Received 28 July 2023; revised 15 September 2023 and 12 December 2023; accepted 7 May 2024. Date of publication 9 May 2024; date of current version 21 May 2024. The review of this article was arranged by Associate Editor D. Mazur.

Digital Object Identifier 10.1109/OJIA.2024.3399057

Explainable AI for Industry 5.0: Vision, Architecture, and Potential Directions

CHANDAN TRIVEDI [1], PRONAYA BHATTACHARYA [2], VIVEK KUMAR PRASAD [1], VIRAJ PATEL [1], ARUNENDRA SINGH [3], SUDEEP TANWAR [1], RAVI SHARMA [4], SRINIVAS ALUVALA [5], GIOVANNI PAU [6], AND GULSHAN SHARMA [7]

[1]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad 382481, India
[2]Department of Computer Science and Engineering, Amity School of Engineering and Technology, and Research and Innovation Cell, Amity University, Kolkata 700135, India
[3]Department of Information Technology, Pranveer Singh Institute of Technology, Kanpur 209305, India
[4]Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun 248001, India
[5]Department of Computer Science and Artificial Intelligence, SR University, Warangal 506371, India
[6]Faculty of Engineering and Architecture, Kore University of Enna, 94100 Enna, Italy
[7]Department of Electrical Engineering Technology, University of Johannesburg, Johannesburg 2006, South Africa

CORRESPONDING AUTHORS: PRONAYA BHATTACHARYA; SUDEEP TANWAR (e-mail: pbhattacharya@kol.amity.edu; sudeep.tanwar@nirmauni.ac.in)

This work was supported by the Kore University of Enna.

ABSTRACT The Industrial Revolution has shifted toward Industry 5.0, reinventing the Industry 4.0 operational process by introducing human elements into critical decision processes. Industry 5.0 would present massive customization via transformative technologies, such as cyber-physical systems (CPSs), artificial intelligence (AI), and big data analytics. In Industry 5.0, the AI models must be transparent, valid, and interpretable. AI models employ machine learning and deep learning mechanisms to make the industrial process autonomous, reduce downtime, and improve operational and maintenance costs. However, the models require explainability in the learning process. Thus, explainable AI (EXAI) adds interpretability and improves the diagnosis of critical industrial processes, which augments the machine-to-human explanations and vice versa. Recent surveys of EXAI in industrial applications are mostly oriented toward EXAI models, the underlying assumptions. Still, fewer studies are conducted toward a holistic integration of EXAI with human-centric processes that drives the Industry 5.0 applicative verticals. Thus, to address the gap, we propose a first-of-its-kind survey that systematically untangles EXAI integration and its potential in Industry 5.0 applications. First, we present the background of EXAI in Industry 5.0 and CPSs and a reference EXAI-based Industry 5.0 architecture with insights into large language models. Then, based on the research questions, a solution taxonomy of EXAI in Industry 5.0 is presented, which is ably supported by applicative use cases (cloud, digital twins, smart grids, augmented reality, and unmanned aerial vehicles). Finally, a case study of EXAI in manufacturing cost assessment is discussed, followed by open issues and future directions. The survey is designed to extend novel prototypes and designs to realize EXAI-based real-time Industry 5.0 applications.

INDEX TERMS Automation, cobots, cyber-physical systems (CPSs), digital twins (DTs), explainable artificial intelligence (EXAI), Industry 5.0.

NOMENCLATURE

3-D	Three-dimensional.
5G	Fifth generation.
6G	Sixth generation.
AAA	American Automobile Association.
AI	Artificial intelligence.
API	Application programming interface.
AR	Augmented reality.
ASC	Agriculture supply chain.
BC	Blockchain.
BFT	Byzantine fault tolerance.
BPM	Business process management.
CAPEX	Capital expenditure.
Cobots	Collaborative robots.

© 2024 The Authors. This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 License. For more information, see https://creativecommons.org/licenses/by-nc-nd/4.0/

COVID-19	Novel coronavirus.
CP	Control process.
CPPS	Cyber-physical production systems.
CPS	Cyber-physical systems.
DApp	Decentralized applications.
DDoS	Distributed denial of service.
DES	Distributed energy system.
DL	Deep learning.
DT	Digital twin.
DVM	Dew virtual machine.
EHD	Electronic health-care data.
EHR	Electronic health record.
eMBB	Enhanced mobile broadband.
EMR	Electronic medical records.
EOS	Electrooptical system.
ESSs	Energy storage systems.
Ex-AI	Explainable artificial intelligence.
FeMBB	Further eMBB.
HACCAP	Hazard analysis and critical control points.
HDG	Health-care data gateway.
HIPPA	Health Insurance Portability and Accountability Act.
ICT	Information and communication technologies.
IoT	Internet of Things.
IoV	Internet of Vehicles.
IPFS	Interplanetary file systems.
JSON	Javascript object notation.
LED	Logic efficiency design.
LSTM	Long short-term memory.
MEC	Multiaccess edge computing.
MHR	Medical health records.
ML	Machine learning.
mMTC	Massive machine-type communications.
P2P	Peer to peer.
PHRD	Personal health-care record.
PoA	Proof of authority.
PoS	Proof of stake.
PoW	Proof of work.
PSNs	Pervasive social networks.
RL	Reinforcement learning.
RPC	Remote procedure call.
SCM	Supply chain management.
SCs	Smart contracts.
SDN	Software-defined networking.
SPM	Supplier performance management.
UAV	Unmanned aerial vehicles.
UBI	Usage-based insurance.
uRLLC	Ultrareliable low-latency communications.
V2I	Vehicle to infrastructure.
V2V	Vehicle to vehicle.
VR	Virtual reality.

I. INTRODUCTION

Industry 4.0 is characterized by customer-driven and personalized manufacturing, which makes it increasingly important for manufacturers to aim for increased promptitude, productivity, and sustainability [1]. Industry 4.0 orients itself with key principles of self-expansion, self-acknowledgment, and self-tracking [2]. According to the National Institute of Standards and Technology publication report 1500-201 [3], cyber-physical systems (CPSs) combine physical manufacturing processes, networking, and computing elements based on business operations, which include feedback and control. Thus, Industry 4.0 introduced the CPS as the core manufacturing framework component, forming highly interconnected engineered process development. In Industry 4.0, the IoT allows sensor-driven control, collecting massive data from sensors attached to machinery and associated devices [4]. AI models analyze the collected real-time data, which improves operational process efficiency by reducing biases. Some common applications includes IoT-based health care [5], inventory and SCM [6], cloud services [7], UAV communication [8], DTs [9], and others.

In Industry 4.0, the vast amounts of big data and the interconnectedness of machines, processes, and systems used for model training mean that there is a diminished human touch. This leads to challenges in operations, performance, and efficiency. As a solution, the shift toward Industry 5.0 is anticipated. This phase emphasizes a harmonious blend of human intelligence and cognitive computing. Industry 5.0 prioritizes personalization through human-to-machine (H2M) and machine-to-human (M2H) engagements. Consequently, it merges the precision and accuracy of intelligent process design with human creativity and intelligence. The potential for human errors leading to accidents is diminished with Industry 5.0 technologies.

The growth in process automation has made AI make resilient decisions. To support the cause, different ML algorithms are presented. For example, regression techniques allow the assessment of relationships between one independent and one dependent variable [10]. For example, machinery prices can be forecasted if sales and other related data are present. Another useful approach is the support vector machine (SVM) algorithm, often used for data classification and regression analysis, such as time-series data prediction, computer vision, and scientific data processing for diagnosing diseases [11]. Another fundamental algorithm is the naive Bayes model, which uses the classical Bayes formula with the firm premise that the characteristics are conditionally unconstrained [12]. Examples include recommendation systems, spam filtering, and sentiment analysis. The k-nearest neighbor (kNN) classification algorithm is one of the most widely used distance-based techniques for forming groups of data [13]. For example, the lung disease diagnosis expert system [14] uses the kNN algorithm—k-means unsupervised learning algorithm splits the unlabeled dataset into a series of groups by clustering. Cyber-profiling criminals [15], delivery store optimization, image segmentation [16], and insurance fraud detection [17] are just several applications. Another method that utilizes feedback is known as RL. In this method, an agent learns how to act in a specific situation by taking actions and observing the outcomes, whether positive or negative, as seen in bid optimization and online recommendations. Recently, multiple DL methods have been extensively used in various industries. DL exploits numerous hidden layers to alleviate multiple issues—for example, image categorization, speech recognition, and language comprehension. Thus, ML and DL models have applications in the automotive arena, such as advanced driver assistance systems and autonomous driving, as well as outside the car, such as during development, production, sales, and after-sales activities [18]. The convolutional neural network (CNN) is a computer vision/image recognition model of an artificial DL neural network (NN). For example, object detection for self-driving cars [18], social media, face recognition, and image analysis are used in health care. Deep neural networks (DNNs) that mimic the movement of cells in the human brain are recurrent neural networks (RNNs). Stock prices, the direction of the stock market, machine translation, and sentiment analysis are just a few examples.

However, the AI models have inherent limitations. Most models require a large amount of high-quality data for accurate predictions. Moreover, the data collection is questionable, as it might be biased, incomplete, or inaccurate, which affects the model accuracy [19]. The AI models are nontransparent and mostly considered to follow the conventional "black box" semantics, making it difficult to understand the model decision. In critical industrial applications, where machinery faults are common, it can be highly problematic, and thus, output interpretability is paramount. Then, models might have a limited scope (designed to cater to a selected set of problems), and in practical use cases, generalization is important (unforeseen operational faults). In such cases, it would require retraining the model, which increases the operating cost. At last, the model is subject to adversarial attacks; thus, a basic understanding of input-to-output mapping is essential. Thus, the models must be trusted, which can avoid unexpected behavior [10]. The quest has generally shifted toward addressing three key principles of AI models, presented as follows.

1) *Valid AI:* refers to the class of AI models that should produce accurate and reliable results [20]. The models are trained on high-quality data, the nature of problem design is specific, and outputs are thoroughly tested and validated to ensure the model prediction. Use cases of valid AI are the automotive industry, medical diagnosis, and fault isolation in CPSs.
2) *Ethical AI:* These AI models are designed to be fair, transparent, and align with ethical principles during training. These models consider the societal impact of privacy, bias, discrimination, and other legal concerns. It should be proved that models are trained in a manner that has preserved these ethical routines. Examples include biometric authentication, certificate generation, and others [21].
3) *Interpretable AI:* These models are designed to be transparent and explainable. The models should explain the output based on the inputs [22]. They focus on the fundamental question: How does the model arrive at the output from the specified input? They are used in financial applications, health care, and credit-rating/scoring platforms.

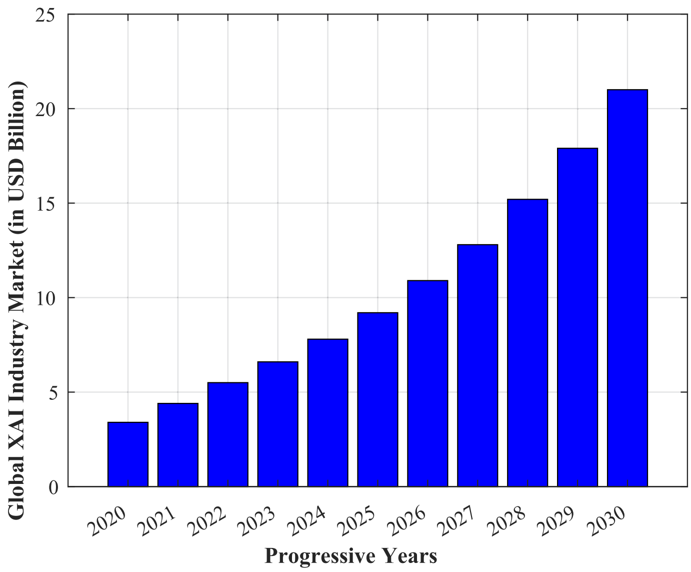

FIGURE 1. Global EXAI investment forecast in Industrial applications by 2030.

EXAI forms the overlap between all these models, and depending on the applicative use cases, it shifts to these approaches (both independently and in combination). EXAI would provide the prediction's self-explanatory basis. EXAI offers several strategies that provide clarification, either through explanations on demand or model illustrations, to avoid restricting the use of the AI model. Fig. 1 shows the global investment in EXAI for industrial applications. By 2030, most of the industry will be using EXAI-based processes in their applications [23].

In EXAI, model transparency is a key concept. It allows the model to reveal its internal operations on the processed inputs. It is critical for Industry 5.0, as it integrates the human element, and thus, models should comprehend how a human works in a particular situation and introspect the outputs. With repeated iterations, the industrial process becomes interpretable [24]. In terms of explanations, there are two forms of the EXAI model: local and global. Local explanations aim to clarify the connection between particular input–output pairs or the logic underlying user inquiry responses. Local explanations are the most useful approach for producing explanations in DNNs. The complete operation of the method is presented in the global explanation. Examples of global explanations include decision trees and model visualization [25].

A. EXAI REALIZATION IN INDUSTRIAL FRONTS

At the industrial front, the potential of EXAI is capped into diverse applicative verticals. For example, the European Union (EU) H2020 project, AI4EU, focuses on designing and developing AI innovations in Europe. One of the front goals is to

TABLE 1. Current Projects and Their Relation With EXAI

Project Name	Duration	Domain	Description	EXAI benefits
Star	2021–2023	Manufacture	STAR assists industrial automation, suppliers, and manufacturer to develop and maintain trusted human-centered AI solutions.	- Inform stakeholders about decisions made with AI. - Explain analysis and autonomy. - EXAI assists in identifying aberrant - behaviors.
XMANAI	2021–present	Manufacture	XMANAI seeks to use EXAI's undeniable power for industrial and human growth.	- Explain concrete (complex) data and model. - Solve manufacturing problems with high impact. - Using hybrid and map moving toward "glass model".
COGLE	2017–2019	Aviation	Users will be able to comprehend the strengths and weaknesses of autonomous systems using COGLE, the system will act in the future and offer suggestions for how to enhance the system's performance.	- AI and human collaboration. - Interpret the accident and provide an explanation for it.
FWF	2020–2023	Medical	They will create a library of explanatory patterns and a novel syntax to combine them based on actual findings. They have.	- Provide an explanation to the human for doing a specific task so that they can work together efficiently.
Cancer prediction on MRI	2022	Medical	Using EXAI identify and segmented affected area in MRI images	- Visual explanation, text-based explanation and Example-based explanation help for understanding reasons for the prediction result.

design EXAI techniques to form a suitable fit for AI models in different industries [26]. Likewise, the United States Defense Advanced Research Project Agency spearheads an EXAI program aimed at elucidating AI models and deciphering outputs for the sensor-driven Internet-of-Battlefield Things [27]. In manufacturing systems, the EU has set up factories of the future, which focuses on design principles of Industry 5.0 in dynamic operating conditions. The project focuses on using EXAI to provide transparency and accountability in the manufacturing process [28].

Similarly, innovative projects are proposed in the autonomous space operations by the National Aeronautics and Space Administration, which proposed EXAI techniques to explain decisions made by autonomous spacecraft [29]. The German Aerospace Center also proposed the explainable robotic assistant project to design explainable rules for robotic assistants on the robot's output actions and decisions while interacting with humans [30]. Thus, EXAI techniques have presented a holistic understanding of the analysis of the behavior of AI models and have addressed novel solutions catered toward the design of valid and ethical systems. Table 1 presents a list of some other real-time projects deployed with EXAI to support different verticals in the Industry 5.0 spectrum.

B. UNIQUENESS OF THE SURVEY

In Industry 5.0, EXAI has become integral to fostering collaboration and partnership between humans and machines. Industry 5.0 is aligned toward hypercustomization, which involves the design of production pipelines that are self-centered, self-healing, and autonomous. Thus, EXAI improves the complex decision-making process of AI models, reduces bias and errors, and improves the safety and reliability of AI-based systems. Recent state-of-the-art (SOTA) surveys on EXAI do not provide a holistic understanding of crucial EXAI models to support Industrial operations. Thus, the motivation for the survey is to present a systematic review of EXAI models, where we present a proposed reference architecture of EXAI in Industry 5.0 and compare it to the traditional AI-driven Industry 4.0 model. We also present a comprehensive solution taxonomy and highlight a unique case study of EXAI for the visualization of machine features to address manufacturing cost assessment.

This survey article covers different aspects of EXAI, which includes techniques, applications, and challenges in implementation in industrial settings. We systematically unfold the benefits of EXAI in Industry 5.0 with large language models (LLMs) and real-time industrial use cases in health care, finance, manufacturing, and transportation. The article will also highlight the challenges and limitations of EXAI, such as the tradeoff between accuracy and interpretability and the need for domain expertise and human input in the EXAI process. Finally, the survey article concludes with a discussion of the future research directions and potential applications of EXAI in Industry 5.0. The survey drives recommendations for future research and development of EXAI in Industry 5.0, including potential areas of innovation and collaboration.

C. SURVEY CONTRIBUTIONS AND ARTICLE STRUCTURE

The key contributions of the survey are as follows.

1) The survey presents different use cases and illustrates the integration of EXAI with Industry 5.0 verticals. Along with a comprehensive review of EXAI advantages, the foundations and core concepts of Industry 5.0 are discussed in detail.
2) An EXAI-assisted Industry 5.0 reference architecture is proposed. We discuss its linked applications' modules and supporting elements with their AI integration.
3) A solution taxonomy of EXAI in Industry 5.0 based on research questions and background study is presented.
4) Research challenges and future directions are outlined, followed by a case study of Industry 5.0 for ad hoc human–machine teaming. Some of the use cases with the key lessons of the survey are outlined.

Fig. 2 presents the structured view of our proposed survey. The Nomenclature presents the list of abbreviations used in the survey preparation.

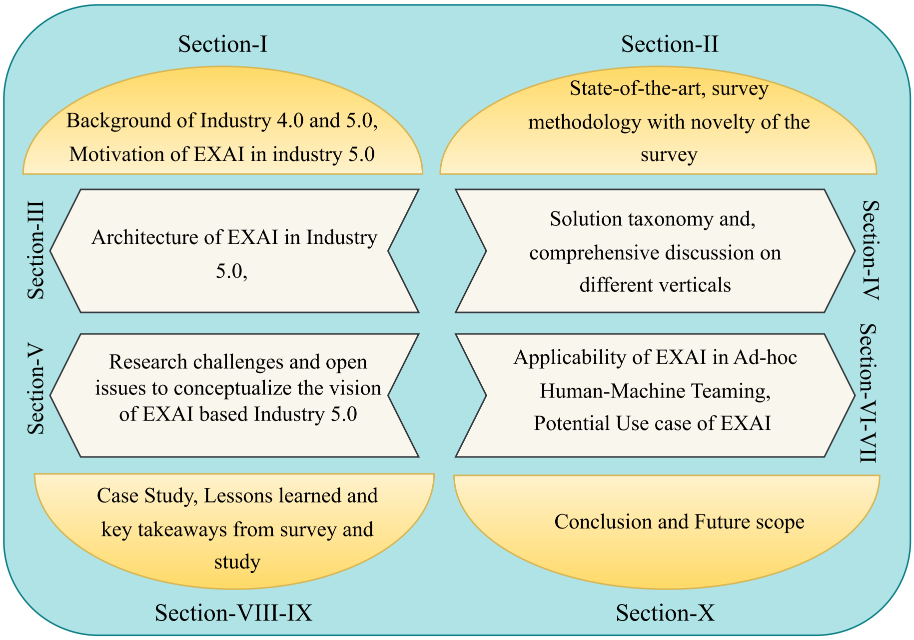

FIGURE 2. Article structure.

II. BACKGROUND AND RELATED WORK

The section presents background terminologies for EXAI, Industry 5.0, and EXAI for CPSs. Then, we present the SOTA surveys on EXAI in industrial applications. Finally, the details are presented as follows.

A. EXPLAINABLE AI

In the 1980s, the term EXAI emerged from expert systems, where we designed rule sets for the expert system to understand and interpret. Due to its simple nature, the concept gained attention among users, but the limited scope of operation could have improved its potential advantage. Thus, it only became popular among the research communities. With the rise of data-driven applications and sophisticated AI models, the need is felt to add understanding to the working of the model [31]. Thus, the need for EXAI interpretation is felt for these models, which became popular among the scientific community. The phrase, EXAI, was initially coined in 2004 by Lent et al. [32], which explained the actions of AI-controlled characters in gaming applications. EXAI research aims to create AI systems that are easy for humans to comprehend, notwithstanding the existing lack of consensus on the idea of explainability and interpretability.

In AI development, stages such as premodeling, model creation, and reverse modeling are instances where explainability can be incorporated. All the phases of AI development can take explainability into account. EXAI market segmentation is done by offering (remedies, facilities), technology (ML, natural language processing (NLP), knowledge application, and others), end-use industry (communications companies, medical services, government sector, retail, transportation, advanced manufacturing, entertainment and media, and others), implementation (cloud), global incentive assessment, and industry forecast, 2020–2030.

The techniques to use EXAI in models mainly include decomposition (breaking a model into its components) and visualization (representing a reason for a model result in the form of a graph or image). Explanation mining is another technique that extracts the relevant data that explains the AI model's prediction. Interpretability approaches distinguish between agnostic models, which are autonomous, and model-specific models, which are only applicable to a particular kind or class of algorithms [33]. Local interpretable model-agnostic explanations (LIME) are used for local interpretability. In contrast, Shapley additive explanations (SHAP) are locally and globally interpretable (the effect of a feature on the target variable). Another technique is a surrogate model, similar to the LIME methodology. In this, the local surrogate models utilize individual data, which explain a complicated model. Another technique is called the partial dependence plot (PDP), which shows how one or more features have little impact on the prediction. These graphs are handy for demonstrating the nonlinearity of the underlying black box model [10]. Rule extraction is used to gain insight into highly complex models. It offers a conclusive and understandable explanation of the knowledge the network gathered throughout its training by placing restrictions on the decision-making process in the artificial neural networks (ANN) by utilizing inputs and outputs [19].

Knowledge distillation is a strategy for training a smaller model (student model) using data from a more extensive model (teacher model). In an NN, layerwise relevance propagation (LRP) propagates a prediction backward using especially created propagation rules. The slight change in feature values that transforms a prophecy into a specified result is a counterfactual explanation. To make a black box model interpretable, prototypes and criticism are utilized. A prototype is a data model representing all data, and criticism is an example of data that a group of prototypes does not effectively describe. The critique aims to provide data and prototypes, particularly data points that prototypes do not adequately depict.

B. INDUSTRY 5.0

Industry 4.0 focused on automating manufacturing processes based on IoT and CPS elements. However, it lacked the involvement of humans in decision processes, and thus, a need is felt to refactor the industrial operations in a human-centric ecosystem. The aim of Industry 5.0 is to model the intelligent and autonomous systems of Industry 4.0 to work alongside humans in a collaborative and complementary manner. It emphasizes including cutting-edge technologies such as AI, robotics, DTs, extended reality (XR), and the metaverse into the production system to augment human abilities and enable human–machine collaboration [34], [35]. The core operations include hypercustomization, massive parallelism, and quality-enabled predictive maintenance. Humans contribute their cognitive and creative abilities to optimize the production process.

In terms of AI models, EXAI adds accountability to the decision capability of CPS elements and human-understandable explanations and reasoning frameworks to model predictions. This is important for Industry 5.0, as it enhances the ability of humans to understand and verify the decisions made by AI systems, thus increasing their trust and adoption. Thus, the CPS has expanded to cyber-physical human systems (CPHSs) in Industry 5.0, allowing seamless collaboration and communication between human workers and intelligent machines. Industry 5.0 presents the concept of long-term sustainability, human–robot collaboration, DTs, and connective interaction for manufacturing and industrial space. For example, certain tasks can be automated in the production pipeline, and inputs can be driven via a twin model. Also, the entire production unit can be visualized in a virtual space created through technologies such as XR and at a large scale, presenting a metaverse-enabled environment for the industrial space. It forms a clear-cut distinction between physical working space and virtual space and the interaction between them.

At the networking front, recent communication networks, such as 5G and beyond networks or 6G network services, offer seamless communication and coordination between different systems and devices in the production process [36]. 5G allows CPHS systems, components, and processes to communicate in real time, with a high density of connection requests. Services such as mMTC in 5G allow large numbers of IoT devices and sensors to collect real-time data on the production process [37], equipment performance, and worker safety. It is particularly useful in applications that require high data throughput and network scalability, such as smart factories [38]. Similarly, 6G services provide more sensing, control, and cognition to CPHS elements. It is expected to enable the deployment of ultradense networks that can support millions of devices per square kilometer (ultra-massive mMTC), thus creating intelligent and autonomous systems that can communicate and coordinate in real time. Moreover, 6G is expected to support deploying advanced technologies such as holographic communications, enabling remote collaboration and training, and immersive AR/VR experiences that can enhance worker safety and productivity in the Industry 5.0 space. 6G supports terahertz communications and thus meets high bandwidth requirements for AR and metaverse space. It also supports holographic communications, which allows control over DTs in Industry 5.0. The service type and network selection depend on the specific use-case application and deployment scenarios.

Industry 5.0 offers massive personalization, enabling customers to choose items tailored to their preferences. This era introduces "Cobots," emphasizing human–robot cooperation. However, the rise in connected systems increases the risk of cyberattacks and data breaches. Thus, a comprehensive security strategy is essential in Industry 5.0, covering all stages, from developing to deploying intelligent systems. The security measures involve traditional crypto-primitives such as encryption, authentication, and access control [39]. In the case of geographically distributed industrial networks, trust is equally important. Thus, BC in Industry 5.0 addresses the issue through trusted ledger maintenance. However, recent trends have suggested using quantum computing in the Industry 5.0 process, which identifies and aligns to secure the most critical assets and implement appropriate security measures to mitigate the risks [40].

C. EXAI FOR INDUSTRY 5.0

AI descriptions play a significant role in helping users comprehend how data are managed [41]. They wish to increase public awareness of potential prejudice and structural flaws. To determine how users perceive fairness in rational judgments, Binns et al. [42] studied explanations in systems for everyday tasks such as calculating premium rates and loan application clearances. Their findings highlight the value of ML explanations in enhancing user understanding and confidence in computational decision-making systems. Rader et al. [43] conducted a crowdsourcing study in related research investigating consumers' awareness of social media algorithms to see how various types of explanations influence users' opinions on news feed algorithmic transparency in a social media network. This study aimed to evaluate algorithmic transparency by measuring user reliability, awareness, and accountability. In addition, they discovered that every explanation raised users' understanding of the system response. Bhattacharya et al. [44] studied fake news classifiers using the fine-grained Q-global vector for word representation semantics and proposed a bidirectional LSTM model. The classified news is stored as a transaction in the BC network. To study the relevant reasons and interactions that ML algorithms could employ to hold people accountable, Stumpf et al. [45] designed experiments. Based on the experiment's findings, they propose utilizing explanations as a potential approach to improve collaboration and information sharing between humans and computers.

D. EXAI FOR CPSS

An expert system, also known as a CPS, is a computer system that utilizes computer-based algorithms to analyze or regulate a mechanism. It includes many industries, such as aviation, automobiles, chemical processes, public utilities, energy, medical, factory, transportation, entertainment, and household products. Radanliev et al. [46] surveyed the application of AI in CPSs. This survey examined how AI prediction in CPSs has evolved. Furthermore, they stated that the model's decision-making process must be transparent to provide more precise information about the industry's revolution.

EXAI has been frequently utilized in CPSs to increase model implementation accuracy and make processes more visible. According to Tao et al. [47], the gaming industry has deployed EXAI in multiview game cheating detection. The goal model combines fraud explainers and fraud classifiers from various perspectives to generate personal, local, and universal explanations that aid in developing facts, reasoning, model debugging, and model compression. The use of EXAI techniques, as suggested by Nascita et al. [48], has revolutionized the use of DL traffic classifiers to evaluate traffic management systems. The goal of developing such a model is to overcome the disadvantages and impossibility of using black box approaches in situations where result reliability or policy explainability is required due to their lack of interpretability. Sarathy et al. [49] investigate the identification of human faces based on their side profile by extracting facial characteristics, assessing feature sets using geometric ratio expressions, and implementing the EXAI technique for efficiency enhancement.

E. RELATED WORK

The existing surveys are discussed in this section, and Table 2 presents the objective, advantages, and disadvantages. Alicioglu and Sun [50] classified EXAI classifiers based on the life circle of a classifier in their survey paper published in 2017. In 2018, Došilović et al. [60] described EXAI in supervised learning and its links with artificial general intelligence. Furthermore, for scholars and practitioners interested in studying the fundamentals of the new and quickly growing research subject on EXAI, Adadi and Berrada [19] provided a starting point. Their survey paper presents EXAI-related concepts, EXAI application domain, and EXAI method taxonomy. In 2019, Clinciu and Hastie [24] introduced EXAI terminologies, such as transparency, intelligibility, interpretability, and explainability. As a result, disagreements about EXAI terminology were resolved.

In 2020, Danilevsky et al. [51] introduced methods for arriving at and visualizing explanations, as well as the main classification of explanations. They discussed the operations and explainability strategies currently available for creating explanations for NLP model predictions to serve as a resource for model developers in the community. Furthermore, Tjoa and Guan [52] have categorized interpretability based on different research, and it helps to learn the complex patterns in interpretability. This survey paper used categorized interpretability in medical research. Moreover, Calegari et al. [53] illustrated symbolic/subsymbolic integration techniques in their survey paper. Prior initiatives were evaluated by Li et al. [10] from the standpoint of data and knowledge engineering (DKE). They categorize the techniques as data-driven explanations derived from task-related data and knowledge-aware algorithms that take into account superfluous knowledge.

In 2021, Čyras et al. [54] developed EXAI approaches founded on computational argumentation methodologies that use a diverse set of reasoning abstractions and explanation delivery mechanisms. On the other hand, a summary of EXAI techniques that have been applied to time series was offered by Rojat et al. [55]. Sahakyan et al. [56] provided an in-depth and current overview of EXAI methods that can be used with tabular data and started the EXAI literature description in the context of tabular data. Islam et al. [57] indicated competitive advantages from various angles (such as local and global), compared and associated the competitive advantage of various EXAI methods, offered EXAI methods with a shared case study, and suggested ways to build responsible or human-centered AI using EXAI as a medium. Vassiliades et al. [61] showed in their survey work how argumentation might contribute to the enhancement of understandable systems in a range of fields, including health care, litigation, the semantic web, cybersecurity, robotic systems, and some general-purpose systems.

In 2022, Ahmad et al. [62] examined AI- and EXAI-based methods used in the context of Industry 4.0. They explained these approaches, how they are used, why, and where they are used in Industry 4.0. Machlev et al. [58] described how EXAI applies to general power systems. Moreover, they analyzed different data where EXAI is used in the power sector. By doing this, they concluded that LIME and SHAP are mostly used for EXAI in power sectors. Monteath and Sheh [59] explained the systematic review for the evaluation of EXAI with a diagram. Furthermore, they demonstrated co-12 properties, which describe different concepts of EXAI. In this way, it helps evaluate old and new EXAI techniques.

F. REVIEW PLAN

In this section, we present a review plan for our article.

The suggested survey was organized thoughtfully. The following steps are taken when reviewing the literature: 1) decide on the study objectives; 2) look at the sources of information; 3) implement search restrictions on the data; 4) impose assessment criteria (inclusion and exclusion); and 5) evaluate quality. This literature assessment includes works in EXAI, industry, research papers, books, and articles. After determining the quality of the data acquired, the survey's pertinent

TABLE 2. SOTA in EXAI-Industry 4.0/5.0

Authors	Year	Objective	1	2	3	4	5	6	Pros	Cons
Alicioglu and Sun [50]	2017	The paper surveys various challenges due to the lack of explainability and visualization for explainable classifiers	N	N	N	N	Y	Y	It describes different explainable classifiers, Visualization for classifiers, and interpretable classifiers	applications of EXAI and relation with industry are not discussed
Danilevsky et al. [51]	2018	This paper summarizes EXAI in supervised learning and recommends future research directions.	N	N	N	N	N	Y	It contains information about explainability, as well as a diagram of the performance transparency trade-off	EXAI architecture, how EXAI is related with the different sectors of the industry is not provided
Clinciu and Hastie [24]	2019	This paper survey is based on the EXAI terminology and Black box problem	N	Y	N	N	N	Y	It shows the black box problem in AI and describes the terminology associated with EXAI	Applications of EXAI, methods of EXAI not shown in this paper
Tao et al. [47]	2020	This paper states EXAI in the domain of NLP	N	Y	N	N	Y	Y	It shows accessible operations and explainability strategies for generating explanations for NLP model predictions	Model of EXAI, implementation of EXAI, EXAI relationship with industry not represented
Tjoa and Guan [52]	2020	This paper provides interpretable information for studies of complex patterns in medical research	N	N	N	N	N	Y	It shows how EXAI is important in the medical sector, different types of interpretability, interpretability via mathematical relation	The model of EXAI, different terminology associated with EXAI, and Application of EXAI in other industrial sectors are not shown
Calegari et al. [53]	2020	This paper provides an overview of the symbolic/ sub-symbolic integration technique using EXAI	Y	N	N	N	Y	Y	It provides a taxonomy of hybrid(symbolic/sub-symbolic) techniques for EXAI	The architecture of EXAI with the industry and application of EXAI in different sectors, are not described
Li et al. [10]	2020	This paper provides cutting-edge evaluation metrics and explanation applications in industrial practice	Y	N	N	N	Y	Y	It shows the reviewing and taxonomizing of existing efforts from a DKE standpoint and summarizing their	lack of explanation about the black box, terminology of EXAI is not shown
Čyras et al. [54]	2021	The paper states that EXAI approaches are built with computational argumentation methods, leveraging its vast array of reasoning abstractions and explanation delivery methods	N	N	N	N	Y	Y	It provides types of explanation (intrinsic and post-hoc), various models for deploying argumentation-based explanations, a roadmap for the future	Application of the EXAI model in different industrial sectors, the future scope of EXAI, and The architecture of EXAI with industry are not shown
Rojat et al. [55]	2021	This paper provides existing EXAI methods applied to time series and the types of explanations produced	N	Y	N	N	N	Y	It provides an analysis of the impact of explanation methods on providing confidence and trust in AI systems	standard EXAI models, Application of EXAI in another area of industry, and the architecture with EXAI are not described
Sahakyan et al. [56]	2021	This paper gives EXAI literature in the context of tabular data and provides a road map that helps the researcher navigate	Y	N	N	N	N	Y	It provides an EXAI literature in the context of tabular data, explains different EXAI models and techniques for analysis	The architecture of EXAI is not shown
Islam et al. [57]	2021	This paper elicits future research direction toward responsible or human-centric AI systems	N	Y	N	N	N	Y	It provides EXAI methods and example-based explanations	The architecture of EXAI, terminology, and taxonomy are not described
Ahmad et al. [53]	2022	This paper provides details of how what, where, and why EXAI is used in industry	N	N	Y	N	N	Y	It provides information about EXAI terminology, high-stakes industry applications	The architecture of EXAI in industry, taxonomy is not shown
Machlev et al. [58]	2022	This paper describes the use of EXAI in the power industry. It demonstrates EXAI's current and future use in the power industry	Y	N	N	N	N	Y	It illustrates EXAI applications in the power Industry	Not given the description of different terminology Architecture of how EXAI beneficial in power Industry
Monteath and Sheh [59]	2022	This paper evaluates practical tools to completely validate, benchmark, and contrast new and current EXAI techniques	Y	Y	N	N	N	Y	It simultaneously optimizes for interpretability by including quantitative measurements as performance during model training.	Taxonomy, relation with industry is not given

1–Blackbox, 2–Terminologies, 3–Industry 5.0, 4–EXAI architecture, 5–Taxonomy, 6–Explainability.

data were obtained. An organized systematic review can assist researchers in producing fair, impartial, and independent outcomes.

G. DATA SOURCES

Digital libraries (data sources), such as IEEE Xplore, Springer, Research-gate, Elsevier, and many others, have been identified. These libraries have a large and diverse collection of literature, which we used to conduct and purpose of this study. Furthermore, we conducted an extensive survey on the connected subject using additional sources, such as papers, blogs, technical reports, and books.

1) SEARCH CRITERIA

The paper is based on the explainability of AI models and implementations in various industries of 5.0. This paper also addresses the EXAI architecture in Industry 5.0. Fig. 3 depicts the keywords and phrases used for relevant topic searches. We refined our search by including Internet publications, blogs, and accumulated papers.

2) INCLUSION AND EXCLUSION

Evaluating a paper as per our relevant topic was how we began the procedure. First, we scanned our search string that included EXAI on different digital libraries and online blogs. Next, we selected a paper that contains the keywords black box and industry. Then, we used the OR keyword to increase our search related to the article. Finally, we collected documents centered on keywords argumentation and EXAI, EXAI in time series, data-driven and knowledged-aware, EXAI, and NLP, symbolic and subsymbolic, and EXAI, Neurosysmolic and EXAI, and others. Afterward, we excluded papers that were far from our interest.

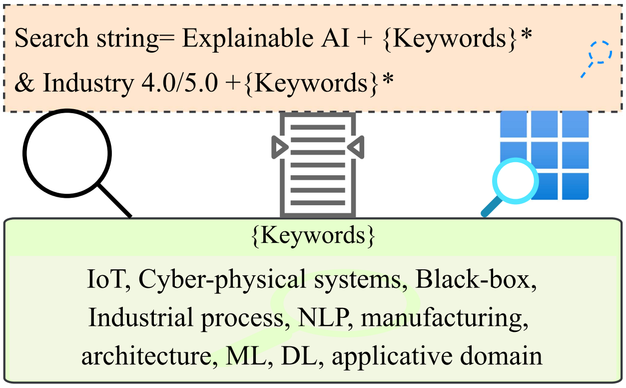

FIGURE 3. Search criteria.

III. EXAI FOR INDUSTRY 5.0: AN ARCHITECTURAL OVERVIEW

This section presents both the traditional AI model and the proposed EXAI-based model from an Industry 5.0 perspective. We also give an overview of the relationship of LLMs in concern with EXAI.

A. TRADITIONAL AI-DRIVEN ARCHITECTURE: A VISION FROM INDUSTRY 4.0 PROCESS CYCLE

Fig. 4 shows traditional AI-driven architecture: a vision from Industry 4.0 process cycle. Here, first of all, an AI-based system is trained based on previous data. If there are no initial data, the process will start with the data collection. In the second stage, different sensors are used in the industry for data gathering. For instance, temperature sensors measure temperature data at different times. Also, humidity and pressure sensors measure water vapors or other gases in the air and pressure levels at a particular time, respectively. Moreover, infrared radiation detects infrared radiation from the surroundings. Optical sensors convert light rays into electrical signals and collect the data in electrical form [63]. One of the latest sensors is the MY THINGS Smart Sensor, a self-contained battery-powered multipurpose IoT sensor that can measure acceleration, temperature, humidity, pressure, and GPS. Afterward, for sending the collected data, use a separate gateway, for example, OPC UA. It is a more secure, open, and dependable data communication method. Furthermore, it is a highly adaptive and versatile technique for transporting data between software systems' controls, monitoring devices, and sensors interacting with real-world data. For collecting large amounts of data, cloud computing is used [64]. This capacity is crucial for storing the data created during a project. It decreases technological resource investment by allowing storage and processing capacity to be contracted on demand, resulting in more flexibility, agility, and adaptability. This allows for cost savings by avoiding the purchase of servers and licenses and hiring specialist employees for maintenance and energy savings. It also allows access to storage from various locations and at various times, regardless of the platform, or connecting devices [65], [66]. After storing the information, it will be sent for data cleaning before applying the artificial model.

In the data cleaning part, data are cleaned using different techniques. For example, remove irrelevant data, clear formatting, fix errors, handle missing values, etc. Afterward, the data will be sent to an AI-based model for training, a model where the system analyzes the data and makes recommendations while improving the correctness of the prediction. For example, the face detection system is trained based on available face images. Here, AI is implemented to train data and improve the performance of the AI model by applying different optimization techniques. A genetic algorithm is an example of it. A genetic algorithm consists of two steps in general. The first step is to choose an individual for the next generation's production. The second step is manipulating that individual using crossover and mutation procedures to create the next generation. The selection mechanism regulates who chooses for reproduction and how many children each chosen individual produces [67]. Simulated annealing (SA) is a popular generic probabilistic solo algorithm for global optimization problems. Each feasible solution in SA is akin to a physical system state, and the fitness function that must be minimized is analogous to the system's internal energy in that state. The ultimate goal is to get the system from an arbitrary (random) initial position to a state where the system's energy is as low as possible [68]. The tabu search algorithm is a metaheuristic algorithm related to evolutionary computing. It solves problems such as the vehicle routing problem, open vehicle routing problem, multitrip vehicle routing and scheduling problem, container loading problem, and job shop problem, among others, [69].

Upon completing the model's training, it will forecast outcomes based on real-world datasets. An alert will be triggered if there are any discrepancies or errors in prediction. For example, the model determines the likelihood of equipment malfunctions during product manufacturing. If equipment fails or a fire occurs, an alert will be activated. The AI system is continuously refined using genuine failure data to enhance its predictive accuracy.

B. PROPOSED REFERENCE ARCHITECTURE: EXAI IN INDUSTRY 5.0 PROCESS CYCLES

Fig. 5 shows the proposed architecture of EXAI in Industry 5.0. It shows how different datasets are gathered from the real world, how they will be processed, how the model will be trained, and how it will explain the prediction of real-world data.

In Industry 5.0, IoT devices and physical assets are used for data collection in the real world. In the industry, there are several sensors, such as temperature sensors, humidity sensors, pressure sensors, and infrared sensors. Many medical instruments, such as computed tomography scans and magnetic resonance imaging (MRI), are used in various medical situations [70], which are part of the IoT. To create a dataset in health care [71], we can collect various patient ailments, the output of medical tools, and doctors' advice [72].

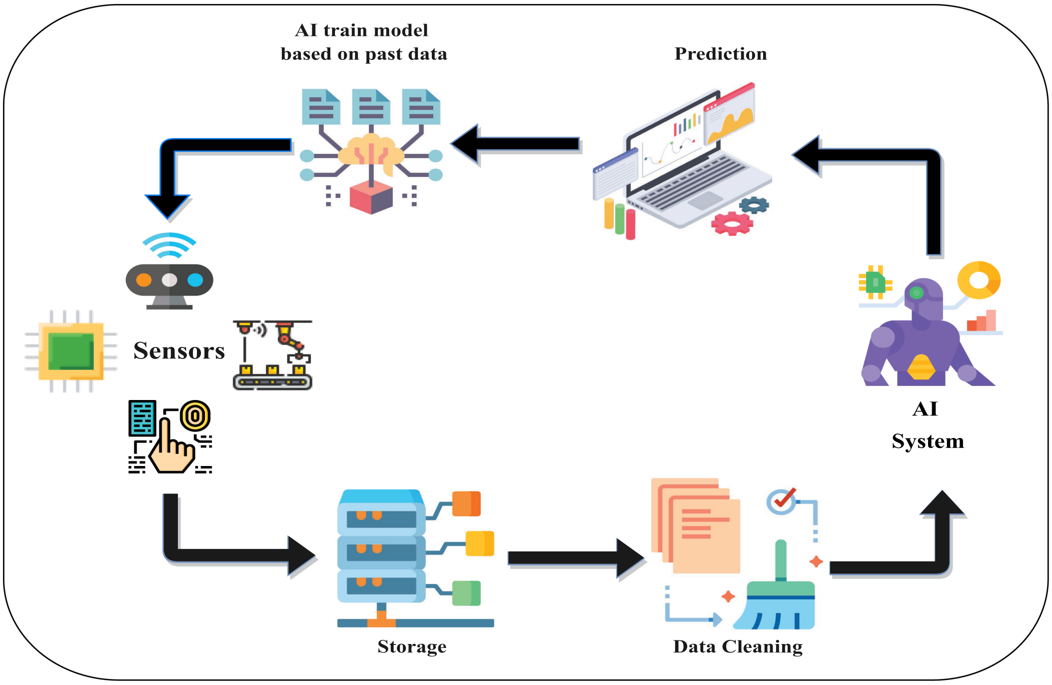

FIGURE 4. Basic architecture of AI.

After the collection of medical data from sensors the next step is preprocessing such that prediction model will be more accurate. The mMTC communication technique has been used for collecting data. It is explicitly designed to collect a large volume of small data packets from a large number of devices at the same time. Moreover, mMTC supports the IoT sensors, which means that data can reduce energy consumption and make work more efficient.

Connection densities of up to one million devices per square kilometer are supported by mMTC. This is more than ten times the capacity of a 4G long-term evolution (LTE) network. As a result, 5G will be able to provide the infrastructure needed to support huge networks of cellular-connected sensors with this capability. Data will be sent to the aggregator via communication. Here, three tasks are performed on the collected data: 1) presentation; 2) cleaning; and 3) collection. It is vital to acquire reliable data when a large volume of data is collected from diverse databases to deliver useful results. Data aggregation can aid in making sensible decisions after being written up as reports; aggregated data can be used to find useful information about a group. By enabling users to comprehend, capture, and visualize data, it also helps with data lineage, which helps monitor the underlying causes of errors in data analytics. It uses two different kinds of data aggregators.

1) *Time Aggregator:* It provides data from a single source over time.
2) *Spatial Aggregation:* It is used as a data point for collecting resources over a predetermined time frame.

The gateway receives aggregated data. Many types of gateways are used for data transfer. There are several types of gateways, including owl4G, wireless local area network (WLAN), low power wide area networks (LPWAN), R3010, industry Wi-Fi, Zigbee, wireless heart, and R3000. The OWL 4G family is specifically designed to provide next-generation security and continuous cellular connectivity in virtually any environment. In WLAN, packets route from a wireless LAN to another network, either a wired or wireless WAN. A WLAN sometimes combines the functions of a wireless access point, a router, and a firewall. The R3010 is an industrial gateway designed for elevator monitoring that offers fast, dependable, and stable Internet connectivity. A ZigBee gateway is a device that allows data to be transferred between a ZigBee network and devices on another network. Finally, WirelessHART provides highly reliable, low latency, and ultralow power connectivity for difficult process monitoring and automation applications.

The data are transmitted from the gateway to the service point. Millimeter wave (mmWave), 5G-eMBB, and 6G are used as service points. The low latency characteristics of 5G mmWave will benefit surveillance and video streaming/broadcasting, as well as the evolution of the 5G smart factory. For cell-edge users, 5G eMBB delivers stable connections with high peak and low data rates. With a packet error rate of the order of 10^{-3} [73], [74], the eMBB service aims to maximize the data rate while retaining a respectable level of dependability. With increased data download speeds, the ability to employ real-time data in industrial operations, teraband reliable low latency communication, and massive network capacity, 6G communication will completely transform international indian ocean expedition (IIoE). Furthermore, a significant amount of data can be retrieved, uploaded, and

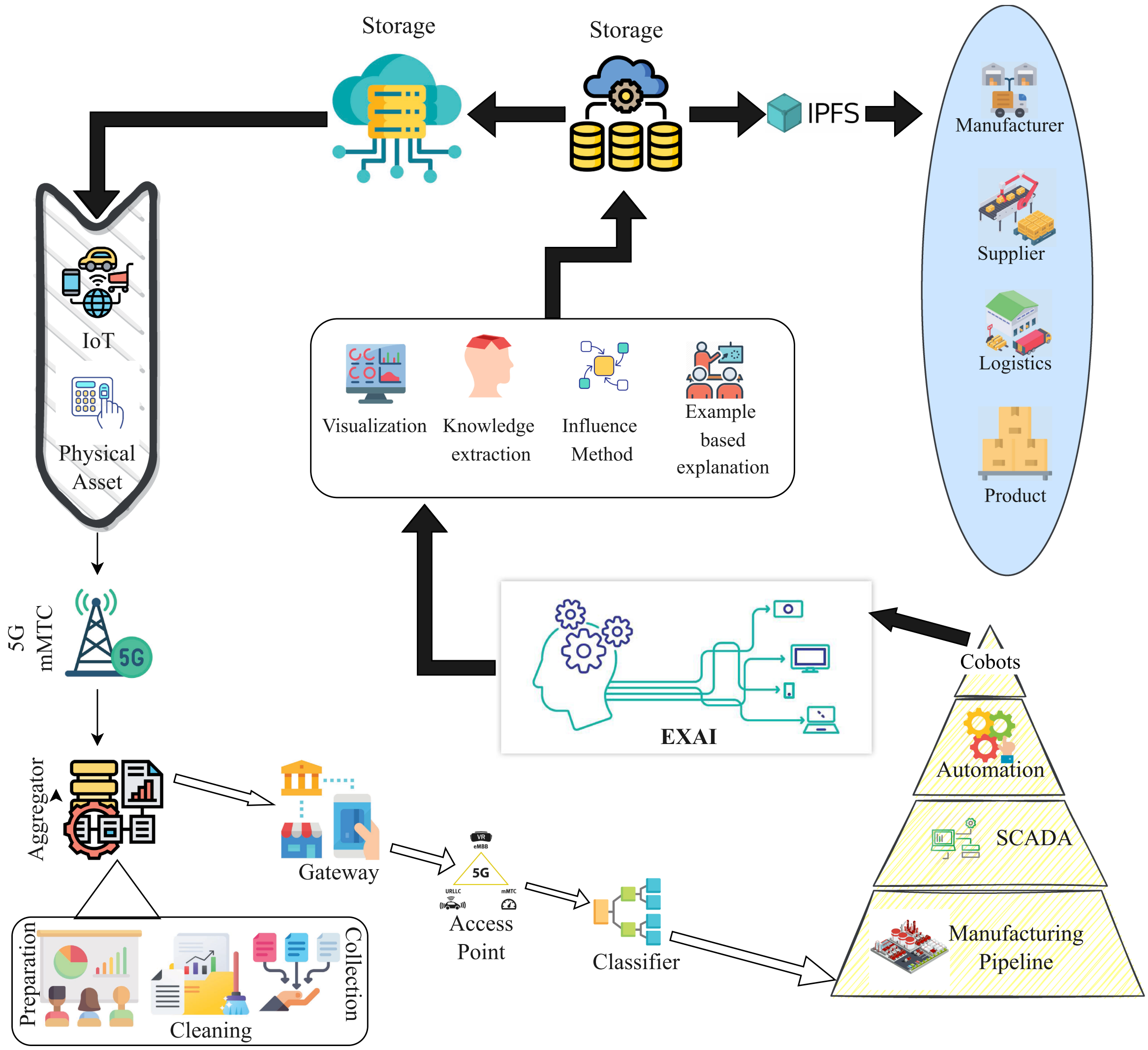

FIGURE 5. **Architecture of EXAI for Industry 5.0.**

exchanged between devices using an edge (cloud) platform [75] at the edge of the 6G network to support integrated and intelligent commercial applications [76].

The service point forwards data to the classifier, which classifies the data based on their specification and feeds them into the required model. It divides into categories: automated, cobots, supervisory control and data acquisition (SCADA), CPSs, process, and manufacturing pipeline. A cobot is a robotic device that manages objects alongside a human operator. It can assist a human operator by producing artificial surfaces that restrict and direct the movement for tasks that require compelling force, such as automobile assembly or critical surgery [77]. The robot operator moves the cobot arms and end-effectors with the touch screen, instructing the cobot how to do different activities and when to switch between them. Cobots can relieve human workers of dangerous, tedious, or laborious tasks. Furthermore, SCADA is a physical and logical solution that enables industrial businesses to control manufacturing processes virtually or locally. Data are monitored, collected, and processed in real time. They are used for data-driven prediction. For instance, Lin and Liu [78] discussed using SCADA system data for current wind energy forecasts, and using a very high-frequency SCADA database with a 1-S sample rate, a DL NN was created to anticipate wind power based on the physical operation of alongshore wind turbines. The manufacturing industry is distinguished by highly sensitive and closely monitored production processes and numerous opportunities for big data analytics. Process operations and controls generate much complex data [79]. Manufacturing pipelines allow companies to build stable yet agile data foundations to build analytics and manufacturing productivity improvement programs. CPSs for industry offer fault-adaptive control systems and human-made analytical models of the system's physics and operations to achieve self-optimization. The CPS can handle massive amounts of data when combined with the right methods [80].

In AI, we care about prediction and do not care about the explanation of prediction. It will be, therefore, problematic for a model to learn from different datasets. That is why architecture provides an EXAI that explains the prediction of the output. There are many techniques to explain the output of the artificial model, e.g., visualization, knowledge extraction, influence methods, and example-based explanation. It is normal to analyze an ML model's representations to investigate the pattern concealed within a neural unit, especially a DNN, to comprehend it. Visualization techniques are primarily used with supervised learning models. There are some visualization techniques in the reviewed literature, including: 1) surrogate models; 2) PDP; and 3) individual conditional expectation. Knowledge extraction is used because it is difficult to explain how ML models work, especially when the models are based on the ANN. However, the underlying layer's cells are altered by learning algorithms, which may result in intriguing internal representations. An explanation must be extracted from the network to obtain the knowledge that an ANN learned during training and encode it as an intelligible internal representation. It primarily responds to two techniques: 1) rule extraction and 2) model distillation. By altering the input or internal components and tracking the impact of the changes in model performance, the influence technique assesses the worth or importance of a feature. Influence-using techniques are typically illustrated. Three alternative methods for determining the input variable's relevance are: 1) sensitivity analysis; 2) LRP; and 3) feature importance. Approaches that use examples to illustrate the performance of ML models choose particular instances from the dataset. Most explanations are model agnostic to encourage the interoperability of any ML model. Two of the most promising example-based interpretability strategies are 1) prototypes and criticisms and 2) counterfactual explanations.

An explanation of the result is stored and further used as a learnable parameter in training a model for IoT devices and other physical devices. Explanation data pass to the storage using the ultrareliable and low-latency communication (URLLC) 5G technique. It consists of eMBB, mMTC that enables a large number of concurrent connections, and URLLC. With end-to-end latencies as low as 5 ms required, the delay budget for individual interfaces can be as low as 1 ms. This means that optimizations must be performed at each uplink and downlink transmission process stage. While not covered by 3GPP specifications, reducing data processing response times drives the development of highly distributed edge computing strategies.

IPFS receives the model prediction. Large digital files should not be shared or stored on BCs [81]. IPFS is a file-sharing technology that can preserve and send huge files more effectively to get around this restriction. It is safe because it uses cryptographic hashes conveniently stored on a BC. Nonetheless, IPFS does not allow users to transfer files to specific individuals [82]. The manufacturer, supplier, logistics, and production departments can use the model's outcome from here. To ensure transparency and control across users, fair consensus mechanisms must be designed where fake node additions are impossible. In this regard, Shymasukha et al. [83] proposed a scheme named *PoRF*, where transactions are ordered and added fairly in the BC network, which reduces the risk of collusion in the network. In Industry 5.0, this offers end-to-end transparency when diverse and heterogeneous users are communicating via the network.

C. EXAI FOR LLMS: INSIGHTS

In Industry 5.0, generative AI has emerged as a game-changing technology with applications spanning from automated design and content creation to data augmentation and simulation [84]. For instance, generative AI algorithms can produce numerous design concepts in the car sector, enabling effective optimization against limitations such as aerodynamics and fuel efficiency. Similar to this, generative models in the pharmaceutical industry can replicate the chemical structures of potential new medication candidates, speeding up the drug discovery process. A subset of generative models emphasizing tasks involving NLP is known as LLMs [85]. Transformer-based models are primarily the architectures that support these LLMs. Transformer designs, along with self-attention mechanisms [86], stand out among these because of their prowess with lengthy sequences and concurrent processing, which makes them extremely scalable. However, due to their complexity, they frequently have many trainable parameters—often hundreds of millions or even billions.

To address the issue of validation of LLMs, EXAI has emerged as a potential solution [87]. EXAI allows us to understand the LLM's inner workings or to give a justification for the predictions they make as they get more complex. In developing industries like Industry 5.0 [88], which aspires to combine human intelligence with machine skills for improved performance in manufacturing, SCM, and quality control, this lack of transparency is especially problematic. Understanding how AI models make decisions in these situations is essential for building trust and responsibility between humans and AI. Techniques such as LIME, SHAP, and attention heatmaps are frequently used as core strategies for obtaining EXAI in LLMs. These techniques are especially pertinent to Industry 5.0, where open decision making can greatly influence productivity and security. For instance, methods such as LIME or SHAP can reveal the characteristics (such as temperature readings or error codes) most important in creating a prediction when a language model is used to evaluate maintenance records to forecast machine breakdowns. Thanks to this transparency, engineers can verify the AI's recommendations, resulting in a stronger collaboration between humans and AI. Other methods include LRP, which backpropagates the output relevance scores to the input layer, allowing for an interpretation of each input feature's contribution. Similarly, counterfactual explanations provide the "what-if" scenarios to show how a slight change in input can lead to a different output, aiding in understanding the model's decision boundary. Other EXAI techniques, such as integrated gradients, involve the path tracing of the model from the baseline input to form a continuous measure of feature importance.

TABLE 3. Overview of LLMs, EXAI Methods, and Their Applications in Industry 5.0

Type of LLM	Model Size	Mechanism	Training Parameters	Hyperparameters	EXAI Method	Industry 5.0 Vertical
GPT-3	175B parameters	Transformer	Batch size: 1024	Learning rate: 0.0001	LRP	Health care: Personalized Treatment
BERT	340M parameters	Transformer	Batch size: 512	Learning rate: 0.002	SHAP	Manufacturing: Process Optimization
XLNet	340M parameters	Transformer	Batch size: 256	Learning rate: 0.0002	LIME	Finance: Portfolio Management
T5	11B parameters	Seq2Seq	Batch size: 512	Learning rate: 0.001	Counterfactual Explanations	Logistics: Supply Chain
ERNIE	29M parameters	Transformer	Batch size: 128	Learning rate: 0.001	Integrated Gradients	Energy: Grid Optimization
RoBERTa	355M parameters	Transformer	Batch size: 1024	Learning rate: 0.0005	Attention-based	Retail: Customer Interaction
ELECTRA	335M parameters	Transformer	Batch size: 256	Learning rate: 0.0002	Decision Attribution	Automotive: Autonomous Driving

Industry 5.0 is significantly improved by integrating EXAI and LLMs, especially in complicated supply chains, predictive maintenance, and real-time decision-making scenarios. These EXAI methods' computational complexity and resource intensiveness, however, raise questions about how scalable and real time they can be. Concerns about data privacy and the moral implications of automated judgments further complicate the implementation of explainable LLMs. While these EXAI techniques are promising, they also have drawbacks that should be considered, particularly in real-time applications. These drawbacks include computational complexity and the accuracy of approximations. For instance, the computational expense of creating explanations using SHAP or LIME would make them unworkable in a manufacturing line that depends on split-second judgments. Furthermore, in high-stakes industrial contexts, these explanations are frequently approximations and might not give a complete picture of the model's decision-making process. Table 3 presents an overview of LLMs and supporting EXAI methods with their applications in Industry 5.0.

The table presents prominent LLMs emphasizing their nuanced attributes, such as model size, inherent mechanisms, and associated hyperparameters. Specifically, models such as GPT-3 and bidirectional encoder representation (BERT) require a massive parameter space, which allows them to form solutions to intricate problem-solving capabilities. These LLMs are paired with corresponding EXAI techniques such as LRP and SHAP to comprehensively view their decision-making processes. It is noteworthy how these EXAI methods are not merely universal but are tailored to the unique computational attributes of each LLM. For instance, LRP is particularly productive for health-care applications where interpretability in personalized treatment plans is imperative. On the other hand, models such as T5, with its sequence-to-sequence architecture, find their EXAI counterpart in counterfactual explanations, offering tangible benefits in logistics and supply chain optimization within the Industry 5.0 framework. XLNet employs a permutation-based training mechanism built upon the transformer architecture, a departure from the masked language modeling approach in BERT. The model's large parameter space and unique mechanism make it compatible with methods such as LIME, which elucidates the predictions by approximating them locally using interpretable models. This pairing is particularly advantageous for complex financial computations, such as portfolio optimization, where interpretability can offer deeper insights into risk assessment.

ELECTRA employs a more efficient pretraining approach, using a replaced token detection task rather than the conventional masked language model or next sentence prediction tasks. This combination benefits the automotive sector, particularly in developing autonomous driving algorithms. Text-to-text transfer transformer (T5) adopts a unified text-to-text framework where every language task is cast as converting input text into target text. Its flexibility and generalized approach make it an ideal candidate for counterfactual explanations, an EXAI method that offers alternate scenarios where the model's decision would change. This makes it exceptionally useful for logistics, specifically in supply chain optimization, where "what-if" scenarios are essential for planning and decision making. Overall, these models, each with their unique characteristics, have the potential to be coupled effectively with specialized EXAI methods, providing not only high performance but also the much-needed interpretability and trust in diverse Industry 5.0 applications.

In a nutshell, the field of EXAI for LLMs is important and challenging. The demand for effective and efficient explanation methods is greater than ever due to the growing usage of these models in various fields, including Industry 5.0. Though current methodologies provide some insights into model behavior, much can still be done to attain complete transparency and reliability in human–AI collaborative contexts.

IV. SOLUTION TAXONOMY

In this section, we present a solution of taxonomy based on EXAI. Fig. 6 shows the proposed solution taxonomy.

A. MACHINE LEARNING

This section presents the ML-based solution for EXAI and its applicative domain. Table 4 represents approaches various literature use with advantages and disadvantages.

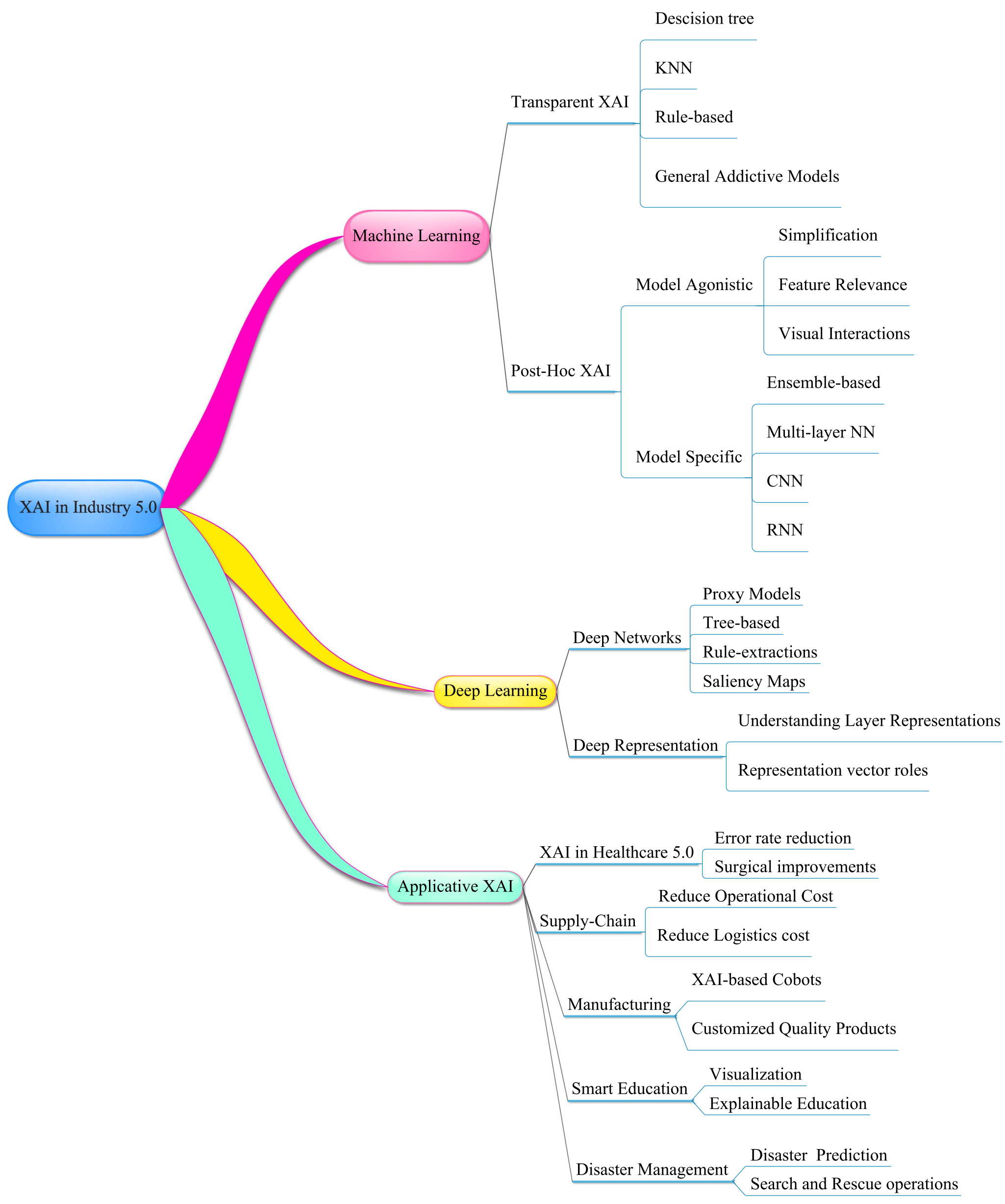

FIGURE 6. **Proposed solution taxonomy of EXAI in Industry 5.0 ecosystems.**

1) TRANSPARENT EXAI

Transparency is the quality of comprehending, justifying, and interpreting an AI system's behavior and its connected components. Furthermore, it categorizes into the following parts.

1) *Decision Tree:* A decision tree provides an explanation based on a hierarchal structure. Monteath and Sheh [59] used decision trees in the EXAI approach to sequential decision support for medical diagnosis, which enables AI systems to collaborate with human specialists while exchanging information and making decisions collaboratively. For example, their system can help doctors decide which testing results are most beneficial based on the available information. The system can also explain a given decision based on the training data. These offer the transparency necessary for fostering patient confidence, adhering to regulations, finding and fixing mistakes, and enhancing patient outcomes.
2) *kNN:* A kNN algorithm works on groups defined by distance or similar quality instances. The information, parameter N, and distance function used to measure similarity impact how transparent a kNN is. A higher value of k affects the user's simulation of the model.
3) *Rule-based:* A rule-based EXAI system aims to produce EXAI with a model exclusively based on predetermined rules. Waa et al. [92] explained that system understanding happens based on two factors. Behavioral factors describe advice prediction and decisive identification

TABLE 4. AI/ML-Based Approaches for EXAI

Approach	Year	Description	Advantage	Disadvantage
Wisdom et al. [89]	2016	The SISTA is proposed for modeling a sequence of correlated observations	SISTA-RNN performs better than state-of-the-art black box RNNs, including long-short term memory (LSTM) RNNs	Outperformance in two black models is insufficient to consider in all cases
Bau et al. [90]	2017	On the interpretability of deep visual representations, networks trained with various initializations	Identify that interpretability is axis dependent phenomenon	Network dissection only investigates the interpretability of CNN network
Shrikumar et al. [91]	2017	Present the DeepLIFT approach, which is a quick and handy way to compute significance scores in a neural network	DeepLIFT models outperform gradient-based techniques by a wide margin	Two similar models with various internal wiring could generate various outputs
Monteath and Sheh	2018	Use a decision tree to relate it with EXAI and apply it to medical diagnosis for decision	EXAI with a decision tree allows us to work together with doctors for medical diagnosis	It only works when we need to do a classification task
Waa et al. [92]	2020	evaluate the effects of two explanation styles on the system in the context of diabetes self-management decision support	Contrastive rule-based explanations pinpoint the situational factor that impacted a system's advice	participants had an overall positive bias toward the system
Yang et al. [93]	2021	An adaptive training algorithm is developed, and GAMI-Net is proposed	GAMI-Net is very understandable and simple to visualize and has the competitive predictive performance to black-box	Consider additional shape constraints for each component function, such as convex or concave

factors. On the other hand, it creates a self-report that illustrates the perceived system understanding. Furthermore, the explanations provided enabled participants to correctly pinpoint the contextual element, which was crucial to a system's recommendation.

4) *General Additive Model:* A general additive model pursues a good balance between prediction accuracy and model interpretability. Several statistically significant constraints are considered to improve the model's interpretability, such as the heredity constraint for requiring structural pairwise interactions, the sparsity constraint for encouraging model similarity, and the marginal clarity constraint for preventing the effects of mixing problems. Yang et al. [93] suggested that the model has a comparable prediction ability to black box ML algorithms. Meanwhile, the model estimated by GAMI-Net is straightforward and understandable to display.

2) POST HOC EXAI

A trained and tested AI model is used as input by a post hoc EXAI approach, which then creates intelligible representations in the form of feature significance ratings, decision rules, plots, or natural language to approximate the inner workings and assessment method of the model.

1) *Model Agonistic:* Model selection is made more flexible by model-agnostic evaluation techniques. Developing a separate assessment framework for each model type is unnecessary. It is divided into three categories.

a) *Simplification:* This refers to the methods used to develop an entirely new system while using the learned model as an explanation. This technique is used to create a new less complex model. Although it is comparable to the original and has the same performance rating, the difficulty level is lower.

b) *Feature relevance:* The inner workings of a model are revealed by computing a relevance score for its controllable variables. These scores indicate the influence a feature has on the model's output. Using feature relevance techniques allows for the indirect explanation of a model.

c) *Visual interaction:* As the section's name implies, the techniques in this segment use graphics to describe opaque models. Many visualization techniques documented in the literature include dimensionality reduction methods that enable a straightforward visualization that is easy for humans to understand.

2) *Model Specific:* Model-specific techniques function based on the particular ML or DL model's applied structures. These methods are employed for model design, such as a CNN. It is divided into the following four categories.

a) *Ensemble-based:* When using a decision tree, the method begins with the set of all the ensemble's rules and discards any that only apply to a few cases. The rule with the lowest error and shortest condition is chosen in each iteration. The cases satisfying this rule are removed from the dataset, and the initial ruleset is updated by the instances still present. Rules that at this point cover few or no instances are discarded and the error of the remaining rules is recalculated. Conversely, a random forest is a set of rules that focuses on the sequential distinction of words or phrases, guiding the expected classification extraction.

b) *Multilayer NN:* Some explainability strategies for multilayer NNs include feature relevance estimators, text explanations, local explanations, and model visualizations. DeepLIFT is a method for calculating significance ratings in a multilayer NN proposed by Shrikumar et al. [91]. It approaches calculating a score based on the

TABLE 5. Approaches Related to DL Domain

Approach	Year	Description	Advantage	Disadvantage
Letham et al. [95]	2015	Employs a novel prior structure to encourage sparsity, making it more difficult to oversimplify the data	model is more accurate but equally interpretable as CHADS2	develops posterior distributions with a strong previous effect. In addition, it has a lot of parameters
Kim et al. [96]	2018	Introduce (CAVs) to enable a human-friendly explanation of an NN's internal state.	provide insight into the forecasts of different classification models	are inadequate on shallower NN's networks
Lundberg et al. [97]	2019	Enhance the readability of tree-based models and use them in the medical field	The enhancement of the interpretability of tree-based ML models has ramifications for a wide range of applications	Not able to provide an accurate result at lower entropy
Doughty et al. [98]	2022	A proxy model technique that is quick to train, accurate when compared to the original model	when access to the model itself is not possible, this method can still be used in situations	It might be less helpful in figuring out how the black box model behaves in edge circumstances

difference between a neuron's activity and the reference activation. Sundararajan et al. [94] raise an important issue with most feature relevance algorithms created for multilayer networks. They demonstrated that most approaches fail two axioms such strategies should uphold: sensitivity and implementation invariance.

c) *CNN:* The internal relationships that make up the CNN structure are incredibly complicated and challenging to understand. Using a different method termed network dissection, Bau et al. [90] quantified the interpretability of the latent representations of CNNs. They analyzed the most prominently active ones by treating each unit as a conceptual detector and assessing each unit further for semantic segmentation. The interpretability of the learned model is another topic covered in this study, along with the consequences of traditional training methods.

d) *RNN:* RNNs are used for prediction problems that are defined over essentially sequential data, and they are especially prominent in time series analysis and NLP. For example, Wisdom et al. [89] suggest an RNN based on the sequential iterative soft-thresholding algorithm (SISTA), which models a series of correlated facts with a sequence of sparse latent vectors, making its weights understandable as the parameters of a systematic statistical model.

B. DEEP LEARNING

This section discusses the DL-based solutions for EXAI and some models supporting the use of DL in EXAI. Table 5 represents DL-based approaches with advantages and disadvantages.

1) DEEP NETWORKS

Key aspects of deep networks include proxy models, tree-based, rule extractions, and saliency maps. We discuss models as follows:

1) *Proxy Models:* A proxy model method trains quickly. It is accurate compared to the prior model and is reliable all the way through. Wood-Doughty et al. [98] used a proxy model to predict time series data. Their proxy model strategy offers unique advantages. Because of its speed, applying table's analysis to a new black box model and dataset is simple, which amply demonstrates the proxy's fidelity to the trained model. After training, the proxy model automatically combines the significance of the features into a comprehensive and understandable explanation.

2) *Tree-based:* To generally explain the individual decisions a tree-based model makes in terms of input contributions, Lundberg et al. [97] introduced TreeExplainer. They contribute an exact Shapley value computation algorithm for tree-based models in polynomial time. For example, it can be essential to distinguish in genetics between (anticipated) outcomes for which two mutations are crucial independently of one another and outcomes for which their combined existence is causal. One can discern between these two situations using TreeExplainer. Finally, using the justifications offered by TreeExplainer, generate a fresh set of tools for a global examination of model behavior.

3) *Rule Extractions:* "If... then..." statements are explanations of rules extraction. To increase interpretability while maintaining accuracy, Letham et al. [95] created Bayesian rule lists (BRLs), a generative model that produces a Gaussian probability over potential decision lists. The IF–THEN rule premise and forecasts are generalized in the rule list as the if, else, and else if rules. The model gets more precise and easier to understand as we add more IF–THEN rules to the decision list. With a lot of conditions, however, explanations lose their backing. BRL attempts to optimize the rules so that the new rule distribution complies with the posterior distribution by selecting a sample rule list from the posterior allocation and continuously adding and changing the rules.

TABLE 6. Applicative EXAI Approaches

Approach	Year	description	Advantage	Disadvantage
Timms [99]	2016	Makes the case that educational Cobots will support instructors in the classrooms of the future and offer instances of present robotics research	exploring innovative interfaces for students, teachers, and supporting technology, and researching the use of the Internet of Things in education	Lack of availability of practical implementation
Defraeye et al. [100]	2019	A digital fruit twin is created to mimic the mango fruit's thermal behavior along the cold chain	By reducing food waste and streamlining operations, DTs can help make the refrigerated supply chain more environmentally friendly	Different surface emissivities and reflections impair accurate temperature measurements
Tremblay et al. [101]	2020	Investigates how AI might help to reduce medical errors	Describes how medical errors are reduced significantly	Not provide any experiment and experimental results
Greif et al. [102]	2020	Develop the concept of a low-cost DT for conventional industries	Uncovering various chances for a seemingly insignificant technical advance to produce informational, automated, and revolutionary corporate benefits	Internet connectivity is necessary across all supply chains and for all technology

New rules may be selected from the prior distribution after optimization.

4) *Saliency Maps:* Among the first attribution techniques created to display the input assessment of the convolutional network is the saliency map. Deep CNNs and the saliency map approach are used in EXAI. Saliency measures can be seen as individual-level contributions using a layer-ordered visualization of information. Numerous EXAI techniques generate saliency maps, emphasizing significant input pixels affecting the prediction. On the other hand, saliency maps concentrate on the input rather than describing how the model decides. This will greatly facilitate their work in particularly data-intensive applications like processing satellite images while maintaining accuracy.

2) DEEP REPRESENTATION

In DL-based solutions, a representation based on the layered approach and vector rule is crucial for understanding its role.

1) *Understanding Layer Representations:* The method used to comprehend each output and explain the output is called understanding layer representations. Representation learning explainability evaluates correlations in the high-dimensional feature space between input and filtered-out versions of itself to offer unambiguous explanations and significantly surpass the gradient-based baselines.
2) *Representation Vector Rule:* A guided line segment is used to represent a vector. The internal representations of AI models, such as DL networks and linear classifiers, are stored as vectors. Therefore, the connection between models and information can be made much simpler by using representation vectors to create predictions. Concept activation vectors (CAVs), which translate an NN's internal state into understandable concepts, are introduced by Kim et al. [96]. The important concept is to see an NN's high-dimensional internal state as a tool rather than a hindrance.

C. APPLICATIVE EXAI

As a part of the solution taxonomy, application-based EXAI can be explored in various verticals, such as health care, SCM, manufacturing, smart education, and disaster management. Table 6 represents the applicative EXAI and its related work.

1) HEALTHCARE 5.0

a) Reduce error rate: There are high hopes for the role of AI in health care, a system in which minor errors can have life-changing consequences. Because future medical errors are expected to result from this interaction, how doctors and AI collaborate becomes a significant component of the clinical environment. Tremblay [101] analyzed the possible role of AI in minimizing medical errors. As developers create AI systems to perform tasks, several hazards and challenges appear, including the danger of fatal injuries brought on by AI system errors, data collection and AI inference violating patient privacy, and more. Some potential solutions include medical practice that will prepare practitioners for transforming roles in an evolving system, integrated monitoring by the food and drug administration (FDA) and other health-care providers, and spending on infrastructure for high-quality representative data. Utilizing EXAI in health care can increase performance and decrease risk.

b) Improve surgery: Most surgeons have doubts about autonomous behaviors during surgery. Yet, unexpectedly, there are a lot of autonomous action examples that already exist and have been around for a while. The emergence of more autonomous operational actions will depend on all of these facets of AI. Yet, only a few surgeons have or are interested in gaining knowledge in this quickly developing subject. In AI surgery, autonomous motions are used. Fortunately, as surgical robots have developed, more doctors are showing an increased understanding of technology and the potential for autonomous activities during procedures, including interventional radiology, endoscopy, and surgery.

2) SUPPLY CHAIN MANAGEMENT

a) Reduce operational cost: Storage and warehousing must be fully integrated into a larger supply chain strategy to reduce costs. Businesses can cut prices and improve space, minimize damage, use much less packaging, and accomplish several other goals by focusing on their processes. Cobots can help SCM companies do this by reducing the cost of ownership. Moreover, automation can help you reduce supply chain expenses and improve the effectiveness of your business operations. We can keep massive sales data without using the cloud computing concepts [103] and retrieve it quickly and affordably by leveraging edge computing technology. A lightweight DT design for use in the construction industry was created by Grief et al. [102]. Their article assessed how DT might lower SCM expenses in the construction sector.

b) Losses during transportation: Defraeye et al. [100] suggested a DT-based approach for simulating the thermal behavior of mango fruit during refrigerated transport. They have also constructed a one-of-a-kind sensing device that mimics fruit to evaluate the temperature model of the fruit pulp. The authors demonstrated the potential of DT in understanding the thermal behavior of fruits throughout the supply chain. Such insights can assist the SCM industry in pinpointing where damages happen when transporting temperature-sensitive fruits, enabling them to implement measures to minimize losses. As a result, DT can aid in improving logistics and refrigeration processes, reducing losses, and achieving a green supply chain.

3) MANUFACTURING PRODUCTION

a) More production using cobots: Workers' injuries and compensation claims are reduced when cobots perform repetitive and arduous jobs efficiently. Workers can then focus on more complex tasks. For example, stamping the same machine part daily might become tedious, resulting in blunders and a loss of interest in the job. In addition, cobots can be placed in hazardous environments (e.g., where gas leaks can be a problem) or inaccessible locations. Cobots inspecting goods in hard-to-reach areas, such as beneath a car's hood, are excellent examples of deployment in challenging settings.

b) Innovation and higher quality product: Introducing a new generation of industrial robots enhances product quality and production line flow, allowing businesses to produce high-quality goods at lower costs. Industry 5.0 enhances manufacturing quality by assigning routine and mundane tasks to robots or machines while reserving tasks requiring critical thinking for humans. Furthermore, it leverages new technologies, such as CPS methods and sophisticated analytical approaches intrinsic to Industry 5.0, to boost production and efficiency.

4) SMART EDUCATION

a) Improve visualization and creativity: Creativity is valued as highly as textbook knowledge in today's schools and colleges. By utilizing EXAI in teaching, we can help students visualize and be more creative. Numerous imaginative activities can be planned where students and robots work together to complete a task, fostering students' inventiveness. For grasping an idea in many subjects, visualizing is necessary. At that point, we can deploy a cobot that supports students' visual learning and collaborates with teachers to assist students in understanding specific topics.

b) Effective, sustainable, and explainable education: Using EXAI techniques, we can analyze at which point students face a problem and how easily we can explain it using different techniques. Timms [99] posits that educational cobots will assist teachers in future classrooms and present examples from existing robotics research. He also imagines smart classrooms with sensors to aid learning and shows how they could be used in new ways if artificial intelligence in education (AIED) apps are integrated into them.

5) DISASTER MANAGEMENT

a) Disaster prediction and mitigation: We can reduce the damage from a disaster if we can forecast it. Sometimes mistakes are made simply by utilizing AI to predict disasters due to a lack of comprehensibility. Their qualitative analysis discovered that Industry 4.0 has few catastrophe recovery and management options. We use EXAI in industry 5.0 to foresee calamities. It will justify the forecast. In addition, it will assist in preparing the prediction models for specific faults. This will make it easier to take certain measures for catastrophe mitigation.

b) Rescue operation: Robots are now used in rescue operations. For example, rescue robots were used in the rescue and response efforts following the 9/11 attacks, the Fukushima Daiichi nuclear disaster, and the Amatrice earthquake in 2016. Humans have no control over a robot. There will be disastrous consequences if the robot's functionality fails. For that, we can use cobots. It works collaboratively with humans. In this way, we can naturalize the disaster more effectively and efficiently.

V. RESEARCH CHALLENGES AND OPEN ISSUES

According to EXAI in the medical field, the problems with the visual and verbal explanations provided by an algorithm such as DNNs need to be exposed. We can resolve this issue by moving the interpretability study away from the algorithm-centric study. Even though it might be the most effective way to reach a consensus, an authoritative organization setting the prerequisites for the deployment of models could impede the advancement of the research itself [104]. In addition, it is easier to deploy complex models on a broad scale with human governance. Therefore, it makes sense to use it at the secondary level. Humanity may only go past this level briefly. Still, it may be time for us to do so since we can gather more data to compare machine forecasts with conventional predictions and solve data ownership problems. In addition, distinct medical environments, such as operating rooms, require special adjustments. It can be solved by developing and implementing a specific framework or auditing approach

TABLE 7. Challenges and Solutions in Various Verticals

verticals	challenges	solutions
Health care	the interpretable algorithm implemented in healthcare needs human supervision. Specialized interfaces for distinct medical settings	Acknowledged that EXAI model is not mature for large scale. Development and use it as a secondary support system. Creating and executing a particular auditing method or framework for EXAI in healthcare.
Manufacture	Due to the complex algorithm, it takes more time to explain predictions. Although cobots can be taught to do intricate and complex tasks, they cannot identify anomalies like variations in raw materials or component location	create encoders or class-specific k-d trees to improve the interpretability of the algorithm. Examine corporate processes for repetitive or simple tasks, and decide whether or not each task is susceptible to cobots
Aviation	trustworthy issues with regulatory authority and organizational leadership. As a coherent explanation of the models, the problem is determining how much testing is required and the success criteria for it	Regulatory authorities must provide sufficient documentation of how specific service schedules use DNNs. XPA is key to identifying the success criteria. depth of explanation, prediction accuracy, approximation, etc., are key XPA
Financial	As models become more extensive and precise, it takes longer to compute explainability algorithms. Due to the lack of applicants with an EXAI background, EXAI specialists will certainly remain sparse. Advice for borrowers whose loans have been declined may or may not be relevant to the borrower's situation	They are recruiting engineering and computer science graduates and training them on explainable AI internally or employing talent outside their businesses. Develop a less time-consuming algorithm so that it will be implemented in this sector more accurately. Train model depends on the borrower's situation; it will help in providing wise advice on the borrower's situation

for EXAI in health care, including protective checks. This will help to ensure the fairness, stability, and fidelity of the EXAI system. We need help explaining projections in the manufacturing sector because of the complicated algorithm. To some extent, we can solve this by developing encoders or class-specific k-d trees to enhance the algorithm's interpretability. The aviation industry depends on trust. Thus, it is crucial to choose more appropriate majors. By applying, we encounter reliable problems with organizational leadership and regulatory power. We can fix a problem if regulatory authorities are required to produce adequate documentation of the process used to generate specific service schedules for predictive maintenance using DNNs. Also, check whether it generates an appropriate result or not. We can develop more suitable algorithms that generate more accurate results for that. Data exchange across various aviation organizations is a further issue in the aviation business. There are several issues when some people use EXAI technologies, and others do not. To stimulate the use of DNNs in predictive maintenance, several aviation organizations and airworthiness authorities must collaborate to promote the EXAI model framework and policies as a mandated design concept. Another issue with a logical explanation of the models in the aviation business is determining the quantity of testing necessary and its success criteria. Exteneded planning and analysis (XPA) is vital to locating the success criteria of a solution. As models become more extensive and precise in the financial sector, it takes longer to compute explainability algorithms. Therefore, we will do further research on developing a less time-consuming algorithm so that it can be implemented in this sector more accurately. Moreover, due to the lack of applicants with an EXAI background, EXAI specialists will certainly remain sparse.

We need more engineers who have appropriate knowledge of EXAI. For that, they are recruiting engineering and computer science graduates straight out of school and training them on EXAI internally or employing talent from outside their businesses. In addition, the borrower's situation may or may not make the advice given to borrowers whose loans have been denied applicable. To fix it, we must train a model dependent on the borrower's circumstances; this will aid in providing sage advice on the borrower's circumstances. In addition, based on pertinent criteria for loan approval, it should produce a result. Table 7 presents the challenges and solutions of different EXAI models for Industry 5.0 verticals.

VI. *AD-HOC-EXAI:* APPLICABILITY OF EXAI IN AD HOC HUMAN–MACHINE TEAMING

This section presents a case study of EXAI in ad hoc human–machine pairing. Team members must be able to coordinate their actions to collaborate effectively. EXAI techniques, which use abstract ideas or descriptions to give the user insight into the AI's logic, strengths and limitations, and behavioral standards, can be used to represent the cobot's behavior policy to the human teammate. This information can help the human teammate predict the future and create a collaboration strategy. As a result, they examine how various abstractions of the AI's policy allow situational awareness (SA) in a scenario where humans and machines collaborate. They also determine if EXAI can support various levels of SA. Finally, they suggest and carry out a study that compares various extractions of the cobot's policy to their stimulated SA levels, evaluating how various explanations can help a human in perceiving the immediate environment (Tier 1), understanding the AI's decision-making framework (Tier 2), and looking ahead to developing a collaborative working strategy (Tier 3). According to their studies, EXAI techniques can help with SA ($p<0.05$).

A. SITUATION

They employ Microsoft's Malmo Minecraft AI Project. In the interactive game Minecraft, users can erect buildings, make tools, and communicate with other players. The cobot has a lot of action primitives and is also coded in a continuous domain. A cobot and a human are charged with building a house that satisfies a set of requirements. The four-story house comprises 89 items, including entrances, steps, a border, fence gates, and two kinds of planks and stone. The following four items require numerous separate foundation ingredients because they are built products. Each agent has a unique set of abilities. The cobot can gather resources and create particular goods. In this area, a policy abstraction that depends on the trial factor allows the cobot and the human to communicate without the person being able to speak directly to the cobot (e.g., via chat). The humans can also see the cobot's display from its first-person perspective, which gives them a limited understanding of its current location, available resources, and course of action. The cobot has complete access to the human state, including its resources, whereabouts, and actions. A static chest that both the cobot and the person can use to store and withdraw resources as required allows both actors to share resources. They include more environmental dynamics. The cobot's policy is hierarchical. A low-level policy governs the cobot's action, while a high-level (which the low-level policy is conditioned upon) policy governs the cobot's inference of human behavior.

B. SITUATIONAL AWARENESS

The harmony of the team's mental models depends on SA, which can also significantly affect efficiency. EXAI brings on the three stages of SA approaches in AI.

1) *Tier 1 (Awareness):* An explanation of the state of the world is provided.
2) *Tier 2 (Understanding):* The agent's current selection is justified.
3) *Tier 3 (Prediction):* It is a statement that enables the user to predict future behavior.

The situation awareness global assessment technique (SAGAT) is frequently utilized to gauge how well SA is upheld.

SA assessment: They respond to a SAGAT assessment in line with the instructions, including recommendations for customizing the SAGAT to evaluate SA in human–machine groups. For Tiers 1, 2, and 3, some questions tested the user's capacity to forecast the cobot's next move, the cobot's response to altered inputs, and the required input to produce a favorable action. In addition, some questions tested the human's comprehension of the cobot's abilities, and the cobot's decision-making policy considers the features. They maintain a sample set of inquiries at every level of SA. Each of the three questions they pose to the user has two to four possible answers and is a multiple-choice question. The supplementary material presents a comprehensive list of questions for each SA level.

C. EXPERIMENTAL CONDITION AND PROCEDURE

They are interested in how the investigation's SA varies depending on the many extractions of the AI's policy. To accomplish this, they employ a 13-in-between design divided into three abstractions: 1) the robot's top-down policy is not explained; 2) the status explains the cobot's hierarchy policy; and 3) a decision tree explains the cobot's hierarchy policy.

1) *No Description:* The human is given no details on the cobot's policy. However, the person with access to both the robot's and the human's first-person displays should point this out, even without an explanation.
2) *Status Explanation:* In the status description, the human is given details about the cobot's policy with the cobot's hierarchical policy results. The information is provided as a brief sentence that encapsulates the details found at the leaf node of the decision tree explanation.
3) *Decision Tree Illustration:* The cobot's policy is displayed to the user, with active links and decision nodes marked. The decision trees display each state feature and aspect that the cobot considers while making a choice.

An online website was used for the experiment, and the initial page gave participants a summary of the human–machine collaboration activity. Users will then engage in three episodes, each representing a different component. Users must watch the 10 min of gameplay in each episode, while the SAGAT interrupts them every 2 min (five trials within each episode). The factor/episode order was varied to alleviate ambiguity.

D. RESULT

Overall, they find that EXAI strategies help people be more aware of their surroundings. Users' capacity to comprehend their environment (tier 1) using or excluding EXAI-based support was similar. This suggests that having a first-person view of teammates and one's gaming is enough to perceive the surroundings. A Friedman's test found significance ($x^2(2)= 34.2$; $p < 0.001$) for Level 2 SA, with pairwise results showing that both status EXAI-based support ($p < 0.05$) and decision-tree EXAI-based support ($p < 0.001$) significantly improve the user's ability to comprehend her environment when compared to cobots without EXAI-based support. Finally, we find that the decision-tree EXAI-based support improves the user's capacity to project cobot behavior into the near future ($p < 0.05$) and has an omnibus significance at tier 3 (F (2, 94) = 4.01; $p < 0.05$). These findings support the theory that EXAI approaches improve SA in human–machine collaboration.

VII. EXAI FOR INDUSTRY 5.0: POTENTIAL USE CASES

The present section deals with potential case studies supporting EXAI's implementation of the latest technology.

A. CLOUD-BASED SMART GOVERNANCE

Innovative governance has been defined as the thoughtful application of ICT and technology to strengthen decision making, management, and citizenry participation through cooperative judgment [105].

1) MOTIVATION AND NECESSITY

Innovative governance ensures the welfare and development of public resources and citizens. Its motivation must be similar to the good governance of the present modern-day democracies [106]. In addition, it must also ensure that ICT is an integral part of the same. The primary issue with the present government is related to unfair policy, corruption, and safety in terms of secure communication, management of resources, transport, and economic infrastructure. These issues above are where intelligent governance can be proven to offer better solutions. To support the decision making for innovative governance, the ICT service is required. The ICT services should be backed with good communications and connections to various devices and support high bandwidth connectivity. The Internet must also support the EXAI, which will lead to better decision making and provide explanations to the system's end users [107]. Explanations and decision-making conditions support real-time actions and operations. For example, the high latency would result in catastrophic conditions related to complex traffic management. This also requires big data operations for safer and faster communications with the various sensor-based devices and ubiquitous gadgets, simultaneously resulting in huge data generation. Smart infrastructure must be given importance to target the outcomes of smart governance. The connected network of the sensor supports the smart infrastructure to communicate the structural information for preservation. This is also used to collect the network data and analysis of the data. The various analysis is done with the help of AI procedures. These AI procedures are helpful for classification, damage detection, prediction of the remaining life, and assessment of condition [108]. This will aim toward minimizing the disruption and saving the maintenance cost. The different case studies where smart monitoring can be associated are wind turbine monitoring, dam monitoring, flood monitoring, ridge monitoring, and maintenance of the subsea valve [109].

2) PRESENT REMEDY AND HOW THE EXAI CAN BE USEFUL

To target the outcomes of smart governance, smart infrastructure must be given importance along with smart monitoring. The connected network of the sensor supports the smart infrastructure to communicate the structural information for preservation [110]. This is also used to collect the network data and analysis of the data. The various analysis is done with the help of AI procedures and smart monitoring. These AI procedures are helpful for classification, damage detection, prediction of the remaining life, and assessment of condition. This will aim toward minimizing the disruption and saving the maintenance cost [111]. The various case studies where intelligent monitoring can be associated are wind turbine monitoring, dam monitoring, flood monitoring, ridge monitoring, and maintenance of the subsea valve [97]. The issue in the current smart monitoring is the need for more transparency in the decision-making process. Hence, the users do not trust the results obtained via the automated decision-making process [112]. Hence, the issue related to the lack of trust can be fulfilled using the enhanced algorithmic decision-making procedure. The utility of the enhanced EXAI will lead to better explanations and provide enhanced transparency in decision-based systems. Making everything transparent to the public is the main objective of governance and governance decision making. Furthermore, EXAI supports the government's accountable public case scenarios such as social cause and public concerns movements. Here, with the accessibility of the EXAI, the public can also get the assurance to the decision-making demand transparency and also the claimed performance [113].

B. EXAI AND DTS

DTs are virtual representations of a system, commodity, or business that allow data processing, simulation-based functioning process monitoring, and performance evaluation [114]. These would effectively manage a product's life cycle while maximizing commercial and manufacturing results. The utility of the EXAI in the DT will provide valuable output for the same and is explained below [115].

1) MOTIVATION AND NECESSITY

DTs are virtual representations of a procedure and a service that allows for operation, data analysis, and monitoring. The performance for the same is carried out through the simulators. This also supports the solutions that manage the business optimization and the product's life cycle. The DTs rely on descriptive capabilities and comprehensive predictive. Afterward, the customers can realize the product's functionality and optimized operational conditions. On the other hand, the manufacturers are associated with maintenance services, which guarantees DTs a profitable and manageable business. The intensive upgrade to the automated CPSs is an advantage of Industry 5.0 [116]. AI incorporates an important role in data accumulation and analysis to maintain its strength. This results in recommendation systems and a good decision-making process [117]. The smarter and faster decision making of the processes and tasks is the primary requirement. Its demand grows when combined with the applications of DTs and robotics, but interpretability and understandability are also required. This indicates and requires explanations of how and why the actions and decisions are taken through the ML/AI models. The sophistication and complexity of AI-based highly automated systems are growing substantially, leading some to believe that humans cannot comprehend the complex mechanics of AI systems, particularly when they offer unanticipated and unexpected conclusions. Cognitive DTs are being introduced to analyze and simulate operation components, resources, technologies, and procedures to advance industrial automation [118]. These technologies include AI, industrial IoT, virtual reality, and big data. Major industries include shipbuilding, steel manufacturing, petroleum, oil, and gas. The DT innovation was also used to improve productivity and safety while lowering operating costs and reducing health hazards and fatalities [119].

2) PRESENT REMEDY AND HOW THE EXAI CAN BE USEFUL

Automation and AI-based decisions can assist system experts and module operators to make decisions properly and quickly when their tasks entail complex systems and procedures. However, the processes and product-related decisions require its authentication in terms of certifications and verification [120]. Therefore, explanations must be provided for these decisions for the AI models associated with the DT and cobots. This should also demonstrate that the consequences are reproducible and traceable. What is the reliability of the prediction? How likely is the system that will crash? And for the AI model, what other stable working circumstances are required? The EXAI can provide the answers to these questions. Hence, the EXAI makes us understand certain features of AI-supported systems and processes. DTs with explanation interfaces and explainable models could be a potential way to increase the credibility of AI in the upcoming industrial era [121]. In experiments, the outcomes of the decision tree and adaptive network-based fuzzy inference model were compared with those of a DL model to verify the effectiveness of EXAI-based explanations. Remarkably, as an accurate virtual version of a major asset, a DT can aid in the explanatory abilities of VR and AR technology [122].

C. UNMANNED AERIAL VEHICLES

UAVs are utilized in various fields, notably defense, business, and public safety. These driverless mechanisms have various benefits over manned aerial transport, including saving fighter jet pilots' lives and lowering costs by preventing human error [121]. In addition, integrating diverse ML approaches is helping the next generation of AI systems to achieve notable success [123].

1) MOTIVATION AND NECESSITY

Recently, automated vehicle machinery has become an integral part of smart cities. These connected autonomous vehicles are important to 6G communications and traffic management. Through wireless communication, autonomous vehicles can connect to other infrastructure and vehicles. This also reduces human involvement in driving scenarios with safety requirements and is thus called nonhuman participation. Various collaborative mechanisms can be incorporated to enhance autonomous vehicle operations [124]. Smart transportation can be identified and implemented in collaboration with automated vehicles and connected vehicle technologies. Autonomous vehicle in transit produces sensor-based information; these data must be collected at the repositories. Based upon this data collection, the surrounding environment of the vehicle communications will be understood, leading to the execution of the activities of the driving forces [121]. In addition to automated vehicles, the development of flying equipment, such as UAVs, also known as drones, has come into existence. The communications of these UAVs are related to adaptive attitude, flexibility, and mobility. Various properties must be mapped to the connectivity scenario, such as reliable data transmission, real-time communication, and high-speed network connectivity. These are to be included in the various subscenarios, such as vehicle-to-everything, vehicle-to-pedestrian, vehicle-to-cloud, V2I, and V2V. Various categories of the autonomous levels of the vehicles are mentioned, such as full automation, high automation, partial automation, assistance during driving, etc. A vehicle must be conscious and know the environmental surroundings during the driving session, and initially, this observes the information and gradually deals with the vehicle control. This is also called a baseline requirement for autonomous vehicles. The driving computing machine works under the data received by the multimodal sensors. These data are then processed for the decision-making process, and it can be achieved using the AI/ML model.

2) PRESENT REMEDY AND HOW THE EXAI CAN BE HELPFUL

With the advent of AI techniques, self-driving must work under various challenges, such as under the conditions of low visibility areas, weather conditions, road surface quality, and some other conditional situations. These different conditions are overcome by using advanced AI concepts and architectures. But still, the trustworthiness of the embedded AI systems in the CAV is uncertain [125]. Hence, an explainable and understandable AI is required to make vehicles (driverless) confident about their routing decisions. Certain conditions lead to the increased risk of accidents, such as the sudden use of the break, encountering changes in the lane, vehicle misjudgment, distraction, and selecting the wrong operation. To deal with such scenarios, autonomous decision-making systems for brake management in emergency conditions were proposed by the researchers [126]. Apart from the mentioned problem, these automated vehicles are also prone to some hidden security issues due to the settings of the AI configuration. These security loopholes can be the main reason for the accidents. Therefore, various AI techniques were implemented for intelligent transportation mechanisms, allocation of resources, interference management, and catching optimization for the UAV-based applications [127].

D. PREDICTIVE MAINTENANCE OF THE INTELLIGENT GRID

The smart grid outperforms the traditional power system regarding renewable energy consumption, information exchange, stability, self-healing, and SA [128], [129]. AI technology, which could also increase decision-making precision and effectiveness, is crucial for supporting the smart grid. EXAI is a prevalent issue in AI because it enables experts to understand, evaluate, and even upgrade ML platforms [130].

1) MOTIVATION AND NECESSITY

In a common power network, electricity is distributed centrally to subscribers. This method is unreliable because electricity or power distribution may be disrupted throughout the network's infrastructure if a power or trip failure occurs along the network's path. To address such difficulties, numerous electric utilities attempt to employ a composite topology of the network or looping in the event of a network failure [49]. Power production has a major environmental impact because it is primarily produced from petroleum-based- fuels, coal, natural sources of gas or biomass from trees and shrubs, and garbage from industry and municipal councils, which results in rising temperatures and also increases the emission of poisonous gas [131]. In addition, some obstacles arise due to increased factors such as electricity theft, security issues, increased demand for power, and power outages. Hence, the traditional electricity grid must be modernized. With the help of an intelligent grid, clients and utilities may communicate in both directions while also being able to sense the transmission system. For the power grid to adapt electronically to the changing demands of end users, the smart grid comprises processors, controllers, new equipment, automation, and technological advances that cooperate. The IoT provides the fundamental framework for the smart grid [50], which is characterized by automated processes, information, and connectivity. This also offers clients a variety of high-quality power supplies safely and efficiently. Handling wireless access for the remote connection among multiple components linked by the smart meters is one of the difficult challenges in making the intelligent grid economically reliable. Implementing a widespread network is necessary for a smart meter. A large range cover is crucial for both the smart grid and the property that satisfies self-healing and averts outages. A primary challenge lies in managing the connectivity systems for the smart grid and the vast data produced from continuous communication. Advanced techniques can be employed to manage these extensive data, facilitating accurate decision making. ML can aid smart grids in collecting data, analyzing data patterns, and determining optimal actions within the smart grid ecosystems. It can also tackle challenges caused by the massive amounts of information generated by connected intelligent and smart grids [132].

2) PRESENT REMEDY AND HOW THE EXAI CAN BE USEFUL

The transmission, generation, regulation, and surveillance of electricity are being transformed by the smart grid's integration which is associated with the communication technology and sensors equipped with the IoT. Therefore, it is essential to deal with the security concerns of the smart grid. To tackle the security challenges of the smart grid, a protected service delivery system that supports the smart grid was presented using the naive Bayes ML algorithm to maintain the smart grid's power usage effectively. The repetitive data in the collected information can be decreased using event-driven random samples. Behara and Saha [133] used the SVM technique to find the capabilities for examining equipment use patterns to solve this problem. Concerning the applicability of 5G-based smart grids, standard AI/ML techniques lack transparency in their forecasts [134]. Keeping the operations of the intelligent grids transparent and accountable through the use of EXAI can be highly beneficial in fostering consumers and generating trust for the prosumers. A forecasting scheme based on an EXAI approach was proposed to predict PV power generation; this also raises the reliability of AI models and thus increases the acceptability of AI in the management of the smart grid. Rožanec et al. [135] identified the Deep-SHAP method; this works on the backpropagation deep explainer system concept. It is based on SHAP that generate an understandable model for emergency process control in intelligent grids [136].

E. XR APPLICATIONS

The term "extended reality" refers to both VR and AR [137]. The technique aims to integrate or replicate the real world with a "DT reality" that can communicate with it. AI successfully self-manages the collaborating equipment in holographic telepresence and multimodal XR technologies enabled by 6G [138].

1) MOTIVATION AND NECESSITY

VR, mixed reality (MR), AR, and other immersive innovations are combined to create XR. By creating an interactive environment or combining the virtual and real worlds utilizing various sensing, these immersive technologies have been applied to extend the real-world experience [137]. In AR, visualizations and content are superimposed on the real world. When used with smart cellphones, laptops, displays, and AR glasses, digital features such as text, animation, and photos improve the users' perspectives as per the actual environment. The Snapchat app's embellishments, which can put hats or spectacles on human heads, the *Pokemon GO* show's overlay of virtual creatures over actual life, and other applications are a few instances of AR. With the head-mounted screen, also known as a VR headset, consumers of VR will be wholly submerged in a digital environment that is a replica and is created to enjoy the view of the artificial world from all angles [139]. XR offers numerous real-world and practical implications in various industries where clients can save time and money traveling, including entertaining, shopping, health care, property investment, advertising, working remotely, and disaster risk management. Holographic telepresence is an advanced method that allows for the projection of an accurate, comprehensive, 3-D model of people far away in space, along with real-time voice communication, giving users the impression that they are speaking with the individuals in person. Other communication-related possibilities, including telemedicine, improved television and film experiences, gaming, promotion, robot control, aeronautical navigation, 3-D modeling, and other simulations, have a great deal of potential to support this technology. Technologies such as holographic telepresence and XR demand a high-speed communication network with almost zero latency and quick sensor data interpretation [140]. 6G can effectively contribute to achieving the genuine benefits of XR thanks to its characteristics, including user-experienced data rate, connection density, dependability, traffic volume density, scalability, and mobility [141].

2) PRESENT REMEDY AND HOW THE EXAI CAN BE USEFUL

AI is employed to autonomously oversee the integrated equipment in holographic telepresence and multimodal XR technologies made possible by 6G. Among the potential applications of AI in these edge devices is the capability to understand the environment through computer vision, analyzing the information derived from the images these devices capture [142]. Permitting AI-based apps in smartphones results in a decrease in network traffic. The following are some possible benefits for AI/ML in holographic telepresence and multisensory XR applications [141].

1) *Virtual Assistance for Evolving Customer Engagements:* Virtual assistants who respond to consumer questions are trained to deliver virtual experience-based assistance, mostly working with the newest trends.
2) *Virtual People:* Training animations to enable real-time responses are an example of the same.
3) *Text Recognition and Translation:* The text extracted from a picture and overlaid into the 3-D environment can be done using the XR programming APIs.
4) *Detection of the Objects:* The entity's size and location within an environment can be approximated to create hitboxes and colliders, which enable interactions between virtual and real entities.
5) *Calculation of Object Position:* The position of the object (such as hands or figures, etc.) could be extrapolated to influence the XR's information.

The predictions and classification made by AI/ML techniques may not inspire much confidence in the decision-making process of the holographic telepresence and mission-critical XR systems because of its black box characteristics [143]. EXAI would close this gap by offering insightful reasoning and justification for the classifications and forecasting process. This might also persuade holographic telepresence applications and XR application developers to decide things based on the outcomes of the AI/ML techniques. For example, the placement of the items can be estimated using AI/ML techniques. However, making judgments based on live AI/ML technology suggestions could occasionally result in erroneous XR material because of incorrect or false positives [144]. According to observations by R. Chengoden et al. [145], EXAI methods and XR models establish safe communication between humans and robots.

F. ZERO-TOUCH NETWORK

Zero-touch communication networks aim to move humans away from the sense–actuation cycle and onto a higher level cycle of intent setting, and monitoring [146]. EXAI in zero-touch networks includes ensuring that client objectives are realized, and contracts are achieved with the utmost care, all while addressing factors of AI security, confidentiality, responsibility, and interpretability, among many others [147].

1) MOTIVATION AND NECESSITY

The introduction of the present business model will unleash the technology behind the software-defined network, network slicing, 5G, and beyond networks. The complexity of the management increases due to the rise in cost efficiency, flexibility, and efficiency due to the incorporation of agility and inter-domain collaborations [148]. To solve the aforementioned problem, the conventional methodologies could be more efficient for management and network-based operations. Hence, this is also foreseeable through closed-loop automation for the operations related to the management. The incorporation of automated management via self-managing mechanisms will lead to the improvement of flexibility and efficiency of the delivered tasks and services. This will also result in the reduction of operating expenses. The 100% automation will be achieved using big data analytics and ML procedures for managing cellular networks [106]. The ML algorithms train themselves from huge data generation. This results in self-management in the network domain (and the same is classified as self-healing, self-configuration, self-optimization, and self-protecting). The outcome of this automation can be observed with a reduction in human errors, a decrease in operational cost, and time to value [106].

2) PRESENT REMEDY AND HOW THE EXAI CAN BE USEFUL

The ML algorithms use the raw data. This also filters the significant measures for the events and is followed by the recognition of the issues in the network. The identified information is then propagated to the upper layers of the ML algorithms. To ensure accountability, transparency, trustworthiness, and reliability in the ML/AI-based mechanisms, the incorporation of ML/AI procedures must be integrated. The advantages of the incorporation will result in fully automated criteria with the invocation of transparency and interpretability [149]. This will also help to achieve to reach the correct decision levels. Ultimately, the end user needs help managing the ML/AI-based result procedures, which happens due to the massive complexity of the ML/AI-based procedures. This mainly happens when a series of updated models are input to the huge volume of the information produced via the 5G and beyond networks models. The incorporation of the EXAI will have the vast potential for dealing with the challenges that occur through the network and the associated ML/AI procedures. The acyclic graph has been proposed for the various taxonomies related to the ML/AI-based models. The concept of the algorithm instant repository was also proposed for identifying the variables related to the output and input and the attributes formation of the ML/AI procedures. The analyst can use these variables and attributes for the reverse engineering process, leading to the result's explanation and solving the problems mentioned before. To make ad hoc decisions, the density-based selection of the feature and neuro-fuzzy-based self-organizing models plays an important role. This will help in the scenarios of speeding up, braking, and changing the lane from right to left and vice versa. Here, the outcome of the model (AI-based) is similar to the if–then rule, which usually will be in the human understandable form. In another study, the rule-based EXAI system has been implemented for different flying techniques of UAVs. Concerning the varying weather conditions, relative enemy location, and the conditions of the nearby environment locations, the decisions of the selected path will be identified with the fuzzy inference technique and with the utility of the if–then rules. Similarly, in the study [150], the random forest technique has been implemented with the EXAI. The merging of the random forest and EXAI results in improved accuracy and self-confident autopilot with the automated and smart navigator scheme. The complicated task in the UAVs is to change in direction and hold itself for a moment to capture some instances. This can be obtained by implementing the AI-based hierarchical model to avoid collision risk and uncertainties. This can also be used for calculating the tracking error, intention recognition, and account perception. Explainability of AI is the primary need for society, customers, manufacturers, and system developers [151].

G. IMPACT OF ZERO TRUST AND ZERO KNOWLEDGE ON EXAI

Zero trust is a security framework that assumes no trust in the underlying infrastructure, requiring verification from anyone trying to access resources. In the context of EXAI, the implementation of zero trust principles can enhance the security and integrity of the explainability mechanisms. This is crucial as EXAI involves the interpretation of complex models, and ensuring the trustworthiness of these explanations is paramount. Implementing zero trust in EXAI could involve rigorous authentication and authorization processes, continuous monitoring of model behavior, and the establishment of secure communication channels. By adopting a zero-trust approach, organizations can mitigate the risk of unauthorized access to sensitive AI models, protecting against potential threats and ensuring that explanations provided by the model are reliable and secure. However, zero knowledge is a cryptographic concept where one party can prove to another that they know a specific piece of information without revealing the information itself. In the context of EXAI, zero knowledge could be applied to enhance privacy and confidentiality. This is particularly relevant when dealing with sensitive data used in AI models [152]. By incorporating zero knowledge techniques, EXAI systems can provide explanations without disclosing the underlying proprietary or confidential information. This is especially important in industries where data privacy is a top priority, such as health care or finance. Zero knowledge ensures that the AI model can convey meaningful insights without compromising the confidentiality of the data used in the training process [153].

The combination of zero trust and zero knowledge has a synergistic impact on EXAI. It establishes a robust security posture by safeguarding against unauthorized access and ensures privacy by allowing for secure and confidential explanations. This is pivotal in gaining the trust of users and stakeholders who rely on the interpretability and reliability of AI models. Moreover, the adoption of these principles aligns with the growing emphasis on ethical AI practices, contributing to the responsible deployment of EXAI systems in various domains. In conclusion, the integration of zero trust and zero knowledge principles significantly enhances the security, privacy, and trustworthiness of EXAI, making it more resilient and reliable in real-world applications [154].

VIII. CASE STUDY: VISUALIZATION OF MACHINING FEATURES AND EXAI FOR MANUFACTURING COST ASSESSMENT

DL-based manufacturing cost prediction learning methods have been used recently, but the models are still employed as a “black box.” This makes it impossible to understand why costs are predicted as they are. This case study uses EXAI to suggest the manufacturing cost-predictive model for 3-D computer-aided design (CAD) models. The suggested method makes it possible to observe how the 3-D CAD model’s machining characteristics affect the rise of manufacturing costs. The suggested method entails three steps: 1) information gathering and preprocessing; 2) exploration of 3-D DL architecture; and 3) visualization of prediction outcomes. The proposed case study discussed the DL model, demonstrating great prediction of manufacturing cost for machine components. By employing the proposed approach, engineering designers can receive design guidance to help cut manufacturing costs while still at the conceptual design stage. In the past few years, there has been a rise in DL usage for cost estimation in manufacturing. To estimate manufacturing costs, Kankanhalli et al. [105] demonstrated that the 3-D CAD models may be trained with a 3-D CNN. Also, for accurate manufacturing cost forecast, this must be clarified which 3-D CAD model aspect mostly contributes to the rise in manufacturing costs. Consequently, the research aims to create an AI model that can suggest to designers which components should be changed in 3-D CAD to save manufacturing costs. To deal with such issues, Wang et al. [106] proposed a scheme where a 3-D CNN is initially constructed using inputs such as materials, volume, and voxelized 3-D CAD. After this, the 3-D gradient-weight class activation mapping (Grad-CAM) is used to illustrate 3-D CAD features that impact manufacturing costs. This case study uses 3-D CAD models, and their pricing information from computer numerical control (CNC)-machined parts sold online to illustrate the effectiveness of the suggested approach. The results demonstrated that the suggested model could distinguish between CNC machining features and machining difficulty. The suggested structure for estimating production costs is depicted in Fig. 7, with a brief explanation of each step as follows.

Step I: This first step gathers 3-D CAD, price, and material data of 1006 machined components. The obtained data should pass through two preprocessing phases before being used as DL input. In the first preprocessing, the 3-D CAD is converted to a mesh file, and the volume is calculated. The data from the voxel and point cloud are then combined. Material, volume, and cost statistics are normalized in the second preprocessing step.

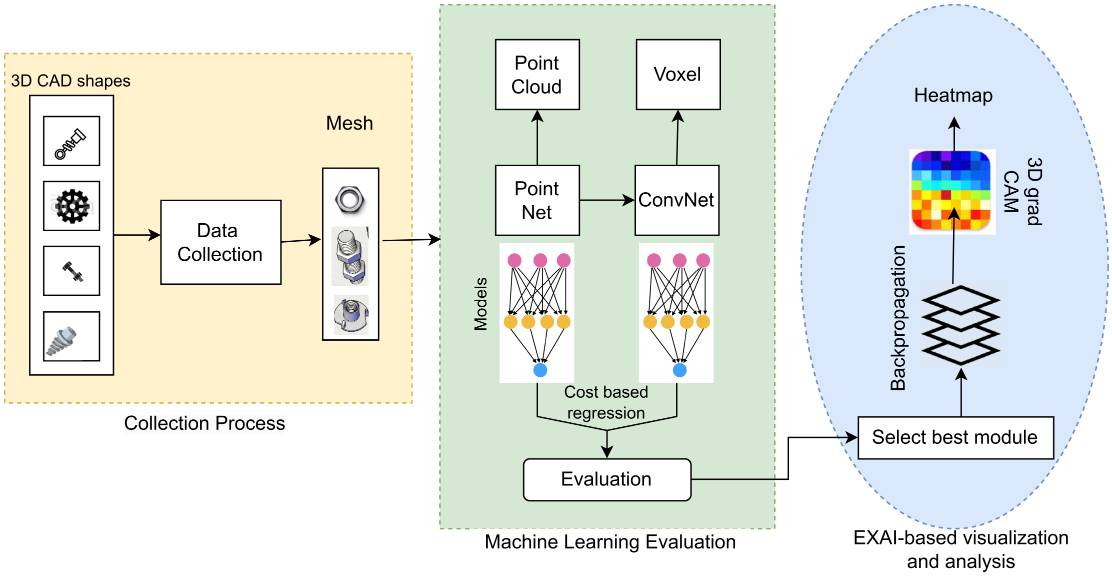

FIGURE 7. **Proposed research framework.**

Step II: To investigate the 3-D DL model, the comparison of the cost prediction capabilities of several PointNet and 3-D CNN-based designs are applied. The best model will be selected after the evaluation.

Step III: The use of 3-D Grad-CAM is made to illustrate the justification for cost prediction results visually. In addition, the suggested model can identify machining difficulties and recognize CNC machining features.

A. DATASET IDENTIFICATION AND PREPROCESSING

The preprocessing of the 3-D CAD is carried out in two ways: 1) the 3-D CAD is converted into the mesh files [109] and 2) the transformation of the point cloud and voxel. First, the 3-D grid will be created using the minimum and maximum values of the vertex coordinates (usually represented as x, y, z) and then converted from mesh to voxel [110].

The study compiles internet pricing information and 3-D CAD data for CNC machined parts from [107]. More than 50% of the overall production cost of CNC-machined parts is typically accounted for material costs. Due to this, the manufacturing costs decrease because of the cutting difficulty and reduced processing time [108]. In the present case study, the collected price data are used. 1006 different types of machinery parts data are gathered in total. There are 34 different parts categories taken into account. Each part's 3-D CAD, material, and cost information is gathered. For usage as a training dataset for DL models, the acquired information of the 3-D CAD and numerical data (which are available in terms of the material, volume, and cost) undergoes the two preprocessing steps.

B. 3-D CAD AND ITS PREPROCESSING

The preprocessing of the 3-D CAD is carried out in two ways: 1) the 3-D CAD is converted into the mesh files [109] and 2) the transformation of the point cloud and voxel. First, the 3-D grid will be created using the minimum and maximum values of the vertex coordinates (usually represented as x, y, z) and then converted from mesh to voxel [110]. Then, the points are transferred from the mesh into another form called the point cloud [111]. After this, the weighted sampling in random order is implemented to identify the 3-D mesh using n points.

C. VOLUME, MATERIALS, COST, AND ITS ASSOCIATED PREPROCESSING

Before being fed into DL models, the data with significant range deviations need to be scaled. Despite the availability of numerous data scaling approaches, the preprocessing of the data is being done using two methods: log normalization and min–max normalization. These methods are represented as follows:

$$y_{new} = \ln y \quad (1)$$

$$y_{new} = y - y_{min}/y_{max} - y_{min}. \quad (2)$$

The dataset representation is mentioned in Table 8. It contains the details of the parts, typically made up of aluminum, stainless steel, and its materials parts. The hot encoding has been used to denote categorical and material data.

D. ARCHITECTURE OF THE 3-D DL: BASELINE MODEL

The DL mechanism was used to explore the architecture and various hyperparameters. The five benchmarks were used to validate the model, and these are VoxNet, PointNet, [97], [111], and [112]. The proposed model configurations are depicted in Table 9.

The fully connected layers execute regression, while the convolutional layers handle feature extraction. Various architectures were tested, using baseline models as a reference.

TABLE 8. Materials Used in All Parts

Precise form	Type	Vector(one -hot)
SUS303	Stainless	[0,0,0,0,0,0,0,0,0,0,0,0,1]
Stainless	Stainless	[0,0,0,0,0,0,0,0,0,0,0,1,0]
SUS304	Stainless	[0,0,0,0,0,0,0,0,0,0,1,0,0]
2000 series aluminum alloys	Aluminum	[0,0,0,0,0,0,0,0,0,1,0,0,0]
A2011	Aluminum	[0,0,0,0,0,0,0,0,1,0,0,0,0]
A5052	Aluminum	[0,0,0,0,0,0,0,1,0,0,0,0,0]
Aluminum alloys	Aluminum	[0,0,0,0,0,0,1,0,0,0,0,0,0]
A6061	Aluminum	[0,0,0,0,0,1,0,0,0,0,0,0,0]
S35C	Steel	[0,0,0,0,1,0,0,0,0,0,0,0,0]
SS400	Steel	[0,0,0,1,0,0,0,0,0,0,0,0,0]
S50C	Steel	[0,0,1,0,0,0,0,0,0,0,0,0,0]
S45C	Steel	[0,1,0,0,0,0,0,0,0,0,0,0,0]
Structural steel	Steel	[1,0,0,0,0,0,0,0,0,0,0,0,0]

TABLE 9. Architecture of the Proposed Model

Type of layers used	Number of filters applied into the model	Function of the activation used	Bias and weight initializer.
Input Shape	$32 \times 32 \times 32$	0	Zeros,Xavier (normal)
3-D Convolution (2)	Two $3 \times 3 \times 3$ (16)	LeakyReLU	Zeros,Xavier (normal)
Max pooling	$2 \times 2 \times 2$	0	0
Dropout	0.3	0	0
3-D Convolution (2)	Two $3 \times 3 \times 3$(32)	LeakyReLU	Zeros,Xavier (normal)
Max pooling	$2 \times 2 \times 2$	0	0
Dropout	0.3	0	0
3-D Convolution	(64)$3 \times 3 \times 3$	LeakyReLU	Zeros,Xavier (normal)
Flatten (×1)	0	0	0
Input 2	16	0	0
Fully connected (×2)	16	0	0
Input 3	1	0	0
Fully connected (×3)	1	0	Zeros,Xavier (normal)
Concatenate layer (×1, ×2, ×3)	0	0	0
Fully connected (5)	2000,300,150,20,16	LeakyReLU	Zeros,Xavier (normal)
Fully connected	1	0	Zeros,Xavier (normal)

The proposed architecture was then constructed with slight adjustments based on the design mentioned in [113]. Exactly two hidden layers of 1024 neurons were used in the regression portion of the aforementioned work. The number of neurons fell to around 2000, 300, 150, 20, and 16 in the suggested model, which comprises five hidden layers. This neuron combination was the winner in the testing phase of the completely connected layer combinations. After this, evaluate several initialization strategies. When the starting bias $= 0$, this was observed that Xavier performs the best for the weight parameters in the model. LeakyReLU is used in place of rectified linear unit (ReLU) for the activation function of the fully connected layer. If a negative result enters ReLU, the output will be set to 0. The neurons become dormant as a result. LeakyReLU, which multiplies the input and sets its hyperparameter, addresses this issue. This is represented using (3). Here, $\alpha = 0.1$

$$f_{\text{act}} = \max(ay, y). \tag{3}$$

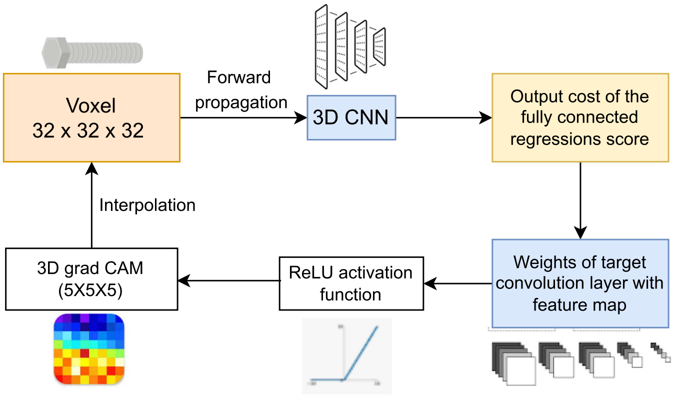

FIGURE 8. Process of 3-D grad CAM.

E. 3-D GRAD-CAM

Although the modeling is employed as a black box, the learned 3-D CNN can estimate the cost of the new parts. However, the predicted outcome must be understandable for the business to be reliable. To visualize the 3-D features, Wang et al. [114] developed a method called 3-D Grad-CAM. This is usually an extension of 2-D Grad-CAM. The process is mentioned in Fig. 8.

$$\alpha_l = 1/B \times \sum_i \sum_j \sum_k \partial x / \partial B^l_{i,j,k} \tag{4}$$

$$L_{\text{3DGCAM}} = \text{ReLU}\left(\sum_l \alpha_l \times B^l\right). \tag{5}$$

The 3-D gradient can be predicted through the backpropagation and regression score x, resulting in $B^l_{i,j,k}$. This is also denoted as a target convolution network. $B^l_{i,j,k}$ is defined as the feature map related to the ith row and jth column of the kth channel and is related to the lth feature map. When the global average pooling is implemented for the mentioned 3D gradient, this will result in (4). This is denoted as α_l; once identified, this will combine linearly with$B^l_{i,j,k}$, then passed through the ReLU, and this results in the 3-D Grad-CAM. Here, this is denoted as L_{3DGCAM}. This is represented as (5).

F. MANUFACTURING COST AND ITS PREDICTION

There are 1006 total data used, of which 80% are utilized for training and 20% for testing. The mean absolute percentage error (MAPE) and the root-mean-square error (RMSE) measure the model's performance. Table 10 displays the outcomes of the top schemes for every architecture. Each model's architecture, input data type, normalization type, and loss function are unique.

G. MACHINING FEATURES AND ITS EXPLANATION

Machined features, such as edges, holes, and teeth, are clearly shown in Maxpooling1. However, the resolution is reduced in deeper layers, as Conv5 and the highlighted parts do not

TABLE 10. Characteristics of the Model

Architecture of the model	Inputs	Normalization techniques used	The utility of the Loss function used	RMSE	MAPE
Voxnet	$Voxel(32)^{1}, Mat, Vol$	Min–Max Log	MAE and MSLE	2728.24	29.88
PointNet	Point $(2048)^{3}$,Mat,vol	Log	MSE	3200.56	19.78
[114]	Voxel(32),Mat,Vol	Min–Max Log	MAE and MSE	1578.89, 2011.34	20.76,10.65
[112]	$Voxel(64)^{2}$,Mat,Vol	Log	MAE	1967.67	12.89
[113]	Voxel(64),Mat,Vol	Log	MSLE	1001.67	14.98
Proposed approach	Voxel(32),Mat,Vol	Min–Max and Log	MAE	1167.67,1187.76	18.67,7.86

accurately depict the key elements. In a research paper, Minerva et al. [115] discussed implementing the CNC features of machining. However, incorporating the CAD model into the CNC has enhanced its clarity and highlighted its machining features. In addition, visualizing cost-critical regions in a CAD system is possible with 3-D Grad-CAM. Thus, engineering designers may identify which elements of conceptual designs must be improved to lower manufacturing costs early in the product design process.

H. EXPLANATIONS FOR THE DEGREE OF DIFFICULTY IN PROCESSING WITH MACHINING

Costs of CNC production are primarily influenced by the complexity of the machining operation, which involves the inclusion of the cutting tool. Drills, which are utilized to drill holes, and end mills, which are utilized to cut sidewalls, are the two main categories of cutting tools. This section mentions the outcomes of the machining issues for the CAD shapes. The dataset used for the experiments includes the degree of depth of the various units, such as 10, 20, and 30 mm. In real-time operations, as the depth size increases, the cost also increases. The reason for the increase in the cost is due to the operation done at the end mill. The end mill is used to cut axially for the given workpiece. Due to this, the increase in the vibration will be recorded, resulting in a decrease in the stability and processing difficulty. The prediction outcome of the model discussed here also indicates the increase in the cost with higher processing depth. Similarly, the round values of the corner of the machinery also impact the outcomes. The larger the rounds, it can be processed with the large diameter end mill in less processing time and cost. Prediction outcomes even indicate that the cost is higher with the smaller round size. The same will be highlighted via the use of the 3-D Grad-CAM. The results confirmed that the proposed model would differentiate between similar shaped-based CAD schemes. The proposed model can train and learn the effect of the various feature types. Even the detail difference in the features for the manufacturing cost for the CAD will be identified without the need for domain knowledge.

IX. LESSONS LEARNED FROM THE SURVEY

The section presents the potential lessons learned from the survey. Finally, this article presents the key lessons learned as follows.

1) The authors discussed the current Industry 5.0 and its relation with the EXAI. The various challenges and their solutions were also mentioned in the article.
2) The authors have highlighted the essential breakthroughs supporting Industry 5.0, such as the AI-ML-based approach for EXAI and various potential case studies. The authors also introduce the connection of LLMs with EXAI and its applications. The case studies include the advantages of using the EXAI in various fields, such as cloud-based smart governance, DTs, UAVs, smart grid, XR applications, and zero-touch networks.
3) As an application of EXAI in manufacturing, cost assessment has been mentioned with its prepossessing to EXAI-based GRAD CAM implementation.

EXAI has enhanced the operational boundaries of AI models, with enhanced trust and transparency of the critical outputs produced with the industrial processes. It becomes a potential tool for industry stakeholders to add accountability to predictive analysis, enhancing industrial processes' maintenance and productivity. Thus, trusted fail-safe systems are built for CPHS, with low bias and improved business logistics, which leverages the rapid prototyping and customization principle of Industry 5.0. In close collaboration, it blends with the H2M and M2H inputs, which forms a resilient ecosystem. Moreover, there is a high availability of EXAI tools and techniques, which makes it easier for organizations to adopt EXAI in diverse industrial verticals.

With the inherent advantages, the adoption of EXAI in Industry 5.0 is still in the nascent stages. Thus, it requires effective collaboration between researchers and industry practitioners to form effective policies to advance EXAI deployment in Industry 5.0. First, the reason for limited adoption is due to the challenges of architectural complexity and computational overheads EXAI brings to the AI models, which might become a challenge for medium industries. Second, the effectiveness of EXAI depends on the type, quantity, and quality of the available data. Effective preprocessing strategies are required in real-time data collection before EXAI integration into the learning model. Finally, the type of bias needs to be identified to improve the model's accuracy, which limits the effectiveness of statistical inferences made by the model.

X. CONCLUSION

Industry 5.0 symbolizes a holistic integration of cutting-edge automation, cyber-physical elements, virtual reality, and resilient communication networks to improve production, mobility, and maintenance systems of industrial processes. EXAI has gained attention in Industry 5.0 due to its key elements (transparency, audibility, and accountability) in production space, which lowers model bias and improves the accuracy of AI models. The article presented a systematic survey of key EXAI principles, specific to Industry 5.0 applications. The integration of EXAI is highlighted in diverse industry verticals such as cloud-based smart governance, DTs, UAVs, smart grid, XR applications, and zero-touch networks. A solution taxonomy of EXAI in Industry 5.0 was discussed, supported through a case study on EXAI application on the manufacturing cost assessment. The article concluded that EXAI validates and extends AI analytics and drives intelligent justification to Industry 5.0 processes. The potential future challenges and key directions are also discussed in the article.

As part of the future scope of the survey, the authors would like to induce privacy preservation with EXAI, which would lead to designing AI models where local data are private and secured against malicious intruders. This leads to the design of private industrial AI systems, where differential-private federated learning systems are built, adding a high degree of security and privacy in the learning process without malicious intruders' release and linkage attacks.

REFERENCES

[1] R. S. Peres, X. Jia, J. Lee, K. Sun, A. W. Colombo, and J. Barata, "Industrial artificial intelligence in industry 4.0—Systematic review, challenges and outlook," *IEEE Access*, vol. 8, pp. 220121–220139, 2020.

[2] I. Vanany, S. Zailani, and N. Pujawan, "Supply chain risk management," *Int. J. Inf. Syst. Supply Chain Manage.*, vol. 2, pp. 16–33, 2011.

[3] E. Griffor, C. Greer, D. Wollman, and M. Burns, "Framework for cyber-physical systems: Volume 2, working group reports, special publication," National Institute of Standards and Technology, Gaithersburg, MD, 2017. Accessed: Mar. 16, 2023. [Online]. Available: https://doi.org/10.6028/NIST.SP.1500-202

[4] P. Bhattacharya, F. Patel, S. Tanwar, N. Kumar, and R. Sharma, "MB-MaaS: Mobile blockchain-based mining-as-a-service for IIoT environments," *J. Parallel Distrib. Comput.*, vol. 168, pp. 1–16, 2022.

[5] V. K. Prasad et al., "ABV-CoVID: An ensemble forecasting model to predict availability of beds and ventilators for COVID-19 like pandemics," *IEEE Access*, vol. 10, pp. 74131–74151, 2022.

[6] T.-M. Choi, S. Kumar, X. Yue, and H.-L. Chan, "Disruptive technologies and operations management in the industry 4.0 era and beyond," *Prod. Oper. Manage.*, vol. 31, no. 1, pp. 9–31, 2022.

[7] H. Mehta, V. K. Prasad, and D. M. Bhavsar, "Efficient resource scheduling in cloud computing," *Int. J. Adv. Res. Comput. Sci.*, vol. 8, no. 3, pp. 809–815, 2017.

[8] M. Zuhair, F. Patel, D. Navapara, P. Bhattacharya, and D. Saraswat, "BloCoV6: A blockchain-based 6G-assisted UAV contact tracing scheme for COVID-19 pandemic," in *Proc. IEEE 2nd Int. Conf. Intell. Eng. Manage.*, London, U.K., 2021, pp. 271–276.

[9] P. Bhattacharya, M. S. Obaidat, S. Sanghavi, V. Sakariya, S. Tanwar, and K.-F. Hsiao, "Internet-of-explainable-digital-twins: A case study of versatile corn production ecosystem," in *Proc. Int. Conf. Commun., Comput., Cybersecur., Informat.*, Dalian, China, 2022, pp. 1–5.

[10] X.-H. Li et al., "A survey of data-driven and knowledge-aware eXplainable AI," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 29–49, Jan. 2022.

[11] T. Evgeniou and M. Pontil, "Support vector machines: Theory and applications," in *Machine Learning and Its Applications*, vol. 2049. Berlin, Germany: Springer, 2001, pp. 249–257.

[12] G. I. Webb, "Naïve Bayes," in *Encyclopedia of Machine Learning*, C. Sammut and G. I. Webb, Eds., Boston, MA, USA: Springer, 2010, pp. 713–714.

[13] N. Ali, D. Neagu, and P. Trundle, "Evaluation of k-nearest neighbour classifier performance for heterogeneous data sets," *SN Appl. Sci.*, vol. 1, Nov. 2019, Art. no. 1559.

[14] P. Bhuvaneswari and B. Therese, "Detection of cancer in lung with K-NN classification using genetic algorithm," *Procedia Mater. Sci.*, vol. 10, pp. 433–440, 2015.

[15] M. Zulfadhilah, Y. Prayudi, and I. Riadi, "Cyber profiling using log analysis and k-means clustering a case study higher education in Indonesia," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 8, pp. 430–435, 2016.

[16] X. Zheng, Q. Lei, R. Yao, Y. Gong, and Q. Yin, "Image segmentation based on adaptive k-means algorithm," *EURASIP J. Image Video Process.*, vol. 2018, no. 1, pp. 1–10, Aug. 2018.

[17] A. Luckow, M. Cook, N. Ashcraft, E. Weill, E. Djerekarov, and B. Vorster, "Deep learning in the automotive industry: Applications and tools," in *Proc. IEEE Int. Conf. Big Data*, 2016, pp. 3759–3768.

[18] S. G. Fashoto, O. Owolabi, J. Sadiku, and J. Gbadeyan, "Application of data mining technique for fraud detection in health insurance scheme using knee-point k-means algorithm," *Aust. J. Basic Appl. Sci.*, vol. 7, no. 8, pp. 140–144, 2013.

[19] A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on explainable artificial intelligence (XAI)," *IEEE Access*, vol. 6, pp. 52138–52160, 2018.

[20] S. Ferrara and S. Qunbar, "Validity arguments for AI-based automated scores: Essay scoring as an illustration," *J. Educ. Meas.*, vol. 59, no. 3, pp. 288–313, 2022.

[21] A. Barredo Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Inf. Fusion*, vol. 58, pp. 82–115, 2020.

[22] D. C. Elton, "Self-explaining AI as an alternative to interpretable AI," in *Artificial General Intelligence*, B. Goertzel, A. I. Panov, A. Potapov, and R. Yampolskiy, Eds., Cham, Switzerland: Springer, 2020, pp. 95–106.

[23] D. Dobrev, "A definition of artificial intelligence," 2012, *arXiv:1210.1568*.

[24] M. Clinciu and H. Hastie, "A survey of explainable AI terminology," in *Proc. 1st Workshop Interact. Natural Lang. Technol. Explainable Artif. Intell.*, 2019, pp. 8–13.

[25] S. Mohseni, N. Zarei, and E. D. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable AI systems," *ACM Trans. Interact. Intell. Syst.*, vol. 11, no. 3/4, pp. 1–45, 2021.

[26] "Science and technology for the explanation of AI decision making-erc-2018-adg grant." Accessed: Jan. 28, 2023. [Online]. Available: https://xai-project.eu/

[27] "Explainable artificial intelligence for defense advanced research projects agency." Accessed: Jan. 28, 2023. [Online]. Available: https://www.darpa.mil/program/explainable-artificial-intelligence

[28] V. Terziyan and O. Vitko, "Explainable AI for industry 4.0: Semantic representation of deep learning models," *Procedia Comput. Sci.*, vol. 200, pp. 216–226, 2022.

[29] "Explainable artificial intelligence-based verification & validation for increasingly autonomous aviation systems, phase 2e." Accessed: Jan. 28, 2023. [Online]. Available: https://techport.nasa.gov/view/113205

[30] "Germany's human-centred approach to AI is inclusive, evidence-based and capacity-building." Accessed: Jan. 28, 2023. [Online]. Available: https://oecd.ai/en/wonk/germany-takes-an-inclusive-and-evidence-based-approach-for-capacity-building-and-a-human-centred-use-of-ai

[31] W. R. Swartout, "Xplain: A system for creating and explaining expert consulting programs," *Artif. Intell.*, vol. 21, no. 3, pp. 285–325, 1983.

[32] M. V. Lent, W. Fisher, and M. Mancuso, "An explainable artificial intelligence system for small-unit tactical behavior," in *Proc. 16th Conf. Innov. Appl. Artif. Intell.*, 2004, pp. 900–907.

[33] P. Bhattacharya, M. S. Obaidat, D. Savaliya, S. Sanghavi, S. Tanwar, and B. Sadaun, "Metaverse assisted telesurgery in healthcare 5.0: An interplay of blockchain and explainable AI," in *Proc. Int. Conf. Comput., Inf. Telecommun. Syst.*, Piraeus, Greece, 2022, pp. 1–5.

[34] J. Kang et al., "Blockchain-based federated learning for industrial metaverses: Incentive scheme with optimal AOI," in *Proc. IEEE Int. Conf. Blockchain*, Espoo, Finland, 2022, pp. 71–78.

[35] P. Bhattacharya et al., “Towards future Internet: The metaverse perspective for diverse industrial applications,” *Mathematics*, vol. 11, no. 4, 2023, Art. no. 941.

[36] P. Bhattacharya et al., “Amalgamation of blockchain and sixth-generation-envisioned responsive edge orchestration in future cellular vehicle-to-anything ecosystems: Opportunities and challenges,” *Trans. Emerg. Telecommun. Technol.*, vol. 35, no. 4, 2021, Art. no. e4410.

[37] V. K. Prasad, S. Tanwar, and M. D. Bhavsar, *Advance Cloud Data Analytics for 5G Enabled IoT*. Cham, Switzerland: Springer, 2021, pp. 159–180.

[38] J. Sanghvi, P. Bhattacharya, S. Tanwar, R. Gupta, N. Kumar, and M. Guizani, “Res6Edge: An edge-AI enabled resource sharing scheme for C-V2X communications towards 6G,” in *Proc. Int. Wireless Commun. Mobile Comput.*, Harbin, China, 2021, pp. 149–154.

[39] C. Trivedi and U. P. Rao, “Secrecy aware key management scheme for internet of healthcare things,” *J. Supercomput.*, vol. 79, no. 11, pp. 12492–12522, 2023.

[40] S. Chourasia, S. M. Pandey, K. Gupta, Q. Murtaza, and R. Walia, “Industry 5.0 for sustainable manufacturing: New product, services, organizational and social information,” in *Surface Engineering*, Boca Raton, FL, USA: CRC, 2023, pp. 243–255.

[41] S. Chaudhary, P. Joshi, P. Bhattacharya, V. K. Prasad, R. Shah, and S. Tanwar, “Untangling explainable AI in applicative domains: Taxonomy, tools, and open challenges,” in *Proc. 4th Int. Conf. Comput., Commun., Cyber-Secur.*, 2023, pp. 857–872.

[42] R. Binns, M. V. Kleek, M. Veale, U. Lyngs, J. Zhao, and N. Shadbolt, “‘it’s reducing a human being to a percentage’; perceptions of justice in algorithmic decisions,” in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2018, pp. 1–14.

[43] E. Rader, K. Cotter, and J. Cho, *Explanations as Mechanisms for Supporting Algorithmic Transparency*. New York, NY, USA: Assoc. Comput. Mach., 2018, pp. 1–13.

[44] P. Bhattacharya, S. B. Patel, R. Gupta, S. Tanwar, and J. J. P. C. Rodrigues, “SaTYa: Trusted Bi-LSTM-based fake news classification scheme for smart community,” *IEEE Trans. Comput. Social Syst.*, vol. 9, no. 6, pp. 1758–1767, Dec. 2022.

[45] S. Stumpf et al., “Interacting meaningfully with machine learning systems: Three experiments,” *Int. J. Human.-Comput. Stud.*, vol. 67, pp. 639–662, Aug. 2009.

[46] P. Radanliev, D. De Roure, M. Van Kleek, O. Santos, and U. Ani, “Artificial intelligence in cyber physical systems,” *AI Soc.*, vol. 36, pp. 783–796, Sep. 2021.

[47] J. Tao et al., “XAI-driven explainable multi-view game cheating detection,” in *Proc. IEEE Conf. Games*, Osaka, Japan, 2020, pp. 144–151.

[48] A. Nascita, A. Montieri, G. Aceto, D. Ciuonzo, V. Persico, and A. Pescapè, “Unveiling mimetic: Interpreting deep learning traffic classifiers via XAI techniques,” in *Proc. IEEE Int. Conf. Cyber Secur. Resilience*, Rhodes, Greece, 2021, pp. 455–460.

[49] N. Sarathy, M. Alsawwaf, and Z. Chaczko, “Investigation of an innovative approach for identifying human face-profile using explainable artificial intelligence,” in *Proc. IEEE 18th Int. Symp. Intell. Syst. Informat.*, Subotica, Serbia, 2020, pp. 155–160.

[50] G. Alicioglu and B. Sun, “A survey of visual analytics for explainable artificial intelligence methods,” *Comput. Graph.*, vol. 102, pp. 502–520, 2021.

[51] M. Danilevsky, K. Qian, R. Aharonov, Y. Katsis, B. Kawas, and P. Sen, “A survey of the state of explainable AI for natural language processing,” in *Proc. 1st Conf. Asia-Pacific Ch. Assoc. Comput. Linguistics/10th Int. Joint Conf. Natural Lang. Process.*, 2020, pp. 447–459.

[52] E. Tjoa and C. Guan, “A survey on explainable artificial intelligence (XAI): Toward medical XAI,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 11, pp. 4793–4813, Nov. 2021.

[53] R. Calegari, G. Ciatto, and A. Omicini, “On the integration of symbolic and sub-symbolic techniques for XAI: A survey,” *Intell. Artif.*, vol. 14, pp. 1–25, 2020.

[54] K. Čyras, A. Rago, E. Albini, P. Baroni, and F. Toni, “Argumentative XAI: A survey,” in *Proc. 13th Int. Joint Conf. Artif. Intell. Surv. Track*, 2021, pp. 4392–4399.

[55] T. Rojat et al., “Explainable artificial intelligence (XAI) on timeseries data: A survey,” 2021, *arXiv:2104.00950*.

[56] M. Sahakyan, Z. Aung, and T. Rahwan, “Explainable artificial intelligence for tabular data: A survey,” *IEEE Access*, vol. 9, pp. 135392–135422, 2021.

[57] S. R. Islam, W. Eberle, S. K. Ghafoor, and M. Ahmed, “Explainable artificial intelligence approaches: A survey,” 2021, *arXiv:2101.09429*.

[58] R. Machlev et al., “Explainable artificial intelligence (XAI) techniques for energy and power systems: Review, challenges and opportunities,” *Energy AI*, vol. 9, 2022, Art. no. 100169.

[59] I. Monteath and R. Sheh, “Assisted and incremental medical diagnosis using explainable artificial intelligence,” in *Proc. 2nd Workshop Explainable Artif. Intell.*, 2018, pp. 104–108.

[60] F. K. Došilović, M. Brčić, and N. Hlupić, “Explainable artificial intelligence: A survey,” in *Proc. 41st Int. Conv. Inf. Commun. Technol., Electron. Microelectronics*, Opatija, Croatia, 2018, pp. 210–215.

[61] A. Vassiliades, N. Bassiliades, and T. Patkos, “Argumentation and explainable artificial intelligence: A survey,” *Knowl. Eng. Rev.*, vol. 36, p. e5, 2021.

[62] I. Ahmed, G. Jeon, and F. Piccialli, “From artificial intelligence to explainable artificial intelligence in industry 4.0: A survey on what, how, and where,” *IEEE Trans. Ind. Informat.*, vol. 18, no. 8, pp. 5031–5042, Aug. 2022.

[63] P. Bhattacharya, A. K. Tiwari, and A. Singh, “Dual-buffer-based optical datacenter switch design,” *J. Opt. Commun.*, vol. 44, no. 2, pp. 155–162, 2023.

[64] V. K. Prasad and M. Bhavsar, “Preserving SLA parameters for trusted IaaS cloud: An intelligent monitoring approach,” *Recent Patents Eng.*, vol. 14, no. 4, pp. 530–540, 2020.

[65] N. Velasquez, E. Estevez, and P. Pesado, “Cloud computing, big data and the industry 4.0 reference architectures,” *J. Comput. Sci. Technol.*, vol. 18, 2018, Art. no. e29.

[66] A. Verma, P. Bhattacharya, U. Bodkhe, D. Saraswat, S. Tanwar, and K. Dev, “FedRec: Trusted rank-based recommender scheme for service provisioning in federated cloud environment,” *Digit. Commun. Netw.*, vol. 9, no. 1, pp. 33–46, 2023.

[67] H. Lingaraj, “A study on genetic algorithm and its applications,” *Int. J. Comput. Sci. Eng.*, vol. 4, pp. 139–143, 2016.

[68] A. Askarzadeh, L. D. S. Coelho, C. E. Klein, and V. C. Mariani, “A population-based simulated annealing algorithm for global optimization,” in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, Budapest, Hungary, 2016, pp. 4626–4633.

[69] V. K. Prajapati, M. Jain, and L. Chouhan, “Tabu search algorithm (TSA): A comprehensive survey,” in *Proc. IEEE 3rd Int. Conf. Emerg. Technol. Comput. Eng.: Mach. Learn. Internet Things*, Jaipur, India, 2020, pp. 1–8.

[70] V. K. Prasad and M. D. Bhavsar, “SLAMMP framework for cloud resource management and its impact on healthcare computational techniques,” *Int. J. E-Health Med. Commun.*, vol. 12, no. 2, pp. 1–31, 2021.

[71] V. K. Prasad and M. D. Bhavsar, “Monitoring IaaS cloud for healthcare systems: Healthcare information management and cloud resources utilization,” *Int. J. E-Health Med. Commun.*, vol. 11, no. 3, pp. 54–70, 2020.

[72] Z. Patel, Y. Velankar, C. Trivedi, and P. Oza, “Wireless implantable medical devices security and privacy: A survey,” in *Smart Energy Advancement Power Technologies*, K. Namrata, N. Priyadarshi, R. C. Bansal, and J. Kumar, Eds., Singapore: Springer, 2023, pp. 69–87.

[73] P. Popovski, K. F. Trillingsgaard, O. Simeone, and G. Durisi, “5G wireless network slicing for eMBB, URLLC, and mMTC: A communication-theoretic view,” *IEEE Access*, vol. 6, pp. 55765–55779, 2018.

[74] A. Singh, A. K. Tiwari, and P. Bhattacharya, “Bit error rate analysis of hybrid buffer-based switch for optical data centers,” *J. Opt. Commun.*, vol. 43, no. 3, pp. 339–346, 2022.

[75] V. K. Prasad, M. D. Bhavsar, and S. Tanwar, “Influence of montoring: Fog and edge computing,” *Scalable Comput.: Pract. Exp.*, vol. 20, no. 2, pp. 365–376, 2019.

[76] F. Charrua-Santos, “6G enabled industrial Internet of everything: Towards a theoretical framework,” *Appl. Syst. Innov.*, vol. 4, 2021, Art. no. 11.

[77] J. Edward, C. W. Wannasuphoprasit, and M. A. Peshkin, “Cobots: Robots for collaboration with human operators,” in *Proc. Int. Mech. Eng. Congr. Expo.*, 1996, pp. 433–440.

[78] Z. Lin and X. Liu, “Wind power forecasting of an offshore wind turbine based on high-frequency SCADA data and deep learning neural network,” *Energy*, vol. 201, 2020, Art. no. 117693.

[79] A. Ismail, H.-L. Truong, and W. Kastner, “Manufacturing process data analysis pipelines: A requirements analysis and survey,” *J. Big Data*, vol. 6, no. 1, pp. 1–26, Jan. 2019.

[80] B. Dafflon, N. Moalla, and Y. Ouzrout, "The challenges, approaches, and used techniques of CPS for manufacturing in industry 4.0: A literature review," *Int. J. Adv. Manuf. Technol.*, vol. 113, pp. 2395–2412, Apr. 2021.

[81] A. Verma et al., "Blockchain for industry 5.0: Vision, opportunities, key enablers, and future directions," *IEEE Access*, vol. 10, pp. 69160–69199, 2022.

[82] C. Trivedi, U. P. Rao, K. Parmar, P. Bhattacharya, S. Tanwar, and R. Sharma, "A transformative shift toward blockchain-based IoT environments: Consensus, smart contracts, and future directions," *Secur. Privacy*, 2023, vol. 6, no. 5, Art. no. e308.

[83] S. Shyamsukha, P. Bhattacharya, F. Patel, S. Tanwar, R. Gupta, and E. Pricop, "PoRF: Proof-of-reputation-based consensus scheme for fair transaction ordering," in *Proc. IEEE 13th Int. Conf. Electron., Comput. Artif. Intell.*, Pitesti, Romania, 2021, pp. 1–6.

[84] Y. K. Dwivedi et al., "Opinion paper: "so what if ChatGPT wrote it?" multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy," *Int. J. Inf. Manage.*, vol. 71, 2023, Art. no. 102642.

[85] Y. Chang et al., "A survey on evaluation of large language models," *ACM Trans. Intell. Syst. Technol.*, vol. 15, no. 3, pp. 1–45, 2023.

[86] A. Vaswani et al., "Attention is all you need," in *Advances in Neural Information Processing Systems*, vol. 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., Glasgow, U.K.: Curran Assoc., Inc., 2017.

[87] D. Saraswat et al., "Explainable AI for healthcare 5.0: Opportunities and challenges," *IEEE Access*, vol. 10, pp. 84486–84517, 2022.

[88] X. Wang et al., "Steps toward industry 5.0: Building '6s' parallel industries with cyber-physical-social intelligence," *IEEE/CAA J. Autom. Sinica*, vol. 10, no. 8, pp. 1692–1703, Aug. 2023.

[89] S. Wisdom, T. Powers, J. Pitton, and L. Atlas, "Interpretable recurrent neural networks using sequential sparse recovery," 2016, *arXiv:1611.07252*.

[90] D. Bau, B. Zhou, A. Khosla, A. Oliva, and A. Torralba, "Network dissection: Quantifying interpretability of deep visual representations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 6541–6549.

[91] A. Shrikumar, P. Greenside, A. Shcherbina, and A. Kundaje, "Not just a black box: Learning important features through propagating activation differences," 2016, *arXiv:1605.01713*.

[92] J. Waa, E. Nieuwburg, A. Cremers, and M. Neerincx, "Evaluating XAI: A comparison of rule-based and example-based explanations," *Artif. Intell.*, vol. 291, 2020, Art. no. 103404.

[93] Z. Yang, A. Zhang, and A. Sudjianto, "GAMI-NET: An explainable neural network based on generalized additive models with structured interactions," *Pattern Recognit.*, vol. 120, 2021, Art. no. 108192.

[94] M. Sundararajan, A. Taly, and Q. Yan, "Axiomatic attribution for deep networks," in *Proc. Int. Conf. Mach. Learn.*, 2017, pp. 3319–3328.

[95] B. Letham, C. Rudin, T. H. McCormick, and D. Madigan, "Interpretable classifiers using rules and Bayesian analysis: Building a better stroke prediction model," *Ann. Appl. Statist.*, vol. 9, pp. 1350–1371, Sep. 2015.

[96] B. Kim et al., "Interpretability beyond feature attribution: Quantitative testing with concept activation vectors (TCAV)," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, pp. 2668–2677.

[97] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," *Nature Mach. Intell.*, vol. 2, pp. 56–67, Jan. 2020.

[98] Z. Wood-Doughty, I. Cachola, and M. Dredze, "Proxy model explanations for time series RNNs," in *Proc. IEEE 20th Int. Conf. Mach. Learn. Appl.*, 2021, pp. 698–703.

[99] M. Timms, "Letting artificial intelligence in education out of the box: Educational cobots and smart classrooms," *Int. J. Artif. Intell. Educ.*, vol. 26, pp. 701–712, 2016.

[100] T. Defraeye et al., "Digital twins probe into food cooling and biochemical quality changes for reducing losses in refrigerated supply chains," *Resources, Conservation Recycling*, vol. 149, pp. 778–794, 2019.

[101] M. Tremblay, "Medical errors in the age of the intelligent machine," *Foresight: Int. J. Appl. Forecasting*, vol. 58, pp. 27–35, 2020.

[102] T. Greif, N. Stein, and C. M. Flath, "Peeking into the void: Digital twins for construction site logistics," *Comput. Ind.*, vol. 121, 2020, Art. no. 103264.

[103] S. Shah, P. Bhattacharya, J. Vaidya, M. Zuhair, and V. K. Prasad, "Administration of Big Data using the cloud as base," in *Proc. IEEE 5th Int. Conf. Contemporary Comput. Informat.*, 2022, pp. 393–398.

[104] P. Bhattacharya, A. K. Tiwari, A. Verma, A. Alabdulatif, S. Tanwar, and R. Sharma, "Lightblocks: A trusted lightweight signcryption and consensus scheme for industrial IoT ecosystems," *Comput. Standards Interfaces*, vol. 88, 2024, Art. no. 103785.

[105] A. Kankanhalli, Y. Charalabidis, and S. Mellouli, "IoT and AI for smart government: A research agenda," *Government Inf. Quart.*, vol. 36, pp. 304–309, 2019.

[106] S. Wang, M. A. Qureshi, L. Miralles-Pechuaán, T. Huynh-The, T. R. Gadekallu, and M. Liyanage, "Explainable AI for B5G/6G: Technical aspects, use cases, and research challenges," 2021, *arXiv:2112.04698*.

[107] J. M. Corchado et al., "Deepint.net: A rapid deployment platform for smart territories," *Sensors*, vol. 21, no. 1, 2021, Art. no. 236.

[108] S. Yang et al., "Cosmos smart intersection: Edge compute and communications for bird's eye object tracking," in *Proc. IEEE Int. Conf. Pervasive Comput. Commun. Workshops*, 2020, pp. 1–7.

[109] T. Alladi, V. Chamola, N. Sahu, and M. Guizani, "Applications of blockchain in unmanned aerial vehicles: A review," *Veh. Commun.*, vol. 23, 2020, Art. no. 100249.

[110] C. T. Wolf and K. E. Ringland, "Designing accessible, explainable AI (XAI) experiences," *ACM SIGACCESS Accessibility Comput.*, no. 125, 2020, Art. no. 6.

[111] S. Matzka, "Explainable artificial intelligence for predictive maintenance applications," in *Proc. IEEE 3rd Int. Conf. Artif. Intell. Ind.*, 2020, pp. 69–74.

[112] Y. Alufaisan, L. R. Marusich, J. Z. Bakdash, Y. Zhou, and M. Kantarcioglu, "Does explainable artificial intelligence improve human decision-making," in *Proc. AAAI Conf. Artif. Intell.*, 2021, vol. 35, pp. 6618–6626.

[113] A. Holzinger, "From machine learning to explainable AI," in *Proc. World Symp. Digit. Intell. Syst. Mach.*, 2018, pp. 55–66.

[114] L. Wang, T. Deng, Z. Zheng, and Z.-J. M. Shen, "Explainable modeling in digital twin," in *Proc. Winter Simul. Conf.*, Phoenix, AZ, USA, 2021, pp. 1–12.

[115] R. Minerva, F. M. Awan, and N. Crespi, "Exploiting digital twin as enablers for synthetic sensing," *IEEE Internet Comput.*, vol. 26, no. 5, pp. 61–67, Sep./Oct. 2022.

[116] X. Zhang, B. Hu, G. Xiong, X. Liu, X. Dong, and D. Li, "Research and practice of lightweight digital twin speeding up the implementation of flexible manufacturing systems," in *Proc. IEEE 1st Int. Conf. Digit. Twins Parallel Intell.*, Beijing, China, 2021, pp. 456–460.

[117] A. Mozo, A. Karamchandani, S. Gómez-Canaval, M. Sanz, J. I. Moreno, and A. Pastor, "B5GEMINI: AI-driven network digital twin," *Sensors*, vol. 22, no. 11, 2022, Art. no. 4106.

[118] J. Bao, D. Guo, J. Li, and J. Zhang, "The modelling and operations for the digital twin in the context of manufacturing," *Enterprise Inf. Syst.*, vol. 13, no. 4, pp. 534–556, 2019.

[119] R. Sahal, S. H. Alsamhi, and K. N. Brown, "Personal digital twin: A close look into the present and a step towards the future of personalised healthcare industry," *Sensors*, vol. 22, no. 15, 2022, Art. no. 5918.

[120] P. Bhatttacharya, K. Patel, M. Zuhair, and C. Trivedi, "A lightweight authentication via unclonable functions for industrial Internet-of-Things," in *Proc. 2nd Int. Conf. Innov. Practices Technol. Manage.*, 2022, vol. 2, pp. 657–662.

[121] W. Guo, "Explainable artificial intelligence for 6G: Improving trust between human and machine," *IEEE Commun. Mag.*, vol. 58, no. 6, pp. 39–45, Jun. 2020.

[122] P. K. Kwok, M. Yan, T. Qu, and H. Y. Lau, "User acceptance of virtual reality technology for practicing digital twin-based crisis management," *Int. J. Comput. Integr. Manuf.*, vol. 34, nos. 7/8, pp. 874–887, 2021.

[123] W. Guo, "Partially explainable big data driven deep reinforcement learning for green 5G UAV," in *Proc. IEEE Int. Conf. Commun.*, Dublin, Ireland, 2020, pp. 1–7.

[124] S. Nahavandi, "Industry 5.0—A human-centric solution," *Sustainability*, vol. 11, no. 16, 2019, Art. no. 4371.

[125] Y. Fu, F. R. Yu, C. Li, T. H. Luan, and Y. Zhang, "Vehicular blockchain-based collective learning for connected and autonomous vehicles," *IEEE Wireless Commun.*, vol. 27, no. 2, pp. 197–203, Apr. 2020.

[126] S. H. Bouk, D. B. Rawat, S. Olariu, and R. Hussain, "Guest editorial introduction to the special issue on communication and computing technologies," *IEEE Trans. Intell. Transp. Syst.*, vol. 23, no. 11, pp. 22436–22441, Nov. 2022.

[127] P. Singh, Z. Elmi, Y.-Y. Lau, M. Borowska-Stefańska, S. Wiśniewski, and M. A. Dulebenets, "Blockchain and AI technology convergence: Applications in transportation systems," *Veh. Commun.*, vol. 38, 2022, Art. no. 100521.

[128] E. Henriksen, U. Halden, M. Kuzlu, and U. Cali, "Electrical load forecasting utilizing an explainable artificial intelligence (XAI) tool on Norwegian residential buildings," in *Proc. Int. Conf. Smart Energy Syst. Technol.*, Eindhoven, The Netherlands, 2022, pp. 1–6.

[129] D. Saraswat, P. Bhattacharya, M. Zuhair, A. Verma, and A. Kumar, "AnSMart: A SVM-based anomaly detection scheme via system profiling in Smart Grids," in *Proc. IEEE 2nd Int. Conf. Intell. Eng. Manage.*, London, U.K., 2021, pp. 417–422.

[130] M. H. Kabir, K. F. Hasan, M. K. Hasan, and K. Ansari, "Explainable artificial intelligence for smart city application: A secure and trusted platform," in *Explainable Artificial Intelligence for Cyber Security*. New York, NY, USA: Springer, 2022, pp. 241–263.

[131] P. Bhattacharya, C. Trivedi, J. Bhatt, P. Desai, and S. Tanwar, "LEAPS: Load and emission performance characteristics for sensor-driven green transport systems," in *Futuristic Trends in Networks and Computing Technologies*. New York, NY, USA: Springer, 2022, pp. 265–278.

[132] C. Xu, Z. Liao, C. Li, X. Zhou, and R. Xie, "Review on interpretable machine learning in smart grid," *Energies*, vol. 15, no. 12, 2022, Art. no. 4427.

[133] R. K. Behara and A. K. Saha, "Artificial intelligence methodologies in smart grid-integrated doubly fed induction generator design optimization and reliability assessment: A review," *Energies*, vol. 15, no. 19, 2022, Art. no. 7164.

[134] P. K. R. Maddikunta et al., "Incentive techniques for the Internet of Things: A survey," *J Netw. Comput. Appl.*, 2022, Art. no. 103464.

[135] J. M. Rožanec, B. Fortuna, and D. Mladenić, "Knowledge graph-based rich and confidentiality preserving explainable artificial intelligence (XAI)," *Inf. Fusion*, vol. 81, pp. 91–102, 2022.

[136] A. Estebsari and E. Werna, "Smart cities, grids, homes and the workforce: Challenges and prospects," in *Current State of Art in Artificial Intelligence and Ubiquitous Cities*. New York, NY, USA: Springer, 2022, pp. 17–39.

[137] T. Morimoto et al., "XR (extended reality: Virtual reality, augmented reality, mixed reality) technology in spine medicine: Status quo and quo vadis," *J. Clin. Med.*, vol. 11, no. 2, 2022, Art. no. 470.

[138] S. H. Choi, K.-B. Park, D. H. Roh, J. Y. Lee, Y. Ghasemi, and H. Jeong, "An XR-based approach to safe human-robot collaboration," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces Abstr. Workshops*, Christchurch, New Zealand, 2022, pp. 481–482.

[139] P. Vijayakumar and E. Dilliraj, "A comparative review on image analysis with machine learning for extended reality (XR) applications," in *Proc. Int. Conf. Ubiquitous Comput. Intell. Inf. Syst.*, 2022, pp. 307–328.

[140] S. Samuel, "Immersive technologies in the healthcare space," in *Augmented Intelligence in Healthcare: A Pragmatic and Integrated Analysis*. Berlin, Germany: Springer, 2022, pp. 433–451.

[141] A. Qayyum et al., "Secure and trustworthy artificial intelligence-extended reality (AI-XR) for metaverses," 2022, *arXiv:2210.13289*.

[142] N. Sharma, L. Saba, N. N. Khanna, M. K. Kalra, M. M. Fouda, and J. S. Suri, "Segmentation-based classification deep learning model embedded with explainable AI for COVID-19 detection in chest X-ray scans," *Diagnostics*, vol. 12, no. 9, 2022, Art. no. 2132.

[143] T. Senevirathna et al., "A survey on XAI for beyond 5G security: Technical aspects, use cases, challenges and research directions," 2022, *arXiv:2204.12822*.

[144] V.-L. Nguyen, R.-H. Hwang, P.-C. Lin, A. Vyas, and V.-T. Nguyen, "Towards the age of intelligent vehicular networks for connected and autonomous vehicles in 6G," *IEEE Netw.*, vol. 37, no. 3, pp. 44–51, May/Jun. 2023.

[145] R. Chengoden et al., "Metaverse for healthcare: A survey on potential applications, challenges and future directions," 2022, *arXiv:2209.04160*.

[146] P. Bhattacharya, A. Mukherjee, S. Tanwar, and E. Pricop, "Zero-load: A zero touch network based router management scheme underlying 6G-IoT ecosystems," in *Proc. IEEE 15th Int. Conf. Electron., Compute. Artif. Intell.*, Bucharest, Romania, 2023, pp. 1–7.

[147] M. Liyanage et al., "A survey on zero touch network and service (ZSM) management for 5G and beyond networks," *J. Netw. Comput. Appl.*, vol. 203, 2022, Art. no. 103362.

[148] Z. A. E. Houda, B. Brik, and L. Khoukhi, "Ensemble learning for intrusion detection in SDN-based zero touch smart grid systems," in *Proc. IEEE 47th Conf. Local Comput. Netw.*, 2022, pp. 149–156.

[149] T. Darwish, G. K. Kurt, H. Yanikomeroglu, G. Senarath, and P. Zhu, "A vision of self-evolving network management for future intelligent vertical HetNet," *IEEE Wireless Commun.*, vol. 28, no. 4, pp. 96–105, Aug. 2021.

[150] B. Dutta, A. Krichel, and M.-P. Odini, "The challenge of zero touch and explainable AI," *J. ICT Standardization*, vol. 9, pp. 147–158, 2021.

[151] Y. Wu, G. Lin, and J. Ge, "Knowledge-powered explainable artificial intelligence (XAI) for network automation towards 6G," in *IEEE Netw.*, vol. 36, no. 3, pp. 16–23, May/Jun. 2022, doi: 10.1109/MNET.005.2100541.

[152] B. Guembe, A. Azeta, and V. Osamor, "Explainable artificial intelligence, the fourth pillar of zero trust security," 2022. [Online]. Available: https://ssrn.com/abstract=4331547

[153] N. Ni and Y. Zhu, "Enabling zero knowledge proof by accelerating zk-SNARK kernels on GPU," *J. Parallel Distrib. Comput.*, vol. 173, pp. 20–31, 2023.

[154] Y. Ge, T. Li, and Q. Zhu, "Scenario-agnostic zero-trust defense with explainable threshold policy: A meta-learning approach," in *Proc. Conf. Comput. Commun. Workshops*, Hoboken, NJ, USA, 2023, pp. 1–6, doi: 10.1109/INFOCOMWKSHPS57453.2023.10225816.

Open Access funding provided by 'Università degli Studi di Enna "KORE"' within the CRUI CARE Agreement