v.7, n.1, 2025, DOI: 10.52664/rima.v7.n1.2025.e256

RIMA
Revista Interdisciplinar e do Meio Ambiente
Interdisciplinary and Environment Journal

A Review of Machine Learning Applications for Energy Consumption Forecasting in Schools

Una Revisión de las Aplicaciones del Aprendizaje Automático para la Previsión del Consumo Energético en las Escuelas

Thu Pham[1]

Abstract: As schools face growing energy demand under constrained budgets, accurate energy forecasting using Machine Learning has become crucial for improving efficiency and planning targeted energy management strategies. This review examines studies that apply ML techniques in forecasting school and campus energy demands. Based on the methodology of these works, a generalized forecasting framework is proposed, which detailedly outlines: data preprocessing, feature selection, model selection, and implementation of results to guide implementation. Across studies, historical load, weather conditions, occupancy and building attributes are among the most reliable predictors of energy demand. Advanced models such as hybrid LSTM architectures or ensemble approaches generally achieve higher accuracy but require a larger complete dataset, increased computational costs and intensive hyperparameter tuning which limits their feasibility in resource-limited school settings. Simpler and more interpretable alternatives such as MLR often offer sufficient accuracy for schools with limited data availability and resources. Future studies should focus on addressing existing gaps by ensuring transparency and consistency in data and methodological reporting.
Keywords: *Machine learning; School energy management; Load forecasting methodology; Review paper.*

Resumen: A medida que las escuelas se enfrentan a una creciente demanda energética con presupuestos limitados, la previsión energética precisa mediante el aprendizaje automático se ha convertido en un factor crucial para mejorar la eficiencia y planificar estrategias de gestión energética específicas. Esta revisión examina los estudios que aplican técnicas de aprendizaje automático en la previsión de la demanda energética de escuelas y campus. Basándose en la metodología de estos trabajos, se propone un marco de previsión generalizado que describe detalladamente: el preprocesamiento de datos, la selección de características, la selección de modelos y la aplicación de los resultados para orientar la implementación. En todos los estudios, la carga histórica, las condiciones meteorológicas, la ocupación y las características de los edificios se encuentran entre los predictores más fiables de la demanda energética. Los modelos avanzados, como las arquitecturas híbridas LSTM o los enfoques de conjunto, suelen alcanzar una mayor precisión, pero requieren un conjunto de datos completo más amplio, mayores costes computacionales y un ajuste intensivo de los hiperparámetros, lo que limita su viabilidad en entornos escolares con recursos limitados. Alternativas más sencillas e interpretables, como el MLR, suelen ofrecer una precisión suficiente para las escuelas con disponibilidad de datos y recursos limitados.

*Author for correspondence
Received for publication on 2025/08/12; approved on 2025/12/25.
[1] * Researcher at National Economics University, Hanoi, Vietnam; email: minhthupham2605@gmail.com
https://orcid.org/0000-0002-8304-1259.

Los estudios futuros deberían centrarse en abordar las lagunas existentes, garantizando la transparencia y la coherencia en los datos y la presentación de informes metodológicos.
Palabras clave: *Aprendizaje automático; Gestión energética escolar; Metodología de previsión de carga; Artículo de revisión.*

INTRODUCTION

Context and Motivation

Schools are among the most significant energy consumers in the education sectors, yet they often operate under strict budgetary and infrastructural constraints. Their energy consumption is majorly influenced by various factors, including academic schedules with breaks and holidays, varying occupancy levels, diverse facility usage patterns, and building characteristics, all of which are further influenced by geographic location and seasonal variations that drive heating and cooling needs [1]. These specific characteristics when combined with aging infrastructure, limited funding for upgrades, and the absence of intelligent energy management systems, often result in unnecessary energy waste and increased operational expenses [2].

This resource inefficiency places a heavy burden on schools' finances, which are already tight due to a lack of governmental funding. In fact, this burden is evident across numerous researches in different countries, with research by Bray et al. [3] showing that school and university buildings consume approximately 60% more energy than commercial office spaces for example. In Saudi Arabia for instance, public educational buildings alone are responsible for approximately 13% of national energy use, largely due to inefficient lighting and air conditioning systems, according to Alshibani (2020) [14]. According to the U.S. Department of Energy [5], K-12 school districts spend "nearly $8 billion annually on energy costs", which accounts for the second largest expense after personnel costs. The U.S. Environmental Protection Agency (EPA) states that nearly 25% of the energy consumed in American schools is wasted, an inefficiency that, if addressed, could save up to $20 billion over a ten-year period [4].

Beyond financial strains and increased operational costs, this inefficiency is detrimental to students' education and wellbeing as wasted spending diverts the already scarce resources due to limited governmental funding that public schools receive away from educational purposes such as instructional or facility qualities. Public schools in socioeconomically disadvantaged areas are particularly vulnerable as they are more likely to operate in older, inefficient buildings [6] and are less likely to have access to energy-efficient upgrades or advanced forecasting tools. Huang et al., in a study of 3,672 schools in

Ontario, found that energy consumption had the strongest negative correlation with students' learning ability [7], which underscores how inefficient energy use perpetuates educational inequity by disproportionately affecting low-income students.

This pressing issue underscores the urgent need for intelligent solutions that enable schools to better manage energy use and allocate resources more effectively. In recent years, machine learning (ML) models have emerged as powerful tools in energy consumption forecasting using historical energy data, weather information, occupancy rates, school schedules, and data that are specific to K-12 school settings. Schools could greatly improve their energy management as accurate forecasting allows them to to make informed decisions on future energy usage based on previous patterns, such as pre-schedule HVAC systems (e.g., heating or cooling the building before occupants arrive), avoid running systems during idle times such as holidays or events, reduce heating at unoccupied hours [9], and predict peak cost hour to shift usage to cheaper off-peak hours.

Literature review

Although public schools are critical infrastructure and major building energy consumers, there is a clear lack of a comprehensive synthesis of ML-based forecasting specifically in school environments. Most data-driven forecasting research has focused on commercial buildings or aggregated public building stocks rather than on K–12 or typical school facilities [7]. This is due to the intrinsic complexity of school buildings: they have highly variable daily and seasonal occupancy driven by class schedules, extracurricular events, and holiday calendars; they also often combine mixed uses (classrooms, gyms, cafeterias, auditoria); and they frequently operate on a strict budget [7].

A growing number of public schools and educational facilities around the world have begun incorporating ML models into their energy management strategies, with varying levels of forecasting accuracy and impact on efficiency. Models such as Multiple Linear Regression (MLR), Artificial Neural Networks (ANN), Support Vector Machines (SVM), Random Forest (RF), Multilayer Perceptron (MLP), and Long Short-Term Memory (LSTM) networks have been tested across diverse educational settings in different countries. For instance, in Croatia's Osijek-Baranja County, multiple linear regression (MLR) and artificial neural network (ANN) models were applied to primary and secondary school electricity data: ANN achieved $R^2 \approx 0.957$ in training (vs. 0.950 for MLR), with CV(RMSE) ≈ 19.8 % which shows high variance in energy consumption and potential inefficiencies in resource use for schools built in differing conditions. In Eastern Province, Saudi Arabia, a regression-based model using 350 school observations demonstrated prediction accuracy over 90 %. The study identified AC capacity and roof area as strongest predictors influencing consumption, suggesting how these models are able to pinpoint the cause being infrastructure design that drives inefficiency. Capozzoli et al. (2015) analyzed annual energy demand for heating across 80 Italian schools using Multiple Linear Regression (MLR) and Classification and Regression Trees (CART); both achieved strong accuracy ($R^2 \approx 0.85$; MAPE ≈ 15%) but CART was able to provide interpretable insights on key drivers of energy use such as floor area and occupancy [68]. Amber et al. (2017) also applied MLR to forecast daily energy use of a university over a five-year period, and the model achieved a RMSE value of around 12-13% [69]. Similarly, many studies utilized tree-based and ensemble techniques in school energy forecasting. Tariq et al. (2024) compared Decision Trees (DT), k-Nearest Neighbors (KNN), Gradient Boosting Regression (GBR), and Long Short-Term Memory (LSTM) networks for energy forecasting in 352 schools, and reported that DT yielded low training error (3.58%), while KNN overfitted, and GBR and LSTM performed better across wide data ranges; school size and AC capacity were the most influential variables [70]. Khaoula Elhabyb et al. (2024) [65] compared GBR, RF and LSTM models using data from three university buildings, and reported that GBR yielded the most accurate forecasts. Recently, more studies have chosen neural network techniques to capture non-linear relationships, such as Ganesh Doiphode and Najafi (2020b) who trained a Multi-Layer Perceptron on three years of monthly data of 25 K-12 schools in Florida [71]. Geraldi and Ghisi (2020) modeled electricity consumption for 90 public schools in Brazil using Bayesian Networks on three years of monthly billing data and survey inputs, and captured key insights such as the number of students being a more reliable predictor than floor-plan [72]. Hybrid models with optimization algorithms have also been commonly used, as Khan et al. (2025) compared ARIMA with a Quantum-Inspired Particle Swarm Optimization (QPSO)–tuned RNN on building-load data, and reported the latter achieve the lowest error (MAE = 15.2; RMSE = 22.8), showing the advantage in accuracy improvement that hyperparameter tuning brings [73]. Current reviews report several consistent limitations that exist within this field of ML applications for load forecasting. Multiple reviews point out the absence of datasets to the public [74], and call for public data to enable generalizability and reproduction. Zhang et al. (2021) also identified that most papers overly focus on the machine learning development while the data side (data resolution, training/testing data split, etc.) are insufficiently discussed [75]. Notably, most studies focus on residential building load forecasting [76], leaving educational buildings underresearched.

This systematic review aims to address the gap on Machine Learning applications in school energy forecasting specifically by synthesizing the methodology of forecasting across different models in different countries; comparing the most commonly employed models in terms of dataset characteristics, input features, and forecasting accuracy; reviewing representative case studies to identify regional patterns; analyzing dataset characteristics and performance results; as well as identifying challenges and opportunities in widespread implementation. The overall purpose of this review is to inform readers of the practicality and effectiveness of ML-based forecasting models, with the broader goal of encouraging the widespread integration of energy management systems in schools worldwide to reduce waste, redirect resources towards improving students' education, and move a step closer to achieving equity in educational opportunity. In addition, this survey paper also seeks to inform and provide accessible guidance by generalizing replicable methods and explaining the various complex concepts in a simplified way so that stakeholders, policymakers, and school administrators without much background on Machine Learning can more easily apply those understandings into the integration of energy management systems.

METHODOLOGY

Review scope and criteria

This systematic review focuses on high quality peer reviewed literature that apply machine learning techniques to forecast energy consumption in school or educational building contexts, including K–12 and higher education buildings where methods are transferable to K–12 settings.

Due to the rapidly changing nature of machine learning methods and data availability, the review only includes recent studies published between 2019 and 2025 to ensure currency. Quality, reliability and relevance to the central topic of school energy forecasting determine final inclusion.

Databases searched and search terms

Databases searched include Google Scholar and major academic publishers and bibliographic databases to maximize coverage of peer reviewed and conference literature: Scopus, MPDI, IEEE Xplore, Sage Journals, ScienceDirect (Elsevier), SpringerLink, JSTOR, and arXiv for preprints. Additionally, PubMed Central was used primarily to gather insights on the impacts of poor energy management on students' health and academic performance.

Example search terms used included combinations of the following key words and phrases

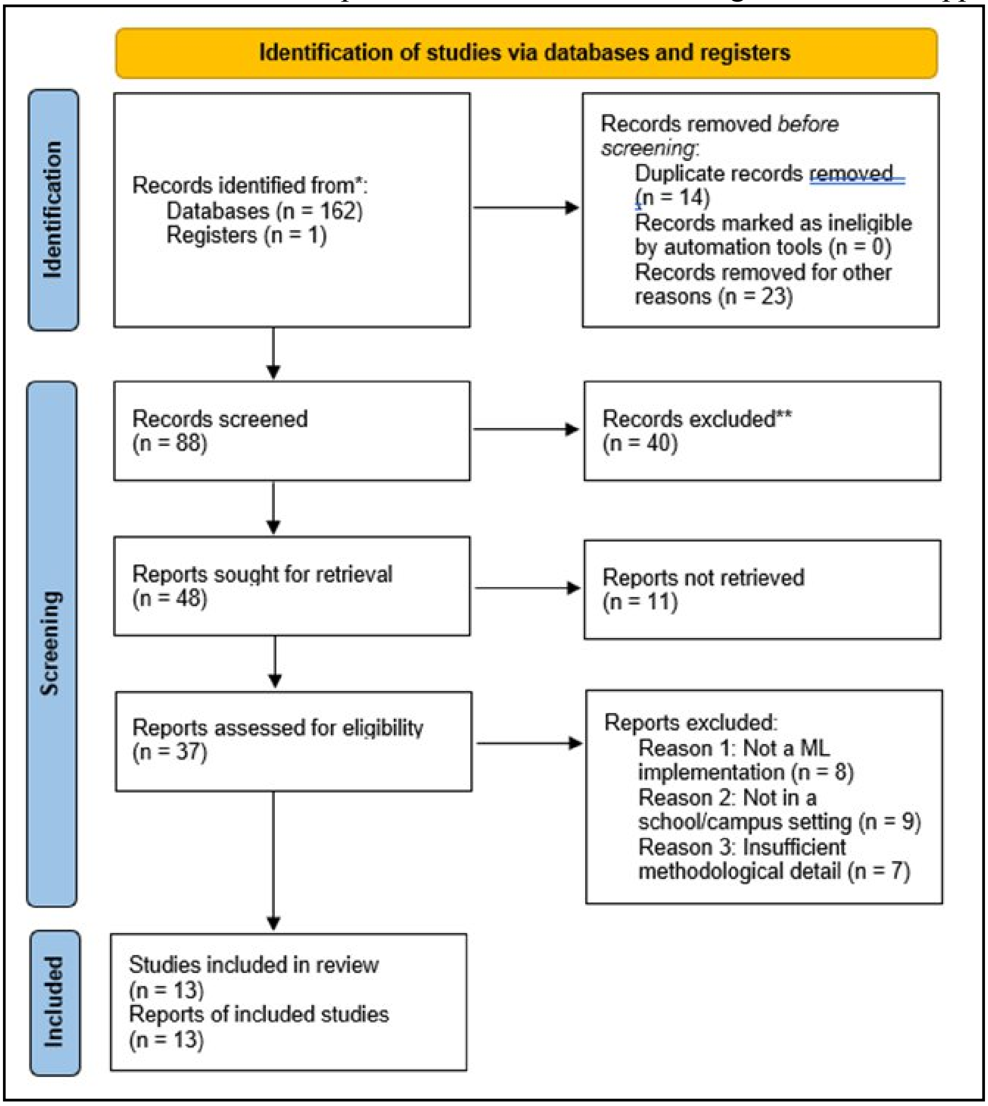
- "machine learning" AND "energy forecasting" AND (school OR "K-12" OR "educational building" OR "school building")
- "deep learning" AND "building energy" AND (school OR "K-12")
    - "LSTM" OR "ANN" OR "random forest" AND "school energy" (keywords change depending on the ML methods being researched for different sections)
- "load forecasting" AND "school" AND ("occupancy" OR "schedule")
- "energy prediction" AND "educational" AND ("machine learning" OR "deep learning")
- "building energy forecasting" AND ("public school" OR "K-12")

- "energy forecasting" AND "school" AND "case studies"

Selection process

This review strictly followed the PRISMA approach, which details the number of sources identified, the screening process, sources included and excluded, and reasons for exclusions.

FIGURE 01: Studies selection process based on criterias using the PRISMA approach.



SOURCE: Prepared by the authors (2025).

During the identification phase, database queries and backward citation retrieved 13 studies for in-depth review, and a total of 76 papers for references. Afterwards, the screening process is carried out, which removes sources with duplicated or similar titles or abstracts. Papers that passed this screen stage underwent full text review for eligibility based on the following inclusion and exclusion criterias:

- Empirical studies, reports, or conference papers published between 2019 and 2025.
- Studies with a high relevance to the topic, meaning they apply machine learning or deep learning algorithms to forecast energy consumption or energy-related variables in school or educational buildings, or other campus studies where models are transferable to K–12 schools.
- Papers that adequately report model architecture or algorithm, input features, dataset size and temporal resolution, and at least one performance metric to allow performance assessment in the review
- Full text in English is available

Conversely, studies that are not relevant to the review, such as papers that focus exclusively on residential or commercial buildings without transferable application to school buildings are excluded. Sources published before 2019, sources without available full text or opinion pieces without adequate methodological details are similarly omitted.

Synthesis method

To synthesize the findings, Table 1 gives a structured qualitative analysis that systematically integrates findings across studies to highlight similarities and differences in results regarding model performance, settings, input features and implementation practicality. Results from the different case studies reviewed are standardized using the generalized framework in Section 4, which outlines some of the most common data features, preprocessing procedures, training and validation methods, and hyperparameter optimization techniques. Finally, in the discussion section each model category is then comparatively analyzed based on the quantitatively and qualitatively reported performance, data demands and computational complexity. This synthesis method allows the review to recognize the similarities and differences across studies, and thus drawing generalizations that could support decision making for school administrators.

RESULTS AND DISCUSSION

Case Studies Analysis

The case studies that form the foundation of this review paper encompass more than a dozen cases of ML-based models being implemented in real-world settings in Asian, European, Latin America and Middle Eastern countries, representing a range of school building levels (K-12 classrooms, secondary schools, university campuses). Different studies in different settings reflect the unique characteristics of the country in which it was conducted (for example, outcomes are often shaped by the country-specific climate conditions that strongly influence energy demands), as well as the specific models used in each study.

Case studies [9] to [21] will be grouped and discussed systematically based on the family of the models used, while also noting their climate and geographical context, then compared and synthesized using standardized tables. Synthesizing the current limited number of case studies in the field will be especially useful for identifying patterns, methodological trends, and general frameworks, since energy forecasting in schools are rather fragmented and context-specific; most papers are case studies rather than large multi-country datasets.

Table 01 provides the overview of recent literature (2019-2025) on energy consumption forecasting in school settings by summarizing studies from multiple countries using different models, presenting results and allowing side-by-side comparisons.

TABLE 01: Summary of published articles on schools using ML applications for energy demand forecasting

Reference	Dataset and Country/region	Objective	Techniques	Input features	Evaluation metrics	Reported results	Limitations/ Future suggestions
[9] Run et al. (2023)	Dataset collected from school buildings in the South of France using sensors: CO_2, indoor temp, indoor humidity sampled every 15 minutes; electricity consumption and outdoor climate data sampled hourly. Analysis is done over five months (November 2021 to April 2022)	Build a preliminary MLR model to predict hourly electricity consumption in winter for school buildings Present case study on three school/university buildings (GEII, GMP, GC) in Southern France	MLR (built two MLR forms: one-way and two-way interaction models)	Level of indoor CO_2 Indoor temperature Indoor humidity Outdoor temperature Outdoor humidity Global solar radiation Day index (weekday/weekend) Time index (occupied/non-occupied) Building net floor area	R^2 MAE (kWh) MSE (kWh) MAPE (%) RMSE (kWh)	Two-way interaction model reported better R^2 for both training set ($R^2 \approx 74\%$) and testing set ($R^2 \approx 77\%$) compared to one-way model, but underestimates results for higher loads (≥ 30kWh/hour) Per-building performance differs: GMP best ($R^2 \approx 76\%$, MAPE≈22%), GC moderate ($R^2 \approx 64\%$, MAPE≈26%). Performed poorly on GEII building ($R^2 \approx 55\%$, MAPE≈69%)	Pre-selected predictors might not be the most influential variables Recommend adding up to 7 hours lagged climate/indoor variables such as delayed solar radiation, and also adding occupancy rate as additional predictors Validation step should be included
[10] Mohammed et al. (202	Dataset collected from active schools in eastern province of Saudi Arabia	Propose a multiple regression-based model for estimating energy	MLR	City (location) AC capacity Number of floors Total roof	R^2	More than 90% of the predictions had errors of less than 15% Standard	Authors imply that future works need to expand input data to include

1)	316 data points were used to train, 35 for validation	consumption in Saudi Arabian schools		area Type of school Number of staff Building age Number of students		deviation for residuals was 28,764.02 kWh Sensitivity analysis revealed that AC capacity and building age had the most impact on the output consumption	other factors (construction materials, roof shape, school orientation, and floor material) and add more training instances
[11] Chung and Yeung (2020)	Obtained 121 questionnaires results (out of 472 secondary schools reach out in Hongkong) with no missing data Target ≈ annual energy use (sum of 12 months of utility bills)	Produce forecasting models with variables that school managers can act on (lighting, AC, management practice) using BS regression as the baseline and compare that with CNLS	BS regression CNLS regression	Building geometry Lighting characteristics (T5/T8/LED shares, lighting control) AC characteristics Number of classes (proxy of ICT counts due to low cross-school variance) Management practices	R^2	BS regression (one variable): $R^2 \approx 0.316$ CNLS regression (five variables): $R^2 \approx 0.444$ Best CNLS regression model recorded adj-$R^2 \approx 0.654$ CNLS regression improved goodness of fit compared to BS models	1-year consumption record limits temporal analysis of weather factors and heating/cooling degree days CNLS models can be difficult to interpret Methodology: Few schools returned their questionnaire late and small sample size Large number of hyperplanes is disadvantageous
[12] Prasad (2024)	Sampled 173 schools in Fiji; 154 sampled schools were grid connected Regression modeling took only 75 data points out of 151 data points (due to missing value) 55 used for constructing regression models, 20 for testing	Determine factors affecting electricity demand in Fijian schools; assess and compare MLR and ANN performance; constructing these models for electricity demand prediction in grid-connected schools	ANN MLR	Electricity demand (dependent variable) Independent variables: Age of schools Number of classrooms Floor area Number of buildings Number of air conditioners Number of students Number of lights Number of teachers	R^2 RMSE	Optimum ANN model had the second highest R 2 of 95.3% and the second lowest RMSE of 59.4 kWh/year, outperforms optimum MLR model ($R^2 \approx$ 95.3% vs $R^2 \approx$ 73.3% and RMSE ≈ 59.4 kWh/year vs RMSE ≈ 0.2248) Determined Light is the most influential variable affecting electricity demand, followed by number of air conditioning systems, school	Capacity of different electrical appliances used was not considered while constructing the regression models Future directions and recommendations: If electricity demand data, floor area data and number of students could be accessed to construct a large dataset, then more robust

						type, and school category	predictive models could be implemented Behavior of students and teachers can also be studied on how those change energy demand Schools are commended to carry out energy audits to keep track of energy use and for comparative studies
[13] Begić Juričić and Krstić (2024)	Dataset retrieved from Energy Management Information System of electrical energy usage for 149 school buildings in Croatia's Osijek-Baranja County 105 buildings (70.5%) training; 44 (29.5%) validation Enhance robustness by accounting for diverse school-building practices (include both primary and secondary schools)	Identify factors influencing annual electrical consumption (AEC) in school buildings in Osijek-Baranja County, Croatia Develop and evaluate accuracy of ANN and MLR prediction models to meet practical needs, compare the trade-offs between model complexity and usability	ANN MLR	Total number of users Total useful surface area Heated volume of the building	R^2 MSE MAPE RMSE CV RMSE	Total area (A) is the strongest predictor of annual electrical consumption (R ≈ 0.944) ANN achieved slightly better performance than MLR model in both training ($R^2 \approx 0.957$ vs 0.950) and validation ($R^2 \approx 0.954$ vs 0.949), with slightly lower RMSE and CVRMSE	Potential limited generalizability to other regions with different climatic conditions and building practices ANN model requires significant computational resources and expertise to implement and interpret Proposed future direction: Validation of the model in various geographical settings, collect data from schools in different regions and consider their different characteristics. Extend the study over a long-term period
[14] Alshibani (2020)	Energy consumption records for 352 schools (nine cities and four different	Identify factors that influence school energy consumption in Eastern Province (hot	ANN	City Number of floors Total built area—all floors (sqm)	MSE	MSE ≈ 0.00542362 (Training) ≈ 0.0159995 (Selection) ≈ 0.0564305	The model can be extended to include factors not in the current dataset

	school types) across Eastern Province of Saudi Arabia.	and humid climate zones), develop a prediction model for school facilities		Total roof area (sqm) Type of school Number of students Number of staff Age of building Number of classrooms Total air-conditioned area (sqm) AC capacity		(Testing) Validation achieved satisfactory results: 87.5% accuracy Weakest correlation: −0.35 between "type of school" and "AC capacity" Strongest correlation: 0.95 between "number of classrooms" and "total air-conditioned area"	(such as schools' orientation in urban areas) and be integrated with a BIM model
[15] Wasesa et al. (2022)	Real-life electricity consumption dataset from a technological university in Bandung, West Java	Enhance forecasting accuracy for microgrid-based buildings during COVID-19 by incorporating internet data	ARIMAX XGBoost SVR	Temporal data: day, date, month Inerta variables: previous record of each variable in the dataset (lag feature) Publicly available data: COVID-19 data from public government website, Google Mobility, and Google Trends	MAE MAPE RMSE	XGBoost 4 model achieved best MAPE score (19.6%) and best MAE (81.494) XGBoost achieved best RMSE score, better capturing electricity consumption dynamics that XGBoost 4 Lagged (inertia variables) improve prediction (MAPE score of XGBoost 4 improved from 28.4% to 19.6%) COVID-19 data boosted performance: MAPE score increases 21.4% → 20.8% for XGBoost and 43.8% → 22.6% for ARIMAX model Internet-based data improved forecasting accuracy	Consumption patterns are somewhat specific and limited to the local context of the study during the COVID-19 period COVID-19 show unpredictable events, so predictors needed to be customized to context-specific COVID policies, internet data only partly represents the public's concern

[16] Nebojša Jurišević et al. (2021)	Data collected over five years (2015-2019) from 11 public kindergartens in Kragujevac, Serbia	Assess predictive models' ability to identify key factors in heat consumption in public kindergarten	SLR MLR DT ANN	Monthly weather data: Heating Degree Days (HDD) Constant values (building characteristics): Building built year Type of built Heating source Number of buildings Heated floor area Heated building volume External walls gross surface External walls net surface (EWNS) Gross fenestration area (GFA) Ceiling surface External walls average U-value (EWA-U) Average fenestration U-value (AF-U) Average ceiling U-value Gross building envelope surface Net building envelope surface (excluding fenestration area) Roof type (flat or pitched) Number of building	R^2 MAPE	Linear models: In terms of R^2, SLR model achieves lower precision than MLR model (0.84 vs 0.89) MAPE the same for both (33%) Low-consumption range (<10 MWh/month): both models had poor accuracy High-consumption rage (>40 MWh/month): MLR 11% better accuracy than the SLR Non-linear models: ANN outperformed DT in both test and training set in R^2 values (0.96 and 0.92 vs 0.92 and 0.84) and MAPE (ANN achieve 10% better than accuracy than DT) High range: ANN (MAPE ≈ 9%) performed better than DT (MAPE ≈ 16%)	Linear models (SLR, MLR) are easy to use, but have restrictions due to multicollinearity and low prediction accuracy in the low heat consumption range Downsides of the ANN model: complexity, making it hard to interpret and develop compared to linear models

				floors Survey-based input: Area of windows used for classroom ventilation			

[17] Cao et al. (2023)	An educational building in Xi'an, Shaanxi province Dataset ranging from October 2021 to December 2021	Propose a prediction model that uses SHAP method and integrated learning	RF XGBoost SVR GRU CNN LSTM SHAP-Stacking CNN-GRU CNN-LSTM	Time features, meteorological features, and historical data Original dataset contained 16 features; after feature analysis, reduced to 4 features: Time Day of the week (Sunday to Saturday) Total solar radiation at the previous moment Energy consumption at the previous moment	R^2 MAE MSE RMSE MAPE CV (Coefficient of Variance)	SHAP-Stacking model achieves best performance consistently: Reduced RMSE by 13.64%-34.55% and MAE by 10.25%-30.54% RMSE was 13.64% lower than second-ranked model, RF Hybrid models CNN-GRU and CNN-LSTM improved performance: RMSE values reduced by 4.98% and 8.40% MAE values reduced by 4.49% and 6.95%	Limitations are not explicitly stated Implications and future direction include: improving the speed of SHAP-Stacking model, using SHAP to better select base models
[18] Muhammad Faiq et al. (2023)	Dataset obtained from a building in Multimedia University, Malacca Campus from January 2018 to July 2021 (during COVID-19 lockdown context) in Malaysia	Evaluate an LSTM energy predictor for an university campus, and compare its performance against SVR and GPR baselines	LSTM SVR GPR	Previous year's energy and forecasted next day weather Pressure Environmental temperature Relative humidity Wind velocity Rainfall duration & amount Type of day Type of lockdown	MAE RMSE	LSTM outperformed SVR and GPR: Proposed LSTM model achieved best RMSE scores (561.692–592.319 kWh) compared to SVR (3135.590–3472.765 kWh) and GPR (1243.307–1334.919 kWh) Across 20 ran simulations, LSTM reported best MAE (165.20) and second-highest RMSE (572.55) Outperformed SVR (MAE ≈ 2851.339 kWh, RMSE ≈ 3270.836 kWh) and GPR (MAE ≈ 999.880 kWh, RMSE ≈ 1310.105 kWh)	Limitations of LSTM model used: requires huge historical data and external variables (environmental or schedule-related variables) for accurate prediction Suggests adding more features to the model (such as occupancy data) to enhance accuracy in future works
[19]	Dataset	Develop and	CNN-BiLSTM	Input	MAE	Higher demand	Supports the

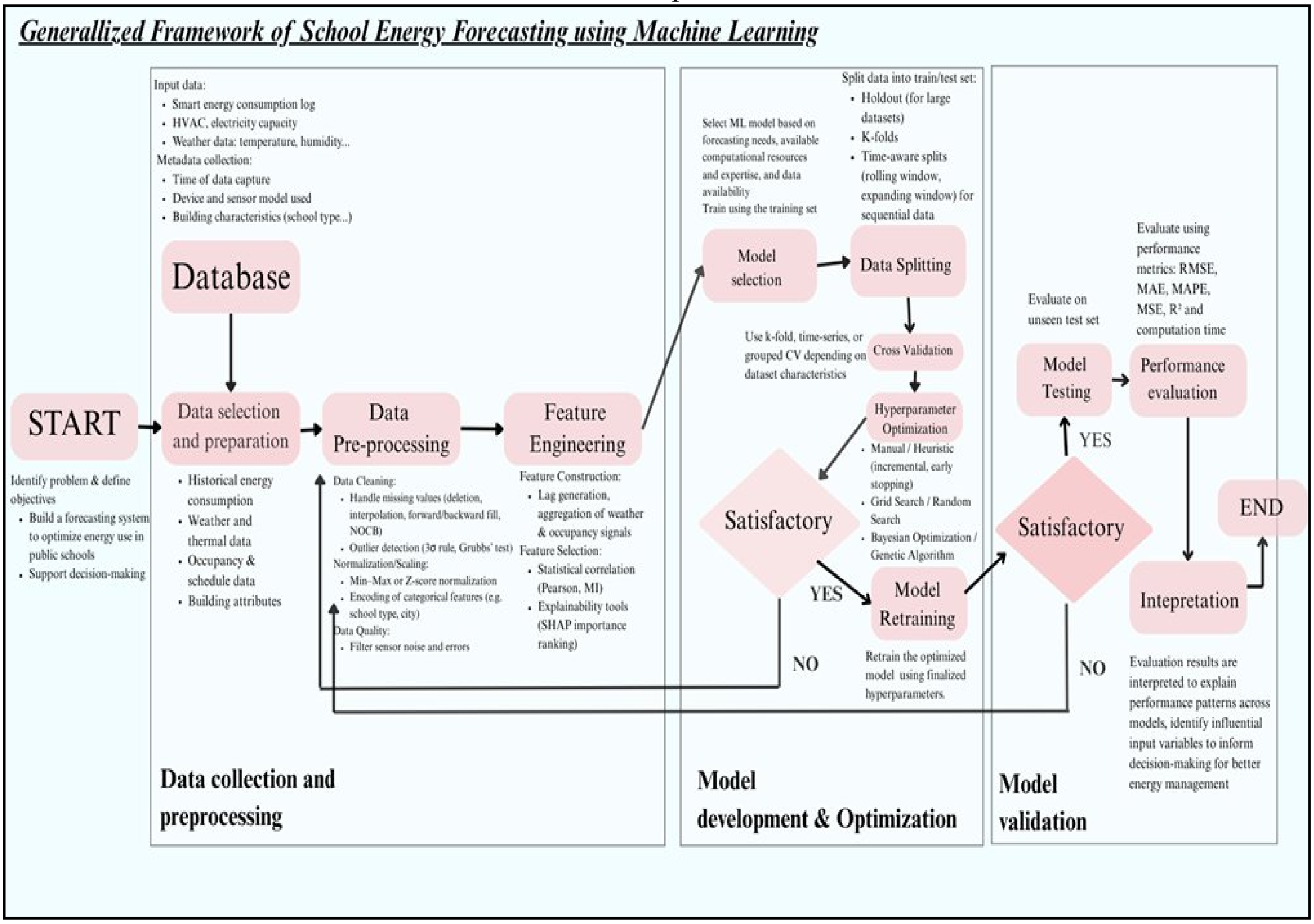
Ahmad et al. (2024)	obtained from UiTM Permatang Pauh campus building in Malaysia, 343 days of data collected at 30-minute intervals, a total of 16,464 data points for analysis	benchmark a hybrid CNN-BiLSTM model for university campus buildings	ANN (benchmarked against)	Day Hour Lagged previous week load consumption Lagged previous day load consumption Holiday Lecture/Non-Lecture week	MSE MAPE RMSE	and larger variability for lecture weeks (1766.95 kW) than non-lecture weeks (1263.95 kW) CNN-BiLSTM outperforms BiLSTM (MAPE ≈ 8.77%) and ANN (MAPE ≈ 13.03%), reporting lowest errors and MAPE ≈ 6.99% ANN model's higher error is due to the simpler algorithm compared to more complex models. BiLSTM and ANN models show under-forecasting on weekdays	utilization of advanced neural network architectures like CNN-BiLSTM in achieving more accurate forecasts and optimize energy resource allocation
[20] Shahid et al. (2023)	Daily consumption (multi-year historic series) for six public schools in Skellefteå municipality (Sweden) Dataset ranging from 2011 to 2022, tested on 2022 dataset	Evaluate RNN-LSTM, CNN, AE for power and heating forecasting in Swedish schools	RNN-LSTM Stacked LSTM-AE CNN-LSTM hybrid, Hybrid LSTM-AE-LSTM	Historical energy consumption District heating values using ADD (Actual Degree Days), NDD (Normal Degree Days), HWDD (Hot Water Degree Days) Weekday and daily temperature Cyclic time encodings	RMSE nRMSE (normalized RMSE)	CNN-LSTM achieved the best accuracy: RMSE ≈ 18-25% (electricity); RMSE≈ 20-30% and nRMSE ~ 5% (district heating) Models using only consumption: RMSE ≈ 60-90% (electricity) RMSE ≈ 35 - 60% (district heating) Average RMSE of weekdays: 45-70%	CNN are unsuitable for capturing long-term temporal sequences (requires LSTM layers), plan to develop an anomaly detection method-based LSTM architecture in future works

[21] Ortega-Díaz et al. (2025)	Case study of a classroom in an educational building in Bucaramanga, Colombia. Duration: 2.5 months, from February 19 to April 30, 2024. After resampling/interpolation, complete dataset contains 20,435 sample	Compare SVR, Decision Tree, and Random forest for predicting classroom AC energy in a tropical-climate classroom	SVR DT RFR	22 input variables (climatological, operational, and temporal): Energy consumption; Outdoor temperature; Outdoor humidity; Dew point; Wind speed; Wind direction; Heat index; Atmospheric pressure; Rain rate; Solar radiation; UV index; Cooling degree days; Door status; Windows status; Indoor temperature; Indoor humidity; Motion; Occupant number; Computer number; Occupancy; Day of the week; Time of the day; Working day	R^2 MAE RMSE	Scenario 4 (March data only) observed best MAE, RMSE, and R^2 Best performing model in sampling: RFR (RMSE = 18.05 Wh; MAE = 4.98 Wh; R^2 = 0.97) During testing: highest R^2 = 0.78, lowest RMSE = 49.77 Wh, achieved by SVR model. Predictive ability of models, especially RFR, decrease when evaluated with new test data Time-stamping: 1-minute sampling improved model performance. RFR obtained an R^2 of 0.95 (highest of all combinations) 90:10 train/test fraction of the SVR model provides the lowest error. 70:30 fraction has the highest RMSE. Using larger training fractions improves pattern learning of the models	Authors acknowledged that 72-day monitoring period is insufficient to generalize the results or develop a high accuracy model Suggests more exhaustive and prolonged monitoring to enhance accuracy in interpreting AC systems; extend to other classrooms, different areas of buildings, or even entire buildings

SOURCE: Research data (2025).

Generalized framework

FIGURE 02: Flowchart of generalized framework of school energy demand forecasting using ML techniques.



SOURCE: Research data (2025).

Data and feature patterns

Across the case studies, the energy consumption is most strongly predicted by the following factors: historical load, weather variables, occupancy levels, as well as other building characteristics such as floor areas, number of classrooms, AC types, etc.. Historical load consumption and lagged features are consistently used to forecast future energy consumption in almost every study mentioned, and appear to be one of the most reliable predictors. For example, in the study by Faiq et al. (2023), it is suggested that the model use “the previous year's energy data and forecasted weather as the input parameter to forecast the next day,” as a core input to the LSTM day-ahead forecast [18]. These features are especially crucial in the study of educational buildings in Bandung, Indonesia in the context of the COVID-19 outbreak, in which Wasesa et al. (2022) [15] and others tested single-lag versus multi-lag by incorporating two types of inertia variable sets: variable 1 (xt-1) and variable 2 (xt−1, xt−2, xt−3, xt−4, xt−5, xt−6, xt−7, and xt−14). The results reported significantly improved prediction accuracy when multi-day lags (1, 2, 3,..., 14-day history) are included; specifically, the MAPE score of the XGBoost 4 model showed an improvement from "28.4% of inertia variable 1 to 19.6% of inertia variable 2" [15]. Similarly, using inertia variable 2 also improves the XGBoost 6 model as the MAPE scores also displays improved accuracy from "35.8% (inertia variable 1) to 24.9% (inertia variable 2)" [15]. While the overall evidence supports that lagged historical data are crucial for reliable forecasting, it is important to consider that not every technique benefits equally from lagged predictors, as seen in the limited gains for SVR models in the same study.

Besides historical consumption, multiple studies also utilize weather and thermal conditions data, such as humidity, temperature or irradiance level, as a common input data for energy forecasting. This is crucial since weather features like rainfall may decrease or increase energy consumption for cooling in hotter climates [45]. Ortega-Diaz et al. (2025) [21] observed that the humidity level fluctuations throughout the day demonstrated an "inversely proportional relationship" with the outside temperature, with that temperature coinciding with AC power consumption in the classroom. Moreover, another weather data type which is indoor humidity, also displays a negative correlation with the AC consumption [21], which shows the link between cooling demand and outdoor thermal conditions. Similarly, in study [18] where LSTM was used to predict consumption in Multimedia University in Malaysia, it was found that environmental variables such as temperature could "greatly affect the accuracy of the model." Specifically, this study tested the importance of each weather variable by removing average pressure, temperature, humidity, wind speed and rainfall levels, and found that it significantly affected the model's performance. By removing average temperature, MAE rises to 212.792 kWh, increasing by approximately 28.8% as compared to the MAE of 165.2 kWh when said data is included. Removing the rainfall amount raises MAE to 181.361 kWh, which is roughly 9.8% higher than full model MAE, and without average temperature and rainfall amount, the model obtained higher RMSE scores around 580 kWh. Overall, this allowed Faiq et al. (2023) to conclude that "temperature, wind speed, rainfall amount and rainfall duration are important variables in increasing the performance of the model," and temperature and rainfall are important parameters for energy forecasting. Most study settings utilize occupancy levels (occupied/unoccupied) or occupancy signals (number of classes, number of occupants, school holidays and weekends, timetables, event schedules, etc..) as one of the main input variables for predicting consumption. Occupancy rate is sometimes reflected using time features (time of day, day of week) [17]. Occupancy significantly influences energy consumption due to the use of electronic equipment, lighting and air conditioning; in other words, higher occupancy levels correlate with an increased use of resources, requiring more energy to operate [13,21]. Ahmad et al. (2024) [19] further enforced this as they reported that lecture weeks consistently saw higher load consumption and more significant fluctuations compared to non-lecture weeks, which emphasizes the significant role of occupancy in forecasting energy demand. In addition, building attributes and features are also crucial variables in energy prediction; as identified by Mohammed et al. (2021) using a sensitivity analysis, building characteristics such as AC capacity and building age were the most important factors affecting energy consumption [10].

Data preprocessing and feature engineering

Data cleaning: Outlier detection and Missing value imputation

Building energy forecasting oftenly faces the issue of poor quality data, namely missing values and outliers due to faults in data collection, transmission and storage and the inherent complexity of building operations, which necessitates the process of data preprocessing to ensure the validity and reliability of analysis results [46]. To enhance data quality and improve prediction, studies have widely employed data cleaning which includes missing value imputations and outliers and noise removals [46]. In building energy prediction, noise, which includes erroneous data values and missing values, implies data points that do not reflect reality such as ones caused by faulty sensors and transmission equipment [47]. On the other hand, outliers are primarily determined by statistical methods or non-statistical methods. In the studies reviewed, two studies that addressed outlier removal both used statistical methods. In educational buildings energy prediction in Shaanxi, Cao et al. (2023) processed the outlier using the standard deviation method, or the 3σ method which measures the distance of the factor from the mean [17]. This method states that if the deviation from the mean value is more than three times, the value of this point is considered an outlier and therefore eliminated [17]. Similarly, Mohammed et al. (2021) [10] used a Grubb's test, a statistical method for detecting outliers in a univariate data set in an approximately normal distribution, and the highest and lowest values would be considered potential outliers.

Regarding handling missing values, Fan et al. (2021) [46] suggested that there are two ways to handle missing values in building operational data, the first being simply removing the data samples that contain missing values, as done in the process of data cleaning of six Swedish schools, in which missing data and outlier values to simply removed before normalizing the datasets [20]. The second method of managing missing values involves applying missing values imputation methods to replace those missing data points with inference ones. Some common methods include mean/median imputation, backward or forward filling, KNN imputation or regression based imputation [46]. Wasesa et al. (2022) reported 29,513 missing values (4.2% of the total 701,280 data) in electricity data in minute intervals, which required data imputation using the NOCB (Next Observation Carried Backward) method of backward filling in which instead of filling the missing value with the last observation, it fills it with the next observed, non-missing value [15]. Cao et al. (2023) deployed linear interpolation, which is the method of curve fitting using linear polynomials to construct new data points within the data range, to estimate the missing value according to two adjacent data points to be interpolated in the sequence [17].

Feature selection

It is critical to note that while adding more input variables can improve prediction in some cases, the more features entered does not necessarily result in higher accuracy; rather, the degree of influence of each feature on the consumption is what is helpful in targeting energy efficiency [17]. In fact, Cao et al. (2023) reported that the model's prediction accuracy "decreases with the increase of number of features," which led the authors to eliminate weaker predictors and enhance performance by limiting their optimal dataset to time, total solar radiation, historical consumption and day of the week. In the context of building load prediction specifically, feature importance analysis to retain relevant and most influential features while discarding redundant or irrelevant features, preventing risks of overfitting, poor generalization capability, reliance on noise, and overall improving the model's performance [48].

To determine the significance or "weight" of specific features within the model, multiple studies reviewed employed powerful feature importance analyses. Some commonly used feature selection techniques include the use of Pearson's correlation analysis which was deployed in the research (Shahid et al., 2023a) in order to determine how strongly variables (such as different date time parameters, energy consumption, and actual degree day) are associated with heating energy use before incorporating them in predictive models [20]. The MI-based feature selection, which evaluates each feature and yields a relevant feature subset, is also commonly used as it can handle data with both categorical and numerical variables. Notably, the SHapley Additive exPlanation (SHAP), an explanation model that determines the credit and impact of an input feature to a model's output prediction accuracy, has been commonly deployed in multiple studies to evaluate the positive or negative impact a variable has on the model which leads to changes in the SHAP value [49]. In [18], the authors deployed this method by removing individual environmental variables across repeated simulations to test the impact on accuracy, and reported that excluding temperature or rainfall significantly worsened performance, which allowed the conclusion that those variables were among the strongest parameters. Similarly, Cao et al.'s study also deployed the SHAP model for feature importance analysis [17], and results showed that the SHAP-Stacking model shows the "best results in the calculation of each evaluation metric." Specifically, the SHAP-Stacking model was the best performing in eight models, having significant reductions from

34.55% to 13.64% in RMSE and 10.25% to 30.54% reduction in MAE, indicating that by adopting feature importance analysis techniques, the researchers were able to obtain more accurate predictions.

Data normalization and scaling

Data normalization is the process of rescaling features to a standardized scale, which ensures that each variable contributes equally and prevents any feature that simply has a larger magnitude from disproportionately dominating the analysis [50,51]. Data normalization is critical in the data preparation process as it mitigates the impacts of scale variations of features in the raw input data and ensures that each feature contributes effectively to the analysis [52]. Normalization improves prediction accuracy as well as the model's convergence speed [17]. Commonly assessed normalization approaches in energy consumption prediction include Min-Max Scaling, Mean, Z-score, IQR and VSS methods. Among the case studies reviewed, most common normalization approaches are Min-max normalization and Z-score normalization.

In particular, Cao et al. (2023) utilized the Min-max scaling approach, a method that rescales features with differing values to a standardized range, typically between 0 and 1 [17]. This method was also adopted by Wasesa et al. (2022) in their study of Indonesian educational buildings to normalize data for both the predictor and the target variables [15]. The specific formula of this scaling approach is as follows, in which x is the raw data retrieved, x scaled is the normalized data between a specific range, xmin is the minimum value from the raw data, and xmax is the maximum value from the raw data.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}}$$

[1]

Similarly, in their study of energy prediction in school buildings, Alshabani (2020) used the Minitab software to automatically scale numeric factors, using the minimum, maximum, mean and standard deviation and reported a good level of accuracy. In the software, the categorical data were transformed into numerical data; specifically, the city factor was encoded numerically on a 1 to 7 and the types of schools on a 1 to 4 scale, and both were defined as categorical variables [14].

Another popular standardization technique is Z-score normalization. Ortega-Diaz et al. (2025) [21], for instance, decided to use StandardScaler to normalize data points from sampled Colombian schools via finding the z-value which measures how far a value x is from the mean of a data set. By calculating the Z-score, observed data points are transformed to change the observed values to have characteristics of a standard normal distribution in which the mean is 0, and the standard deviation is 1. The transformed data is equally distributed above and below the mean value, which makes the variance equal to one [53]. This method follows the following formula:

$$x_{standardized} = \frac{x - \mu}{\sigma}$$

[2]

Where μ is mean and σ is standard deviation.

Model training, data splitting, tuning and validation

Data splitting

The model training process in educational buildings energy forecasting involves feeding historical school-specific data into chosen machine learning algorithms which then can learn the patterns and uncover the relationship between input variables and energy consumption, gradually adjusting their internal parameters to minimize errors and improve prediction accuracy. Typically, model training involves the following main stages: a training phase where the chosen algorithm learns patterns from input data, a validation phase where performance is assessed using evaluation metrics, and hyperparameters are tuned; and a testing phase where the model is tested on unseen data to measure its generalization ability beyond the specific study setting.

Arlot, S., & Celisse, A. (2010) emphasize that, since real-world datasets are limited, data splitting is essential to mitigate overfitting, or the case where the model performs well on training data but performs poorly on unseen data [54]. Data splitting involves allocating a portion of the data for training, and the remaining portion for testing and validation. One of the most common and simplest data partitioning techniques is the holdout method, which involves randomly holding out the test dataset from the training process while the rest is reserved for testing [55]. Generally, although the exact proportions may differ, a large portion of roughly 70-80% of the data is allocated to the training phase, though this share can increase to as much as 95% when larger datasets are available [55]. In fact, among the case studies reviewed, multiple studies utilized this holdout validation as their data splitting method due to its simplicity. For instance, Orgeta-Diaz et al (2025) experimented with different split proportions of 50:50, 60:40, 70:30, 80:20, 90:10, and found that the 80:20 and 90:10 ratios yielded the lowest RMSE values across models, while the 70:30 fraction performed worst, indicating that a larger training set generally yields better accuracy [21]. In the study by Faiq et al. (2023) [18] in Malaysia applying CNN-BiLSTM models, 70% of the dataset was used for training while the remaining 30% was used for testing, while another study in Malaysia by Ahmad et al. (2024) [19] selected a 85-15 split. Similar approaches were chosen in Saudi Arabian researches: Mohammed et al.'s study trained a regression model with 316 data points while holding out 35 (11%) for validation [10], and Alshabani et al.'s study partitioned data into 60% for training, 20% for selection, and 20% for testing [14]. Such findings demonstrate the flexibility of the holdout method as the proportions can be adjusted differently from study to study based on available data and model evaluation needs. While this approach is simple, it is generally suitable for larger datasets with years of measured data, and may not reflect the patterns reliably for smaller datasets [55]. Additionally, the nature of school energy presents unique challenges such as being limited in duration, and relying on a single split may lead to inefficient use of scarce data while also making models highly sensitive to that specific split, worsening the model's ability to generalize.

Hyperparameter Optimization

Hyperparameters,such as number of hidden layers, neurons per layer, or learning rates, unlike parameters which are learned during the training, are set by the user prior to the training process. Hyperparameters optimization (HPO) is the process of selecting and tuning the parameters of the forecasting algorithms to the best configuration possible [56]. The accuracy of an energy prediction model is dependent on the configuration of its hyperparameters [57], playing an important role in the forecasting accuracy [56]. The most commonly used HPO methods include Grid search, Random search, Bayesian optimization, heuristic optimization, more advanced evolutionary algorithms, etc..

Random search involves training and testing the model based on random combinations of the numeric, integer, or categorical hyperparameters [58], and according to Hossain et al. (2021) [57], this method could better identify new combinations of the parameters or better discover new hyperparameters in order to improve the optimization, leading to improved performance thought taking more time. Random search typically has much better performance than grid search in higher-dimensional HPO settings despite being computationally costly [58].

Another widely used HPO approach is grid search, which involves the user setting a fixed grid of hyperparameters and the model is trained exhaustively based on every possible combination within that grid [57]. In the study of Indonesian educational buildings [15], this grid search approach was applied across several algorithms. For the XGBoost models, Wasesa et al. (2022) optimized six parameters: maximum depth, learning rate, minimum child weight, objective, sub sample and tree method. For SVR models, they optimized four parameters: C, gamm, kernel and epsilon. Similarly, the optimization for ARIMAX models in this study considered AR, I and MA [15], which demonstrates the ability of grid search to handle multiple different model types for linear ARIMA models to boost algorithms. However, the author also stated that the training time varied considerably with XGBoost taking the longest runtime, acknowledging the common limitation of grid search of being computationally expensive for exhaustive evaluation.

Bayesian Optimization (BO), an automated tuning method, is becoming widely used due to its ability to find the optimal hyperparameters in fewer steps and higher efficiency than grid-based methods [57]. Bayesian Optimization models the mapping between hyperparameters and past performance [62], essentially creating a surrogate model using Gaussian process or a random forest [58], allowing the optimizer to focus its search on the most promising regions of the hyperparameter space [59,60]. This approach of creating a surrogate model could be effective for handling hybrid models such as CNN-BiLSTM used for education buildings load forecasting, because training those networks is computationally heavy and the hyperparameter space is large. In [19], the authors utilized the Bayesian optimization algorithm to facilitate the identification of the best hyperparameter values and therefore enhancing the performance of the CNN-BiLSTM model. Specifically, they include the unique layering strategy of CNN as the foundational layer with 128 filters and four kernel sizes; followed by seven additional BiLSTM layers. The BiLSTM was configured with 180, 80, 80, 50, 10, 15, and 1 neurons across its layers, while the ANN block contained eight layers with neuron counts of 100, 100, 80, 120, 100, 30, 90, and 1. They emphasized that they were able to obtain the best hyperparameter values using Bayesian optimization and ultimately achieve improved accuracy with less errors. Overall, Wasesa et al. (2022) noted that facilitating optimal hyperparameters values using BO algorithm led to enhanced performance and efficiency of the CNN-BiLSTM mode [15].

Evolutionary algorithms stimulate natural evolutionary processes of genetic improvements in humans or animals to solve optimization problems [61]. In simple words, this approach is based on the concept that when individuals of a population compete for scarce resources, only the fittest individuals could survive [63]. This mutation process of selecting fittest value could be applied to optimization where evolutionary algorithms are used for hyperparameter tuning. In the study of Serbian schools [16], a Genetic Algorithm was used to develop an Evolutionary Assembled Artificial Neural Network (EANNN) which was utilized for heat consumption forecasting in kindergartens to configure optimal ANN parameters, which include: the number of neurons in a hidden layer, the type of activation functions in the layers, the number of learning epochs, learning rate, and momentum. According to Nebojša Jurišević et al. (2021), the optimization using GA involves automatically and iteratively configuring and evaluating the ANN performances, improving performance via essentially updating the populations of candidate solutions until convergence [16].

Apart from the common optimization approaches reviewed above, heuristic and manual tuning approaches are evident in several case studies. For example, for their NN model, Alshabani (2020) [14] employed incremental and stepwise optimization rather than exhaustive searches in order to figure out the optimal number of hidden neurons. According to Alshabani, the algorithm for order selection was "incremental order" which involves beginning with the minimum number of neurons (order), gradually increasing with a certain number of perceptions in each iteration, and finally selecting the optimal order with the lowest selection lost; the final result was an ANN with three hidden neurons. The authors also note that in a complex model like such in the study, the error of selection increased with the number of neurons. Other studies, though not explicitly stated by the authors, relied on manual adjustment combined with early stopping to manage overfitting. Importantly, Begić Juričić and Krstić (2024) also stated that after 10 consecutive epochs, the training is stopped if there was no improvement in validation loss, and the training is stopped at 100 epochs to avoid excessive computation [13]. Such approaches demonstrate how incremental or early stopping criterias can be effective compared to other search methods, especially in resource-limited contexts.

Evaluation, Validation and Cross Validation

The evaluation of these ML-based models relies on quantitative evaluation metrics that measure the difference between the predicted and the actual consumption values, evaluating whether the model's performance was satisfactory or not. Some common evaluation analysis that measure the performance's error include: Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and the Mean Absolute Percentage Error (MAPE); the lower these metrics are, the better the model has performed [9]. MAE, which measures the average magnitude of prediction errors, is less sensitive to outliers, meaning it is suitable for noisy data sets, whereas Root Mean Squared Error (RMSE) is also commonly used as it penalizes larger deviations more strongly [64]. Other metrics such as R^2 indicates model fit (the higher the R^2, the better the model performance is) or MSE, which is common for optimization tasks but less interpretable than other metrics [64].

Cross-validation (CV) is a resampling procedure that improves the reliability of ML models by ensuring that the performance is evaluated across multiple partitions of the dataset. By randomly shuffling the data into diverse subsets of train and test sets, in which each contains a representative sample of the data [55], it mitigates the risks of obtaining biased results from relying on a single train-test split. According to Hasanov et al. (2022), one of the most widely used CV approaches is k-fold cross-validation where the dataset is repeatedly split into k folds, or subsets, to be trained on k-1 folds and tested repeatedly, and results are then averaged [55]. This method not only reduces variance and improves generalization ability, but for school settings which usually have scarce datasets, k-fold CV could effectively maximize the amount of data available by allowing each data point to be used for both training and validation which provides a more robust and generalizable estimate of model's performance [65]. Several studies in the educational energy forecasting literature have applied k-fold CV to strengthen their model, such as Run et al. (2023) [9] who assessed model performance using 10-fold CV. Similarly, Muhammad Faiq et al. (2023) applied a 5-fold CV combination with a grid search to tune the hyperparameters of a SVR model. Ortega-Diaz et al. (2025) study in Colombia slightly differs in the way it initially splitted the data into an 80:20 training-testing split but also applied k-fold CV within only the training set. The authors identified that, due to the limited data available, the four-fold ($k \approx 4$) cross validation allowed them to best maintain low computational cost while achieving a balance between bias and variance [21]. Chung and Yeung (2020) [11] applied Leave-One-Out Cross-Validation (LOOCV), which is another CV technique that is effective when datasets are small. In their study, Chung et al. used LOOCV to check for overfitting in stepwise BS models by comparing training errors with LOOCV sum of squared errors, and found that they were similar which indicated no serious overfitting. While LOOCV could offer enhanced accuracy for specific models compared to k-fold CV, which could be beneficial for some school settings studies with limited data, it requires greater computational resources [66].

A large portion of case studies on energy forecasting in educational buildings rely on sequential time series data such as historical electricity load, occupancy schedules, lagged loads, or weather variables in which the latter observations are dependent on the previous patterns. Because of this, in such cases, the common k-folds CV which assumes data points are independent of each other and allows random reshuffling, cannot be applicable as randomly splitting these time-dependent sequences could risk data leakage, or future information leaking into the training set [55]. Instead, time series require CV approaches that preserve the chronological order of data by training the model on past observations and validating it on the future observations while still ensuring multiple rounds of training and validation. One of the most commonly used time-aware approaches is the rolling window method where the dataset is divided into subsets and the training set is gradually expanded and tested on subsequent folds of the data, which ensures that the model is always trained on past observations and evaluated on future ones. Among the case studies, this method could be seen in research by Shahid et al. (2023) which used 3-8 years of historical data of sampled Swedish school buildings for training and 2022's data as the testing set. Wasesa et al.'s study, which also depends on time series data, used the dataset from 1 March 2020 to 30 April 2021 for training, the dataset from 1 May 2021 to 31 May 2021 for validation, and data from 1 June 2021 until 30 June 2021 to test for accuracy scores [15]. It is important to note, however, that while time-aware CV methods are essential for sequential datatasets, many case studies reviewed in this paper, including ones that have time-dependent input variables, do not explicitly state in their methodology whether the chronological nature of data is preserved in the validation process. In energy forecasting studies of school settings, it is crucial to address whether the temporal dependencies were accounted for, as that may raise concerns about potential biases in the models' performances.

Discussion of common models

This section will discuss some of the most commonly employed ML approaches and most outstanding models from different model families for school load forecasting in recent years, spanning from traditional linear regression models and neural networks such as ANN, to advanced architectures like XGBoost or hybrid deep learning techniques. When assessing and comparing each approach, multiple factors including applicability, data availability, computational capacity and resource constraints are taken into account. Overall, through providing an overview of each model, summarizing their strengths and weaknesses, and examining their performance as reported in recent empirical studies of real-world school settings, this discussion section aims to inform school administrators in selecting a suitable energy management model for their own schools.

Multiple Linear Regression (MLR)

In recent years, linear techniques have been popularly utilized for load forecasting, especially load forecasting in school settings. Linear regression analysis, which uses the independent variable to predict the independent variable, estimates the coefficients of the linear equation that best predicts the independent value by fitting a straight line that minimizes discrepancies between predicted and actual values [22]. In this section, the paper will discuss one of the most common linear techniques used in school energy prediction: Multiple Linear Regression (MLR).

Despite the emergence of numerous comprehensive techniques, statistical and hybrid methods, linear techniques remain a popular and suitable choice in school energy forecasting, especially serving as a method of preliminary energy assessment. Across current literature, those complex methods are reported as requiring specialised softwares, user expertise as well as a model calibration; on the other hand, linear regression analysis techniques overcome such difficulties by providing a reliable alternative for unskilled users or resource-limited settings [23]. Namely, MLR method, which incorporates multiple explanatory features, does not require calculation tools such as personal computers or software programs [23], making it one of the most simple, low-cost, and interpretable prediction methods. In the context of school energy forecasting, typical applications include benchmarking annual consumption across institutions, short-term load prediction using climatic variables and operational schedules, or being used as baseline in school building energy prediction against other techniques. Several studies have reported moderate accuracy in MLR performances. Specifically, in their study across schools in Fiji, Prasad et al. [12] reported moderate accuracy in MLR model (8c) with an $R^2 \approx 73.3\%$ and an RMSE ≈ 0.2248. MLR models usually surpass Simple Linear Regression (SLR) techniques, as Nebojša Jurišević et al. (2021) reported SLR model achieving lower precision than MLR model ($R^2 \approx 0.84$ vs $R^2 \approx 0.89$) [16]. Juričić & Krstić's study [13] which had a large multi-building dataset (149 schools) also reported MLR model's achieving good performance of R^2 values of 0.950 and 0.949 for the training and validation sets, respectively; ANN models used in the same study achieved only modest improved accuracy of $R^2 \approx 0.957$, indicating that with a larger cross-sectional dataset of school districts with multiple schools, MLR can be competitive in its accuracy and predictability.

Despite being low-computational and interpretable, the MLR method has major limitations, the first clear drawback being its inability to model non-linear relationships between independent and dependent variables [12]. This is crucial when considering the context of load forecasting in schools: school energy use is often influenced by non-linear effects such as temperature thresholds, occupancy levels, or user behaviors, which linear regression alone cannot capture. Secondly, linear methods like MLR face the issue of multicollinearity among independent variables, restricting some variables that could influence the output to be utilized; for example, regarding influential but non-linear variables as non-significant [16]. Thirdly, linear fits can underestimate high loads, exhibiting poor performance at extreme high or low consumption ranges, as results in study by Nebojša Jurišević et al. (2021) indicate both the MLR and SLR models achieved low accuracy in the high and low-consumption range, though MLR did exceeded SLR's accuracy by 11% in the high range [16]. Similarly, study by Run et al. [9] of schools in France which used two MLR models: first order and two-way interaction models also showed that even though the two-way interaction model did improve in accuracy for both training ($R^2 \approx 74\%$) and testing set ($R^2 \approx 77\%$) compared to one-way model, it still underestimates results for higher loads (≥ 30kWh/hour). Moreover, results of the same study also suggest high variability in performances across different buildings, with the model's best performance being GMP building ($R^2 \approx 76\%$, MAPE≈22%), GC moderate ($R^2 \approx 64\%$, MAPE≈26%), and performed unacceptably poorly on GEII building ($R^2 \approx 55\%$, MAPE≈69%) , demonstrating MLR's sensitivity to building heterogeneity. For this reason, Run et al. (2020) deemed MLR as mainly a preliminary model that resolves immediate energy management needs. Similarly, Jurišević et al. [16] also suggested that in low heat consumption, linear methods achieve relatively low prediction, suggesting that when school energy demand falls into a lower range (e.g., summer break), regressions forecasts are less reliable.

In short, linear regression techniques such as MLR are especially helpful as baseline models in school energy forecasting due to their low computational demand, interpretability and modest data requirements; such characteristics are especially useful when considering an energy management solution for schools with limited IT resources or schools where complete metered datasets or complex computing softwares are unavailable, allowing even unskilled school administrators to gain not highly accurate but still helpful insights on their school's energy demand. Nevertheless, compared to other models, linear regression methods usually achieve much lower accuracy, and such techniques are not to replace a dynamic simulation model; rather, they provide a simple tool for determining energy needs [23] that is applicable to most resource-limited schools. Rather than the final forecasting solution, these techniques often serve as a preliminary model [9], or a benchmarking model, like in the study by Begić Juričić and Krstić (2024), ANN models were benchmarked against MLR which effectively served as a simple and transparent baseline that more advanced models can utilize to improve upon [13].

Artificial Neural Networks (ANN)

Neural networks are computational architectures that follow the way neuronal structurals of brains process information [24], composed of layers of interconnected nodes called neurons where each connection carries a weight and each neuron applies an activation function and bias term to determine the output [25]. Thus, the layered architecture allows such models to capture complex and non-linear relationships [24], making them very suitable for energy forecasting in schools where relationships within data are oftenly non-linear. Among neural network techniques, Artificial Neural Networks (ANN) are among the most widely used Machine Learning models for load forecasting in schools. ANNs are typically split into an input layer which receives predictor variables; one or multiple hidden layers depending on the complexity of the task, which transform data through weighted connections; and an output layer that produces the final prediction [25]. Notably, in forecasting tasks, ANNs are commonly implemented as feedforward networks where data flows in the forward direction [25], fitting for school forecasting where many data variables are sequential and time-dependent.

There are numerous advantages to using ANNs in school energy prediction, one of which is that the multilayered structure enable that to model highly non-linear and complex relationships between data that some traditional regression approaches fail to capture [24], which allow them to align with the complex dynamics of school building energy systems. ANN models are flexible and able to integrate vast amounts of data [25] from multiple sensors, weather inputs, and occupancy data, which could effectively support energy management and scheduling. Moreover, according to Runge, J., & Zmeureanu, R. (2019), ANNs have the ability to learn data directly without deep understanding of the physical system or complex programing [25]; rather, they automatically adjust their internal weights during training and flexibly adapt to different contexts and robustness accordingly [25,27]. Multiple studies using ANNs for handling diverse and irregular load patterns show better predictive performances compared to those of time-series and regression-based models, reporting that they do not only achieve higher accuracy but might also require fewer computational resources than other approaches in some cases [26]. For instance, in the study of Serbian schools, Nebojša Jurišević et al. (2021) [16] applied ANN to hourly consumption data from multiple kindergartens which saw significant gains compared to Decision Tree and surpasses all other models in terms of accuracy in all consumption ranges: achieving high accuracy of $R^2 \approx 0.96$ (training) and $R^2 \approx 0.92$ (testing). Most importantly, ANNs maintained its accuracy in across different consumption ranges: in the higher range, the models achieved MAPE value of just 9% compared to that of DT with 16%; in the low and medium consumption range, the MAPE value were 28% vs. 16%, and 24% vs. 12% for ANN and DT, respectively, demonstrating ANN's ability to handle irregular load patterns typical of schools while maintaining accuracy across various consumption ranges. Similarly, Begić Juričić and Krstić (2024) [13] trained an ANN model configured with a 3-5-1 architecture and also achieved significantly highly accurate performance of $R^2 \approx 0.957$ (training) and $R^2 \approx 0.954$ (validation) [13] with RMSE values of 3024.25 and 3415.75, respectively, outperforming multiple linear regression (MLR) benchmarks and generalize "reasonably well to unseen data."

On the other hand, ANN models also have a few limitations. The most major one, according to Runge, J., & Zmeureanu, R. (2019), is limited generalization outside of their training set [67], meaning ANN models trained on season or context-specific data may not always transfer well to other conditions (e.g., model trained in summer failing in winter), which requires continuous retraining strategies to maintain accuracy that can be computationally intensive [25]. This limitation is also acknowledged the study by Begić Juričić and Krstić (2024) where the authors noted that since the models are trained on a single country dataset specific to that setting, their generalizability to "other regions with different climatic conditions and building practices are limited" [13]. Compared to other models such as CNN-BiLSTM hybrids in the study by Ahmad et al. (2024), standard ANN models were found to under-forecast weekday peak loads due to the simpler algorithm compared to more complex models, while CNN-BiLSTM hybrids reduced the error rates in nearly half [19]. In [19], ANN was mainly used as a baseline model for more complex models that are suitable for complex load dynamics such as CNN-BiLSTM and LSTM to be benchmarked against. Moreover, ANNs also suffer from overfitting that stems from learning too closely from training data thereby capturing noise rather than generalizable patterns, which reduces accuracy, especially for long-horizon predictions [25]. Study by Prasad (2024) highlighted this risk as they emphasized how ANN models' performance were sensitive to the number of parameters, available data points and missing values as the models improved significantly and yielding higher RMSE values for modeling with 75 data points than compared to 100 data points [12]; this shows how smaller, incomplete school datasets with missing variables require more caution in implementing ANNs.

Developing and implementing effective ANN architectures for school load forecasting still requires high expertise and careful hyperparameter tuning [25] to ensure high performance. In the study in Serbia, feature selection and hyperparameter configuration using a genetic algorithm called Evolutionary Assembled Artificial Neural Network for optimization required 50 generations and 200 population size [16]. This development of the EAANN model for configuration required the MATLAB software, which again underlines the computational cost of robust ANN development. In the Croatian study, ANN training relied on early stopping and fixed epochs to avoid excessive computation [13]. When schools are considering energy management solutions, computational costs and expertise requirements are crucial factors to be mindful of, and accordingly, schools with limited technical resources may struggle to implement and maintain ANN systems. Finally, due to the "black-box" nature of ANNs, they are often difficult to interpret and develop [12,16], which makes interpreting more difficult and reduces transparency for decision-makers in school, especially ones without much expertise in the field. Overall, ANNs are most suitable for schools with moderate or large but well-structured and complete datasets, the resource capability for hyperparameter optimization, and a need to capture non-linear energy demand behaviors.

Extreme Gradient Boosting (XGBoost)

Ensemble learning is a family of ML techniques that combine multiple diverse base learners to each high predictive performance, as when those models are put together, they can compensate for each other's errors, thereby reducing the bias and reliance as well as producing more robust predictions [29,30,31] that cannot be achieved by any learning algorithm alone. Ensemble learning approaches are usually divided into bagging where the model learns independently in parallel, and boosting, where models are trained sequentially so that the following model corrects the individual errors of the previous one, thereby iteratively reducing bias and improving performance [33].

For the purposes of this paper, this section will discuss XGBoost (Extreme Gradient Boosting), an emerging technique that is gaining popularity in load forecasting tasks for its superiority compared to traditional gradient boosting methods due to its robustness and high predictive ability. More specifically, XGBoost's improved performance can be attributed to additional optimizations in computational efficiency, such as incorporating algorithm for finding optimal splits in the tree [29], which allowed faster training time; and the incorporation of a regularization term in the loss function that allows an increased ability in handling missing values [33] and most notably, significantly prevents overfitting, an issue prominent in boosting-based methods that learn too closely on noisy samples [29,31]. For those reasons, the XGBoost algorithm minimizes the need for feature engineering, including data normalization and data scaling [31]. Moreover, according to Moon et al. (2024), robust XGBoost models are especially effective for large, high-dimensional datasets and complex modeling tasks [29], which renders them especially effective and scalable for capturing complex, non-linear relationships of large, complex school datasets that are dependent on various drivers (weather, occupancy, schedules). This could be shown in the study by Wasesa et al. (2022) on technological university buildings in Bandung in the context of COVID-19, where XGBoost models consistently outperform the other two models used in the study, SVR and ARIMAX, and achieved the best predictive accuracy across multiple experimental settings. Namely, the XGBoost-3 model achieved the lowest MAPE value of 11.9% while ARIMAX and SVR reported MAPE values of 13.5% and 12.5% [15], as well as minimized absolute errors with the lowest MAE value of 23.9 compared to ARIMAX (32.3) and SVR (25.) [15] despite having a larger, more complex set of predictors. The study also highlights XGBoost models' strength in handling multiple predictors and non-linear relationships as the author notes that by including a combination of temporal lags, electricity consumption, COVID-19 data, Google mobility trends as predictors, the XGBoost-8 model saw a significant improvement in its RMSE. Thus, such findings emphasize XGBoost's ability to handle complex, high-dimensional inputs while maintaining accuracy and proves it suitable for forecasting tasks in school settings with large datasets with many predictors.

On the other hand, despite the advantages, XGBoost models also face several limitations that are crucial to be taken into account when considering school energy demand forecasting applications, one of which is their tendency to overfit when hyperparameters are not carefully tuned, causing the performance to potentially decrease when inputs are highly irregular [29,34]. Unlike simpler models, XGBoost often requires intensive hyperparameter optimization of multiple interacting hyperparameters such as learning rate or maximum depth. In the study by Wasesa et al. (2022), the authors noted that their XGBoost models had to be optimized using a hyperparameter grid search that optimized maximum depth, learning rate, minimum child weight, objective, subsample, and tree method. Even though XGBoost achieved superior accuracy with the best model achieving significantly low error (MAPE = 19.6%, MAE = 81.494), it also came at the cost of significantly higher training time and computational resources compared to SVR and ARIMAX models used in the same study.Moreover, the model is also highly sensitive to sample size and input dimension, as Si et al. (2024) stating that "blindly increasing the input dimension" will increase the difficulty of capturing important interactions [35]. Overall, such findings suggest that XGBoost not only requires large, structured datasets but also careful hyperparameter tuning to prevent overfitting, which is computationally demanding and requires specialized techniques [34], making the development process more difficult and resource-intensive compared to simpler models. When these factors are taken into account, XGBoost techniques become less suitable for schools with large and irregular datasets, as well as schools with limited computational resources and expertise for intensive hyperparameter tuning.

Long short-term memory (LSTM) and hybrid LSTM architectures

In recent years, deep learning approaches such as CNNs and LSTMs have gained significant attention in the field of load forecasting for school settings with multiple schools implementing those methods. Long Short-Term Memory (LSTM) networks are a specialized form of recurrent neural networks (RNN) which solves the vanishing and exploding gradient issues that occur within conventional RNNs [36]. The architecture includes memory cells regulated by the input gate, output gate, and forget gate [39], which manage the memorization, passing and discard of information [38], allowing LSTM models to preserve long-range temporal dependencies while handling non-linear data of complex, dynamic relationships in energy consumption data [36,39]. LSTM models are also shown to consistently outperform traditional RNNs and linear approaches in capturing long-term dependencies [39]. Muhammad Faiq et al. (2023) [18] applied LSTM models to forecasting a Malaysian university campus and reported that the model outperformed both SVR and GPR across 20 simulations run as well as achieved a low MAE value of 165.20 kWh and a RMSE of 572.55 kWh, compared to SVR (MAE 2851.34 kWh, RMSE 3270.84 kWh) and GPR (MAE 999.88 kWh, RMSE 1310.11 kWh), demonstrating the model's ability to effectively extracting long-term and nonlinear temporal dependencies in education building settings. Due to their ability to retain important historical information while removing irrelevant data, LSTM models are also highly effective for capturing patterns within sequential data, making them suitable for energy forecasting tasks in schools with large historical datasets that encompasses multiple variables (such as average pressure, temperature, humidity, rainfall, etc., [19]) or time-dependent data.

However, despite their strong efficacy in time-series forecasting, LSTM, as standalone models, still face several limitations that undermine their practicality in load forecasting for school settings. The first major limitation comes from the fact that LSTM models' complex architecture, which is built around multiple gates and memory cells [37], makes them highly computationally intensive while requiring increased training time and increased memory consumption [37], especially when the number of layers and hidden state size rise [38]. Moreover, LSTMs are also highly sensitive to hyperparameter selection, and performance often strongly depends on the choices of the number of layers, hidden units, batch size, and learning rate [35,37] which greatly affect the training stability; however, fine-tuning those hyperparameters often requires extensive trial and error as well as ML expertise. LSTM models are also vulnerable to overfitting risks [37], especially when data are sparse or noisy, which is a common situation in school energy data where historical, high-resolution datasets may be limited. They also present interpretability issues [37] due to their black-box nature, which may complicate explanation among school administrators who need transparency for effective decision making. In [19], Muhammad Faiq et al. (2023) also emphasized the fact that LSTM requires huge historical data to make accurate predictions, and also depends greatly on external features such as schedule and weather information to achieve optimal accuracy, showing that LSTM is especially vulnerable to limited data. Overall, when considering LSTM models for school energy management, such constraints mentioned above should be taken into account as schools with limited computational resources or expertise in the implementation and interpretation of LSTM models, schools with sparse historical data that are incomplete without external features, or schools without the budget for intensive hyperparameter tuning might be unfit for this model.

Another drawback of LSTM comes from their difficulty in handling very long sequences, because even though single LSTMs were designed to address gradient issues in conventional RNNs, studies [37,39] emphasize that they still struggle to handle extremely lengthy input sequences and may lose some key information when processing such sequences; they may only perform well on specific types of load data and perform poorly on others due to their fixed structures [39]. This motivates researchers to often adopt hybrid methods combining different algorithms (e.g., CNN-LSTM, RNN-LSTM, etc.,) that complement or preprocess the data to reduce errors presented in individual models in order to improve accuracy and robustness [39].

Hybrid LSTM architectures have become increasingly popular among researchers and building administrators in overcoming weaknesses of standalone LSTM models, particularly in complex forecasting tasks such as school energy demand forecasting. Hybrid LSTM architectures combine the LSTM models with other algorithms that excel at complementary tasks while preserving LSTM's core advantage of modeling temporal dependencies, thus addressing the limitations of single LSTMs in handling complex data patterns [38,39,40]. For example, CNN models have the outstanding ability to effectively extract local patterns, fluctuations [41], and spatial information, so when combined with LSTM's long-term modeling capabilities, the CNN-LSTM technique is able to capture both daily trends and sharp variations in energy consumption, thereby utilizing each model's strengths as well as overcome individual limitations to improve prediction accuracy [42]. Shahid et al. (2023) also emphasized this operational robustness in handling fluctuations as they stated that the CNN-LSTM values are "very narrow to the actual load values" [20]. Cao et al. (2023) [17] also reported in their study that by adding convolutional preprocessing, the CNN-LSTM model reduced RMSE by 8.40% and MAE by 6.95%

relative to LSTM without convolution, showing the benefits of incorporating the strengths of CNN in extracting and reducing errors in school energy prediction. Load data of schools usually contains both complex temporal and non-temporals patterns that single LSTMs might struggle to fully capture [44], but hybrid approaches could effectively model complex non-linear relationships [40,42], allowing the system to adapt to dynamic environmental and behavioral variables such as weather, occupancy, or seasonal variations that drive school energy use. Incorporating the strengths of another model such as CNNs, RNNs, or even fuzzy logic and gradient boosting models [40] in the preprocessing or feature-extraction stage before the LSTM allows hybrid models to filter redundant information, learn complex data patterns and improve temporal feature extraction [40] in order to improve consistency as well as generalizability across various settings [43]. Multiple studies [35,39] have consistently reported that hybrid LSTMs, by filtering noise and outliers to reduce biases, outperform other single LSTM techniques in terms of accuracy [40], achieving lower MAPE and RMSE values in both short and long-term forecasting tasks [41,43]. In real-world school datasets, such as one of electricity demand in six Swedish schools [20], CNN-LSTM was reported to achieve good accuracy with RMSE and nRMSE of roughly 18% to 25% and 5% to 6% respectively, and that weekday RMSEs feels from ≈45%–70% to ≈19%–24% after CNN-LSTM tuning. More notably, it was also noted that hybrid approaches such as RNN-LSTM or CNN-LSTM still exhibit robustness even in noisy or incomplete datasets [40], a feature that may be especially valuable for schools where metering data may be inconsistent or contain occasional irregular patterns such as holidays [42,43]. For instance, Ahmad et al. (2024) used an advanced neural hybrid architecture CNN-BiLSTM on a compass dataset, and after Bayesian hyperparameter optimization and feature engineering, obtained a remarkable RMSE = 165.87 kW and MAPE = 6.99%, outperforming both BiLSTM (RMSE 198.12 kW, MAPE 8.77%) and ANN. The same study also reported that the model still remains accurate and generalizable across different days of the week including the weekend, which shows how hybrid models could consistently perform well in school load forecasting tasks, even with varying load patterns [19].

However, these performance improvements come at the cost of significantly greater model complexity and data demand which may limit implementation in resource-limited or public schools. The most obvious shortcoming of hybrid LSTM models is that their complex structures could significantly complicate preprocessing and hyperparameter tuning, thus requiring more training time, memory and computational resources, even more than conventional LSTMs [39,43], and although mitigation strategies such as regularization are possible, they demand further costs and expertise [42,43]. Moreover, due to that complexity, hybrid LSTM structures also typically require larger and more high-resolution training datasets and additional data for learning long-term patterns [40], as insufficient data raises the risk of overfitting. For instance, Ahmad et al.'s campus data spanned over 343 days and contained 16,474 data points. Additionally, same as with single LSTM models, interpreting prediction results of hybrid LSTMs could be difficult due to the complex internal mechanisms, potentially posing challenges for schools that need immediate decision-making [40,43]. Finally, while hybrids frequently show improvements in short-term errors, in some cases, hybrid architectures are only comparable to the single LSTM model [20], which implies that the additional complexity does not guarantee significant gains in every school context, therefore administrators should take into account such tradeoffs when considering more complex architectures.

CONCLUSIONS

This review synthesizes recent works on ML methods for forecasting energy demand in schools and campus buildings, as well as identifying consistent patterns and assessing the practicality of implementation across studies. Overall, ML methods show clear ability to improve energy management in schools. Historical load, weather/thermal variables, occupancy/schedule signals, and simple building attributes (floor area, AC capacity, etc.) appear to be the most reliable predictors of school energy use. Hybrid and ensemble approaches such as LSTMs generally achieve the lower accuracy, but their improved performance comes at the cost of greater data requirements, computational complexity, hyperparameter tuning, and being more difficult to interpret for school administrators without much expertise.

Several limitations of this paper include the limit in scope: since the review intentionally mostly covers peer-reviewed literature published between 2019 and 2025, this timeframe may have excluded relevant earlier or non-English works. Additionally, most papers reviewed are context-specific; the ML models are only applied to schools and campuses within specific settings, rely on small single-site datasets, use different temporal resolutions, and apply different preprocessing and validation techniques, which limits the generalizability and comparability across different models and settings. Finally, several papers do not explicitly report steps in their methodology, omitting crucial details such as the train/test split techniques or cross validation techniques used, making it more difficult to fairly assess and reproduce the method. Moving forward, future research should focus on clearly addressing the methodology, which includes data resolution, data preprocessing, feature engineering processes, hyperparameter optimization procedures, as well as noting resource requirements and interpretability in order to enhance reproducibility and remain accessible and actionable for school administrators.

REFERENCES

[1] JURIČIĆ, H. B.; KRSTIĆ, H. Analyzing Patterns and Predictive Models of Energy and Water Consumption in Schools. Sustainability, [s. l.], v. 17, n. 12, (2025), p. 5514. Available from: https://doi.org/10.3390/su17125514. Accessed on: 21 Jan. 2025.

[2] QUISPE, E. C.; MIRA, M. V.; DÍAZ, M. C.; MENDOZA, R. C.; VIDAL, J. R. Energy Management Systems in Higher Education Institutions' Buildings. Energies, [s. l.], v. 18, n. 7, (2025), p. 1810. Available from: https://doi.org/10.3390/en18071810. Accessed on: 21 Jan. 2025.

[3] BRAY, R.; FORD, R.; MORRIS, M.; HARDY, J.; GOODING, L. The co-benefits and risks of smart local energy systems: A systematic review. Energy Research & Social Science, [s. l.], v. 115, (2024), p. 103608. Available from: https://doi.org/10.1016/j.erss.2024.103608. Accessed on: 21 Jan. 2025.

[4] BETTER BUILDINGS INITIATIVE. K-12 School Districts. Energy.gov, (2023). Available from: https://betterbuildingssolutioncenter.energy.gov/sectors/k-12-school-districts. Accessed on: 21 Jan. 2025.

[5] ENVIRONMENTAL PROTECTION AGENCY (EPA). Energy Efficiency Programs in K-12 Schools: A Guide to Developing and Implementing Greenhouse Gas Reduction Programs. [S. l.], (2011). Available from: https://www.epa.gov/sites/default/files/2015-08/documents/k-12_guide.pdf. Accessed on: 21 Jan. 2025.

[6] BRYCHKOV, D.; GOGGINS, G.; DOHERTY, E. et al. A systemic framework of energy efficiency in schools: experiences from six European countries. Energy Efficiency, [s. l.], v. 16, n. 21, (2023). Available from: https://doi.org/10.1007/s12053-023-10099-4. Accessed on: 21 Jan. 2025.

[7] HUANG, Z.; GOU, Z.; CAI, S. Energy justice in education sector: The impact of student demographics on elementary and secondary school energy consumption. Heliyon, [s. l.], v. 9, n. 5, (2023), p. e16191. Available from: https://doi.org/10.1016/j.heliyon.2023.e16191. Accessed on: 21 Jan. 2025.

[8] JASIM, N. I.; GUNASEKARAN, S. S.; ALDAHOUL, N. et al. Toward Sustainable Campus Energy Management: A Comprehensive Review of Energy Management, Predictive Algorithms, and Recommendations. Energy Nexus, [s. l.], (2025), p. 100435. Available from: https://doi.org/10.1016/j.nexus.2025.100435. Accessed on: 21 Jan. 2025.

[9] RUN, K.; CÉVAËR, F.; DUBÉ, J.-F. Preliminary Multiple Linear Regression Model to Predict Hourly Electricity Consumption of School Buildings. Green Energy and Technology, [s. l.], (2023), p. 119–127. Available from: https://doi.org/10.1007/978-3-031-33906-6_10. Accessed on: 21 Jan. 2025.

[10] MOHAMMED, A.; ALSHIBANI, A.; ALSHAMRANI, O.; HASSANAIN, M. A regression-based model for estimating the energy consumption of school facilities in Saudi Arabia. Energy and Buildings, [s. l.], v. 237, (2021), p. 110809. Available from: https://doi.org/10.1016/j.enbuild.2021.110809. Accessed on: 21 Jan. 2025.

[11] CHUNG, W.; YEUNG, I. M. H. A study of energy consumption of secondary school buildings in Hong Kong. Energy and Buildings, [s. l.], v. 226, (2020), p. 110388. Available from: https://doi.org/10.1016/j.enbuild.2020.110388. Accessed on: 21 Jan. 2025.

[12] PRASAD, R. D. School Electricity Consumption in a Small Island Country: The Case of Fiji. Energies, [s. l.], v. 17, n. 7, (2024), p. 1727. Available from: https://doi.org/10.3390/en17071727. Accessed on: 21 Jan. 2025.

[13] BEGIĆ JURIČIĆ, H.; KRSTIĆ, H. Comparing MLR and ANN models for school building electrical energy prediction in Osijek-Baranja County in Croatia. Energy Reports, [s. l.], v. 12, (2024), p. 3595–3606. Available from: https://doi.org/10.1016/j.egyr.2024.09.039. Accessed on: 21 Jan. 2025.

[14] ALSHIBANI, A. Prediction of the Energy Consumption of School Buildings. Applied Sciences, [s. l.], v. 10, n. 17, (2020), p. 5885. Available from: https://doi.org/10.3390/app10175885. Accessed on: 21 Jan. 2025.

[15] WASESA, M.; ANDARIESTA, D. T.; AFRIANTO, M. A. et al. Predicting Electricity Consumption in Microgrid-Based Educational Building Using Google Trends, Google Mobility, and COVID-19 Data in the Context of COVID-19 Pandemic. IEEE Access, [s. l.], v. 10, (2022), p. 32255–32270. Available from: https://doi.org/10.1109/access.2022.3161654. Accessed on: 21 Jan. 2025.

[16] JURIŠEVIĆ, N.; GORDIC, D.; VUKIĆEVIĆ, A. M. Assessment of predictive models for the estimation of heat consumption in kindergartens. Thermal Science, [s. l.], v. 26, (2021). Available from: https://doi.org/10.2298/TSCI201026084J. Accessed on: 21 Jan. 2025.

[17] CAO, W.; YU, J.; CHAO, M. et al. Short-term energy consumption prediction method for educational buildings based on model integration. Energy, [s. l.], v. 283, (2023), p. 128580. Available from: https://doi.org/10.1016/j.energy.2023.128580. Accessed on: 21 Jan. 2025.

[18] FAIQ, M.; TAN, K. G.; LIEW, C. P. et al. Prediction of energy consumption in campus buildings using long short-term memory. Alexandria Engineering Journal, [s. l.], v. 67, (2023), p. 65–76. Available from: https://doi.org/10.1016/j.aej.2022.12.015. Accessed on: 21 Jan. 2025.

[19] AHMAD, M. Z.; DAHLAN, N. Y.; MAT YASIN, Z. Enhanced Load Forecasting Using CNN-BiLSTM Models in University Buildings with Solar PV. International Journal of Electrical and Electronics Engineering, [s. l.], v. 11, n. 10, (2024), p. 61–70. Available from: https://doi.org/10.14445/23488379/ijeee-v11i10p107. Accessed on: 21 Jan. 2025.

[20] SHAHID, Z. K.; SAGUNA, S.; ÅHLUND, C. Forecasting Electricity and District Heating Consumption: A Case Study in Schools in Sweden. In: IEEE Green Technologies Conference (GreenTech), (2023), p. 169–175. Available from: https://doi.org/10.1109/GreenTech56823.2023.10173792. Accessed on: 21 Jan. 2025.

[21] ORTEGA-DIAZ, L.; JARAMILLO-IBARRA, J.; OSMA-PINTO, G. Estimation of the air conditioning energy consumption of a classroom using machine learning in a tropical climate. Frontiers in Big Data, [s. l.], v. 8, (2025). Available from: https://doi.org/10.3389/fdata.2025.1520574. Accessed on: 21 Jan. 2025.

[22] PHAN, A. M.; HUNG, H. Using Linear Regression Analysis to Predict Energy Consumption. [S. l.]: Research Square, (2024). Available from: https://doi.org/10.21203/rs.3.rs-4590592/v1. Accessed on: 21 Jan. 2025.

[23] CIULLA, G.; D'AMICO, A. Building energy performance forecasting: A multiple linear regression approach. Applied Energy, [s. l.], v. 253, (2019), p. 113500. Available from: https://doi.org/10.1016/j.apenergy.2019.113500. Accessed on: 21 Jan. 2025.

[24] MICHAILIDIS, P.; MICHAILIDIS, I.; GKELIOS, S.; KOSMATOPOULOS, E. Artificial Neural Network Applications for Energy Management in Buildings: Current Trends and Future Directions. Energies, [s. l.], v. 17, n. 3, (2024), p. 570. Available from: https://doi.org/10.3390/en17030570. Accessed on: 21 Jan. 2025.

[25] RUNGE, J.; ZMEUREANU, R. Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. Energies, [s. l.], v. 12, n. 17, (2019), p. 3254. Available from: https://doi.org/10.3390/en12173254. Accessed on: 21 Jan. 2025.

[26] MYSTAKIDIS, A.; KOUKARAS, P.; TSALIKIDIS, N.; IOANNIDIS, D.; TJORTJIS, C. Energy Forecasting: A Comprehensive Review of Techniques and Technologies. Energies, [s. l.], v. 17, n. 7, (2024), p. 1662. Available from: https://doi.org/10.3390/en17071662. Accessed on: 21 Jan. 2025.

[27] JING, J.; DI, H.; WANG, T.; JIANG, N.; XIANG, Z. Optimization of power system load forecasting and scheduling based on artificial neural networks. Energy Informatics, [s. l.], v. 8, n. 1, (2025). Available from: https://doi.org/10.1186/s42162-024-00467-4. Accessed on: 21 Jan. 2025.

[28] SHRIVASTAVA, D.; GOSWAMI, P. Artificial Neural Network and Regression Analysis Techniques for Campus Load Forecasting. In: 4th Asian Conference on Innovation in Technology (ASIANCON), (2024), p. 1–7. Available from: https://doi.org/10.1109/asiancon62057.2024.10837798. Accessed on: 21 Jan. 2025.

[29] MOON, J.; MAQSOOD, M.; SO, D. et al. Advancing ensemble learning techniques for residential building electricity consumption forecasting: Insight from explainable artificial intelligence. PLoS ONE, [s. l.], v. 19, n. 11, (2024), p. e0307654. Available from: https://doi.org/10.1371/journal.pone.0307654. Accessed on: 21 Jan. 2025.

[30] GONZÁLEZ, S.; GARCÍA, S.; DEL SER, J.; ROKACH, L.; HERRERA, F. A practical tutorial on bagging and boosting based ensembles for machine learning: Algorithms, software tools, performance study, practical perspectives and opportunities. Information Fusion, [s. l.], v. 64, (2020), p. 205–237. Available from: https://doi.org/10.1016/j.inffus.2020.07.007. Accessed on: 21 Jan. 2025.

[31] MIENYE, I. D.; SUN, Y. A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. IEEE Access, [s. l.], v. 10, (2022), p. 99129-99149. Available from: https://doi.org/10.1109/ACCESS.2022.3207287. Accessed on: 21 Jan. 2025.

[32] SHAFIUZZAMAN, M.; ISLAM, M. S.; BASHAR, T. M. R. et al. Enhanced Very Short-Term Load Forecasting with Multi-Lag Feature Engineering and Prophet-XGBoost-CatBoost Architecture. Energy, [s. l.], (2025), p. 137981. Available from: https://doi.org/10.1016/j.energy.2025.137981. Accessed on: 21 Jan. 2025.

[33] SHABBIR, N.; AHMADIAHANGAR, R.; ROSIN, A. et al. XgBoost based Short-term Electrical Load Forecasting Considering Trends & Periodicity in Historical Data. In: IEEE International Conference on Energy Technologies for Future Grids (ETFG), (2023), p. 1–6. Available from: https://doi.org/10.1109/etfg55873.2023.10407926. Accessed on: 21 Jan. 2025.

[34] BURNWAL, Y.; JAISWAL, R. C. A comprehensive survey on prediction models and the impact of XGBoost. International Journal for Research in Applied Science and Engineering Technology, [s. l.], v.

11, n. 12, (2023), p. 1560–1566. Available from: https://doi.org/10.22214/ijraset.2023.57625. Accessed on: 21 Jan. 2025.

[35] SI, B.; NI, Z.; XU, J.; LI, Y.; LIU, F. Interactive effects of hyperparameter optimization techniques and data characteristics on the performance of machine learning algorithms for building energy metamodeling. Case Studies in Thermal Engineering, [s. l.], (2024), p. 104124. Available from: https://doi.org/10.1016/j.csite.2024.104124. Accessed on: 21 Jan. 2025.

[36] LIU, H.; LI, Z.; LI, C.; SHAO, L.; LI, J. Research and application of short-term load forecasting based on CEEMDAN-LSTM modeling. Energy Reports, [s. l.], v. 12, (2024), p. 2144–2155. Available from: https://doi.org/10.1016/j.egyr.2024.08.035. Accessed on: 21 Jan. 2025.

[37] KANDADI, T.; SHANKARLINGAM, G. Drawbacks of LSTM Algorithm: A Case Study. [S. l.]: SSRN, (2025). Available from: https://doi.org/10.2139/ssrn.5080605. Accessed on: 21 Jan. 2025.

[38] VAN HOUDT, G.; MOSQUERA, C.; NÁPOLES, G. A review on the long short-term memory model. Artificial Intelligence Review, [s. l.], v. 53, n. 8, (2020). Available from: https://doi.org/10.1007/s10462-020-09838-1. Accessed on: 21 Jan. 2025.

[39] GUO, W.; LIU, S.; WENG, L.; LIANG, X. Power Grid Load Forecasting Using a CNN-LSTM Network Based on a Multi-Modal Attention Mechanism. Applied Sciences, [s. l.], v. 15, n. 5, (2025), p. 2435. Available from: https://doi.org/10.3390/app15052435. Accessed on: 21 Jan. 2025.

[40] WAHEED, W.; XU, Q.; AURANGZEB, M. et al. Empowering Data-driven Load Forecasting by Leveraging Long Short-Term Memory Recurrent Neural Networks. Heliyon, [s. l.], v. 10, n. 24, (2024), p. e40934. Available from: https://doi.org/10.1016/j.heliyon.2024.e40934. Accessed on: 21 Jan. 2025.

[41] SHAWON, S. M.; HAIDER, S. N.; BARUA, A. et al. Hybrid CNN-LSTM model for urban energy load forecasting with IGA-XAI for smart grids: Peak and off-peak variability insights. Results in Engineering, [s. l.], v. 28, (2025), p. 107245. Available from: https://doi.org/10.1016/j.rineng.2025.107245. Accessed on: 21 Jan. 2025.

[42] SALMAN, D.; DIREKOGLU, C.; KUSAF, M.; FAHRIOGLU, M. Hybrid deep learning models for time series forecasting of solar power. Neural Computing & Applications, [s. l.], (2024). Available from: https://doi.org/10.1007/s00521-024-09558-5. Accessed on: 21 Jan. 2025.

[43] KIM, B.-J.; NAM, I.-W. A Review of Hybrid LSTM Models in Smart Cities. Processes, [s. l.], v. 13, n. 7, (2025), p. 2298. Available from: https://doi.org/10.3390/pr13072298. Accessed on: 21 Jan. 2025.

[44] LU, N.; OUYANG, Q.; LI, Y.; ZOU, C. Electrical Load Forecasting Model Using Hybrid LSTM Neural Networks with Online Correction. [S. l.]: Arxiv, (2024). Available from: https://arxiv.org/html/2403.03898v1. Accessed on: 21 Jan. 2025.

[45] ABUKU, M.; JANSSEN, H.; ROELS, S. Impact of wind-driven rain on historic brick wall buildings in a moderately cold and humid climate: Numerical analyses of mould growth risk, indoor climate and energy consumption. Energy and Buildings, [s. l.], v. 41, n. 1, (2009), p. 101–110. Available from: https://doi.org/10.1016/j.enbuild.2008.07.011. Accessed on: 21 Jan. 2025.

[46] FAN, C.; CHEN, M.; WANG, X.; WANG, J.; HUANG, B. A Review on Data Preprocessing Techniques Toward Efficient and Reliable Knowledge Discovery From Building Operational Data. Frontiers in Energy Research, [s. l.], v. 9, (2021). Available from: https://doi.org/10.3389/fenrg.2021.652801. Accessed on: 21 Jan. 2025.

[47] ZHAO, T.; SUN, Y.; CHAI, Z.; LI, K. An outlier management framework for building performance data and its application to the power consumption data of building energy systems in non-residential buildings. Journal of Building Engineering, [s. l.], (2022), p. 105688. Available from: https://doi.org/10.1016/j.jobe.2022.105688. Accessed on: 21 Jan. 2025.

[48] MILÁ MURILLO, A. Building energy load profile prediction. (2024). Master's thesis (Master's in Energy Engineering) – Universitat Politècnica de Catalunya, Barcelona, 2024. Accessed on: 21 Jan. 2025.

[49] ALI, S.; AKHLAQ, F.; IMRAN, A. S. et al. The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. Computers in Biology and Medicine, [s. l.], (2023), p. 107555. Available from: https://doi.org/10.1016/j.compbiomed.2023.107555. Accessed on: 21 Jan. 2025.

[50] KIM, Y.-S.; KIM, M. K.; FU, N. et al. Investigating the Impact of Data Normalization Methods on Predicting Electricity Consumption in a Building Using different Artificial Neural Network Models. Sustainable Cities and Society, [s. l.], v. 118, (2024), p. 105570. Available from: https://doi.org/10.1016/j.scs.2024.105570. Accessed on: 21 Jan. 2025.

[51] SINGH, D.; SINGH, B. Investigating the impact of data normalization on classification performance. Applied Soft Computing, [s. l.], v. 97, (2019), p. 105524. Available from: https://doi.org/10.1016/j.asoc.2019.105524. Accessed on: 21 Jan. 2025.

[52] BLOKHINTSEV, L. D.; SAVIN, D. A. Study of the Influence of Different Methods of Taking into Account the Coulomb Interaction on Determining Asymptotic Normalization Coefficients within the Framework of Exactly Solvable Model. Physics of Atomic Nuclei, [s. l.], v. 84, n. 4, (2021), p. 401–407. Available from: https://doi.org/10.1134/s1063778821040098. Accessed on: 21 Jan. 2025.

[53] SHAPI, M. K. M.; RAMLI, N. A.; AWALIN, L. J. Energy Consumption Prediction by Using Machine Learning for Smart Building: Case Study in Malaysia. Developments in the Built Environment, [s. l.], v. 5, (2020), p. 100037. Available from: https://doi.org/10.1016/j.dibe.2020.100037. Accessed on: 21 Jan. 2025.

[54] ARLOT, S.; CELISSE, A. A survey of cross-validation procedures for model selection. Statistics Surveys, [s. l.], v. 4, (2010), p. 40–79. Available from: https://doi.org/10.1214/09-ss054. Accessed on: 21 Jan. 2025.

[55] HASANOV, M.; WOLTER, M.; GLENDE, E. Time Series Data Splitting for Short-Term Load Forecasting. In: Power and Energy Student Summit (PESS + PELSS), Kassel, (2022), p. 1-6. Accessed on: 21 Jan. 2025.

[56] KHALID, R.; JAVAID, N. A survey on hyperparameters optimization algorithms of forecasting models in smart grid. Sustainable Cities and Society, [s. l.], v. 61, (2020), p. 102275. Available from: https://doi.org/10.1016/j.scs.2020.102275. Accessed on: 21 Jan. 2025.

[57] HOSSAIN, M. R.; TIMMER, D.; MOYA, H. Machine learning model optimization with hyper-parameter tuning approach. [S. l.]: ICAETA, (2021). Available from: https://www.researchgate.net/publication/354495368. Accessed on: 21 Jan. 2025.

[58] BISCHL, B.; BINDER, M.; LANG, M. et al. Hyperparameter optimization: Foundations, algorithms, best practices, and open challenges. WIREs Data Mining and Knowledge Discovery, [s. l.], v. 13, n. 2, (2023). Available from: https://doi.org/10.1002/widm.1484. Accessed on: 21 Jan. 2025.

[59] HUTTER, F.; HOOS, H. H.; LEYTON-BROWN, K. Sequential Model-Based Optimization for General Algorithm Configuration. Lecture Notes in Computer Science, [s. l.], (2011), p. 507–523. Available from: https://doi.org/10.1007/978-3-642-25566-3_40. Accessed on: 21 Jan. 2025.

[60] SNOEK, J.; LAROCHELLE, H.; ADAMS, R. P. Practical Bayesian Optimization of Machine Learning Algorithms. [S. l.]: NIPS, (2012). Available from: https://www.researchgate.net/publication/225307589. Accessed on: 21 Jan. 2025.

[61] DEB, K. Multi-objective Optimisation Using Evolutionary Algorithms: An Introduction. Multi-Objective Evolutionary Optimisation for Product Design and Manufacturing, [s. l.], (2011), p. 3–34. Available from: https://doi.org/10.1007/978-0-85729-652-8_1. Accessed on: 21 Jan. 2025.

[62] AMIRABADI, M. A.; KAHAEI, M. H.; NEZAMALHOSSEINI, S. A. Novel suboptimal approaches for hyperparameter tuning of deep neural network. Physical Communication, [s. l.], v. 41, (2020), p. 101057. Available from: https://doi.org/10.1016/j.phycom.2020.101057. Accessed on: 21 Jan. 2025.

[63] PERERA, D.; SIRIMANNA, M. P. G. A novel simulation based evolutionary algorithm to optimize building envelope for energy efficient buildings. In: ICIAFS, (2014). Available from: https://doi.org/10.1109/iciafs.2014.7069623. Accessed on: 21 Jan. 2025.

[64] HOSAMO, H.; MAZZETTO, S. Performance Evaluation of Machine Learning Models for Predicting Energy Consumption and Occupant Dissatisfaction in Buildings. Buildings, [s. l.], v. 15, n. 1, (2024), p. 39. Available from: https://doi.org/10.3390/buildings15010039. Accessed on: 21 Jan. 2025.

[65] ELHABYB, K.; BAINA, A.; BELLAFKIH, M.; DEIFALLA, A. F. Machine Learning Algorithms for Predicting Energy Consumption in Educational Buildings. International Journal of Energy Research, [s. l.], (2024), p. 1–19. Available from: https://doi.org/10.1155/2024/6812425. Accessed on: 21 Jan. 2025.

[66] LUMUMBA, V.; KIPROTICH, D.; MPAINE, M.; MAKENA, N.; KAVITA, M. Comparative Analysis of Cross-Validation Techniques: LOOCV, K-folds Cross-Validation, and Repeated K-folds Cross-Validation in Machine Learning Models. American Journal of Theoretical and Applied Statistics, [s. l.], v. 13, n. 5, (2024), p. 127–137. Available from: https://doi.org/10.11648/j.ajtas.20241305.13. Accessed on: 21 Jan. 2025.

[67] RUNGE, J.; ZMEUREANU, R. Forecasting Energy Use in Buildings Using Artificial Neural Networks: A Review. Energies, [s. l.], v. 12, n. 17, (2019), p. 3254. Available from: https://doi.org/10.3390/en12173254. Accessed on: 21 Jan. 2025.

[68] CAPOZZOLI, A. et al. Estimation Models of Heating Energy Consumption in Schools for Local Authorities Planning. Energy and Buildings, [s. l.], v. 105, (2015), p. 302–313. Available from: https://doi.org/10.1016/j.enbuild.2015.07.024. Accessed on: 21 Jan. 2025.

[69] AMBER, K. P. et al. Energy Consumption Forecasting for University Sector Buildings. Energies, [s. l.], v. 10, n. 10, (2017), p. 1579. Available from: https://doi.org/10.3390/en10101579. Accessed on: 21 Jan. 2025.

[70] TARIQ, R. et al. Complex Artificial Intelligence Models for Energy Sustainability in Educational Buildings. Scientific Reports, [s. l.], v. 14, n. 1, (2024), p. 15020. Available from: https://doi.org/10.1038/s41598-024-65727-5. Accessed on: 21 Jan. 2025.

[71] DOIPHODE, G.; NAJAFI, H. A Machine Learning Based Approach for Energy Consumption Forecasting in K-12 Schools. [S. l.]: ASME, (2020). Available from: https://doi.org/10.1115/imece2020-24128. Accessed on: 21 Jan. 2025.

[72] GERALDI, M. S.; GHISI, E. Mapping the Energy Usage in Brazilian Public Schools. Energy and Buildings, [s. l.], v. 224, (2020), p. 110209. Available from: https://doi.org/10.1016/j.enbuild.2020.110209. Accessed on: 21 Jan. 2025.

[73] KHAN, S. et al. Optimizing Load Demand Forecasting in Educational Buildings Using Quantum-Inspired Particle Swarm Optimization (QPSO) with Recurrent Neural Networks (RNNs): A Seasonal Approach. Scientific Reports, [s. l.], v. 15, n. 1, (2025). Available from: https://doi.org/10.1038/s41598-025-04301-z. Accessed on: 21 Jan. 2025.

[74] VILLANO, F. et al. A Review on Machine/Deep Learning Techniques Applied to Building Energy Simulation, Optimization and Management. Thermo, [s. l.], v. 4, n. 1, (2024), p. 100–139. Available from: https://doi.org/10.3390/thermo4010008. Accessed on: 21 Jan. 2025.

[75] ZHANG, L. et al. A Review of Machine Learning in Building Load Prediction. Applied Energy, [s. l.], v. 285, (2021), p. 116452. Available from: https://doi.org/10.1016/j.apenergy.2021.116452. Accessed on: 21 Jan. 2025.

[76] PATSAKOS, I. et al. A Survey on Deep Learning for Building Load Forecasting. Journal of Sensors, [s. l.], v. 2022, (2022), p. 1–25. Available from: https://doi.org/10.1155/2022/1008491. Accessed on: 21 Jan. 2025.

.