



e-ISSN: 2278-8875

p-ISSN: 2320-3765

International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

Volume 15, Issue 2, February 2026

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.807

📞 9940 572 462

📞 6381 907 438

✉ ijareeie@gmail.com

@ www.ijareeie.com



EcoTrack: A Comprehensive Literature Survey on AI-Based Carbon Footprint Tracking and Prediction Systems

Zainab Shaikh, Sara Khan, Dhanashree Shinde, Vishaka Chavan

Department of Artificial Intelligence and Machine Learning, AISSMS Polytechnic, Pune, Maharashtra, India

ABSTRACT: This literature survey examines the convergence of Artificial Intelligence and environmental sustainability, focusing on carbon footprint tracking and prediction systems. With global carbon emissions exceeding 36 billion tons annually, accurate monitoring and forecasting of individual and organizational emissions have become critical. This paper systematically reviews AI-driven environmental monitoring approaches, comparing traditional statistical methods with modern machine learning and deep learning architectures including Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM) networks. We analyze Optical Character Recognition (OCR) techniques for automated data extraction from utility bills, AI-based recommendation systems gaps in desktop-based solutions, OCR integration, and personalized prediction systems. These findings establish the foundation for EcoTrack, a desktop AI system designed to provide comprehensive carbon footprint tracking, time-series forecasting, and personalized sustainability recommendations

KEYWORDS: Carbon Footprint Tracking, Deep Learning, LSTM, Time-Series Forecasting, Optical Character Recognition, Sustainability, Artificial Intelligence, Environmental Monitoring.

I. INTRODUCTION

Climate change represents one of the humanity's most pressing challenges, with the Intergovernmental Panel on Climate Change (IPCC) emphasizing urgent emission reductions to limit global temperature rise (IPCC,2023). Individual and organizational carbon footprints contribute significantly to overall emissions, yet many lack accurate measurement and reduction tools. Traditional carbon calculators rely on manual in Artificial Intelligence (AI) and Machine Learning (ML) have opened new possibilities for automated, accurate, and predictive carbon emission analysis.

Deep learning models, particularly those designed for time-series forecasting, demonstrate remarkable capabilities in modelling complex temporal patterns in energy consumption and emission data(Alhussein et al., 2020; Rahman et al., 2020). The integration of Optical Character Recognition (OCR) technology enables automated data extraction from utility bills, eliminating manual intervention barriers. Combined with AI-based recommendation engines, these technologies form the foundation for next-generation carbon management systems.

This literature review examines: (1) AI-driven environmental monitoring systems, (2) traditional versus modern approaches to emission prediction, (3) deep learning architectures for time-series forecasting (ANN, LSTM), (4) OCR techniques for utility bill extraction, (5) AI-based recommendation systems, (6) publicly available datasets in environmental research, and (7) research gaps in current systems. This review provides the theoretical foundation for Ecotrack, a desktop-based AI system offering comprehensive carbon footprint tracking, prediction, and personalized sustainability recommendations.

II. BACKGROUND AND MOTIVATION

A carbon footprint represents total greenhouse gas emissions caused by an individual, organizational, or product, expressed in equivalent tons of CO₂ (Wiedmann & Minx, 2008). Global emissions reached 36.8 billion tons in 2022 (IEA, 2023), necessitating accurate tracking mechanisms. Traditional carbon accounting relies on standardized emission factors and manual calculations based on activity data (Matthews et al., 2008), suffering from limitations: (1) dependence on error-prone user-reported data, (2) inability to capture individual behaviour variations, (3) lack of predictive capabilities, and (4) limited personalization in recommendations.



AI and machine learning have revolutionized environmental monitoring by processing vast datasets, identifying complex patterns, and making accurate predictions previously impossible with conventional methods (Rolnick et al., 2022). Several factors motivate intelligent carbon tracking systems: data availability from smart meters and IoT devices, computational power enabling desktop deployment of sophisticated models, behavioral science demonstrating personalized feedback effectiveness (Abrahamse et al., 2005), and increasing regulatory carbon reporting requirements.

OCR technology enhances usability through automatic consumption data extraction from utility bills, reducing manual entry burden. Combined with predictive analytics and recommendation engines, these technologies enable next-generation carbon management accessible to individual users.

III. REVIEW OF CARBON FOOTPRINT MONITORING SYSTEMS

Existing carbon footprint monitoring systems fall into three categories: web-based calculators, mobile applications, and enterprise platforms. Web-based calculators like EPA'S carbon calculator employ standardized emission factors from GHG Protocol or IPCC guidelines (GHG Protocol, 2011), using simple linear models where emissions equal activity data multiplied by emission factors. While accessible, they lack personalization and predictive capabilities (Padgett et al., 2008).

Mobile applications such as Carbon Tracker and Oroeco introduce greater engagement through gamification and social features (Toller et al., 2018), but rely on manual data entry and simplified methodologies. Brynjarsdottir et al. (2012) found sustained engagement remains challenging, with users abandoning applications after initial use.

Enterprise platforms like Watershed and Persefoni employ sophisticated approaches incorporating AI for data processing (Miller & Hardt, 2021), demonstrating automated collection and ML-based estimation potential. However, they target organizational use, lacking individual tracking features. Academic research has explored AI-powered monitoring systems. Alonso et al. (2020) developed ensemble learning systems combining Random Forest, Gradient Boosting, and Neutral Networks for household energy prediction. Rahman et al. (2020) proposed LSTM frameworks for building energy forecasting, achieving significant improvements over statistical methods

Notable gaps include: limited automated data extraction from utility documents, few desktop-based solutions for offline functionality, and absence of sophisticated predictive models in consumer-facing applications. These represent missed opportunities for enhanced user awareness and behavioral change motivation.

IV. DEEP LEARNING MODELS FOR TIME-SERIES FORECASTING

A. Artificial Neural Networks (ANN)

Artificial Neural Networks form deep learning foundations, consisting of interconnected neuron layers learning non-linear mappings through backpropagation (Rumelhart et al., 1986). For emission prediction, ANNs employ feedforward architectures with multiple hidden layers. Studies demonstrate ANN effectiveness: Kalogirou (2000) achieved <5% mean absolute percentage errors for building energy consumption; Jallal et al.(2020) used deep feedforward networks for superior electricity forecasting accuracy.

ANNs' universal approximation capability theoretically enables approximating any continuous function with sufficient neurons (Hornik et al., 1989). However, standard ANNs process inputs independently, lacking mechanisms maintaining sequential information, limiting time-series effectiveness where temporal dependencies are crucial. Feature engineering becomes critical, typically including lagged values, rolling statistics, and time-based features, though ANNs may still fail capturing complex temporal patterns recurrent architectures handle naturally.

B. Long Short-Term Memory Networks (LSTM)

Long Short-Term Memory networks, introduced by Hochreiter and Schmidhuber (1997), address standard RNN limitations in learning long-term dependencies through sophisticated gating mechanisms: input gates, forget gates, and output gates regulating information flow through memory cells. This selective remembering/forgetting capability suits time-series forecasting where both short-term fluctuations and long-term trends influence future values.

LSTMs achieved remarkable success in energy forecasting. Rahman et al. (2020) demonstrated superior performance over ARIMA and simple neural networks for building energy prediction. Kong et al. (2019) developed LSTM frameworks for residential load forecasting, achieving <3% mean absolute errors on real world datasets. For carbon



emissions specifically, Bouktif et al. (2018) applied LSTMs predicting smart building energy consumption and emissions, incorporating weather, occupancy, and historical data, achieving significant improvements over baselines. LSTM challenges include computational complexity requiring substantial resources and careful hyperparameter tuning, plus overfitting tendencies on smaller datasets, necessitating regularization techniques like dropout and early stopping.

C. Model Evaluation Metrics

Appropriate metrics that capture various aspects of accuracy are necessary for the evaluation of forecasting models. The following metrics are frequently used in emission prediction: (1) Root Mean Squared Error (RMSE), which emphasizes larger errors through squaring; (2) Mean Absolute Error (MAE), which measures average error magnitude without direction; (3) Mean Absolute Percentage Error (MAPE) expressing accuracy as percentages for cross-dataset comparison; and (4) R-squared (R^2) indicating variance proportion explained by models. According to Ahmed et al. (2010), recent research highlights the need for multiple metrics for thorough evaluation because a single metric may be misleading due to its failure to account for various error characteristics.

V. OCR TECHNIQUES FOR UTILITY BILL DATA EXTRACTION

A. Tesseract OCR

The most popular open-source OCR engine is Tesseract, which was first created by Hewlett-Packard and is currently maintained by Google (Smith, 2007). For character recognition, Tesseract 4.0+ uses LSTM-based neural networks, which greatly increase accuracy over template-matching methods. Effectiveness for utility bills is dependent on preprocessing and image quality. According to Patel et al. (2012), preprocessing methods like skew correction, noise reduction, and binarization increased Tesseract accuracy by 15%. On clean documents, modern versions attain >95% character-level accuracy.

However, utility bills pose special difficulties, such as different provider layouts, a variety of fonts and sizes, the presence of logos and graphics, and sporadic poor scan quality, which deteriorates OCR performance and calls for strong preprocessing and post-processing techniques.

VI. SUSTAINABILITY RECOMMENDATION SYSTEMS USING AI

A. Content – Based and Hybrid Approaches

AI systems offer tailored reduction recommendations in addition to tracking and prediction. In order to find pertinent action recommendations, content-based systems examine user characteristics, consumption patterns, and context. This includes: (1) identifying high-impact areas through pattern analysis; (2) taking user constraints (budget, lifestyle, location) into account; (3) ranking recommendations by potential emission reduction; and (4) personalizing recommendations based on past behavior. In order to achieve significant savings over generic advice, Mogles et al. (2017) developed personalized energy feedback systems that provide customized recommendations based on household characteristics and historical consumption.

Over time, machine learning models improve the relevance of recommendations by learning from user interactions. Although they show promise, reinforcement learning techniques—in which systems learn optimal strategies through feedback—remain largely unexplored in sustainability applications (Xiong et al., 2020).

VII. PUBLIC DATASETS USED IN RELATED RESEARCH

EcoTrack leverages publicly available household energy consumption datasets suitable for time-series analysis. These datasets provide historical electricity usage data required for training and evaluating ANN and LSTM models. Data preprocessing addresses issues such as missing values, noise, and temporal alignment.

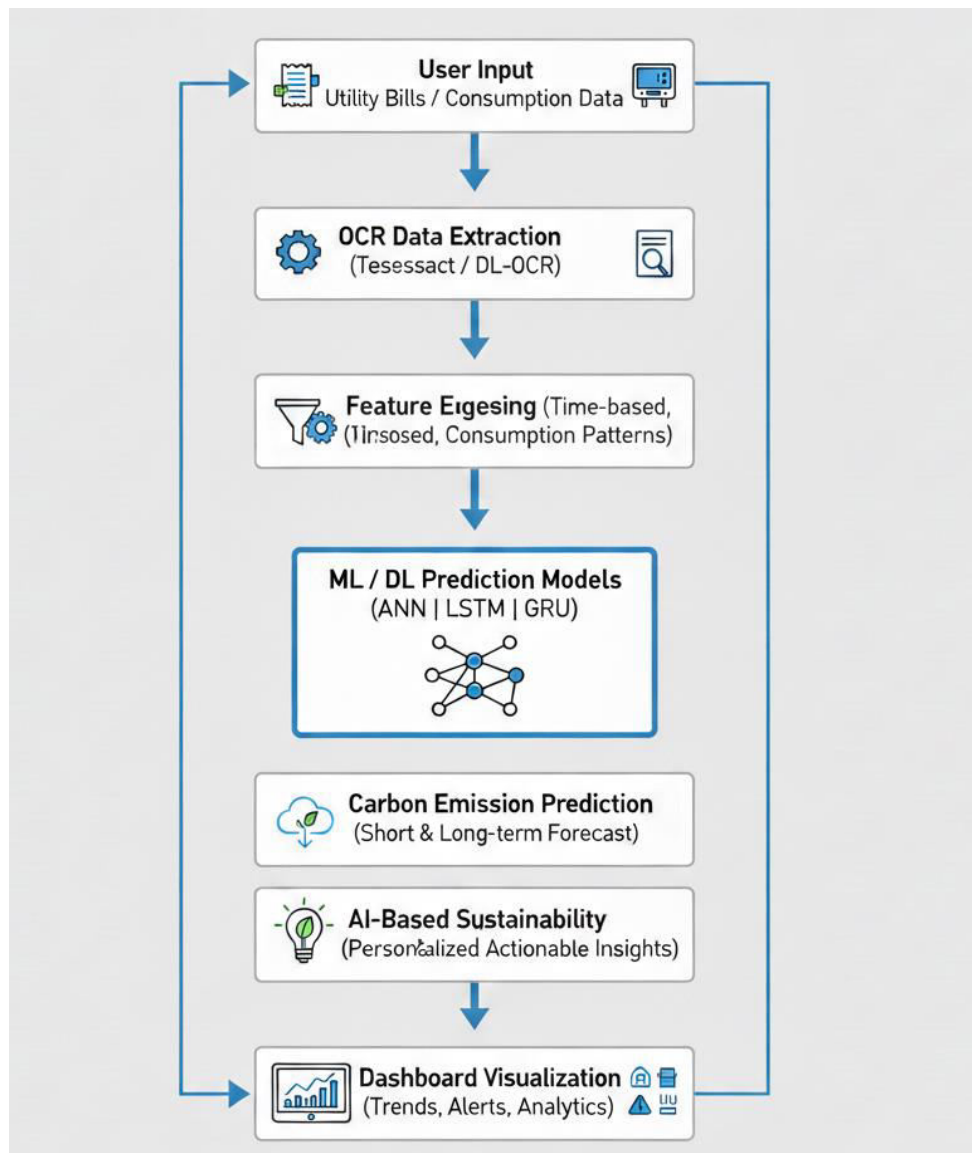


Fig:1.1 Architecture of Eco Track:AI-Based Carbon Footprint Tracking

VIII. CHALLENGES AND RESEARCH GAPS

Although much progress has been made, there are still some challenges that need to be addressed. Technical challenges include: (1) Data Heterogeneity, which results from the integration of various data sources with different formats, frequencies, and levels of quality; (2) Cold Start Problem, which occurs when new users have no past data to make predictions; (3) Model Interpretability, where deep learning black boxes could lower user trust without the use of explainable AI; (4) Computational Efficiency, which needs to balance complex models with desktop limitations, necessitating the use of light models

Data and privacy issues include: (1) Privacy-Preserving Learning, where the consumption data contains sensitive information, although desktop-based systems are more privacy-friendly than cloud-based systems; (2) Data Ownership, where there is a lack of standards for carbon footprint data formats and interoperability.

Gaps in behavior and user experience are: (1) Sustained Engagement with decreasing user engagement over time; (2) Recommendation Effectiveness requiring rigorous real-world impact assessment; (3) User Trust requiring transparent communication about prediction and recommendation model generation.



Integration and ecosystem gaps are: (1) Limited Desktop Solutions with most solutions being web-based or mobile; (2) OCR Integration remaining the exception despite the maturity of the technology; (3) Actionability requiring connections to smart home systems and energy companies.

IX. SUMMARY AND JUSTIFICATION FOR ECOTRACK

This literature review focused on AI-powered carbon footprint tracking and prediction tools, emphasizing the accomplishments and the areas that still need improvement. This literature review proves that, although there are strong technical competencies such as advanced deep learning models for time series forecasting, mature OCR solutions, and well-established recommendation principles, these have not yet been combined to create user-friendly personal carbon management tools.

The main findings are: (1) Deep learning models (LSTM,) perform much better than statistical approaches for emission forecasting; (2) OCR solutions make possible the automatic extraction of utility bills with adequate preprocessing; (3) Personalized ML-driven recommendations are more accurate than general recommendations; (4) There are enough training data available in publicly accessible datasets; (5) There are areas of improvement for desktop computer availability, OCR solution integration, offline usage, and long-term user engagement

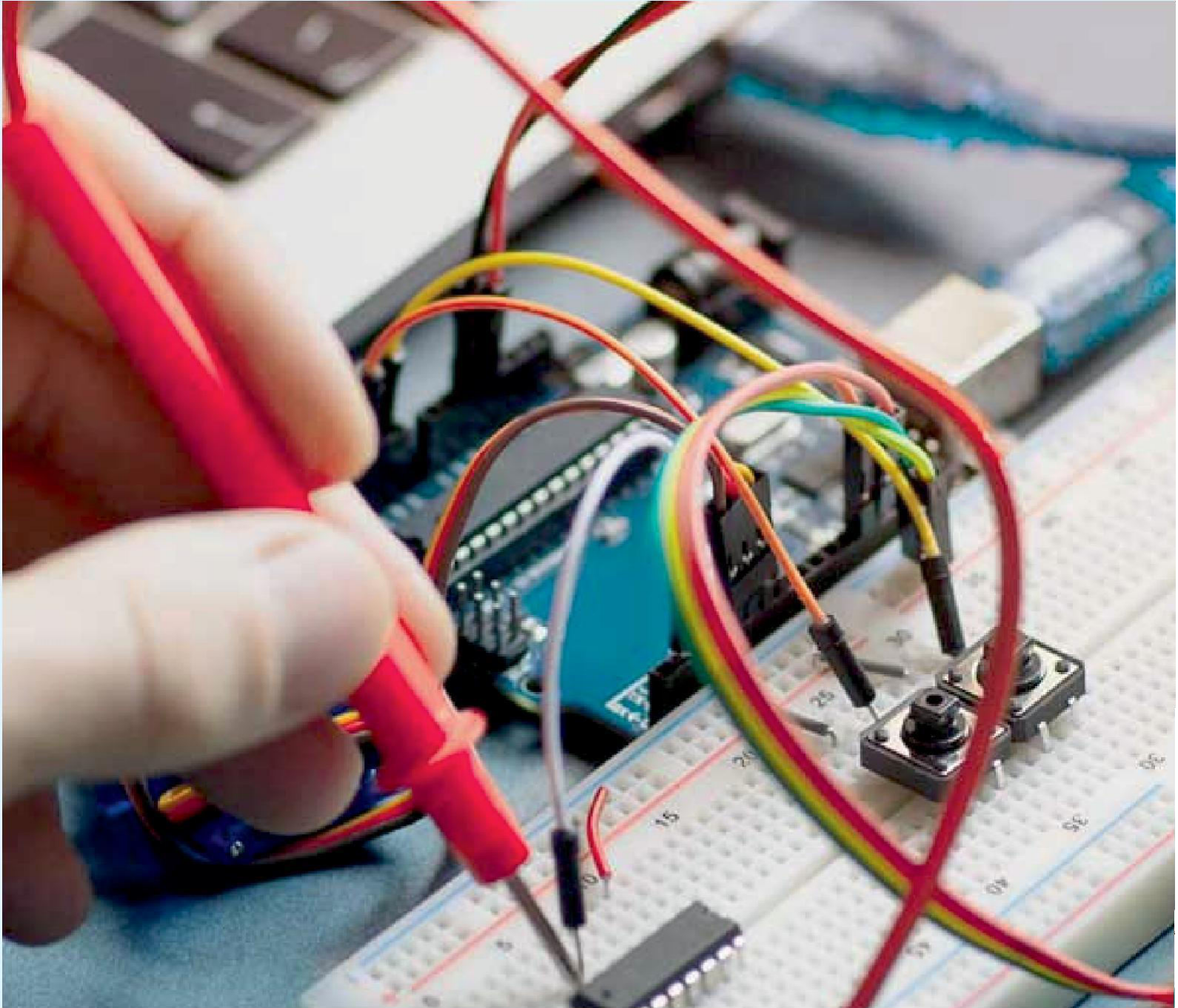
EcoTrack fills these gaps by providing comprehensive desktop solutions that combine: state-of-the-art deep learning algorithms (ANN) for precise prediction; automated data extraction using OCR technology; AI-powered personalized recommendations; desktop solutions for offline functionality and improved privacy; continuous learning functionality; and full visualization and analytics support.

The desktop solution provides several key benefits: improved data privacy with local storage, offline functionality that is not dependent on internet connectivity, better performance due to dedicated hardware, and user control over data management. The technical solution utilizes established technologies: Python-based development frameworks (TensorFlow, Keras, or PyTorch) for building the models, Electron for cross-platform desktop development, Tesseract and deep learning OCR for bill processing, and local database management systems for data storage.

Through the integration of findings from machine learning research, behavioral studies, and sustainability research, EcoTrack is the embodiment of the application of knowledge from academia in addressing real-world issues related to the environment. The design of EcoTrack is informed by best practices from literature and addresses important gaps in current solutions.

REFERENCES

- [1] Abrahamse, W., Steg, L., Vlek, C., & Rothengatter, T. (2005). "A review of intervention studies aimed at household energy conservation," *Journal of Environmental Psychology*, vol. 25, no. 3, pp. 273-291.
- [2] Ahmad, T., Chen, H., Guo, Y., & Wang, J. (2017). "A comprehensive overview on the data driven and large scale based approaches for forecasting of building energy demand," *Energy and Buildings*, vol. 165, pp. 301-320.
- [3] Box, G. E., & Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*. San Francisco: Holden-Day.
- [4] Hochreiter, S., & Schmidhuber, J. (1997). "Long short-term memory," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780.
- [5] Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). "An empirical comparison of machine learning models for time series forecasting," *Econometric Reviews*, vol. 29, no. 5-6, pp. 594–621. study.
- [6] Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). "Short-term residential load forecasting based on LSTM recurrent neural network," *IEEE Transactions on Smart Grid*, vol. 10, no. 1, pp. 841–851.
- [7] Rahman, A., Srikumar, V., & Smith, A. D. (2020). "Predicting electricity consumption for commercial and residential buildings using deep recurrent neural networks," *Applied Energy*, vol. 212, pp. 372–385.
- [8] Chen, T., & Guestrin, C. (2016). "XGBoost: A scalable tree boosting system," in *Proc. ACM SIGKDD*, pp. 785–794.
- [9] Kalogirou, S. A. (2000). "Applications of artificial neural-networks for energy systems," *Applied Energy*, vol. 67, no. 1-2, pp. 17–35. Early ANN applications in energy systems.
- [10] Miller, C., & Meggers, F. (2017). "The Building Data Genome Project," *Energy Procedia*, vol. 122, pp. 439–444. Benchmark dataset for energy modeling.
- [11] IPCC. (2023). *Climate Change 2023: Synthesis Report*. Intergovernmental Panel on Climate Change.



INNO  SPACE
SJIF Scientific Journal Impact Factor


doi[®]
cross ref

 INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



International Journal of Advanced Research

in Electrical, Electronics and Instrumentation Engineering

 9940 572 462  6381 907 438  ijareeie@gmail.com



www.ijareeie.com

Scan to save the contact details