

XAIoT: A Conceptual Model for XAI Solutions on IoT

Humberto Parejas-Llanovarced , Marta Caro-Martínez , Juan A. Recio-Garcia  *

Department of Software Engineering and Artificial Intelligence, Universidad Complutense de Madrid, Madrid (Spain)

* Corresponding author: hparejas@ucm.es (H. Parejas-Llanovarced), martcaro@ucm.es (M. Caro-Martínez), jareciog@fdi.ucm.es (J. A. Recio-Garcia).

Received 24 May 2024 | Accepted 4 August 2025 | Published 31 August 2026



ABSTRACT

Artificial intelligence (AI) systems are part of our current world, helping us to perform a wide variety of tasks. In parallel, eXplainable Artificial Intelligence (XAI), which obtains explanations to help users understand AI systems, has increased in significance in the last few years. This relevance comes from users' necessity to trust new complex AI models. However, AI is not confined to computers but is also applied to other devices to use current internet connectivity to make our lives easier. The already well-known Internet of Things (IoT) refers to this digitally connected universe of smart devices. Hence, XAI should also be applied to the IoT, although there is a gap in the literature concerning conceptualizing explanation methods applied to AI models on the IoT. In this work, we propose a conceptual model named XAIoT (eXplainable Artificial Intelligence of Things) that structures and formalizes the main features of such systems. The ultimate goal is to provide a framework that guides the selection of the most suitable XAI solution for a given intelligent IoT application. This conceptual model is validated by classifying up to 100 approaches obtained from a systematic review of the state-of-the-art and formalized in an ontology to guide XAIoT system design.

KEYWORDS

Explainable Artificial Intelligence, Explanation Methods, Internet Of Things, Ontology, Systematic Review.

DOI: 10.9781/ijimai.2026.6503
eLocator: e6503

I. INTRODUCTION

SOME decades ago, intelligent systems emerged to help us complete and carry out many tasks, such as making movie recommendations, detecting faces, and translating languages [1]. Artificial intelligence (AI) solves tasks in critical domains, such as medicine or defense, where understanding why or how intelligent systems work is especially important for users [2]. AI models can be transparent, such as tree-based methods, where users can easily understand the decision-making process. However, other popular AI techniques, such as deep learning models, are black boxes that do show users how a prediction was obtained [3]. That is why we have seen high interest in eXplainable Artificial Intelligence (XAI) during the last few years. XAI is an AI subfield in which the main goal is to explain the underlying behavior of AI systems to the final users [2]. One of the main XAI challenges is to create AI systems that can provide clear and understandable explanations for their decisions and actions, increasing users' trust in those systems and making them more prone to accept their predictions [4].

The analysis presented in this paper shows that during the last four years, research papers on XAI have increased from 177 documents in 2018 to 2702 in 2022¹ prove the increase in interest in explainability in several domains that are looking for "human understanding" to

achieve a higher acceptance from users, comply with regulations and, most importantly, with ethics (trust, transparency, fairness, robustness, privacy).

In parallel to AI, the Internet of Things (IoT) is another rising field that transforms devices into interconnected virtual objects having the capacity to be smart, i.e., obtaining knowledge, learning about it, and acting in consequence [5]. In this way, the IoT has a very tight relation to AI, as AI models are being applied to the data collected from IoT devices. The impact of the massive data generation and enhancement of communications and technologies is increasing the demand for AI models to make intelligent decisions in the different domains where IoT is being applied. The use of AI to increase the usefulness of data collected by IoT sensors is rapidly gaining importance [6]–[8]. Evidently, intelligent IoT applications have the same requirements regarding explainability as any other standalone AI system. Therefore, including XAI in the IoT can improve these devices' utility if users can better understand how things are learning and acting [9]. We denote the resulting joint field between XAI and IoT as XAIoT (eXplainable Artificial Intelligence of Things).

Moreover, likewise the research papers on XAI in all the domains, the analysis shows that between 2018 and 2022 the publications on the applicability of XAI to the IoT field have increased from 2 to 52¹ respectively, which proves the rise in attention in this domain. Therefore, there are already many XAIoT solutions to apply in

¹ Scopus. Further details in Section II.

Please cite this article as:

H. Parejas-Llanovarced, M. Caro-Martínez, J. A. Recio-Garcia. XAIoT: A Conceptual Model for XAI Solutions on IoT, *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 10, no. 1, e6503, 2026, <http://doi.org/10.9781/ijimai.2026.6503>

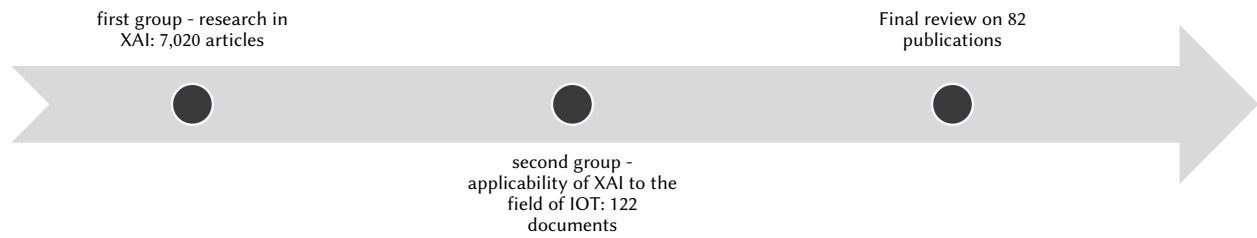


Fig. 1. Quantitative analysis of the literature review on IoT in XAI: number of results found in each search.

different situations for different problems. As a consequence, picking the best XAIoT solution for a specific situation has become a difficult decision-making task. Some works in the literature have discussed general features regarding explainability in IoT [9], [10]. However, a more complete, structured, and formal guide to design XAIoT systems is required to allow developers to choose the best explanation solution for their intelligent IoT applications. In the literature, we can find some works that establish a formalized guide to develop XAI solutions [11], [12], but not specifically for the IoT field. To overcome this research gap, this paper presents a systematic review, analyzing, structuring, and combining existing surveys and works on XAI and IoT. The main result is a conceptual model describing the most relevant explanation requirements of an XAIoT system to guide its successful design. This *conceptual model* is proposed as a theoretical framework where the features of an XAIoT system are defined and delimited. Furthermore, this model is validated by classifying and characterizing 100 different XAIoT solutions. Finally, we built an ontology, using the OWL-DL language, that allows making semantic queries to the model and obtaining the most suitable explanation solution for a given domain, task, and AI model. These contributions will help designers bridge the application of XAI in the IoT.

This paper is organized as follows. Section II reviews the literature in the XAI and IoT fields and presents its methodology. Section III describes the conceptual model resulting from analyzing the works found by the literature review. The conceptual model implementation is presented in Section IV, and its validation is described in Section V. Finally, Section VI provides the most relevant conclusions and proposes future work.

II. SYSTEMATIC LITERATURE REVIEW

To build our conceptual model, we conducted a systematic review analyzing relevant publications. This literature review on IoT in XAI was organized into two groups. In the first group, we studied the current research in XAI quantitatively, whereas the second group focused on existing works on the applicability of XAI to the IoT field, which was later analyzed in-depth to generate our conceptual model. Fig. 1 shows the total number of publications found in this analysis in this research and chosen to generate the conceptual model. Our initial goal is to analytically check whether there is a correlation between the evolution of the global research on XAI and the specific research on XAIoT.

For both groups, we defined the search criteria regarding terms and years. The database used for the review is SCOPUS². Next, we present some quantitative figures of the results.

First, we focused our search on the XAI literature from the last four years (i.e., 2018-2022), where we searched publications that include the terms “explainable” or “explainability” and “artificial intelligence”,

“machine learning” and “deep learning”³. As a result, we obtained 7,020 research documents, 71% of which were published in the last two years. This increase in publications shows the community’s interest in the XAI field, specifically in machine learning and deep learning. Moreover, 74% of the documents are enclosed in other related fields, such as computer science (38%), engineering (14%), mathematics (11%), medicine (6%), and decision support (5%). See further details in Fig. 2. Regarding the affiliation of the papers, the United States has the largest number of publications (16%), followed by Germany (8%) and China, and the United Kingdom (7% both). Next, we added the terms “Internet of Things” or “IoT” to the previous search to identify the publications concerning XAI applied to IoT⁴. The term “Internet of Things (IoT)” was used to identify the documents related to XAI and IoT. We found a total of 122 documents, 76% of which were published in 2021 and 2022. Regarding the main areas related to XAI and IoT, we can conclude that computer science is the most representative (42% of the papers belongs to it), followed by engineering (18% of the papers) and mathematics (14% of the papers).

Fig. 2 shows the quantitative analysis of both the generic XAI and the specific XAIoT research areas. We can clearly see a correlation between the interest in both areas.

Regarding the geographical distribution of XAIoT papers, the United States is the country with the largest percentage of publications in this field (10%), followed by South Korea (5%) and the United Kingdom, Australia, China, and India (fewer than 5%). The analysis of the publication dates also shows that 41% were published in 2020, 22% in 2021, 18% in 2019, 13% in 2022, and 5% in 2018. We present the details in . As we identified before, we can observe the progressive increment of interest in XAI and IoT in the last few years.

Another relevant feature of the review is that it is not limited to specific papers in XAI and IoT; it also includes publications presenting surveys or taxonomies in these areas. Thus, we considered 16 additional survey papers [1],[4], [9], [10], [13]–[24]. Due to the revision of these papers, further references were evaluated (46 papers in total).

After removing irrelevant references, the final revision set contained 82 publications that were analyzed to define and validate our conceptual model and the proposed XAIoT ontology.

III. MODEL DESCRIPTION

The XAIoT model is a conceptual framework to structure and formalize the main features and vocabulary of XAI systems in IoT applications. To find these terms and relationships and generate this model, we studied the current XAI methods taxonomies [4], [13]–[23], [25]–[27], and we carried out a literature review of the papers

³ We used the following query: QUERY:TITLE-ABS-KEY ((explainable OR explainability) AND ((artificial AND intelligence) OR (machine AND learning) OR (deep AND learning))) AND PUBYEAR > 2017 AND PUBYEAR < 2023.

⁴ In this case, we used the query: QUERY: TITLE-ABS-KEY (IoT AND (explainable OR explainability) AND ((artificial AND intelligence) OR (machine AND learning) OR (deep AND learning))) AND PUBYEAR > 2017 AND PUBYEAR < 2023.

² <https://www.scopus.com/home.uri>

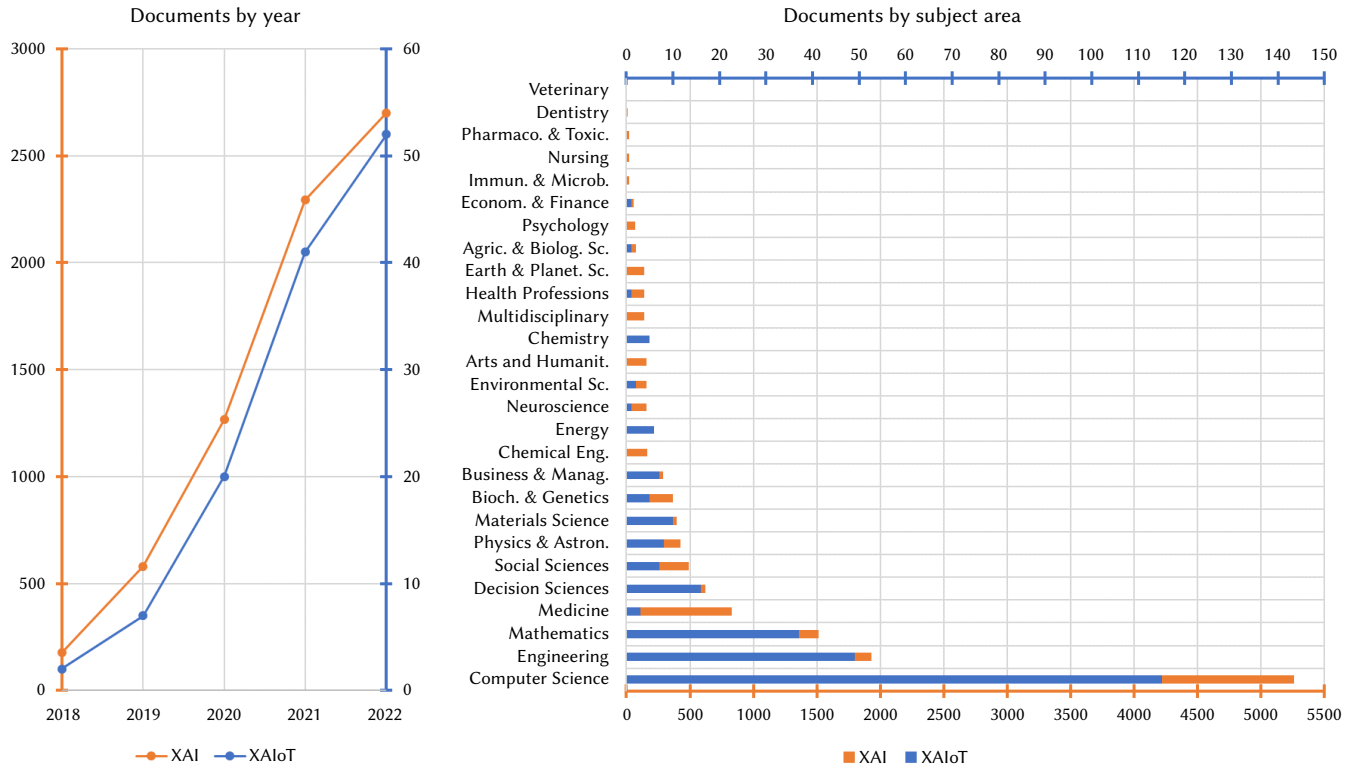


Fig. 2. Quantitative analysis of the systematic review on explainable AI and IoT. The chart on the left (Documents by year) shows the number of publications resulting from our search between 2018 and 2022, a total of 7,022 documents that include terms related to XAI, and 122 documents that in addition to XAI included the term IoT. While the chart on the right (Documents by subject area) shows the different subject areas in which the research was conducted, in some cases an article has been categorized into more than one subject area. Source data available in Table IV.

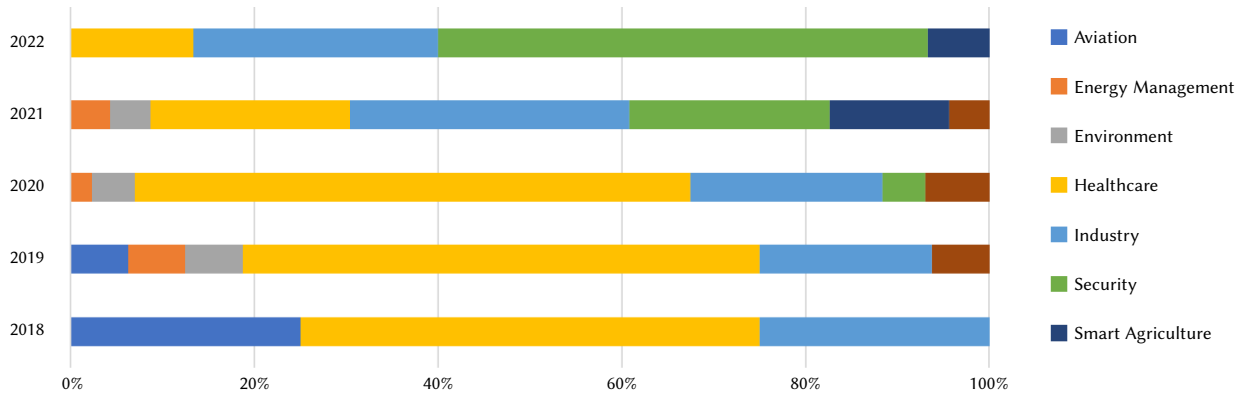


Fig. 3. Number of papers in XAIoT published per year and domain, over the 82 out of 122 XAIoT articles chosen for their relevance to perform this research. Source data available in Table V.

found with the systematic process described in the previous section. The resultant model defines the predominant domains, tasks, and XAI techniques used in the IoT field. From this analysis, we propose the conceptual model represented in Fig. 4, where the central concept is the *AIoT task* performed and explained. *AIoT tasks* are grouped by their application *domain* and use an *AI model*. The *AI model* solves a type of problem and is explained by an *XAI method*. The model also describes an *XAI method* through some properties well known in the literature: *concurrence*, *scope*, *input format*, *output format*, and *portability*.

The following sections detail these concepts and their instances (types of individuals we found for each concept).

A. Domains and Tasks

This subsection presents an exhaustive overview of *AIoT tasks*. To achieve this, we review the studies found in the literature and group them by their application domain. Eight *XAIoT domains* (i.e., instances)

have been identified following the terminology found in the reviews by [1], [10], [24], [28]: *aviation*, *energy management*, *environment*, *health care*, *industry*, *security*, *smart agriculture*, and *transportation*.

From the reviewed papers, we propose sixteen *XAIoT tasks* (instances) previously grouped by the identified domains (including their prevalence percentage): *anomaly detection* (7%), *assistance* (1%), *automated maneuvering* (1%), *autonomous processes and robotics* (2%), *business management* (5%), *cyberattack detection* (1%), *decision support* (43%), *facial recognition* (1%), *image processing* (11%), *improve process quality* (1%), *internet of behavior* (1%), *intrusion detection* (6%), *modeling* (1%), *predictive maintenance* (7%), *recommendation* (5%), and *risk prediction* (5%).

Fig. 5 shows the distribution of these *AIoT tasks* grouped by their *AIoT Domain*. The largest number of *AIoT tasks* are from *health care* with 41%, 23% from *industry*, 15% from *security*, 6% from *transportation*,

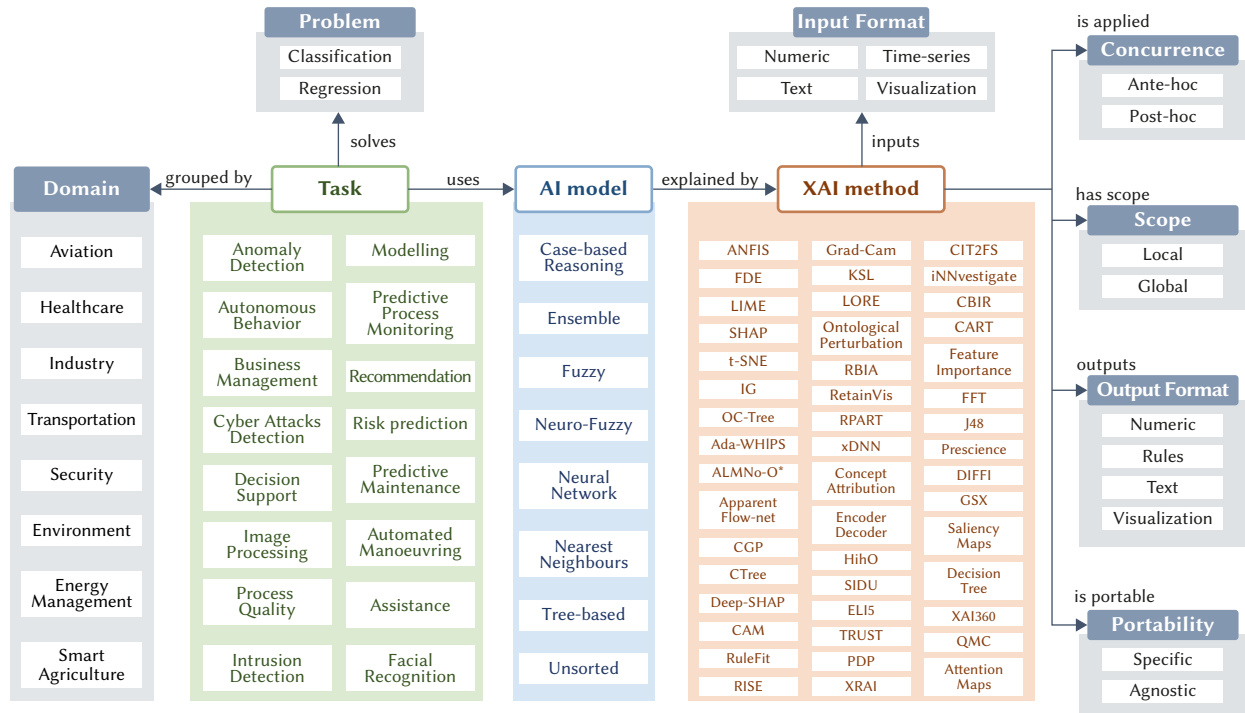


Fig. 4. Overview of the XAIoT conceptual model.

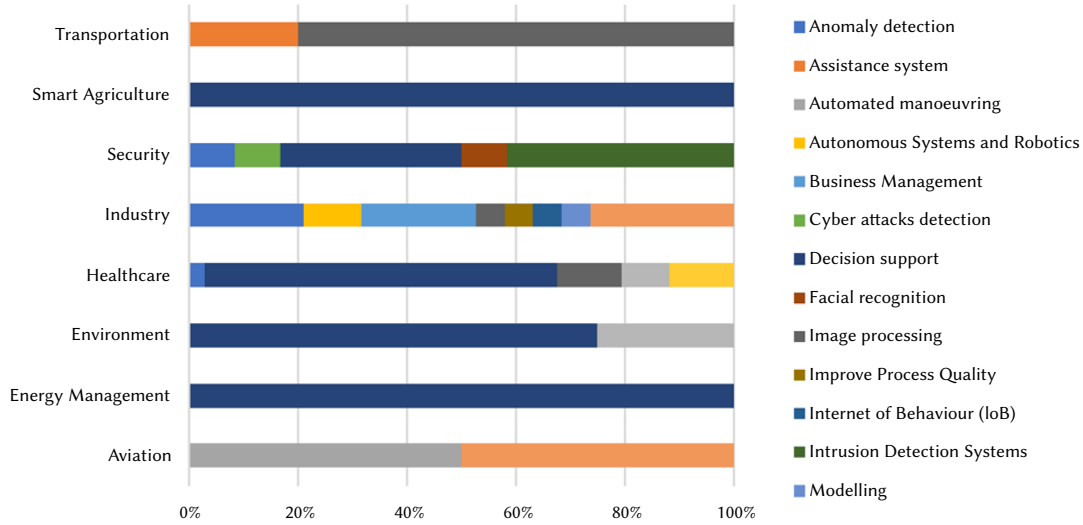


Fig. 5. Normalized distribution of AI Task by IoT Domain. The chart shows the number of articles published by each of the 16 AIoT tasks and each of the 8 IoT domains identified in our research. Source data available in Table VI.

4% from *energy management* and *smart agriculture* each, and *aviation* with 2%. The *health care* domain has been further developed from the point of view of XAIoT than other domains [1]. Moreover, the *Industry* domain is relevant and mainly focuses on predictive maintenance and anomaly detection tasks. Globally, the *decision support task* is the most common task found in the XAIoT literature.

Next, we review these *AIoT tasks* from the point of view of their application *AIoT domain*:

Aviation. Different intelligent applications applied, for example, on *automated maneuvering tasks*, are emerging from unmanned aerial vehicle (UAV) development and are used for observation and tactical planning. Here, machine learning models are used to give the logic behind UAV decisions. In this field, we can find the XAI proposal by [29] based on a visual platform in the format of if-then rules derived from the Sugeno-type fuzzy inference model.

Moreover, for *predictive maintenance tasks*, failure diagnosis explainability (FDE) is being used as an XAI method to provide transparency and interpretability of the assessed diagnosis for automated failure diagnosis in aviation maintenance. The assessed diagnosis is explained by checking whether a new failure matches the expected values of a certain diagnosis (class) [30].

Energy Management. AI-powered energy decision systems used in smart cities, smart grids, and smart home applications provide decision-makers with critical predictions about energy use and emissions [10].

In this domain, *decision support* is the principal task where XAI methods have been applied. The predominant models to which the XAI methods (i.e., local interpretable model-agnostic explanations (LIME), SHapley Additive ex-Planations (SHAP), and t-distributed stochastic neighbor embedding (t-SNE)) have been applied are

neural networks and tree-based model-agnostic models, resolving classification problems.

Environment. Environmental applications in the IoT focus on many issues and problems in areas such as smart cities, air monitoring, water management, and climate monitoring. To solve these tasks, mostly AI models with high complexity and low interpretability are used [10].

The main tasks identified in this domain are *decision support tasks* and *recommendation tasks*, which have been implemented principally by neural network models (75%) and ensemble models (25%). In addition, the SHAP XAI method has been applied in 50% of the papers reviewed, followed by LIME and integrated gradients with 25% each.

Health care. AI decisions in health care are critical, and XAI plays an essential role in health care. Decision-making in health care affects people directly, making it very difficult to compensate for their negative consequences [10].

The task with the highest percentage of IoT applications is the *decision support task*, which appears in 65% of the papers studied, followed by *risk prediction* and *image processing* (12% of the papers each), *recommendation task* (9%), and *anomaly detection task* (3%). Among the thirty-four papers reviewed, the three main XAI methods applied in the health care IoT domain are SHAP with 14% and LIME and GRAD-CAM with 11% each.

Industry. AI systems in the Industrial Internet of Things (IIoT) enable machines to perform self-monitoring, interpretation, diagnosis, and analysis autonomously in production lines and manufacturing processes or logistics [10].

Studying the state-of-the-art, we identified three main groups of tasks applied to *industry* to be included in our conceptual model. The task with the highest percentage of experimentation is *predictive maintenance* (26% of the papers explained this task), followed by *anomaly detection* and *business management* (21% of the papers). This analysis shows that the industry is leveraging AI to enhance operations that avoid damage to their infrastructure and improve the quality of the products by identifying production anomalies. Among the twenty-four papers reviewed, the two main XAI methods applied in the IIoT domain are SHAP (17% of the papers) and feature importance (13%), followed by LIME, GRAD-CAM, ELI5, DeepSHAP, and CAM (8% each).

Security. Security and privacy are remarkable domains in the IoT. AI algorithms and models are widely used for intrusion detection, anomalous traffic detection, authentication, and malware detection in IoT security systems [10].

The most significant tasks, which we included in our XAIoT conceptual model, are the *intrusion detection task* (42% of the papers that we studied explained this type of task, which shows the criticality of making the monitoring systems that detect suspicious activities and generate alerts transparent) and the *decision support task* (33% of the papers). The three main XAI methods applied in the *security* XAIoT domain among the fifteen papers reviewed are SHAP (27% of the papers), LIME (20% of the papers), and decision trees (13%).

Smart Agriculture. Smart agriculture opens the door to sustainable and responsible agriculture by making better management decisions with AI-powered decision support systems [10]. Accordingly, the predominant task in this domain is the *decision support task*. The XAI methods applied in this domain are decision trees, Grad-CAM, LIME, and PDP.

Transportation. An autonomous vehicle provides capabilities such as sensing the environment without human interventions, identifying optimized routes from source to destination, providing human-free decision capabilities, decreasing accidents, and enhancing mobility in traffic and accident situations [31].

The most developed task in this domain is the *image processing task* (80% of the papers studied this task), followed by *assistance* (20% of the papers). This shows the importance of the treatment of images in this sector. Here, the XAI methods applied are attention maps (applied in 40% of the cases) and randomized input sampling for explanation (RISE), SHAP, and XRAI (in 20% of the cases each).

B. AI Models

The *AI task* concept is related not only to the *domain* but to two more concepts: the *problem* that the *AI task* has to solve and the *AI model* used to solve it. Regarding the *problems* to solve, we defined two instances: the well-known *classification* and *regression* problems. There are other problems that AI can solve, as demonstrated in the literature, for example, clustering [32], link profiling [33], similarity matching [34], and causal modeling [35]. However, we have not found these problems applied to XAIoT in the papers we studied. Therefore, these problems are not included in the *problem* concept in our conceptual model. Regarding the *AI models*, we also defined several instances considering the common models used in the literature. The most used *AI models* in IoT solutions are *neural networks* (NNs) (in 50% of the papers studied), *ensemble models* (EM) (in 13%), and *tree-based models* (TB) (in 11%). These are the instances of the *AI model* concept. We show their analysis in Table I. In this analysis, we can mention that within the NN group, the most commonly used models are convolutional neural networks (CNNs) and long/short-term memory (LSTM). For the EMs, the most applied model is random forest (RF), and with regard to TB models, the most experimented with is decision tree (DT).

Although we do not include the AI model architecture concept within the conceptual model, we want to briefly discuss this aspect, as we consider it relevant in the IoT field since distributed systems are related to the real future of IoT applications [102]. As with any intelligent system, explanations should be included in ML/AI models that are trained in a distributed way, such as federated learning. Moreover, according to the literature, distributed systems seem to be applied in health care and security domains or other typical IoT domains more frequently. These domains are very critical since users need to trust intelligent systems to be sure about the proper working of the system [103], [104]. The main feature of a distributed system is that the data can be stored in multiple machines where each machine trains the model with that part of the data, sharing the knowledge between them. Then, we need to explain the behavior of the AI model for each distribution, and at least a specific number of distributions should be trustworthy to consider the system trustworthy as a whole [104]. Furthermore, another important factor to consider is not only to obtain explanations for each machine but also to obtain all those explanations in a proper delivery time [105]. We find an example in smart cities and autonomous cars, traffic control systems, or unmanned aerial vehicles (UAVs), as noted by Azetta et al. (2020). This kind of system obtains a wide variety of data as input and needs to process them in a highly dynamic and real-time fashion. It is also critical to obtain explanations that way. Finally, to obtain explanations for distributed architectures, we need to consider their scalability, latency, topology, synchronization, convergence speed, etc. Therefore, we can use these features to justify the system's behavior [106].

C. XAI Methods

Forty-seven *XAI methods* were distinguished from the reviewed papers, which were used to explain the *AI models* applied to the *AIoT domains*. These methods, which are the instances of the *XAI method*

TABLE I. AI MODELS FOUND IN THE LITERATURE REVIEW. CBR (CASE BASE REASONING), EM (ENSEMBLE MODEL), FM (FUZZY MODEL), NFM (NEURO-FUZZY MODEL), NN (NEURONAL NETWORK), NNN (NEAREST NEIGHBOURS MODEL), TB (TREE-BASED MODEL), UM (UNSORTED MODEL)

AI Model	AI Method	Count	References
CBR	CBR-kNN, CBR-WkNN, CBR-PRVC (Pattern Recognition, Validation and Contextualisation) Methodology	2	[36], [37].
EM	Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), Gradient Boosting Machines (GBM), Isolation Forest (IF), Random Forest (RF), CatBoost (based on Gradient boosting library), Deep Feed Forward (DFF)	13	[30], [38], [39], [40], [41], [42], [43], [44], [45], [46], [47], [48], [49].
FM	Constrained Interval Type-2 Fuzzy System (CIT2FS), Fuzzy inference system	2	[50], [51].
NFM	Adaptive Network-Based Fuzzy Inference System (ANFIS), Zero-Order Autonomous Learning Multiple-Model (ALMMo-0)	3	[29], [52], [53].
NN	ApparentFlow-net, Convolutional Neural Network(CNN), Deep Neural Network (DNN), Neuronal Network (NN), Deep Reinforcement Learning (DRL), Global-Local Capsule Networks (GLCapsNet), Knowledge-Shot Learning (KSL), Long/Short-Term Memory (LSTM), Multilayer Perceptrons (MLP), Recurrent Neural Network (RNN), Region-Based CNN (RCNN), RestNet, Temporal Convolutional Network (TCN), VGG-16, YOLO, Bidirectional GRU Model (BGRU), Autoencoder Model	44	[54], [55], [56], [57], [58], [59], [60], [61], [62], [63], [64], [65], [66], [67], [68], [69], [70], [71], [72], [73], [44], [74], [75], [76], [77], [78], [28], [79], [51], [80], [81], [82], [83], [84], [85], [86], [87], [88], [89], [90], [91], [92], [93], [94].
NNN	k-Nearest Neighbours (kNN)	1	[95].
TB	Classification and Regression Tree (CART), Conditional Inference Tree (CTree), Decision Tree (DT), Fast and Frugal Trees (FFTs), J48, One-Class Tree (OCTree), Recursive Partitioning and Regression Trees (RPART)	7	[96], [60], [97], [98], [99], [100], [88].
UM	Cartesian Genetic Programming (CGP)	1	[101].

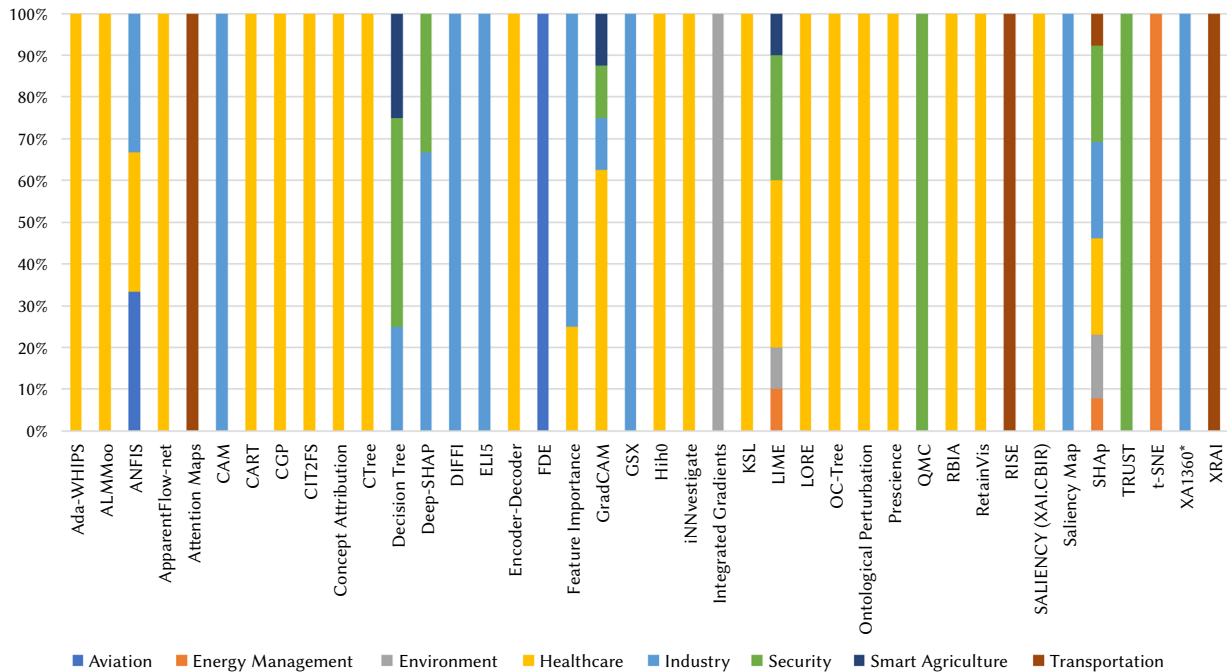


Fig. 6. Normalized distribution of XAI method by Domain. The chart shows that, of the total of 82 selected publications, a total of 101 experiments carried out with 47 different XAI methods were identified. Source data available in Table VII.

concept, are presented in Fig. 6. The three most popular XAI methods to explain the models are SHAP (in 18% of the papers studied), LIME (13%), and Grad-CAM (9%).

We need to characterize the *XAI methods* to define how and when they can be applied to explain an *AI model*. These properties are, according to our literature review (see Fig. 4): *concurrency*, *scope*, *output format* and *portability* [11]. Next, we depict these properties.

Concurrency. The *concurrency* property highlights the transparency of the *AI model* that is explained regarding the *XAI method* [107]. The *XAI method* is *ante-hoc* if the *AI model* is transparent (it is a white box) and can access the underlying AI reasoning process to create explanations. It can also be *ante-hoc* if the *AI model* creates the explanations or if it is explainable by design. When the *AI model* is not transparent (it is a black box) for the *XAI method*,

then the *XAI method* has to be independent of the *AI model*. In this case, our *XAI method* is *post hoc* and it is a surrogate interpretable model, so it approximates the behavior of the black box.

Scope. The *scope* considers the amount of knowledge explained and the portion of data utilized by the *XAI method* [108]. When an *XAI method* is *global*, it explains the whole *AI model*. It can also mean that the *XAI method* uses all the knowledge available to create explanations. If the *XAI method* is *local*, then it only explains a specific prediction made by the *AI model*. In this way, the method only uses specific dataset instances to generate the explanation.

Output format. We can also classify an *XAI method* according to the *output format* of the generated explanation [109]. It can be *numerical* when the method displays the explanation as a set of numerical data understandable and usable for users. Explanations

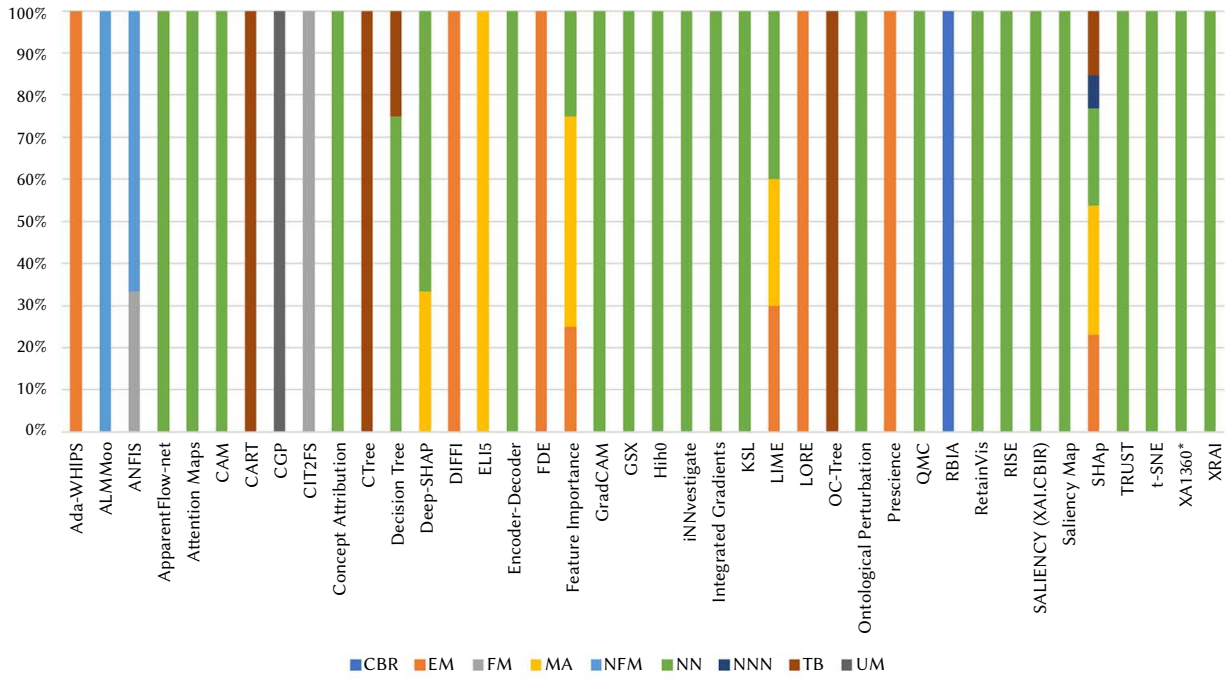


Fig. 7. Normalized distribution of XAI Method by AI model. The chart shows that, of the total of 82 selected publications, a total of 101 experiments carried out with 47 different XAI methods were identified. Source data available in Table VIII.

are sometimes shown as *rules*, for example, when they show the reasoning process performed by the *AI model* (that needs to be ante-hoc) [110], or when the *AI model* can also be treated as a black box from which symbolic rules can be extracted (post hoc XAI method) [111]. Finally, we find *text*, *chart* and *visualization*, which are the most common output formats [11]. The text format refers to an explanation displayed in natural language [112]. Chart format alludes to the visual data representation of the explanation. The visualization format mainly includes explanations presented through images or video, although it can include more interactive alternatives such as virtual reality [11], [113].

Input format. The input format for an XAI method is one of the key attributes that we can use to define this type of method. We cannot apply a specific XAI method if we do not have a dataset type accepted by that method. The types we can define here are *visualization*, (*tabular text*), *numeric (tabular numeric)*, and *time series*. The *visualization* format is the one that we have when the XAI method accepts images as input. *Numeric* refers to tabular data, and *text* refers to any type of writing [4]. *Time series* is also an input format type that we found in the IoT field, although it was less common [114]. It refers to a set of data collected over a range of times. Finally, as we had for the output format, here we also have *rules*.

Portability. The *portability* property determines to which type of *AI models* the *XAI method* can be applied to [115]. If the *XAI method* can explain any *AI model*, then we define our *XAI method* as *model-agnostic*. If the *XAI method* is *model-specific*, then the *XAI method* can only generate explanations for a specific *AI model*.

Once we describe the properties of *XAI methods*, we can analyze their relationship with *AI models*. The distribution of *XAI methods* by *AI models* in the IoT field is presented in Fig. 7. The *AI model* with more implementations in the reviewed literature is neural networks, and the *XAI methods* used to explain this model are Grad-CAM (18% of the cases), LIME (10%), and SHAP (8%). Additionally, the second most predominant *AI model* is EM, and the two main *XAI methods* applied to it are SHAP (29%) and LIME (21%). This outcome clearly shows the

versatility of LIME and SHAP as explanation methods, both solving classification problems. Both methods are model-agnostic and post hoc. LIME is always local, but SHAP can be local or global.

D. Model Relationships

The previous sections describe the properties and relationships between the main concepts of the XAIoT conceptual model. However, we need to identify the relationships between their corresponding instances (possible values) to provide a framework that guides the selection of the most suitable XAI solution for a given IoT application. These relationships between the model instances are visually represented in Fig. 8.

The relationships that we defined in our conceptual model are: **group by** and **groups** (the relationships between *domain* and *AI task*), **solves** (between *problem* and *AI task*), **uses** and **isUsedBy** (between *AI task* and *AI model*), **explained by** and **explains** (between *XAI method* and *AI model*), **is applied** (between *XAI method* and *portability*), **has scope** (between *XAI method* and *scope*), **outputs** (between *XAI method* and *output format*), and **is portable** (between *XAI method* and *portability*). These relations are represented in Table II.

TABLE II. RELATIONSHIPS BETWEEN CONCEPTS INCLUDED IN THE CONCEPTUAL MODEL

Relation	Inverse	Origin concept	Destination concept
-	group-by & group	domain	AI-task
-	solves	problem	AI-task
uses	isUsedBy	AI-task	AI-model
-	explained-by & explains	XAI-method	AI-model
is-applied	-	XAI-method	portability
has-scope	-	XAI-method	scope
outputs	-	XAI-method	output-format
is-portable	-	XAI-method	portability

The goal of representing these relationships is to enable the model implementation described next, which supports selecting the most suitable XAI approach for a given IoT system.

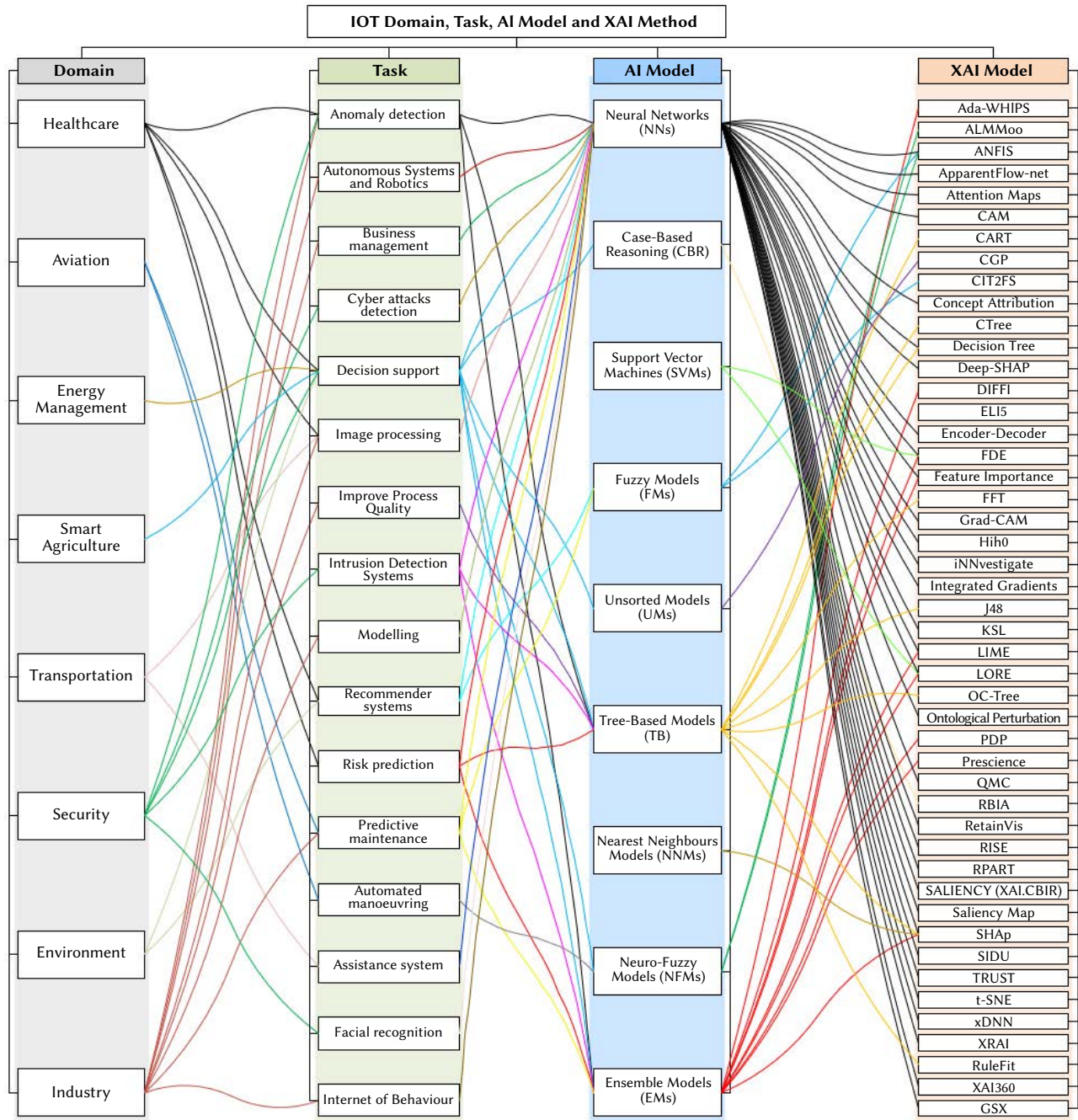


Fig. 8. Relationships between AIoT Domains, AIoT Tasks, AI Models, and XAI Methods.

IV. MODEL IMPLEMENTATION

This section presents the XAIoT ontology, which is the OWL-DL implementation of our conceptual model. The XAIoT ontology encompasses the same concepts and relationships we defined previously, describing the vocabulary defining XAI methods applied to IoT contexts. Fig. 9 shows a screenshot of the interactive visualization website of the ontology⁵. The corresponding source code of the ontology is also publicly available at the GitHub repository⁶.

The aim of the ontology is to guide developers in designing XAIoT systems. Therefore, we implemented the ontology using OWL, a well-known language for representing complex knowledge in the semantic

web, using the DL (description logics) logic language, fragments of first-order logic that allow reasoning using the concepts and relationships described in the ontology. We can describe our ontology by considering some metrics. We have 10 classes (the concepts defined in the conceptual model), 12 object properties that establish the relationships between these classes (described in Section D), and 90 instances that belong to one of our classes (shown in Fig 4).

As a result, we can make queries to extract interesting and useful knowledge. XAIoT allows us to create any query that involves a question about the concepts defined in the ontology considering the rest of the concepts we have and the relationships defined. Therefore, the ontology can provide a classification of the best XAIoT solutions to apply in a specific situation. In our case, any instance of a concept in the XAIoT ontology can be used as a parameter in the query to define this specific situation. We do not need to perform parameter optimization

⁵ Available at <https://gaia.fdi.ucm.es/ontologies/#XAIoT>

⁶ <https://github.com/UCM-GAIA/XAIoT>

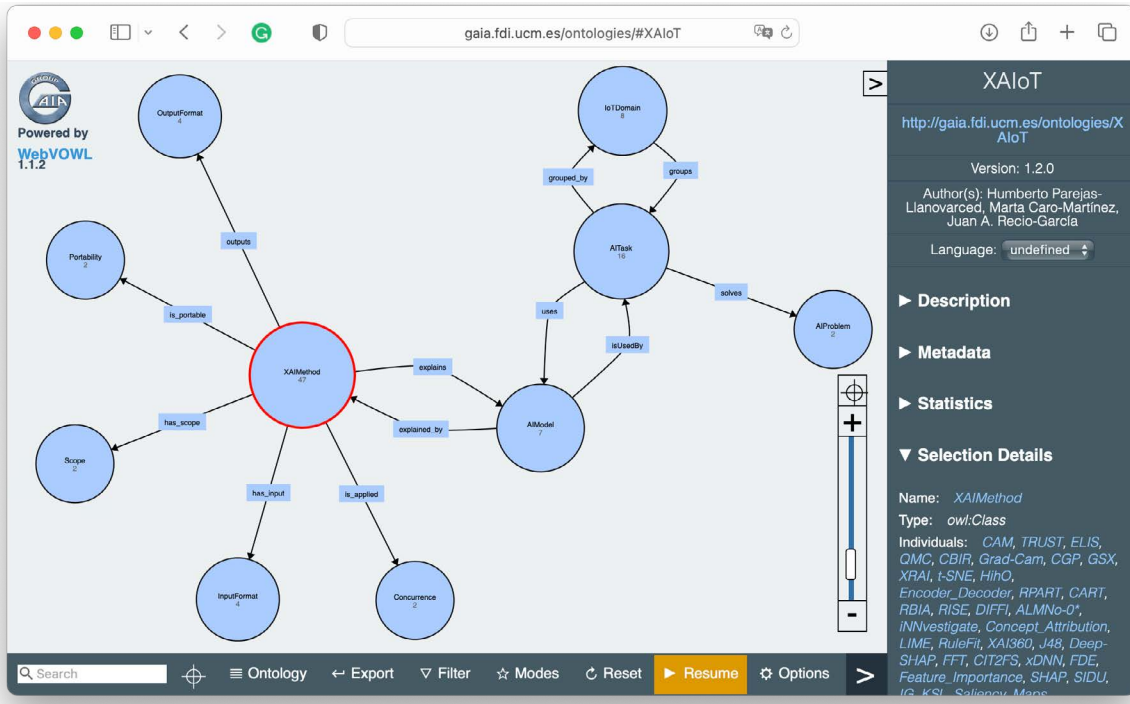


Fig. 9. Visualization of the XAIoT ontology, available on our website⁵.

DL query:

Query (class expression)

AIModel and explained_by value Ada-WHIPS

Execute Add to ontology

Query results

Instances (1 of 1)

Ensemble

DL query:

Query (class expression)

AITask and (grouped_by value SmartAgriculture)

Execute Add to ontology

Query results

Instances (1 of 1)

Decision Support

DL query:

Query (class expression)

XAIMethod and (has_scope value local) and is_portable value post-hoc and explains value CaseBasedReasoning

Execute Add to ontology

Query results

Instances (1 of 1)

Query for RBIA ? ☐ Direct superclasses

(a) Getting the AI models explained by the *Ada-WHIPS* explanation method.

(b) Getting the AI tasks applied to the *smart agriculture* domain.

(c) Obtaining the *local* and *post-hoc* XAI methods that can explain *CBR*.

Fig. 10. Examples using the XAIoT ontology for guiding the selection of the most suitable XAI method for an IoT application.

since the query result is going to retrieve all the solutions that satisfy the requirements included in the query. If those requirements are satisfied with several solutions, then there is not one better solution over another, and the design user can decide which one to use. As a consequence, we do not have a fixed list of queries to use on XAIoT because the design user can create any query according to their needs. We illustrate some examples of the available knowledge extraction in Fig. 10. The DL reasoner used to create these example queries and in validating the ontology was Pellet. Considering the concepts and instances represented in our ontology, we can infer that *ensemble methods* are the only *AI models* the *Ada-WHIPS* explainable method can explain (see Fig. 10a). We have another example in Fig. 10b where we can observe that *decision support tasks* are the tasks we found in the *smart agriculture* domain. We can make more complex queries such as the one shown in Fig. 10c. With that query, we obtain the *XAI method* (RBIA) that is *local* and *post hoc* and can be applied to explain a *CBR* system. Through this kind of reasoning process, our ontology can help to select what type of methods should be applied to explain an IoT application. As long as the ontology reasoners work (introducing a

query to obtain a prediction that satisfies the requirements included in the query), the only task to make use of XAIoT we implemented is classification. However, this is only in the case that we use a reasoner with the ontology. Actually, ontology can be used in other ways that involve knowledge-intensive tasks, whereas XAIoT can be used as a knowledge source.

V. MODEL VALIDATION THROUGH SURVEYED APPROACHES

The conceptual model was created following an iterative design process, in which every intermediate model was validated, replicating the validation process used by and . The validation aims to determine whether a subset of XAIoT systems can be classified with our model. The whole validation set comprises 100 XAIoT systems found in the systematic review presented in Section II. IoT experts perform this validation process, proposing refinement of the model when these XAIoT systems that cannot be classified. The process ends when all the approaches in the validation set are classified according to the proposed model.

In detail, the development of the XAIoT model was an iterative process, in which we have included several concepts in the model, as well as their relationships and instances. In every step, we validated this set of components to confirm that the classification performed by the model was correct. If the classification did not perform as expected, we added new components (concepts, instances, and relationships) to the model. Again, with this new version, we validated the classification performance. This process continued until the classification produced results as expected. At that moment, we considered the XAIoT model complete and finished. Therefore, we have carried out an ablation study, since we have validated the model at every step without several components included in the final version of the model.

Moreover, the validation process was carried out developing the ontology and performing the classification using Protégé. For each version of the XAIoT model (each step in the validation process), we developed the conceptual model and the OWL-DL code for the ontology. The classification process was validated using the Pellet reasoner and the Protégé interface. The whole development and validation process of the XAIoT model has been represented in Fig. 11 as a scheme, and detailed in Algorithm 1.

Algorithm 1. Algorithm to build and validate the XAIoT model

Input: *correctClassification* = *False*, *XAIoT model* = {}, *XAIMethods*

Output: *XAIoTmodel*

```

1  while correctClassification == False do
2    nConcepts, nInstances, nRelationships,
      expClassification = sys_lit_review()
3    newComponent =
      (nConcepts, nInstances, nRelationships)
4    add(newComponent, XAIoT model)
5    classification =
      apply(XAIoT model, XAIMethods)
6    if classification == expClassification then
7      correctClassification = True
8  end
9  return XAIoTmodel

```

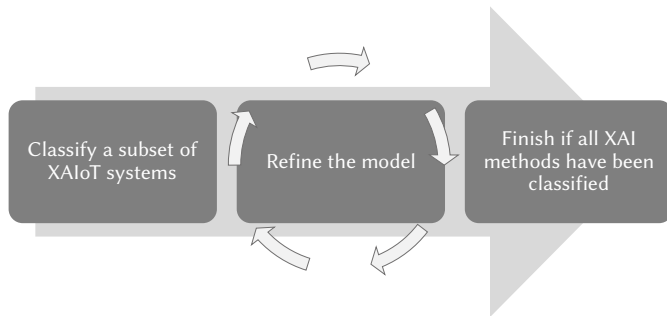


Fig. 11. XAIoT development and validation process scheme.

This classification is presented in Appendix. Next, we describe the most relevant systems included in this validation table, classified according to our conceptual model and organized by domain and task.

In the **energy management** domain, [54] present a transportation energy model (TEM) that uses XAI methods to predict household transportation energy consumption. [55] also proposed a method to predict energy demand in various situations using a deep learning model based on an autoencoder. In their experiment, they explain prediction results with the t-SNE algorithm. [38] proposed two novel workflows to predict occupancy: one that was based on computer vision and one based on machine learning. They used several state-

of-the-art explanation methods (such as SHAP) to provide a detailed analysis of the algorithm parameters and how the image properties influence the resulting performance.

Among the *decision support* tasks in the **environment** domain, we can highlight the work by [39], which uses XAI to build a toolbox for better interpreting species distribution models (SDMs). [56] also proposed a CNN to address adaptation space reduction. They explored using integrated gradients as an XAI method for learning and prediction. [57] presented a novel visual analytic platform, targeting two goals: a) an architecture for creating custom interactive visual analytic dashboards using well-defined components linked to each other and b) the inclusion of components specifically for making AI methods more explainable (i.e., SHAP). As an example of the *recommendation* task in the environment domain, [58] developed a microfluidic platform capable of monitoring temporal gene expression. By coupling deep NN and XAI algorithms (i.e., SHAP), they showed how to explain which genes contribute to identifying patterns.

As mentioned previously, **health** is the most popular domain for XAIoT applications. Therefore, we find many works that our model must be able to classify. Regarding the *decision support* task, [40] focused on AdaBoost, a black-box model that has been widely adopted in the CAD literature. They addressed the challenge of explaining AdaBoost classification with a novel Ada-WHIPS algorithm that extracts simple, logical rules from AdaBoost models. [52] proposed a new extended zero-order autonomous learning multiple-model (ALMMo-0) neurofuzzy approach to classifying different heart disorders through sounds. ALMMo-0 has a learning engine composed of a massively parallel set of fuzzy rules, which can self-adapt and provide transparent and human-understandable if-then representations. ALMMo-0 as XAI produced transparent fuzzy rules, which are humanly interpretable and may help specialists provide a more accurate diagnosis. [59] proposed a method to classify cardiac pathology based on a novel approach to extract image-derived explanation features to characterize the shape and motion of the heart. ApparentFlow-net is applied to generate pixelwise apparent flow given a pair of image frames on the same slice as input. [101] proposed an ante-hoc classification method based on a Cartesian Genetic Programming (CGP) approach, which automatically identifies the presence of the disease and concurrently provides the explicit classification model used by the system. [60] defined another ante-hoc cardiovascular prediction system that can be easily interpreted based on two principles: a) reduction in attributes without degrading the prediction system performance and b) selection of recursive partitioning and regression trees (RPART) and conditional inference trees (CTree) as techniques to interpret the final prediction system. [61] proposed adopting deep learning for COVID-19 detection from X-rays. The explanation aims to localize the areas in the X-ray symptomatic of COVID-19 presence. Additionally, to provide explainability, they proposed visualizing class activation maps (Grad-CAM), a technique usually considered for debugging deep neural networks. [64] adapted state-of-the-art XAI algorithms (Grad-CAM and SHAP) for hand gesture classification, evaluating the contribution of each input feature to the prediction. [65] proposed an ante-hoc electrocardiogram classification method named knowledge-shot learning (KSL). KSL constructs a novel neural network architecture that can extract interpretable feature vectors and provide support cases as result explanations. [118] examined the influence of different input representations on a trained model for hip arthroplasty (THA) using LIME. Another application of LIME was proposed by [119] to generate explanations on how a CNN detects tumor tissue in patches extracted from histology whole slide images.

[42] proposed a conceptual framework for building human-centered, decision-theory-driven XAI (SHAP for attribution, LORE for counterfactual rules) by designing and implementing an explainable

clinical diagnostic tool for intensive care phenotyping. Regarding counterfactual explainability, there are other studies such as [120] with their method named RELAX (reinforcement learning agent explainer for arbitrary predictive models), [121] with FOCUS (flexible optimizable counterfactual explanations for tree ensembles) method, and [122] with their feature tweaking algorithm.

[66] introduced DoctorXAI, a post hoc explainability technique that deals with multilabel, sequential, ontology-linked data. [123] presented an XAI model fit on a large longitudinal dataset using electrodes that can reveal which are most important. To determine the relative importance of the different features for each model prediction, they adopted SHAP.

Concerning *risk prediction* tasks in the health domain, [97] focused on applying XAI techniques (CART, CTree, FFT, and J48) to health care data as decision trees to predict the falling risk in older adults. [124] proposed an interpretable machine learning framework (feature importance) for image classification problems to make high-quality explanations. [43] reported the development and testing of a machine learning-based system that predicts the risk of hypoxemia and provides real-time explanations (prescience) of the risk factors during general anesthesia.

We can highlight the following proposals regarding *image processing* tasks for health IoT applications. [68] presented the main components of the framework for attribution to concepts and introduced a spatial pooling operation on top of the feature maps to obtain a solid interpretability analysis through concept attribution, which is a complementary technique to saliency heatmaps giving concept-based explanations rather than pixel-based explanations. [69] proposed a novel in-model joint architecture composed of an explainer and a classifier (the explainer is based on an encoder-decoder network, and the classifier is based on VGG-16). This architecture outputs a class label and a visual explanation of such a decision without needing additional labeled data to train the explainer in addition to the image class. [70] showed how the expert-level evaluation of XAI methods (Grad-CAM and SIDU) in a medical application can be utilized and aligned with the actual explanations generated by the clinician. [71] presented a framework for augmenting statistical findings in medical imaging workflows. A novel visualization technique (hierarchical occlusion: HiHO) was proposed to increase the portability of machine learning interpretability to IoT applications.

We can also select a couple of relevant systems for *recommendation* tasks in the health domain. [72] presented an explainable AI early warning score (XAI-EWS) system for the early detection of acute critical illness. XAI-EWS potentiates clinical translation by accompanying a prediction with information on the data explaining it. Explanations are calculated with the high-level library for explaining neural networks *iNNvestigate*. [73] presented a human-in-the-loop XAI system for content-based image retrieval (CBIR) of video frames similar to a query image from minimally invasive surgery (MIS) videos for surgical education.

Finally, for the *anomaly detection* task, [96] proposed a novel one-class model that uses an explainable one-class decision tree (OC-Tree).

Industry is also a popular domain for IoT applications. *Anomaly detection* is an important AI task with enormous applicability in industrial scenarios. [45] proposed an approach for defining “feature importance” explanations in anomaly detection problems applied to quality monitoring. [46] reported a methodology to incorporate XAI-based insights (feature importance) into the data science process of developing a high-precision classifier (for fault diagnosis of steel plates). [74] developed a novel explainable method called the Gumbel-Sigmoid eXplanator (GSX). First, they proposed a differentiable and trainable selector complemented by a regularization layer to control the number of features for each instance. [95] proposed a novel

approach for fault detection and diagnosis in rotating machinery. The methodology consists of feature extraction, fault detection, and fault diagnosis. SHAP, the popular technique to interpret black-box models, was used in fault diagnosis.

For the *autonomous processes and robotics* task, we found the work by [75] that proposed using deep reinforcement learning (DRL) to optimize both aggregate and minimum service provisioning. To achieve trusted autonomy, DRL agents must be able to explain their actions for transparent human-machine interrogation through feature importance. Additionally, [76] presented a method that incorporates explicit object recognition processing into DRL models. Then, object saliency maps visualize the internal states of DRLNs, thus enabling the formation of explanations.

For *business management* tasks in the industrial environment, there are several systems in the literature. [77] proposed a novel local post hoc explanation (decision tree) approach for a deep learning classifier that is expected to justify the model decisions for experts. [125] presented the use case of power generation forecasting using the XGBoost algorithm. Feature relevance was explained with the ELI5 XAI tool. [126] illustrated the opportunities and challenges of process predictions and XAI for Industry 4.0 with a fully automated factory prototype. They addressed a binary classification problem where global feature importance presented as textual explanations was applied to deep neural networks.

Regarding industrial *image processing* tasks, S. [78] presented an ante-hoc approach for analyzing a classification procedure for fiber layout defects. Smoothed integrated gradients, guided gradient class activation mapping, and DeepSHAP were applied to a CNN classifier. For the *process quality improvement* task, [98] proposed using nonlinear modeling with SHAP to infer how a set of production parameters and the process quality of a manufacturing system are related.

[28] introduced and discussed the concept of *Internet of Behavior* (IoB) tasks and their integration with XAI techniques (XAI explainer platforms such as IBM AI Explainability 360) to provide trusted and evident experience in the process of influencing user behavior in an electrical power consumption use case.

Regarding *modeling* tasks, [79] presented a modeling tool, NeuroCAD, using sensitivity analysis, layerwise relevance propagation, the Grad-CAM method, and LIME.

Regarding *predictive maintenance* tasks, we can find several model-dependent systems using CNNs. [51] introduced a post hoc XAI approach for CNNs on vibration signal analysis classification using Grad-CAM. Similarly, [80] demonstrated how CNN and CAM could be employed in real-world vibration video to characterize the machine's status, representing normal or fault conditions. [81] demonstrated that CNN can be used for making predictions based on multivariate time-series data and explaining (using CAM) these predictions. Finally, [127] investigated why deep neural network (DNN) models cannot identify the problematic component of a turbofan engine due to its black-box property. This study employs dimensionality reduction and SHAP techniques.

The following domain is **security**. With regard to *anomaly detection* tasks, [82] proposed an autoencoder-based detection framework using convolutional and recurrent networks to discover cyber threats in IIoT networks. It also explains employing LIME in the model through latent representations of malicious events.

For *cyber attack detection* tasks, [83] designed an approach using a set of LSTM modules integrated into an ensemble of detectors for cyber attacks against IIoT systems. These modules were then merged using a decision tree to obtain an explainable output at the final stage.

We can also highlight some *decision support* applications for the security domain. [84] investigated trustworthiness and interpretability

via XAI-based techniques (DeepSHAP) to understand, interpret and improve the behavior of state-of-the-art multimodal traffic classifiers. [85] proposed a universal XAI model named transparency relying upon statistical theory (TRUST), which is model-agnostic, high-performing, and suitable for numerical applications.

Concerning the *facial recognition* task, [86] investigated human face identification based on extracting the facial features and diagnosing and explaining the feature sets with geometric ratio expressions (comparing the derived quadratic mean QM of two similar vectors).

We can also find many examples of *intrusion detection* tasks, such as the proposal by [88] that design a new XAI-based framework to explain any critical decisions for IoT intrusion detection systems. Their framework uses three main XAI techniques (RuleFit, LIME, and SHAP) on top of their DNN-based model. [100] also explained the outcomes of intrusion detection systems, in this case, through decision tree algorithms that can be easily read with the goal of facilitating decision-making. [87] compared the detection of IoT device malware using CNN models. Then, Grad-CAM along the t-SNE was used to understand the overall performance of the proposed methods. Finally, we can cite [82], who proposed a highly efficient model named XSRU-IoMT for the effective and timely detection and explanation of sophisticated attack vectors in IoT networks.

The **smart agriculture** domain mostly focuses on *decision support* applications. For example, [89] exposed an interpretable machine learning approach based on a multivariate decision tree to forecast the evaporation rate using data from an IoT infrastructure. [90] developed the automatic and intelligent data collector and classifier framework by integrating IoT and deep learning. The Grad-CAM algorithm is employed to visualize the features and their importance extracted by the sensors. [49] proposed a model framework to identify factors that potentially explain agricultural land use for different plantations. By developing a machine learning model coupled with a model-agnostic explanation approach (LIME and partial dependence plots PDP), their framework provides both global and local interpretations of the most influential factors.

The last domain is **transportation**. With respect to *assistance* tasks, [91] demonstrated how explanations using attention maps can be provided for a deep learning system that detects specific objects (charging posts) for driver assistance in an electric bus.

Regarding *image processing* for transportation, there are several systems in the literature, such as the one proposed by [92] that presented a hierarchical approach for predicting eye fixations in autonomous driving scenarios. They built an encoder-decoder network that merges visual feature information to obtain global-local explanations employing attention maps. [93] developed a vision-based system to detect and identify several objects and predict the intention of pedestrians in traffic scenes. This study applied the RISE algorithm to make the final classification explainable by generating saliency maps. [128] proposed different object detection algorithms and the SHAP method to analyze and explain their performance. Finally, [94] proposed using an explainable intelligence technique (XRAI) to understand some of the tasks involved in developing advanced driver-assistance systems (ADAS).

As a conclusion to our validation, we can say that we described and classified the previous works obtained from the literature using our conceptual model. Then, we concluded that our conceptual model is complete enough to characterize XAIoT systems and, therefore, can help XAIoT developers as a tool to consult different alternatives to apply explanation systems in IoT scenarios. We also can conclude that our model has overcome previous related works [9], [10], whose main contributions were to define key features in XAIoT approaches. These works were studies and surveys, they did not proposed a conceptual

model or a formalized guide. In our work, we provide the community with a formalized model that allow design users to develop XAI solutions specifically through an ontology-based methodology. Some of our previous works also offer a formalized conceptual model but for the design of XAI approaches in general, not specifically for the IoT requirements [11], [12]. Therefore our work is a novel contribution to the current state-of-the-art. Moreover, after the analysis of the different conceptual and terminological approaches found in the XAIoT literature, we can conclude that the definitions adopted in the proposed model are task- and context-related. Our main concept is the AI task to be explained, and all the remaining concepts try to elaborate further on the definition of this task. Both IoT tasks and domains have been studied in-depth to find the explanation requirements for each IoT scenario. There are other concepts in the existing literature, like user's related ones, that are not so relevant or there is no clear consensus on their usefulness for the IoT field.

VI. CONCLUSIONS

The increasing complexity of data generated by IoT devices makes it difficult for humans to understand the decisions made by the AI models that interpret these data. To address this issue, we propose the XAIoT model, a theoretical framework based on an in-depth study of the state-of-the-art that defines and delimits the features of explainable AI applied to the IoT. This conceptual model allows developers to examine the different concepts, attributes, requirements, and relationships to consider when developing this type of system.

However, the model may have some limitations, the XAI and IoT domains are growing and evolving very quickly, and the emergence of new innovative and/or disruptive proposals will include aspects that have not been considered in our model. Nevertheless, these aspects can be incorporated through the iterative design process used during its validation.

The ultimate goal of this conceptual model is to provide a framework that guides the selection of the most suitable XAI solution for a given intelligent IoT application. This model was validated by classifying and characterizing up to 100 XAIoT solutions. Moreover, we built an ontology using the OWL-DL language that supports making semantic queries to the model and obtaining the most suitable explanation solution for a given domain, task, and AI model.

This systematic review and the validation process clearly demonstrate that health care is the domain with the most applications in both XAI and IoT areas, followed by industrial solutions. Furthermore, the predominant XAI methods are SHAP and LIME. Additionally, the most significant AI models are neuronal networks and ensemble learning, which mainly address classification problems.

Exploiting the XAIoT ontology is the most relevant line of future work, as we plan to implement a recommender system that, by using the knowledge represented in the ontology, may provide the most suitable XAI approach for the needs of an intelligent IoT system. There are also some extensions to the conceptual model that can be explored, such as the impact of distributed architectures since these kinds of systems are inherent to the IoT domain [102].

In addition, as future work we can study and consider other characteristics of the IoT, such as the type of device [129], and semantic communication (which integrates user and application requirements and the meaning of information in data processing and transmission) [130]. Those characteristics were not key concepts in our validation, as they were not decisive for the classification of XAIoT methods. However, as we have mentioned, the XAI and the IoT are fields that are evolving fast nowadays, so we need to consider new characteristics in our model for future work.

APPENDIX

A. Validation Tables

The Validation Table (see Table III) consolidates the classification that has been made on the 82 publications selected in this research, and where we have identified 101 XAI experiments applied to the IoT domain (XAIoT). AIoT tasks (16) are grouped by their application domain (8) and use an AI model (9), where the AI model solves a type of problem and explains it using an XAI method. Likewise, the properties of XAI methods are detailed, such as concurrence, scope, output format, input format, and portability. AI Model (refer to Table I), problem type (C: classification, R: regression), portability (MS: specific, MA: agnostic), concurrence (Ah: ante hoc; Ph: post hoc), scope (L: local; G: global), input format (N: tabular numeric; T: tabular text; V: visual; TS: time series), output format (N: numeric; R: rules; T: textual; V: visual; C: chart).

TABLE III. CLASSIFICATION OF XAIoT EXPERIMENTS

IoT Domain	Task	Author	AI Model	Type	Portab.	XAI Method	Conc.	Scope	Input Format	Output Format
Aviation	Automated manoeuvring	Keneni et al., 2019	NFM, NN	C R	MS	ANFIS	Ph	L G	N, T	N, R, T, C
	Predictive maintenance	Ten Zeldam et al., 2018	EM, NN, NNN, SVM	C R	MS	FDE	Ph	L G	N, T	N, C
Energy Management	Decision support	Amiri et al., 2021	NN	C	MA	LIME	Ph	L	N, T	N, C
		Sirmacek et al., 2020	TB	C	MA	SHAP	Ph	L G	V	N V
		Kim et al., 2019	NN	C	MA	t-SNE	Ph		N, TS	N
Environment	Decision support	Diallo et al., 2020	NN	C	MA	Integrated Gradients	Ph	L	N, TS	N, C
		Ryo et al., 2021	EM	C	MA	LIME	Ph	L	N	N, C
		Kalamaras et al., 2019	NN	C	MA	SHAP	Ph	L G	N, TS	N, C
	Recommender system	Graham et al., 2020	NN	C	MA	SHAP	Ph	L G	N, TS	N, C
Healthcare	Anomaly detection	Itani et al., 2020	TB	C R	MS	OC-Tree	Ah	G	N	N
	Decision support	Hatwel et al., 2020	EM	C R	MS	Ada-WHIPS	Ph	L	N, T	N, R, T
		Soares et al., 2020	NFM	C R	MS	ALMMo-0	Ah	L G	N, T	N, R
		Aghamohammadi et al., 2019	NFM, NN	C R	MS	ANFIS	Ph	L G	N, T	N, T, C
		Zheng et al., 2019	NN	C R	MS	ApparentFlow-net	Ah	L	V	V
		Senatore et al., 2019	UM	C R	MS	CGP	Ah	G	V	N, V
		Porto et al., 2021	TB	C R	MS	CTree	Ah	L	N, T	N, R, C
		Brunese et al., 2020	NN	C R	MS	Grad-CAM	Ph	L	V	V
		Tabik et al., 2020	NN	C R	MS	Grad-CAM	Ph	L	V	V
		Hossain et al., 2020	NN	C R	MA	Grad-CAM	Ph		N, T, V	N, V
		Gozzi et al., 2022	NN	C R	MS	Grad-CAM	Ph	G	N, V, TS	N, V
		Chou et al., 2020	NN	C R	MS	KSL	Ah Ph		N, T, TS	N, C
		Dindorf et al., 2020	MA	C	MA	LIME	Ph	L G	N, T, TS	N, C
		Palatnik de Sousa et al., 2019	MA	C	MA	LIME	Ph	L G	V	N, V
		Lage et al., 2018	MA	C	MA	LIME	Ph	L G	T	N, C
		Hossain et al., 2020	NN	C	MA	LIME	Ph	L	N, T, V	N, V
		Dave et al., 2020	EM	C	MA	LIME	Ph	L	N	N, C
		Wang et al., 2019	EM, NN, SVM	C R	MS	LORE	Ph	L	N, T	N, R, T
		Panigutti et al., 2020	NN	C R	MS	Ontological Perturbation	Ph	L	N, T	N, R, T
		Lamy et al., 2020	CBR	C R	MS	RBIA	Ph	L G	N	T, C
		Lamy et al., 2019	CBR	C R	MS	RBIA	Ph	L G	N, T	N, T, C
		Kwon et al., 2019	NN	C R	MS	RetainVis	Ph	L G	N, T	N, T, C
		Porto et al., 2021	NN	C R	MS	RPART	Ah	L	N, T	N, R, C
		Hu et al., 2021	MA	C	MA	SHAP	Ph	L G	N	N, C
		Lin et al., 2020	MA	C	MA	SHAP	Ph	L G	N	N, C
		Wang et al., 2019	MA	C	MA	SHAP	Ph	L G	N, T	N, R, T
		Dave et al., 2020	EM	C	MA	SHAP	Ph	L G	N	N, C
		Gozzi et al., 2022	NN	C	MS	SHAP	Ph	L G	N, V, TS	N, V
		Soares et al., 2020	NN	C R	MS	xDNN	Ah	G	N, T	N, R
	Image processing	Graziani et al., 2020	NN	C R	MS	Concept Attribution	Ph	L G	V	V
		Rio-Torto et al., 2020	NN	C R	MS	Encoder-Decoder	Ah	L	V	V
		Muddamsetty et al., 2021	NN	C R	MS	Grad-CAM	Ph	L	V	V
		Monroe et al., 2020	NN	C	MA	HihO	Ph	G	V	V
		Muddamsetty et al., 2021	NN	C R	MS	SIDU	Ph	L	V	V

IoT Domain	Task	Author	AI Model	Type	Portab.	XAI Method	Conc.	Scope	Input Format	Output Format
	Recommender system	D'Alterio et al., 2020	FM	C R	MS	CIT2FS	Ah	L	N	T, C
		Lauritsen et al., 2020	NN	C R	MS	iNNvestigate	Ph	L G	N	N, C
		Chittajallu et al., 2019	NN	C	MA	SALIENCY (XAI-CBIR)	Ph		V	V
	Risk prediction	Lindsay et al., 2020	TB	C R	MS	CART	Ah	L	N, T	R, T
		Lindsay et al., 2020	TB	C R	MS	CTree	Ah	L	N, T	R, T
		Pintelas et al., 2020	MA	C	MA	Feature Importance	Ph	L G	V	N, T, V
		Lindsay et al., 2020	TB	C R	MS	FFT	Ah	L	N, T	R, T
		Lindsay et al., 2020	TB	C R	MS	J48	Ah	L	N, T	R, T
		Lundberg et al., 2018	EM	C R	MS	Prescience	Ph	L	N, T, TS	N, C
		Pnevmatikakis et al., 2020	NN	C	MA	SHAP	Ph	L G	N, T	N, C
Industry	Anomaly detection	Carletti et al., 2019	EM	C R	MS	DIFFI	Ph	G	N, TS	N, C
		A. Khara et al., 2020	EM	C	MA	Feature Importance	Ph	L G	N, T	N, R, T, C
		Mansouri et al., 2022	NN	C	MA	GSX	Ph		N, TS	N
		Brito et al., 2022	NNN, kNN	C	MA	SHAP	Ph	L G	N	N, C
	Autonomous Systems and Robotics	Guo et al., 2020	NN	C	MS	Feature Importance		L	N	N, T, C
		Iyer et al., 2018	NN	C R	MA	Saliency Map		L	V	N, V
	Business Management	Mehdiyev et al., 2021	NN, TB	C R	MA	Decision Tree	Ph	L	N, T	N, R
		Zhang et al., 2021	MA		MA	Deep-SHAP	Ph	L G	N	N, C
		Sarp et al., 2021	MA	C R	MA	ELI5	Ph	L G	N, T, TS	N, C
		Rehse et al., 2019	MA	C	MA	Feature Importance	Ph	L G	N, TS	N, T, C
	Image processing	S. Meister et al., 2022	NN	C	MA	Deep-SHAP	Ph	L G	V	N, V
		S. Meister et al., 2021	NN	C	MA	Integrated Gradients	Ph	L	V	N, V
	Improve Process Quality	Senoner et al., 2021	TB	C	MA	SHAP	Ph	L G	N	N, C
	Internet of Behaviour Modelling	Elyan et al., 2022	NN	C	MS MA	XAI360	Ph	L G	N, TS	R, T
		Schönhof et al., 2021	NN	C R	MS	Grad-CAM	Ph	L	V	V
		Schönhof et al., 2021	MA	C	MA	LIME	Ph	L G	V	V
	Predictive maintenance	Chen et al., 2020	FM, NN	C R	MS	ANFIS	Ph	L G	V	N, R, T, V
		Sun et al., 2020	NN	C R	MS	CAM	Ph	G	V	V
		Assaf et al., 2019	NN	C R	MS	CAM	Ph	G	V	V
		Serradilla et al., 2020	MA	C R	MA	ELI5	Ph	L G	N, TS	N, C
		Chen et al., 2020	NN	C R	MS	Grad-CAM	Ph	L	V	V
		Serradilla et al., 2020	MA	C	MA	LIME	Ph	L G	N, TS	N, C
		Hong et al., 2020	MA	C	MA	SHAP	Ph	L G	N, TS	N, C
		Serradilla et al., 2020	MA	C	MA	SHAP	Ph	L G	N, TS	N, C
Security	Anomaly detection	Khan et al., 2021	NN	C	MA	LIME	Ph	L	N, T, TS	N, C
	Cyber attacks detection	Saharkhizan et al., 2020	NN	C R	MS	Decision Tree	Ph	G	N, TS	N, C
	Decision support	Nascita et al., 2021	NN		MA	Deep-SHAP	Ph	G	N, T	N, C
		Sarhan et al., 2021	EM	C	MA	SHAP	Ph	L G	N, T	N, C
		Le et al., 2022	TB	C	MA	SHAP	Ph	L G	N, T, TS	N, C
		Zolanvari et al., 2021	NN	C	MA	TRUST	Ah	L	N, T, TS	N
	Facial recognition	Sarathy et al., 2020	NN	C R	MS	QMC	Ph	G	V	N, V
	Intrusion Detection Systems	Mahbooba et al., 2021	TB	C R	MS	Decision Tree	Ah	G	N, T, TS	N, R
		Naeem et al., 2022	NN	C R	MA	Grad-CAM	Ph	L	V	V
		Houda et al., 2022	NN	C	MA	LIME	Ph	L	N, T, TS	N, C
		Khan et al., 2022	NN	C	MA	LIME	Ph	L	N, TS	N, C
		Houda et al., 2022	TB		MS	RuleFit	Ah	G	N, T, TS	N, C
		Houda et al., 2022	MA	C	MA	SHAP	Ph	L G	N, T, TS	N, C
		Alani et al., 2022	EM	C	MA	SHAP	Ph	L G	N, TS	N, C
		Naeem et al., 2022	NN	C	MA	t-SNE	Ph		V	N, V
Smart Agriculture	Decision support	Garrido et al., 2022	NN	C R	MS	Decision Tree	Ah	G	N, TS	R, C
		Kundu et al., 2021	NN	C R	MA	Grad-CAM	Ph	L	V	V
		Viana et al., 2021	EM	C	MA	LIME	Ph	L	N, T	N, C
		Viana et al., 2021	EM		MA	PDP	Ph	G	N, T	N, C
Transport.	Assistance system	Nowak et al., 2019	NN	C R	MS	Attention Maps	Ph	L	V	V
	Image processing	Martinez-Cebrian et al., 2020	NN	C R	MS	Attention Maps	Ph	L	V	V
		Li et al., 2020	NN	C R	MS	RISE	Ph	L	V	V
		Ponn et al., 2020	MA	C	MA	SHAP	Ph	L G	V	N, V
		Lorente et al., 2021	NN	C R	MS	XRAI	Ph	L	V	V

B. Data Tables

The following tables are the data source of the Fig. 2, Fig. 3, Fig. 5, Fig. 6, and Fig. 7.

TABLE IV. QUANTITATIVE ANALYSIS OF THE SYSTEMATIC REVIEW ON EXPLAINABLE AI AND IoT. THE TABLE ON THE LEFT (DOCUMENTS BY YEAR) SHOWS THE NUMBER OF PUBLICATIONS RESULTING FROM OUR SEARCH BETWEEN 2018 AND 2022, A TOTAL OF 7,022 DOCUMENTS THAT INCLUDE TERMS RELATED TO XAI, AND 122 DOCUMENTS THAT IN ADDITION TO XAI INCLUDED THE TERM IoT. WHILE THE TABLE ON THE RIGHT (DOCUMENTS BY SUBJECT AREA) SHOWS THE DIFFERENT SUBJECT AREAS IN WHICH THE RESEARCH WAS CONDUCTED, IN SOME CASES AN ARTICLE HAS BEEN CATEGORIZED INTO MORE THAN ONE SUBJECT AREA (REFER TO FIG. 2)

Year	XAIoT	XAI
2018	2	177
2019	7	581
2020	20	1,267
2021	41	2,293
2022	52	2,702
Total	122	7,020

Subject area	XAIoT	XAI
Computer Science	115	5,258
Engineering	49	1,928
Mathematics	37	1,515
Medicine	3	827
Decision Sciences	16	624
Social Sciences	7	490
Physics & Astronomy	8	424
Materials Science	10	397
Biochemistry & Genetics	5	364
Business & Management	7	290
Chemical Engineering	0	163
Energy	6	163
Neuroscience	1	160
Environmental Science	2	157
Arts and Humanities	0	156
Chemistry	5	156
Multidisciplinary	0	142
Health Professions	1	141
Earth & Planetary Sciences	0	137
Agricultural & Biological Sci.	1	74
Psychology	0	71
Economics & Finance	1	57
Immunology & Microbiology	0	22
Nursing	0	20
Pharmacology & Toxicology	0	20
Dentistry	0	8
Veterinary	0	2
Total	274	13,766

TABLE V. NUMBER OF PAPERS IN XAIoT PUBLISHED PER YEAR AND DOMAIN, OVER THE 82 OUT OF 122 XAIoT ARTICLES CHOSEN FOR THEIR RELEVANCE TO PERFORM THIS RESEARCH (REFERS TO FIG. 3)

Year	2018	2019	2020	2021	2022	Total
Aviation	1	1				2
Energy Management		1	1	1		3
Environment		1	2	1		4
Healthcare	2	8	20	3	1	34
Industry	1	3	6	5	4	19
Security			2	5	5	12
Smart Agriculture				2	1	3
Transportation		1	3	1		5
Total	4	15	34	18	11	82

TABLE VI. NORMALIZED DISTRIBUTION OF AI TASK BY IoT DOMAIN. THE TABLE SHOWS THE NUMBER OF ARTICLES PUBLISHED BY EACH OF THE 16 AIoT TASKS (ROWS) AND EACH OF THE 8 IoT DOMAINS (COLUMNS) IDENTIFIED IN OUR RESEARCH (REFERS TO FIG. 5)

Application	Aviation	Energ	Env	Healthc	Indust	Sec	Agric	Transp	Total
Anomaly detect				1	4	1			6
Assistance syst								1	1
Automated M.	1								1
Autonomous Syst					2				2
Business Manag.					4				4
Cyber attacks D.						1			1
Decisión support		3	3	22		4	3		35
Facial recognition						1			1
Image processing				4	1			4	9
Improve Proc Qlty					1				1
IoB					1				1
Intrusion Detect.						5			5
Modelling					1				1
Predictive maint	1				5				6
Recommender S.			1	3					4
Risk prediction				4					4
Total	2	3	4	34	19	12	3	5	82

TABLE VII. NORMALIZED DISTRIBUTION OF XAI METHOD BY DOMAIN. THE TABLE SHOWS THAT, OF THE TOTAL OF 82 SELECTED PUBLICATIONS, A TOTAL OF 101 EXPERIMENTS CARRIED OUT WITH 47 DIFFERENT XAI METHODS WERE IDENTIFIED (SOME OF THE ARTICLES EXPERIMENTED WITH MORE THAN ONE XAI METHOD). THE TABLE ALSO DISAGGREGATES THE NUMBER OF EXPERIMENTS PERFORMED BY EACH OF THE 8 IoT DOMAINS -COLUMNS- AND THE 47 XAIoT METHODS -ROWS- (REFER TO FIG. 6)

XAI Method	Aviat	Energ	Env	Health	Indust	Sec	Agric	Transp	Total
Ada-WHIPS				1					1
ANFIS	1			1	1				3
ApparentFlow.				1					1
Att. Maps								2	2
CAM					2				2
CART				1					1
CGP				1					1
CIT2FS				1					1
Conc. Attr.				1					1
CTree				2					2
Decis. Tree					1	2	1		4
Deep-SHAP					2	1			3
DIFFI					1				1
ELI5					2				2
Encod.-Decod.				1					1
FDE	1								1
Feat. Imp.				1	3				4
FFT				1					1
Grad-CAM				5	2	1	1		9
GSX					1				1
HihO				1					1
iNNvestigate				1					1
IG			1		1				2
J48				1					1
KSL				1					1
LIME		1	1	5	2	3	1		13
LORE				1					1
OC-Tree				1					1
Ontol. Pert.				1					1
PDP							1		1
Prescience				1					1
QMC						1			1
RBIA				2					2
RetainVis				1					1
RISE								1	1
RPART				1					1
RuleFit						1			1
SALIENCY(CBIR)				1					1
Saliency Map					1				1
SHAP		1	2	6	4	4		1	18
SIDU				1					1
TRUST						1			1
t-SNE		1				1			2
XAI360					1				1
xDNN				1					1
XRAI								1	1
ALMMo-0				1					1
Total	2	3	4	44	24	15	4	5	101

TABLE VIII. NORMALIZED DISTRIBUTION OF XAI METHOD BY AI MODEL. THE TABLE SHOWS THAT, OF THE TOTAL OF 82 SELECTED PUBLICATIONS, A TOTAL OF 101 EXPERIMENTS CARRIED OUT WITH 47 DIFFERENT XAI METHODS WERE IDENTIFIED (SOME OF THE ARTICLES EXPERIMENTED WITH MORE THAN ONE XAI METHOD). THE TABLE ALSO DISAGGREGATES THE NUMBER OF EXPERIMENTS PERFORMED BY EACH OF THE 9 AI MODELS -COLUMNS- AND THE 47 XAIoT METHODS -ROWS- (REFER TO FIG. 7)

XAI Method	CBR	EM	FM	MA	NFM	NN	NNN	TB	UM	Total
Ada-WHIPS		1								1
ANFIS			1		2					3
ApparentFlow-net						1				1
Attention Maps						2				2
CAM						2				2
CART								1		1
CGP									1	1
CIT2FS			1							1
Concept Attribution						1				1
CTree								2		2
Decision Tree						3		1		4
Deep-SHAP				1		2				3
DIFFI		1								1
ELI5				2						2
Encoder-Decoder						1				1
FDE		1								1
Feature Importance		1		2		1				4
FFT								1		1
Grad-CAM						9				9
GSX						1				1
HihO						1				1
iNNvestigate						1				1
Integrated Gradients						2				2
J48								1		1
KSL						1				1
LIME		3		5		5				13
LORE		1								1
OC-Tree								1		1
Ontological Perturbation						1				1
PDP		1								1
Prescience		1								1
QMC						1				1
RBIA	2									2
RetainVis						1				1
RISE						1				1
RPART						1				1
RuleFit								1		1
SALIENCY (XAI-CBIR)						1				1
Saliency Map						1				1
SHAP		4		7		4	1	2		18
SIDU						1				1
TRUST						1				1
t-SNE						2				2
XAI360						1				1
xDNN						1				1
XRAI						1				1
ALMMo-0					1					1
Total	2	14	2	17	3	51	1	10	1	101

CREDiT AUTHORSHIP CONTRIBUTION STATEMENT

Humberto Parejas-Llanovarcad, Marta Caro-Martínez, and Juan A. Recio-García: Conceptualization; Methodology; Data curation; Formal analysis; Software; Writing – Original Draft; Writing – Review & Editing.

DATA STATEMENT

The data supporting the findings of this study are publicly available on Github⁷.

DECLARATION OF CONFLICTS OF INTEREST

The authors confirm that there are no financial, personal, or professional conflicts of interest that could have influenced the work reported in this manuscript.

ACKNOWLEDGMENT

Supported by Grant PID2023-150566OB-I00 (AUDITIA-X) funded by MCIN/AEI/10.13039/501100011033 and by ERDF/EU and the BOSCH-UCM Honorary Chair on Artificial Intelligence applied to Internet of Things. Supported by the SHOU-X project (PR17/24-31890), funded by Comunidad de Madrid through the UCM program that supports projects led by early-career researchers (Line A). Action funded by the Community of Madrid through the Multiannual Agreement for the regulation of the cooperation framework within the Regional System of Scientific Research and Technological Innovation, signed between the Community of Madrid and the UCM, within the framework of the 6th PRICIT (Regional Plan for Scientific Research and Technological Innovation for the 2022–2025 period). Supported by the SI4/PJI/2024-00032 project funded by the Comunidad de Madrid through the UAM.

REFERENCES

- [1] M. R. Islam, M. U. Ahmed, S. Barua, S. Begum, “A Systematic Review of Explainable Artificial Intelligence in Terms of Different Application Domains and Tasks,” *Applied Sciences*, vol. 12, p. 1353, jan 2022, doi: <http://doi.org/10.3390/app12031353>
- [2] D. Gunning, M. Stefik, J. Choi, T. Miller, S. Stumpf, G. Z. Yang, “Xai—explainable artificial intelligence,” *Science Robotics*, vol. 4, 12 2019, doi: <http://doi.org/10.1126/SCIROBOTICS.AAY7120>
- [3] O. Loyola-Gonzalez, “Black-box vs. white-box: Understanding their advantages and weaknesses from a practical point of view,” *IEEE Access*, vol. 7, pp. 154096–154113, 2019, doi: <http://doi.org/10.1109/ACCESS.2019.2949286>
- [4] A. B. Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. García, S. Gil-López, D. Molina, R. Benjamins, *et al.*, “Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai,” *Information fusion*, vol. 58, pp. 82–115, 2020
- [5] S. Madakam, V. Lake, V. Lake, V. Lake, *et al.*, “Internet of things (iot): A literature review,” *Journal of Computer and Communications*, vol. 3, no. 05, p. 164, 2015.
- [6] S. Biswas, D. Devi, M. Chakraborty, “A hybrid case based reasoning model for classification in internet of things (iot) environment,” *Journal of Organizational and End User Computing*, vol. 30, pp. 104–122, 10 2018, doi: <http://doi.org/10.4018/JOEUC.2018100107>
- [7] D. Thakker, B. K. Mishra, A. Abdullatif, S. Mazumdar, S. Simpson, “Explainable artificial intelligence for developing smart cities solutions,” *Smart Cities*, vol. 3, pp. 1353–1382, 12 2020, doi: <http://doi.org/10.3390/smartcities3040065>. Number: 4Publisher: Multidisciplinary Digital Publishing Institute.
- [8] I. García-Magariño, R. Muttukrishnan, J. Lloret, “Human-centric ai for trustworthy iot systems with explainable multilayer perceptrons,” *IEEE Access*, vol. 7, pp. 125562–125574, 2019, doi: <http://doi.org/10.1109/ACCESS.2019.2937521>. Conference Name: IEEE Access.
- [9] S. K. Jagatheesaperumal, Q.-V. Pham, R. Ruby, Z. Yang, C. Xu, 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, doi: <http://doi.org/10.1109/OJCOMS.2022.3215676>
- [10] I. Kok, F. Y. Okay, O. Muyanli, S. Ozdemir, “Explainable Artificial Intelligence (XAI) for Internet of Things: A Survey,” *Arxiv:2206.04800*, vol. Preprint, jun 2022, doi: <http://doi.org/10.48550/arxiv.2206.04800>
- [11] M. Caro-Martínez, G. Jiménez-Díaz, J. A. Recio-García, “Conceptual modeling of explainable recommender systems: An ontological formalization to guide their design and development,” *Journal of Artificial Intelligence Research*, vol. 71, pp. 557–589, 2021.
- [12] M. Caro-Martínez, A. Wijekoon, J. A. Recio-García, D. Corsar, I. Nkisi-Orji, “Conceptual modelling of explanation experiences through the iseeonto ontology,” in *CEUR Workshop Proceedings*, vol. 3389, 2023, CEUR Workshop Proceedings.
- [13] T. Speith, “A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods; A Review of Taxonomies of Explainable Artificial Intelligence (XAI) Methods,” p. 12, 2022, doi: <http://doi.org/10.1145/3531146.3534639>
- [14] D. Minh, H. X. Wang, Y. F. Li, T. N. Nguyen, “Explainable artificial intelligence: a comprehensive review,” *Artificial Intelligence Review*, vol. 55, pp. 3503–3568, jun 2022, doi: <http://doi.org/10.1007/s10462-021-10088-y>
- [15] M. Langer, D. Oster, T. Speith, H. Hermanns, L. Kästner, E. Schmidt, A. Sesing, K. Baum, “What do we want from Explainable Artificial Intelligence (XAI)? – A stakeholder perspective on XAI and a conceptual model guiding interdisciplinary XAI research,” *Artificial Intelligence*, vol. 296, p. 103473, jul 2021, doi: <http://doi.org/10.1016/j.artint.2021.103473>
- [16] V. Belle, I. Papanonis, “Principles and Practice of Explainable Machine Learning,” *Frontiers in Big Data*, vol. 4, p. 39, jul 2021, doi: <http://doi.org/10.3389/fdata.2021.688969>
- [17] J. A. McDermid, Y. Jia, Z. Porter, I. Habli, “Artificial intelligence explainability: the technical and ethical dimensions,” *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 379, p. 20200363, oct 2021, doi: <http://doi.org/10.1098/rsta.2020.0363>
- [18] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, P. M. Atkinson, “Explainable artificial intelligence: an analytical review,” *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, vol. 11, sep 2021, doi: <http://doi.org/10.1002/widm.1424>
- [19] G. Vilone, L. Longo, “Classification of Explainable Artificial Intelligence Methods through Their Output Formats,” *Machine Learning and Knowledge Extraction*, vol. 3, pp. 615–661, aug 2021, doi: <http://doi.org/10.3390/make3030032>
- [20] Moamar Sayed-Mouchaweh, “Explainable AI Within the Digital Transformation and Cyber Physical Systems,” *Explainable AI Within the Digital Transformation and Cyber Physical Systems*, 2021, doi: <http://doi.org/10.1007/978-3-030-76409-8>
- [21] W. Samek, K. R. Müller, “Towards Explainable Artificial Intelligence,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11700 LNCS, Springer Verlag, 2019, pp. 5–22, doi: http://doi.org/10.1007/978-3-030-28954-6_1
- [22] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, “A survey of methods for explaining black box models,” *ACM Computing Surveys*, vol. 51, aug 2018, doi: <http://doi.org/10.1145/3236009>
- [23] K. Sokol, P. Flach, “Explainability fact sheets: A framework for systematic assessment of explainable approaches,” *FAT* 2020 - Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 56–67, 1 2020, doi: <http://doi.org/10.1145/3351095.3372870>
- [24] I. Ahmed, G. Jeon, F. Piccialli, “From Artificial Intelligence to Explainable Artificial Intelligence in Industry 4.0: A Survey on What, How, and Where,” *IEEE Transactions on Industrial Informatics*, vol. 18, pp. 5031–5042, aug 2022, doi: <http://doi.org/10.1109/TII.2022.3146552>
- [25] I. Tiddi, M. d’Aquino, E. Motta, “An ontology design pattern to define explanations,” in *Proceedings of the 8th International Conference on*

⁷ <https://github.com/UCM-GAIA/XAIoT>

- Knowledge Capture*, 2015, pp. 1–8.
- [26] S. Chari, O. Seneviratne, D. M. Gruen, M. A. Foreman, A. K. Das, D. L. McGuinness, “Explanation ontology: a model of explanations for user-centered ai,” in *The Semantic Web—ISWC 2020: 19th International Semantic Web Conference, Athens, Greece, November 2–6, 2020, Proceedings, Part II*, 2020, pp. 228–243, Springer.
- [27] S. Chari, O. Seneviratne, M. Ghalwash, S. Shirai, D. M. Gruen, P. Meyer, P. Chakraborty, D. L. McGuinness, “Explanation ontology: A general-purpose, semantic representation for supporting user-centered explanations,”
- [28] H. Elayan, M. Aloqaily, F. Karray, M. Guizani, “Internet of Behavior (IoB) and Explainable AI Systems for Influencing IoT Behavior,” *IEEE Network*, vol. (early access), 2022, doi: <http://doi.org/10.1109/MNET.009.2100500>
- [29] B. M. Keneni, D. Kaur, A. Al Bataineh, V. K. Devabhaktuni, A. Y. Javaid, J. D. Zaiantz, R. P. Marinier, “Evolving Rule-Based Explainable Artificial Intelligence for Unmanned Aerial Vehicles,” *IEEE Access*, vol. 7, pp. 17001–17016, 2019, doi: <http://doi.org/10.1109/ACCESS.2019.2893141>
- [30] S. Ten Zeldam, A. de Jong, R. Loendersloot, T. Tinga, “Automated Failure Diagnosis in Aviation Maintenance Using Explainable Artificial Intelligence (XAI),” *PHM Society European Conference*, vol. 4, jul 2018, doi: <http://doi.org/10.36001/PHME.2018.V4I1.432>
- [31] D. Saraswat, P. Bhattacharya, A. Verma, V. K. Prasad, S. Tanwar, G. Sharma, P. N. Bokoro, Sharma, “Explainable AI for Healthcare 5.0: Opportunities and Challenges,” *IEEE Access*, vol. 10, pp. 84486–84517, 2022, doi: <http://doi.org/10.1109/ACCESS.2022.3197671>
- [32] M. G. Omran, A. P. Engelbrecht, A. Salman, “An overview of clustering methods,” *Intelligent Data Analysis*, vol. 11, no. 6, pp. 583–605, 2007.
- [33] M. Van Otterlo, “A machine learning view on profiling,” in *Privacy, Due Process and the Computational Turn*, Routledge, 2013, pp. 41–64.
- [34] M. Zheng, S. You, L. Huang, F. Wang, C. Qian, C. Xu, “Simmach: Semi-supervised learning with similarity matching,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 14471–14481.
- [35] B. Y. White, J. R. Frederiksen, “Causal model progressions as a foundation for intelligent learning environments,” *Artificial intelligence*, vol. 42, no. 1, pp. 99–157, 1990.
- [36] J. B. Lamy, K. Sedki, R. Tsopra, “Explainable decision support through the learning and visualization of preferences from a formal ontology of antibiotic treatments,” *Journal of Biomedical Informatics*, vol. 104, p. 103407, apr 2020, doi: <http://doi.org/10.1016/J.JBI.2020.103407>
- [37] J. B. Lamy, B. Sekar, G. Guezennec, J. Bouaud, B. Séroussi, “Explainable artificial intelligence for breast cancer: A visual case-based reasoning approach,” *Artificial Intelligence in Medicine*, vol. 94, pp. 42–53, mar 2019, doi: <http://doi.org/10.1016/J.ARTMED.2019.01.001>
- [38] B. Sirmacek, M. Riveiro, “Occupancy prediction using low-cost and low-resolution heat sensors for smart offices,” *Sensors (Switzerland)*, vol. 20, pp. 1–22, sep 2020, doi: <http://doi.org/10.3390/s20195497>
- [39] M. Ryo, B. Angelov, S. Mammola, J. M. Kass, B. M. Benito, F. Hartig, “Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models,” *Ecography*, vol. 44, pp. 199–205, feb 2021, doi: <http://doi.org/10.1111/ecog.05360>
- [40] J. Hatwell, M. M. Gaber, R. M. Atif Azad, “Ada-WHIPS: Explaining AdaBoost classification with applications in the health sciences,” *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, 2020, doi: <http://doi.org/10.1186/s12911-020-01201-2>
- [41] D. Dave, H. Naik, S. Singhal, P. Patel, “Explainable AI meets Healthcare: A Study on Heart Disease Dataset,” *Arxiv:2011.03195*, vol. Preprint, nov 2020, doi: <http://doi.org/10.48550/arxiv.2011.03195>
- [42] D. Wang, Q. Yang, A. Abdul, B. Y. Lim, “Designing theory-driven user-centric explainable AI,” in *Conference on Human Factors in Computing Systems - Proceedings*, 2019.
- [43] S. M. Lundberg, B. Nair, M. S. Vavilala, M. Horibe, M. J. Eisses, T. Adams, D. E. Liston, D. K. W. Low, S. F. Newman, J. Kim, S. I. Lee, “Explainable machine-learning predictions for the prevention of hypoxaemia during surgery,” *Nature biomedical engineering*, vol. 2, pp. 749–760, oct 2018, doi: <http://doi.org/10.1038/S41551-018-0304-0>
- [44] A. Pnevmatikakis, S. Kanavos, G. Matikas, K. Kostopoulou, A. Cesario, S. Kyriazakos, “Risk Assessment for Personalized Health Insurance Based on Real-World Data,” *Risks* 2021, Vol. 9, Page 46, vol. 9, p. 46, mar 2021, doi: <http://doi.org/10.3390/RISKS9030046>
- [45] M. Carletti, C. Masiero, A. Beghi, G. A. Susto, “Explainable machine learning in industry 4.0: Evaluating feature importance in anomaly detection to enable root cause analysis,” *Conference Proceedings - IEEE International Conference on Systems, Man and Cybernetics*, vol. 2019-October, pp. 21–26, oct 2019, doi: <http://doi.org/10.1109/SMC.2019.8913901>
- [46] A. Kharal, “Explainable artificial intelligence based fault diagnosis and insight harvesting for steel plates manufacturing,” *arXiv:2008.04448*, vol. Preprint, 2020.
- [47] M. Sarhan, S. Layeghy, M. Portmann, “An Explainable Machine Learning-based Network Intrusion Detection System for Enabling Generalisability in Securing IoT Networks,” *arXiv:2104.07183*, vol. Preprint, 2021.
- [48] M. M. Alani, “BotStop : Packet-based efficient and explainable IoT botnet detection using machine learning,” *Computer Communications*, vol. 193, pp. 53–62, sep 2022, doi: <http://doi.org/10.1016/J.COMCOM.2022.06.039>
- [49] C. M. Viana, M. Santos, D. Freire, P. Abrantes, J. Rocha, “Evaluation of the factors explaining the use of agricultural land: A machine learning and model-agnostic approach,” *Ecological Indicators*, vol. 131, p. 108200, nov 2021, doi: <http://doi.org/10.1016/j.ecolind.2021.108200>
- [50] P. D’Alterio, J. M. Garibaldi, R. I. John, “Constrained interval type-2 fuzzy classification systems for explainable AI (XAI),” *IEEE International Conference on Fuzzy Systems*, vol. 2020-July, jul 2020, doi: <http://doi.org/10.1109/FUZZ48607.2020.9177671>
- [51] H. Y. Chen, C. H. Lee, “Vibration Signals Analysis by Explainable Artificial Intelligence (XAI) Approach: Application on Bearing Faults Diagnosis,” *IEEE Access*, vol. 8, pp. 134246–134256, 2020, doi: <http://doi.org/10.1109/ACCESS.2020.3006491>
- [52] E. Soares, P. Angelov, X. Gu, “Autonomous Learning Multiple-Model zero-order classifier for heart sound classification,” *Applied Soft Computing Journal*, vol. 94, p. 106449, sep 2020, doi: <http://doi.org/10.1016/j.asoc.2020.106449>
- [53] M. Aghamohammadi, M. Madan, J. K. Hong, Watson, “Predicting Heart Attack Through Explainable Artificial Intelligence,” in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 11537 LNCS, 2019, pp. 633–645, Springer Verlag.
- [54] S. Shams Amiri, S. Mottahedi, E. R. Lee, S. Hoque, “Peeking inside the black-box: Explainable machine learning applied to household transportation energy consumption,” *Computers, Environment and Urban Systems*, vol. 88, p. 101647, jul 2021, doi: <http://doi.org/10.1016/j.compenvurbysys.2021.101647>
- [55] J. Y. Kim, S. B. Cho, “Electric energy consumption prediction by deep learning with state explainable autoencoder,” *Energies*, vol. 12, p. 739, feb 2019, doi: <http://doi.org/10.3390/en12040739>
- [56] A. B. Diallo, H. Nakagawa, T. Tsuchiya, “An Explainable Deep Learning Approach for Adaptation Space Reduction,” in *Proceedings - 2020 IEEE International Conference on Autonomic Computing and Self-Organizing Systems Companion, ACSOS-C 2020*, aug 2020, pp. 230–231, Institute of Electrical and Electronics Engineers Inc.
- [57] I. Kalamaras, I. Xygonakis, K. Glykos, S. Akselsen, A. Munch-Ellingsen, H. T. Nguyen, A. J. Lepperod, K. Bach, K. Votis, D. Tzovaras, “Visual analytics for exploring air quality data in an ai-enhanced iot environment,” in *Proceedings of the 11th International Conference on Management of Digital EcoSystems, MEDES '19*, New York, NY, USA, 2020, p. 103–110, Association for Computing Machinery.
- [58] G. Graham, N. Csicsery, E. Stasiowski, G. Thouvenin, W. H. Mather, M. Ferry, S. Cookson, J. Hasty, “Genome-scale transcriptional dynamics and environmental biosensing,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, pp. 3301–3306, feb 2020, doi: <http://doi.org/10.1073/pnas.191300311>
- [59] Q. Zheng, H. Delingette, N. Ayache, “Explainable cardiac pathology classification on cine MRI with motion characterization by semi-supervised learning of apparent flow,” *Medical Image Analysis*, vol. 56, pp. 80–95, nov 2018, doi: <http://doi.org/10.48550/arxiv.1811.03433>
- [60] R. Porto, J. M. Molina, A. Berlanga, M. A. Patricio, “Minimum Relevant Features to Obtain Explainable Systems for Predicting Cardiovascular Disease Using the Statlog Data Set,” *Applied Sciences* 2021, Vol. 11, Page 1285, vol. 11, p. 1285, jan 2021, doi: <http://doi.org/10.3390/APP11031285>
- [61] L. Brunese, F. Mercaldo, A. Reginelli, A. Santone, “Explainable Deep Learning for Pulmonary Disease and Coronavirus COVID-19 Detection from X-rays,” *Computer Methods and Programs in Biomedicine*, vol. 196, p.

- 105608, nov 2020, doi: <http://doi.org/10.1016/J.CMPB.2020.105608>
- [62] S. Tabik, A. Gomez-Rios, J. L. Martin-Rodriguez, Sevillano-Garcia, M. Rey-Area, D. Charte, E. Guirado, J. L. Suarez, J. Luengo, M. A. Valero-Gonzalez, P. Garcia-Villanova, E. Olmedo-Sanchez, F. Herrera, "COVIDGR Dataset and COVID-SDNet Methodology for Predicting COVID-19 Based on Chest X-Ray Images," *IEEE Journal of Biomedical and Health Informatics*, vol. 24, pp. 3595–3605, dec 2020, doi: <http://doi.org/10.1109/JBHI.2020.3037127>
- [63] M. Shamim Hossain, G. Muhammad, N. Guizani, "Explainable AI and mass surveillance system-based healthcare framework to combat COVID-19 like pandemics," *IEEE Network*, vol. 34, no. 4, pp. 126–132, 2020, doi: <http://doi.org/10.1109/MNET.011.2000458>
- [64] N. Gozzi, L. Malandri, F. Mercurio, A. Pedrocchi, "XAI for myo-controlled prosthesis: Explaining EMG data for hand gesture classification," *Knowledge-Based Systems*, vol. 240, p. 108053, mar 2022, doi: <http://doi.org/10.1016/J.KNOSYS.2021.108053>
- [65] Y. Hsiu Chou, S. Hong, Y. Zhou, J. Shang, M. Song, H. Li, "Knowledge-shot learning: An interpretable deep model for classifying imbalanced electrocardiography data," *Neurocomputing*, vol. 417, pp. 64–73, dec 2020, doi: <http://doi.org/10.1016/J.NEUCOM.2020.07.045>
- [66] C. Panigutti, A. Perotti, D. Pedreschi, "Doctor xai: An ontology-based approach to black-box sequential data classification explanations," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, FAT* '20*, New York, NY, USA, 2020, p. 629–639, Association for Computing Machinery.
- [67] B. C. Kwon, M. J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, J. Choo, "RetainVis: Visual Analytics with Interpretable and Interactive Recurrent Neural Networks on Electronic Medical Records," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 299–309, jan 2019, doi: <http://doi.org/10.1109/TVCG.2018.2865027>
- [68] M. Graziani, V. Andrearczyk, M. M. S., H. Müller, "Concept attribution: Explaining CNN decisions to physicians," *Computers in Biology and Medicine*, vol. 123, p. 103865, aug 2020, doi: <http://doi.org/10.1016/J.COMPBIOMED.2020.103865>
- [69] I. Rio-Torto, K. Fernandes, L. F. Teixeira, "Understanding the decisions of CNNs: An in-model approach," *Pattern Recognition Letters*, vol. 133, pp. 373–380, may 2020, doi: <http://doi.org/10.1016/J.PATREC.2020.04.004>
- [70] S. M. Muddamsetty, M. N. Jahromi, T. B. Moeslund, "Expert Level Evaluations for Explainable AI (XAI) Methods in the Medical Domain," *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 12663 LNCS, pp. 35–46, 2021, doi: <http://doi.org/10.1007/978-3-030-68796-03>
- [71] W. S. Monroe, F. M. Skidmore, D. G. Odaibo, M. M. Tanik, "HiHO: accelerating artificial intelligence interpretability for medical imaging in IoT applications using hierarchical occlusion: Opening the black box," *Neural Computing and Applications*, vol. 33, pp. 6027–6038, jun 2021, doi: <http://doi.org/10.1007/s00521-020-05379-4>
- [72] S. M. Lauritsen, M. Kristensen, M. V. Olsen, M. S. Larsen, K. M. Lauritsen, M. J. Jørgensen, Lange, B. Thiesson, "Explainable artificial intelligence model to predict acute critical illness from electronic health records," *Nature Communications* 2020 11:1, vol. 11, pp. 1–11, jul 2020, doi: <http://doi.org/10.1038/s41467-020-17431-x>
- [73] D. R. Chittajallu, B. Dong, P. Tunison, R. Collins, Wells, J. Fleshman, G. Sankaranarayanan, Schwaizberg, L. Cavuoto, A. Enquobahrie, "XAI-CBIR: Explainable ai system for content based retrieval of video frames from minimally invasive surgery videos," in *Proceedings - International Symposium on Biomedical Imaging*, vol. 2019-April, apr 2019, pp. 66–69, IEEE Computer Society.
- [74] T. Mansouri, S. Vadera, "A Deep Explainable Model for Fault Prediction Using IoT Sensors," *IEEE Access*, vol. 10, pp. 66933–66942, 2022, doi: <http://doi.org/10.1109/ACCESS.2022.3184693>
- [75] W. Guo, "Partially Explainable Big Data Driven Deep Reinforcement Learning for Green 5G UAV," *IEEE International Conference on Communications*, vol. 2020-June, jun 2020, doi: <http://doi.org/10.1109/ICC40277.2020.9149151>
- [76] R. Iyer, Y. Li, H. Li, M. Lewis, R. Sundar, K. Sycara, "Transparency and Explanation in Deep Reinforcement Learning Neural Networks," in *AIES 2018 - Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 2018, pp. 144–150.
- [77] N. Mehdiyev, P. Fettke, "Explainable Artificial Intelligence for Process Mining: A General Overview and Application of a Novel Local Explanation Approach for Predictive Process Monitoring," in *Studies in Computational Intelligence*, vol. 937, Springer Science and Business Media Deutschland GmbH, sep 2021, pp. 1–28, doi: <http://doi.org/10.48550/arXiv.2009.02098>
- [78] S. Meister, M. Wermes, J. Stüve, R. M. Groves, "Investigations on Explainable Artificial Intelligence methods for the deep learning classification of fibre layup defect in the automated composite manufacturing," *Composites Part B: Engineering*, vol. 224, p. 109160, nov 2021, doi: <http://doi.org/10.1016/j.compositesb.2021.109160>
- [79] R. Schönhof, A. Werner, J. Elstner, B. Zopcsak, R. Awad, M. Huber, "Feature Visualization within an Automated Design Assessment leveraging Explainable Artificial Intelligence Methods," *Procedia CIRP*, vol. 100, pp. 331–336, jan 2022, doi: <http://doi.org/10.1016/j.procir.2021.05.075>
- [80] K. H. Sun, H. Huh, B. A. Tama, S. Y. Lee, J. H. Jung, S. Lee, "Vision-Based Fault Diagnostics Using Explainable Deep Learning with Class Activation Maps," *IEEE Access*, vol. 8, pp. 129169–129179, 2020, doi: <http://doi.org/10.1109/ACCESS.2020.3009852>
- [81] R. Assaf, A. Schumann, "Explainable deep neural networks for multivariate time series predictions," *IJCAI International Joint Conference on Artificial Intelligence*, vol. 2019-August, pp. 6488–6490, aug 2019, doi: <http://doi.org/10.24963/IJCAL2019/932>
- [82] I. A. Khan, N. Moustafa, D. Pi, K. M. Sallam, A. Y. Zomaya, B. Li, "A New Explainable Deep Learning Framework for Cyber Threat Discovery in Industrial IoT Networks," *IEEE Internet of Things Journal*, vol. 9, pp. 11604–11613, jul 2022, doi: <http://doi.org/10.1109/JIOT.2021.3130156>
- [83] M. Saharkhizan, A. Azmoodeh, A. Dehghantanha, S. Member, K.-K. Raymond Choo, R. M. Parizi, "An Ensemble of Deep Recurrent Neural Networks for Detecting IoT Cyber Attacks Using Network Traffic," *IEEE Internet of Things Journal*, vol. 7, no. 9, 2020, doi: <http://doi.org/10.1109/JIOT.2020.2996425>
- [84] A. Nascita, A. Montieri, G. Aceto, D. Ciunzio, V. Persico, A. Pescapé, "XAI Meets Mobile Traffic Classification: Understanding and Improving Multimodal Deep Learning Architectures," *IEEE Transactions on Network and Service Management*, vol. 18, pp. 4225–4246, dec 2021, doi: <http://doi.org/10.1109/TNSM.2021.3098157>
- [85] M. Zolanvari, Z. Yang, K. Khan, R. Jain, N. Meskin, "TRUST XAI: Model-Agnostic Explanations for AI With a Case Study on IIoT Security," *IEEE Internet of Things Journal*, vol. (early access), 2021, doi: <http://doi.org/10.1109/JIOT.2021.3122019>
- [86] N. Sarathy, M. Alsawwaf, Z. Chaczko, "Investigation of an Innovative Approach for Identifying Human Face-Profile Using Explainable Artificial Intelligence," in *2020 IEEE 18th International Symposium on Intelligent Systems and Informatics (SISY)*, sep 2020, pp. 155–160, Institute of Electrical and Electronics Engineers Inc.
- [87] H. Naeem, B. M. Alshammari, F. Ullah, "Explainable Artificial Intelligence-Based IoT Device Malware Detection Mechanism Using Image Visualization and Fine-Tuned CNN-Based Transfer Learning Model," *Computational Intelligence and Neuroscience*, vol. 2022, 2022, doi: <http://doi.org/10.1155/2022/7671967>
- [88] Z. A. E. Houda, B. Brik, 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, doi: <http://doi.org/10.1109/OJCOMS.2022.3188750>
- [89] M. C. Garrido, J. M. Cadenas, A. Bueno-Crespo, R. Martínez-España, J. G. Giménez, M. Cecilia, "Evaporation Forecasting through Interpretable Data Analysis Techniques," *Electronics (Switzerland)*, vol. 11, feb 2022, doi: <http://doi.org/10.3390/ELECTRONICS11040536>
- [90] N. Kundu, G. Rani, V. S. Dhaka, K. Gupta, S. C. Nayak, S. Verma, M. F. Ijaz, M. Woźniak, "Iot and interpretable machine learning based framework for disease prediction in pearl millet," *Sensors*, vol. 21, p. 5386, aug 2021, doi: <http://doi.org/10.3390/s21165386>
- [91] T. Nowak, M. R. Nowicki, K. Cwian, P. Skrzypczynski, "How to improve object detection in a driver assistance system applying explainable deep learning," in *IEEE Intelligent Vehicles Symposium, Proceedings*, vol. 2019-June, jun 2019, pp. 226–231, Institute of Electrical and Electronics Engineers Inc.
- [92] J. Martinez-Cebrian, M. A. Fernandez-Torres, Diaz-De-Maria, "Interpretable Global-Local Dynamics for the Prediction of Eye Fixations in Autonomous Driving Scenarios," *IEEE Access*, vol. 8, pp. 217068–217085, 2020, doi: <http://doi.org/10.1109/ACCESS.2020.3041606>

- [93] Y. Li, H. Wang, L. M. Dang, T. N. Nguyen, D. Han, Lee, I. Jang, H. Moon, "A deep learning-based hybrid framework for object detection and recognition in autonomous driving," *IEEE Access*, vol. 8, pp. 194228–194239, 2020, doi: <http://doi.org/10.1109/ACCESS.2020.3033289>
- [94] M. P. S. Lorente, E. M. Lopez, A. L. Florez, A. L. Espino, J. A. I. Martínez, A. S. de Miguel, "Explaining Deep Learning-Based Driver Models," *Applied Sciences* 2021, Vol. 11, Page 3321, vol. 11, p. 3321, apr 2021, doi: <http://doi.org/10.3390/APP11083321>
- [95] L. C. Brito, G. A. Susto, J. N. Brito, M. A. Duarte, "An explainable artificial intelligence approach for unsupervised fault detection and diagnosis in rotating machinery," *Mechanical Systems and Signal Processing*, vol. 163, p. 108105, jan 2022, doi: <http://doi.org/10.1016/j.ymssp.2021.108105>
- [96] S. Itani, F. Lecron, P. Fortemps, "A One-Class Classification Decision Tree Based on Kernel Density Estimation," *Applied Soft Computing Journal*, vol. 91, may 2018, doi: <http://doi.org/10.1016/j.asoc.2020.106250>
- [97] L. Lindsay, S. Coleman, D. Kerr, B. Taylor, A. Moorhead, "Explainable Artificial Intelligence for Falls Prediction," *Communications in Computer and Information Science*, vol. 1244 CCIS, pp. 76–84, 2020, doi: <http://doi.org/10.1007/978-981-15-6634-98>
- [98] J. Senoner, T. Netland, S. Feuerriegel, "Using Explainable Artificial Intelligence to Improve Process Quality: Evidence from Semiconductor Manufacturing," vol. 68, pp. 5704–5723, dec 2021, doi: <http://doi.org/10.1287/MNSC.2021.4190>
- [99] T. T. H. Le, H. Kim, H. Kang, H. Kim, "Classification and Explanation for Intrusion Detection System Based on Ensemble Trees and SHAP Method," *Sensors*, vol. 22, p. 1154, feb 2022, doi: <http://doi.org/10.3390/s22031154>
- [100] B. Mahbooba, M. Timilsina, R. Sahal, M. Serrano, "Explainable Artificial Intelligence (XAI) to Enhance Trust Management in Intrusion Detection Systems Using Decision Tree Model," *Complexity*, vol. 2021, 2021, doi: <http://doi.org/10.1155/2021/6634811>
- [101] R. Senatore, A. D. Cioppa, A. Marcelli, "Automatic Diagnosis of Neurodegenerative Diseases: An Evolutionary Approach for Facing the Interpretability Problem," *Information* 2019, Vol. 10, Page 30, vol. 10, p. 30, jan 2019, doi: <http://doi.org/10.3390/INFO10010030>
- [102] K. Iwanicki, "A distributed systems perspective on industrial iot," in *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*, 2018, pp. 1164–1170, IEEE.
- [103] J. H. Yoo, H. Jeong, J. Lee, T.-M. Chung, "Federated learning: Issues in medical application," in *Future Data and Security Engineering: 8th International Conference, FDSE 2021, Virtual Event, November 24–26, 2021, Proceedings* 8, 2021, pp. 3–22, Springer.
- [104] G. Schryen, M. Volkamer, S. Ries, S. M. Habib, "A formal approach towards measuring trust in distributed systems," in *Proceedings of the 2011 ACM symposium on applied computing*, 2011, pp. 1739–1745.
- [105] F. Alzetta, P. Giorgini, A. Najjar, M. I. Schumacher, D. Calvaresi, "In-time explainability in multi-agent systems: Challenges, opportunities, and roadmap," in *Explainable, Transparent Autonomous Agents and Multi-Agent Systems: Second International Workshop, EXTRAAMAS 2020, Auckland, New Zealand, May 9–13, 2020, Revised Selected Papers 2*, 2020, pp. 39–53, Springer.
- [106] R. Dwivedi, D. Dave, H. Naik, S. Singhal, R. Omer, P. Patel, B. Qian, Z. Wen, T. Shah, G. Morgan, et al., "Explainable ai (xai): Core ideas, techniques, and solutions," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–33, 2023.
- [107] G. Friedrich, M. Zanker, "A taxonomy for generating explanations in recommender systems," *AI Magazine*, vol. 32, no. 3, pp. 90–98, 2011.
- [108] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature machine intelligence*, vol. 2, no. 1, pp. 56–67, 2020.
- [109] F. Gedikli, D. Jannach, M. Ge, "How should I explain? A comparison of different explanation types for recommender systems," *International Journal of Human-Computer Studies*, vol. 72, no. 4, pp. 367–382, 2014.
- [110] H. Gómez-Vallejo, B. Uriel-Latorre, M. Sande-Meijide, B. Villamarín-Bello, R. Pavon, F. Fdez-Riverola, D. Glez-Pena, "A case-based reasoning system for aiding detection and classification of nosocomial infections," *Decision Support Systems*, vol. 84, pp. 104–116, 2016.
- [111] I. Hatzilygeroudis, J. Prentzas, "Symbolic-neural rule based reasoning and explanation," *Expert Systems with Applications*, vol. 42, no. 9, pp. 4595–4609, 2015.
- [112] J.-P. Poli, W. Ouerdane, R. Pierrard, "Generation of textual explanations in xai: the case of semantic annotation," in *2021 IEEE International Conference on Fuzzy Systems (FUZZ-IEEE)*, 2021, pp. 1–6, IEEE.
- [113] L. A. Hendricks, Z. Akata, M. Rohrbach, J. Donahue, B. Schiele, T. Darrell, "Generating visual explanations," in *European conference on computer vision*, 2016, pp. 3–19, Springer.
- [114] P. Ivaturi, M. Gadaleta, A. C. Pandey, M. Pazzani, S. R. Steinhubl, G. Quer, "A comprehensive explanation framework for biomedical time series classification," *IEEE journal of biomedical and health informatics*, vol. 25, no. 7, pp. 2398–2408, 2021.
- [115] A. M. Antoniadis, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, C. Mooney, "Current challenges and future opportunities for xai in machine learning-based clinical decision support systems: a systematic review," *Applied Sciences*, vol. 11, no. 11, p. 5088, 2021.
- [116] W. W. W. Consortium, et al., "Owl 2 web ontology language document overview," 2012.
- [117] E. Sirin, B. Parsia, B. C. Grau, A. Kalyanpur, Y. Katz, "Pellet: A practical owl-dl reasoner," *Journal of Web Semantics*, vol. 5, no. 2, pp. 51–53, 2007.
- [118] C. Dindorf, W. Teufl, B. Taetz, G. Bleser, M. Fröhlich, "Interpretability of Input Representations for Gait Classification in Patients after Total Hip Arthroplasty," *Sensors* 2020, Vol. 20, Page 4385, vol. 20, p. 4385, aug 2020, doi: <http://doi.org/10.3390/S20164385>
- [119] I. P. De Sousa, M. M. B. R. Vellasco, E. C. da Silva, "Local Interpretable Model-Agnostic Explanations for Classification of Lymph Node Metastases," *Sensors (Basel, Switzerland)*, vol. 19, jul 2019, doi: <http://doi.org/10.3390/S19132969>
- [120] Z. Chen, F. Silvestri, J. Wang, H. Zhu, H. Ahn, Tolomei, "ReLAX: Reinforcement Learning Agent Explainer for Arbitrary Predictive Models," *International Conference on Information and Knowledge Management, Proceedings*, pp. 252–261, oct 2022, doi: <http://doi.org/10.1145/3511808.3557429>
- [121] A. Lucic, H. Oosterhuis, H. Haned, M. de Rijke, "FOCUS: Flexible Optimizable Counterfactual Explanations for Tree Ensembles," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 5313–5322, jun 2022, doi: <http://doi.org/10.1609/AAAI.V36I5.20468>
- [122] G. Tolomei, F. Silvestri, "Generating Actionable Interpretations from Ensembles of Decision Trees," *IEEE Transactions on Knowledge and Data Engineering*, vol. 33, pp. 1540–1553, apr 2021, doi: <http://doi.org/10.1109/TKDE.2019.2945326>
- [123] Z. Hu, M. Beyeler, "Explainable AI for retinal prostheses: Predicting electrode deactivation from routine clinical measures," *International IEEE/EMBS Conference on Neural Engineering, NER*, vol. 2021-May, pp. 792–796, may 2021, doi: <http://doi.org/10.1109/NER49283.2021.9441194>
- [124] E. Pintelas, M. Liaskos, I. E. Livieris, S. Kotsiantis, P. Pintelas, "Explainable Machine Learning Framework for Image Classification Problems: Case Study on Glioma Cancer Prediction," *Journal of Imaging* 2020, Vol. 6, Page 37, vol. 6, p. 37, may 2020, doi: <http://doi.org/10.3390/JIMAGING6060037>
- [125] S. Sarp, M. Kuzlu, U. Cali, O. Elma, O. Guler, "An interpretable solar photovoltaic power generation forecasting approach using an explainable artificial intelligence tool," in *2021 IEEE Power & Energy Society Innovative Smart Grid Technologies Conference (ISGT)*, 2021, pp. 1–5.
- [126] J.-R. Rehse, N. Mehdiyev, P. Fetteke, "Towards Explainable Process Predictions for Industry 4.0 in the DFKI-Smart-Lego-Factory," *KI - Künstliche Intelligenz*, vol. 33, pp. 181–187, 2019, doi: <http://doi.org/10.1007/s13218-019-00586-1>
- [127] C. W. Hong, C. Lee, K. Lee, M. S. Ko, D. E. Kim, K. Hur, "Remaining Useful Life Prognosis for Turbofan Engine Using Explainable Deep Neural Networks with Dimensionality Reduction," *Sensors* 2020, Vol. 20, Page 6626, vol. 20, p. 6626, nov 2020, doi: <http://doi.org/10.3390/S20226626>
- [128] T. Ponn, T. Kröger, F. Diermeyer, "Identification and Explanation of Challenging Conditions for Camera-Based Object Detection of Automated Vehicles," *Sensors* 2020, Vol. 20, Page 3699, vol. 20, p. 3699, jul 2020, doi: <http://doi.org/10.3390/S20133699>
- [129] A. Schmidt, K. V. Laerhoven, "How to build smart appliances?," *IEEE Personal Communications*, vol. 8, pp. 66–71, 8 2001, doi: <http://doi.org/10.1109/98.944006>
- [130] D. Gunduz, Z. Qin, I. E. Aguerri, H. S. Dhillon, Z. Yang, A. Yener, K. K. Wong, C. B. Chae, "Beyond transmitting bits: Context, semantics, and task-oriented communications," *IEEE Journal on Selected Areas in Communications*, vol. 41, pp. 5–41, 2023, doi: <http://doi.org/10.1109/JSAC.2022.3223408>



[Humberto Parejas-Llanovarced](#)

Humberto Parejas-Llanovarced is a seasoned IT professional, MBA, Lecturer, specialized in Digital Transformation, Agentic AI and Cloud Technologies, and Researcher at the Complutense University of Madrid, Spain. He has got his Ph.D. in Artificial Intelligence and Computer Science at the same university, his work is focused on eXplainable AI (XAI), Case-Based Reasoning (CBR) and Deep Learning (DL).



[Marta Caro-Martínez](#)

Marta Caro-Martínez is a Researcher and Lecturer at the Complutense University of Madrid, Spain. She got her Ph.D. in Artificial Intelligence and Computer Science in 2022 at the same university. With eight years of experience in research, her work focuses on Artificial Intelligence, specifically on XAI, recommender systems, and case-based reasoning. Her expertise also includes machine learning and ontologies. Her publications have appeared at several international conferences and international journals.



[Juan A. Recio-Garcia](#)

Juan A. Recio-Garcia is Professor at the Department of Software Engineering and Artificial Intelligence at the Complutense University of Madrid. Currently he holds the BOSCH-UCM Honorary Chair on Artificial Intelligence applied to Internet of Things. His research is focused on Case-Based Reasoning, Recommender Systems and eXplainable AI (XAI). He is lead investigator of several national-founded projects and has conducted several contracts with companies in the area of AI.