



# IJRASET

International Journal For Research in  
Applied Science and Engineering Technology



---

# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

---

**Volume:** 14    **Issue:** IV    **Month of publication:** April 2026

**DOI:** <https://doi.org/10.22214/ijraset.2026.81260>

[www.ijraset.com](http://www.ijraset.com)

Call:  08813907089

E-mail ID: [ijraset@gmail.com](mailto:ijraset@gmail.com)

# A Survey on Energy Management for Smart Grids Using Deep Lambda Architecture

Shwetha AB, Shiva Balaji S Katta, Surabhi Desai, Varshini Patil, Vishalakshi Koti  
Dept. of CSE Sapthagiri College of Engineering

**Abstract:** Smart grids are popping up everywhere now, and that brings in a flood of data from all sorts of sources in energy systems. Some of it is neat and organized, but a lot is just messy unstructured stuff. It pushes the need for ways to store everything without losing track, process it fast, and even analyze right away. Traditional setups for managing energy often struggle with this, like they can't scale easily or handle delays well, and predicting things for power networks gets tricky too.

We are looking at recent ideas in smart grid energy management here, focusing on things like Data Lakes for storage and that Lambda Architecture approach. It seems like they mix in big data tools with machine learning, including deeper stuff like CNNs and LSTM models, plus different forecasting methods. Batch processing works alongside real-time stuff, and that combo helps with load forecasts, better demand responses, and distributing energy more evenly. Not sure if it covers everything perfectly, but it stands out how these pieces team up.

Current systems still have issues, you know, like not scaling right or missing that quick response at the edge. So moving toward these smarter setups feels important. Pulling together Data Lakes for storage, Lambda for handling layers, and AI for predictions could make future grids way more efficient and reliable, maybe even greener over time. Some people might argue it's not fully there yet.

The hybrid models get technical fast, and I think that part is a bit unclear without digging deeper. Overall though, the way integration ties it all seems promising for energy handling, even as it evolves. That catches my attention mostly.

## I. INTRODUCTION

Over the past decade, the electrical power sector has evolved significantly due to rising energy demand, increased integration of renewable sources, and advances in digital technologies. Unlike traditional centralized systems, modern power grids operate in a dynamic and decentralized environment, where solar energy, wind power, electric vehicles, and smart appliances introduce new operational complexities. Ensuring stability, efficiency, and reliability has therefore become a major challenge.

Smart grids have emerged as an advanced solution, enabling real-time monitoring, bidirectional communication, and automated control. These systems rely on data-driven decision-making but generate massive and diverse datasets from sources such as smart meters, sensors, and market platforms. Traditional data management systems struggle to handle this volume, velocity, and variety, leading to scalability and latency issues.

To address these challenges, big data technologies such as Data Lakes and distributed computing frameworks are widely adopted. Data Lakes allow storage of raw, heterogeneous data, while Lambda Architecture combines batch and real-time processing to balance accuracy and low latency, making it suitable for applications like load forecasting and fault detection.

Additionally, artificial intelligence techniques—including Decision Trees, CNNs, and LSTM networks—enhance predictive analytics, anomaly detection, and energy optimization. The integration of big data and AI enables efficient, scalable, and intelligent energy management systems.

This survey focuses on recent advancements in smart grid energy management using Data Lake and Lambda Architecture, analyzing existing approaches, identifying limitations, and highlighting future research directions for building scalable and low-latency systems.

## II. BACKGROUND AND MOTIVATION

Traditional electric power systems were built on a centralized model with unidirectional energy flow, limited monitoring, and static decision-making based on predictable demand. While effective in the past, this approach lacks the flexibility required for today's dynamic energy landscape.

Modern power systems have shifted toward decentralization due to the integration of distributed energy resources like solar panels and wind turbines, along with the rise of electric vehicles and dynamic pricing.

These changes introduce variability and uncertainty, requiring real-time, intelligent, and adaptive grid operations. However, traditional infrastructures struggle with bidirectional energy flow and large-scale data handling, leading to inefficiencies and reduced reliability.

Smart grids address these challenges by enabling real-time monitoring, automation, and data-driven decision-making. However, they generate massive, diverse, and high-speed data from multiple sources, which conventional systems cannot efficiently manage.

To overcome this, advanced solutions such as Data Lakes and Lambda Architecture are used. Data Lakes allow flexible storage of raw, heterogeneous data, while Lambda Architecture combines batch and real-time processing for accurate and low-latency insights. Additionally, AI techniques like Decision Trees, CNNs, and LSTMs enhance prediction, fault detection, and energy optimization.

The integration of big data and AI provides a scalable and intelligent framework for smart grid energy management, improving efficiency, reliability, and sustainability while supporting the development of future smart cities.

### III. LITERATURE SURVEY

#### A. Energy Load Forecasting Techniques in Smart Grids: A Cross-Country Comparative Analysis

Author and Year: R. Hachache et al. (2024) Methodology: This study provides a detailed comparative evaluation of different energy load forecasting methods using smart grid datasets collected from multiple countries. It examines conventional statistical techniques such as ARIMA, alongside machine learning models like Random Forest and Support Vector Machines, and advanced deep learning approaches including CNN, LSTM, and combined CNN-LSTM architectures. The analysis focuses on how these models perform under varying geographical conditions and consumption behaviors. The findings indicate that hybrid models deliver superior results compared to standalone methods. In particular, CNN is effective in capturing spatial patterns in the data, while LSTM excels at modeling temporal relationships, and their combination leads to higher forecasting accuracy and greater reliability across diverse datasets. Limitation: Although the study offers a thorough comparison of forecasting techniques, its primary focus remains on prediction accuracy, with limited attention given to data storage and processing aspects. It does not incorporate big data frameworks such as Data Lakes or Lambda Architecture, which are crucial for managing and analyzing real-time smart grid data. Furthermore, the use of hybrid models increases computational complexity, making them more resource-intensive and potentially challenging to implement in large-scale systems.

#### B. Short-Term Electricity Load Forecasting Based on Neural-Prophet and CNN-LSTM

Author and Year: Shuai Lu and Taotao Bao (2024)

Methodology: This study introduces a hybrid forecasting approach that integrates NeuralProphet with a CNN-LSTM framework to enhance short-term electricity load prediction. NeuralProphet is used to model underlying trends and seasonal variations, while CNN is applied to capture spatial features and LSTM to learn temporal dependencies in the data. By combining these components, the model achieves improved accuracy and is better suited for handling complex, nonlinear patterns. The proposed approach is tested on real-world datasets and shows superior performance when compared to conventional methods and individual deep learning models. Limitation: The study mainly concentrates on prediction accuracy and does not consider challenges related to data management or real-time data processing. It also does not incorporate scalable big data frameworks such as Data Lakes or Lambda Architecture, which are essential for handling large-scale smart grid data. As a result, its practical applicability in real-world, large-scale smart grid environments remains limited.

#### C. Machine Learning Techniques for Load Forecasting in Smart Grids with Renewable Energy Integration

Author and Year: Cyrus Tinny Plavila (2025)

Methodology: This paper provides a comprehensive overview of machine learning and deep learning methods applied to load forecasting in smart grid systems, particularly under the influence of renewable energy integration. It examines a range of models, including ARIMA, Random Forest, CNN, LSTM, and various hybrid techniques. The study emphasizes that the inclusion of renewable energy sources increases variability and uncertainty, thereby making accurate forecasting more complex. In addition, it evaluates the performance of different models and identifies the most suitable approaches for specific operating conditions and scenarios.

Limitation: The paper is primarily survey-oriented and does not include any practical implementation or experimental validation of the discussed methods. It also does not address real-time data processing aspects and fails to present a complete system architecture tailored for smart grid applications.

#### D. *Big Data Reference Architecture for the Energy Sector*

Author and Year: K. Wehrmeister et al. (2025)

**Methodology:** This paper presents a reference architecture tailored for big data applications in the energy domain, with a particular focus on smart grid systems. It outlines a layered framework that includes components for data acquisition, storage, processing, analytics, and application services. The proposed architecture places strong emphasis on secure data sharing, seamless interoperability among diverse systems, and the ability to scale analytics as data volume grows. Additionally, it underscores the role of integrating IoT technologies, cloud computing platforms, and distributed processing frameworks to enable efficient and reliable energy management. **Limitation:** The paper is mainly conceptual in nature, emphasizing architectural design rather than practical deployment. It does not incorporate specific forecasting techniques or provide experimental validation using real-world datasets. Consequently, its direct application to real-time smart grid forecasting remains limited without additional implementation and evaluation.

#### E. *Data Lakes: A Survey of Concepts and Architectures*

Author and Year: S. Azzabi et al. (2024) **Methodology:** This paper presents a comprehensive overview of Data Lake technologies, covering key aspects such as architecture, data ingestion methods, storage mechanisms, governance strategies, and processing approaches. It examines modern frameworks, including Lambda Architecture and Lakehouse models, and explains their benefits in managing large-scale data environments. Additionally, the study addresses important challenges associated with Data Lakes, such as maintaining data quality, effective metadata management, and ensuring data security.

**Limitation:** The study is primarily theoretical in nature and does not specifically address applications within smart grid systems. It does not incorporate forecasting techniques, nor does it illustrate how Data Lake frameworks can be practically applied to energy management scenarios.

#### F. *Data Integration and Storage Strategies in Heterogeneous Analytical Systems*

Author and Year: P. Koukaras (2025)

**Methodology:** This paper reviews different data integration and storage strategies used in large-scale analytical systems. It covers techniques such as data federation, lakehouse architecture, metadata management, and interoperability between systems. The study emphasizes the importance of scalable and flexible data management solutions for handling heterogeneous data sources.

**Limitation:** The paper does not specifically address smart grid systems or energy forecasting. It also lacks implementation of machine learning or deep learning models, making it less relevant for predictive analytics in energy management.

#### G. *Enhancing Short-Term Power Load Forecasting With a TimesNet-Crossformer-LSTM Approach*

Author and Year: Jun He et al. (2024)

**Methodology:** This paper presents an advanced hybrid deep learning model that combines TimesNet, Crossformer, and LSTM for short-term load forecasting. TimesNet captures temporal patterns, Crossformer models complex dependencies in multivariate time-series data, and LSTM handles sequential learning. The model processes multidimensional smart grid data and achieves high forecasting accuracy. The integration of transformer-based architectures improves the model's ability to capture long-term dependencies and complex relationships.

**Limitation:** The model is computationally expensive and complex, requiring high processing power and resources. It also lacks integration with real-time data processing frameworks, making it difficult to deploy in practical smart grid environments with continuous data streams.

#### H. *Data Lake Lambda Architecture for Smart Grids Big Data Analytics*

Author and Year: A. A. Munshi and Y. Mohamed (2018)

**Methodology:** This paper introduces a big data framework for smart grid systems that integrates Data Lake storage with Lambda Architecture to efficiently manage large volumes of generated data. The Data Lake component allows flexible storage of diverse data types—both structured and unstructured—collected from sources such as sensors, smart meters, and operational platforms. To handle data processing, the framework employs Lambda Architecture, which utilizes both batch processing for historical data analysis and a speed layer for real-time data handling. In addition, the system incorporates analytical tools and visualization dashboards to support informed decision-making, making it well-suited for large-scale smart grid analytics and management.

Limitation: The study primarily emphasizes data architecture and does not integrate contemporary machine learning or deep learning approaches for forecasting and prediction tasks. As it is based on earlier work, it does not reflect recent advancements in AI, and it overlooks key challenges such as scaling learning models and incorporating modern neural network techniques.

#### IV. METHODOLOGY

The proposed methodology for smart grid energy management leverages a combination of Data Lake and Lambda Architecture to tackle the increasing complexities of modern power systems, including rising electricity demand, integration of renewable energy sources, large-scale data generation, and the need for real-time decision-making. Smart grid environments generate vast amounts of data from various operational components, and managing this data effectively requires scalable storage solutions, efficient processing mechanisms, and intelligent predictive models.

To address these requirements, the methodology brings together Data Lake for flexible data storage, Lambda Architecture for handling both historical and real-time data processing, and machine learning techniques such as Decision Tree and Convolutional Neural Networks (CNN). This integrated approach aims to create a reliable, scalable, and future-oriented framework for efficient smart grid energy management.

##### A. Data Acquisition and Collection

The first stage of the methodology involves collecting data from multiple smart grid sources. Modern energy systems generate information from smart meters, electricity usage records, renewable energy production units, market pricing systems, weather services, and utility operational logs. These sources provide valuable insights into energy generation, transmission, consumption, and demand behavior.

The collected data can be of different types:

- Structured Data: Numerical tables such as hourly energy usage and pricing data
- Semi-Structured Data: JSON, XML, and logs generated from devices and systems
- Unstructured Data: Text reports, maintenance records, or external documents

Continuous data collection is essential because forecasting and decision-making depend on both historical patterns and real-time changes in the grid environment.

##### B. Centralized Storage Using Data Lake

After collection, all raw data is stored in a centralized Data Lake. A Data Lake is preferred over traditional relational databases because smart grid datasets are massive, diverse, and continuously growing. Traditional databases require predefined schemas and often struggle with scalability, whereas a Data Lake stores data in raw format until processing is required.

Key benefits of using a Data Lake include:

- Storage of structured, semi-structured, and unstructured data in one platform
- Support for petabyte-scale storage
- Easy integration of multiple data sources
- Preservation of raw historical data for future analytics
- Cost-effective and scalable storage solution

Possible technologies for implementation include HDFS, AWS S3, Azure Data Lake, or Google Cloud Storage.

This stage ensures that the system can efficiently store large datasets without losing important information.

##### C. Data Preprocessing and Quality Improvement

Raw smart grid data may contain errors, missing values, duplicates, inconsistent formats, and irrelevant attributes. Therefore, preprocessing is performed before applying analytics and machine learning algorithms.

The preprocessing process includes:

- Missing Value Handling: Replacing or estimating unavailable records
- Duplicate Removal: Eliminating repeated entries
- Noise Filtering: Removing abnormal or corrupted data
- Normalization: Scaling numerical values to improve model training
- Feature Selection: Selecting relevant variables such as load, weather, time, and price

- Transformation: Converting data into a machine-readable format

For example, date and time values can be transformed into features such as hour, day, month, and season to improve demand forecasting accuracy.

Libraries such as Pandas, NumPy, and Scikit-learn are used for preprocessing tasks.

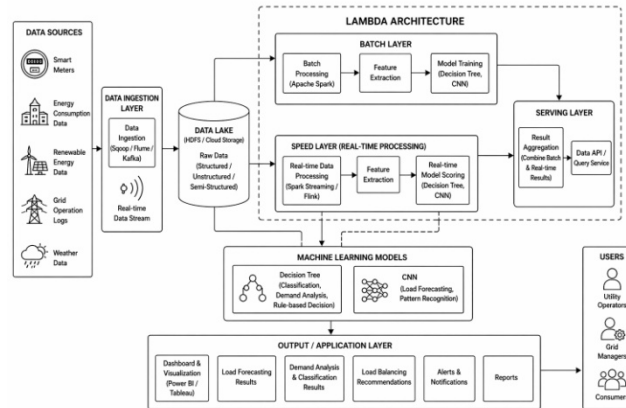


Fig. 1. Proposed System Architecture for Energy Management using DeepLake Lambda Architecture

#### D. Lambda Architecture for Data Processing

To handle both historical and real-time smart grid data efficiently, the system adopts Lambda Architecture. This architecture combines batch processing and stream processing to achieve high accuracy with low latency.

The architecture consists of three layers:

- 1) Batch Layer: The batch layer processes large volumes of historical data stored in the Data Lake. It performs deep analysis on completed datasets and generates long-term insights.

Main functions of the batch layer:

- Historical load pattern analysis
- Seasonal demand trend identification
- Consumer behavior analysis
- Long-term forecasting model training
- Energy consumption reporting

For example, the batch layer can identify that electricity demand increases during summer afternoons or winter nights.

Technologies used: Apache Hadoop, Apache Spark. The batch layer ensures accurate and comprehensive analytics by using full historical data.

- 2) Speed Layer: The speed layer is responsible for processing real-time data streams with minimal latency. It plays a crucial role in enabling immediate monitoring and rapid decision-making in smart grid systems. Unlike batch processing, this layer focuses on handling continuously incoming data and generating instant insights.

The primary functions of the speed layer include:

- Real-time monitoring of electricity demand
- Detection of sudden load spikes
- Instant identification of faults and anomalies
- Monitoring fluctuations in renewable energy generation
- Providing immediate recommendations for load balancing

For instance, if there is a sudden surge in electricity demand within a specific region, the speed layer quickly detects the change and generates alerts for system operators, allowing timely corrective actions. Common technologies used for implementing the speed layer include: Apache Kafka, Apache Spark Streaming, Apache Flink.

Overall, the speed layer ensures rapid response capabilities, which are essential for maintaining stability, reliability, and efficiency in modern smart grid operations.

3) **Serving Layer:** The serving layer acts as the interface between the processing components and end-user applications. It integrates the outputs generated by both the batch layer and the speed layer to deliver final, consistent, and queryable results. This layer is responsible for making processed data easily accessible for visualization, reporting, and decision-making purposes.

The key functions of the serving layer include:

- Providing unified query results by combining batch and real-time outputs
- Enabling fast retrieval of processed and analyzed data
- Supporting data visualization for dashboards and monitoring tools
- Offering decision support services for operators and stakeholders

By merging historical insights with real-time updates, the serving layer ensures that users receive accurate, up-to-date, and comprehensive information. This capability is essential for effective monitoring, analysis, and control in smart grid systems.

Common technologies used in the serving layer include: Apache Cassandra, Elasticsearch

Overall, the serving layer plays a critical role in delivering reliable and efficient access to analytics results, thereby supporting informed decision-making in modern energy management systems.

### E. Machine Learning Models in Smart Grid

To enhance the intelligence and adaptability of smart grid systems, machine learning models are integrated into data processing pipelines. These models enable predictive analysis, pattern recognition, and automated decision-making based on large volumes of energy data.

1) **Decision Tree Algorithm:** The Decision Tree algorithm is a supervised learning technique widely used for classification and decision-making tasks. It works by splitting data into branches based on specific conditions, ultimately forming a tree-like structure that represents decision rules. Its simplicity and interpretability make it suitable for various smart grid applications.

Applications of Decision Trees in smart grids include:

- Classification of energy demand levels (high, medium, low)
- Identification of peak load conditions
- Support for demand response strategies
- Detection of abnormal or irregular consumption patterns
- Rule-based operational decision-making

The key advantages of the Decision Tree algorithm are:

- Easy to interpret and explain
- Fast execution and low computational complexity
- Effective performance on structured datasets
- Minimal requirement for data preprocessing

For example, a Decision Tree model can classify electricity demand as high when conditions such as elevated temperature and peak usage hours (e.g., evening time) are met.

2) **Convolutional Neural Network (CNN):** Convolutional Neural Networks (CNNs) are a class of deep learning models designed to automatically extract meaningful features from data. Although initially developed for image processing tasks, CNNs have demonstrated strong performance in analyzing time-series data, including energy consumption patterns.

Applications of CNN in smart grids include:

- Short-term load forecasting
- Prediction of peak electricity demand
- Recognition of consumption patterns
- Forecasting of renewable energy generation

Extraction of hidden and complex data features The main advantages of CNN models are:

- High accuracy in forecasting tasks
- Automatic feature extraction without manual intervention
- Strong capability in identifying complex patterns
- Improved performance compared to many traditional techniques
- Common frameworks used for implementing CNN models include:

- TensorFlow
- Keras
- PyTorch

For instance, a CNN model can analyze historical energy consumption data to learn daily, weekly, or seasonal usage patterns, thereby improving prediction accuracy in smart grid systems.

#### F. Forecasting and Energy Optimization

The outputs generated from machine learning models such as Decision Tree and Convolutional Neural Networks (CNN) are utilized to enable intelligent energy management. Based on predicted demand patterns and the current state of the grid, the system recommends optimized control actions to improve overall efficiency.

Key optimization outcomes include:

- Balanced distribution of electrical load across different regions
- Implementation of peak shaving techniques to reduce demand spikes
- Improved utilization of renewable energy sources
- Reduced reliance on high-cost backup power generation
- Minimization of energy losses and wastage
- Enhanced cost efficiency in grid operations

This stage plays a vital role in ensuring that available energy resources are allocated and utilized in the most efficient and sustainable manner.

#### G. Visualization and Monitoring

The processed results and analytical outputs are represented through interactive dashboards and reports, enabling operators and decision-makers to effectively monitor and understand system behavior.

Commonly used visualization tools include:

- PowerBI
- Tableau
- Plotly
- Matplotlib
- Typical dashboard features include:
  - Real-time visualization of energy demand
  - Forecasting graphs and predictive insights
  - Analysis of historical consumption trends
  - Alerts for faults and abnormal conditions
- –Monitoring of key performance indicators (KPIs) Effective visualization enhances system transparency and supports informed, data-driven decision-making.

#### H. Performance Evaluation

The effectiveness of the proposed system is assessed using standard evaluation metrics that measure both prediction accuracy and system performance.

Important evaluation metrics include:

- Accuracy: Measures the correctness of model predictions
- RMSE (Root Mean Square Error): Evaluates prediction error magnitude
- MAE (Mean Absolute Error): Measures average deviation between predicted and actual values
- Latency: Indicates the response time for processing real-time data
- Scalability: Assesses the system's ability to handle increasing data volume
- Reliability: Reflects consistent system performance under varying load conditions
- The proposed approach is expected to deliver improved performance compared to conventional systems, particularly in terms of forecasting precision and real-time data processing capabilities.

### I. Expected Impact

The integration of Data Lake storage, Lambda Architecture, and machine learning models such as Decision Tree and CNN provides a comprehensive framework for advanced smart grid energy management.

The anticipated benefits of this approach include:

- Enhanced accuracy in energy demand forecasting
  - Faster and more efficient decision-making processes
  - Better utilization of renewable energy resources
  - Reduction in operational and energy costs
  - Improved reliability and stability of the power grid
  - Decreased energy wastage
  - Scalable infrastructure capable of future expansion
  - Strong support for the development of smart cities
- Overall, the proposed methodology offers a robust, scalable, and intelligent solution for next-generation energy management systems, addressing both current challenges and future requirements of smart grid environments.

### V. CONCLUSION

The transition from traditional power systems to smart grids has introduced both opportunities and challenges, including rising energy demand, renewable integration, dynamic pricing, and massive data generation. Conventional energy management systems struggle with scalability, speed, and forecasting accuracy.

This study highlights the role of big data and AI in addressing these issues. Machine learning and deep learning models—such as Decision Trees, CNNs, LSTMs, and hybrid approaches—offer improved forecasting accuracy by capturing complex consumption patterns and trends. Scaled data architectures are equally important.

Data Lakes enable flexible storage of diverse data types, while Lambda Architecture combines batch and real-time processing to balance accuracy and low latency, ensuring timely and reliable insights.

The integration of Decision Trees and CNNs enhances decision-making and load forecasting, creating a more adaptive system. Overall, this approach improves efficiency, reliability, cost-effectiveness, and renewable energy utilization.

In conclusion, combining Data Lake, Lambda Architecture, and AI provides a scalable and future-ready solution for smart grid energy management, supporting sustainable energy systems and smart city development.

### REFERENCES

- [1] R. Hachache, M. Labrahmi, A. Grilo, et al., "Energy Load Forecasting Techniques in Smart Grids: A Cross-Country Comparative Analysis," *Energies*, vol. 17, no. 10, pp. 2251, 2024.
- [2] S. Lu and T. Bao, "Short-Term Electricity Load Forecasting Based on Neural Prophet and CNN-LSTM," *IEEE Access*, 2024.
- [3] C. T. Plavila, "Machine Learning Techniques for Load Forecasting in Smart Grids with Renewable Energy Integration," *MDPI Energies*, 2025.
- [4] K. Wehrmeister, et al., "Big Data Reference Architecture for the Energy Sector," *Springer Journal of Big Data*, 2025.
- [5] S. Azzabi, et al., "Data Lakes: A Survey of Concepts and Architectures," *Springer*, 2024.
- [6] P. Koukaras, "Data Integration and Storage Strategies in Heterogeneous Analytical Systems," *MDPI*, 2025.
- [7] J. He, K. Yuan, Z. Zhong, and Y. Sun, "Enhancing Short-Term Power Load Forecasting With a TimesNet-Crossformer-LSTM Approach," *IEEE Access*, vol. 12, 2024.
- [8] A. A. Munshi and Y. A. I. Mohamed, "Data Lake Lambda Architecture for Smart Grids Big Data Analytics," *IEEE Access*, vol. 6, pp. 40463–40474, 2018.



10.22214/IJRASET



45.98



IMPACT FACTOR:  
7.129



IMPACT FACTOR:  
7.429



# INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24\*7 Support on Whatsapp)