Chapter 1

Explainable Artificial Intelligence Methods Applied to Unmanned Aerial Vehicles: A Conceptual Framework

Mustafa Meliksah Ozmen[1]

Bekir Aksoy[2]

Abstract

The study examined the use of explainable artificial intelligence (XAI) methods for unmanned aerial vehicles (UAVs). The aim of the study is to make the artificial intelligence decision processes in the autonomous systems of UAVs more understandable and transparent. This requirement has gained importance as UAVs are used in increasingly complex tasks. In order to increase the reliability of UAVs especially in areas such as security, military operations and agriculture, and to reinforce users' trust in these systems, the integration of XAI methods has become critical.

In the study, both model-based and post-hoc explanation methods were used. Model-based methods such as decision trees and explainable neural networks were preferred to ensure transparency of UAVs' decision-making processes. Post-hoc methods such as Local Intrepretable Model - Adnogtics Explanations (LIME) and SHapley Additive Explanations (SHAP) were used to increase the explainability of deep learning models . In the study, these techniques were evaluated to understand why UAVs make certain decisions in their complex operational processes.

As a result of the study, it was observed that the integration of Explainable Artificial Intelligence (XAI) methods into UAV systems made decision processes more transparent and thus increased the reliability of UAVs. It was emphasized that this transparency, especially in terms of public safety and ethical practices, plays a critical role in expanding the use of UAVs and making them safer.

---

1 PhD, Isparta University of Applied Sciences, cadet.2045@gmail.com Isparta, 0000-0003-3585-0518

2 Associate Professor, Isparta University of Applied Sciences, bekiraksoy@isparta.edu.tr Isparta, 0000-0001-8052-9411

https://doi.org/10.58830/ozgur.pub575.c2358

1. INTRODUCTION

With the rapid development of technology today, the use of Unmanned Aerial Vehicles (UAVs) has become widespread. UAVs are frequently used in many important areas such as facilitating social life and performing tasks that are harmful to human health. One of the most important advantages of UAVs is their high success rates in performing tasks. These success rates also minimize possible human-related errors.

UAVs in the realized book section A conceptual framework study based on academic literature has been carried out on the use of Artificial Intelligence (AI) technologies, which are one of the frequently used technologies together with AI technologies. The structure of the study has been examined by considering the general usage areas and importance of UAVs, AI and UAVs, Explainable artificial intelligence (XAI) structure, and the use of XAI in UAVs in the introduction section. The second section , XAI methods and algorithms, has been discussed under two main headings as model-based and post-hoc based methods. Under the title of model-based method, decision trees and explainable neural network models have been examined. In post-hoc based methods, LIME, SHAP models were discussed in detail . In the third part of the study, academic literature-based reviews of UAVs' XAI applications were conducted. In the final stage, the results of the study were discussed in detail.

1.1. General Areas of Use and Importance of Unmanned Aerial Vehicles (UAVs)

With the advancement of technology, the use of UAVs has also begun to be used frequently in social life. UAVs are frequently used especially in military and civilian life. (Thiels et al., 2015). It is possible to define UAVs as vehicles that can be remotely controlled and have the ability to perform autonomous flight. In addition to their high maneuverability due to their small size, their low cost makes the use of UAVs more widespread and effective (Shadiev & Yi , 2023).

of UAVs is in military areas. UAVs have successfully performed many tasks such as reconnaissance, surveillance, intelligence gathering and attack in military operations. In particular, UAVs are of great importance in terms of reducing loss of life and cost effectiveness since they do not require a pilot (Santos et al., 2023).

UAVs have a wide range of applications in civilian areas, from agriculture to mapping and environmental monitoring. (Muhmad Kamarulzaman et al., 2023). It has been determined that UAVs used in the agricultural sector provide great efficiency in areas such as crop monitoring, spraying and water management and reduce the workload of farmers (Istiak et al., 2023). In addition, the use of UAVs in environmental monitoring and natural disaster management allows for fast and effective interventions. The rapid data collection capacity of UAVs in disasters such as forest fires and floods has great advantages in many points such as the importance of saving human lives (Devoto et al., 2020). Another area of use of UAVs is cargo transportation and logistics. One of the important advantages of UAVs is that they provide fast delivery services, especially in regions where road transportation is difficult, and are used in vital tasks such as the transportation of emergency medical supplies (Çorbacı & Doğan, 2023). As can be understood from the examples given above, it is an inevitable fact that UAVs are an important technological tool not only for military operations but also in all areas of civilian life.

As a result, the advantages offered by UAV technologies in a wide range increase the strategic importance of UAVs day by day. The economic and operational efficiency provided by UAVs in both military and civilian applications causes this technology to become widespread at the global level.

1.2. Artificial Intelligence and UAVs

AI is a branch of science and technology that creates the characteristics and abilities of human intelligence through computer programs (PK, 1984). AI enables machines to perform human-like tasks by modeling the complex functions of the human mind, such as learning, reasoning, problem solving, and language understanding.

UAVs have evolved significantly in recent years thanks to the integration of AI technologies. Its application to UAVs has increased the ability of these vehicles to move autonomously, perceive their environment and adapt to dynamic conditions. This integration has allowed UAVs to perform more complex and sensitive missions in military, civil and commercial areas (Thiels et al., 2015).

AI's One of the most obvious examples of its use in UAVs is autonomous navigation. While traditional UAV systems are dependent on pre-programmed routes, AI-supported UAVs can avoid obstacles and reach their targets in the most effective way by performing real-time data analysis (Talwandi et al., 2024). For example, image processing systems developed using deep learning algorithms can It helps them perceive their environment better and respond quickly to sudden changes (Pandey et al., 2024).

Additionally, AI-based object recognition and tracking systems allow UAVs to detect specific objects or people in search and rescue operations or security applications (Devoto et al., 2020). This offers great advantages, especially in critical missions such as finding missing persons in natural disasters or ensuring border security.

Swarm technology is also used in AI It is another important area of use in UAVs . Due to AI algorithms, multiple UAVs can act in a coordinated manner without a central control mechanism (Campion et al., 2018). This technology enables fast and effective scanning of large areas or large-scale data collection.

AI technologies with UAVs has significantly expanded the capabilities and application areas of these vehicles. AI-supported UAVs are shaping the technological developments of the future by offering innovative solutions in many areas, from agriculture to logistics, from security to environmental monitoring.

1.3. What is Explainable Artificial Intelligence (XAI)?

XAI refers to a set of methods and techniques designed to make the decision-making processes of AI systems transparent and understandable to humans (Adadi & Berrada , 2018). Unlike traditional "black box" models that offer little insight into how inputs are transformed into outputs, XAI provides clear explanations of how and why an AI system reached a particular decision. The importance of XAI lies in its ability to build trust between users and AI systems, facilitate compliance with regulatory standards that require accountability, and enable the identification and reduction of biases within models (Gunning et al., 2019). By making AI models interpretable, XAI improves ethical decision-making and supports the responsible deployment of AI technologies in critical areas such as healthcare, finance, and autonomous vehicles (Doshi-Velez & Kim, 2017).

1.4. XAI's Importance and Necessity of Use in UAVs

XAI has become increasingly critical in the development and deployment of UAVs as these systems grow in complexity and autonomy. UAVs , commonly known as drones , have advanced significantly through the integration of AI, particularly in the areas of autonomous navigation, surveillance, and data analysis (Floreano & Wood , 2015). However, the complexity of AI algorithms often causes them to operate as "black boxes," where decision-making processes are not transparent or easily understood

(Lipton, 2018). This lack of transparency raises concerns about safety, ethical concerns, and trust among users and regulators.

XAI's Integrating into UAVs addresses these challenges by making AI systems more transparent and interpretable (Keneni, 2018). By providing clear explanations of how AI systems reach certain decisions, XAI increases the reliability and safety of UAV operations. This transparency is essential for operators and stakeholders to trust the actions of UAVs , especially in applications involving public safety and security (Adadi & Berrada , 2018).

Safety is a top concern in UAV operations; failures or errors can lead to property damage or loss of life. XAI facilitates the detection of potential errors and biases in AI algorithms, contributing to the development of more robust and reliable UAV systems (Keneni, 2018). By making decision-making processes transparent, engineers can more effectively diagnose errors and quickly implement corrective measures. This capability is especially important in complex environments where UAVs must make instantaneous decisions based on real-time data.

Incorporating explainable AI into UAVs is both important and necessary for the safe, ethical, and effective deployment of these technologies. XAI increases transparency, security, and compliance with regulatory and ethical standards—critical factors for the continued growth and acceptance of UAV applications. By making AI systems in UAVs understandable, we not only improve their functionality, but also build the trust necessary for their integration into various aspects of society.

2. EXPLAINABLE ARTIFICIAL INTELLIGENCE METHODS AND ALGORITHMS

Internet technology and the ability to make sense of the large data sets that increase due to internet technology have been one of the main factors in the rapid development of AI technologies. However, investigating the reasons for the results obtained from AI training has also been a phenomenon that has developed together with AI technologies. For this reason, the concept of XAI has emerged as the meaning of the results obtained from AI training. XAI is a field of research that aims to make the internal workings and decision-making processes of machine learning models understandable to humans (Gunning , 2017).

The history of explainable AI has followed a parallel course with the general developments in the field of AI. Since the first AI systems were built on symbolic structures and rule-based models, the decision-making processes of the models could be easily understood by humans (Russell &

Norvig , 2010). Since the internal workings and logic of the systems were transparent in this period, explainability became a natural feature. However, in the late 1990s and early 2000s, rapid developments in machine learning and especially deep learning significantly increased the complexity of AI models (LeCun , Bengio & Hinton , 2015). These models, which showed high performance on large data sets, became "black boxes" whose internal structures and decision-making mechanisms were difficult to understand by humans . This has raised concerns about the reliability and accountability of models. Traditional "black box" models, despite their high performance, have made it difficult to understand the logic behind their decisions due to the complexity they contain (LeCun , Bengio , & Hinton , 2015). This leads to reliability and accountability problems in many fields, especially in medicine, finance, and law. Therefore, explainable AI addresses these problems by explaining how and why the predictions obtained with models reach certain results and can be considered to provide transparency for AI technology (Ribbon , Singh, & Guestrin , 2016). Explainability is an ethical and legal obligation, beyond a technical requirement. The European Union's General Data Protection Regulation (GDPR) emphasizes the importance of explainability in automated decision-making processes (Wachter , Mittelstadt , & Floridi , 2017). In this context, explainable AI aims to prevent possible risks and misunderstandings while increasing the social acceptance of AI systems. For this purpose, the US Defense Advanced Research Projects Agency (DARPA) first initiated the XAI program in 2016 to encourage research in this field (Gunning , 2017). The program aimed for AI systems to both demonstrate high performance and ensure that their decisions were understandable by humans. The decisions of AI systems can profoundly affect the lives of individuals. For this reason, explainability is considered an ethical imperative, beyond being a technical need. For example, in medical diagnostic systems, the explainability of decisions ensures that doctors and patients trust these decisions (Doshi-Velez & Kim, 2017). Similarly, explainability is of critical importance in financial decision support tools in terms of detecting and correcting errors.

2.1. Model-Based Methods

In the field of XAI, model-based methods involve using models that are directly understandable and interpretable. Since these models are transparent in nature, decision-making processes can be easily followed by humans. Examples of model-based methods include models such as decision trees and explainable neural networks (Molnar , 2019). The advantage of model-based methods is that they integrate explainability into the design of the model. However, when working with complex data or high-dimensional data sets, their performance may be lower than that of deep learning models, so in some cases a balance may need to be established between explainability and performance (Rudin , 2019).

2.1.1. Decision Trees

Decision trees are models that are frequently used in the field of AI and machine learning and are naturally explainable due to their internal structures. These models perform classification or regression tasks by branching the data according to certain features (Quinlan , 1986). The explainability of decision trees stems from the fact that their structures are easily understandable by humans; each node and branch explains the logic behind the decisions taken by the model (Hastie , Tibshirani , & Friedman, 2009). Since decision trees are an internally interpretable model, they can be considered a model-based model (Molnar , 2019). In these models, since the decision processes and the features used are clearly visible, it is understandable why the model reached a certain conclusion (Breiman , Friedman, Olshen , & Stone, 1984). For example, a decision tree model can use features such as income level, credit score, and existing debts to approve a customer's loan application, and the effect of each of these features on the decision can be read directly from the structure of the tree (Rokach & Maimon , 2005).

Decision trees are frequently preferred in applications requiring explainability in health, finance, marketing and many other areas (Witten , Frank, Hall & Pal, 2016). Especially in sectors where ethical and legal regulations are important, it is critical that the models are transparent and the justifications for the decisions are understandable (Murthy , 1998). The advantages of decision trees are listed below.

- Easy Understandability: Decision trees can be understood even by non-experts thanks to their visually representable structures (Quinlan , 1986).
- Feature Selection: It can be easily seen from the structure of the tree which features the model attaches more importance to (Breiman , Friedman, Olshen & Stone, 1984).
- Fast Computation: Decision trees can generally be trained and predicted quickly, making them practical when working with large data sets (Witten , Frank, Hall & Pal, 2016).

In addition to the advantages mentioned above, decision trees also have some limitations. In particular, the overfitting problem can cause the model to adapt too much to the data it was trained on and to perform poorly on new data (Rudin , 2019). Pruning techniques are used to reduce this problem (Hastie , Tibshirani & Friedman, 2009). In addition, decision trees can have performance problems in complex data structures and high-dimensional data sets (Murthy , 1998).

To overcome these limitations, ensemble methods such as Random Forests and Gradient Boosting Trees have been developed (Rokach & Maimon , 2005). These methods offer higher accuracy and generalization ability by combining multiple decision trees. However, the explainability of these ensemble models is lower than a single decision tree (Molnar , 2019). In this case, it may be necessary to use additional techniques and tools for explainability (Breiman , Friedman, Olshen , & Stone, 1984).

2.1.2. Explainable Neural Networks

Explainable neural networks are approaches that prioritize explainability when designing the structure of the model and the learning process (Zhang & Zhu, 2018). These models aim to present decision-making processes and the information they represent internally in a human-understandable form. Some approaches aim to increase explainability by simplifying the architecture of neural networks or by imposing certain restrictions. For example, linear functions are used as activation functions in Linear Neural Networks to make the behavior of the model more understandable (Kawaguchi , 2016). In model-based methods, certain layers or activation functions of neural networks can be designed with explainability in mind. For example, ReLU (Rectified Linear Unit) activation function can be used instead of more explainable activation functions (Maas , Hannun & Ng, 2013). In addition, the explainability of the model is increased by showing which features it focuses on with attention mechanisms (Bahdanau , Cho & Bengio , 2015). Embedded explanation layers can be added to make the information in the neural networks more understandable. These layers ensure that the intermediate outputs of the model are interpretable by humans (Alain & Bengio , 2017). Prototype-based neural networks provide justifications for their decisions using prototype examples for each class while classifying (Li , Liu , Chen & Rudin , 2018). Explainable neural networks are frequently used in important fields such as medicine and autonomous driving. In the field of medical image analysis, explaining why neural networks make a certain diagnosis allows doctors to trust these decisions (Esteva et al., 2017). Similarly, in autonomous vehicles, the explainability of decisions is very important in terms of safety and legal responsibilities (Kim & Canny , 2017). Explainable neural networks provide reliability and accountability by making the internal workings of the models more understandable (Montavon , Samek & Müller, 2018). However, while increasing explainability, the performance of the model may be compromised (Chen et al., 2019). Therefore, establishing a balance between explainability and performance is also very important.

2.2. Post-Hoc Methods

Post-hoc methods are one of the important techniques used in the field of XAI to explain the decisions and predictions of complex and often difficult-to-understand models (e.g. deep neural networks). These methods are applied after the model has been trained and aim to understand its outputs and decision processes without changing the internal structure of the model (Lipton, 2016).

Modern machine learning methods, and especially deep learning-based approaches, are characterized as "black boxes" due to their inherent complexity, even though they achieve successful results with high accuracy rates (Guidotti et al., 2018). This reduces the confidence in the results obtained from the models and can lead to critical problems, especially in fields such as medicine, finance and law (Doshi-Velez & Kim, 2017). Post-hoc methods increase the reliability and accountability feature by making the decisions of complex models more understandable (Samek , Wiegand & Müller, 2017).

The biggest advantage of post-hoc methods is that they help understand the decisions of high-performance but unexplainable models (Guidotti et al., 2018). Since these methods can be applied without changing the internal structure of the model, they can be used on existing models without any additional effort (Alvarez -Melis & Jaakkola , 2018). However, post-hoc explanations may not always fully reflect the real decision processes of the model and can sometimes be misleading (Rudin , 2019). Therefore, these explanations should be interpreted carefully. The LIME and SHAP methods, which are frequently used post-hoc methods, are discussed in detail below.

2.2.1. LIME (Local Interpretable Model - Agnostic Explanations)

As the complexity of AI and machine learning models increases, it becomes increasingly difficult to understand the decision-making processes of the models (Samek , Wiegand , & Müller, 2017). In particular, the "black box" nature of deep learning models makes it difficult for users and developers to understand how XAI models work and why they make certain decisions (Adadi & Berrada , 2018). In this context, the field of XAI has emerged as a research area that aims to make the inner workings and decision-making mechanisms of AI models compatible with human understanding (Gilpin et al., 2018).

LIME method is a prominent technique in the XAI field that can produce local explanations independent of the model (Molnar , 2019). First developed by Ribeiro et al. in 2016, LIME aims to explain the individual predictions of any machine learning model using a simple and understandable model (Ribeiro , Singh & Guestrin , 2016). This method is an effective tool for interpreting the decisions of especially complex and difficult-to-understand models.

of LIME starts by creating synthetic data samples around the data point of interest (Garreau & Luxburg , 2020). These synthetic samples are data that are similar to the original data point but show variations in certain features. The black-box model makes predictions on these synthetic data and is trained with a simple model such as linear regression to examine the relationship between these predictions and the synthetic data (Lundberg & Lee, 2017). Thus, the trained simple model approximately represents the behavior of the black-box model locally and determines the extent to which the decision is affected by which features (Arras et al., 2017).

One of the most important advantages of LIME is that it is model-agnostic (Guidotti et al., 2018). This means that LIME can be used with any machine learning model. Therefore, it is effectively used in explaining AI models such as deep neural networks, random forests or support vector machines (Ribbon , Singh & Guestrin , 2018).

LIME, which has a wide range of applications, has been successfully applied in different areas such as text classification (Hendricks et al., 2016), image recognition (Selvaraju et al., 2017), medical diagnosis (Tjoa & Guan , 2020) and financial forecasting (Chen et al., 2018). For example, in the field of medical diagnosis, LIME can be used to understand how a model reveals a certain diagnosis, and thus doctors can better evaluate the model's decisions (Holzinger et al., 2019). Similarly, in financial forecasting models, risk assessment can be made more transparent with LIME (Bussmann et al., 2020).

The explainability of AI models (Miller, 2019). It plays an important role in complying with ethical, legal and practical requirements by making the decisions of the models more understandable and increasing the trust of users in these models (Barredo In the future, with the further development of LIME and similar methods, it is expected that AI applications will become more transparent and accountable (Gunning & Aha, 2019).

2.2.2. Shapley Additive exPlanations

Another important method in the field of XAI is SHAP method (Lundberg & Lee, 2017). First proposed by Lundberg and Lee in 2017, this method aims to calculate the contribution of each feature to the model's prediction in a fair and consistent way (Lundberg & Lee, 2017). The working principle of SHAP is based on the Shapley values used in cooperative game theory [51,52]. The Shapley value is used to fairly determine the marginal contribution of each player to the total gain in a game (Roth, 1988). SHAP adapts this concept to AI models and calculates the contribution of each feature to the model (Štrumbelj & Kononenko, 2014). In this way, it can be understood to what extent which features are effective in the decision-making process of the model (Sundararajan & Najmi , 2019).

SHAP method is used to determine the marginal contribution of each feature to the model's prediction by considering all possible feature combinations (Janzing, Minorics & Bloebaum, 2020). However, in practice, since computing all combinations is computationally costly, SHAP calculations are performed with different approximation methods (Chen et al., 2018). In particular, its variants such as Kernel SHAP and Tree SHAP optimize the calculations for different model types (Lundberg, Erion & Lee, 2020).

SHAP is that it can provide both local and global explanations (Kumar et al., 2020). While local explanations show the contribution of features to the model for a single data point, global explanations reveal the behavior and trends of the model in general (Lundberg & Lee, 2017). In addition, the mathematical foundations of SHAP ensure the consistency and fairness of the explanations (Sundararajan, Taly, & Yan, 2017). The SHAP method has been applied in different disciplines such as medicine (Lundberg et al., 2018), finance (Bussmann et al., 2020), energy (Dong et al., 2019), and social sciences (Štrumbelj & Kononenko, 2014). For example, in medical diagnostic models, SHAP can identify the factors that contribute most to disease risk, which helps in the development of clinical decision support systems (Tonekaboni et al., 2019). In financial models, which factors are more effective in credit risk assessments can be analyzed with SHAP (Chen et al., 2018).

In conclusion, the SHAP method is a powerful and flexible tool for improving the explainability of AI models (Gunning & Aha, 2019). It increases the transparency and reliability of models by calculating the contribution of features to the model in a fair and consistent manner (Miller, 2019). In the future, with the further development of SHAP and similar methods, it is expected that AI applications will better comply with ethical and legal requirements (Barredo (Arrieta et al., 2020).

3. LITERATURE REVIEW: STUDIES ON THE USE OF XAI IN UAVS

In the study titled "Assuring Safe and Efficient Operation of UAV Using Explainable Machine Learning" by Alharbi et al. (2023), a demand and capacity management system based on explainable machine learning was developed to ensure the safe and efficient operation of UAVs. In the study, a model was created to predict airspace capacity and determine congestion levels to enable UAVs to operate safely in the airspace. The model aimed to balance performance and explainability by combining deep learning techniques with fuzzy rule-based systems. The system helps UAVs choose the most optimal routes by analyzing air traffic. As a result of testing the developed system in a simulation environment, an increase of over 23% in airspace availability was observed. Additionally, the system's maximum capacity increase was identified as 65%, while the minimum safety gain was found to be 35%. The system's 70% explainability aids UTM (Unmanned Traffic Management) authorities in making more effective decisions (Alharbi , Petrunin & Panagiotakopoulos , 2023)

In the study by Ekramul Haque and colleagues, a solution was proposed using Zero Trust Architecture (ZTA) and Deep Learning (DL) methods to enhance UAV security. The study aimed to detect and classify UAVs using radio frequency signals. Moreover, the model's transparency and explainability were ensured using XAI tools such as SHAP and LIME. The method achieved an accuracy rate of 84.59% using RF signals. As a result, it was demonstrated that ZTA could enhance UAV security, and the integration of DL and XAI provided both security and explainability (Haque et al., 2024).

In the study by Goyal et al. (2024), an XAI-based security solution was developed to enhance the security of 5G-supported UAV networks. In the study, network traffic was monitored and analyzed using XAI to detect nodes attacking UAV networks. The research results showed that the proposed method detected attacking nodes with an accuracy rate of 98.4%.

The method performed better compared to AI/DL-based methods. The results provide a reliable security solution to enhance data transfer security in 5G-based UAV networks (Goyal et al., 2024).

In the study by Zhu et al. (2024), an XAI-based edge computing framework was developed to monitor the safety of UAVs during air surveillance in extreme conditions and to process large-scale image data. In the study, an alert system was designed using the random forest algorithm to monitor drone health and identify security concerns, and a MapReduce-based image processing module was proposed for large-scale image classification and object detection. The research increased the transparency of traditional AI systems using SHAP and provided an effective mechanism for monitoring the operational safety of drones. Experimental results show that the model achieved high accuracy (99.21%) in drone health monitoring and object detection (Zhu et al., 2024).

In the study by Javeed et al. (2024), an Intrusion Detection System (IDS) for UAVs was developed. The system was designed using advanced deep learning techniques to defend UAVs against cyber threats. The method employed the Hierarchical Attention-based Long Short-Term Memory (H-LSTM) architecture, which can model complex temporal dependencies in UAV data. The H-LSTM architecture was effective in detecting short-term anomalies and long-term deviations, and explainability was provided through the SHAP mechanism. SHAP values made it possible to understand the decisions of the IDS transparently, enabling security analysts to explain the system's decisions. Experiments were conducted using the N-BaIoT dataset, and the proposed system achieved high accuracy and low false positive rates in threat detection. The study provided an explainable and efficient cybersecurity solution to enhance UAV security (Javeed et al., 2024).

In the study by Hong and Yoo (2024), a model was developed to detect multiple attacks on the Control Area Network (CAN) protocol of UAVs. In the study, a heterogeneous model capable of detecting multiple types of attacks simultaneously was proposed, and explainability was provided using the SHAP method. The model was used to detect attacks such as DoS and GPS signal spoofing. In the study, feature importance measures were used to distinguish between attack and normal data, thereby improving the model's accuracy. Experimental results showed that the model was successful in detecting attacks with an accuracy rate of 97% (Hong & Yoo , 2024).

In the study by Souripalli et al. (2024), a model was developed for the autonomous navigation of UAVs in dense fog environments using the

Explainable Deep Reinforcement Learning (DRL) method. In the study, Twin Delayed DDPG (TD3) and Proximal Policy Optimization (PPO) algorithms were trained in the AirSim simulation environment, and an image adaptation module optimizing navigation in foggy environments was integrated. Image adaptation used a deep learning-based defogging technique to recover image details lost due to fog. The TD3 algorithm performed better than the PPO algorithm, with a high success rate (77%) and a low collision rate (16%). Additionally, the explainability of the model was ensured using SHAP and LIME. Thus, the transparency of the model's decision-making processes was increased, and it was observed that performance in foggy environments was significantly improved (Sayed , Souripalli & Chiddarwar , 2024).

4. CONCLUSION

Nowadays, the concept of AI, which is one of the concepts that entered our lives with the Industry 4.0 revolution, is. AI is frequently used in many areas such as health, agriculture, security and engineering. With the development of AI technologies , the reliability and transparency of the results obtained from AI models are very important in critically important cases such as medicine and security. In the study, a conceptual study was carried out on UAV technologies, which is one of the areas where AI technologies are frequently used. In the study, a study was carried out on the use of XAI methods, which are a frequently used method in determining the reliability and transparency of UAV and AI .

integrated use of AI technologies has become increasingly important, especially as autonomy and complexity increase. The use of AI in UAVs provides great advantages in critical areas such as autonomous navigation, environmental perception and data analysis. However, with this integration, transparency and security issues have also come to the fore. XAI plays a critical role in making the decision-making processes of AI systems understandable by addressing reliability and transparency issues. The reliability of UAVs is particularly important in terms of public safety and ethical applications. XAI Its use in UAVs increases operators' confidence in these vehicles, while also facilitating the fulfillment of regulatory requirements and the establishment of social acceptance. Therefore, XAI stands out as an important technological development that will contribute to the safer, more accountable and widespread use of UAV technologies in the future. The study was carried out using UAV and XAI methods based on academic literature, and a conceptual study was carried out for applications.

5. References

Adadi , A., & Berrada , M. (2018). Peeking inside the black-box : a survey on explainable artificial intelligence (XAI). IEEE access , 6, 52138-52160.

Alain, G., & Bengio , Y. (2017). Understanding Intermediate Layers Using Linear Classifier Probes . In Proceedings of the International Conference on Learning Representations .

Alharbi , A., Petrunin , I., & Panagiotakopoulos , D. (2023). assuring Safe and Efficient Operation of UAV Using Explainable Machine Learning. Drones , 7(5), 327.

Alvarez -Melis, D., & Jaakkola , T. S. (2018). on the Robustness of Interpretability Methods .arXiv preprint arXiv:1806.08049.

Arras , L., Horn , F., Montavon , G., Müller, KR, & Samek , W. (2017). " What is relevant in a text document ?: An interpretable machine learning approach . PloS one , 12(8), e0181142.

Bahdanau , D., Cho , K., & Bengio , Y. (2015). Neural Machine Translation by Jointly Learning to Align and Translate . In Proceedings of the International Conference on Learning Representations .

Barredo Arrieta , A., et al. (2020). Explainable artificial intelligence (XAI): Concepts , taxonomies , opportunities and challenges about responsible AI. Information Fusion , 58, 82-115.

Breiman , L., Friedman, J., Olshen , R., & Stone, C. (1984). Classification and Regression Trees. Wadsworth .

Bussmann , N., Giudici , P., Marinelli , D., & Papenbrock , J. (2020). Explainable AI in fintech risk management . Frontiers in Artificial Intelligence, 3, 26.

Campion , M., Ranganathan , P., & Faruque , S. (2018). UAV swarm communication and control architectures : a review . Journal of Unmanned Vehicle Systems , 7(2), 93-106.

Chen, C., Li , O., Tao, D., Barnett, A., Rudin , C., & Su, J. K. (2019). this Looks Like That : Deep Learning for Interpretable Image Recognition . Of Advances in Neural Information Processing Systems (pp . 8930–8941).

Chen, H., et al. (2018). explaining neural networks semantically and quantitatively .arXiv preprint arXiv:1806.04220.

Chen, J., Song , L., Wainwright , M. J., & Jordan, M. I. (2018). Learning to explain : An information-theoretic perspective on model interpretation . In International Conference on Machine Learning (pp . 883-892). PMLR.

Çorbacı, FK, & Doğan, YE (2023). An Approach to Preliminary Design of a Quadrotor Cargo UAV. International Journal of Aviation Science and Technol. , 4(02), 63-74.

Devoto , S., Macovaz , V., Mantovani , M., Soldati , M., & Furlani , S. (2020). Advantages of using UAV digital photogrammetry in the study of slow moving Coastal landslides . Remote Sensing , 12(21), 3566.

Dong, Y., et al. (2019). SHAP- based feature importance ranking for time series forecasting In 2019 IEEE International Conference on Big Data (pp . 3785-3791). IEEE.

Doshi-Velez , F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning .arXiv preprint arXiv:1702.08608.

Esteva , A., Kuprel , B., Novoa , RA, et al. (2017). Dermatologist -Level Classification of Skin Cancer with Deep Neural Networks. Nature, 542(7639), 115–118.

Floreano , D., & Wood , R. J. (2015). science , technology and the future of small autonomous drones . nature , 521(7553), 460-466.

Garreau , D., & Luxburg , U.V. (2020). explaining the explainer : A first theoretical analysis of LIME. In International Conference on Artificial Intelligence and Statistics (pp . 1287-1296). PMLR.

Gilpin , L.H., et al. (2018). explaining explanations : An overview of interpretability of machine learning In 2018 IEEE 5th International Conference on data science and advanced analysis (DSAA) (pp . 80-89). IEEE.

Goyal , SB, Rajawat , AS, Solank , RK, Patil, D., & Potgantwar , A. A Trustable Security Solutions Using XAI for 5G-Enabled UAV.

Guidotti , R., et al. (2018). A survey of methods for explaining Black box models . ACM Computing Surveys (CSUR), 51(5), 93.

Guidotti , R., Monreale , A., Ruggieri , S., Turini , F., Giannotti , F., & Pedreschi , D. (2018). A Survey of Methods for Explaining Black Box Models . ACM Computing Surveys , 51(5), 93.

Gunning , D. (2017). Explainable Artificial Intelligence (XAI). DARPA.

Gunning , D., & Aha, D. W. (2019). DARPA's explainable artificial intelligence (XAI) program. AI Magazine, 40(2), 44-58.

Gunning , D., Stefik , M., Choi , J., Miller, T., Stumpf , S., & Yang, G. Z. (2019). XAI—Explainable artificial intelligence . Science robotics , 4(37), eaay7120.

Haque , E., Hasan, K., Ahmed, I., Alam, MS, & Islam , T. (2024). Enhancing UAV Security Through Zero Trust Architecture: An Advanced Deep Learning and Explainable AI Analysis. arXiv preprint arXiv:2403.17093.

Hastie , T., Tibshirani , R., & Friedman, J. (2009). the Elements of Statistical Learning. Springer .

Hendricks , L.A., et al. (2016). generating visual explanations . In European Conference on Computer Vision (pp . 3-19). Springer .

Holzinger , A., et al. (2019). causability and explainability of artificial intelligence in medicine . Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery , 9(4), e1312.

Hong, Y. W., & Yoo , D. Y. (2024). Multiple Intrusion Detection Using Shapley Additive Explanations and a Heterogeneous Ensemble Model in an Unmanned Aerial Vehicle's Controller Area Network. Applied Sciences , 14(13), 5487.

Istiak , MA, Syeed , MM, Hossain , MS, Uddin , MF, Hasan, M., Khan , RH, & Azad, NS (2023). Adoption of Unmanned Aerial Vehicle (UAV) imagery in agricultural management : A systematic literature review . Ecological Informatics , 102305.

Janzing , D., Minorics , L., & Bloebaum , P. (2020). Feature Relevance Quantification in explainable AI: A causal problem. In International Conference on Artificial Intelligence and Statistics (pp . 2907-2916). PMLR.

Javeed , D., Gao , T., Kumar, P., Shoukat , S., Ahmad , I., & Kumar, R. (2024, June). An Intelligent and Interpretable Intrusion Detection System for Unmanned Aerial Vehicles . In ICC 2024-IEEE International Conference on Communications (pp . 1951-1956). IEEE.

Kawaguchi , K. (2016). without Deep Learning Poor Local Minimum . In Advances in Neural Information Processing Systems (pp . 586–594).

Keneni, B. M. (2018). Evolving Rule Based Explainable Artificial Intelligence for Decision Support System of Unmanned Aerial Vehicles (Master's thesis , University of Toledo).

Kim, J., & Canny , J. (2017). Interpretable Learning for Self- Driving Cars by Visualizing Causal Attention . In Proceedings of the IEEE International Conference on Computer Vision (pp . 2942–2950).

Kumar, I.E., Venkatasubramanian , S., Scheidegger , C., & Friedler , S.A. (2020). problems with Shapley-value-based Explanations as featured importance measures . In International Conference on Machine Learning (pp . 5491-5500). PMLR.

LeCun , Y., Bengio , Y., & Hinton , G. (2015). Deep learning Nature, 521(7553), 436-444.

Li , O., Liu , H., Chen, C., & Rudin , C. (2018). Deep Learning for Case- Based Reasoning Through Prototypes : A Neural Network That Explains Its Predictions . In Proceedings of the AAAI Conference on Artificial Intelligence (pp . 3530–3537).

Lipton, Z. C. (2016). The Mythos of Model Interpretability . arXiv preprint arXiv:1606.03490.

Lipton, Z. C. (2018). The mythos of model interpretability : In machine learning , the concept of interpretability is both important and slippery . Queue, 16(3), 31-57.

Lundberg , S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions . Of Advances in Neural Information Processing Systems (pp . 4765-4774).

Lundberg , S. M., Erion , G., & Lee, S. I. (2020). From local explanations to global understanding with explainable AI for trees . Nature Machine Intelligence, 2(1), 56-67.

Lundberg , S.M., et al. (2018). Explainable machine learning predictions for the Prevention of hypoxemia during surgery . Nature Biomedical Engineering , 2(10), 749-760.

Maas , AL, Hannun , AY, & Ng, AY (2013). Rectifier Nonlinearities Improve Neural Network Acoustics Models . In Proceedings of the 30th International Conference on Machine Learning.

Miller, T. (2019). Explanation in artificial intelligence : Insights from the social sciences . Artificial Intelligence, 267, 1-38.

Molnar , C. (2019). Interpretable Machine Learning. https://christophm.github.io/interpretable-ml-book/

Molnar , C. (2019). Interpretable Machine Learning. leanpub .

Montavon , G., Samek , W., & Müller, K.-R. (2018). methods for Interpreting and Understanding Deep Neural Networks.Digital Signal Processing , 73, 1–15.

Muhammad Kamarulzaman , AM, Wan Mohd Jaafar , W.S., Mohd Said, MN, Saad , SNM, & Mohan , M. (2023). UAV Implementations in Urban Planning and Related Sectors of Rapidly Developing Nations: A Review and Future Perspectives for Malaysia . Remote Sensing , 15(11), 2845.

Murthy , S. K. (1998). Automatic Construction of Decision Trees from Data: A Multi- Disciplinary Survey . Data Mining and Knowledge Discovery , 2(4), 345–389.

Pandey , B.K., Pandey , D., & Sahani , S.K. (2024). autopilot control unmanned aerial vehicle system for sewage defect Detection using deep learning . Engineering Reports , e12852.

PK, FA (1984). What is Artificial Intelligence? Success is no accident It is hard work , perseverance , learning , studying , sacrifice and most of all , love of what You are doing or learning to do, 65.

Quinlan , J. R. (1986). Induction of Decision Trees. Machine Learning, 1(1), 81–106.

Ribeiro , M.T., Singh, S., & Guestrin , C. (2016). “ Why I should trust “you ?”: Explaining the predictions of any classifier . In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp . 1135-1144). ACM.

Ribeiro , M.T., Singh, S., & Guestrin , C. (2018). Anchors : High- precision model -agnostic explanations . In Proceedings of the AAAI Conference on Artificial Intelligence (Vol . 32, No. 1).

Rokach , L., & Maimon , O. (2005). Decision Trees. In Data Mining and Knowledge Discovery Handbook (pp . 165–192). Springer .

Roth , A. E. (1988). The Shapley value : Essays in honor of Lloyd S. Shapley . CambridgeUniversity Press .

Rudin , C. (2019). Stop Explaining Black Box Machine Learning Models for High -Stakes Decisions and Use Interpretable Models Instead . Nature Machine Intelligence, 1(5), 206–215.

Russell, S., & Norvig , P. (2010). Artificial Intelligence: A Modern Approach . prentice Hall .

Samek , W., Wiegand , T., & Müller, K.-R. (2017). Explainable Artificial Intelligence: Understanding , Visualizing and Interpreting Deep Learning Models .arXiv preprint arXiv:1708.08296.

Santos, NP, Rodrigues , VB, Pinto , AB, & Damas , B. (2023, April). automatic Detection of Civilians and Military Personnel in Reconnaissance Missions using a UAV. In 2023 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC) (pp . 157-162). IEEE.

Sayed , L.J., Souripalli , PK, & Chiddarwar , SS (2024, June). autonomous Navigation of Drones Using Explainable Deep Reinforcement Learning in Foggy Environment. In 2024 IEEE Students Conference on Engineering and Systems (SCES) (pp . 1-6). IEEE.

Selvaraju , R.R., et al. (2017). Grad -CAM: Visual explanations from deep networks via gradient-based localization . In Proceedings of the IEEE International Conference on Computer Vision (pp . 618-626).

Shadiev , R., & Yi , S. (2023). A systematic review of UAV applications to education Interactive Learning Environments , 31(10), 6165-6194.

Shapley , L. S. (1953). A value for n- person games . In Contributions to the Theory of Games (pp . 307-317). Princeton University Press .

Štrumbelj , E., & Kononenko , I. (2014). explaining prediction Models and individual predictions with feature contributions . Knowledge and Information Systems , 41(3), 647-665.

Sundararajan , M., & Najmi , A. (2019). the many Shapley values for model explanation . In International Conference on Machine Learning (pp . 9269-9278). PMLR.

Sundararajan , M., Taly , A., & Yan, Q. (2017). Axiomatic attribution for deep networks . In International Conference on Machine Learning (pp . 3319-3328). PMLR.

Talwandi , N.S., Thakur , P., & Khare , S. (2024, February). An Automatic Navigation System for New Technical Advanced Drones for Different Alpplications . In 2024 IEEE International Conference on Computing, Power and Communication Technologies (IC2PCT) (Vol . 5, pp . 736–741). IEEE.

Thiels , C.A., Aho , J.M., Zietlow , S.P., & Jenkins, D.H. (2015). Use of unmanned aerial Vehicles for medical product transport.Air medical Journal , 34(2), 104-108.

Tjoa , E., & Guan , C. (2020). A survey on explainable artificial intelligence (XAI): Towards medical IEEE Transactions on Neural Networks and Learning Systems , 32(11), 4793-4813.

Tonekaboni , S., et al. (2019). What clinicians want : Contextualizing explainable machine learning for clinical end use . In Machine Learning for Healthcare Conference (pp . 359-380). PMLR.

Wachter , S., Mittelstadt , B., & Floridi , L. (2017). Why a right to explanation of automated decision making Does not exist in the General Data Protection Regulation . International Data Privacy Law , 7(2), 76-99.

Witten , IH, Frank, E., Hall , M.A., & Pal, C.J. (2016). Data Mining : Practical Machine Learning Tools and Techniques . Morgan Kaufmann.

Zhang, Q. S., & Zhu, S. C. (2018). Visual Interpretability for Deep Learning: A Survey . Frontiers of Information Technology & Electronic Engineering , 19(1), 27–39.

Zhu, H., Demirbaga , U., Aujla , GS, Shi , L., & Zhang, P. (2024). Explainable Edge AI Framework for IoD-Assisted Aerial Surveillance in Extreme Scenarios . IEEE Internet of Things Journal .