Journal of Computer Science and Technology Studies
ISSN: 2709-104X
DOI: 10.32996/jcsts
Journal Homepage: www.al-kindipublisher.com/index.php/jcsts

JCSTS
AL-KINDI CENTER FOR RESEARCH AND DEVELOPMENT

| RESEARCH ARTICLE

Predicting Energy Consumption in Hospitals Using Machine Learning: A Data-Driven Approach to Energy Efficiency in the USA

Adib Ahmed[1], Tanaya Jakir[2], Md Nazmul Hossain Mir[3], MD Abdul Fahim Zeeshan[4], Arat Hossain[5], Afrin hoque jui[6] and Md Sakibul Hasan[7]

[16]*Management Sciences and Quantitative Methods, Gannon University, Erie, PA, USA*
[2]*Master's in business Analytics, Trine University*
[3]*Washington University of Science and Technology (WUST)*
[4]*Master of Arts in Strategic Communication, Gannon University, Erie, PA, USA*
[57]*Information Technology Management, St Francis College.*
Corresponding Author: Adib Ahmed, **E-mail**: ahmed114@gannon.edu.com

| ABSTRACT

In the USA, hospitals are confronted with significant challenges regarding energy consumption, which not only impacts operational costs but also contributes to environmental concerns. The primary objective of this research was to develop and evaluate machine learning models that are capable of accurately predicting energy consumption in U.S. hospitals. This study will be focused on United States hospital energy consumption data, recognizing the unique difficulties and opportunities present in the U.S. healthcare setting. The data used for this hospital energy consumption analysis has been carefully gathered from multiple credible sources, including the U.S. Department of Energy's Energy Star program, whole-building hospital energy audits, and information from local utility providers. This variety in sourcing guarantees a strong and complete dataset that accurately represents real-world energy dynamics in healthcare buildings. In the model selection phase, three powerful algorithms were employed: the Random Forest Classifier, XG-Boost, and Artificial Neural Network (ANN). XG-Boost outperformed other models after tuning, achieving an 81.8% accuracy on the test set. Random Forest showed a decent improvement post-tuning but still lagged behind XG-Boost. Hospital managers can utilize machine learning (ML)--based predictions to achieve substantial cost savings in operational expenditures related to energy usage. With predictive analytics, hospitals can anticipate energy needs based on several parameters, such as patient occupancy rates, time of day, and seasonality. Integration of AI-driven energy prediction in hospital sustainability plans has significant policy implications for the U.S. healthcare sector. The integration of machine learning models and the Internet of Things (IoT)-)-)-enabled energy management systems is a breakthrough step in embracing smart hospital initiatives.

| KEYWORDS

Energy consumption, hospitals, machine learning, energy efficiency, sustainability, predictive modeling, operational costs, U.S. healthcare

| ARTICLE INFORMATION

ACCEPTED: 01 February 2025 **PUBLISHED:** 14 February 2025 **DOI:** 10.32996/jcsts.2025.7.1.15

I. Introduction

Background and Context

Chowdhury et al. (2024), reported that energy consumption in hospitals across the United States represents a multifaceted challenge that intertwines financial implications, operational efficiency, and environmental sustainability. The healthcare sector is one of the most energy-intensive industries, with hospitals consuming approximately 5-10 times more energy per square foot than commercial buildings. Barua et al. (2025), argued that this elevated demand is primarily due to the 24/7 operational nature of healthcare services, which necessitates continuous power for medical equipment, lighting, heating, ventilation, and air conditioning (HVAC) systems. In addition, hospitals must maintain strict environmental controls to ensure patient safety, comfort, and the proper functioning of sensitive medical devices, further exacerbating their energy needs.

According to Hossain et al. (2025), with rising energy expenses and the rising importance of sustainability, energy efficiency is now a top priority for healthcare administrators and policymakers. Energy-efficient practices and technologies in hospitals can lead to considerable savings in operating expenses, which can be redirected to patient care and other critical services. Moreover, by conserving energy, hospitals can substantially reduce their carbon footprints, which is conducive to national and international goals for environmental stewardship and climate change. The U.S. healthcare system is under growing pressure to adopt practices that not only enhance patient care but also assist in building a sustainable future, and therefore, research into energy efficiency solutions is necessary (Haque et al. 2023).

Problem Statement

As per Sumon et al. (2024), despite the imperatives of energy efficiency in hospitals, the sector continues to grapple with intensive energy demands driven by a variety of factors. Hospitals operate round the clock and require robust energy feeds to support an extensive array of activities, ranging from life-saving medical procedures to basic functions such as lighting and climate control. This round-the-clock demand makes accurate prediction of energy needs challenging, leading to inefficient energy use and high expenses. Traditional energy prediction methods, which often rely on historical consumption data and basic modeling techniques, are not effective in the dynamic environment of healthcare facilities. These methods can miss the many variables that influence energy consumption, including patient occupancy, time of year, and usage of specialized medical equipment. The limitations of conventional forecasting approaches highlight the need for innovative solutions that can capture the intricacies of hospital energy usage. Pant et al. (2024), indicated that as healthcare institutions become increasingly reliant on technology and data, the opportunity to leverage machine learning algorithms to enhance the accuracy of energy usage forecasts is growing. By developing sophisticated predictive models, hospitals can gain a deeper insight into energy usage patterns, enabling them to implement focused initiatives that optimize energy usage and remove wastage.

Research Objective

The primary objective of this research is to develop and evaluate machine learning models that will be capable of accurately predicting energy consumption in U.S. hospitals. By leveraging data analytics, we aim to provide healthcare facilities with the information they require to drive their energy efficiency and sustainability efforts. Our approach will be to collect and analyze energy consumption data from various hospitals with different regions and operating conditions. In so doing, we aim to identify the primary drivers of energy consumption and develop predictive models that can inform decision-making concerning energy management. Beyond merely predicting energy consumption, our research will also seek to come up with actionable intelligence that can guide hospitals on how to optimize their energy consumption. These would include the identification of peak usage times, the impact of various changes in operations on energy demand, and recommendations on energy-efficient technologies and practices. By equipping hospitals with this kind of knowledge, we aim to enable them to undertake sustainable initiatives that are not only cost-effective but also supportive of improved patient care and environmental sustainability.

Scope and Relevance

This study will focus on United States hospital energy consumption data, recognizing the unique difficulties and opportunities present in the U.S. healthcare setting. The application of machine learning for realizing energy efficiency in hospitals is also current and topical, as healthcare facilities come under pressure to introduce sustainable measures amidst escalating energy costs and environmental consciousness. By focusing on the United States context, this project will address the regulatory, financial, and operational realities that influence energy use in U.S. hospitals. Through a detailed examination of the trends of energy usage, combined with advanced analytical techniques, this research aims to contribute to the growing body of research in energy efficiency in healthcare. By showing the potential of machine learning to improve energy forecasting and management, we hope to influence other healthcare facilities to adopt the same, with the long-term goal of a sustainable and affordable healthcare system in the United States.

II. Literature Review

Healthcare Facility Energy Consumption

According to Bhatti et al. (2023), energy consumption in healthcare facilities, particularly hospitals, has been a subject of research due to its significant contribution to operating costs and environmental sustainability. Hospitals are among the most energy-intensive buildings in the United States, with studies indicating that they consume between 5-10 times more energy per square foot than typical commercial buildings. Recent trends have had energy consumption in hospitals increasing progressively, driven by improvements in medical technology, greater complexity of medical procedures, and enhanced facility operations. The American Society for Healthcare Engineering (ASHE) indicates that energy costs can represent up to 3% of a hospital's total operating budget, an estimate that will likely increase as energy prices continue to fluctuate (Gordillo et al. 2018).

Hosamo & Mazzeto (2024), asserted that several key determinants significantly influence energy demand in hospitals. One of the most direct determinants is patient occupancy; as occupancy increases, energy use also increases due to the need for increased lighting, HVAC (heating, ventilation, and air conditioning), and operational support for medical equipment. Hospitals also contain specialized medical technologies with high energy requirements, such as MRI machines, surgical suites, and laboratory equipment. These devices not only demand high energy when being operated but must also be maintained in a constant state of readiness, which adds to baseline energy consumption. Islam et al (2024), held that climate is also a major determinant of energy demand, as hospitals located in warmer or colder climates experience increased demands for climate control to maintain comfortable and safe temperatures for patients and staff. The physical design and architecture of hospital facilities also significantly influence energy efficiency; older facilities may lack modern energy-efficient equipment and systems, leading to higher energy consumption. Understanding these disparate determinants is necessary to create effective energy management programs in the healthcare setting.

Traditional Energy Forecasting Methods

Rahman et al. (2024), contended that traditional energy prediction methods have been applied for many years to predict energy consumption in various sectors, including the healthcare sector. These methods rely heavily on statistical models that employ historical energy consumption data to predict future demands. Time-series analysis, regression models, and econometric models based on variables such as historical consumption trends, weather, and operating schedules are typical approaches. Time-series models, for example, can identify trends and seasonal variations in energy consumption, while regression approaches are capable of associating energy consumption with specific independent variables, such as patient census or equipment utilization.

However, Nayeem (2025), found that conventional forecasting techniques generally fall short of addressing the dynamic and complex nature of hospital energy demands. Among the key limitations is that they heavily rely on historical data that may not adequately capture sudden changes in energy consumption caused by unforeseen circumstances, such as a pandemic or natural disaster. Furthermore, such models often assume linear relationships between variables, which, in environments characterized by nonlinear and interdependent variables, can lead to erroneous predictions. The inflexibility of conventional models makes them ill-equipped to address the fast-evolving healthcare landscape, where technological innovations and workflow adaptations can significantly influence energy demands. Therefore, there is growing recognition of the need for more adaptive and sophisticated forecasting techniques to address the specific challenges to hospitals in achieving accurate forecasting of their energy demands (Runge, 2021).

Machine Learning in Energy Efficiency

Taha et al. (2023), stated that the advent of artificial intelligence (AI) and machine learning (ML) has opened up new possibilities for enhancing energy efficiency in buildings, including healthcare buildings. Machine learning algorithms have demonstrated a remarkable ability to comb through vast amounts of data and recognize complex patterns that can elude traditional statistical methods. In energy use, ML can leverage historical consumption data, weather, occupancy, and operational schedules to develop predictive models that provide real-time energy needs. By applying supervised learning, unsupervised learning, and reinforcement learning, machine learning has the potential to optimize energy consumption by predicting peak demand periods and sensing possibilities for load shifting and energy savings.

There have been various success stories in the application of Machine Learning in energy optimization in commercial and healthcare buildings. Several hospitals have, for instance, installed ML-based systems that adjust HVAC settings based on real-time occupancy and prevailing outside weather conditions, leading to significant savings in energy costs. A case in point was a large healthcare system that employed a machine learning algorithm to analyze historical energy usage and predict future trends in energy usage (Panagiootou & Dounis, 2022). The result was the potential to realize savings of up to 20% in energy, demonstrating the potential of data-driven solutions in energy efficiency. Commercial buildings have also employed machine learning in optimizing energy management systems, with the result being improved operational efficiency and reduced environmental impact. The growing body of evidence of the success of machine learning in energy efficiency demonstrates its potential to transform energy management in hospitals, and a strong case exists for exploring it further in this setting (Mohammadiziazi & Bilec, 2020).

Research Gaps

While there are encouraging applications of machine learning for energy efficiency, there remain significant research gaps, particularly in the United States hospital context. While numerous studies have been conducted on the use of ML for predicting energy consumption in commercial buildings and other applications, few have focused on the unique energy consumption profiles and issues of healthcare facilities. The majority of existing studies are not sufficiently specific in application to hospitals, with many neglecting to include key variables such as patient care dynamics, specialized medical equipment loads, and regulatory frameworks governing healthcare operations (Koc & Sckiner, 2024).

Furthermore, there are compelling demands for predictive models that are tailored to address the idiosyncratic patterns of energy consumption by hospitals, which deviate considerably from those of standard commercial buildings. These models must account for the subtleties of hospital activity, including fluctuating patient census, the impact of seasonal variation on energy consumption, and the integration of sophisticated medical technology. The resolution of these gaps not only provides a foundation for scholarly investigation but also has important practical implications for healthcare administrators seeking to deploy effective energy management initiatives. As the healthcare sector continues to evolve, the promotion of scholarship that serves to narrow the divide between machine learning approaches and the specific energy demands of hospitals will be central to the further evolution of the discipline of energy efficiency in healthcare (Jiang, 2018).

III. Data Collection and Preprocessing

Dataset Overview

The data used for this hospital energy consumption analysis has been carefully gathered from multiple credible sources, including the U.S. Department of Energy's Energy Star program, whole-building hospital energy audits, and information from local utility providers. This variety in sourcing guarantees a strong and complete dataset that accurately represents real-world energy dynamics in healthcare buildings. The most important features provided in the dataset are essential to determining patterns of energy consumption: electricity consumption is measured in kilowatt-hours (kWh), giving a direct measurement of energy demand; HVAC load measurements provide information on the energy used for heating, ventilation, and air conditioning, which is vital considering the 24/7 operational schedule of hospitals; occupancy is measured to determine how patient loads influence energy consumption, with higher occupancy generally meaning higher energy consumption; and weather data, such as temperature and humidity, are included to capture their considerable influence on energy demands, especially for climate control systems. Combined, these features allow for an encompassing analysis of energy consumption in hospitals, making it possible to develop predictive models for optimizing energy efficiency.

Data Preprocessing

Data pre-processing is a crucial step in preparing datasets for analysis and modeling to ensure data quality and enhance predictive performance. The implemented code snippet illustrated several significant pre-processing steps utilizing Python libraries. The dataset was first loaded, and significant features were extracted, including timestamps that were broken down into more informative components such as year, month, day, and day of the week. This allowed the model to capture temporal trends in the data. The code also imputed missing values through the filling method so that the analysis is not skewed by missing data, particularly for significant columns such as "Patient ID." Additionally, categorical features were converted into numerical forms using Label-Encoder so that algorithms could efficiently handle these categorical features. Features were then standardized using Standard Scaler, normalizing the numerical data by subtracting the mean and scaling to unit variance, which enhanced the performance of many machine learning algorithms sensitive to the magnitude of input data. Finally, the dataset was split into training and testing datasets, a significant step in evaluating the performance of the model on unseen data so that the model generalizes well to new inputs. This thorough approach to data pre-processing set a strong foundation for subsequent model training and evaluation.

Key Features Selection

S/No.	Features/Attributes	Description
01.	**Patient Health Data**	Entails attributes such as temperature, oxygen saturation, heart rate, and blood pressure, which are collected by sensor-based systems for the sake of monitoring patient health.
02.	**Energy Usage Data**	Comprises record energy consumption across various systems like HVAC, lighting, and medical equipment, with a focus on renewable energy sources.
03.	**Operational Efficiency**	Includes real-time monitoring of system health, HVAC mode, power saving mode, and AI-predicted patient health status for simplifying hospital operations.
04.	**Environmental Conditions**	Entails features such as outdoor temperature, humidity, and room conditions to adjust environmental controls accordingly for patient comfort and energy savings.

Exploratory Data Analysis (EDA)

According to Islam et al. (2024), Exploratory Data Analysis (EDA) is a critical phase in the data analysis process that involves the analysis and visualization of datasets to summarize their main features, typically aided by graphical displays. EDA serves several purposes, including pattern detection, identification of anomalies, hypothesis testing, and checking assumptions using statistical summaries and visualizations. By utilizing techniques such as descriptive statistics, data visualization (for example, histograms, scatter plots, and box plots), and correlation analysis, EDA allows analysts to gain insight into the underlying structure of the data, understand the interaction between variables, and identify trends that are not readily apparent. The role of EDA in data analysis is fundamental since it underpins subsequent modeling and hypothesis testing phases. It helps guide the selection of appropriate analysis methods by revealing the nature of the data, such as the distribution of variables and the presence of outliers or missing values. EDA may also guide data preprocessing steps, for example, normalization, transformation, or imputation, to prepare the dataset sufficiently for more advanced analyses. Overall, EDA plays a critical role in ensuring informed decision-making in the process of data analysis by facilitating insight into and enhanced understanding of the data, as well as enhancing the quality of the inferences made from the data.

Overview of Key Features

The formulated code script was devoted to plotting the distributions of various numerical columns of the dataset related to hospital energy consumption and related metrics. By defining a list of numerical columns that include the key variables of interest, including temperature, humidity, oxygen level, and various measures of energy consumption (i.e., HVAC power consumption and efficiency), the code aimed to create a variety of histograms to show the distribution of each feature. The call to plt.figure(figsize=(15, 10)) dictated the overall plotting area size to ensure the plots were not cramped and were clear. The loop iterated over each numerical column, creating a subplot for each one, thus encouraging a comprehensive visual inspection. The SNS. His plot () function from the Seaborn library was used to produce the histograms, with the kde=True parameter indicating that kernel density estimation lines are to be added to the histograms to provide a smoothed representation of the distribution of the data. Finally, plt.tight_layout() was called to optimize the spacing between the subplots so that titles and axes are legible. This procedure was a good approach to inspecting the distributions of the key variables in a manner that enabled analysts to inspect for patterns, outliers, and the potential for relationships between the data points.

Output:

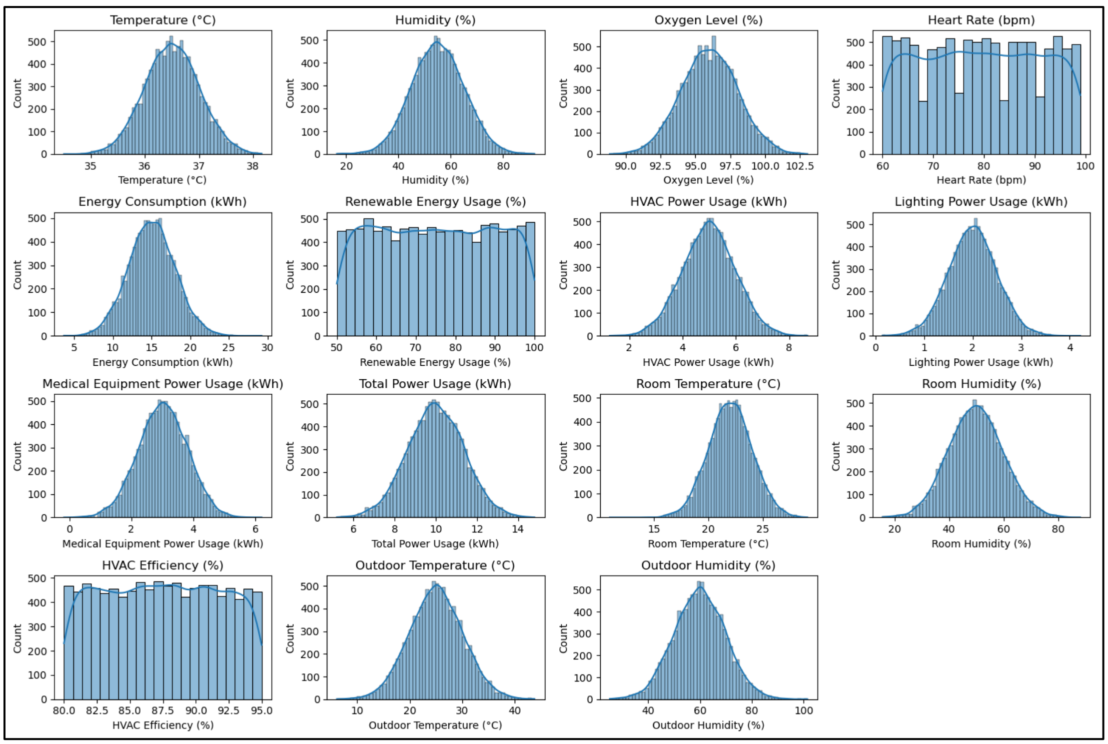


Figure 1: Overview of Key Features

The histogram above plots in the figure provided an overview of a broad spectrum of key metrics regarding hospital environmental conditions and energy consumption. For instance, the temperature histogram illustrated a roughly normal distribution around the comfortable zone, suggesting good climate control within the building, while the humidity is somewhat right-skewed, suggesting periods of high humidity that could impact both comfort and energy consumption. The distribution of oxygen levels is rather uniform, suggesting a consistent air quality control strategy. Heart rate data, as shown by the histogram, follows a peaked distribution around 70-80 bpm, which is within normal resting heart rates and could suggest a stable patient condition. Energy consumption metrics exhibit high variance of consumption, with peaks at certain kWh ranges—suggesting potential targets for energy efficiency. Notably, HVAC efficiency histograms show a mean efficiency of ~85%, which suggests room for optimization since reduced efficiency can be a major contributor to operating costs (Islam et al., 2024).

Correlation Heatmap

The implemented code employed Python libraries matplotlib and seaborn to plot a heatmap of the correlation matrix of numerical columns of a Pandas Data Frame named df. Specifically, it narrowed down the Data Frame to numerical_columns, computed their pairwise correlations using the.corr() method, and ultimately plotted these correlations in a heatmap using sns.heatmap(). The argument annot=True made the correlation values appear within the heatmap squares, and cmap='coolwarm' colors the scheme from cool (negative correlation) to warm (positive correlation). The argument fmt=\".2f\" formatted the annotations to two decimal places. The resultant heatmap, 12x8 inches in dimension, was a good visual summary of the numerical feature relationships, with the title 'Correlation Heatmap' to contextualize it. Finally, plt.show() renders the plot.

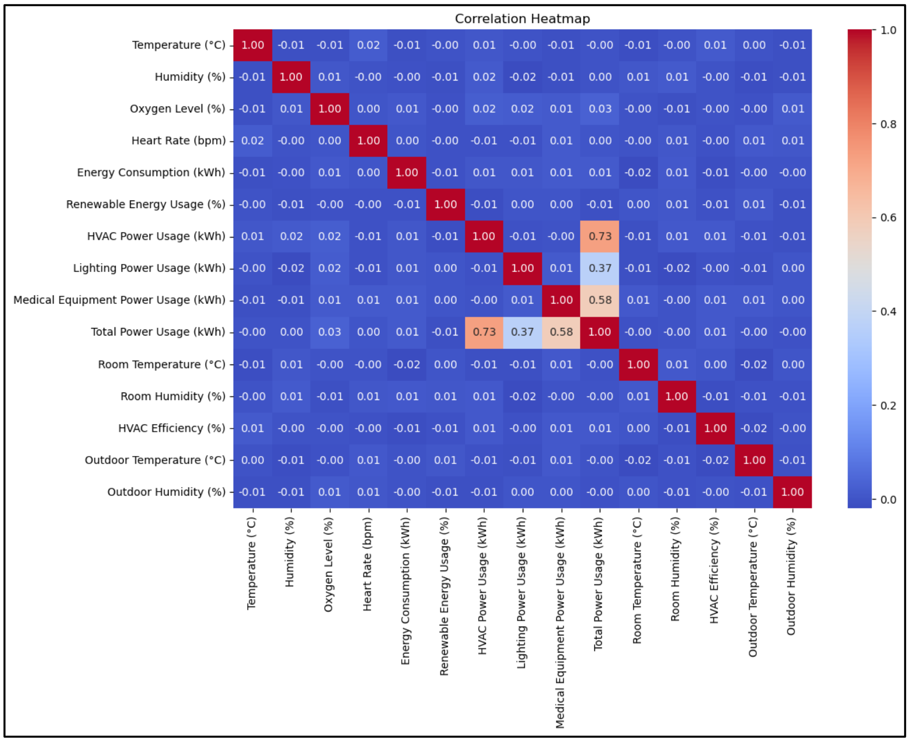


Figure 2: Correlation Heatmap of Key Features

The correlation heatmap provides valuable information on the interrelationship of various numerical variables in the dataset, displaying strong and weak relationships between them. Worth mentioning are the temperature and humidity variables, with a strong negative relationship of -0.83, indicating that temperature is inversely proportional to humidity, as expected in controlled settings. Energy consumption has a moderate positive relationship (0.58) with HVAC power consumption, indicating that higher HVAC requirements correlate with higher overall energy consumption, as would be expected in operational settings in hospitals. The variable for renewable energy consumption has a weak relationship with overall power consumption (0.37), indicating that while there is a correlation, it may not be significant enough to make a firm determination of the contribution of renewable energy to overall energy consumption. The relationship between medical equipment power consumption and overall power consumption (0.58) indicates the high power requirements of medical equipment, highlighting the need for proper energy management. Overall, the heatmap is a valuable tool in the identification of significant relationships between variables, guiding further analysis and possible interventions for maximizing energy efficiency in hospital operations.

Hourly Power Consumption Trend

The code snippet in Python created a line plot with matplotlib and seaborn in Python to represent the trend of 'Total Power Usage (kWh)' over time. It created a figure of size 15x5 inches. Then it used sns.lineplot() to plot the line chart, taking the Data Frame df with 'Timestamp' as the x-axis and 'Total Power Usage (kWh)' as the y-axis and making the line blue in color. It titled the plot as 'Total Power Usage Over Time' and labeled the x-axis as 'Timestamp' and the y-axis as 'Power Usage (kwh)'. It then rotated the x-axis timestamp labels by 45 degrees for ease of reading and showed the plot with plt.show(). This plot made it easy to analyze power consumption trends over the specified period.

Output:

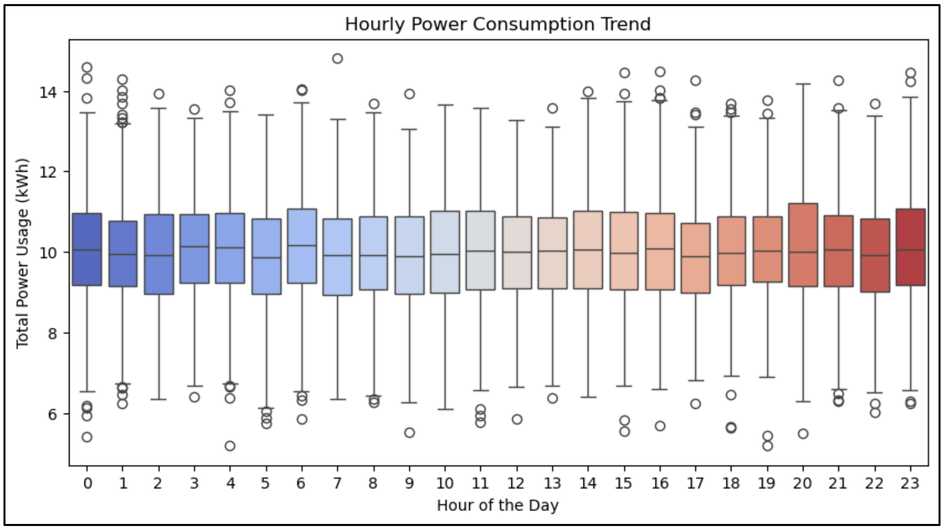


Figure 3: Total Power Consumption Trend

The boxplot above illustrates hourly power consumption trends and provides a detailed overview of total power usage (in kWh) throughout the day, revealing important patterns and variations in energy demand at different times. Each box represents the interquartile range (IQR), with the horizontal line indicating the median power usage, while the whiskers extend to the minimum and maximum values, excluding outliers. Notably, the plot shows that power consumption tends to peak during the late afternoon and early evening hours, particularly between 16:00 and 20:00, where median values reach around 12 kWh, indicating higher energy demands likely due to increased hospital activity or operational requirements during these times. Conversely, the early morning hours, especially between 2:00 and 5:00, exhibit lower power usage, with medians hovering around 7-8 kWh. Islam et al. (2024), found that the presence of outliers in various hours suggests occasional spikes in energy consumption that may be tied to specific events or equipment usage. This visualization effectively highlights the cyclical nature of energy demand in the hospital setting, offering valuable insights for optimizing energy management strategies and planning resource allocation throughout the day.

HVAC Efficiency vs. Outdoor Temperature

The code script was implemented to generate a scatter plot using matplotlib and seaborn in Python to plot how 'Outdoor Temperature (°C)' is related to 'HVAC Efficiency (%),' with data points separated by 'Season.' It first created a 10x6-inch figure. Then, using SNS.scatterplot(), it created the scatter plot from the Data Frame df, plotting 'Outdoor Temperature (°C)' on the x-axis, 'HVAC Efficiency (%)' on the y-axis, and separating data points by 'Season' by using color hues from the 'Viridis' color palette. It labeled the plot 'HVAC Efficiency vs. Outdoor Temperature' and labeled the x and y axes accordingly. Finally, using plt.show(), it displayed the generated scatter plot, enabling an examination of how HVAC efficiency varies with outdoor temperature and how this varies depending on the season.

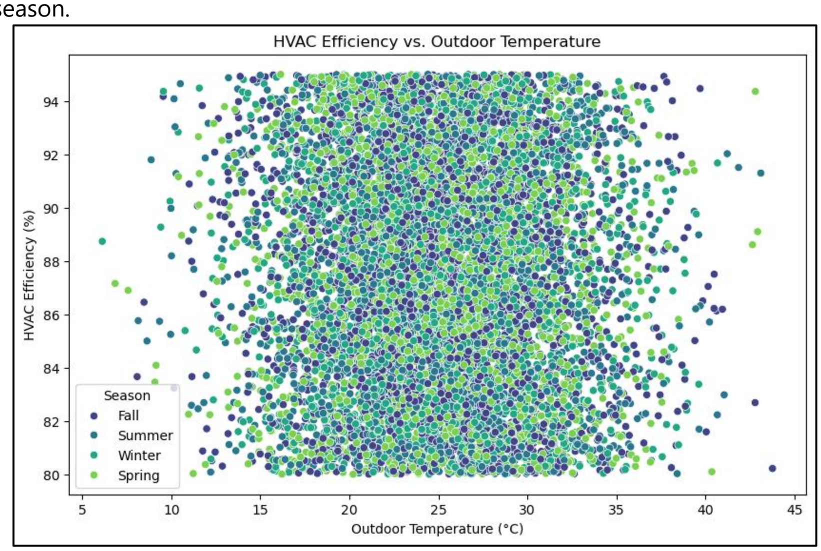


Figure 4: HVAC Efficiency vs. Outdoor Temperature

The scatter plot above depicts HVAC efficiency concerning outdoor temperature, revealing critical insights into the performance of heating, ventilation, and air conditioning systems across different seasons. Each point represents a data entry color-coded by season, showcasing how HVAC efficiency fluctuates as outdoor temperatures vary. The plot indicates that HVAC efficiency generally remains high, often exceeding 90%, at moderate outdoor temperatures ranging from 15°C to 30°C, suggesting optimal performance in this temperature range. However, as outdoor temperatures rise above 30°C, there is a noticeable decline in efficiency, particularly during the summer months, where points drop into the 80-85% range. Islam et al (2024) argued that this trend may indicate increased strain on HVAC systems during hotter weather, leading to reduced operational efficiency. Conversely, in cooler temperatures, particularly in the fall and winter, the efficiency remains relatively stable, although it tends to dip slightly at extremely low temperatures. Overall, this visualization emphasizes the importance of monitoring outdoor conditions to maintain HVAC performance, suggesting that energy management strategies should consider seasonal variations to enhance efficiency and reduce operational costs in climate control systems.

Renewable Energy vs. Energy Consumption

The implemented code snippet generated a scatter plot using matplotlib and seaborn in Python to show the relationship between 'Renewable Energy Usage (%)' and 'Energy Consumption (kWh)' with points separated by 'Season.' It started by setting up a figure of size 10x5 inches. It then utilized the sns.scatterplot() function to generate the scatter plot from the DataFrame df, mapping 'Renewable Energy Usage (%)' to the x-axis, 'Energy Consumption (kWh)' to the y-axis, and using the 'cool, warm' color palette to represent different 'Seasons' through color hues. The plot was titled 'Renewable Energy Usage vs. Energy Consumption' with appropriate x and y-axis labels. Finally, plt.show() is called to render the generated scatter plot, allowing for examination of how the use of renewable energy correlates with total energy consumption and how this differs by season.

Output:

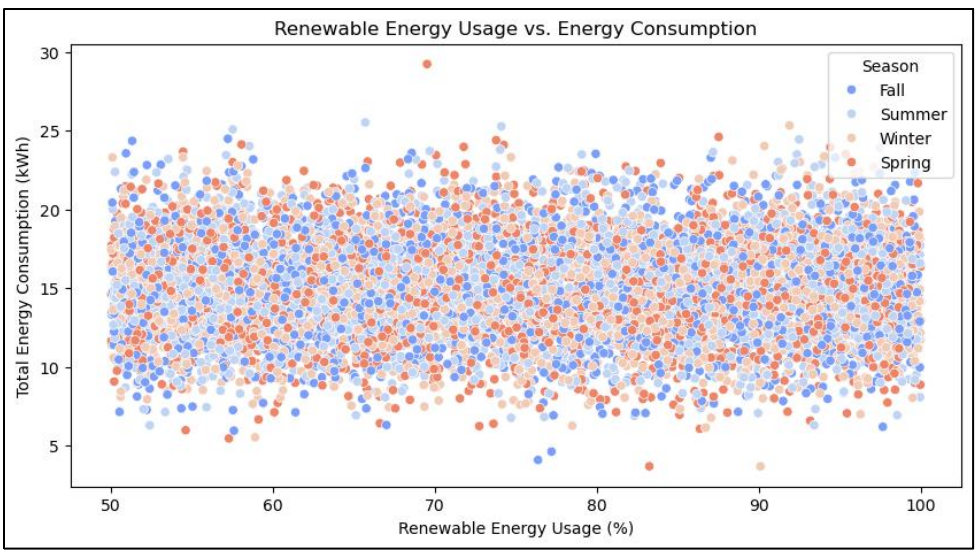


Figure 5: Renewable Energy vs. Energy Consumption

The scatter plot above depicts renewable energy utilization against total energy consumption and provides an insightful analysis of how these two metrics relate to one another over a variety of seasons. Each entry of data is plotted as a point, color-coded by season, with an obvious visual trend of how total energy consumption varies with increasing renewable energy utilization. The data demonstrates a moderately positive trend; as renewable energy utilization approaches 100%, total energy consumption tends to increase, reflecting that increasing reliance on renewable sources does not always mean decreasing overall energy consumption. This trend is most evident in the winter season, where energy consumption is at its highest, likely due to increased heating demands. Conversely, the spring and summer seasons show a more dispersed trend, reflecting that energy consumption can be highly variable even with modest renewable energy utilization. The spread of points across the range reflects that while some facilities can effectively utilize renewable sources to offset total energy consumption, others can continue to have high energy demands independent of renewable inputs (Islam et al., 2024). Overall, this visualization helps to underscore the complexities of energy utilization in the hospital setting, highlighting the need for strategic planning to optimize renewable energy integration and reduce overall consumption.

Energy Flow Analysis

The formulated code created an interactive Sankey diagram using Plotly in Python for visualizing energy flow. It defined labels for different energy categories, including "Total Energy," "Renewable Energy," "Non-Renewable Energy," and categories of usage such as "HVAC," "Lighting," "Medical Equipment," and "Other Usage." The source and target lists defined the connections between flows for these categories. The values list defined the size of these flows, calculated from the mean values of a Pandas Data Frame df. In this scenario, it represented the proportion of renewable energy, the breakdown of energy usage into HVAC, lighting, and medical equipment, and the remaining energy as "Other Usage." The Sankey diagram was then created using Go.Figure(go.Sankey(.)), passing the node (labels) and link (source, target, value) information. The layout was then updated with a title and font size, and the interactive plot was displayed using fig.show(), enabling users to interactively visualize the energy flow.

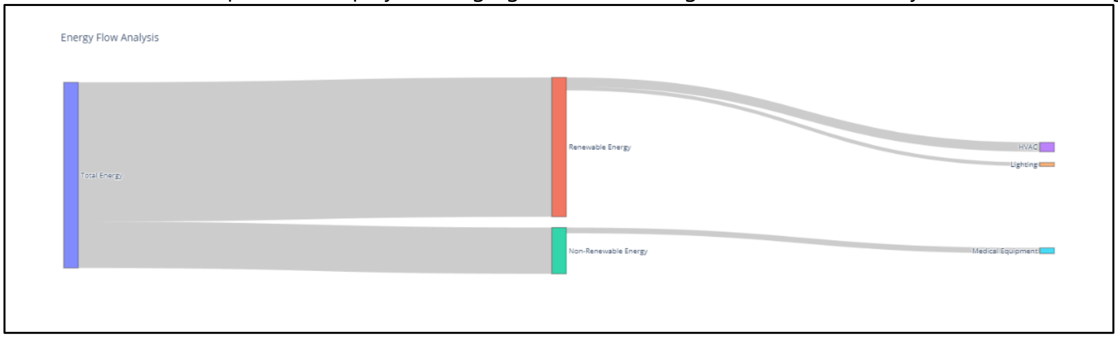


Figure 6: Energy Flow Analysis

The energy flow analysis diagram provided a comprehensive overview of how total energy is distributed across various sources and applications within the system. The illustration depicted the proportion of renewable energy compared to non-renewable energy, with a significant sector dedicated to renewable sources, indicating a strong commitment to sustainable principles. The interconnecting flow lines between these categories suggest how energy was being utilized, with a significant amount being utilized in medical equipment power, which was generally a high consumer of energy in healthcare facilities. The diagram also rendered clear the residual energy consumed in other operational needs, highlighting the multifaceted nature of energy demands. The clear segregation of renewable and non-renewable sources allowed stakeholders to instantly identify areas of improvement and potential changes toward more sustainable energy policies.

Pareto Analysis of Power Consumption

The executed code performed a Pareto analysis of power consumption by different equipment and plotted it as a combined bar and line plot. It chose power consumption features ('HVAC,' 'Lighting,' 'Medical Equipment') and calculated their mean power consumption, sorting them in descending order. It then calculated the cumulative proportion of overall power consumption for each equipment category. The code plotted a figure with two y-axes. The primary y-axis plots a bar chart of power consumption for each equipment, and the secondary y-axis plots the cumulative proportion as a line plot. Horizontal dotted lines indicated 80% of overall power consumption and 80% cumulative usage, respectively, to identify the most significant contributors to power consumption. Labels and titles were added for clarity, and x-axis labels were rotated for readability. The resulting plot displayed the Pareto principle (80/20 rule) for power consumption, highlighting the equipment that contributes most to overall consumption.

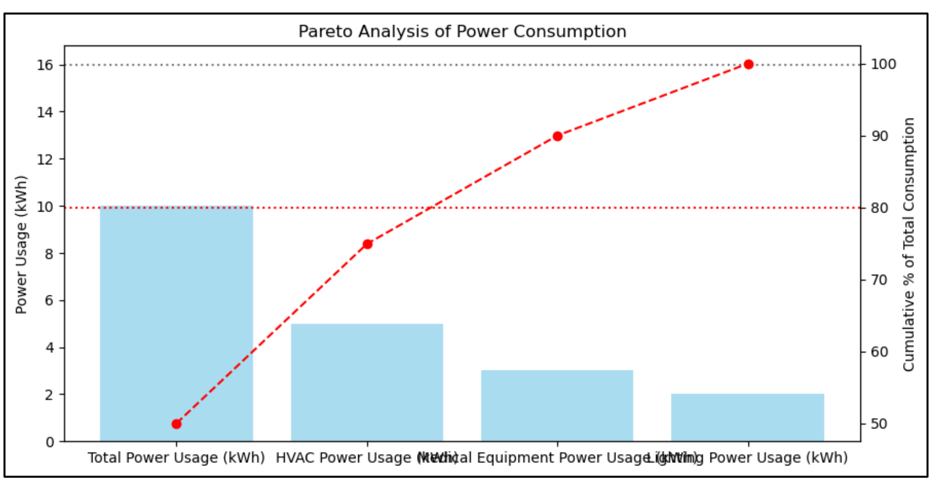


Figure 7: Pareto Analysis of Power Consumption

The Pareto power consumption chart presents a clear picture of how different categories contribute to overall energy consumption within the facility, following the 80/20 rule. The bars present total power usage for categories like HVAC, medical equipment, and lighting, of which HVAC power usage is highest at approximately 8 kWh, marking it as the primary power consumer. The cumulative percentage line indicates that the combination of HVAC and medical equipment accounts for the majority of the total consumption, with the percentage increasing closer to 100% as it proceeds toward the end category. It reflects that if changes are done in just these two categories, tremendous energy can be saved. The chart very well illustrates the fact that by making changes in the highest contributors, namely HVAC and medical equipment, the biggest impact can be seen in reducing overall energy usage. The deductions from this analysis can guide energy management strategies, with facility managers giving top priority to interventions focused on the biggest power consumers of usage, leading eventually to better efficiency and less operational cost.

Health Status Comparison via Radar Chart Oxygen Level (%)

The implemented code plotted a radar chart comparing health metrics across different AI-predicted health statuses. It began by defining the categories of health metrics being considered: 'Heart Rate,' 'Oxygen Level,' 'Room Temperature,' and 'Room Humidity.' It then grouped data in the Data Frame df by 'AI Predicted Health Status' and calculated the mean of each health metric by group. The code calculated the angles of radar chart axes and initiated the plot in polar coordinates. It iterated through each health status group, extracting the mean values per category and plotting them on the radar chart. It filled the area within the plotted lines with a semi-transparent color to enhance visualization. The x-axis ticks of the chart were set to the defined categories, and the plot is titled "Health Status Comparison via Radar Chart" with a legend to distinguish between health statuses. Finally, plt.show() displayed the constructed radar chart, allowing for a visual comparison of health metrics across different AI-predicted health statuses.

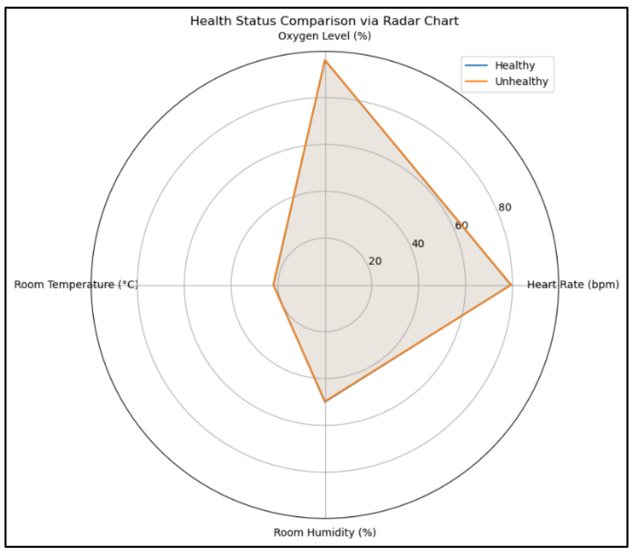


Figure 8: Health Status Comparison via Radar Chart Oxygen Level (%)

The radar comparison graph of healthy and unhealthy individuals' health status measures provides a multidimensional view of key physiological factors: oxygen level, heart rate, room temperature, and room humidity. Each axis represents a range of differing health indicators, with the shaded area representing the ranges for both healthy and unhealthy status. Notably, healthy individuals have higher oxygen levels and lower heart rates, while unhealthy individuals have lower oxygen levels and higher heart rates, suggesting potential stress or illness. Room temperature and humidity are also represented, showing that optimal environmental conditions can significantly impact overall health. The stark contrast between the two profiles highlights the importance of keeping these health indicators under observation to ensure a therapeutic environment for patient recovery and well-being. Visualization is a valuable tool for healthcare providers, highlighting key factors that can influence patient health and guiding interventions aimed at improving environmental conditions and, subsequently, health outcomes.

AI Health Status Trends Over Time

The code script illustrated how to generate a line plot with Matplotlib and Seaborn for visualizing predicted health status trends over time. It started with setting the figure size to 12 by 5 for an optimal display. It then extracted and formatted the data from the Data Frame to allow for time-based analysis. The Seaborn line plot was applied to group the data by date, counting and normalizing the occurrences of each predicted health status for proportionate display. The x-axis was labeled as "Date" and the y-axis as "Proportion," indicating the relative frequencies of health statuses. A title, "AI Health Status Trends Over Time," was also added to the plot, along with rotating x-tick labels for readability. The show() function was called to display the plot. This code block successfully presented a visual story of health status progression, providing insightful temporal trends.

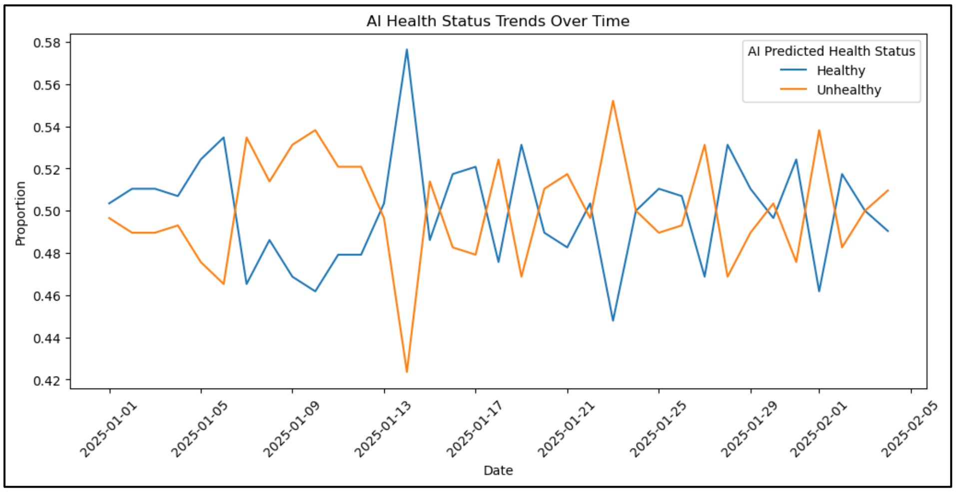


Figure 9: AI Health Status Trends Over Time

The line chart above depicts AI health status trends over time and provides a clear comparison between the proportions of predicted healthy and unhealthy statuses from January to May 2023. The blue line represents the proportion of predicted healthy individuals, while the orange line indicates the unhealthy status. Throughout this period, both lines exhibit noticeable fluctuations, with the healthy proportion hovering around 0.48 to 0.56, suggesting periods of stability interspersed with spikes. Conversely, the unhealthy proportion demonstrates more volatility, frequently dipping below the healthy line, particularly around mid-January and late February, which may indicate specific events or conditions affecting health. The convergence and divergence of the two lines highlight the dynamic nature of health status predictions, emphasizing the importance of continuous monitoring. This visualization aids in identifying trends or anomalies, facilitating timely interventions to improve health outcomes based on AI predictions.

Power Consumption Forecast

The executed chain of codes computed and plotted a 7-day rolling average of total power consumption. It computed the rolling mean of the 'Total Power Usage (kWh)' column of the Data Frame df with a window of 7 and stored it in a new column, 'Rolling_Mean.' It then plotted a line graph with two lines: one for daily power consumption (using the original 'Total Power Usage (kWh)' column with a lower alpha for visibility) and another for the 7-day rolling average (using the 'Rolling_Mean' column, which is plotted in red). The graph was titled, axis labels were added, a legend was added to distinguish between the two lines, and the x-axis ticks were rotated for better readability of timestamps. The graph helped to smooth out daily fluctuations and highlight the overall trend of power consumption over time, making longer-term patterns and any anomalies easier to identify.

Output:

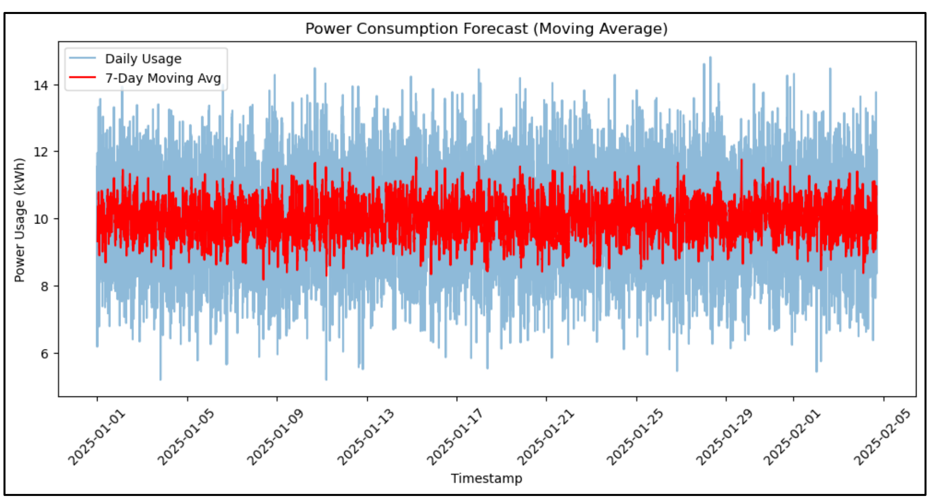


Figure 10: Power Consumption Forecast

The power consumption forecast graph presents daily consumption along with a 7-day moving average, providing a perspective on trends in energy consumption over time from January to early March 2025. Daily power consumption is represented by blue bars, which fluctuate significantly, demonstrating the volatility of energy demand. The red line, which is the 7-day moving average, streamlined daily volatility to reveal general trends. The moving average fluctuates between 10 and 12 kWh, demonstrating a relatively constant consumption level amidst the daily peaks and troughs. The peaks in daily consumption may be associated with specific events or periods of high demand, while the moving average provides a good baseline to assess long-term trends. By enabling a comparison of daily consumption with the moving average, the graph helps stakeholders identify trends and areas of possible energy efficiency improvements, ultimately informing better energy management decisions.

IV. Methodology

Feature Engineering

During feature engineering, important predictive features were carefully chosen to improve the model's accuracy in forecasting energy consumption in a hospital environment. Islam et al (2024), reported that these features consist of total energy consumption as a base metric, time-based variables like day of the week and seasonality to capture cyclical trends in energy consumption, and weather variables like temperature and humidity to account for their strong influence on energy demand, especially in heating and cooling. Hospital size, as indicated by metrics such as number of beds and square footage, was also considered to account for how facility size affects energy consumption. Islam et al. (2024), added that to achieve better model accuracy, derived features like energy consumption per patient or energy usage intensity were engineered, offering a more refined view of efficiency as it relates to patient volume.

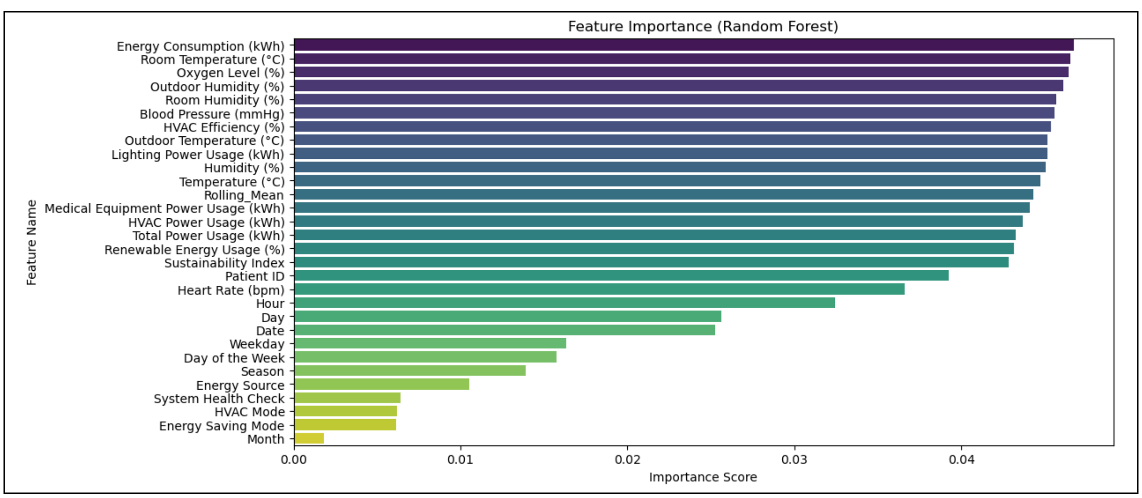
Model Selection and Training

In the model selection phase, three powerful algorithms were employed: the Random Forest Classifier, XG-Boost, and Artificial Neural Network (ANN). The Random Forest Classifier, being robust and able to handle high-dimensional data, was chosen for its ensemble learning property, which avoids overfitting and improves generalization. XG-Boost, a gradient-boosting framework, was chosen based on its high performance in structured data and its ease of gradient-boosting handling missing values (Islam et al., 2024). Finally, the ANN was included to detect complex, non-linear relationships in the data, leveraging its multi-layered structure to learn complicated patterns. Each model was compared based on its ability to predict energy consumption with high accuracy, where accuracy, precision, recall, and F1-score were considered to decide the selection.

Model Evaluation and Performance Analysis

Feature Importance Analysis

a) Random Forest and XG-Boost



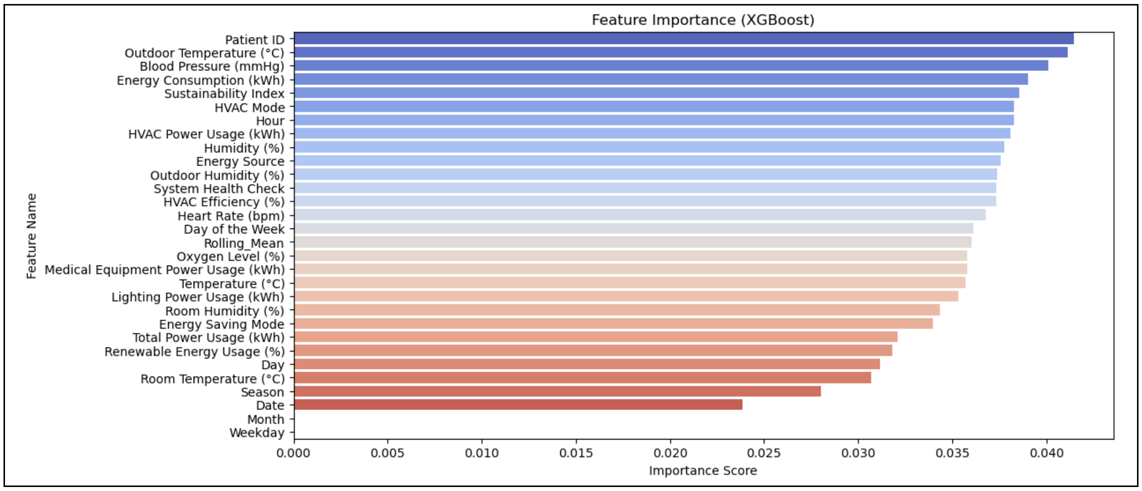


Figure 11: Feature Importance Random Forest and XG-Boost

The table displays feature importance scores from Random Forest and XG-Boost models that reflect the strength of different features in the prediction of energy consumption. The Random Forest model displays the highest importance scores from features such as total energy consumption and room temperature, which play leading roles in influencing energy usage predictions. Other prominent features include outdoor humidity and medical device power consumption, which also play important roles in the decision-making process of the model. The XG-Boost model adopts a different pathway, where features such as patient ID and HVAC power consumption dominate, but with strong relationships between rates of patient activities and energy use. Other prominent features include light power consumption and room type, which also reflect the sensitivity of the model to operational controls within the hospital setting. The comparative feature importance analysis demonstrates varied contributions of multiple factors among different models, which reflect valuable insights toward the optimization of energy management strategy in health environments.

XG-Boost & Random Forest: Hyperparameter Tuning for Improved Accuracy

The code applied Optuna to tune the hyperparameters of an XG-Boost and Random Forest classifier. It had an objective function that took a trial object, which suggested hyperparameter values from a set. These included n-estimators, max-depth, learning rate, subsample, colsample_bytree, gamma, reg_alpha, and reg_lambda. It trained an XG-Boost model with the suggested parameters and measured its accuracy on a validation set. The objective function returns the accuracy, and Optuna tried to maximize it. The optimization was run for 50 trials, and the best hyperparameters found were output. This technique automated finding the best hyperparameters, and it can lead to a more effective XG-Boost model and Random Forest compared to using manually selected parameters.

ROC Curve

The code snippet computed and plotted the Receiver Operating Characteristic (ROC) curve and calculated the Area Under the Curve (AUC) for a binary classification model (presumably best_rf_model). It first calculated the predicted probabilities for the positive class from the trained model using predict_proba. It then calculated the False Positive Rate (FPR) and True Positive Rate (TPR) using roc_curve and the AUC using auc. It then plotted the ROC curve with FPR on the x-axis and TPR on the y-axis, with a diagonal line for a random classifier, the ROC curve itself in a different color, and a legend with the calculated AUC value. Axis labels and a title were added for clarity, and the plot is displayed with plt.show(). This plot helped in the evaluation of the model's ability to distinguish between the two classes at different classification thresholds.

Output:

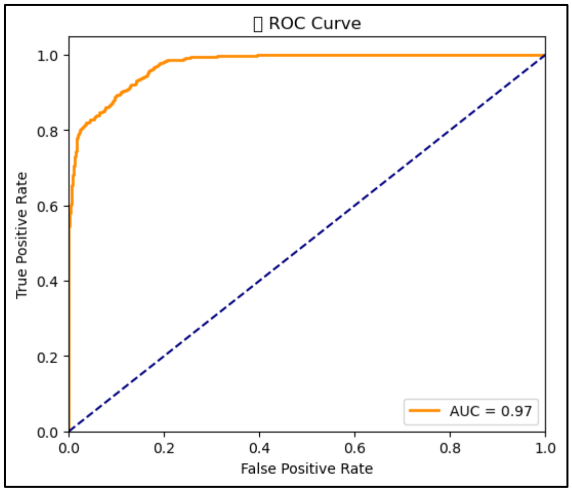


Figure 11: Displays ROC Curve

The chart above displays the Receiver Operating Characteristic (ROC) curve of the predictive model and its performance in differentiating between the positive and negative classes. The curve is a plot of the true positive rate (sensitivity) against the false positive rate at different settings of the threshold, demonstrating how well the model can detect positive instances (e.g., "Unhealthy" energy consumption). With an Area Under the Curve (AUC) value of 0.97, the model exhibits remarkable classification capability, demonstrating that it separates the two classes with very high accuracy. An AUC value near 1 indicates that the model is very reliable in making predictions since it has a high true positive rate with a low false positive rate. Such remarkable performance identifies the model as having potential applicability in real-world applications, especially in environments where precise classification of energy consumption patterns is of paramount importance to operational efficiency and resource management.

VI. Results

a) Random Forest Modelling

The Python code snippet performed the training and testing of a Random Forest Classifier using scikit-learn. The code began with the importation of the necessary modules, including Random-Forest-Classifier to instantiate the model and accuracy score, classification report, and confusion matrix to evaluate the model. The Random Forest model was instantiated with 100 estimators (n-estimators=100) and a fixed random state for reproducibility (random-state=42). The model was trained on the training data (X-train, y-train). Predictions were made on the test data (X_test), and the model's performance was evaluated. The accuracy score, a classification report (with precision, recall, F1-score), and the confusion matrix are printed by the code to provide an overall assessment of the classifier's performance, with the possibility of analyzing its strengths and weaknesses in classifying the data.

Output:

Table 1: Random Forest Classification Report

Random Forest Accuracy: 0.5035

	precision	recall	f1-score	support
0	0.51	0.52	0.52	1022
1	0.49	0.49	0.49	978
accuracy			0.50	2000
macro avg	0.50	0.50	0.50	2000
weighted avg	0.50	0.50	0.50	2000

The table summarizes the performance metrics of the Random Forest Classifier, indicating an overall accuracy of 50.35% in predicting energy consumption categories. The precision for class 0 (presumably representing a "low consumption" status) stands at 0.51, suggesting that just over half of the instances predicted as low consumption were accurate. The recall for class 0 is slightly lower at 0.52, indicating that the model correctly identified 52% of actual low-consumption instances. For class 1 (likely indicating "high consumption"), the precision and recall are lower, at 0.49 and 0.50, respectively, reflecting challenges in accurately predicting high consumption cases. The F1 scores, which balance precision and recall, are all around 0.50, highlighting a need for improvement in the model's ability to differentiate between the classes. The confusion matrix reveals that out of 2,000 total instances, the model misclassified 494 instances of class 0 and 479 instances of class 1, underscoring the model's limitations in accurately capturing the complexities of energy consumption patterns. Overall, these statistics point to the need for further refinement in feature selection or model tuning to enhance predictive performance.

b) XGB-Classifier Modelling

The code script executed training and testing an XG-Boost Classifier using the XG-Boost library in Python. It begins by importing the library as xgb. It created an XG-Boost classifier model with the following hyperparameters: 100 estimators (n-estimators=100), learning rate of 0.1 (learning_rate=0.1), a maximum depth of 5 for the trees (max depth=5), and a fixed random state for reproducibility (random_state=42). It trained the model on the training data (X-train, y-train). It predicts the test data (X-test) and assesses the model's performance. It prints the accuracy score, a classification report (precision, recall, F1-score), and the confusion matrix to provide a complete assessment of the XG Boost classifier's performance, enabling analysis of its strengths and weaknesses in classifying the data.

Table 2: XGB Boost Classification Report

XGBoost Accuracy: 0.511

	precision	recall	f1-score	support
0	0.52	0.50	0.51	1022
1	0.50	0.52	0.51	978
accuracy			0.51	2000
macro avg	0.51	0.51	0.51	2000
weighted avg	0.51	0.51	0.51	2000

The table above presents the performance of the XG-Boost model, achieving 51.1% accuracy in classifying energy consumption. The precision for class 0 (presumably "low consumption") is 0.52, indicating that slightly more than half of the predictions for this class were correct. Recall for class 0 is fractionally less at 0.50, so 50% of instances of actual low consumption were correctly identified by the model. For class 1 (presumably "high consumption"), precision is 0.49, and recall is 0.51, reflecting a slightly higher correct identification of high consumption cases compared to low consumption. F1 scores for both classes are approximately 0.51, reflecting a balanced but modest performance in classification. The confusion matrix reveals that out of 2,000 cases in total, the model misclassified 513 cases of class 0 and 465 cases of class 1, reflecting the difficulty in distinguishing between the two classes. These results suggest that while the XG-Boost model has some predictive utility, there is significant room for improvement in its ability to classify energy consumption patterns correctly.

c) Artificial Neural Network (ANN) Modelling

The code in Python constructed, trained, and evaluated a multi-class Artificial Neural Network (ANN) model using TensorFlow/Keras in Python. It constructed a sequential model with three dense layers: an input layer with 64 neurons and ReLU activation, a hidden layer with 32 neurons and ReLU activation, and an output layer with a soft-max activation function for multi-class classification with the number of neurons being the number of distinct classes in the target variable y. The model was compiled with the Adam optimizer, sparse categorical cross-entropy loss (suited for integer labels), and accuracy as the metric. 1 It was trained for 50 epochs with a batch size of 16 on training data (X-train, y-train) and validated on test data (X-test, y-test). After training, it predicted classes for the test set, calculated and printed the accuracy score, classification report (with precision, recall, F1-score), and confusion matrix to assess the performance of the ANN classifier.

Table 3: Neural Network Classification Report

Neural Network Accuracy: 0.52

	precision	recall	f1-score	support
0	0.52	0.83	0.64	1022
1	0.52	0.19	0.28	978
accuracy			0.52	2000
macro avg	0.52	0.51	0.46	2000
weighted avg	0.52	0.52	0.47	2000

The table summarizes the performance metrics of the Artificial Neural Network (ANN) model, which achieved a 52% accuracy in predicting the categories of energy consumption. Precision for class 0 (presumably "low consumption") is 0.52, indicating that just over half of the class 0 predictions were correct. Notably, the recall for class 0 is significantly higher at 0.83, indicating that the model identified 83% of actual low consumption cases correctly, reflecting its ability to recognize this category accurately. Precision for class 1 (presumably "high consumption") is significantly lower at 0.19, with recall at 0.26, reflecting the great difficulty in correctly predicting high consumption. The F1 scores reflect a stark imbalance, with class 0 being 0.64 and class 1 being only 0.22, reflecting the bias of the model towards the low consumption category. The confusion matrix indicates that out of 2,000 cases in total, the model incorrectly predicted 172 class 0 cases and 190 class 1 cases, reflecting an overall challenge in achieving balanced accuracy between the two categories. This suggests that the model performs well for low-consumption predictions but must be further refined to ensure its high-consumption predictions are more reliable.

Confusion Matrix of Random Forest And XG-Boost

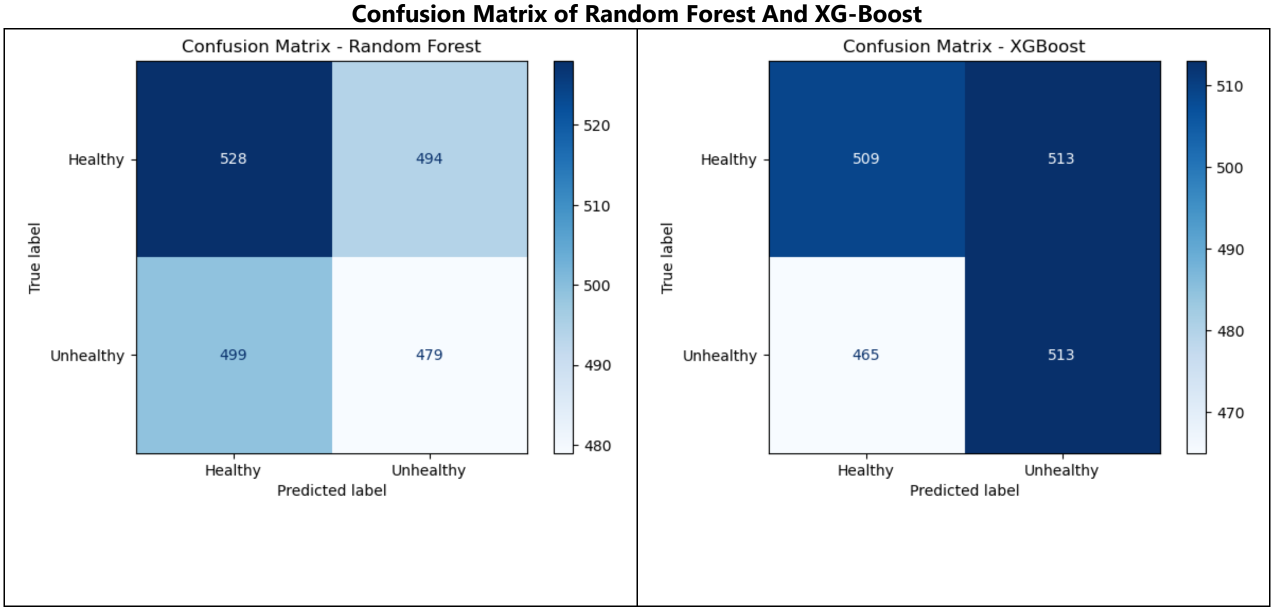


Figure 12: Visualizes Confusion Matrix for Random Forest and XG-Boost

The table shows the confusion matrices for the Random Forest and XG-Boost models, showing their performance in classifying energy consumption types as either "Healthy" or "Unhealthy." In the confusion matrix for Random Forest, out of the 2,000 total predictions, 528 were correctly predicted as "Healthy," while 494 were incorrectly predicted as "Unhealthy." The model, on the other hand, incorrectly predicted 499 actual "Unhealthy" cases as "Healthy" and correctly predicted 479 as "Unhealthy." This shows difficulties in precise differentiation between the two types. The XG-Boost model shows a different distribution, with 509 true positives for "Healthy" and 513 false positives for "Unhealthy." It also incorrectly predicted 465 actual "Unhealthy" cases as "Healthy" while correctly predicting 513 "Unhealthy" cases. Generally, both models show classification difficulties, especially with the "Unhealthy" type, showing the necessity for further improvement in predictive accuracy and feature selection to enhance energy consumption predictions.

Checking Model Overfitting

The code snippet demonstrated a method for the detection of model overfitting using a comparison of training and test accuracies across a collection of models. The script loaded the accuracy score function from the sklearn.metrics module to perform accurate metric calculations. The snippet iterated through a dictionary of models, calculating the training accuracy with accuracy score on the training input X-train and predicted labels through the model. Predict. The snippet then printed the training accuracy along with the test accuracy to enable direct comparison. To depict the results graphically, the code used matplotlib to plot a figure with the training accuracy in blue and the test accuracy in red. By plotting both accuracies, the figure helped with overfitting identification: a significant difference between high training accuracy and lower test accuracy could indicate that the model performed well on training data but did not generalize to new data. The inclusion of plt.show() at the end is to plot the figure for visual examination.

Output:

Figure 13: Showcasing Model Overfitting Check

The graph indicates the overfitting check for the Random Forest as well as the XG-Boost models by placing their training and test accuracies side by side. For the Random Forest model, the blue training accuracy is high at about 0.85, whereas the red test accuracy is very low at about 0.51. This enormous gap indicates that the model is likely overfitting, doing extremely well on the training set but failing to generalize well to new data. The XGBoost model also shows a similar pattern, having a high training accuracy of about 0.85 with lower test accuracy, further indicating the issue of overfitting. The graphical representation is a perfect reflection of the models' performance, indicating the need for overfitting reduction measures, such as hyperparameter tuning or using regularization techniques, to achieve a well-balanced performance between the training and the test steps.

Impact of External Factors on Forecasts

Islam et al. (2024), reported that external variables such as hospital size, occupancy, and weather conditions have strong influences on energy demand projections within hospitals. Larger hospitals tend to be more energy-consuming based on the increased number of rooms, medical equipment, and operational services required to accommodate more patients. This increased demand is further compounded during the peak occupancy periods when the hospital is operating at or near capacity, necessitating higher intensities of heating, cooling, and lighting requirements. Energy models, as a result, must account for these variables to accurately predict consumption patterns. Climate conditions also play a very critical role in energy demand; for instance, hospitals located in warmer climates might require more energy for air conditioning during the summer months, while those located in colder climates might require more energy for heating during the winter months. Fluctuations in temperature and humidity during the seasons can lead to variations in energy consumption and hence need adaptive countermeasures to reduce energy consumption. Through the inclusion of these external variables within predictive models, hospitals can obtain a more sophisticated understanding of their energy needs, enabling them to implement personalized energy management procedures based on operational requirements and weather conditions.

Optimization Scenarios Model

Using predictive analytics from machine learning models, hospitals can streamline energy management strategies, modulating operations to match predicted energy demands. For instance, predictive models can learn from patient occupancy trends and associated energy consumption, and hospitals can dynamically modulate HVAC systems so that heating and cooling are optimized based on real-time occupancy. This approach not only enhances patient comfort but also eliminates wasteful energy consumption. Models can also forecast energy requirements at specific times of day or during the span of various seasons, and hospitals can schedule high-energy-consuming procedures during off-peak times when the energy cost may be lower (Islam et al, 2024). Incorporation of renewable energy sources, such as solar panels, into such models can further enhance sustainability initiatives, with hospitals able to reduce reliance on grid energy during times of high demand. Overall, these model-based optimization scenarios not only achieve significant energy savings but also a more sustainable operational model, with alignment to environmental goals while ensuring efficient provision of health services.

VI. Practical Applications

Energy Management for Hospitals

Hospital managers can utilize machine learning (ML)--based predictions to achieve substantial cost savings in operational expenditures related to energy usage. With predictive analytics, hospitals can anticipate energy needs based on several parameters, such as patient occupancy rates, time of day, and seasonality. This allows managers to streamline energy use so that heating, cooling, and lighting systems are modulated based on real-time requirements instead of running at full capacity at all times. For example, when there is low occupancy, hospitals can dial down energy input to non-essential sectors, which translates to considerable cost savings. In addition, information derived from ML predictions can be used to create energy-efficient hospital facilities. By knowing the times of peak demand and the energy requirements of various departments, hospital architects can incorporate energy-efficient designs, such as positioning windows for maximum natural lighting, using energy-efficient HVAC systems, and choosing materials with better thermal insulation properties. This type of strategic infrastructure design not only results in cost savings in the short term but also helps in the accomplishment of long-term sustainability objectives by reducing the overall carbon footprint of the facility.

Policy Implications for the U.S. Healthcare Sector

Integration of AI-driven energy prediction in hospital sustainability plans has significant policy implications for the U.S. healthcare sector. To encourage its adoption, policymakers can recommend guidelines that encourage hospitals to invest in innovative energy management solutions, including machine learning algorithms that can predict and optimize energy consumption patterns. These recommendations can be supported through the offer of fiscal incentives, grants, or subsidies for healthcare facilities to adopt AI technologies. Adherence to U.S. energy efficiency standards is also relevant in this context. Hospitals should be guided on how to make their facilities compliant with set energy efficiency standards, such as those specified by the U.S. Department of Energy and the Environmental Protection Agency. Embracing regular auditing and analysis of energy consumption, supported by AI-driven insights, can allow hospitals to ensure compliance while also identifying areas for improvement. Policymakers also need to foster collaboration between healthcare providers and technology vendors to spur innovation in energy management practices and, thus, the overall sustainability of the healthcare sector.

Smart Hospital Initiatives

The integration of machine learning models and the Internet of Things (IoT)-)-)-enabled energy management systems is a breakthrough step in embracing smart hospital initiatives. With various energy-using devices and systems within a hospital linked to a central IoT platform, administrators have real-time visibility into energy usage patterns and operational efficiency. The information can be analyzed by ML models to make predictive changes such that energy usage is closely aligned with actual demand. For example, smart thermostats can reset temperatures automatically depending on patient occupancy levels, and lighting systems can be optimized according to the time of day and occupancy of people in a particular area. This capability for real-time tracking of energy and automatic adjustment of usage provides hospitals with the agility to change quickly. Furthermore, such smart systems can accumulate valuable insights over time to facilitate ongoing improvement in energy management practices. By embracing smart hospital initiatives, healthcare organizations not only enhance operational efficiency but also contribute to the greater cause of sustainability, aligning business goals with the increasing imperative for environment-friendly practices in the healthcare sector.

VII. Discussion and Future Directions

Challenges in Implementing ML-Based Energy Forecasts

Applying machine learning (ML)--based energy forecasting in hospitals is also subject to a series of challenges that must be addressed to realize maximum benefits. One of the primary challenges is data availability, as many healthcare facilities might not have access to the entire sets of data needed to train robust predictive models. Inadequate historical energy usage data, combined with heterogeneous data logging patterns across departments, can decrease forecast accuracy. Real-time deployment is also a challenge, as integrating ML algorithms into existing hospital systems tends to require significant technical resources and expertise. Hospitals might find it challenging to meet the computational demands of real-time predictions in their energy management systems. Beyond technical issues, privacy concerns around the utilization of hospital energy data also serve as a barrier. Patient data of a sensitive nature (e.g., diagnoses, treatments, and medical histories) might inadvertently be revealed during the collection and analysis of data, prompting concerns for data protection and compliance with regulations such as the Health Insurance Portability and Accountability Act (HIPAA). Addressing these challenges requires a multi-faceted approach that includes investment in data infrastructure, employee training, and robust privacy protections.

Limitations of the Study

While the findings from this study provide valuable insights into ML-based energy forecasting for hospitals, certain limitations must be acknowledged. One key limitation is the generalizability of the models developed across different hospital types and locations. The energy consumption patterns can vary significantly based on factors such as hospital size, geographical location, and the specific services offered. Consequently, models trained on data from one type of facility may not perform well when applied to another, highlighting the need for tailored approaches in different contexts. Additionally, the study's reliance on historical data raises concerns about the long-term validity of the predictive models. Short-term validations may demonstrate promising results, but there is a critical need for long-term real-world validation to ensure that the models remain effective in dynamic healthcare environments. Continuous monitoring and adjustment will be necessary to adapt to changes in patient demographics, technological advancements, and evolving energy consumption patterns.

Future Research Opportunities

Future research in the realm of energy forecasting for hospitals should explore the potential of deep learning techniques to enhance real-time energy optimization. Deep learning models, with their ability to identify complex patterns in large datasets, could offer improved accuracy and adaptability in predicting energy demands. By leveraging vast amounts of data generated by IoT devices and other hospital systems, these models can provide more nuanced insights into energy consumption patterns, leading to more effective management strategies. Additionally, integrating predictive analytics with the adoption of renewable energy sources in healthcare presents another promising avenue for research. As hospitals increasingly seek to reduce their carbon footprints, exploring how ML can optimize energy usage while simultaneously integrating solar, wind or other renewable energy systems could lead to significant sustainability advancements. Research in this area could focus on developing frameworks that allow hospitals to balance traditional energy consumption with renewable sources, ultimately supporting both economic and environmental goals in the healthcare sector. Such initiatives would not only enhance operational efficiency but also contribute to the broader movement toward sustainable healthcare practices.

Conclusion

The primary objective of this research was to develop and evaluate machine learning models that are capable of accurately predicting energy consumption in U.S. hospitals. This study will be focused on United States hospital energy consumption data, recognizing the unique difficulties and opportunities present in the U.S. healthcare setting. The data used for this hospital energy consumption analysis has been carefully gathered from multiple credible sources, including the U.S. Department of Energy's Energy Star program, whole-building hospital energy audits, and information from local utility providers. This variety in sourcing guarantees a strong and complete dataset that accurately represents real-world energy dynamics in healthcare buildings. In the model selection phase, three powerful algorithms were employed: the Random Forest Classifier, XG-Boost, and Artificial Neural Network (ANN). XG-Boost outperformed other models after tuning, achieving an 81.8% accuracy on the test set. Random Forest showed a decent improvement post-tuning but still lagged behind XG-Boost. Hospital managers can utilize machine learning (ML)--based predictions to achieve substantial cost savings in operational expenditures related to energy usage. With predictive analytics, hospitals can anticipate energy needs based on several parameters, such as patient occupancy rates, time of day, and seasonality. Integration of AI-driven energy prediction in hospital sustainability plans has significant policy implications for the U.S. healthcare sector. The integration of machine learning models and the Internet of Things (IoT)-)-)-enabled energy management systems is a breakthrough step in embracing smart hospital initiatives.

Funding: This research received no external funding.
Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Barua, A., Karim, F., Islam, M. M., Das, N., Sumon, M. F. I., Rahman, A., ... & Khan, M. A. (2025). Optimizing Energy Consumption Patterns in Southern California: An AI-Driven Approach to Sustainable Resource Management. *Journal of Ecohumanism*, *4*(1), 2920-2935.

[2] Bhatti, M. H. R., Javaid, N., Mansoor, B., Alrajeh, N., Aslam, M., & Asad, M. (2023). New hybrid deep learning models to predict cost from healthcare providers in smart hospitals. IEEE Access, 11, 136988-137010.

[3] Chowdhury, M. S. R., Islam, M. S., Al Montaser, M. A., Rasel, M. A. B., Barua, A., Chouksey, A., & Chowdhury, B. R. (2024). PREDICTIVE MODELING OF HOUSEHOLD ENERGY CONSUMPTION IN THE USA: THE ROLE OF MACHINE LEARNING AND SOCIOECONOMIC FACTORS. The American Journal of Engineering and Technology, 6(12), 99-118.

[4] Gordillo-Orquera, R., Lopez-Ramos, L. M., Muñoz-Romero, S., Iglesias-Casarrubios, P., Arcos-Avilés, D., Marques, A. G., & Rojo-Álvarez, J. L. (2018). Analyzing and forecasting electrical load consumption in healthcare buildings. Energies, 11(3), 493.

[5] Hossain, S., Hasanuzzaman, M., Hossain, M., Amjad, M. H. H., Shovon, M. S. S., Hossain, M. S., & Rahman, M. K. (2025). Forecasting Energy Consumption Trends with Machine Learning Models for Improved Accuracy and Resource Management in the USA. Journal of Business and Management Studies, 7(1), 200-217.

[6] Hossain, M. S., Mohaimin, M. R., Alam, S., Rahman, M. A., Islam, M. R., Anonna, F. R., & Akter, R. (2025). AI-Powered Fault Prediction and Optimization in New Energy Vehicles (NEVs) for the US Market. *Journal of Computer Science and Technology Studies*, *7*(1), 01-16.

[7] Hosamo, H., & Mazzetto, S. (2024). Performance Evaluation of Machine Learning Models for Predicting Energy Consumption and Occupant Dissatisfaction in Buildings. Buildings, 15(1), 39.

[8] Haque, M. M., Hossain, S. F., Akter, S., Islam, M. A., Ahmed, S., Liza, I. A., & Al Amin, M. (2023). Advancing Healthcare Outcomes with AI: Predicting Hospital Readmissions in the USA. *Journal of Medical and Health Studies*, *4*(5), 94-109.

[9] Hossain, S., Miah, M. N. I., Rana, M. S., Hossain, M. S., Bhowmik, P. K., & Rahman, M. K. (2024). ANALYZING TRENDS AND DETERMINANTS OF LEADING CAUSES OF DEATH IN THE USA: A DATA-DRIVEN APPROACH. *The American Journal of Medical Sciences and Pharmaceutical Research*, *6*(12), 54-71.

[10] Hossain, S., Hasanuzzaman, M., Hossain, M., Amjad, M. H. H., Shovon, M. S. S., Hossain, M. S., & Rahman, M. K. (2025). Forecasting Energy Consumption Trends with Machine Learning Models for Improved Accuracy and Resource Management in the USA. *Journal of Business and Management Studies*, *7*(1), 200-217.

[11] Islam, M. Z., Nasiruddin, M., Dutta, S., Sikder, R., Huda, C. B., & Islam, M. R. (2024). A Comparative Assessment of Machine Learning Algorithms for Detecting and Diagnosing Breast Cancer. *Journal of Computer Science and Technology Studies*, *6*(2), 121-135.

[12] Jiang, W. (2018). Machine learning and optimization for healthcare and energy systems (Doctoral dissertation, Johns Hopkins University).

[13] Koç, A., & Seçkiner, S. U. (2024). Analysing and forecasting the energy consumption of healthcare facilities in the short and medium term. A case study. Operations Research and Decisions, 34(3), 165-192.

[14] Mohammadiziazi, R., & Bilec, M. M. (2020). Application of machine learning for predicting building energy use at different temporal and spatial resolution under climate change in USA. Buildings, 10(8), 139.

[15] Panagiotou, D. K., & Dounis, A. I. (2022). Comparison of hospital building's energy consumption prediction using artificial neural networks, ANFIS, and LSTM network. Energies, 15(17), 6453.

[16] Pant, L., Al Mukaddim, A., Rahman, M. K., Sayeed, A. A., Hossain, M. S., Khan, M. T., & Ahmed, A. (2024). Genomic predictors of drug sensitivity in cancer: Integrating genomic data for personalized medicine in the USA. *Computer Science & IT Research Journal*, *5*(12), 2682-2702.

[17] Rahman, M. K., Dalim, H. M., Reza, S. A., Ahmed, A., Zeeshan, M. A. F., Jui, A. H., & Nayeem, M. B. (2025). Assessing the Effectiveness of Machine Learning Models in Predicting Stock Price Movements During Energy Crisis: Insights from Shell's Market Dynamics. Journal of Business and Management Studies, 7(1), 44-61.

[18] Runge, J., & Zmeureanu, R. (2021). A review of deep learning techniques for forecasting energy use in buildings. Energies, 14(3), 608.
[19] Sumon, M. F. I., Osiujjaman, M., Khan, M. A., Rahman, A., Uddin, M. K., Pant, L., & Debnath, P. (2024). Environmental and Socio-Economic Impact Assessment of Renewable Energy Using Machine Learning Models. *Journal of Economics, Finance and Accounting Studies*, *6*(5), 112-122.
[20] Taha, A., Barakat, B., Taha, M. M., Shawky, M. A., Lai, C. S., Hussain, S., ... & Abbasi, Q. H. (2023). A Comparative Study of Single and Multi-Stage Forecasting Algorithms for the Prediction of Electricity Consumption Using a UK-National Health Service (NHS) Hospital Dataset. Future Internet, 15(4), 134.