doi: 10.1016/j.enconman.2019.111799.
[35] P. D. D. Phan, M. A. N. N. M. Anh, and B. H. Hu. Nguyen, "Using Linear Regression Analysis to Predict Energy Consumption." Jul. 03, 2024, doi: 10.21203/rs.3.rs-4590592/v1.
[36] R. Mohan, S. Devneni, S. Sumpreet, V. Mohan, and N. Pachauri, "Comparative Analysis of Machine Learning Algorithms for the Building Energy Prediction," in *2024 2nd International Conference on Device Intelligence, Computing and Communication Technologies (DICCT)*, Dehradun, India: IEEE, Mar. 2024, pp. 409–413, doi: 10.1109/DICCT61038.2024.10532823.
[37] Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, "Random Forest based hourly building energy prediction," *Energy and Buildings*, vol. 171, pp. 11–25, Jul. 2018, doi: 10.1016/j.enbuild.2018.04.008.
[38] Q. Yin, C. Han, A. Li, X. Liu, and Y. Liu, "A Review of Research on Building Energy Consumption Prediction Models Based on Artificial Neural Networks," *Sustainability*, vol. 16, no. 17, p. 7805, Sep. 2024, doi: 10.3390/su16177805.
[39] C. Marcy, T. Goforth, D. Nock, and M. Brown, "Comparison of temporal resolution selection approaches in energy systems models," *Energy*, vol. 251, pp. 1–13, Jul. 2022, doi: 10.1016/j.energy.2022.123969.
[40] K. Yu, "Adaptive Bi-Directional LSTM Short-Term Load Forecasting with Improved Attention Mechanisms," *Energies*, vol. 17, no. 15, p. 3709, Jul. 2024, doi: 10.3390/en17153709.
[41] H. Luo, Y. Zhang, X. Gao, Z. Liu, X. Meng, and X. Yang, "Multi-scale electricity consumption prediction model based on land use and interpretable machine learning: A case study of China," *Advances in Applied Energy*, vol. 16, p. 100197, Dec. 2024, doi: 10.1016/j.adapen.2024.100197.
[42] C. Kühnert, N. M. Gonuguntla, H. Krieg, D. Nowak, and J. A. Thomas, "Application of LSTM networks for water demand prediction in optimal pump control," *Water*, vol. 13, no. 5, p. 644, Feb. 2021, doi: 10.3390/w13050644.
[43] D. Chi, "Research on electricity consumption forecasting model based on wavelet transform and multi-layer LSTM model," *Energy Reports*, vol. 8, pp. 220–228, Jul. 2022, doi: 10.1016/j.egyr.2022.01.169.
[44] S. H. Godasiaei, O. A. Ejohwomu, H. Zhong, and D. Booker, "Interpretability in machine learning for IAQ and HVAC optimisation: A response to Oka et al," *Building and Environment*, vol. 284, p. 113494, Oct. 2025, doi: 10.1016/j.buildenv.2025.113494.
[45] R. T. Moyo, M.S. Thesis, "Artificial intelligence based solar/diesel hybrid water pumping system," Department of Mechanical Engineering, Durban University of Technology, Durban, South Africa, 2021.

## BIOGRAPHIES OF AUTHORS



**Joshua B. Ancheta** is professional electronics engineer for 15 years and 23 years academe experience. He obtained his masters of Engineering from Technological University of the Philippines, Manila and now he is studying for Doctor of Philosophy in Electronics Engineering at De La Salle University, Manila. Currently he is working as an engineering professor at Lyceum of the Philippines-Cavite, and Asia Pacific College. His research interests are electronics, robotics, mechatronics, and AI. He can be contacted at email: joshua_ancheta@dlsu.edu.ph.



**Aaron Don M. Africa** is a researcher in the field of communications network engineering. His expertise includes communications network engineering, control systems, expert systems, network design, and optimization. His work involves communication systems optimization for more efficient functionalities. This is for these systems to adapt effectively in the industry. He creates different simulation models to reproduce different scenarios in network design. He can be contacted at email: aaron.africa@dlsu.edu.ph.



**Luigi Carlo M. De Jesus** an electronics engineer with a master of Engineering in Computer Engineering Degree from Asia Pacific College and a bachelor's degree in Electronics Engineering from FEU – Institute of Technology. He is now studying for a Doctor of Philosophy at De La Salle University, Manila. He is the head of the Engineering and Science Laboratory Office at Asia Pacific College. His research interests are image processing, machine learning, IoT-based monitoring systems, and energy optimization. He can be contacted at email: luigid@apc.edu.ph; luigi_dejesus@dlsu.edu.ph.



**Stanley Glenn E. Brucal** professional electronics engineer with more than 19 years of experience in the academe, and an ASEAN Chartered Professional Engineer (ACPE), with master's degree in Electronics and Communications Engineering. He graduated Doctor of Philosophy in Electronics and Communications Engineering at De La Salle University, Manila. Currently, he is working for Asia Pacific College-School of Engineering as an Associate Professor and is serving as 2nd member of the Professional Regulations Commission Continuing Professional Development (CPD) Council of Electronics Engineering. His research areas include data and building power consumption optimization, and energy and indoor air quality monitoring and control systems. He can be contacted at email: stanley_brucal@dlsu.edu.ph.