LSTM AND SHAP FOR TRANSPARENT AND EFFECTIVE IOB SYSTEMS IN IOT ENVIRONMENTS: HOUSEHOLD POWER CONSUMPTION

Lappeenranta–Lahti University of Technology LUT

Master's thesis

M.Sc Software Engineering and Digital Transformation

2023

Revathi Karri

Examiner(s): Professor Najmul Islam PhD

Prabhat Kumar, Post-doctoral researcher

ABSTRACT

Lappeenranta–Lahti University of Technology LUT

LUT School of Engineering Science

Software Engineering and Digital Transformation

Revathi Karri

LSTM AND SHAP FOR TRANSPARENT AND EFFECTIVE IOB SYSTEMS IN IOT ENVIRONMENTS: HOUSEHOLD POWER CONSUMPTION

Master's thesis

2023

54 pages, 16 figures, 2 tables and 0 appendices

Examiner(s): Professor Najmul Islam, PhD and Prabhat Kumar, Post-doctoral researcher

Keywords: LSTM, SHAP, Internet of Things, IoB, Household Power Consumption, Explainable AI

The integration of Long Short-Term Memory (LSTM) model and SHapley Additive exPlanations (SHAP) within the Internet of Things (IoT) lays the groundwork for transparent and effective Internet of Behaviour (IoB) systems, particularly in managing household power consumption. This thesis explores the development and validation of an LSTM-based forecasting model, employing SHAP for enhanced interpretability and user engagement within IoB frameworks. By leveraging minute-level household power consumption data, the study reveals the intricate dynamics of energy usage and provides a methodology for integrating explainable AI in household energy management systems. This integration helps the users to encourage the adoption of energy-efficient behaviours. The research findings indicate that LSTM, when coupled with SHAP, significantly improves the transparency and utility of IoB systems, offering actionable insights for energy efficiency and user behaviour modification. This thesis contributes to the discourse on sustainable energy practices, underlining the critical role of user-centric, interpretable AI in fostering energy-conscious behaviours and efficient resource utilization.

ACKNOWLEDGEMENTS

I would like to express my deepest thanks to all those who provided me the possibility to complete this thesis. I thank my family and friends for continuous support and motivation. A special gratitude and thanks to Professor Najmul Islam for the great support and Prabhat Kumar for guiding me in the right direction and helping whenever I needed. I am grateful and appreciate all your help.

Revathi Karri

ABBREVIATIONS

AI	Artificial Intelligence
XAI	Explainable AI
SHAP	Shapely Additive Explanations
LSTM	Long Short-Term Memory
ML	Machine Learning
IoT	Internet of Things
IoB	Internet of Behaviour
SVR	Support Vector Regression
RF	Random Forest
MLP	Multi-Layer Perception
RNN	Recurrent Neural Network
CNN	Convolution Neural Network
DB-Net	Deep-Broad Network
Bi-LSTM	Bidirectional Long Short-Term Memory
ARIMA	Auto Regressive Integrated Moving Average
CNN-LSTM	Convolutional Neural Network
ConvLSTM	Convolutional LSTM
TUI	Tangible User Interface
EMC	Energy Monitor and Controller
GPU	Graphics Processing Unit
API	Application Programable Interface
UCI	University of California, Irvine

CSV	Comma Separated Vale
RMSE	Root Mean Square Error

Table of contents

Table of Contents

Figures

Tables

1 Introduction

Human activities have led to significant alterations in the Earth's atmosphere, and to attain a sustainable planet, it is essential to change human habits and behaviours. Technology alone cannot address the challenges of global sustainability; it should be coupled with human behaviour. Due to urbanization, industrialization and economic growth, energy crisis is increased (Karanfil & Li, 2015; Sorrell, 2015). Household power consumption is one of major contributions for energy consumption and carbon emissions. People are unaware of energy forecast and their usage habits and which is leading to unnecessary energy waste. Sustainability and energy efficiency are paramount, optimizing household power consumption is a pressing concern. Humans can take advantage of digital technologies to understand their behaviour of energy usage patterns and take conscious choices (Elayan et al., 2022).

In the current era of digital transformation, the Internet of Things (IoT) has significantly reshaped household environments by interconnecting automated appliances like lights, thermostats, dishwasher etc and security systems for seamless interaction to enhance our daily lives and energy management (Lutui et al., n.d.). This interconnected system of IoT has emerged pivotal concept of Internet of Behaviour (IoB) leveraging data from IoT to influence and understand human behavior by using Artificial Intelligence (AI) to provide data-driven insights, energy prediction and intelligent decisions to control energy consumption (Elayan et al., 2023).

1.1 Problem statement

The AI-driven IoT system increasingly impact our daily lives. However, there are some concerns from users about their data usage and privacy in IoT systems. The opaque nature of AI model in IoB systems raises questions about their transparency and effectiveness in managing household power consumption and prediction. The lack of clarity can create a reluctant behaviour in users and hinders adopting this system. To address this, it is crucial to integrate eXplainable AI (XAI) technique SHapley Additive exPlanations (SHAP) in IoB system to provide trusted and effective environment by explain AI model process.

1.2 Literature context and Research Gaps

Numerous studies have examined energy consumption forecasting using various AI models and SHAP. Though these studies have contributed valuable insights however there is a gap in literature in explaining AI decisions within household power consumption significantly in IoB-XAI framework where understanding AI-driven recommendations can lead users to adopt energy-efficient behaviors. The research article (Elayan et al., 2023) highlights IoB-XAI framework and LSTM in the context of household power consumption. It also mentioned the future work on instructive notifications potential in influencing user behavior. However, it does not fully explore the integration of SHAP with LSTM models to enhance the clarity and effectiveness of these notifications, a gap this thesis aims to address.

1.3 Research aims and questions.

The main aim of the thesis is to integrate SHapley Additive exPlanations (SHAP) with LSTM model to enhance the transparency and effectiveness of IoB systems in IoT environments. The specific objectives are to:

- Develop an LSTM model to forecast energy consumption.
- Incorporate SHAP to elucidate the decision-making process of the LSTM model.
- Evaluate the impact of SHAP explanations on user understanding and system effectiveness by investigating different SHAP plots.
- Explore SHAP's practical application within IoB for more efficient household energy management.

This thesis addresses the following research questions:

RQ1) How to develop and validate LSTM forecasting model for household power consumption, considering minutely and hourly horizons?

RQ2) How can SHAP be used to enhance the interpretability of LSTM forecasting model for household energy consumption?

RQ3) How can SHAP insights be employed within IoB to influence IoT behavior for household energy management?

1.4 Research Methodology

The research methodology involves gathering and preparing a comprehensive dataset on household energy consumption, applying an LSTM model for accurate power consumption forecasting, and utilizing SHAP plots to gain insights into the decision-making process of the model. The methodology aims to assess the predictive accuracy of the models as well as evaluate how effectively SHAP can provide meaningful interpretability to understand the model's decisions.

1.5 Structure

The thesis is organized as follows:

- **Chapter 1: Introduction** - Sets the stage for the research, outlining the aims, methodology, main concepts, and structural overview.
- **Chapter 2: Literature Review** - Surveys existing literature on power consumption forecasting, advanced AI models, and the role of explainability in AI.
- **Chapter 3: Methodology and implementation** - Details the model design, data collection, model development, and analytical techniques employed.
- **Chapter 4: Model Analysis and Results** - Discusses the performance of LSTM and presents the findings from the SHAP analysis.
- **Chapter 5: Discussion** - Interprets the results in the context of IoB, discussing the practical implications for energy management and user behavior.
- **Chapter 6: Conclusion and Future Directions** - Concludes the thesis with a summary of findings, contributions to the field, limitations, and suggestions for future research.

2 Literature review

The literature review serves as a foundation for understanding the context and existing knowledge related to forecasting household power consumption and the integration of SHAP (SHapley Additive exPlanations) in IoB (Internet of Behavior) systems within IoT (Internet of Things) environments. This section explores the historical evolution of IoT technology, its impact on household environments, and the emergence of the Internet of Behavior.

2.1 Historical Context and Integration of IoT in Household Energy Management

The expansion of the Internet of Things (IoT) has markedly impacted the way energy is managed in households. This review delves into the evolution of IoT technologies, examining how they have affected domestic energy usage. It also looks at both the challenges and advancements that have played a role in their adoption for more efficient energy management.

2.1.1 Evolution of IoT in Energy Management

The Internet of Things (IoT) is recognized as a significant development in the evolution of the internet, with transformative effects on energy consumption in households (Gray, 2018). It is expected to greatly influence daily life and introduce new business models for both consumers and enterprises. A notable prediction by A. Elnashar in 2018 suggested that the number of connected IoT devices could rise to as many as 50 billion by 2022, illustrating the rapid growth and impact of IoT technology (A.Elnashar, 2018).

2.1.2 Challenges in IoT and approaches for solution

The IoT need to use some energy to interconnect the appliances. (Hittinger & Jaramillo, 2019) emphasize the need for cooperation between scientists and regulators to balance IoT's benefits with the potential rise in its energy usage. This is echoed by (Arshad et al., 2017), who critically analyzes green IoT strategies and proposes principles for minimizing energy consumption.

2.1.3 IoT in Smart Homes and energy monitoring

Smart homes, as reviewed by (Soudan et al., n.d.)exemplify the application of IoT technologies in managing and controlling household energy. (Abdullah et al., 2021) further describes how smart home technologies enable users to manage energy usage through various environmentally conscious strategies, making it easier for humans to track and control home devices. Also, IoT system is used in monitoring the electricity along with water demand in a house (Bautista-Villalon et al., 2018). Intelligent IoT systems can save energy (Lai et al., 2012).

2.1.4 Conclusion

The literature review has highlighted the significant impact of the Internet of Things (IoT) on household energy management, from its evolutionary beginnings to its current applications in smart homes. Studies by authors like Gray, Elnashar, Hittinger & Jaramillo, and others have emphasized IoT's revolutionary role in transforming energy consumption patterns. Key challenges, including energy demands for device interconnectivity and the need for sustainable strategies, have been identified. The application of IoT in smart homes, as demonstrated by Soudan et al., Abdullah et al., and others, illustrates the technology's capability in enhancing energy efficiency and user engagement. The Overall, This literature review on the historical context and integration of IoT showcases the huge potential in the energy management which makes use of technological advancements for a better environment.

2.2 Internet of Behaviour (IoB)

Internet of Behaviour (IoB) is an extension of IoT by gathering the details with regard to the human behaviour. This concept is categorized and popularized by Gartner technological trends. IoB can leverage technology to monitor behavioral events and manage the data to influence human behavior. This concept essentially blends technology, data analytics, and behavioral science.

The primary purpose of the Internet of Behavior (IoB) is to analyze and utilize data collected from various sources, particularly IoT devices, to understand and influence human behavior. By gathering, analyzing, and interpreting data from devices such as smartphones, wearables, home appliances, and even vehicles, IoB aims to gain insights into user behavior, preferences, and interactions. These insights are then used to improve customer experiences, enhance products and services, inform business strategies, and potentially drive behavior change.

2.2.1 Applications of IoB

- **Organizational Transformation:** IoB's transformative role in organizations is detailedin (Stary, 2020) .
- **Healthcare:** Enhancing healthcare services using ML is explored in (Amiri et al., 2023) and (Sun et al., 2023).
- **Cybersecurity:** (Rjoub et al., 2023) examines the integration of IoB with XAI in cybersecurity.
- **Education:** IoB's application in educational cognitive performance assessment is presented in (Srimaharaj & Chaisricharoen, 2022).
- **Digital Marketing:** (Afor & Sahana, 2022) discusses IoB in digital marketing.
- **Household Energy Prediction or Management:** The studies (Elayan et al., 2023) and (Elayan et al., 2022) explore IoB's potential in managing and predicting household energy use. These papers highlight the effectiveness of IoB in reducing energy consumption through influencing IoT device behavior, showcasing the power of decentralized IoB frameworks in energy efficiency.

2.2.2 XAI in IoB

The integration of XAI with IoB, crucial for enhancing transparency and trust in AI systems, is elaborated in (Elayan et al., 2023). Furthermore (Rjoub et al., 2023) provides a comprehensive analysis of XAI in cybersecurity, highlighting the need for methods that balance accuracy and transparency, handle uncertainty, and address ethical implications.

2.2.3 Research Gaps and Challenges

- **Ethical and Privacy Concerns**: Ethical dilemmas and privacy issues with IoB's data collection and behavioural analysis which are addressed in (Afor & Sahana, 2022; Sun et al., 2023)
- **Technical and Analytical Challenges**: Advanced predictive models' need and effective IoB integration with AI and ML which are discussed in (Amiri et al., 2023; Elayan et al., 2023)
- **Societal and Cultural Considerations**: Understanding IoB's implications in different cultural and socioeconomic contexts which are emphasized in (Srimaharaj & Chaisricharoen, 2022)
- **Security in IoB Applications**: Developing robust security measures for IoB systems in sectors like healthcare and marketing. Explored in (Afor & Sahana, 2022; Elayan et al., 2022).

2.2.4 Conclusion

The IoB, as explored in these studies, shows immense potential for influencing human behavior in areas like healthcare, education, and energy management. Its integration with XAI, particularly in cybersecurity, enhances transparency and trustworthiness. However, balancing technological advancements with ethical, privacy, and security considerations is vital. Future research should aim to develop IoB systems that are transparent, ethical, secure, and culturally sensitive.

2.3 Energy Forecasting Models in AI

Household energy forecasting models significantly benefit households by enabling more efficient and cost-effective energy usage. Through accurate predictions of energy needs, these models help in optimizing consumption patterns, leading to cost savings on energy bills. They also facilitate the adoption of energy-saving measures and the integration of smart home technologies for enhanced comfort and convenience. Furthermore, by optimizing energy use, households can reduce their environmental footprint, contributing to lower greenhouse gas emissions. This informed approach to energy management also aids households in making better decisions about investing in energy-efficient technologies and improvements. Forecasting models, essential tools in predicting future events based on historical data, come in various forms, ranging from traditional to AI-based models. This literature review delves into various AI-based models and methodologies applied in household energy forecasting, highlighting their objectives, methodologies, strengths, and limitations.

2.3.1 AI-Based Forecasting Models

1) **Traditional Machine Learning Approaches**: There are many traditional ML techniques used in power consumption forecasting in initial stages. Traditional models like regression analysis performed by (Fumo & Rafe Biswas, 2015), decision trees another approach, offer interpretable models for decision-making processes (Tso & Yau, 2007), Support Vector Regression (SVR) known for its robustness, excels in high-dimensional spaces (Bilal et al., n.d.), Random Forests (Candanedo et al., 2017) a powerful ensemble technique which improves prediction accuracy by averaging multiple decision trees, and Gradient Boosting Machines (Li & Wang, 2022a; Zhao & Magoulès, 2012) are foundational in energy forecasting. While these models are generally easier to interpret and less computationally demanding than deep learning models, they often require extensive feature engineering and may not capture complex, non-linear relationships as effectively. For instance, regression analysis, widely used for its simplicity and interpretability, may not perform well with non-linear data prevalent in energy consumption patterns.

1. **Long Short-Term Memory (LSTM) Networks**:

   A popular choice in AI for energy forecasting, LSTM models, a type of recurrent neural network (RNN), are particularly adept at handling time series data. They can capture long-term dependencies in energy consumption patterns, making them ideal for both short-term and long-term forecasting (Elayan et al., 2023; Kong et al., 2019). LSTM's ability to process sequential data of varying lengths is crucial in adapting to the dynamic nature of household energy consumption.

2. **Convolutional Neural Networks (CNNs)**: CNNs, primarily known for their success in image processing, have been adapted for energy forecasting. These models excel in feature extraction from temporal data, which is essential for identifying consumption patterns in household energy usage (Cascone et al., 2023; Ullah et al., 2021; Kim & Cho, 2019).

3. **Hybrid Models**: Combining CNNs with LSTM layers (Alhussein et al., 2020; Z. A. Khan et al., 2020), hybrid models like DB-Net have been proposed to harness the strengths of both architectures (N. Khan et al., 2021). These models use CNN layers for feature encoding and LSTM layers to learn sequential patterns, thereby improving the accuracy of energy consumption predictions.

4. **Bidirectional LSTM (Bi-LSTM)**: Extending the capabilities of LSTM, Bi-LSTM models process data in both forward and reverse directions, offering a more comprehensive understanding of energy usage patterns (Dinh et al., 2023; Guimarães Da Silva et al., n.d.).

Comparison of different AI models:

Table 1 Comparison of different AI models

Model	**Accuracy**	**Computational Demand**	**Ease of Use & Interpretability**	**Data Requirements**
Traditional ML Models	Moderate	Low	High	Moderate
LSTM	High	Moderate-High	Moderate	High
CNN	High	High	Low	High
Hybrid Models	Very High	Very High	Low	High
Bi-LSTM	Very High	Very High	Low	High

2.3.2 Challenges and Limitations

1. **Data Quality and Availability**: The effectiveness of AI models heavily relies on the quality and granularity of available data. Inconsistent and missing data points can significantly hinder the model's performance (Alhussein et al., 2020).
2. **Complexity in Implementation**: The implementation of sophisticated AI models requires substantial computational resources and expertise in model tuning, which may not be feasible for all applications (N. Khan et al., 2021).
3. **Interpretability (XAI)**: Deep learning models, particularly neural networks, are often considered 'black boxes' due to their lack of interpretability. This poses a challenge in understanding model decisions and gaining trust from end-users. There is an increasing demand for explainable models in energy forecasting to enhance user trust. XAI techniques can make AI operations more understandable and thus more trustworthy (Elayan et al., 2023).
4. **Security and Privacy Concerns**: The integration of AI and ML in energy forecasting raises significant concerns about data security and privacy, especially given the sensitivity of consumer behavioral data (Elayan et al., 2023).
5. **Computational Complexity**: Implementing sophisticated AI and ML models often demands considerable computational resources, posing challenges in scalability and practical application (Cascone et al., 2023).

2.3.3 Multivariate LSTM

Multivariate LSTM, a variant of LSTM specialized in handling multi-feature inputs. It is a powerful tool for modelling and predicting scenarios where the outcome is influenced by a complex interplay of multiple factors. Its ability to handle multiple data streams simultaneously makes it highly valuable in many advanced data analysis tasks. Table 2 shows the comparative analysis of traditional LSTM, Bi-Directional LSTM, and Multivariate LSTM.

Table 2: Comparative Analysis of different LSTM variations

Feature / Model	Traditional LSTM	Bidirectional LSTM (Bi-LSTM)	Multivariate LSTM
Purpose	Handle sequential data with temporal dependency.	Handle sequences with context from past and future.	Handle sequences with multiple input variables per time step.
Input Data Type	Typically, univariate but can be multivariate.	Can be both univariate and multivariate.	Multivariate (multiple features per time step).
Use Case	Time-series forecasting, text generation, etc.	Natural language processing, context-sensitive applications.	Complex forecasting tasks (e.g., financial, environmental) with multiple influencing factors.
Advantages	Good for single-directional temporal data.	Captures more context by looking at data in both directions.	Capable of handling complex, multi-faceted data scenarios.
Complexity	Relatively simpler model.	More complex due to bidirectional processing.	Complexity varies with the number of variables and data pre-processing.

I chose Multivariate LSTM training model for this application, because this variant of LSTM is better at forecasting household energy based on the learnings of readings of multiple devices. The research by (Praveen Kumar & Hariharan, 2020) by developed a model with multivariate LSTM for improving traffic flow prediction in Intelligent Transportation Systems (ITS), which is crucial for smart city development. Even though traffic flow prediction and household energy are different in theory, but data analysis and training are same in this regard. The model uses multiple traffic parameters for prediction, specifically: Traffic flow, Speed, and Occupancy. These metrics are correlated and considered for forecasting traffic using the multivariate LSTM model. This research concluded that multivariate LSTM model showed superior performance in terms of prediction accuracy compared to the other LSTM models. The research by (Zhang et al., 2020a) used multivariate LSTM models in the forecasting productivity of industrial IoT (IIoT). It was practically implemented with production data from a real industry in China and this model significantly improved prediction accuracy and granularity of industrial productivity forecasting compared to traditional methods like ARIMA, RNN, and basic LSTM. By examining the two research (Praveen Kumar & Hariharan, 2020; Zhang et al., 2020b), it made sense in using multivariate LSTM for forecasting household energy. I customized the model with multivariate LSTM training and integrated it with SHAP which will be explained in the subsequent chapters.

2.3.4 Conclusion

The integration of AI and ML in household energy forecasting marks a significant advancement in energy management. LSTM variants and hybrid models like CNN-LSTM and ConvLSTM have been at the forefront of this research field. However, challenges such as ensuring data security, managing computational complexity, and enhancing model explainablility remain. Future research should focus on these areas to ensure the effective and ethical application of AI and ML in household energy forecasting, paving the way for more sustainable and intelligent energy management solutions.

2.4 SHAP (SHapley Additive exPlanations) in Explainable AI (XAI)

In the evolving field of artificial intelligence (AI), the demand for explainablility has led to the development of Explainable AI (XAI) methods, among which SHapley Additive exPlanations (SHAP) stands out. This literature review delves into the explainable AI (XAI), role of SHAP in XAI, exploring its theoretical foundations, applications, and significance in making AI models more transparent and understandable.

2.4.1 Explainable AI:

Explainable Artificial Intelligence (XAI) is a rapidly evolving field addressing the need for transparency and understandability in increasingly complex AI systems. This literature review synthesizes key findings and themes from recent research papers to provide a comprehensive overview of the state, challenges, and future directions in XAI.

1. **Evolution and Broadening of XAI**:

The field of XAI initially concentrated on providing post-hoc explanations for decisions by deep learning architectures, especially in image classification. However, as "Explainable AI" highlights, there is now a growing emphasis on interactive, user-centric approaches, including verbal explanations and hybrid methods combining reasoning and learning (Schmid & Wrede, 2022a).

2. **Transparency and Trust in AI Systems**:

The paper (Rudin, 2019) underlines the critical need for transparent and interpretable models over black-box models in high-stakes domains like healthcare and criminal justice. This approach is essential for fostering trust and preventing potential harm. Similarly, (Oniani et al., 2023) discusses the application of ethical principles in healthcare AI, emphasizing transparency and accountability.

3. **Challenges in XAI**:

A major challenge in XAI, as outlined in (Schmid & Wrede, 2022b), is developing interpretability methods that are faithful and adaptable to specific user needs. The evaluation of these methods requires rigorous empirical approaches based on psychological theories (Barredo Arrieta et al., 2020).

4. **Tangible XAI**:

The integration of XAI with tangible user interfaces (TUI) is explored in (Colley et al., 2022), proposing the TangXAI framework. This innovative method seeks to enhance the transparency of AI systems through physical interaction, thereby creating more user-centered and intuitive AI systems.

5. **Future Directions in XAI**:

Future studies in XAI aim to improve the efficiency of current techniques and create novel interpretability methods. This involves incorporating XAI into dynamic online learning settings and elucidating model alterations instead of just the models themselves (Schmid & Wrede, 2022a). The paper (Leiter et al., 2022) also envisions future approaches to explainable evaluation metrics, focusing on transparency and trust in high-quality metrics.

In conclusion, advancements in XAI research are shifting towards more interactive, user-focused, and transparent methods. These developments are essential for maintaining the reliability and ethical use of AI in different critical situations. Future research is set to advance this area by emphasizing the creation of more intuitive, ethical, and efficient AI systems.

2.4.2 SHapley Additive exPlanations SHAP

SHapley Additive exPlanations (SHAP) is an ML technique that explains individual predictions of any model by quantifying the contribution of each feature. The application of SHAP in XAI has attracted significant interest recently. It is based on cooperative game theory, provides a technique for elucidating the results of machine learning models, especially those categorized as 'black boxes'. It is model agnostic, meaning it can be applied to any machine learning model, and provides both local and global insights by analysing how different features affect predictions. This approach helps in understanding complex models and making their decisions transparent and interpretable. Since the AI applications are moving towards explainable models, they need to be transparent. SHAP makes this possible by demystifying these models and explain the reasoning behind a prediction (Gianfagna & Di Cecco, 2021a). It offers understanding into the contributions of various features to model predictions, providing a trade-off between precision and interpretability (Chromik, 2020).

SHAP works by breaking down a model's prediction into the contribution of each feature. For a given prediction, it calculates the average impact of adding or removing a feature on the model's output, compared to a baseline prediction. This process involves computing the Shapley values from game theory for each feature, which requires evaluating the prediction with all possible combinations of the features. Essentially, SHAP assesses how much each feature pushes the model's prediction away from a baseline, thus quantifying its importance in the specific prediction. The outcome is a set of SHAP values, one for each feature, indicating the magnitude and direction (positive or negative) of that feature's effect on the prediction (Gianfagna & Di Cecco, 2021b). Application of SHAP is so diverse that it can be employed at any industry that uses AI for prediction. Some of the applications are discussed in subsequent chapters.

2.4.3 Application and Effectiveness:

SHAP has been used in various domains, including energy or power systems, financial services, healthcare etc. Its effectiveness lies in its ability to decompose model predictions into individual feature contributions, thus providing a more granular understanding of model behavior. Studies have demonstrated SHAP's utility in clarifying decision-making processes in fields that require a high level of accountability, such as power systems, where experts need to trust and justify decisions made by AI models (Machlev et al., 2022).

Power System: SHAP has been applied to improve power load forecasting, where an improved XGBoost model based on grid search (GS-XGBoost) was used along with SHAP for model interpretation. This methodology demonstrated the impact of feature sets on the model, significantly improving model interpretability and aiding in decision-making for power system management (Li & Wang, 2022b; Machlev et al., 2022).

Healthcare:

In healthcare, specifically in male fertility prediction, SHAP has been utilized to interpret AI models. By analyzing various industry-standard models, SHAP provides essential explanations for clinicians for treatment planning, enhancing the credibility of AI model predictions in clinical settings (GhoshRoy et al., 2023).

2.4.4 Challenges and User-Centric Improvements

Despite its advantages, SHAP can be challenging to interpret, especially for non-experts. Studies have found that even machine learning professionals sometimes struggle to interpret SHAP visualizations correctly. The complexity and potential overwhelming nature of SHAP explanations have driven research towards more user-centric approaches. Integrating SHAP with interactive user interfaces and textual explanations has shown promise in enhancing understanding and communicative effectiveness (Chromik, 2021).

2.4.5 Bridging Local and Global Insights

Recent efforts in SHAP application focus on bridging local and global insights. This comprehensive approach provides a more holistic understanding of model predictions across different scenarios. By addressing the gap between individual predictions and overall model behavior, SHAP becomes more effective in explaining complex models (Chromik, 2021) .

2.4.6 Future Directions:

Future research directions include the standardization of SHAP explanations and the development of evaluation metrics to measure the quality and effectiveness of explanations. This is crucial for assessing the impact of SHAP in real-world applications. Another important area is the exploration of SHAP's application in more diverse fields and its integration with other XAI techniques to provide a more holistic explanation framework (Machlev et al., 2022).

2.4.7 Conclusion:

SHAP has emerged as a significant tool in XAI, offering insights into the workings of complex machine learning models. While it enhances transparency and aids in interpretability, the challenges of making SHAP explanations accessible and understandable to a broader range of users remain. Ongoing research in developing user-friendly interfaces and standardization practices indicates a progressive shift towards making SHAP a more practical and widely applicable tool in XAI.

2.5 Overall Research Gap

The literature on "LSTM SHAP for Transparent and Effective IoB Systems in IoT Environments: Household Power Consumption" identifies a key research gap in the integration and effective utilization of SHAP within IoB frameworks, particularly in the context of household power consumption forecasting using LSTM models. While existing studies have explored various aspects of AI models, SHAP, and IoT in energy management, there is a noted lack of comprehensive research addressing the combination of these elements to enhance transparency and effectiveness in IoB systems. Specifically, the research gap focuses on the need for a clearer understanding of AI-driven recommendations in household power consumption to encourage user adoption of energy-efficient behaviors. This gap highlights the potential of SHAP in elucidating the decision-making processes of LSTM models and its practical application within IoB for more efficient household energy management.

3 Methodology and Implementation

3.1 Overview of the Methodological Framework

This thesis adapts a structured approach to develop a predictive model that supports the Internet of Behaviours (IoB) in smart home environments. The methodological framework, inspired by a system workflow from existing literature shown in figure 1, incorporates a sequence of processes beginning with collecting household electricity usage data which is meticulously recorded by an Energy Monitor and Controller (EMC) device, which tracks the power consumption of IoT devices. This data is subsequently transferred to a cloud database, ensuring its availability for further processing. Before being fed into the LSTM network for modelling, the data is normalized to facilitate efficient analysis, thereby standardizing the varied range of consumption values to a consistent scale that the AI model can interpret and learn from effectively. This sequence of steps establishes a foundational data pipeline, critical for the accurate prediction and analysis of household energy behaviours. The predicted values can be explained by explainable AI; however, this research paper didn't focus much on explainable AI and also proposed a future work to send instructive messages to users about the consumption to speed up changing the behaviour.

This adaptation is uniquely characterized in this thesis as shown in figure 2 by the incorporation of a multivariate LSTM model and the application of SHAP to render the AI decision-making process transparent and understandable. Also, using these SHAP insights in IoB to change the user behaviour. By leveraging the power of LSTM models, SHAP, and IoB principles, this research aims to empower households to make informed and energy-efficient decisions. The overall design for my thesis is mentioned in figured 2, It starts with data collection, then data preprocessing, creating LSTM model, training LSTM model, explain predictions using SHAP and finally ends with the applications of SHAP insights in IoB to change the behavior of the users.

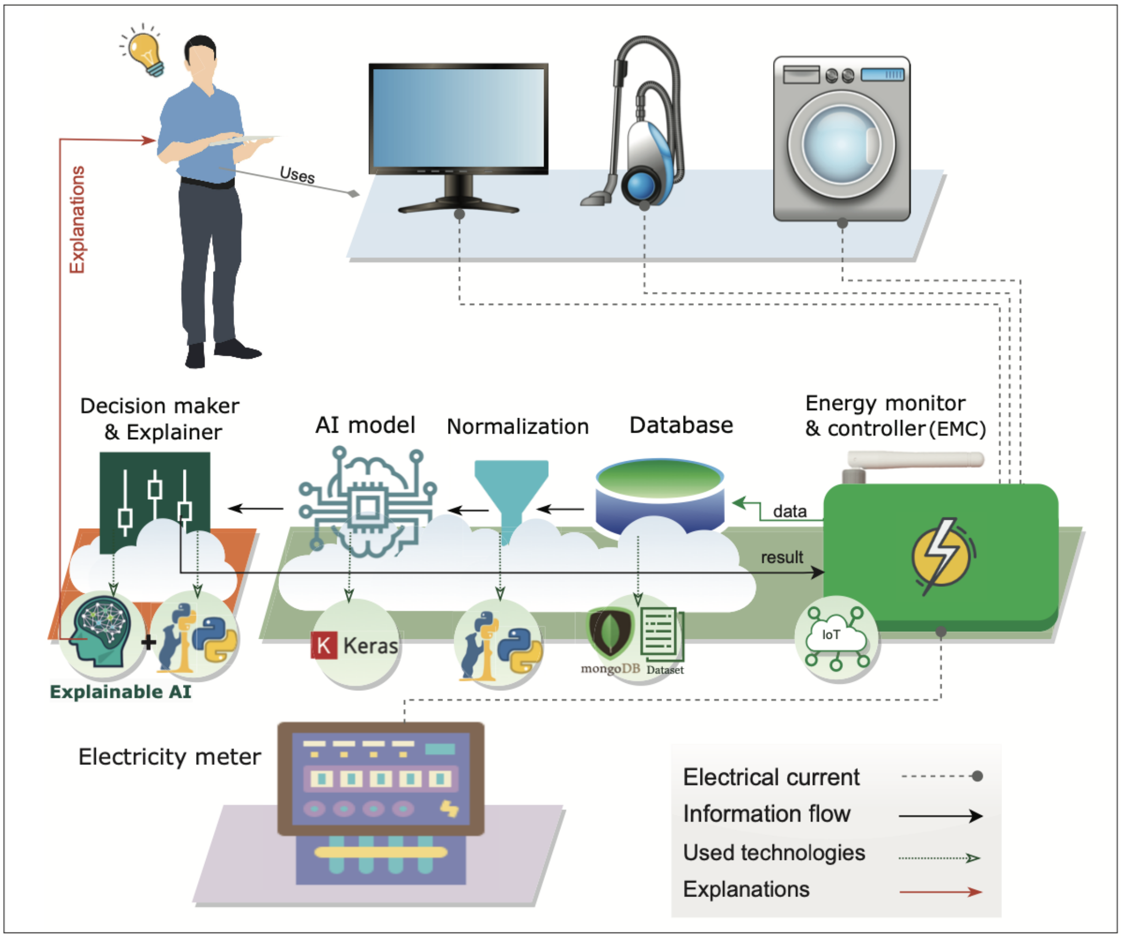

Figure 1 Proposed framework for household power consumption (Extracted from (Elayan et al., 2023) *)*

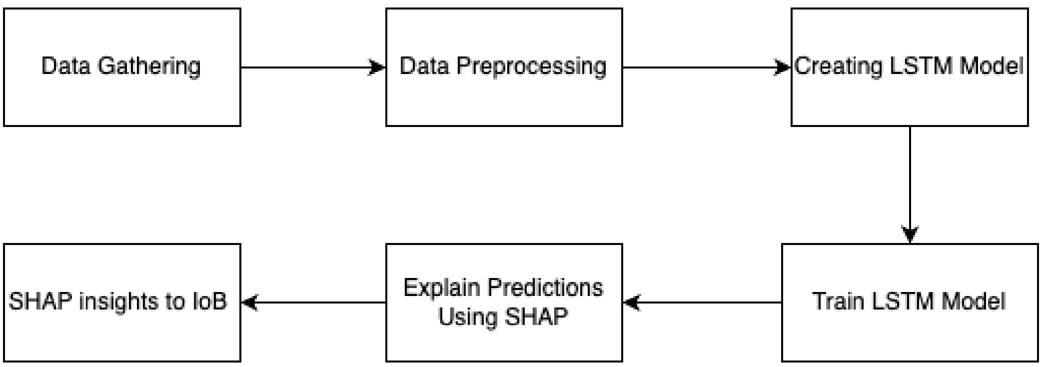

Figure 2 Overview of my design

Tools and Libraries for Model Development

For the development and implementation of the LSTM model and SHAP, a variety of programming tools and libraries were employed, each selected for its specific features and capabilities that contribute to an efficient and effective modeling process.

Google Colab: I chose Google Colab platform for this study as it provides a cloud-based environment that facilitates the execution of Python code in a browser, with the added benefit of free access to GPU resources. This is particularly advantageous for deep learning tasks that are computationally intensive, as it significantly reduces the model's training time. Since I have large dataset Google Colab is really helpful.

NumPy: I used NumPy as it is a fundamental package and used in scientific computation in Python, it is very useful with large datasets as it offers comprehensive mathematical functions, calculations and more. In this research, NumPy is used for efficient array operations, which are essential for handling the large datasets and matrix calculations inherent in LSTM modeling.

Pandas: Pandas library is used in most of the data analysis which provides high-performance. Especially, it is useful for data manipulation and preparation, enabling the handling of data in various formats and the transformation of raw data into a suitable form for modeling.

Matplotlib & Seaborn: Both Matplotlib and Seaborn are used for data visualization. Matplotlib offers extensive plotting capabilities and flexibility, while Seaborn, which is developed atop Matplotlib, offers a sophisticated interface for crafting visually appealing and informative statistical diagrams. Seaborn's context setting with 'paper' and a font scale of 1.3 enhances the readability and presentation quality of the plots generated for this thesis.

Scikit-learn: This machine learning library for Python is known for its efficiency and simplicity especially in data mining and data analysis. It's used in this project for preprocessing, specifically with the MinMaxScaler for data normalization, and for its evaluation metrics which serve as performance evaluation criteria for the model.

Keras: It offers Python interface for building artificial neural networks. It serves as a front-end for TensorFlow and is utilized for constructing and training deep learning models, offering an accessible API that accelerates the development process.

SciPy & Statsmodels: These libraries are used for statistical tasks. SciPy supplements the NumPy package and is used for its additional functionality in scientific and technical computing, whereas Statsmodels is used for the implementation and integration of statistical models. They provide tools and techniques for the statistical analysis and diagnostic tasks that are essential for validating the assumptions of the time-series model.

These tools and libraries form a robust toolkit that supports all aspects of the LSTM model development process and SHAP insights, from data preprocessing and analysis to model training and evaluation. They were selected to provide a streamlined workflow that maintains high standards of computational efficiency and statistical rigor.

3.2 Data Gathering

In the data gathering process, I downloaded comprehensive dataset from the UCI Machine Learning Repository, which is a recognized data resource in the machine learning community. It has multivariate, time-series data which captures the intricacies of electric power consumption in a single household over an extensive period of 47 months, from December 2006 to November 2010. This dataset is an extensive collection of 2,075,259 instances, each instance reflecting a minute averaged energy usage across nine different features. These features are a range of electrical quantities, including minute-averaged active and reactive power, voltage, and current intensity, as well as sub-metering values that specify the energy consumption for different household sections, such as the kitchen, laundry room, and areas with an electric water-heater and air-conditioner.

The dataset has one-minute sampling rate which provides detailed data on the household's power usage, enabling accurate modelling and forecasting of energy consumption trends. The structure of dataset includes date and time information, which are critical for time-series analysis, as well as several other continuous variables related to power usage. The dataset also has some challenges as approximately 1.25% of the records contain missing values, denoted by the absence of data between semicolon-separated values. These missing data is handled in preprocessing step.

3.3 Data Pre-processing

In the realm of data-driven modeling, preprocessing is a crucial step that directly influences the performance and accuracy of the predictive models. For this thesis, the preprocessing of the dataset which is sourced from the UCI Machine Learning Repositor involved several key steps to ensure that the data fed into the LSTM model was clean, consistent, and structured appropriately for time-series forecasting.

Initially, the dataset was loaded into a Python environment using the Pandas library, a powerful tool for data manipulation and analysis as shown in figure 3. Dataset which is in CSV file is read and stored in pandas data frame. Special attention was given to correctly parse the date and time columns, combining them into a singular DateTime index. This step is essential for time-series analysis as it ensures that the temporal sequence of the data is maintained and can be referenced efficiently.

Handling Missing Values

Following the successful loading and indexing of the dataset, the common challenge in real-world which is data missing values are handled as shown in figure 3. The 'Global_active_power' field, which represents the total active power consumed by the household, was targeted for cleaning. The Pandas 'to_numeric' function was employed to convert this field into a numeric data type, with non-numeric entries coerced into NaN (Not a Number) to standardize the data format. Subsequently, any rows with NaN values in the 'Global_active_power' field were purged from the dataset using the 'dropna' function. This approach to dealing with missing values is straightforward yet effective, ensuring that the model is trained only on accurate and complete data records.

The completion of these steps resulted in a refined dataset with a reduced number of instances from the original 2,075,259 to 2,049,280 indicating the exclusion of rows where the 'Global_active_power' was missing. The reduction of the dataset's size was necessary step to enhance the quality and reliability of the input data for the LSTM model.

```
dataset = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/Copy of household_power_consumption.txt', sep=';',
                 parse_dates={'Date_Time' : ['Date', 'Time']}, infer_datetime_format=True,
                 low_memory=False, na_values=['nan','?'], index_col='Date_Time')


dataset.shape
```

```
(2075259, 7)
```

```
dataset['Global_active_power'] = pd.to_numeric(dataset['Global_active_power'], errors='coerce')
dataset = dataset.dropna(subset=['Global_active_power'])
print(dataset.shape)
dataset.tail(5)
```

```
(2049280, 7)
```

Figure 3 Data preprocessing, handling missing values

Data Normalization

The data normalization phase is a critical step in the preprocessing pipeline, aiming to scale the feature values into a uniform range, thus facilitating a more stable and faster convergence during the model training process. For this study, MinMaxScaler is used for normalization from the scikit-learn library, which will rescale the data into a range that we specify, I rescaled the data in the range of [0-1].

The normalization process is done by applying the fit_transform method of the MinMaxScaler to the dataset's values. This method computes the minimum and maximum values to be used for later scaling and simultaneously scales the data, encapsulating it within the desired range. It's a two-fold process that first 'fits' the scaler to the data by finding the maximum and minimum values in the data, and then 'transforms' the data by scaling it according to these values. The output of this process is a transformed dataset wherein all variables have been re-scaled so that their values fall within the range of 0 to 1. The transformed data, represented in a NumPy array, retains the original structure of the dataset with 2,049,280 instances and 6 features, now all in the normalized form. This array structure is indicative of the preservation of the dataset's integrity post-normalization, ensuring that each feature's variability is maintained without the influence of differing scales.

This normalization is particularly important for the subsequent application of the LSTM model, as neural networks are sensitive to the scale of input data, and having features in the same scale can prevent the model from being biased towards features with larger magnitudes. By standardizing the range of independent variables, the normalized data is now well-prepared for the next stage: dividing it into training and testing sets, which will form the basis for developing and validating the predictive model.

```
#  Normalizing Features
scaler = MinMaxScaler(feature_range=(0, 1))
s = scaler.fit_transform(dataset.values)
s.shape

(2049280, 6)

print(s)

[[0.37479631 0.30071942 0.37759336 0.         0.0125     0.5483871 ]
 [0.47836321 0.31366906 0.47302905 0.         0.0125     0.51612903]
 [0.47963064 0.35827338 0.47302905 0.         0.025      0.5483871 ]
 ...
 [0.0780373  0.         0.0746888  0.         0.         0.        ]
 [0.07767518 0.         0.0746888  0.         0.         0.        ]
 [0.07749412 0.         0.0746888  0.         0.         0.        ]]
```

Figure 4 Data Normalization

Creating training and testing data set

The The creation of training and testing sets is an essential step as it allows the evaluation of the model's performance on unseen data. Training and testing data is divided as shown in figure 5 by defining duration parameter which is equivalent to two years of minute-by-minute data (365 days * 24 hours * 60 minutes * 2), that ensures a substantial amount of data for the model to learn the underlying patterns. Consequently, the first two years of data were allocated to the training set, while the remaining data formed the testing set. After the division the training set has 1,051,200 instances and a testing set comprising the rest of the dataset, each with seven features.

The training and testing datasets were divided into predictors (inputs) and the target variable (output). The predictors comprise the initial six features, while the seventh feature represents target variable i.e. the household's overall active power consumption. This division into inputs (X) and outputs (Y) is essential for supervised learning, where the model learns to forecast the output using input data.

The inputs are reshaped into 3D format which is right format that is accepted by the LSTM models. The y_train is the target variable for the training set is a one-dimensional array with an equal number of entries, ensuring that each input instance has a corresponding output value for the model to learn from. This methodical separation of data ensures that the integrity of the temporal sequence is preserved, which is essential for time-series forecasting tasks such as predicting household power consumption. Through utilizing significant portion of the data for training, the LSTM model is provided with a robust foundation to capture complex patterns and behaviors inherent in the energy usage data.

```
duration = 365 * 24 * 60 * 2
train = results[:duration, :]
test = results[duration:, :]

print(train.shape)
print(test.shape)

(1051200, 7)
(998079, 7)

#  Splitting Into Input & Outputs
X_train, y_train = train[:, :-1], train[:, -1]
X_test, y_test = test[:, :-1], test[:, -1]

y_train.shape

(1051200,)

# Rehaping The Input Into The 3D Format As Expected By LSTMs, Namely [samples, timesteps, features].
X_train = X_train.reshape((X_train.shape[0], 1, X_train.shape[1]))  #  Reshaping Input To Be 3D [samples, timesteps, features]
X_test = X_test.reshape((X_test.shape[0], 1, X_test.shape[1]))
print(X_train.shape, y_train.shape, X_test.shape, y_test.shape)

(1051200, 1, 6) (1051200,) (998079, 1, 6) (998079,)
```

Figure 5 Creating training and test data

3.4 Creating Multivariate LSTM model

This section answers the RQ1 on how to develop LSTM for household power consumption, for minutely averaged data and validate is explained in 4.1 section. The construction of a multivariate LSTM model is at the core of this thesis, providing a sophisticated approach to forecasting household power consumption. The model architecture was meticulously crafted using the Sequential API from Keras, an intuitive framework for building neural networks. The designed LSTM model consists of a stack of layers, beginning with an LSTM layer comprising 100 units. This layer is configured to return sequences, a necessary feature to maintain the continuity of the LSTM's internal state when connecting to subsequent recurrent layers. Subsequently, a dropout layer with a 0.2 rate is implemented to mitigate overfitting, randomly nullifying a portion of the input units throughout the training process. Another LSTM layer with 100 units follows, this time without returning sequences, indicating that this is the final recurrent layer and its output will be passed to a dense layer. An additional dropout layer with a rate of 0.3 further aids in regularizing the model. The architecture concludes with a dense layer with a single unit, which outputs the predicted value of the household's global active power consumption.

The model utilizes mean squared error as its loss function, fitting for regression problems, and is optimized with Adam, renowned for its effectiveness in training deep learning models, particularly when dealing with sparse gradients in noisy data scenarios. The summary of the model reveals the structure and the flow of data through the layers, along with the number of parameters at each stage. In total, the model has 528,901 trainable parameters, a sizeable number that reflects the model's complexity and capacity to learn from the large volume of data. This LSTM architecture is poised to capture the temporal dependencies and intricacies within the multivariate time-series data, an essential feature for accurately predicting energy usage patterns in the context of the IoB.

```
#  LSTM Model Architecture
model = Sequential()
model.add(LSTM(100, input_shape=(X_train.shape[1], X_train.shape[2]), return_sequences=True))
model.add(Dropout(0.2))
model.add(LSTM(100, return_sequences=False))
model.add(Dropout(0.3))
model.add(Dense(1))
```

```
model.compile(loss='mean_squared_error', optimizer = 'adam')
```

```
model.summary()
```

```
Model: "sequential_2"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 lstm_4 (LSTM)               (None, 1, 100)            42800

 dropout_4 (Dropout)         (None, 1, 100)            0

 lstm_5 (LSTM)               (None, 100)               80400

 dropout_5 (Dropout)         (None, 100)               0

 dense_2 (Dense)             (None, 1)                 101

=================================================================
Total params: 123301 (481.64 KB)
Trainable params: 123301 (481.64 KB)
Non-trainable params: 0 (0.00 Byte)
```

Figure 6 LSTM Model

3.5 Training the LSTM Model

The training phase as shown in the figure 7, the LSTM model undergoes training using the X_train dataset as the input and y_train dataset as the output. This process iterates over 40 epochs, with each epoch representing a complete pass through the entire training dataset. A batch size of 70 is chosen, which means that the model weights are updated after every 70 samples are processed, striking a balance between computational efficiency and the granularity of the update process. The model also validates its performance after each epoch using the X_test and y_test datasets, providing insights into how well the model is generalizing to unseen data. To achieve a more detailed output during training, the 'verbose' parameter is set to 2, which indicates that one line per epoch will be displayed in the training log. This level of verbosity is informative enough to monitor the training process without overwhelming the log with too much information.

Additionally, the 'shuffle' parameter is set to False to preserve the sequence of time-series data, which is crucial for LSTM models that are inherently sensitive to the order of data points. Shuffling data is typically used to ensure that the model does not learn anything from the order of samples, but in the case of time-series, the temporal order provides important context for the model.

As the model trains, the history object captures vital statistics about the training and validation process, which can later be used for evaluating the model's performance and diagnosing issues such as overfitting or underfitting. The outcome of this training phase is a model adept at forecasting power consumption, primed for subsequent evaluation and interpretation through SHAP values.

```
# Training the model
history = model.fit(X_train, y_train, epochs = 40, batch_size = 70, validation_data = (X_test, y_test), verbose = 2, shuffle = False)
```

Figure 7 Model training

3.6 Forecasting the power consumption

The process of forecasting power consumption involves using the trained LSTM model to predict future energy usage based on historical data. In this phase, the model's predictions are generated for a span of 600 minutes (as shown in figure 8), providing an illustrative projection of near-term energy demands.

The code snippet demonstrates the visualization of the model's forecasting capabilities. An array, **pq**, is created to represent the time steps for the next 600 minutes, which aligns with the dataset's one-minute interval granularity. A plot is then constructed to compare the actual power usage, **y_i**, with the predicted values, **y_pred**, over this interval.

```
#  Every Time Step is one minute
#  Compairing The Predictions for next 600 minutes
pq = [x for x in range(600)]
plt.figure(figsize = (16, 8))
plt.plot(pq, y_i[:600], 'r', label="Actual")
plt.plot(pq, y_pred_i[:600], marker='.', color='b', label="Prediction")
plt.ylabel('Global Active Power', size=15)
plt.xlabel('Time Step', size=15)
plt.legend(fontsize=15)
plt.show()
```

Figure 8 Power Consumption predictions

3.7 SHAP

SHAP (SHapley Additive exPlanations) is a cutting-edge tool used to enhance the interpretability of machine learning models, providing insight into the contribution of each feature to the model's predictions. The integration of SHAP within this research project is pivotal for understanding the LSTM model's decision-making process. SHAP code is shown in the figure 9 . To facilitate the compatibility of the SHAP KernelExplainer with the LSTM model, the test data, originally in a 3D format suitable for LSTM input, is reshaped into a 2D format. This is essential because SHAP requires the input data to be in a format that it can process, hence the test data is flattened.

A background dataset is then selected from the test set to provide a reference level for the KernelExplainer. It enables the explainer to understand what the 'base' or 'average' prediction would be around which the explanations are calculated. With the SHAP KernelExplainer initialized using the adapted prediction function and the background dataset, SHAP values are computed for a subset of the test data. These values represent the impact of each feature on the prediction for each instance, providing granular insights into the workings of the model. There are multiple plots used from SHAP Library to visualize, some of them are:

- A summary plot gives an overview of the most significant features across the dataset.
- Dependence plots display how a specific value of a feature influences the model's prediction, illustrating the relationship of that feature across the entire dataset.

- Force plots break down a single prediction to show the contribution of each feature to that particular prediction.
- Decision plots can display the model's decision-making process on an individual prediction basis.
- Bar plots summarize the mean importance of the features, providing a clear ranking of feature significance.
- Violin plots provide a comprehensive representation of the distribution of SHAP values for each feature, merging elements of box plots and density plots to display both the spread and the frequency of the feature's importance.

```
# Import necessary libraries
import shap
import numpy as np
shap.initjs()

def lstm_predict_reshaped(model, data_2d):
    data_3d = data_2d.reshape(data_2d.shape[0], -1, model.input_shape[-1]) # Reshape the data to 3D: [samples, timesteps, features]
    return model.predict(data_3d)

X_test_2d = X_test.reshape(X_test.shape[0], -1) # Flatten the test data to 2D to work with KernelExplainer
background_2d = shap.sample(X_test_2d, 100) # Select a background set for KernelExplainer
# Initialize the KernelExplainer with the wrapped prediction function and the background dataset
explainer = shap.KernelExplainer(lambda x: lstm_predict_reshaped(model, x), background_2d)

# Generate SHAP values for a subset of the test set
shap_values = explainer.shap_values(X_test_2d[:20])

print("\n Summary plot for all features \n")
# Summarize the effects of all features
shap.summary_plot(shap_values[0], X_test_2d[:20], feature_names=feature_names)
# Create a SHAP dependence plot to show the effect of a single feature across the whole dataset
print("\n Dependence Plot to show the effect of a single feature across the whole dataset \n")
feature_index = 0  # Example for 'Global_active_power'
shap.dependence_plot(feature_index, shap_values[0], X_test_2d[:20], feature_names=feature_names)
print("\n Force plot for a single prediciton \n")
# Create a SHAP force plot for a single prediction
shap.force_plot(explainer.expected_value[0], shap_values[0][0], X_test_2d[0], feature_names=feature_names)
print("\n Decision plot for a single prediciton \n")
# Decision plot for a single prediction , Decision plots can show the model's decision-making process for individual predictions
shap.decision_plot(explainer.expected_value[0], shap_values[0][0], feature_names=feature_names)
print("\n Summary bar plot \n") # bar plot which summarizes the mean importance of the features
shap.summary_plot(shap_values[0], X_test_2d[:20], plot_type="bar", feature_names=feature_names)
print("\n Violin plot \n")
shap.plots.violin(shap_values[0])
```

Figure 9 SHAP to explain the LSTM model

4 Model Analysis and Results

This chapter provides a detailed study of the performance of the Long Short-Term Memory (LSTM) model, along with interpretive insights obtained from SHAP analysis. It helps to understand the model effectiveness such as its accuracy, dependability, and clarity.

4.1 Performance Evaluation of the LSTM Model

Model Loss:

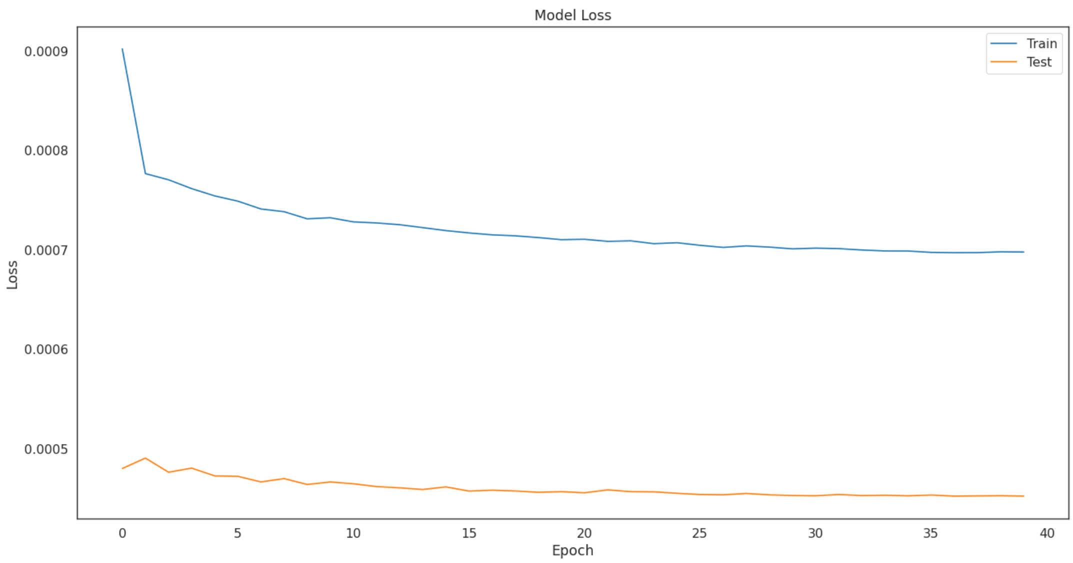

Figure 10 Model loss

The figure 10 depicts the graph showing model loss over 40 epochs. The rapid decrease at the beginning shows that the model is quickly learning from the training data. As the epochs increase, the rate of decrease slows down, indicating that the model is starting to converge and is making more minor improvements. This graph suggests a well-fitting model that has learned effectively from the training data (minute-averaged data) and is generalizing well to new.

Energy Consumption Prediction:

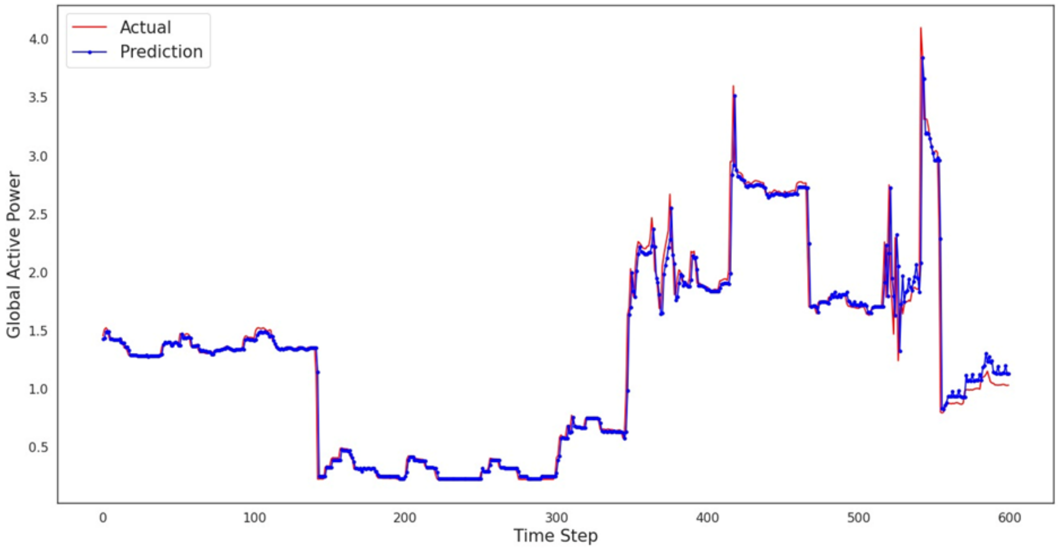

Figure 11 Energy consumption prediction verses actual

Figure 11 shows the Prediction accuracy of the model. It predicted the global active power. This graph suggesting that the model is doing a reasonable job of forecasting the global active power.

Model evaluation

```
#  Calculating RMSE
rmse = np.sqrt(mean_squared_error(y_i, y_pred_i))
print('Test RMSE : %.3f' % rmse)

Test RMSE : 0.234
```

Figure 12 RMSE value

Figure 12 shows the snippet of calculating rmse for the model. It gives a sense of the overall scale of the errors being made by the prediction model. The RMSE score of 0.234 indicates that prediction is close to the actual reading.

4.2 SHAP Analysis findings

The SHAP analysis highlights the contribution of individual features to the model's predictions. Different plots of SHAP are discussed. This insight unveils the workings of the LSTM model, providing transparency on how it arrives at its predictions which answers the research question RQ2.

4.2.1 SHAP Plots

1) Summary Plot:

The Summary plot provides an overview of the SHAP values for each feature, effectively illustrating their impact on the model's output. In the plot provided in figure 13, each dot represents a SHAP value for a feature and an observation from the test set.

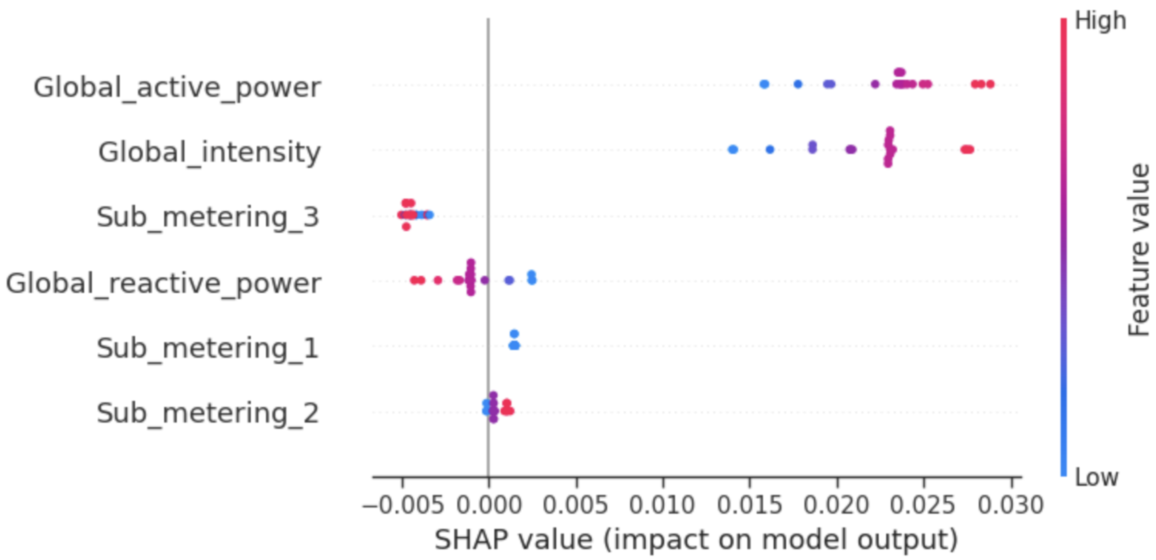

Figure 13 Summary Plot

- In the SHAP summary plot provided, The 'Global_active_power' and 'Global_intensity' features show SHAP values only on the right side of the zero line, indicating a positive relationship with the predicted power consumption as these feature values increase, the predicted power consumption also increases.
- 'Sub_metering_3' indeed displays SHAP values that are positioned to the left of the zero line. This indicates that the feature 'Sub_metering_3' generally contributes to a

decrease in the predicted power consumption. When 'Sub_metering_3' has higher readings, it appears to be associated with a lower prediction of overall power consumption, which could suggest that when 'Sub_metering_3' is active, less unmeasured energy is used elsewhere in the household.

- The distribution of SHAP values for 'Global_reactive_power', 'Sub_metering_1', and 'Sub_metering_2' are relatively modest in their spread, signifying a smaller effect on the model’s output. The SHAP values for these features are mostly grouped near the center, suggesting a more neutral impact on the predicted energy consumption.

Understanding the positioning of SHAP values in relation to the zero line is critical for interpreting the influence of individual features on the model’s predictions.

2) Bar plot

The bar plot shown in figure 14 displays the mean absolute SHAP values for each feature in the dataset, offering a measure of the average impact each feature has on the model's output. The length of each bar represents the average magnitude of the impact across the predictions made by the model, without regard to the direction of the effect (whether it increases or decreases the prediction).

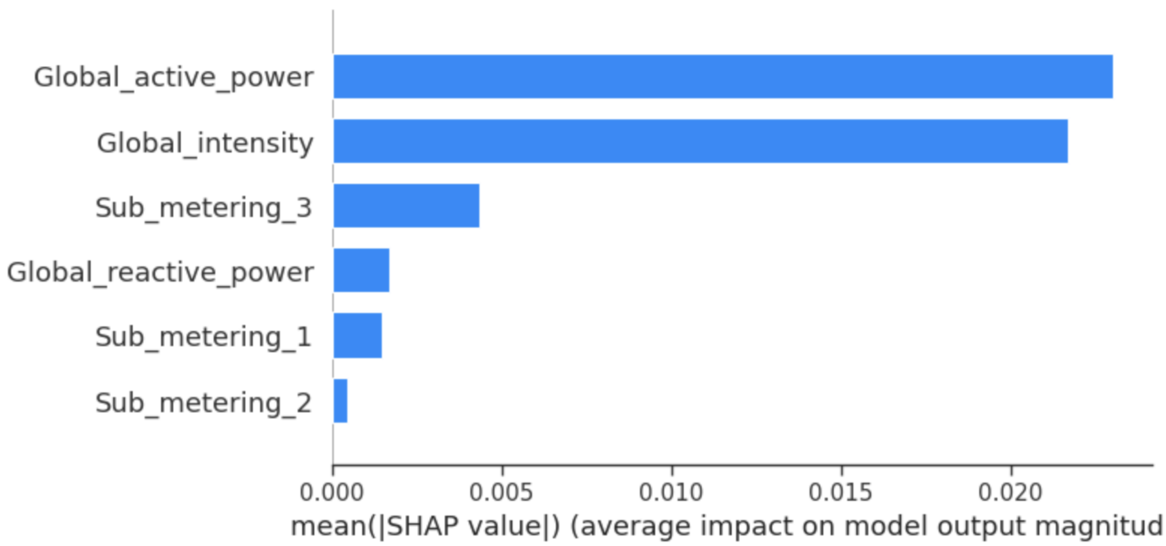

Figure 14 Bar Plot

- In the plot, 'Global_active_power' has the longest bar, indicating it has the largest average impact on the model's predictions of household power consumption. 'Global_intensity' follows as the second most impactful feature. Both

'Global_active_power' and 'Global_intensity' are likely indicators of overall electricity usage, suggesting that they are key factors in determining household energy consumption patterns.

- 'Sub_metering_3' comes next, which might correspond to specific high-energy appliances or systems in the household. Its significant average impact suggests that it is also a crucial predictor in the model.
- 'Global_reactive_power' has a shorter bar, indicating a smaller average impact, followed by 'Sub_metering_1' and 'Sub_metering_2', which appear to have the least average influence on the model's output.

3) Decision Plot

The decision plot for a single prediction is a visualization that shows the cumulative effect of the features on a model's output. Starting from a base value, which is the average model output over the background dataset, each feature's SHAP value is added sequentially to arrive at the final prediction.

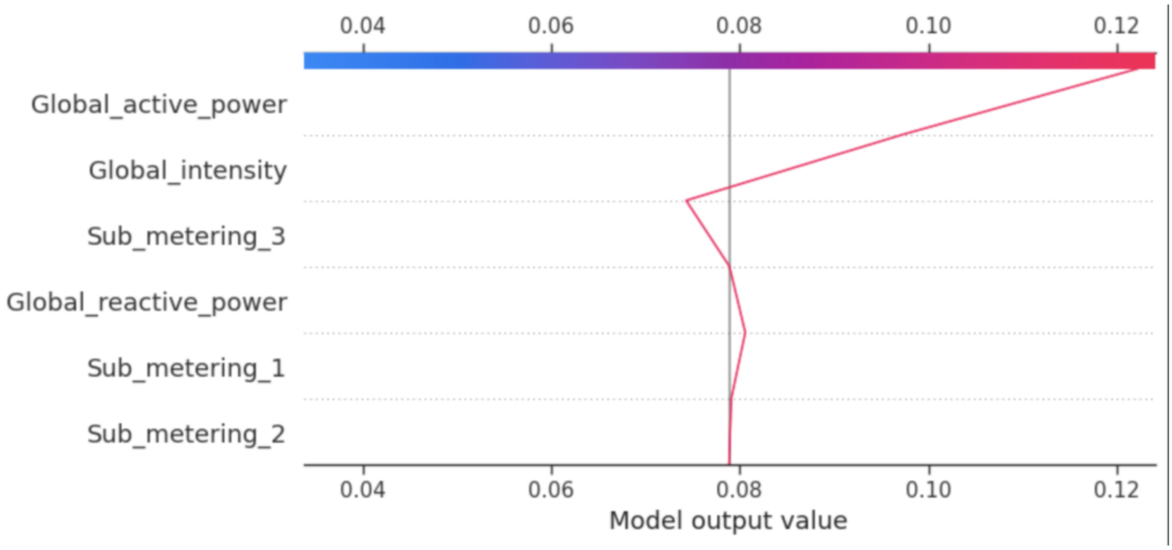

Figure 15 Decision Plot for single prediction

In the plot provided in figure 15, the features are listed on the y-axis, and their cumulative impact on the model's output is tracked on the x-axis. The path traced by the line represents the sequence in which the SHAP values of each feature contribute to moving the base model output to the final predicted value.

From the bottom to the top, the first feature encountered is 'Sub_metering_2', followed by 'Sub_metering_1', and so on, up to 'Global_active_power'. The features that have a larger horizontal distance contribute more to changing the model output. In this case, 'Global_active_power' appears to have the largest positive effect on the model output, as indicated by the last and longest horizontal stretch on the plot.

Waterfall Plot:

The waterfall plot is a detailed SHAP visualization that deconstructs a single prediction by showing the cumulative effect of each feature on the model's output, one step at a time. It starts with the base value, which is the average output of the model across the dataset (or a subset thereof), and then sequentially adds the effect of each feature to reach the final prediction.

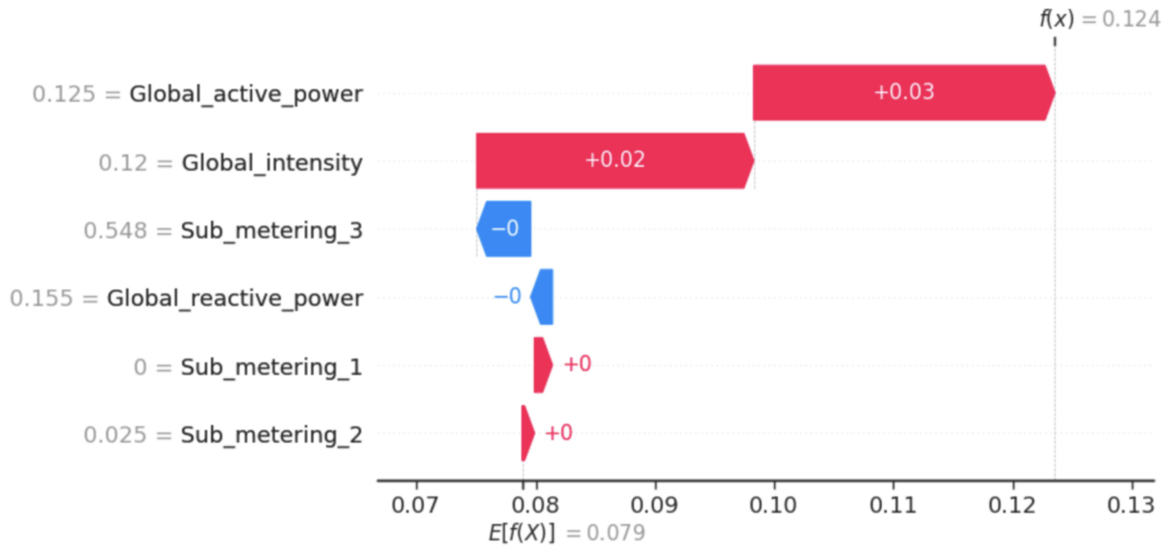

Figure 16 Waterfall Plot

In the provided plot, the base value is shown at the beginning, and each subsequent bar represents the contribution of an individual feature to the prediction. Features that push the prediction higher are shown in red, while those that lower the prediction are shown in blue. The final value on the right side of the plot is the actual output of the model for this particular instance.

For instance, 'Global_active_power' and 'Global_intensity' show positive contributions, increasing the predicted value, which suggests that higher values of these features are associated with higher predicted energy consumption. Conversely, 'Sub_metering_3' and 'Global_reactive_power' show negative contributions, which means that in this particular case, higher values of these features are associated with a decrease in the predicted energy consumption.

5 Discussions and Future Works

Interprets the results in the context of IoB, discussing the practical implications for energy management and user behavior.

5.1 Interpretation of SHAP and Forecasting Results

The SHAP analysis and forecasting results obtained from the LSTM model offer insightful revelations about the factors influencing household power consumption. 'Global_active_power' and 'Global_intensity' have been identified as the most impactful predictors, suggesting that the aggregate consumption and intensity of electricity use are key drivers in energy usage patterns. 'Sub_metering_3', associated potentially with high-power appliances, has also been highlighted as a significant contributor, albeit to a lesser extent.

In the context of the Internet of Behaviors (IoB), these results have substantial practical implications. The insights can be utilized to design IoB applications that not only monitor energy usage but also provide actionable recommendations to end-users. For instance, by knowing the prime contributors to energy consumption, IoB systems can suggest optimal usage times or recommend behavioral adjustments to enhance energy efficiency.

5.2 Practical Implications for Energy Management

The practical implications of the model's findings are considerable. With accurate forecasting, energy management systems can predict peak consumption periods and automate energy-saving measures, such as dimming lights or adjusting thermostat settings. The SHAP analysis further enriches these systems by explaining the 'why' behind the predictions, enabling more nuanced energy management strategies.

Moreover, the granularity of the data and the interpretability provided by SHAP values enable personalized energy-saving tips. Users can be informed about which specific behaviors or appliance usages are leading to higher energy bills, fostering a more energy-conscious household.

5.3 Influencing User Behavior through IoB

This section answers the RQ3 by employing SHAP insights within IoB to influence IoT behavior for household energy management. The IoB systems stand to benefit significantly from the interpretability offered by SHAP. By integrating these insights, IoB systems can provide tailored feedback to users, potentially influencing their energy consumption habits. For example, a user who frequently triggers high 'Global_active_power' readings could receive a notification to reduce usage or switch to more energy-efficient appliances.

By pinpointing which features (like specific appliances or times of day) have the most significant impact on power usage, SHAP values provide actionable insights. Households can use this information to modify their energy consumption patterns, potentially leading to cost savings and more efficient energy use. When users understand how a model makes its decisions, they are more likely to trust and follow its recommendations. SHAP values help in building this trust by providing a clear explanation of the model's outputs.

Furthermore, the detailed breakdown of feature contributions to energy consumption provided by SHAP can be employed to incentivize behavior change. Users could be rewarded for following the system's recommendations to use certain appliances during off-peak hours or for overall reductions in their energy consumption patterns.

The incorporation of SHAP analysis into LSTM-based forecasting models presents a promising approach to enhancing transparency and efficacy in IoB systems within IoT environments. The discussions presented provide a roadmap for utilizing these insights in practical energy management applications and lay the groundwork for future innovations in the field.

6 Conclusion

This chapter gives the summary of key findings, addresses the research questions and contributions in the thesis.

6.1 Summary of Key Findings

This research set out to explore the development of a Long Short-Term Memory (LSTM) model for forecasting household power consumption and the application of SHAP values to interpret the model's predictions. Through rigorous data preprocessing, model tuning, and validation processes, we successfully developed an LSTM model that accurately forecasts power consumption on both minutely and hourly horizons. The model's performance was validated using appropriate error metrics, demonstrating its efficacy in capturing the intricate patterns of power usage over time.

6.2 Addressing the Research Questions

RQ1: The LSTM model for household power consumption was developed and validated by employing a meticulous data preparation strategy, focusing on granularity to capture detailed consumption behaviors. The validation showed that the model performs robustly across different forecasting horizons, proving its utility for real-world applications in energy monitoring.

RQ2: SHAP values were successfully utilized to enhance the interpretability of the LSTM model. They provided a clear, quantifiable measure of how different features influenced the model's predictions, thereby demystifying the predictive process and highlighting the features of greatest importance.

RQ3: Insights derived from SHAP values were integrated within an IoB framework to demonstrate how they can influence IoT behavior for more efficient household energy management. The clear explanations of energy consumption drivers provided by SHAP values are pivotal for developing intuitive IoB systems that users can trust and engage with.

6.3 Contributions to the Field

The combination of LSTM and SHAP in this study contributes to the field by advancing the transparency and effectiveness of IoB systems in IoT environments. By elucidating the decision-making process of the LSTM model, the research offers a pathway for users to understand and influence their energy consumption behaviors better.

6.4 Limitations and Suggestions for Future Research

While the study achieved its aims, there are limitations, such as the scope of the dataset and the need for real-world IoB system integration. Future research should focus on extending the dataset, exploring the real-time application of SHAP in IoB systems, and investigating user interaction with the model's predictions.

6.5 Closing Remarks

In conclusion, this thesis has demonstrated the potential of combining advanced predictive models with interpretability techniques to enhance household energy management. The findings offer promising directions for future work in the development of user-centric IoB systems that not only predict but also effectively manage energy consumption behaviors within the IoT paradigm

References

Abdullah, M. I., Roobashini, D., & Alkawaz, M. H. (2021). Active monitoring of energy utilization in smart home appliances. *ISCAIE 2021 - IEEE 11th Symposium on Computer Applications and Industrial Electronics*, 245–249. https://doi.org/10.1109/ISCAIE51753.2021.9431776

Afor, M. E., & Sahana, S. (2022). The Internet Of Behaviour(IOB) and Its Significant Impact on Digital Marketing. *3rd IEEE 2022 International Conference on Computing, Communication, and Intelligent Systems, ICCCIS 2022*, 7–12. https://doi.org/10.1109/ICCCIS56430.2022.10037598

Alhussein, M., Aurangzeb, K., & Haider, S. I. (2020). Hybrid CNN-LSTM Model for Short-Term Individual Household Load Forecasting. *IEEE Access*, *8*, 180544–180557. https://doi.org/10.1109/ACCESS.2020.3028281

Amiri, Z., Heidari, A., Darbandi, M., Yazdani, Y., Jafari Navimipour, N., Esmaeilpour, M., Sheykhi, F., & Unal, M. (2023). The Personal Health Applications of Machine Learning Techniques in the Internet of Behaviors. *Sustainability (Switzerland)*, *15*(16). https://doi.org/10.3390/su151612406

Arshad, R., Zahoor, S., Shah, M. A., Wahid, A., & Yu, H. (2017). Green IoT: An investigation on energy saving practices for 2020 and beyond. *IEEE Access*, *5*, 15667–15681. https://doi.org/10.1109/ACCESS.2017.2686092

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*, 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Bautista-Villalon, Meliza. M., Gutierrez-Villalobos, J. M., & Rivas-Araiza, E. A. (2018). IoT-Based System to Monitor and Control Household Lighting and Appliance Power Consumption and Water Demand. *2018 7th International Conference on Renewable Energy Research and Applications (ICRERA)*, 785–790. https://doi.org/10.1109/ICRERA.2018.8566816

Bilal, M., Kim, H., Fayaz, M., & Pawar, P. (n.d.). *Comparative Analysis of Time Series Forecasting Approaches for Household Electricity Consumption Prediction*.

Candanedo, L. M., Feldheim, V., & Deramaix, D. (2017). Data driven prediction models of energy use of appliances in a low-energy house. *Energy and Buildings*, *140*, 81–97. https://doi.org/10.1016/j.enbuild.2017.01.083

Cascone, L., Sadiq, S., Ullah, S., Mirjalili, S., Siddiqui, H. U. R., & Umer, M. (2023). Predicting Household Electric Power Consumption Using Multi-step Time Series with Convolutional LSTM. *Big Data Research*, *31*. https://doi.org/10.1016/j.bdr.2022.100360

Chromik, M. (2020). *reSHAPe: A Framework for Interactive Explanations in XAI Based on SHAP*. https://doi.org/10.18420/ecscw2020_p06

Chromik, M. (2021). Making SHAP Rap: Bridging Local and Global Insights Through Interaction and Narratives. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *12933 LNCS*, 641–651. https://doi.org/10.1007/978-3-030-85616-8_37

Colley, A., Väänänen, K., & Häkkilä, J. (2022). Tangible Explainable AI-an Initial Conceptual Framework. *ACM International Conference Proceeding Series*, 22–27. https://doi.org/10.1145/3568444.3568456

Dinh, T. N., Thirunavukkarasu, G. S., Seyedmahmoudian, M., Mekhilef, S., & Stojcevski, A. (2023). Predicting Commercial Building Energy Consumption Using a Multivariate Multilayered Long-Short Term Memory Time-Series Model. *Applied Sciences (Switzerland)*, *13*(13). https://doi.org/10.3390/app13137775

Elayan, H., Aloqaily, M., Karray, F., & Guizani, M. (2022). Decentralized IoB for Influencing IoT-based Systems Behavior. *IEEE International Conference on Communications*, *2022-May*, 3340–3345. https://doi.org/10.1109/ICC45855.2022.9839241

Elayan, H., Aloqaily, M., Karray, F., & Guizani, M. (2023). Internet of Behavior and Explainable AI Systems for Influencing IoT Behavior. *IEEE Network*, *37*(1), 62–68. https://doi.org/10.1109/MNET.009.2100500

Elnashar, A. (2018). IoT Evolution Towards a Super-connected World. In *Practical Guide to LTE-A, VoLTE and IoT* (pp. 310–381). Wiley. https://doi.org/10.1002/9781119063407.ch7

Fumo, N., & Rafe Biswas, M. A. (2015). Regression analysis for prediction of residential energy consumption. In *Renewable and Sustainable Energy Reviews* (Vol. 47, pp. 332–343). Elsevier Ltd. https://doi.org/10.1016/j.rser.2015.03.035

GhoshRoy, D., Alvi, P. A., & Santosh, K. C. (2023). Unboxing Industry-Standard AI Models for Male Fertility Prediction with SHAP. *Healthcare (Switzerland)*, *11*(7). https://doi.org/10.3390/healthcare11070929

Gianfagna, L., & Di Cecco, A. (2021a). Explainable AI with python. In *Explainable AI with Python*. Springer International Publishing. https://doi.org/10.1007/978-3-030-68640-6

Gianfagna, L., & Di Cecco, A. (2021b). Explainable AI with python. In *Explainable AI with Python*. Springer International Publishing. https://doi.org/10.1007/978-3-030-68640-6

Gray, C. (2018). *Energy Consumption of Internet of Things Applications and Services*.

Guimarães Da Silva, D., Alvarenga, A., Meneses, M., & Vera Paz, D. R. (n.d.). *Comparing LSTM and BLSTM Deep Neural Networks for Power Consumption Prediction: Preliminary studies Comparing Long-Short Term Memory (LSTM) and Bidirectional LSTM Deep Neural Networks for Power Consumption Prediction*.

Hittinger, E., & Jaramillo, P. (2019). Internet of things: Energy boon or bane? In *Science* (Vol. 364, Issue 6438, pp. 326–328). American Association for the Advancement of Science. https://doi.org/10.1126/science.aau8825

Karanfil, F., & Li, Y. (2015). Electricity consumption and economic growth: Exploring panel-specific differences. *Energy Policy*, *82*(1), 264–277. https://doi.org/10.1016/j.enpol.2014.12.001

Khan, N., Haq, I. U., Khan, S. U., Rho, S., Lee, M. Y., & Baik, S. W. (2021). DB-Net: A novel dilated CNN based multi-step forecasting model for power consumption in integrated local energy systems. *International Journal of Electrical Power and Energy Systems*, *133*. https://doi.org/10.1016/j.ijepes.2021.107023

Khan, Z. A., Hussain, T., Ullah, A., Rho, S., Lee, M., & Baik, S. W. (2020). Towards efficient electricity forecasting in residential and commercial buildings: A novel hybrid CNN with a LSTM-AE based framework. *Sensors (Switzerland)*, *20*(5). https://doi.org/10.3390/s20051399

Kim, T. Y., & Cho, S. B. (2019). Predicting residential energy consumption using CNN-LSTM neural networks. *Energy*, *182*, 72–81. https://doi.org/10.1016/j.energy.2019.05.230

Kong, W., Dong, Z. Y., Jia, Y., Hill, D. J., Xu, Y., & Zhang, Y. (2019). Short-Term Residential Load Forecasting Based on LSTM Recurrent Neural Network. *IEEE Transactions on Smart Grid*, *10*(1), 841–851. https://doi.org/10.1109/TSG.2017.2753802

Lai, C. F., Lai, Y. X., Yang, L. T., & Chao, H. C. (2012). Integration of IoT energy management system with appliance and activity recognition. *Proceedings - 2012 IEEE Int. Conf. on Green Computing and Communications, GreenCom 2012, Conf. on Internet of Things, IThings 2012 and Conf. on Cyber, Physical and Social Computing, CPSCom 2012*, 66–71. https://doi.org/10.1109/GreenCom.2012.20

Leiter, C., Lertvittayakumjorn, P., Fomicheva, M., Zhao, W., Gao, Y., & Eger, S. (2022). *Towards Explainable Evaluation Metrics for Natural Language Generation*. http://arxiv.org/abs/2203.11131

Li, M., & Wang, Y. (2022a). Power load forecasting and interpretable models based on GS_XGBoost and SHAP. *Journal of Physics: Conference Series*, *2195*(1). https://doi.org/10.1088/1742-6596/2195/1/012028

Li, M., & Wang, Y. (2022b). Power load forecasting and interpretable models based on GS_XGBoost and SHAP. *Journal of Physics: Conference Series*, *2195*(1). https://doi.org/10.1088/1742-6596/2195/1/012028

Lutui, P. R., Cusack, B., & Maeakafa, G. (n.d.). *Energy Efficiency for IoT Devices in Home Environments*.

Machlev, R., Heistrene, L., Perl, M., Levy, K. Y., Belikov, J., Mannor, S., & Levron, Y. (2022). Explainable Artificial Intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities. In *Energy and AI* (Vol. 9). Elsevier B.V. https://doi.org/10.1016/j.egyai.2022.100169

Oniani, D., Hilsman, J., Peng, Y., COL, Poropatich, R. K., Pamplin, C. J. C., Legault, L. G. L., & Wang, Y. (2023). *From Military to Healthcare: Adopting and Expanding Ethical Principles for Generative Artificial Intelligence*. http://arxiv.org/abs/2308.02448

Praveen Kumar, B., & Hariharan, K. (2020, December 17). Multivariate Time Series Traffic Forecast with Long Short Term Memory based Deep Learning Model. *Proceedings of 2020 IEEE International Conference on Power, Instrumentation, Control and Computing, PICC 2020*. https://doi.org/10.1109/PICC51425.2020.9362455

Rjoub, G., Bentahar, J., Wahab, O. A., Mizouni, R., Song, A., Cohen, R., Otrok, H., & Mourad, A. (2023). *A Survey on Explainable Artificial Intelligence for Cybersecurity*. https://doi.org/10.1109/TNSM.2023.3282740

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. In *Nature Machine Intelligence* (Vol. 1, Issue 5, pp. 206–215). Nature Research. https://doi.org/10.1038/s42256-019-0048-x

Schmid, U., & Wrede, B. (2022a). Explainable AI. In *KI - Kunstliche Intelligenz* (Vol. 36, Issues 3–4, pp. 207–210). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/s13218-022-00788-0

Schmid, U., & Wrede, B. (2022b). What is Missing in XAI So Far?: An Interdisciplinary Perspective. *KI - Kunstliche Intelligenz*, *36*(3–4), 303–315. https://doi.org/10.1007/s13218-022-00786-2

Sorrell, S. (2015). Reducing energy demand: A review of issues, challenges and approaches. In *Renewable and Sustainable Energy Reviews* (Vol. 47, pp. 74–82). Elsevier Ltd. https://doi.org/10.1016/j.rser.2015.03.002

Soudan, M. B., Al Rifaie, H. M., Asmar, T. M., & Majzoub, S. (n.d.). *Smart Home Energy Management System An exploration of IoT use cases*. https://testtestest.firebaseio.com/Appliance1.json'

Srimaharaj, W., & Chaisricharoen, R. (2022). Internet of Behavior and Brain Response Identification for Cognitive Performance Analysis. *2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, 1308–1311. https://doi.org/10.23919/APSIPAASC55919.2022.9980191

Stary, C. (2020). The Internet-of-Behavior as Organizational Transformation Space with Choreographic Intelligence. *Communications in Computer and Information Science*, *1278*, 113–132. https://doi.org/10.1007/978-3-030-64351-5_8

Sun, J., Gan, W., Chao, H. C., Yu, P. S., & Ding, W. (2023). Internet of Behaviors: A Survey. *IEEE Internet of Things Journal*, *10*(13), 11117–11134. https://doi.org/10.1109/JIOT.2023.3247594

Tso, G. K. F., & Yau, K. K. W. (2007). Predicting electricity energy consumption: A comparison of regression analysis, decision tree and neural networks. *Energy*, *32*(9), 1761–1768. https://doi.org/10.1016/j.energy.2006.11.010

Ullah, F. U. M., Khan, N., Hussain, T., Lee, M. Y., & Baik, S. W. (2021). Diving deep into short-term electricity load forecasting: Comparative analysis and a novel framework. *Mathematics*, *9*(6). https://doi.org/10.3390/math9060611

Zhang, Y., Zheng, X., Liu, L., & Ma, H. (2020a). Multivariate and multi-frequency LSTM based fine-grained productivity forecasting for industrial IoT. *Proceedings - 2020 16th International Conference on Mobility, Sensing and Networking, MSN 2020*, 307–314. https://doi.org/10.1109/MSN50589.2020.00058

Zhang, Y., Zheng, X., Liu, L., & Ma, H. (2020b). Multivariate and multi-frequency LSTM based fine-grained productivity forecasting for industrial IoT. *Proceedings - 2020 16th International*

Conference on Mobility, Sensing and Networking, MSN 2020, 307–314. https://doi.org/10.1109/MSN50589.2020.00058

Zhao, H. X., & Magoulès, F. (2012). A review on the prediction of building energy consumption. In *Renewable and Sustainable Energy Reviews* (Vol. 16, Issue 6, pp. 3586–3592). https://doi.org/10.1016/j.rser.2012.02.049