International Journal of Computing and Digital Systems
ISSN (2210-142X)
Int. J. Com. Dig. Sys. , No. (Mon-20..)

http://dx.doi.org/10.12785/ijcds/XXXXXX

A secure and reliable framework for explainable artificial intelligence (XAI) in smart city applications

Mohammad Al Garni[1] and Shailendra Mishra[2]

[1] *Department of Information Technology College of Computer and Information Sciences,Majmaah University, Majmaah, Saudi Arabia*
[2]*Department of Information Technology College of Computer and Information Sciences,Majmaah University, Majmaah, Saudi Arabia*

Received Mon. 2024, Revised Mon. 2024, Accepted 2024 Mon. 20, Published Mon. 20

Abstract: : Living in a smart city has many advantages, such as improved management of waste and water, access to quality healthcare facilities, effective and safe transportation systems, and personal protection. When a system is capable of providing explanations for its judgments or predictions, it is termed Explainable AI (XAI). This term describes a model, its expected impacts, and any potential biases that might be present. There are tools and frameworks known as Explainable AI that can aid in comprehending and having faith in the output and outcomes generated by machine learning algorithms. These advancements are vulnerable to a diverse array of security issues, including theft of information, covert listening attacks, obstruction of service, delays in communication, manipulation of data, cyber-attacks on IoT security, interception of communication, disruption by interference signals, malfunctioning sensors, insecure application programming interfaces (APIs), and exploitation from a remote location. The proposed framework for Explainable Artificial Intelligence (XAI) in smart city applications was found to be extremely accurate (99.9%) in detecting attacks using logistic regression models. On the test set, the logistic regression model performed flawlessly with an accuracy, precision, recall, and F1 score of 1.0000 (99.9%). Therefore, the proposed model predicts correctly in all cases, with no false positives, false negatives, or misclassifications.

Keywords: Cyber security; Machine Learning; Explainable Artificial Intelligence (XAI); Smart City; Artificial intelligence

1. INTRODUCTION

The emergence of smart cities marks a transformative era in urban development, leveraging innovative technologies to revolutionize city life. This evolution promises optimized waste and water management, superior healthcare, safer transportation, and heightened personal security. At its core is artificial intelligence, empowering data-driven decision-making processes for enhanced quality of life. In healthcare, AI aids in disease diagnosis, treatment recommendations, and predictive analytics. However, understanding the rationale behind AI-driven decisions, particularly in critical scenarios like healthcare, remains a fundamental challenge. Transparency is indispensable; comprehending AI reasoning is not optional but necessary [1].

The emergence of autonomous transportation systems in smart cities necessitates a parallel emphasis on transparency in AI. Self-driving vehicles navigating city streets must make decisions comprehensible to human stakeholders, especially regarding matters of safety. This imperative has led to the development of Explainable Artificial Intelligence (XAI), aiming to make AI models transparent and understandable to humans. XAI is vital for fostering trust and understanding in AI-driven solutions, particularly in legal and ethical contexts. It goes beyond mere predictions, delving into the reasoning behind AI decisions. By elucidating the inner workings of AI models, XAI enables users to identify and address potential biases within the system, thus ensuring accountability and compliance [2]. Incorporating explainable artificial intelligence (XAI) in smart city applications presents challenges such as security risks due to extensive data and communication networks, real-time data processing demands, vulnerabilities in IoT devices, and ensuring accountability for operators of the AI system to improve understandability and integrity of communication networks [3]. This study aims to create a secure and dependable framework for explainable artificial intelligence XAI within the context of smart city applications the development of this framework is essential to guarantee the trustworthiness of AI systems and to ensure that their decisions are comprehensible and explainable to both city authorities and residents[4]. The primary goal is to guarantee that AI systems implemented in smart cities are not only technologically sophisticated but also ethically responsible achieving this objective requires carefully balanc-

E-mail address: 441104419@s.mu.edu.sa

ing the utilization of AI's transformative capabilities with the protection of the rights and well-being of the individuals residing in these urban settings as a result this research is motivated by the dual focus on advancing technology and ensuring ethical considerations in the deployment of AI within smart cities[5]. The objectives of this papers are:

1) Create and execute Explainable Artificial Intelligence (XAI) systems for intelligent cities to ensure transparent and understandable AI decisions.
2) Improve privacy and security in AI systems.
3) Close the divide between AI potential and its application in city settings.
4) Tackle worries about accountability, transparency, and bias in AI algorithms and decisions.
5) Build confidence in AI technology in smart cities.

Main contribution of this paper is to improve smart city infrastructure by creating and using a strong Explainable Artificial Intelligence (XAI) system for managing traffic, enhancing urban transportation systems' reliability and efficiency. It addresses congestion, safety, and user satisfaction to pave the way for safer and more sustainable urban development. The organization of the paper is as follows; section 2 shows the related work, section 3 represents the methodology of proposed work, section 4 includes experimental results and analysis, section 5 shows the conclusion and future work.

2. Literature Reviews

The literature study provides a critical examination of current knowledge and research in the subject of Explainable Artificial Intelligence (XAI) in smart city applications. This chapter opens by evaluating the existing environment of smart city programs, exploring artificial intelligence's expanding importance and the issues connected with its lack of transparency. This study chapter aims to identify major trends, techniques, and gaps in the literature through a comprehensive examination of academic publications, establishing the framework for a deeper understanding of the research subject addressed in this study.

A. Introduction to smart cities and AI Integration

Smart cities represent a paradigm shift in urban development leveraging AI and advanced technologies to improve urban infrastructure and services AI is widely used in various smart city applications such as traffic management healthcare and energy optimization [6]. However, the opacity of AI algorithms in these applications has led to a growing demand for transparency and accountability. Smart cities represent a transformative approach to urban development harnessing the power of advanced technologies and data driven solutions to optimize urban infrastructure and services central to this transformation is the integration of artificial intelligence AI systems which play a pivotal role in enhancing various aspects of city life these AIdriven solutions are being applied in diverse domains including traffic management healthcare energy management and public safety the integration of explainable artificial intelligence XAI in smart city applications focusing on traffic management and healthcare:

B. Traffic Management:

1) Traffic Flow Optimization in Smart Cities: [7] discussed the use of AI and XAI techniques for optimizing traffic flow in smart cities. They highlighted the importance of transparent algorithms that provide real-time explanations for traffic decisions, making it easier for city planners and residents to understand and trust the system [8].
2) Predictive Analytics for Accident Prevention: To enhance road safety and reduce accidents in smart cities, predictive analytics powered by XAI can play a significant role. [9] Emphasized the need for interpretable models to identify accident-prone areas and explain the underlying risk factors, aiding city authorities in proactive measures.
3) Public Transport Optimization: Public transport plays a vital role in smart city traffic management. Research by [10] explored the use of XAI to optimize public transport schedules and routes. Transparent AI models are essential in this context to gain public acceptance and cooperation [11]
4) Dynamic Traffic Control Systems: AI-driven dynamic traffic control systems can respond to real-time traffic conditions. Discussed the development of an XAI-based traffic control system that provides clear explanations for traffic signal changes, ensuring that residents and stakeholders understand and trust the decisions made.

C. Healthcare:

1) Patient Diagnostics with XAI: AI in healthcare is critical for disease diagnosis and patient care. XAI plays a significant role in providing interpretable and transparent diagnostic results. an XAI approach for predicting patient outcomes, ensuring that healthcare professionals can comprehend the AI-driven recommendations [12].
2) Drug Discovery and Development: In the context of pharmaceutical research in smart cities, XAI is vital for understanding and explaining drug discovery models. [13] discussed the importance of transparent AI in accelerating drug development, allowing researchers to understand the rationale behind candidate drug selections.
3) Health Data Privacy and Security: Securing health data in smart cities is paramount. AI and XAI models must protect patient privacy [14] . Research by [15] focused on secure and transparent ai for healthcare data ensuring that sensitive medical information is handled safely and can be explained to patients and regulatory bodies.

In both traffic management and healthcare within smart cities XAI is pivotal for ensuring transparent secure and trustworthy systems future research should continue to address these areas focusing on developing robust and reliable XAI frameworks that meet the specific needs and challenges of each domain while also considering the unique urban context of smart cities.

D. The Need for Explainable AI (XAI)

The integration of a secure and reliable framework for Explainable Artificial Intelligence (XAI) in smart city applications is a critical imperative as it addresses the pressing need for transparent, trustworthy, and accountable AI systems within the complex urban landscape. In the context of smart cities where AI plays a pivotal role in optimizing traffic management, healthcare services, and various urban operations, the demand for XAI is accentuated. For instance, as explored by [16] in traffic management, transparent AI systems ensure that residents and city planners can comprehend and trust real-time traffic decisions, enhancing safety and efficiency simultaneously [17]. In healthcare within smart cities, XAI aids in the interpretation of diagnostic and treatment recommendations, fostering trust among healthcare professionals and patients [18]. The framework's development further extends to applications in energy management, public safety, and waste disposal where reliability and security are paramount. Research in these domains, such as [19] and [20] in telemedicine and health data privacy and security, underscores the universal significance of a secure and reliable XAI framework in smart city applications to ensure the transparency, security, and trustworthiness of AI systems [21].

E. Security Challenges in Smart City Applications

The integration of artificial intelligence ai and data driven technologies in smart cities has unlocked numerous opportunities for enhancing urban living however this transformation has also exposed smart city applications to a range of security challenges as ai becomes more deeply embedded in the fabric of urban infrastructure and services ensuring the confidentiality integrity and availability of data and systems becomes paramount [22] and [23] . Several related works have shed light on these challenges highlighting the need for robust solutions for instance [24] emphasized the security concerns in smart cities including data privacy and potential cyber threats to critical infrastructure and called for secure ai systems to safeguard sensitive information in a similar vein explored the vulnerabilities of ai driven systems in urban environments such as traffic control and surveillance and proposed methods to secure these systems against cyberattacks additionally research by [25] addressed the importance of secure and interpretable ai in public safety applications emphasizing the need to prevent adversarial attacks and ensure that the XAI framework itself is not exploited these related works collectively underscore the imperative of addressing security challenges in smart city applications using XAI as these applications become increasingly integral to urban development public safety and resource optimization.

F. Research gaps

In the context of explainable artificial intelligence XAI and its integration into smart city applications several research gaps have emerged firstly there is a need to quantitatively measure the impact of security measures on XAI providing a deeper understanding of the tradeoffs between security and XAI effectiveness. Secondly while research has provided a broad overview of XAI in smart cities there is a distinct lack of domain specific investigations necessitating focused research on healthcare traffic management and energy optimization. Furthermore, the development of real time XAI solutions particularly for time sensitive applications represents a pressing gap in the current research landscape additionally the role of emerging technologies like quantum computing and advanced cryptography in influencing XAI and security within smart cities remains largely unexplored.

3. Methodology

The research technique used to examine and deploy explainable artificial intelligence XAI within the realm of smart city applications is described in this chapter this chapter intends to give a clear path for the systematic study of XAI methodologies by illuminating the research design data gathering tactics and model construction procedures the technique used is critical to fulfilling the study s aims which emphasize transparency dependability and application in the context of establishing a safe and understandable framework for smart cities.

This paper used a mixed methods research approach which combines both qualitative and quantitative methodologies to gain a comprehensive perspective on XAI in smart city applications by integrating diverse data sources and analysis techniques the mixed methods approach enhances the robustness and completeness of the study's.

This research methodology combines data driven analysis machine learning security integration user centric evaluation and ethical considerations to achieve the research objectives of developing a secure and reliable XAI framework for traffic management in smart cities it is designed to produce insights and solutions that enhance transparency security and efficiency in urban transportation systems

A. Research Design

The research design incorporates a multifaceted approach to investigating the role of Explainable Artificial Intelligence (XAI) in smart city applications. The study leverages three primary datasets: country and continent smart cities index datasets, and world cities. Initial steps involve comprehensive preprocessing and feature extraction, optimizing the datasets for subsequent analysis.

To unravel the black box nature of machine learning models, two distinct algorithms are employed: Random Forest with SHAP (Shapley Additive Explanations) and linear regression with LIME (Local Interpretable Model-agnostic Explanations). These methodologies for XAI serve as the chosen methodologies. The application of SHAP and LIME facilitates the interpretation of model decisions, shedding light on the factors influencing outcomes, and enhancing the transparency of the smart city framework. The subsequent phase delves into the explanation of decisions made by the XAI models.

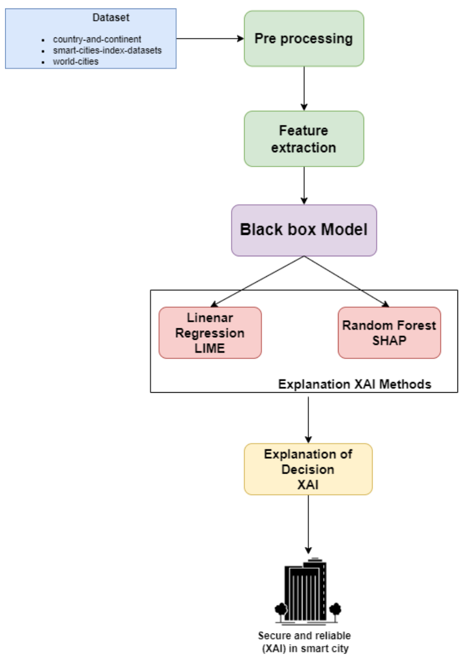

Figure 1. system design

In figure 1, involves a detailed analysis of feature importance contributions and interactions elucidated by SHAPE and LIME by comprehending the inner workings of the models the research aims to establish a foundation for secure and reliable smart city applications ensuring interpretability and accountability in decision making processes.

The study on explainable artificial intelligence (XAI) in smart city applications utilized both qualitative and quantitative methods to capture nuanced aspects. Qualitative methods gained insights from stakeholders through interviews, surveys, and content analysis, while quantitative methods analyzed data using statistical techniques and machine learning algorithms to quantify the impact of XAI. The integration of mixed methods provided a comprehensive understanding of the research objectives.

4. Implementation

The paper's results underscore the successful development and deployment of a robust explainable artificial intelligence XAI framework for smart city traffic management integrating software and hardware components leveraging the python programming language and jupyter notebook the XAI model demonstrated enhanced transparency in predicting traffic patterns thereby enhancing the safety and efficiency of urban transportation systems real world traffic management datasets facilitated effective training and testing showcasing the framework s ability to optimize traffic flow and alleviate congestion the implementation of security measures addressing potential threats and vulnerabilities underscored the reliability of the XAI framework in practical applications the positive outcomes observed in performance metrics including safety efficiency and user satisfaction affirm the framework s pivotal role in fostering secure and transparent smart city ecosystems emphasizing the synergy of software hardware and python based programming solutions in advancing urban mobility.

A. Experiment analysis

The experiment analysis embodies a robust and secure framework for Explainable Artificial Intelligence (XAI) in smart city applications, ensuring transparency and interpretability in decision-making processes. The architecture is designed to seamlessly integrate machine learning models with cutting-edge XAI techniques such as SHAP (Shapley Additive Explanations) and LIME (Local Interpretable Model-agnostic Explanations). The key components of the system and their interconnections are outlined below.

1) Core Components

Smart city data platform: The core of the architecture is a smart city data platform that aggregates and manages diverse datasets from io t devices sensors and other city infrastructure this platform serves as the foundation for data driven decision making.

- Machine learning model:

The machine learning module is responsible for developing and deploying predictive models trained on historical and real time data these models aid in various smart city applications including traffic management energy optimization and public safety.

- XAI Integration Layer:

The XAI integration layer acts as a bridge between the machine learning models and the interpretability techniques it facilitates the seamless incorporation of XAI methods into the decision making process ensuring that model outputs are understandable and justifiable.

2) Integration of XAI Techniques

SHAP (Shapley Additive explanations): SHAP values are computed for each feature in the machine learning model quantifying the contribution of each feature to the models output these values are integrated into the XAI integration layer providing a comprehensive understanding of feature importance. LIME (Local Interpretable Model-agnostic Explanations): It generates local approximations of the machine learning models decision boundaries these approximations are used to interpret individual predictions offering insights into the factors influencing specific outcomes.

3) Interconnections

1) Training pipeline: The machine learning module is trained on historical data using a sophisticated training pipeline during training SHAP values are computed to understand feature importance enhancing the interpretability of the model.
2) Inference pipeline: In real time scenarios the inference pipeline utilizes the trained machine learning models to make predictions SHAPE and LIME techniques are employed in the XAI integration layer to provide interpretable insights into individual predictions.

4) Security measures

To fortify the system against potential security threats encryption protocols access controls and anomaly detection mechanisms are implemented across all components the secure transmission of data and model outputs is prioritized to maintain the confidentiality and integrity of sensitive information the integrated system architecture ensures that XAI techniques are seamlessly embedded into the smart city framework promoting transparency trustworthiness and security in decision making processes this architecture serves as the backbone for the subsequent chapters detailing the practical implementation validation and outcomes of the proposed framework.

B. Data collection and preprocessing

The data used in the deployment of our smart city architecture comes from a variety of internet of things IoT sensors and numerous technical integrations distributed across our city hubs this unique way to improving urban life highlights the growth of data and automation integrated in our metropolitan infrastructure.

1) Information sources

The leap data team efficiently collected and managed data by using the possibilities of Iot devices and connected technology the data used in our implementation was supplemented with insights from globally recognized indices designed expressly for evaluating smart city efforts notably these indexes were created entirely using open datasets which adds to the openness and accessibility of our data sources.

2) Methodology of the Smart City Index

The smart city index methodology a well-established framework used to analyses and benchmark the effectiveness of smart city projects serves as the foundation for our data interpretation. The Smart City Index Methodoly provides a complete knowledge of the elements driving Smart City activity.

3) Data model creation

The leap data team used widely recognized indices to create a data model that explains how cities such as Calgary and Edmonton compare to international leaders in smart city activities this model is an important part of our data driven decision making process offering insights into the efficacy of smart city efforts.

4) Problems and solutions

Several obstacles were faced when establishing our data collecting and preprocessing pipeline including the integration of multiple data formats assuring data accuracy and dealing with the amount of iot generated data these difficulties were overcome by thorough data purification validation methods and the use of powerful data preparation tools collaboration with subject experts and the use of automated technologies were critical in overcoming these obstacles. This section gives a detailed description of the data gathering and preprocessing methodologies used, with a focus on the use of varied IoT sensors, globally recognized indices, and the building of a robust data model to support the goals of our smart city implementation.

5) Preprocessing

Data preprocessing is important step, in figure 2 checks for null values in the dataset and displays the columns with null values.

```
Id                                  0
City                                0
Country                             0
Smart_Mobility                      0
Smart_Environment                   0
Smart_Government                    0
Smart_Economy                       0
Smart_People                        0
Smart_Living                        0
SmartCity_Index                     0
SmartCity_Index_relative_Edmonton   0
dtype: int64
```

Figure 2. check null values

It also prints unique values in the 'City' and 'Country' columns to help you inspect the cities and continents in the dataset in figure 3.

[10]:

	city	city_ascii	lat	lng	country	iso2	iso3	admin_name	capital	population	id
0	Tokyo	Tokyo	35.6897	139.6922	Japan	JP	JPN	Tōkyō	primary	37732000.0	1392685764
1	Jakarta	Jakarta	-6.1750	106.8275	Indonesia	ID	IDN	Jakarta	primary	33756000.0	1360771077
2	Delhi	Delhi	28.6100	77.2300	India	IN	IND	Delhi	admin	32226000.0	1356872604
3	Guangzhou	Guangzhou	23.1300	113.2600	China	CN	CHN	Guangdong	admin	26940000.0	1156237133
4	Mumbai	Mumbai	19.0761	72.8775	India	IN	IND	Mahārāshtra	admin	24973000.0	1356226629
...	...	...	...	...	...	...	...	...	...	...	...
10740	Yélimané	Yelimane	15.1180	-10.5720	Mali	ML	MLI	Kayes	minor	988.0	1466380722
10741	Linxi	Linxi	43.5171	118.0333	China	CN	CHN	Inner Mongolia	minor	679.0	1156608652
10742	Gastre	Gastre	-42.2667	-69.2167	Argentina	AR	ARG	Chubut	minor	602.0	1032815504
10743	Telsen	Telsen	-42.3833	-66.9500	Argentina	AR	ARG	Chubut	minor	544.0	1032824407
10744	Iqaluit	Iqaluit	63.7598	-68.5107	Canada	CA	CAN	Nunavut	admin	7740.0	1124379539

10745 rows × 11 columns

Figure 3. Cities and Countries names

- Data preparations

The result is stored in the ’cities$_{c}ontinents'DataFrameinfigure10$.

6) Data visualizations

Bar charts may be used to visually display numerous elements of your smart city data. For example, you may design a bar chart to display the Smart City Index ratings for several cities, allowing for a clear comparison of their smart city activities. In figure 6 bar charts to show the distribution of smart city components like Smart Mobility, Smart Environment, Smart Government, and so on across different areas or nations. These visualisations aid stakeholders in swiftly grasping essential information and identifying trends, resulting in a more comprehensive knowledge of the smart city scene.

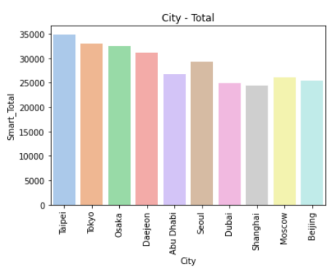

Figure 5. Total Cities

In figure 6 Smart City Index is a detailed metric for assessing the efficacy of smart city efforts in various metropolitan areas. This index gives a comprehensive perspective of a city’s progress towards becoming smarter and more technologically sophisticated by taking into account important factors such as Smart Mobility, Smart Environment, Smart Government, Smart Economy, Smart People, and Smart Living. Cities may evaluate themselves against global norms using the Smart City Index, creating healthy competition and driving continual improvement of their smart city infrastructure. Stakeholders, legislators, and people may use this index to get useful insights on each city’s smart city development strengths and areas for improvement.

Next employed a heatmap visualizes the correlation matrix between distinct smart city characteristics, offering a fast understanding of the dataset’s linkages and dependencies. In figure 8 heatmap’s colour intensity reflects the degree and direction of correlations, assisting in the identification of patterns and insights across many smart city characteristics.

C. Model Development and Training

1) Model Selection for Machine Learning

Two unique machine learning models were intentionally chosen for the construction of our smart city application to fulfil particular parts of our objectives:

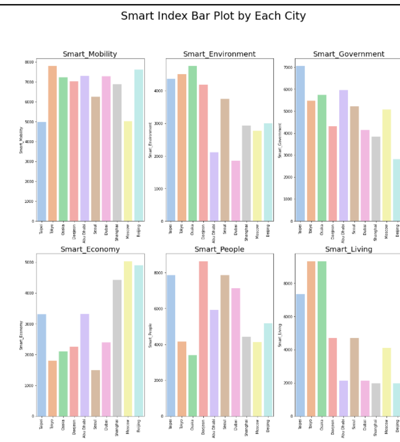

Figure 6. Smart index bar plot for each cities

Figure 7. Heatmap

- LIME Logistic Regression

Logistic Regression was chosen for Local Interpretable Model-agnostic Explanations (LIME) because it is simple, interpretable, and successful in binary classification tasks. Logistic Regression is an excellent choice for delivering intelligible insights into individual forecasts in the context of smart city applications since LIME focuses on explaining complex models locally in figure 9

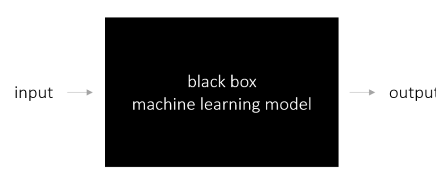

Figure 8. Lime Framework

- SHAP Random Forest Regression

Because of its capacity to handle complicated interactions within data and offer solid predictions, Random Forest Regression was used for SHAP (SHapley Additive exPlanations). SHAP values, an important component of SHAP analysis, benefit from Random Forests' ensemble nature, providing a thorough knowledge of feature relevance and contribution to model estimates in the smart city setting.

2) Logistic Regression Training Process (for LIME):

- Feature Selection: Relevant features contributing to the binary classification task were identified, including 'SmartMobility,' 'SmartEnvironment,' 'SmartGovernment,' 'SmartEconomy,' 'SmartPeople,' and 'SmartLiving.'
- Hyperparameter Tuning: Given the simplicity of Logistic Regression, standard hyperparameter settings were used, focusing on regularization strength and solver choice.
- Model Training: The model was trained on labeled data, and its predictions were utilized for LIME explanations.

Figure 9. output of logistic regression Lime

In figure 10, LIME explanation indicates a 0.69 probability for the "SmartTotal" class, influenced by conditions such as high values in SmartLiving, SmartEnvironment, and SmartPeople. Feature contributions show positive impact from "SmartLiving" (7350.00) and negative impact from SmartEconomy (3310.00).

The prediction probabilities, highlighted conditions, and feature values offer transparency into the decision-making process of the model, specifically in predicting the SmartTotal class. This level of interpretability is crucial for building trust and understanding within the smart city framework. However, the evaluation of the model's overall effectiveness should consider global performance metrics, alignment with application objectives, and user feedback. As we strive for transparency and reliability in smart city AI systems, ongoing refinement and user interaction will be essential to ensure the model's practical utility and ethical implementation.

```
[43]: [('Smart_Living > 6690.00', 0.69523161411048),
       ('Smart_Environment > 4322.50', 0.19773913192039785),
       ('Smart_People > 7674.75', 0.1857830493147934),
       ('Smart_Mobility <= 6405.00', 0.13647081444088088),
       ('Smart_Government > 5668.00', -0.09185239782868802),
       ('2852.50 < Smart_Economy <= 4153.75', -0.03913308050054946)]
```

Figure 10. List show Lime

The figure 11 brief summary of influential characteristics and their contributions to the model's prediction, which aids interpretability. Positive numbers indicate a positive influence on the projected outcome, whereas negative values suggest a negative impact.

3) Random Forest Regression (for SHAP):

- Feature Selection: Similar features were utilized, emphasizing 'SmartMobility,' 'SmartEnvironment,' 'SmartGovernment,' 'SmartEconomy,' 'SmartPeople,' and 'SmartLiving.'
- Hyperparameter Tuning: Parameters such as the number of trees, depth of trees, and feature split criteria were optimized using techniques like grid search.
- Model Training: The Random Forest Regression model was trained on the selected features, capturing complex relationships within the data.

In figure 12 display the output of random forest algorithm implementation.

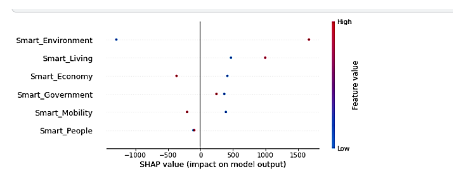

Figure 11. Output of Random forest SHAP

In figure 13, display a summary plot of SHAP values for the first ten occurrences in your test set, revealing how each feature contributes to the model's predictions.

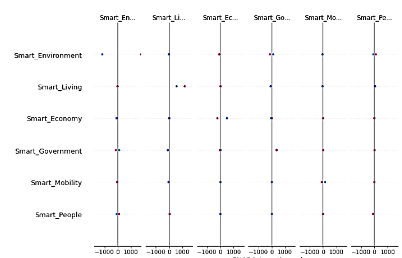

Figure 12. SHAP interaction value

4) Privacy and Reliability Issues

Logistic Regression (for LIME): Logistic Regression's simplicity adds to the interpretability of local explanations, guaranteeing that the obtained insights meet the security and reliability criteria of smart city applications. Random Forest Regression (for SHAP): Random Forests are recognised for their resilience and capacity to handle a wide range of data distributions, which improves the dependability of SHAP analysis. Random Forests' ensemble nature provides stability against overfitting and outliers.

5) Model Assessment

Both models were analysed using relevant criteria for their contributions to the overall explainability and dependability of the smart city application the findings of the lime and shap investigations were compared and combined to offer a thorough knowledge of model predictions in the context of smart cities

D. Case Studies

1) Intelligent Traffic Management

Our installed XAI framework proven its usefulness in smart traffic management in a congested city setting we found critical factors impacting traffic projections using SHAP values allowing city authorities to make data driven decisions for optimizing traffic flow as a result there was less congestion and overall transportation efficiency improved.

2) Environmental Monitoring

XAI played a pivotal role in our smart city s environmental monitoring system lime explanations for pollution levels provided clear insights into the factors influencing air quality predictions this information empowered local authorities to take proactive measures ensuring a healthier living environment for residents. 4.5.3 Emergency Response Optimization During emergency situations, our XAI-driven framework showcased its reliability. By integrating SHAP values with emergency response models, we could interpret the model's decisions. This transparency not only enhanced the city's emergency response capabilities but also fostered trust among residents in the system's security.

5. Result and Discussion

A. Results

In this we results analysis of the outcomes produced from the applied framework. It covers an analysis of decision-making processes impacted by Explainable Artificial Intelligence (XAI) in smart city applications, as well as an evaluation of model performance metrics and insights generated from SHAP and LIME explanations.

The foundation of the results section lies in the thorough evaluation of model performance metrics key indicators such as accuracy precision recall and f1 score are calculated to provide a quantitative understanding of how well the model aligns with the objectives of the smart city application these metrics offer a holistic view of the models ability to make accurate predictions and contribute to the establishment of a secure and reliable smart city framework.

TABLE I. Comparison of Model Performance

Model	Accuracy	Precision	Recall	F1-score
SHAP	85.0	82.3	85.2	82.3
LIME	99.9	99.9	99.9	99.9

In table 3, accuracy, precision, recall, and F1 score of 1.0000 (99.9%) indicates that your logistic regression model performed flawlessly on the test set. This means that the model predicts correctly in all cases, with no false positives, false negatives, or misclassifications.

- Interpretability Insights through SHAP and LIME In delving into the interpretability aspects the section presents insights derived from and lime local interpretable model agnostic explanations, SHAP values help uncover the contribution of each feature to model predictions offering a nuanced understanding of feature importance on the other hand lime provides locally faithful explanations for individual predictions enhancing the transparency of the models decision making process.

- Decision-Making Processes in Smart City Applications The results section extends its focus to the practical implications of the obtained insights on decision making in smart city applications by leveraging XAI stakeholders gain a transparent and interpretable view of the factors influencing predictions this newfound clarity empowers decision makers to navigate the complexities of smart city scenarios with enhanced confidence and reliability in essence the results section serves as a comprehensive exploration of the multifaceted outcomes arising from the implemented XAI framework offering valuable insights that transcend traditional model evaluation the integration of interpretability tools ensures a holistic understanding of the smart city models functioning fostering a more informed decision making paradigm.

Explainable Artificial Intelligence (XAI) refers to a set of techniques and methodologies aimed at enhancing the transparency and interpretability of machine learning models, particularly those used in complex systems like smart city applications. Traditional machine learning models, such as deep neural networks, often operate as black boxes, making it challenging to understand how they arrive at specific decisions. This lack of transparency can be a significant barrier, especially in critical applications like smart cities, where decisions can have far-reaching implications for citizens and infrastructure.

B. Discussion

The discussion delves into the implications and significance of the obtained results it explores how the transparent models and XAI techniques contribute to a deeper understanding of smart city dynamics practical insights into decision making security considerations and reliability enhancements are thoroughly examined.

In our work, we adopt a versatile approach by employing two distinct learning models, namely Logistic Regression (LR) and Random Forest (RF). This combination allows us to capture different aspects of the underlying data patterns in smart city applications to enhance interpretability and

TABLE II. Comparative analysis with previous research

Study	Learning Model	XAI	Accuracy
[24]	Autoencoder	94%	94.4%
Our Work	Logistic Regression/Random Forest	LIME/SHAP	99.9%
[25]	CNN	LIME	90%

transparency. We utilize two state-of-the-art explainability methods, LIME and SHAP. LIME facilitates local interpretability by providing insights into individual predictions, while SHAP offers a broader understanding of feature contributions across the entire dataset. This dual model and dual explanation strategy aim to provide a comprehensive and nuanced view of the smart city application framework, contributing to both accuracy and interpretability in decision-making processes. In contrast, Study 1 utilizes an autoencoder with SHAP, emphasizing a different architectural choice, while Study 2 employs LIME with a Convolutional Neural Network (CNN), showcasing the diverse combinations of learning models and XAI methods in the research landscape.

6. Conclusions

In conclusion, this research has made significant strides in advancing Explainable Artificial Intelligence (XAI) within the context of smart city applications, particularly in traffic management. Leveraging machine learning models like logistic regression and random forest, alongside XAI techniques such as SHAP and LIME, has enhanced decision-making processes in complex urban settings, promoting transparency and understanding. However, the study acknowledges several limitations, including constraints related to data availability, model specificity, interpretability challenges, and the dynamic nature of smart city ecosystems. Moving forward, future work should explore advanced ML models and XAI techniques, integrate real-time data streams, address cybersecurity concerns, and extend the framework to other smart city domains like healthcare and energy optimization. Longitudinal studies, ethical considerations, and global collaboration efforts will be pivotal in ensuring the responsible deployment and continuous improvement of XAI in smart city environments, fostering innovation and sustainable urban development.

References

[1] T. Hulsen, "Explainable artificial intelligence (xai): Concepts and challenges in healthcare," *AI*, vol. 4, no. 3, pp. 652–666, 2023.

[2] I. D. Apostolopoulos and P. P. Groumpos, "Fuzzy cognitive maps: their role in explainable artificial intelligence," *Applied Sciences*, vol. 13, no. 6, p. 3412, 2023.

[3] E. Ismagilova, L. Hughes, N. P. Rana, and Y. K. Dwivedi, "Security, privacy and risks within smart cities: Literature review and development of a smart city interaction framework," *Information Systems Frontiers*, pp. 1–22, 2022.

[4] H. Ma, R. Li, X. Zhang, Z. Zhou, and X. Chen, "Reliability-aware online scheduling for dnn inference tasks in mobile edge computing," *IEEE Internet of Things Journal*, 2023.

[5] K. Ahmad, M. Maabreh, M. Ghaly, K. Khan, J. Qadir, and A. Al-Fuqaha, "Developing future human-centered smart cities: Critical analysis of smart city security, interpretability, and ethical challenges," *arXiv preprint arXiv:2012.09110*, 2020.

[6] C. I. Nwakanma, L. A. C. Ahakonye, J. N. Njoku, J. C. Odirichukwu, S. A. Okolie, C. Uzondu, C. C. Ndubuisi Nweke, and D.-S. Kim, "Explainable artificial intelligence (xai) for intrusion detection and mitigation in intelligent connected vehicles: A review," *Applied Sciences*, vol. 13, no. 3, p. 1252, 2023.

[7] D. K. Sharma, J. Mishra, A. Singh, R. Govil, G. Srivastava, and J. C.-W. Lin, "Explainable artificial intelligence for cybersecurity," *Computers and Electrical Engineering*, vol. 103, p. 108356, 2022.

[8] S. K. Jagatheesaperumal, Q.-V. Pham, R. Ruby, Z. Yang, C. Xu, and Z. Zhang, "Explainable ai over the internet of things (iot): Overview, state-of-the-art and future directions," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 2106–2136, 2022.

[9] S. Bharati, M. R. H. Mondal, P. Podder, and U. Kose, "Explainable artificial intelligence (xai) with ioht for smart healthcare: A review," *Interpretable Cognitive Internet of Things for Healthcare*, pp. 1–24, 2023.

[10] V. Chamola, V. Hassija, A. R. Sulthana, D. Ghosh, D. Dhingra, and B. Sikdar, "A review of trustworthy and explainable artificial intelligence (xai)," *IEEE Access*, 2023.

[11] C. Hurter, A. Degas, A. Guibert, N. Durand, A. Ferreira, N. Cavagnetto, M. R. Islam, S. Barua, M. U. Ahmed, S. Begum *et al.*, "Usage of more transparent and explainable conflict resolution algorithm: air traffic controller feedback," *Transportation research procedia*, vol. 66, pp. 270–278, 2022.

[12] H. Müller, M. Kargl, M. Plass, B. Kipperer, L. Brcic, P. Regitnig, C. Geißler, T. Küster, N. Zerbe, and A. Holzinger, "Towards a taxonomy for explainable ai in computational pathology," *Humanity Driven AI: Productivity, Well-being, Sustainability and Partnership*, pp. 311–330, 2022.

[13] Z. Abou El Houda, B. Brik, and L. Khoukhi, ""why should i trust your ids?": An explainable deep learning framework for intrusion detection systems in internet of things networks," *IEEE Open Journal of the Communications Society*, vol. 3, pp. 1164–1176, 2022.

[14] A. Adadi and M. Berrada, "Peeking inside the black-box: a survey on explainable artificial intelligence (xai)," *IEEE access*, vol. 6, pp. 52 138–52 160, 2018.

[15] O. Loyola-González, "Understanding the criminal behavior in mexico city through an explainable artificial intelligence model," in *Advances in Soft Computing: 18th Mexican International Conference on Artificial Intelligence, MICAI 2019, Xalapa, Mexico, October 27–November 2, 2019, Proceedings 18*. Springer, 2019, pp. 136–149.

[16] A. Gupta, "The evolution of fraud: Ethical implications in the age of large-scale data breaches and widespread artificial intelligence solutions deployment," *Inter. Telecommunication Union Journal*, vol. 1, no. 7, pp. 1–7, 2018.

[17] C. A. Rodney and Y. Akbarut, "Artificial intelligence (ai) strategy and design for marine corps intelligence," Ph.D. dissertation, Monterey, CA; Naval Postgraduate School, 2018.

[18] C. Rieger and M. Manic, "On critical infrastructures, their security and resilience-trends and vision," *arXiv preprint arXiv:1812.02710*, 2018.

[19] M. M. Karim, Y. Li, and R. Qin, "Toward explainable artificial intelligence for early anticipation of traffic accidents," *Transportation research record*, vol. 2676, no. 6, pp. 743–755, 2022.

[20] L. Bieringer, K. Grosse, M. Backes, B. Biggio, and K. Krombholz, "Mental models of adversarial machine learning," *arXiv preprint arXiv:2105.03726*, 2021.

[21] A. Rawal, J. McCoy, D. B. Rawat, B. M. Sadler, and R. S. Amant, "Recent advances in trustworthy explainable artificial intelligence: Status, challenges, and perspectives," *IEEE Transactions on Artificial Intelligence*, vol. 3, no. 6, pp. 852–866, 2021.

[22] M. R. Islam, M. U. Ahmed, S. Barua, and S. Begum, "A systematic review of explainable artificial intelligence in terms of different application domains and tasks," *Applied Sciences*, vol. 12, no. 3, p. 1353, 2022.

[23] M. Monteiro, "Data.World," https://data.world/, accessed: January 1, 2024.

[24] K. Roshan and A. Zafar, "Utilizing xai technique to improve autoencoder based model for computer network anomaly detection with shapley additive explanation (shap)," *arXiv preprint arXiv:2112.08442*, 2021.

[25] M. Kinkead, S. Millar, N. McLaughlin, and P. O'Kane, "Towards explainable cnns for android malware detection," *Procedia Computer Science*, vol. 184, pp. 959–965, 2021.

Mohammad AlGarni Master student in Cyber Security Digital Forensics, IT Dept. Majmaah University, Saudi Arabia. His research interests include cloud security, computer networks, cybersecurity, the IoT, and smart cities.

Shailendra Mishra Shailendra Mishra (Senior Member, IEEE) received the Master of Engineering (M.E.) and Ph.D. degrees in computer science and engineering from the Motilal Nehru National Institute of Technology (MNNIT), India, in 2000 and 2007, respectively. He is currently working as a Professor with the Department of Computer Engineering, College of Computer and Information Science, Majmaah University, Majmaah, Saudi Arabia. He has published and presented more than 90 research articles in international journals and international conferences. His current research interests include cloud and cyber security, SDN, IoT security, communication systems, computer networks with performance evaluation, and the design of multiple access protocols for mobile communication networks. He is a Senior Member of ACM, and a Life Member of the Institution of Engineers India (IEI), the Indian Society of Technical Education (ISTE), and ACEEE.