IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 25 July 2022, accepted 6 August 2022, date of publication 8 August 2022, date of current version 16 August 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3197671

SURVEY

Explainable AI for Healthcare 5.0: Opportunities and Challenges

DEEPTI SARASWAT[1], PRONAYA BHATTACHARYA[1], (Member, IEEE), ASHWIN VERMA[1], VIVEK KUMAR PRASAD[1], SUDEEP TANWAR[1], (Senior Member, IEEE), GULSHAN SHARMA[2], PITSHOU N. BOKORO[2], AND RAVI SHARMA[3]

[1]Department of Computer Science and Engineering, Institute of Technology, Nirma University, Ahmedabad, Gujarat 382481, India
[2]Department of Electrical Engineering Technology, University of Johannesburg, Johannesburg 2006, South Africa
[3]Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun 248001, India

Corresponding authors: Sudeep Tanwar (sudeep.tanwar@nirmauni.ac.in) and Pronaya Bhattacharya (pronoya.bhattacharya@nirmauni.ac.in)

ABSTRACT In the healthcare domain, a transformative shift is envisioned towards Healthcare 5.0. It expands the operational boundaries of Healthcare 4.0 and leverages patient-centric digital wellness. Healthcare 5.0 focuses on real-time patient monitoring, ambient control and wellness, and privacy compliance through assisted technologies like artificial intelligence (AI), Internet-of-Things (IoT), big data, and assisted networking channels. However, healthcare operational procedures, verifiability of prediction models, resilience, and lack of ethical and regulatory frameworks are potential hindrances to the realization of Healthcare 5.0. Recently, explainable AI (EXAI) has been a disruptive trend in AI that focuses on the explainability of traditional AI models by leveraging the decision-making of the models and prediction outputs. The explainability factor opens new opportunities to the black-box models and brings confidence in healthcare stakeholders to interpret the machine learning (ML) and deep learning (DL) models. EXAI is focused on improving clinical health practices and brings transparency to the predictive analysis, which is crucial in the healthcare domain. Recent surveys on EXAI in healthcare have not significantly focused on the data analysis and interpretation of models, which lowers its practical deployment opportunities. Owing to the gap, the proposed survey explicitly details the requirements of EXAI in Healthcare 5.0, the operational and data collection process. Based on the review method and presented research questions, systematically, the article unfolds a proposed architecture that presents an EXAI ensemble on the computerized tomography (CT) image classification and segmentation process. A solution taxonomy of EXAI in Healthcare 5.0 is proposed, and operational challenges are presented. A supported case study on electrocardiogram (ECG) monitoring is presented that preserves the privacy of local models via federated learning (FL) and EXAI for metric validation. The case-study is supported through experimental validation. The analysis proves the efficacy of EXAI in health setups that envisions real-life model deployments in a wide range of clinical applications.

INDEX TERMS Explainable AI, healthcare 5.0, metrics, deep learning.

I. INTRODUCTION

Recently, there is a shift from hospital-centric to patient-centric view in the healthcare industry, which allows the patient to control the health operations. The shift is realized and supported through emerging disruptions in artificial intelligence (AI), Internet-of-Things (IoT), big-data, and assisted fog and edge networks. In a nutshell, digital health is equipped with smart sensors that form real-time prediction models and business analytics [1]. This patient-centric and sensor-driven analytical view is termed Healthcare 4.0, allowing smart and connected care to patients [2]. The healthcare industry has aligned its operations according to the healthcare 4.0 vision, but soon, the health industry will be at the dawn of

The associate editor coordinating the review of this manuscript and approving it for publication was R. K. Tripathy.

another paradigm shift [3]. The shift, termed Healthcare 5.0, would involve smart control, interpretable healthcare analytics, three-dimensional view models, and augmented and virtual reality [4]. Thus, healthcare would be pervasive, highly personalized, dynamic, and reason-based analytics, which would drive innovative business solutions in the health sector.

In healthcare 5.0, medical science technology foresees the interconnection of millions of IoT-based sensors that would communicate data through fifth-generation (5G) network infrastructure to provide digital wellness, smart healthcare and improved healthcare metrics. 5G and IoT combined with AI form a scenario where smart mobile wearables are integrated with mobile communication and medical technologies for easy and remote healthcare delivery. Advanced IoT devices attached to patients collect medical vitals, monitor progress, and diagnose health conditions [5], [6] to the doctor/medical institutions without significant human interaction. 5G in IoT promises 10 Gbps throughput, $<$ 10ms latency, secure future communications, increased cellular coverage, enhanced network performance, and enhanced battery lifetime by almost 90%. AI algorithms, like convolutional neural networks (CNN) or deep neural networks (DNN), perform complex operations on generated huge data sets such as image and text recognition, imaging and enable accurate disease prediction and detection and remote health treatment [7].

Due to the massive disruption of various AI technologies worldwide, questions have arisen regarding its impact on societal and individual issues. AI, therefore, needs to be utilized appropriately to ensure ethics, transparency, and accountability. This has led to the creation of Responsible AI. Explainable artificial intelligence (EXAI) is a technique that can be used for AI-enabled diagnosis and analysis to ensure the characteristics above. This will result in result tracing and model improvement in healthcare. EXAI is based on feature extraction to enable explainability and interpretability of the model [8]. EXAI defined a self-explanatory framework with design principles to understand and predict the behavioural aspects of ML/DL models. Medical AI induces a variety of objectives to technology designers, healthcare professionals, and social lawmakers as it imposes stringent requirements for reconsidering roles and responsibilities. EXAI finds various applications such as transportation, sales, finance, human resource, and healthcare, such as clinical support systems, disease detection and classification, drug delivery, medical image segmentation, maintenance and evolution in health care technologies, robotic-assisted surgery, and many others [9]. Explainability has to be defined to address specific perspectives such as technological, legal, medical, and patient for a clinical support system [10] to ensure end-to-end patient-doctor transparent operation.

In healthcare operations, EXAI is applied over different clinical decision models to address trusted analytics. It is used to manage medical data, clinical diagnosis, healthcare sensor bias reduction, disease classification, and segmentation [12].

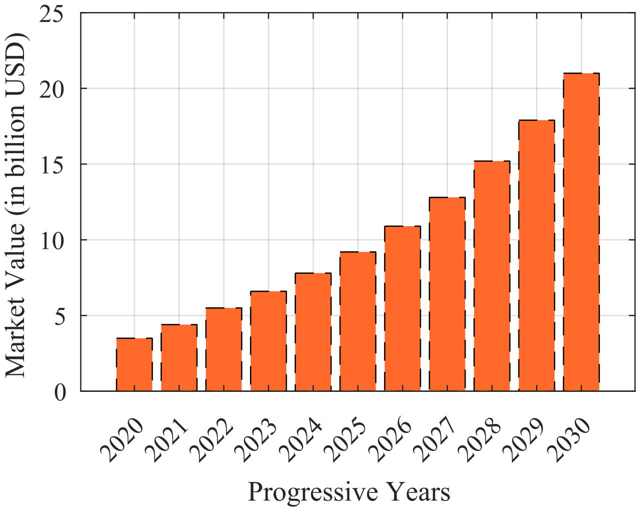

FIGURE 1. Global EXAI forecast [11].

It enables easy debugging and improves the performance of trained models through an added module that justifies the output decision of the model. Thus, EXAI adds transparency to AI algorithms to justify the model predictions. EXAI provides perfect explanations to end users behind its decisions and predictions and ensures compliance of ML/DL algorithms per the set parameters. It also allows the system to optimize the algorithm to reduce bias by overriding the system's decision. Thus, it brings safety and fairness to healthcare models and empowers people to believe in decision-making. EXAI applies to various ML techniques such as random forest, artificial neural networks, decision trees, and many others. The scope of the explanation in the EXAI model is categorized into local and global approaches. The global approach requires an explanation of the whole model, whereas the local approach requires an explanation of only the individual prediction [13]. The explainability requirement should be compatible with the set healthcare parameters and patient personalized health conditions to justify the alignment of the algorithm to the patient use-case, which preserves the personalization characteristics in healthcare 5.0 ecosystems. FIGURE 1 show the global EXAI forecast from the year 2020-2030 which indicates a compound annual growth rate (CAGR) of 18.4% [11]. TABLE 1 shows the list of abbreviations and the associated description used throughout the article.

A. MARKET TRENDS AND RESEARCH STATISTICS

EXAI is gaining importance in various application verticals like healthcare, retail and marketing, media & entertainment, aerospace & defence, insurance, financial services, industrial Internet of Things (IIoT) and many more. In the market, EXAI offers advantages like higher customer retention rate, better inventory management, high design interpretability, high performance and scalability, and reduced cost estimation. E.g. EXAI in the retail industry can predict the upcoming

TABLE 1. Abbreviations and their descriptions.

Abbreviations	Descriptions	Abbreviations	Descriptions
5G	Fifth Generation	EXAI	Explainable AI
6G	Sixth Generation	FL	Federated Learning
AE	Autoencoder	FTL	Federated Transfer Learning
AGI	Artificial General Intelligence	GAS	Global Aggregation Server
AI	Artificial Intelligence	Grad-CAM	Gradient-Weight Class Activation Mapping
AR	Augumented Reality	HCI	Human-Computer Interaction
AUPRC	Area Under the Precision-Recall Curve	IIoT	Industrial Internet-of-Things
AUROC	Area Under the Receiver Operating Characteristic	IoT	Internet-of-Things
B5G	5G and beyond	LIME	Local Interpretable Model-agnostic Explanations
BB	Black-box	LRP	Layerwise Relevance Propagation
CAGR	Compounded Annual Growth Rate	LSTM	Long-Short Term Memory
CAM	Class Activation Mapping	MIL	Multiple Instance Learning
CNN	Convolutional Neural Networks	MIT-BIH	Massachusetts Institute of Technology - Boston's Beth Israel Hospital
COVID-19	Novel Coronavirus disease-2019	ML	Machine Learning
CPHS	Comprehensive Personalized Healthcare Services	NFV	Network Function Virtualization
CRD	Center of Reviews and Dissemination	NHAI	National Highway Authorities of India
CT	Computerised Tomography	NLP	Natural Language Processing
CV	Computer Vision	NN	Neural Networks
DARE	Database of Abstracts of Reviews of Effects	PPV	Positive Predicted Value
DARPA	Defence Advanced Research Projects Agency	RNN	Recurrent Neural Networks
DL	Deep Learning	SDN	Software Defined Networking
DNN	Deep Neural Networks	SHAP	SHapley Additive exPlanations
DWS-CNN	Depth-wise Separable CNN	SVM	Support Vector Machines
DWT	Discrete Wavelet Transform	TI	Tactile Internet
ECG	Electrocardiogram	VGG	Visual Geometry Group
EHR	Electronic Health Records	XDM	Explainable Diagnostic Module

TABLE 2. Real-world industry projects of EXAI in healthcare domain.

Project Name	Objective	Duration	Company	Potential Outcomes
FWF Project on EXAI in Medical Domain [14]	Design library-based explainability patterns through grammar-based rules. This would map machine explanations on gathered data, and the project would impart an open framework and tool design	2020-2023	Holzinger Group, Vienna	Benchmarking tools and available libraries to build EXAI modules in model design
IBM Cloud Pak [15]	IBM and Geisinger Health System have designed a prototype that integrates an AI model and identifies the severity of high risk among novel coronavirus disease-2019 (COVID-19) patients	2021-Present	IBM	Specific EXAI modules to study the impact of sepis and COVID-19 mortality rates
Cancer Predictions on magnetic resonance imaging (MRI) images [16]	EXAI module design on CNN model to detect cancerous lesions based on specific MRI images, where suspicious areas would be identified and segmented	2019-2022	SAS Analytics and Solutions	Potential EXAI analytics on provided MRI images
ArgumeNtaTIon-Driven explainable artificial intelligence fOr digiTal mEdicine (Antidote) project [17]	Integrated EXAI module to comprehend low-level features of DL algorithms which are combined with high-level feature sets of human augmentation in healthcare datasets	2022-Present	Wimmics with a consortium of six universities in Europe	A dialogue-based module would be set up to collect data from patients and provide interpretable analysis and predictions that justify the healthcare diagnostics.
Centre of Healthcare Informatics, Australian Institute of Health Innovation [18]	The project aims to streamline basic EXAI requirements in healthcare and design models and frameworks to validate the EXAI outputs over DL models	2021-Present	Macquarie University, Sydney, Australia	Validation tools over standard healthcare datasets
Human-Centred AI Project [19]	Design methods to reenact the ML decision-making process, to comprehend learning and knowledge extraction on effective healthcare interfaces	2020-Present	University of Natural Resources and Life Sciences, Vienna, Austria	EXAI-based knowledge extraction and decision-learning module via human-computer interaction (HCI) interface
Sant' Anna School of Advanced Studies, Pisa [20]	Design augmentative abstractions on different AI methods that allow explanations to be drawn from dialogues between humans and machines	October 2020- 2025	A collaborative project between the Royal Academy of Engineering and JP Morgan Research Chair Francesca Toni	Augmentative dialogue generator via symbolic AI
Society of Medical Decision Making [21]	Project on EXAI module for cancer risk prediction algorithms and its importance in clinical analysis	February 2020- April 2023	Imperial College, UK, London	Analysis of EXAI as an interpretable tool for risk prediction over clinical cancer data

fashion trends and allows retailers/stockers to display the latest merchandise. In the e-commerce industry, EXAI can enable product search based on their stored recommendation. In business strategy development, EXAI provides accountability & insights into key business fundamentals such as sales, customer behaviour patterns, and employee turnover strengthens ethical business norms and prevents bias and loss of brand reputation. Recent market trends in EXAI observe significant advantages such as increased customer retention, enhanced management, and flaw detection. For example, in 2019, a third-party application developer leaked over 500 million Facebook profiles on Amazon Cloud Service in a fraud case [23]. In 2020, there was a cyberattack on the central server of the National Highway Authorities of India (NHAI) due to weak cybersecurity infrastructure [24]. In such scenarios, EXAI can provide insights into why such

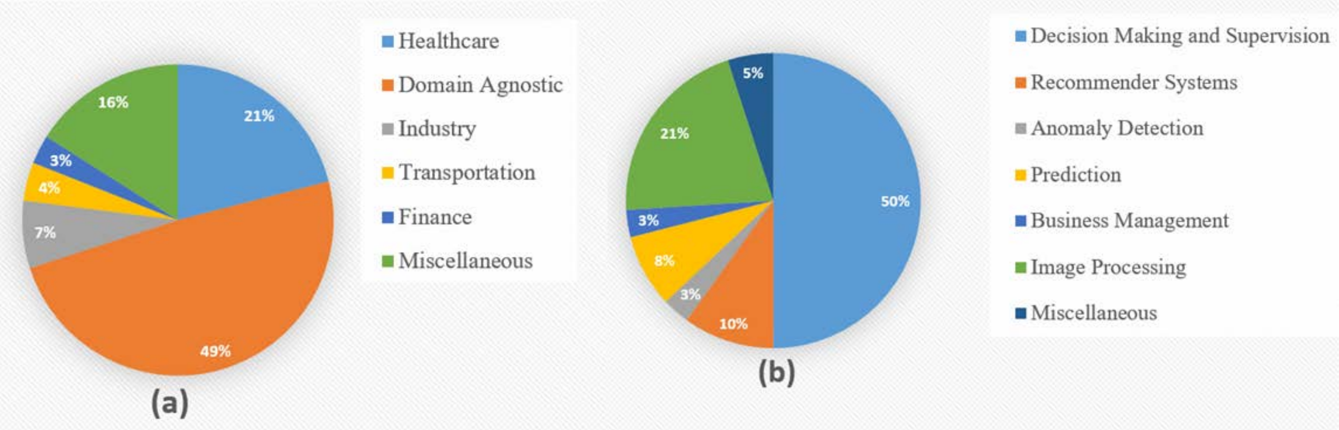

FIGURE 2. EXAI research statistics for different applications (a) Domain and (b) Task applied [22].

incidents happen and what steps could be taken to avoid such cyber attacks in future.

The field of EXAI has continuously gained momentum, producing knowledge from different outlooks, viz. philosophy, taxonomy, and development. FIGURE 2 represents the distribution of research articles about different application domains and tasks. FIGURE 2a shows domain agnostic distribution. The healthcare sector has gained influence, and EXAI is critical in providing explanations. FIGURE 2b shows the importance of EXAI in task-driven AI/ML applications to support decision-making in different applications like recommender systems, prediction, image processing, and business management.

B. SURVEY SCOPE AND ORGANIZATION MAP

FIGURE 3 pictures the organization of the sections and survey reading map. The remainder of this article is arranged as follows. Section II provides the methodology proposed for the literature survey. Section III presents the use cases and realization of EXAI in various applications. Section IV details the overview of EXAI, the requirement of data analytics and processing concerning healthcare aspects, and the existing state-of-the-art. Section V briefs the rise of ML/DL techniques and applicability of EXAI to enable transparency in the decision-making process, followed by the emergence of EXAI metrics in healthcare. Section VI details the proposed EXAI-enabled classification and segmentation architecture for healthcare 5.0. Section VII presents the solution taxonomy of EXAI specifically concerning the healthcare applications. Section VIII describes the future scope and research directions in EXAI. Section IX presents a proposed case study *ExoCOVID*, an EXAI-assisted ECG monitoring architecture in the healthcare ecosystem to derive explanations and interpretations understandable to the user. Section X presents the lessons learned from the overall survey. Finally, Section XI concludes the article.

C. NECESSITY OF THE SURVEY

In healthcare 5.0, analytics plays an important role in driving business solutions to patient needs. AI-based prediction models act as black boxes, where the health inputs are fed to the model and prediction output is gained. Generally, ML and DL algorithms work on complex and interrelated data; thus, the comprehension of the output is not straightforward, even for AI experts. In healthcare, the situation is critical, as it involves patient health and future clinical decisions to be based on current prediction outputs. Thus, it is imperative to understand exactly what these algorithms code. The algorithms consist of various test cases defined over the AI layer to cater to diverse healthcare use-cases. Thus, the interpretability and explainability of AI models are much needed in healthcare 5.0. Interpretable ML provide techniques to understand and validate how the ML model works and allows stakeholders to understand the basic knowledge of the decision and confidence in the former [25]. This makes AI algorithms have a white-box nature and allows transparency in analytics. Usually, humans easily understand linear models, but non-linear models are complex, for example, DNN models. In healthcare, the patient's future predictions depend on his current lifestyle and health-record indicators: time-varying, multi-connected, and non-linear. Therefore, it is required to derive understandable human explanations of the complex trained models.

Thus, the explainability of AI models needs to be qualitatively and quantitatively explored [26]. Medical AI applications require transparency and trust factor with doctors/medical practitioners for their acceptance and integration into actual practice. EXAI guides medical practitioners in interpreting black-box models and their decision-making process to verify particular decisions by ML algorithms which is very important in the medical field. A federated environment coupled with explainability will ensure data privacy & availability as well as real-time classification application to the doctors under critical conditions.

D. NOVELTY OF THE PROPOSED SURVEY

In existing state-of-the-art schemes, researchers have proposed explainability solutions in various practical healthcare applications such as COVID-19 detection and prediction [27], [28], cardio-vascular disease [29], [30],

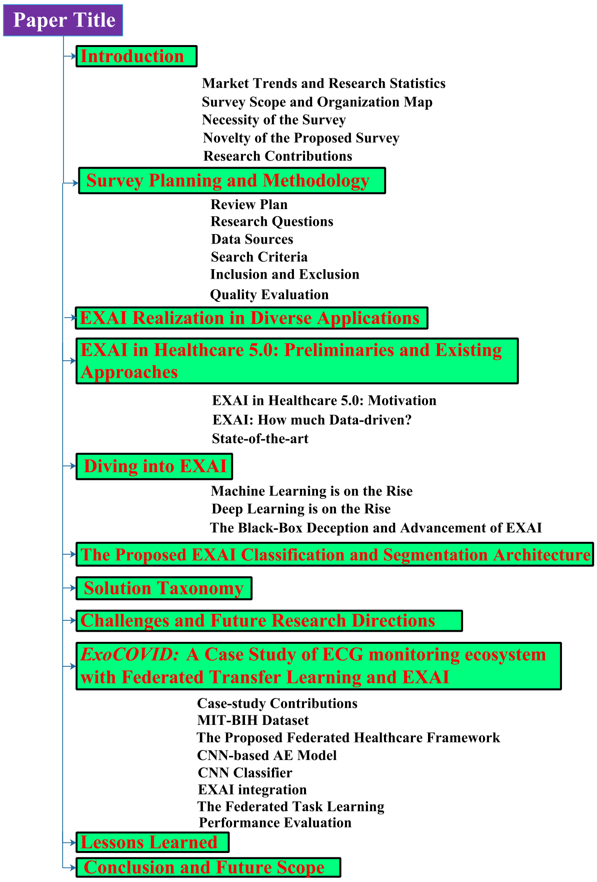
Paper Title
Introduction
Market Trends and Research Statistics
Survey Scope and Organization Map
Necessity of the Survey
Novelty of the Proposed Survey
Research Contributions
Survey Planning and Methodology
Review Plan
Research Questions
Data Sources
Search Criteria
Inclusion and Exclusion
Quality Evaluation
EXAI Realization in Diverse Applications
EXAI in Healthcare 5.0: Preliminaries and Existing Approaches
EXAI in Healthcare 5.0: Motivation
EXAI: How much Data-driven?
State-of-the-art
Diving into EXAI
Machine Learning is on the Rise
Deep Learning is on the Rise
The Black-Box Deception and Advancement of EXAI
The Proposed EXAI Classification and Segmentation Architecture
Solution Taxonomy
Challenges and Future Research Directions
ExoCOVID: A Case Study of ECG monitoring ecosystem with Federated Transfer Learning and EXAI
Case-study Contributions
MIT-BIH Dataset
The Proposed Federated Healthcare Framework
CNN-based AE Model
CNN Classifier
EXAI integration
The Federated Task Learning
Performance Evaluation
Lessons Learned
Conclusion and Future Scope

FIGURE 3. Survey and reading map.

biomedical engineering [31], structural health monitoring [32], mental health diagnosis [33] and many more. The solutions also explore techniques to utilize DL techniques and the implementation of algorithms to provide global and local explanations. However, no research focuses on explainability through an integrated approach per the built specifications. Traditional DL techniques rely on the central server, which is prone to attacks and raise security and privacy issues in critical healthcare. Moreover, to avoid violations of healthcare laws [34] A distributed AI-based learning framework needs to be adopted for healthcare data sharing concerns and healthcare data size limitations. The proposed research scheme provides end-to-end explainability for medical imaging applications through AI and federated transfer learning (FTL) in a healthcare 5.0 scenario supported by a case study of cardiovascular disease classification and its interpretability through the implementation of the EXAI environment.

E. RESEARCH CONTRIBUTIONS

The following are the research contributions of the paper.

- The paper presents a survey and explains the key concepts & attributes of EXAI. The survey also highlights the applicability in the healthcare 5.0 ecosystem.
- The paper proposes an end-to-end EXAI-enabled medical image classification and segmentation architecture. The architecture combines CNN-based DL techniques and FTL for COVID-19 detection. The proposed FTL-assisted CNN autoencoder classifier effectively denoises raw data collected from COVID-19 patients and classifies the former into five classes. Using the proposed explainability module, the explainable diagnostic module (XDM), the scheme interprets the classifier's prediction and provides the decision-making process.
- A solution taxonomy of EXAI in various industrial and societal applications is presented. A particular use-case taxonomy in healthcare is also proposed, integrating different AI techniques with EXAI.
- Future research scope and directions are discussed, and a case study, *ExoCOVID* for ECG monitoring with FTL and EXAI is presented. The case study uses the Arrhythmia dataset to train the proposed architecture using the clean and noisy version of the dataset. The framework is evaluated using precision, recall, and F1 score performance metrics and accuracy of ≈ 98% is achieved. The scheme provides desired interpretability and classification with an additional privacy protection feature due to the adoption of FL.

II. SURVEY PLANNING AND METHODOLOGY

The survey planning and methodology follow the standards and regulations by Kitchenman *et al.* [35], [36]. The survey is divided into six essential and analytic steps in the subsections below.

A. REVIEW PLAN

The survey highlights and exercises the problem definition and outline systematically. The critical points encompassed in the literature are (i) pointing to the research question, (ii) discerning the sources of data, research studies, and publication, (iii) logical conditions adopted for the search of keywords for desired research, (iv) insertion/deletion criteria identification (v) standardization and evaluation of the search and research writing.

B. RESEARCH QUESTIONS

The first step in the survey is jotting down the research question to assess the survey's objectives. These questions mainly focus on (i) Evolution and technology trends of EXAI in healthcare 5.0 applications and (ii) seamless integration of technological advancements like AI, 5G and beyond (B5G) networks, and FL in various applications to ensure a quality experience and interaction to the user (iii) lessons learnt from the survey and identification of the future scope of the survey in various human-centric applications. TABLE 3 highlights the identified research questions with corresponding objectives to assist the survey.

C. DATA SOURCES

The literature database employed for research is most relevant to computer science and medicine/healthcare. The literature explored IEEE Xplore, ACM Digital Library, PubMed etc.,

TABLE 3. Research questions of the proposed study.

Q. No.	Research Question	Objective
RQ 1	What is the need and importance of EXAI in healthcare applications?	To understand the emergence and features of EXAI to improve the understanding, trust, and efficiency of results of AI.
RQ 2	What are the limitations of current AI models, and how does EXAI help gaining trust to the required industry standards?	To explore and incorporate EXAI properties in trending DL and ML "black box" models for healthcare ecosystems to build user trust, satisfying legal and ethical requirements.
RQ 3	How does EXAI enable augmenting AI engines?	To discuss the insights of EXAI technology to achieve better AI modelling and interpretable explanations in natural language in healthcare 5.0 scenario
RQ 4	What are the challenges and open research for EXAI inclusion in various applications?	To formulate open research and challenges for seeking interpretability in AI systems.
RQ 5	How does EXAI benefit AI-healthcare?	To conceptualize the integration of EXAI and other AI technologies with superior privacy for healthcare 5.0 through a possible use-case scenario.

TABLE 4. Literature databases used to extract the raw literature corpus.

Name	Address
ACM Digital Library	https://dl.acm.org
IEEE Xplore	https://xplore.ieee.org
Springer Link	https://link.springer.com
ScienceDirect	https://www.sciencedirect.com
PubMed	https://www.ncbi.nlm.nih.gov/pubmed
arXiv	https://arxiv.org
SPIE Digital Library	https://www.spiedigitallibrary.org/
Wiley Online Library	https://onlinelibrary.wiley.com/

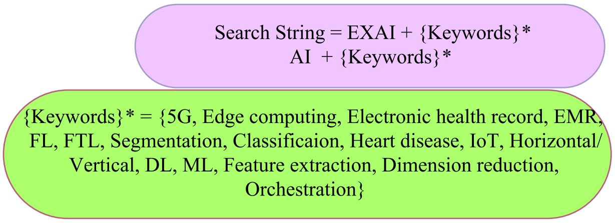
Search String = EXAI + {Keywords}*
AI + {Keywords}*
{Keywords}* = {5G, Edge computing, Electronic health record, EMR, FL, FTL, Segmentation, Classificaion, Heart disease, IoT, Horizontal/ Vertical, DL, ML, Feature extraction, Dimension reduction, Orchestration}

FIGURE 4. Keywords and search strings.

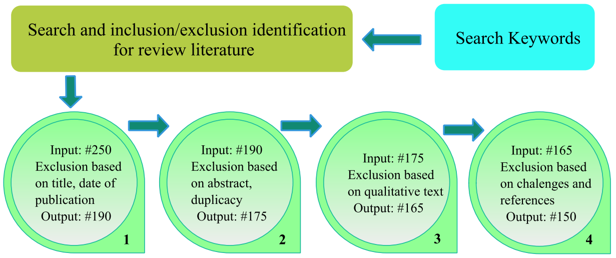
Search and inclusion/exclusion identification for review literature
Search Keywords
Input: #250
Exclusion based on title, date of publication
Output: #190
1
Input: #190
Exclusion based on abstract, duplicacy
Output: #175
2
Input: #175
Exclusion based on qualitative text
Output: #165
3
Input: #165
Exclusion based on chalenges and references
Output: #150
4

FIGURE 5. Narrowed-down search in inclusion/exclusion step.

for research. This database provides a rich source of information content. The study [35], [36] also recommends other electronic sources such as books, web blogs, preprints, articles, and patents for incorporation in the survey of interest. TABLE 4 shows the literary databases which are used to extract the articles related to our study.

D. SEARCH CRITERIA

An extensive literature survey was carried out on technologies related to EXAI, AI (deep learning, machine learning), FL & FTL, their combination and integration in the healthcare 5.0 scenario. The search is further narrowed down by collecting the initial corpus for the proposed survey and subsequent filtering. The searched online articles and corresponding references of the cited paper were included in the study. FIGURE 4 frames the search criteria for topics and articles, including the common synonyms.

E. INCLUSION AND EXCLUSION

After the defined search for the literature is completed, the other process involves filtering to eliminate results based on publication date, title and abstract review, and duplicate removal. Firstly, the academic repositories were searched for a combination of strings such as EXAI and healthcare. Then, papers with the keywords "XAI in healthcare", "explainable artificial intelligence in healthcare", "XAI and AI", "XAI for classification", and "XAI and FL" were searched. Finally, the database was searched with keywords XAI in healthcare for classification, XAI and FL in healthcare. OR keywords were utilized to aid fast database search. In parallel, the literature database search was optimized with keywords like FL, XAI, edge intelligence with XAI, CNN, classification, segmentation, and society 5.0. The papers with no potential interest and not containing all elements expected in the proposed work were excluded. FIGURE 5 represents the addition and deletion criteria for the proposed survey.

F. QUALITY EVALUATION

Finally, the evaluation was carried out on filtered as well as reference literature as per guidelines furnished by the Database of Abstracts of Reviews of Effects (DARE) and Center for Reviews and Dissemination (CRD) [35]. The evaluation carried out assures the quality of the assessment.

III. EXAI REALIZATION IN DIVERSE APPLICATIONS

In this section, we present the adoption of EXAI as use-cases in diverse applications. FIGURE 6 presents the specific areas where EXAI acts as a potential and interpretable mechanism to drive AI analytics. This section addresses the RQ 3 that presents how EXAI augments the functionality of traditional AI models. The details are presented as follows.

A. CUSTOMER REVIEW APPLICATIONS

Owing to the vast development of recommender systems in e-commerce applications, the customer review data has also significantly increased. Mostly, the data distribution contains

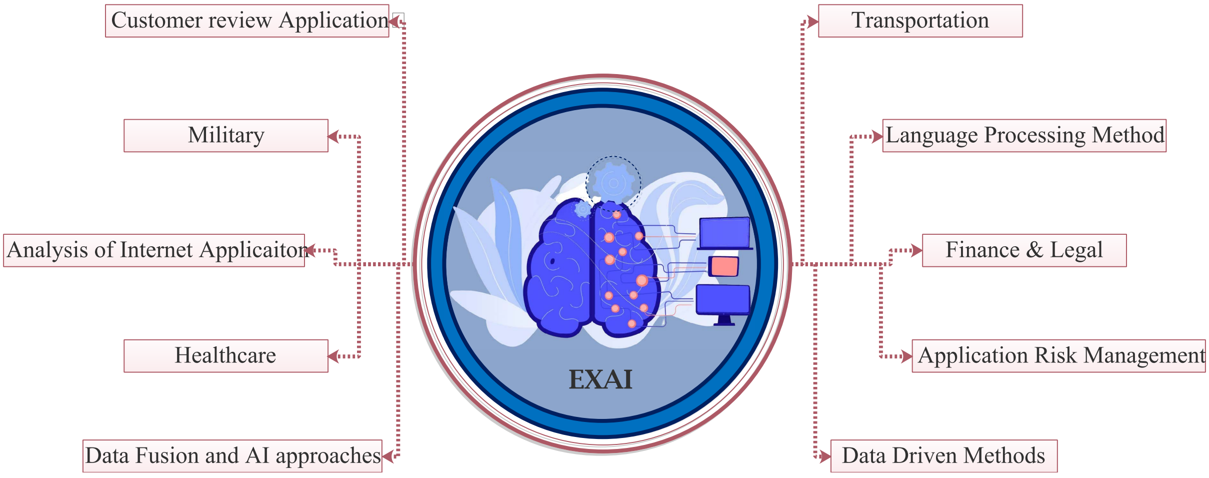
Customer review Application
Military
Analysis of Internet Applicaiton
Healthcare
Data Fusion and AI approaches
EXAI
Transportation
Language Processing Method
Finance & Legal
Application Risk Management
Data Driven Methods

FIGURE 6. EXAI realization in diverse applications.

positive, neutral, and negative product classification, which might be contextual and subjective [37]. In such cases, the AI models employ natural language processing and semantic analysis of contextual information to classify emotions and sentiments. In this direction, EXAI is a potential tool to derive meanings from syntactic data and map them to the semantic content. This would improve the natural language engines to classify emotions with higher accuracy [38].

B. MILITARY

In military applications, AI plays a major role in supporting training, defence simulation, and surveillance operations. Recently, Internet-of-Military-Things (IoMT) added IoT functionalities-connectivity, data, and services to military systems like army wearables (sensors to measure pulse and heart rate of soldiers in battlegrounds), UAVs for surveillance (motion and camera sensors), and proximity sensors to detect unwanted movements [39]. Thus, the military gathers a large amount of data which can be analyzed by ML and DL models that form self-control and adaptive IoMT environments. However, real-time sensor data is interval-based, contains noise, and might be inaccurate owing to sensor lags and malfunctions. In such cases, the prediction from ML and DL models would be inaccurate. The use of the autonomous system for defence or military operation is the same as using the autonomous vehicle and an automated system to treat any disease that may be life-threatening. Thus, EXAI is a potential tool that can capture biases in the data distribution, dependency, and inaccuracies in data balancing [40]. This provides timely information to military setups to reassess the collected data and apply data cleaning techniques to minimize the bias.

C. ANALYSIS OF INTERNET APPLICATIONS

The modern internet applications are data-driven and connect millions of socially networked users. Thus, these applications generate massive data, and therefore social networking analytics is on the rise, which helps us to investigate the decision process such as trust relationship [41], [42], and different opinions [43]. AI engines perform advanced ML to present content (posts, profiles, and advertisements) based on interests, demographics, and subscribed information. For example, applications like LinkedIn use DL for job recommendations and present social circles (depending on previous connections) to send you aligned posts and connect requests. Applications like Snapchat combines augmented reality (AR) toolboxes with computer vision (CV) algorithms to track your face movements and embed virtual objects in face space. Filters are applied while clicking photos in real-time through pixel correction and light enhancement engines. However, to increase the precision and accuracy of target models, EXAI modules like SHapley Additive exPlanations (SHAP), visual activation layers to support convolutional layers, occlusion sensitivity (to capture abstract image features in ConvNet models), and gradient-weighted class activation mapping (Grad-CAM) (makes CNN predict the importance of visual features) are used. This determines the decision-making logistics and makes the problem applicable to real-setups [44].

D. HEALTHCARE

In healthcare, AI has deep penetration to improve the analytics and prediction models and identify anomalies and diagnosis patterns. Thus, AI use-cases in healthcare image classification, segmentation, and disease predictions [45]. However, AI decisions in healthcare are critical, and EXAI plays an essential role in healthcare. EXAI techniques like Bayesian teaching, saliency maps, and others facilitate transparency in screening decisions regarding how the AI model arrived at the prediction and provide traceability in clinical output [46]. This is useful for deep models (imaging analysis) for applications like tumour segmentation, where data collection, labelling, and augmentation are critical. With EXAI, essential features are highlighted, which allows accurate predictions in the medical domain.

E. DATA FUSION AND AI APPROACHES

In big-data analytics, data from heterogeneous sources are fused and categorized as structured, semi-structured, or unstructured. Big-data techniques allow for effective policy designs, improving models' business intelligence. Tools like Apache Spark and Hadoop are used for effective processing and storing tools for voluminous data. Once data is aligned, data transformation (normalization and scaling) techniques are employed, which is followed by reduction principles (dimensionality, numerosity, and statistical) [47]. However, the fusion and reduction principles might lower the data complexity, which reduces the interpretability and observability of the data. In these cases, the employed ML or DL models have low precision. Thus, explainability in big-data analysis and data fusion models is critical to identifying the data's essential features, predicting the most important parts, and understanding the interaction between feature sets that collectively accounts for a prediction made by the model.

F. TRANSPORTATION

An autonomous vehicle provides capabilities such as sensing the environment without human interventions, identifying optimized routes from source to destination, providing human-free decision capabilities, decreasing accidents, and enhanced mobility in traffic & accident situations. It comes with several challenges for explainability to the AI, such as criteria for object detection, identification through sensors etc. The recent incident of Uber taxi has reported one human death due to the miss-classification of the human subject as a polythene bag [48]. The addition of explainability techniques can help prevent such incidents in future and requires a lot of research in the explainability enhancement of autonomous vehicles [49], [50].

G. LANGUAGE PROCESSING METHODS

Natural language processing (NLP) is an AI-based tool to process information based on linguistic distribution and different data from the users for opinion mining to recognize the accurate behaviour [51] and credit risk assessment [52]. Due to the advancement of data science, NLP based approach can handle a wide range of linguistic data [53], [54]. Challenges of NLP include sentiment analysis and the complexity of linguistic data processing. In such cases, EXAI techniques can help practitioners perform sentiment analysis to analyze the decision-makers, which helps reduce the complexity of linguistic data that allows for real-time and accurate conversions of natural language in different data analytics use-cases.

H. FINANCE & LEGAL

Financial services use AI to benefit customers through investment advice and portfolio management. The service provider's access the private information raises an issue of data security and transparent landing. Financial organizations are highly regulated and must follow specific laws to ensure fair and transparent transactions and portfolio management. Thus, challenges to the AI-based system in credit score are why the user has given such a credit score. Organizations like Equifax are working on an AI-based credit score model that generates automated code to explain the computed credit score. AI also has the potential to identify the repetition at the court. Transparency and fairness in making a decision are necessary. The researchers are working towards an automated decision-making system that explains what verdict is generated [55], [56]. EXAI is a powerful tool that classifies the discrepancies in user portfolios, risk assessment, and credit evaluation in these cases. In risk assessment, model agnostic methods are applicable, which explains the logic behind the correlation of the predicted variables. Giudici and Raffinetti [57] exploited Shapley values to perform statistical normalization, based on Lorenz Zonoids, to assess the financial risk. Thus, such applications greatly bring auditability to the credit systems.

I. APPLICATION RISK MANAGEMENT AND EFFECT ANALYSIS

To eliminate failure from the system, product, and services, systematic risk management, failure and effect analysis are required to analyze and evaluate the effect of loss. Due to several subjective and objective conditions, it is challenging to analyze and identify the failure node and its assessment. Thus, EXAI based approach is suitable to provide labels for opinion [58]. The work in [59] applied linguistic distribution to represent members that analyze and employed methods to determine the risk priority of a failure node.

J. DATA DRIVEN LEARNING METHODS

With the rise of AI and data science, the preference for learning methodology has also increased. Preference learning combines decision-making and ML, focusing on a group of attributes or individuals and models multi-group learning functions with historical data. With EXAI, preference learning is widely used for linguistic models [60], [61], [62]. In conjunction with preference learning methods, DL methods are used to set and estimate parameters for decision models, identify optimal weights of multiple attributes for the linguistic model, and identify the parameter for fusion aggregation function and distributed data. These methods are widely used in online recommendation [63] and financial risk assessment [64].

IV. EXAI IN HEALTHCARE 5.0: PRELIMINARIES AND EXISTING APPROACHES

This section explains the background of EXAI in various applications verticals like industry and automation, healthcare, industrial IoT, etc., and existing state-of-the-art in EXAI for multiple applications. There are three subsections, each explaining the necessity of EXAI. The first subsection explains the EXAI needs & features, EXAI important events along with the subsequent acquisition of wireless networks, from 1G to 5G deployments. The second subsection discusses the data-driven capability of EXAI. The third subsection discusses existing state-of-the-art concerning various healthcare applications. Therefore, the research question RQ 1 is addressed through this section put forward in survey planning and methodology. RQ 1 is addressed through the inception and potential of EXAI in AI-enabled healthcare applications. The data-driven capability of EXAI provides feedback and intelligent recommendations when AI is put into production.

A. EXAI IN HEALTHCARE 5.0: MOTIVATION

1) NEED AND BACKGROUND

Explainable artificial intelligence (EXAI) is defined as the system developed to make an AI system coherent to humans. The term "EXAI" was coined by Van Lent *et al.* [71] in 2004 that demonstrates the systems' capability that explains the behaviour of AI in gaming applications. However, with significant progress in ML applications, the focus has shifted toward implementing models and algorithms. EXAI, however, emerged as a decision-making entity due to the extremely high emergence of AI/ML across various sectors and industry verticals which may affect the critical decision-making process and effective recommendations, predictions or actions at the end of the user. EXAI minimizes the social and legal ethical issues and enables decision-making models to be explainable and understandable [72]. EXAI can demystify the black-box nature of AI/ML models by acting as a Responsible AI. However, there is a decision boundary between accuracy and interpretability of the model as it largely depends on the quality and quantity of training data sets. EXAI acts as a third-wave generation of AI to generate algorithms than can explain themselves precisely. TABLE 5 tabulates the key concepts of EXAI.

EXAI associates the interface between decision-maker (models) and humans. The gap synchronizes the comprehension of humans and accurate representation of decision-maker [73]. EXAI provides an explainability of AI/ML models to the end-user. There are different types of application domains categorized by [74] as explained below.

- *Stage:* This explainability describes the time process that a model undergoes to explain the decision. There are two types of stages. The first stage is called *Antehoc* where transparent explanations are generated from the beginning of the data training process to achieve optimal performance (e.g., fuzzy logic, tree-based modelling). The other stage is called *Post hoc* where an external model imitates the base model's behaviour to provide explanations to the user, for example, the support vector machines (SVM), and neural networks (NNs).
- *Scope:* This explainability explains the extent of the scope of explainable methods. In *global* scope, the model technique is transparent to the user, while in *local* scope, the model explains the single instant to the user.
- *Input-Output:* This explainability explains the model in the form of images, numbers, texts, numeric, rules etc., to the user.

EXAI is essential for benefits encompassing commercial, ethical and regulations if users understand and manage AI results effectively. There are four primary reasons for EXAI requirements, viz. to justify, control, improve, and discover.

2) EVOLUTION TIMELINE

FIGURE 7 timelines the important events in EXAI adoption in healthcare verticals to subsequent integration with EXAI along with a shift from fourth-generation (4G) to 5G wireless networks. In 1956, John McCarthy coined the term AI which describes the science behind intelligent machines [75], [76]. AI has evolved continuously over the last five decades with the inclusion of ML/DL techniques to change the algorithm-only perspective to a personalized medicine perspective. Some of the advantages of the use of AI in medicine include disease diagnosis, therapeutic response predication, improved workflow, procedure accuracy, therapeutic monitoring & clinical operations, significant improvement in overall patient outcome, and the potential creation of preventive medicine in the near future [77].

In the early 2000s, the era of DL commenced, which has marked a significant advancement to enable classification over individually trained datasets compared to traditional ML methods. Till 2020, there was subsequent evolution of various DL training algorithms viz. CNN, Le-NET, AlexNet, visual geometry group (VGG), GoogLeNet, and ResNet. At the same time, healthcare 2.0 and 3.0 came into existence that focused on electronic health records (EHR) and its decentralization, efficient communication, intelligent wearables, human-to-machine intervention, and data-driven & cost-efficient applications to form a patient-centric ecosystem. In 2016, advancement in AI technology took place along with the inception of healthcare 4.0 with goals of real-time data collection, multimedia, personalized medicine & connected care, and integration of both big data analytics and AI for effective decision making for analyzing and storing EHRs. Healthcare 5.0 proposes patients are at the centre of the healthcare ecosystem and introduces lightweight IoT solutions with integration of 5G/6G communication and are coupled with a security mechanism to form a patient-centric tactile model [78], [79].

A variety of research took place mid-2016 in EXAI. It came into full-fledged analysis-oriented existence in 2017 when Defense Advanced Research Projects Agency (DARPA) launched its EXAI program to develop an AI system to enable trust and understanding to end-users through a projection of learned models and their decisions [80]. A group of academicians called FAT also carried out a study in 2018 to bring transparency to explainable system [81]. EXAI can drive various AI-relying domains such as transportation, healthcare, legal, finance, military, industrial IoT, etc. The year 2020 and beyond marked advanced AI and its explainability in the healthcare sector

TABLE 5. Key terminologies in EXAI.

Term	Description
Interpretable ML	A system which enables a user to visualize and study how inputs are mathematically mapped to outputs. The term is often interchangeable with explainability in certain ML contexts [65], [66].
Black-box (BB) problem	In the context of computing and quality assurance, the black-box subsystem does not reveal anything about the internal design, structure and implementation. The black-box concept protects intellectual property and maintains competitiveness. In the context of AI, the black-box problem refers to the difficulty of the system to provide suitable explanations to the answer; the former has arrived.
Responsible AI	Takes into account societal, ethical and moral values. There are three entities of responsible AI viz. accountability, responsibility, and transparency [67].
Data Science	Data science field unifies statistics, data science and ML to understand the actual phenomenon with the data.
Social Science	Social science deals with social relationships and among individuals and society through explanations [68].
Third-wave AI	Third-wave AI refers to the next generation of AI systems, where constructed AI models explain real-world phenomenon. Some of the examples are intelligent AI, brain-computer interface, and human symbiosis [69].
Artificial General Intelligence (AGI)	Ability of an agent or machine to understand or learn any intellectual task like a human being. It is also referred as strong AI or general intelligent action [70]

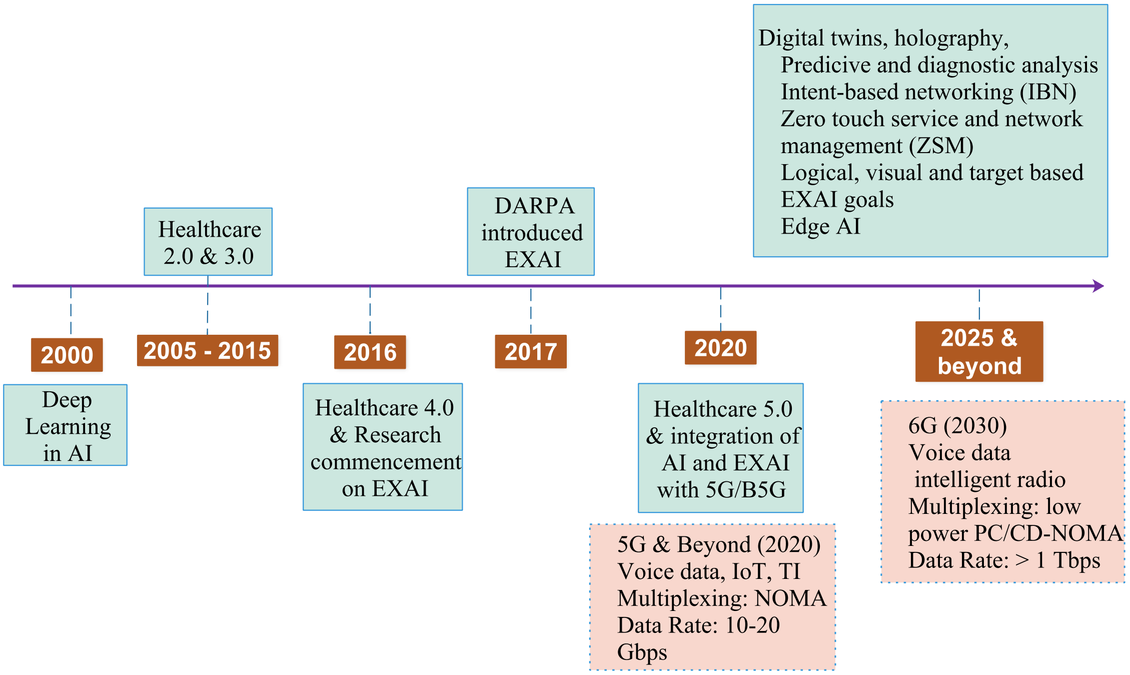
Digital twins, holography,
Predicive and diagnostic analysis
Intent-based networking (IBN)
Zero touch service and network management (ZSM)
Logical, visual and target based EXAI goals
Edge AI
Healthcare 2.0 & 3.0
DARPA introduced EXAI
2000
2005 - 2015
2016
2017
2020
2025 & beyond
Deep Learning in AI
Healthcare 4.0 & Research commencement on EXAI
Healthcare 5.0 & integration of AI and EXAI with 5G/B5G
5G & Beyond (2020)
Voice data, IoT, TI
Multiplexing: NOMA
Data Rate: 10-20 Gbps
6G (2030)
Voice data
intelligent radio
Multiplexing: low power PC/CD-NOMA
Data Rate: > 1 Tbps

FIGURE 7. Timeline of events in EXAI.

to address various domains like heart, cancer, skin disease, diabetes, asthma, COVID-19 [82] etc. through DL, feature extraction, dimension reduction, visual relevance, and trust-based decision-making algorithm to enable vertical healthcare applications like drug development, disease prediction & diagnosis, population health, automated healthcare, practice and consultation. 5G standards have resulted in massive IoT phenomenon into reality. 5G supports high device density, data rates, reliability, fully softwarized network infrastructure, like software-defined networking (SDN) and network function virtualization (NFV), and an AI-powered automated network management system to support critical operations like autonomous driving and healthcare. 5G supports tactile internet (TI) and utilizes an efficient multiplexing scheme and successive interference cancellation at the Tx/Rx system to ensure low power multiplexing is primarily used in IoT-based systems. EXAI-enabled 5G/B5G network would facilitate real-time massive data collection through additional information of AI-powered black box to ensure transparency and authenticity in upstream and downstream data processing and bridges the gap between end-user and service provider in an AI-powered 5G network.

B. EXAI: HOW MUCH DATA-DRIVEN?

Data-driven EXAI methods generate explanations solely from data analysis and interpretation without any external input information, or prior knowledge [83]. Data-driven strategies are initiated to provide explanations by selecting a dataset (distribution in the form of global or local). The dataset chosen or its variation is given as an input to the black-box model. Human-friendly explanations are then generated based on some definite analysis of predictions of the black-box model. There are various forms of data-driven explanations, e.g., features importance and decision rules. The data-driven explanations are classified in three states, i.e. global, local, and instance-based [84]. Global methods are classified into three subclasses, viz. model extraction, feature-based, and transparent model design. Local methods are classified into two subclasses, viz. Local approximation and Propagation-based methods while instance-based methods are classified into two subclasses, namely, the prototypes and criticism and counterfactuals.

C. STATE-OF-THE-ART

This section presents the existing state-of-the-art EXAI systems concerning various applications. Authors in [85] proposed EXAI as an effective tool for effective prediction and analysis of the diagnosis of disease through AI-enabled health data. They have explained EXAI as a generalized tool in conjunction with clinical knowledge to improve AI systems' predictions to improve traceability and accountability. Authors in [86] surveyed recent and current trends in surgical applications and medical diagnosis using EXAI based on findings across various research platforms, described various EXAI-enabled methods for medical EXAI applications and finally discussed the achieved challenges and research directions. The authors also propose a case study for accurate breast cancer predictions and diagnosis. Authors in [10] describe EXAI in healthcare in a multidisciplinary manner to analyze the relevance from legal, medical, patient, and technological perspectives. The authors deduce a set of outcomes for applicability of perspectives and conclude the importance of EXAI in the clinical system from an ethical and individual point of view. The work in [87] describes a survey of recent advances in healthcare applications and the current status of EXAI. The authors explain how EXAI leverages multi-modal and multi-centre data fusion through two descriptive clinical-level case studies. COVID-19 classifications and hydrocephalus segmentation and conducted analysis from both quantitative and qualitative points of view. Authors in [88] present a comprehensive survey on EXAI concerning the motivations and applicability in intelligent systems like transportation, healthcare, legal, military, and finance. The survey also presents various angles of EXAI, such as taxonomy, measurement strategies, human perception, and prediction, and focuses on open issues and challenges. Authors in [89] survey the interpretability and explainability of ML algorithms and critically interpret the suggestions into two distinct categories, perceptive interpretability and mathematical structure interpretability in the medical context and also discuss open challenges and prospects. Authors in [90] outline EXAI's importance in 6G wireless networks. The methodology includes public & legal motivations, trade-offs between performance & explainability, and XAI algorithms and adopts a case study on wireless MAC and PHY layer optimization. Authors in [91] present a survey on comprehensive, personalized healthcare services (CPHS) in healthcare 5.0 using different AI approaches. The author presented a three-tier IoT-based healthcare architecture and a discussion of the security aspect. Finally, the authors propose the methodology to combat existing approaches based on three factors: reliability, resilience, and personalization. The work in [92] describes existing AI techniques like ML/DL/NLP in and extends the survey to explain the importance of EXAI in future medicine and biomedical applications. Authors in [29] implements analyze the medical ECG data on proposing a generalized model, *ST-CNN-GAP-5*, that implements DNN algorithms using online available two ECG datasets with the achieved accuracy of 95.8% and AUC value of 99.46%. The dataset is analyzed using SHAP for explainability. Authors in [93] develop an explainable machine to predict patients with nerve sparing, enabling surgical planning and patient counselling without any risk. The clinicopathological data samples were taken from 900 lobes, modelled using logistic regression techniques and validated with the existing data sets. The accuracy of the performance metrics was 81% in the area under the receiver operating characteristic (AUROC) and 69% in the area under the precision-recall curve (AUPRC). Authors in [28] discuss AI prospects and framework for EXAI-based challenges to tackle COVID-19 disease. Authors in [94] present *EXAM*, an explainable attention-based method for automatic diagnosis of COVID-19 utilizing graphical interpretation. The proposed approach properly calibrates the model for implicit explainability. It effectively differentiates main and redundant traits through spatial and channel-wise focus processes and robust performance to model chest X-ray and CT-scan image datasets.

Authors in [103] presents a generalized taxonomy of EXAI based on current challenges and future directions. The proposed taxonomy incorporates the EXAI database models, reviewed taxonomies, and a decision tree approach to select the best taxonomy for desired applications. The authors in [104] present a data collection methodology for EXAI-assisted responsible AI healthcare. They present a case study for pregnancy and risk prediction and found that high-risk women patients are likely to adopt intelligent solutions when health provider is in the public domain. The authors in [105] propose a cloud-centric cybertwin-based DL multi-modal system for ECG pattern detection in the backdrop of a 6G communication network. The system monitors data from different IoT body sensor networks and smart devices, detects the motion patterns, processes them, and provides health reports to the clinicians. TABLE 6 compares the existing state-of-the-art research with the proposed survey.

V. DIVING INTO EXAI

EXAI is a novel research area in ML that targets how AI systems respond to black-box decisions. Through EXAI, one can create recommendation systems for the healthcare systems also. Many ML algorithms cannot explain how and why a choice or decision was reached. This is especially true of the most widely used deep neural network techniques today. As a result, the lack of understandability in these black-box models can undermine our trust in AI technologies. Although deep neural networks can provide a tremendous payoff in performance, the EXAI is becoming increasingly crucial for DL-based powered applications, particularly in medical

TABLE 6. Comparison of existing state-of-the-art research and the proposed survey.

Author	Year	Objective	1	2	3	4	Method/Technique	Pros	Cons
Karim M. *et al.* [95]	2020	Exploits explainable prediction for COVID-19 based on chest X-ray images	Y	Y	N	N	EXAI, DNN, class activation mapping (CAM), layerwise relevance propagation (LRP)	The proposed model classifies COVID-19 with positive predicted value (PPV) and recall metrics of 96.12% and recall 94.3% and outperforms traditional methods	Does not considers external factors and symptoms for COVID-19 possible causes
Ahsan *et al.* [96]	2020	Development of DNN for COVID-19 prediction using CT and chest X-ray images and interpretability through local interpretable model agnostic explanations (LIME)	Y	Y	N	N	DNN, LIME	Provides accuracy of around 95% for chest and X-ray with 95% confidence level	EXAI metrics are not considered
Nazar *et al.* [97]	2021	Studies human-computer interaction with EXAI in healthcare using AI techniques	Y	Y	N	N	AI, HCI, EXAI	Presents a discussion on HCI, AI, and EXAI based on state-of-the-art literature. The survey also discusses the relevance of EXAI in healthcare and challenges in past literature surveys	Does not discuss solution taxonomy for EXAI in healthcare for use case scenario
Le *et al.* [98]	2021	COVID-19 diagnosis using CNN using deep support vector machine in 5G-enabled IoT environment	Y	N	N	Y	IoT, depth-wise separable CNN (DWS-CNN), Gaussian filtering	Propose DWS-CNN model detects binary and multiple classes of COVID-19 disease using data acquisition, preprocessing using Gaussian filter, feature extraction, and classification processes	Does not explain IoT communication model assisted by 5G and lacks solution taxonomy
Tan *et al.* [99]	2021	Deep-learning and 5G-assisted real-time cardiovascular monitoring for COVID-19 patients	Y	N	N	Y	5G, DL, CNN, long short term memory (LSTM)	Propose cardiovascular system monitoring using 5G-enabled IoT wearables and data processing using DNN	Does not explores integration of DNN, LSTM and CNN for effective prediction and generalization
P. Angelov *et al.* [100]	2021	Reviews state-of-the-art explainable AI techniques in ML/DL domain	N	Y	Y	N	Prototype-based models, Surrogate models	Provides an analytical view and challenges of explainable AI based on national standards and proposes a solution taxonomy	Does not explores explainability aspect through a case-specific healthcare application
Taimoor *et al.* [91]	2022	Discusses reliable and resilient AI and IoT enabled personalized healthcare services in healthcare 5.0 domain	Y	N	Y	N	AI and non-AI approach, IoT	Propose comprehensive, personalized healthcare services by contextualizing Healthcare IoT to support clinical personalization, discuss AI and non-AI based techniques as well as use case example	Lacks evaluation of the scheme for a particular medical case as well as integration of EXAI
Fuhram *et al.* [101]	2022	Reviews explainability and interpretability for COVID-19 imaging applications	N	Y	N	N	Feature-based distribution, Region identification	Provides identification of several tasks, an overview of recent explainable techniques in imaging scenarios, and recommends best practices in explainable/interpretable AI implementation	Does not discuss existing AI techniques for COVID-19 classification and its solution using EXAI. The survey is also limited to particular healthcare scenario
Jagatheesa perumal *et al.* [102]	2022	Survey on provisioning secure healthcare system using emerging technologies	Y	N	Y	Y	IoT, 5G, AI, Big-data	Provides a deep discussion of IoT, 5G, AI, and big data analytics for secure healthcare system along with the presentation of case study	Does not provide an integrated approach for secure healthcare system
Speith *et al.* [103]	2022	Reviews approaches and challenges of EXAI using taxonomy	N	Y	Y	N	Functioning-Based, Conceptual, and Result based approach, Decision Tree fitting	Enables researchers to have current knowledge of state-of-the-art challenges in EXAI and overcome the challenges through a proposed taxonomy using decision fit approach	Does not discuss subcategories for feature relevance methods
Oprescu *et al.* [104]	2022	Presents a data collection approach for responsible AI intelligence in healthcare application	Y	Y	N	N	AI, Responsible AI	Proposed finding helps in providing a trustworthy AI-based solution for pregnant women and concludes the importance of explainability while adopting AI-based solutions and applications	Data collection is limited by participants' geographical diversity, personal information, and COVID-19 scenarios
Qi *et al.* [105]	2022	Presents a DL cybertwin-enabled multimodal network for ECG data monitoring	Y	N	N	Y	Cybertwin, DL, 6G, Data fusion	Proposed IoT sensor-based architecture for real-time collection, processing, monitoring of ECG patterns & activity through a 6G-enabled cybertwin with robust accuracy and lower computation time	Does not provide explanations to the accuracy of the DNN model
Proposed	2022	Proposes a integrated EXAI-assisted architecture for classification of COVID-19 patients in healthcare 5.0 environment	Y	Y	Y	Y	5G, AI, EXAI, CNN, Federated transfer learning	Utilizes AI-enabled integrated architecture that explains performance metrics using EXAI at the backdrop of 5G communication network	-

1-AI,2-EXAI,3-Solution Taxonomy,4-5G/6G, Y-parameter is taken into account, N-parameter is not taken into account.

and pharmaceutical investigations. Inadequate explainability and transparency in most of the existing AI systems are the major reasons for the infrequency of successful integration and implementation of AI tools into clinical practice. With increasing advancements in computational capabilities and extremely massive data rise, AI has made significant advances in increasing intellect, richness, and automated functionality in recent years. AI also covers a wide range of computer and signal processing-related studies. The research question RQ 2 thus addresses the limitations of traditional modelling through a unified framework to help understand and interpret the decisions of ML/DL models.

A. ML IS ON THE RISE

Hard-coded rules are typically used in traditional AI methods, which explain how to solve a specific problem. In contrast, an ML framework uses the power of a massive amount of data (both instances and metadata). The former then figures the attainment of results in an effective manner. In supervised ML, labelled data is utilized for learning, while unsupervisedlearning does not require labels [106]. Reinforcement learning uses knowledge and experience of past environments to apply to the present environment. In traditional reinforcement learning scenarios, the agent interacts with the surrounding, obtains a reward function and optimizes the strategy to obtain maximum rewards for significant improvement.

ML requires high computational capabilities to process a large amount of data to gain final results. Statistical techniques are integrated with ML to maximize the probability of decision-making. On the other hand, statistics frequently involve areas of research unrelated to the creation of algorithms that can learn to make projections or judgments based on data. Complicated numerical ML algorithms do not automatically converge despite dependencies on data science and analytics [107]. Symbolic approaches are also used in AI that employs logic and inference to generate problem interpretations and arrive at a solution.

B. DL IS ON THE RISE

DL is a collection of methodologies that have revolutionized ML in recent years. DL is a collection of methods that create neural networks with deep layers rather than a single algorithm. DL networks are complex and require extremely high computational capabilities and the implementation of node clusters. DL algorithms consistently outperform other algorithms in image detection, natural language processing, and speech synthesis [108]. The accuracy of DL models entirely depends on data size. However, DL poses disadvantages such as expensive data model training, sophisticated hardware requirements, lengthy development cycles, unavailability of skills to develop a model for user-defined applications, unavailability of knowledge of how DL algorithm converges, and the black-box nature.

C. THE BLACK-BOX DECEPTION AND THE ADVANCEMENT OF EXAI

Today's DL methods produce a highly reliable result; however, the procedures are opaque, incomprehensible, untrustable, and provide a black-box nature [109]. A trustable AI decisive network is required to incorporate ethical standards, policy discussions and various forms of AI (responsible, EXAI, valid, machine-level, human-level etc.) EXAI represents various system classes of how an AI system makes judgments and predictions. EXAI simplifies decision, accountability, and priority and enables researchers to grasp the insights of research outcomes by providing understandable descriptions of how AI systems conduct their research. There are specific desired properties/terminologies of EXAI which are widely used in scientific, public, and strategy conversations.

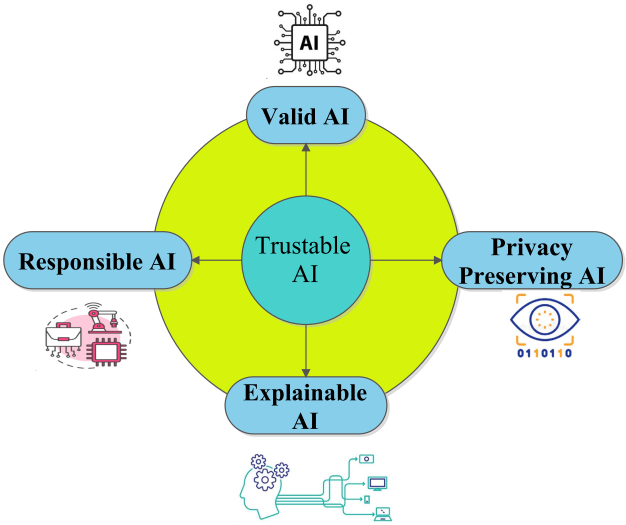

FIGURE 8. Attributes of the trustable AI.

- *Interpretability*: It refers to a sense of understanding of how Artificial intelligence works.
- *explainability*: It explains how a choice was made to a wider variety of consumers.
- *transparency*: It assesses the information or model's degree of availability.
- *Justifiability*: It denotes a grasp of the facts to support a definite conclusion.
- *Contestability*: It indicates how consumers can challenge a judgement.

FIGURE 8 represents the attributes of trustable AI. AI can no longer be regarded as a "black box" that receives input and generates output without a clear grasp of what is happening within. People must understand how an AI system arrives at its conclusions and suggestions to trust its decisions, whether ethical or not. Creating trustworthiness in AI entails establishing transparency, ethics, and responsible AI. For transparency, humans must be able to see how the AI makes judgments and what data it uses. Without visibility, it is impossible to analyze and examine the rationale behind AI decisions. Transparency allows people to improve their processes by seeing where they fail and make errors. The reputable and trusted AI framework is intended to assist businesses in identifying and mitigating potential risks linked to AI ethics at all stages of the AI life cycle.

Trustworthy AI is a concept used to define AI that is legal, ethical, and technically sound. The trustable AI will be, classified as Valid AI, Responsible AI, Privacy AI, and Explainable AI. Valid AI is significant because it gives organizations insights into their operations that they may not have been aware of earlier. In many situations, a valid AI can execute tasks better than people.

Responsible AI is the discipline of designing, developing, and implementing AI to empower employees and enterprises while having a fair influence on customers and communities. Data without privacy and the unlawful use of AI can negatively impact both reputations and the system. Hence, privacy-preserving plays a vital role in the Trusted AI. To implement the same, the Companies must incorporate confidentiality, transparency, and security into their AI initiatives and ensure that data is acquired, used, managed, and kept securely and responsibly.

EXAI is a set of techniques and strategies that enable human consumers to understand and trust ML algorithms' results and output. It refers to a model's projected influence and potential biases. Many ML concepts are inherently understandable and explainable, such as linear models, decision trees, and rule-based models. They are referred to as seamless models or white-box modelling techniques. However, these models are pretty less effective. A model-specific and post-hoc EXAI methodologies can be developed using advanced ML and DL models. Explanation by simplistic analysis, architecture reconfiguration, feature relevance elaboration, and visual elaboration are commonly used approaches. Precisely, more sophisticated models can be more effective while decreasing explainability. The model-agnostic methodology utilizes a replacement mechanism for forecast explanation. They are frequently used post-hoc to describe deep neural networks using local and global surrogacy. There are different models of EXAI, as explained below.

- *Global EXAI for Specific Models:* These techniques can restrict interpretability to improve understandability. The structural constraints such as sparsity and monotonicity utilize fewer inputs and are monotonous in nature. Prior meaningful information may also be used to limit the higher-level functions that can be formed from the information.
- *Local EXAI for Specific Models:* Local EXAI approach provides an interpretation for a given occurrence. New attention techniques explain the role of various multi-dimensional data in explaining a sample instance. For example, an attention module for the recurrent neural networks (RNN) will also be used to describe the visuals of the model in the clinical description.
- *Model-agnostic Global EXAI:* A proxy representation is built in model-agnostic global EXAI to approximate a definable subsystem for the behaviour-based model. For example, an interpretable decision tree model on how clinical symptoms influence therapy response may be utilized for the approximation of a complicated DL model. The "IF-THEN" condition can clarify the significant role of dynamic values in the diagnosis of clinical parameters. Clinical specialists can trust the DL model if specific clinical signs are proven to be rational, and confounding sounds can be precisely eliminated. Diagnostic techniques can also be used to get insight into the importance of individual attributes in the model's projections. Individual conditional expectation can be used to clarify the impact of a feature on different instances and observe the variation of impact in different instances. A partially dependent graph, for example, can reveal the involvement of medical manifestations in a patient's favourable reply to therapeutic intervention, as detected by a computer-enabled healthcare system.
- *Model-agnostic Local EXAI:* The goal of this type of EXAI technique is to generate model-agnostic explanations for a specific case or the area around a specific instance. The well-validated tool, LIME, can explain a complex DL model in the vicinity of an exemplar. Considering a DL system that categorizes high risk for a particular human feature for specific diseases or mortality rates requires a post-doc explanation from the doctor. The interpretable modules are tampered with to see how the changes in physiological attributes affect the forecasts. A linear model is learned for this perturbation dataset, and more weights are assigned to the instances near the physiological characteristic. The essential subsystems of the above model might reflect the impact of a specific human property, which demonstrates more risk or vice-versa. This might provide clinicians with a clear way to interpret the classification.

VI. THE PROPOSED EXAI CLASSIFICATION AND SEGMENTATION ARCHITECTURE

This paper describes two common yet crucial uses of EXAI combined with feature extraction and classification. An EXAI algorithm is utilized for CT image classification for COVID-19 patients and hydrocephalus segmentation using MRI & CT datasets. This section addresses the research question RQ 5 through EXAI and AI integration proposal in the healthcare ecosystem.

Dataset: To preserve privacy, the CT data is acquired from four local hospitals in China. The CT cross-centre data from a total of 804 patients (380 quantities - COVID-19 positive, 424 quantities - COVID-19 negative (normal)) was collected from four different hospitals (A, B, C, D) to ensure fair comparison and the corresponding model was trained. A publically available data set of 2,034 CT volumes and 130,511 images from *CC-CCII* was used for verification of models' performance for unbiased independent testing.

Data augmentation, pre-processing, and standardization: The CT images were segmented using the U-Net segmentation network, as defined in the protocol. An arbitrary rectangular area (element size was selected randomly in [3/4, 4/3]) was cropped, sampled randomly in [90%, 100%], transformed into a dimension size of 224 × 224, and inverted the input volumes with a probability of 0.5. This sequence of CT

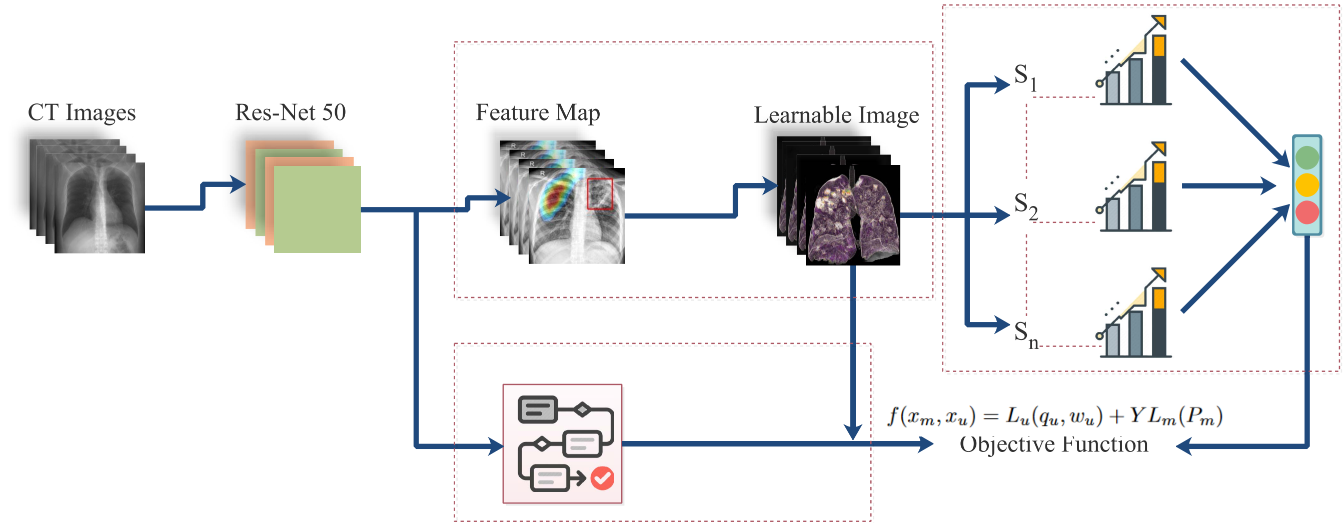

FIGURE 9. **Proposed EXAI classification model with XDM module.**

volumes consisting of consecutive CT image slices serves as the input data.

The proposed EXAI architecture is divided into two phases, viz. classification and image segmentation. The details are presented as follows.

A. CLASSIFICATION MODEL

A CT stack consisting of a dataset of COVID-19 patients and other lung infection images is considered. As data is collected from different hospitals $H = \{H_1, H_2, H_3, \ldots, H_n\}$, the scanning parameters and visual parameters for CT machines are different, thereby degrading the classification accuracy. Moreover, the images are primarily non-labelled; they only consist of patient annotations and labelling. In such cases, weakly supervised models are not a correct fit [110]. Therefore, an explainable diagnosis module (XDM) is added as follows.

B. XDM MODEL

FIGURE 9 shows the proposed EXAI diagnosis model (XDM) for providing the explanation. A class activation mapping (CAM) approach [111] is considered, which highlights the infected section, and improves the visual interpretability. C generates the local feature and maps $\{M_1, M_2, \ldots, M_k\}$ for CT images with corresponding weights $\{W_1, W_2, \ldots, W_k\}$ for CT images. The weighted sum is fed to a backbone ResNet Model. A particular cross-section of the k^{th} map $M_k \in R^{A' \times B'}$ is considered, where $A' \times B'$ is the shape, with $w^{fc} \in R^{K \times C}$, K represents map count. The score S_c is computed by Eq. (1) for a particular class c is given by following expression [87]

$$S_c = \sum_{i=1}^{k} w^{fc}_{k,c}(1/A'B' \sum_{i=1}^{A'} \sum_{j=1}^{B'} F^k_{i,j}) \tag{1}$$

For the S_c score for c class, we define the activation map M_c^{fc} for the k^{th} feature map for shape $A' \times B'$ is presented in Eq. (2) as follows [87]

$$(M_c^{fc})(i,j) = \sum_{i=1}^{k} w^{fc}_{k,c} F^k_{i,j} \tag{2}$$

In the generation of CAM, the network is densely trained till the last fully connected layer. This approach increases the computational overhead. To address this, an XDM approach is considered, where dense layer is replaced by $|x|$ convolution (L_{conv}) layer with weights (W^{conv}), with some form of (W^{fc}). The XDM approach for class c for Eq. (1) is reformulated into Eq. (3) as follows.

$$\begin{aligned} X(S_c) &= 1/A'B' \sum_{i=1}^{A'} \sum_{j=1}^{B'} (\sum_{k=1}^{k} W^{conv}_{k,c} F^k_{i,j}) \\ &= 1/A'B' \sum_{i=1}^{A} \sum_{j=1}^{B} (A(L_c^{conv}))_{i,j} \end{aligned} \tag{3}$$

where $A(L_c^{conv})$ is the activation map for C^{th} class, which can be trained adaptively. $A(L_c^{conv})$ can accurately predict the CT images infection with successive iterations.

To address the lesson severity from the CT volume stack, the details of the image slice and integration module are addressed below.

C. SLICE MODULE

Some CT images are classified with large lesions, and positive cases only capture a small portion of the CT stack. To address the same, a slice module is presented. The slice module considers a joint distribution $J(D)$ of image slices and the mode of infection probability. A multiple instance learning (MIL) framework [112] and a sample set are considered as a bag of instances. The positive bag B^+ must have at least one positive instance, and the negative bag B^- might consist of more than one instance. The bags are labelled as $l_1, l_2, l_3 \ldots, l_n$ for a given patient E_p. A total of n CT slices as $S_1, S_2, S_3, \ldots, S_n$ is considered which is divided into $|S|$ disjoint sections $S_{CT\,n} = \{C_i\}_{i=1}^{|S|}$, and can be expressed using Eq. (4) as follows [87],

$$|S| = max(1, [n/S_l]) \tag{4}$$

here S_l is the length of one section on a designated slice of P. The joint probability defined in [113] can be expressed using Eq. (5) in the following manner

$$\begin{aligned} P(C|P) &= P(C|\{C_i\}_{i=1}^{|S|}) \\ &= \frac{1}{1 + \Pi_{i=1}^{|S|}(1/P(c|P_i - 1)} \end{aligned} \tag{5}$$

here $P(c|P_i)$ is the probability of section i in C. The k-move probability of all the classes to complete section probability can be expressed using Eq. (6) as follows [87]

$$P(c|P_i) = \sigma(1/kmaxC^{(j)} \in M \sum_{i=1}^{k} S_c^{(j)} \tag{6}$$

where $M \subset S_i$, $|m| = k$, and S_c^j is the j^{th} top class score for the i^{th} sector and $\sigma()$ is the sigmoid function. The patients' annotations $A(P)$ are sent to classification loss function using Eq.(7) [87] as follows.

$$L_{class} = -\sum_{c=0}^{1}[y_c log(P)(C|P) + (1 - Y_c)log(1 - P(C|P))] \tag{7}$$

To improve prediction accuracy of the image, a noise connection strategy is applied based on noise distribution and posterior distribution $p(y_i/I)$ is completed over noise distribution $N(Z_c|I)$ using marginal probabilities expressed in Eq. (8) and Eq. (9) presented as follows [87].

$$P(L_c = i/y_c = j_i, I) \tag{8}$$

$$P(Z_c = i/I) = \sum_j p(Z_c = 1/y_c = j, I](p(y_c = j/I)) \tag{9}$$

The classification loss due to impact of noise is summarized as $L_{noisy} = -1/N \sum_{i_1}^{N} \sum c = 0^1 [y_c^n logP(Z_c = 1/I_n) + (1 - y_n^n)logP(Z_c = 0|I_n)]$. The total loss of COVID-19 EXAI classification can be expressed using Eq. (10) as follows.

$$L_{(t)} = L_{class} + \lambda L_{noisy} \tag{10}$$

where λ is the tunable hyperparameter to balance distribution between L_{class} and L_{noisy}.

D. SEGMENTATION BY EXAI

The MRI multimodal data acquired for hydrocephalus patients is considered. Previous studies have suggested images with slice thickness of < 3mm. With small slice thickness, more images can be equipped at the cost of degraded image acquisition accuracy and large imbalance. As a result, thick slices are generally preferred. To address the limitations, an EXAI segmentation model for both the thick and thin slice image is presented, where annotations on only thick images are considered. FIGURE 10 shows the segmentation architecture for EXAI model.

The image dataset $I = \{I_{thick}, I_{thin}\}$ consists of thick and thin slices where thick slice can be represented as $s(I\_thick) = (x_u, y_u)|x_u \in R^{A \times B \times 3}, y_u \in R^{A \times B}$ and thin slices $s(I\_thin) = \{x_m | x_m \in R^{A \times B \times 3}\}$ is considered. We considered the unlabelled $S(I_{thin})$ to minimize the model performance gap with post-hoc EXAI scheme [114]. A segmentation network based on the V-net [115] is considered, where encoder is replaced by ResNet-50 [116] performed on ImageNet dataset [117]. The decoder network considered for subpixel convolution for segmentation is expressed using Eq. (11) as follows.

$$P^L = s(W_L.M^{L-1} + \propto_L) \tag{11}$$

where $s()$ is a tensor $A \times B \times C \times r^2$ transformed into $rA \times rB \times C$, where r is a scaling factor. M^{L-1} and M^L are the input and output feature maps respectively, W_L and $\propto_L$ are parameters for sub-pixel $s()$ operator at the L^{th} layer. A multimodal training process is conducted that jointly optimize both the I_{thick} and I_{thin}. The objective function is constructed using Eq. (12) as follows.

$$f(x_m, x_u) = L_u(q_u, w_u) + YL_m(P_m) \tag{12}$$

Here, Y is the hyper-parameter, q_u and w_u are the prediction and segmentation maps, and L_u is the cross entropy function. L_u is defined using Eq. (13) as follows.

$$L_u(q_u, w_u) = -1/ABC \sum_{c=1}^{A} B \sum_{c=1}^{C} y_u^{n,c} log P_u^{n,c} \tag{13}$$

For slice, L_m portion is selected for the decision boundary features to achieve alignment. The distance can be minimized based on prediction distribution d_p and uniform distribution $u = 1/c$ that eliminates prediction uncertainty. The slice objective f_n defined through f-divergence is expressed using Eq. (14) as follows.

$$L_m(P_m) = 1/ABC \sum_{n=1}^{A} B \sum_{c=1}^{C} D_f(P_m^{n,c}||u) \tag{14}$$

which yields to Eq. (15) presented as follows [87]

$$L_m(P_m) = 1/ABC \sum_{n=1}^{A} B \sum_{c=1}^{C} f(CP_m^{n,c})) \tag{15}$$

where $f()$ is a divergence function. Small gradients are arranged to a smaller sample set to address the imbalance. x^2 divergence with $f() = a^2 - 1$ is normally preferred. Finally, the encoder module can interpret feature space visualization, and the decoder decomposes them into a 2D space via principle component analysis. The whole space with multi-layer perception then fits the decomposed samples.

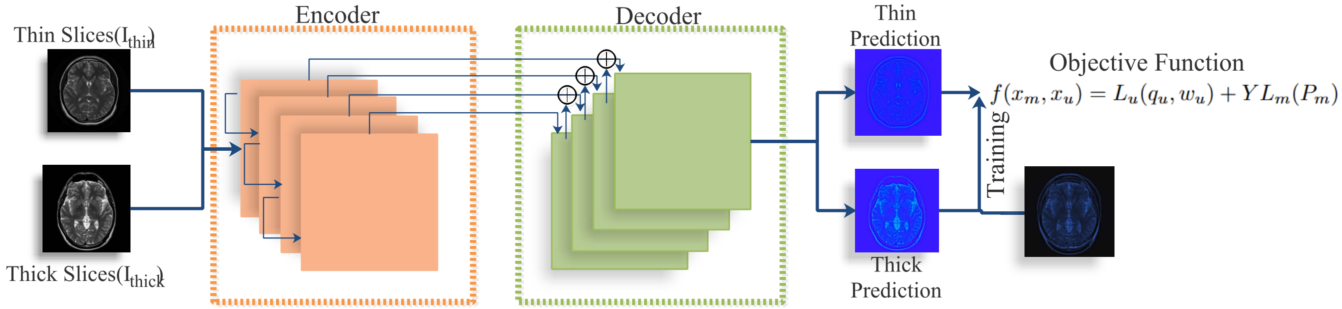

FIGURE 10. Segmentation of EXAI model.

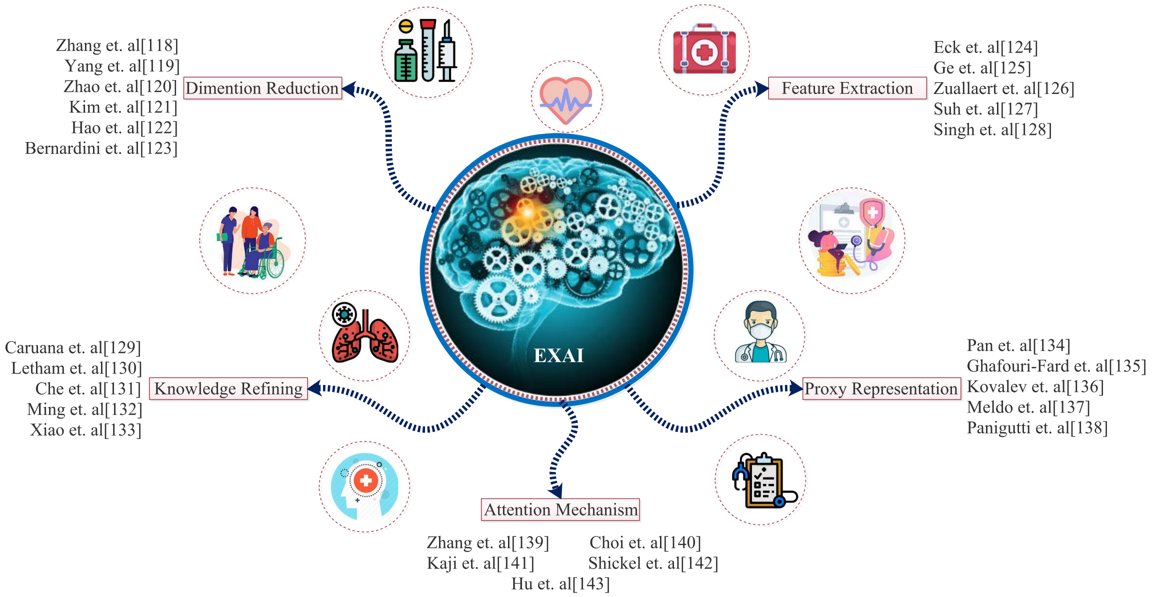

FIGURE 11. Solution taxonomy of EXAI in healthcare applications.

VII. SOLUTION TAXONOMY

The development of EXAI is the need of social and scientific importance. The technology is mimicking to solve a complex problem, and AI now achieves things that were earlier accomplished by human approach and reasoning. The sophisticated explanation limits the expansion of AI, and the responsibility of the decision is unclear because it is based on prediction and classification. This makes it hard to agree with highly consequential system decisions. EXAI is a powerful descriptive tool that provides insights to achieve higher accuracy with traditional linear models.

This subsection discusses the solution taxonomy of EXAI in healthcare applications to incorporate various features for diagnosis and surgery. FIGURE 11 depicts the same.

A. DIMENSION REDUCTION

Classification in AI/ML depends on the number of variables called features. It is hard to visualize and increase the complexity of the training dataset. Sometimes, many of the features are correlated and redundant and thus require feature reduction to improve visualization and explainability of the outcomes. There are many approaches for dimension reduction, such as independent component and principal component analysis. In the same direction in the healthcare domain, Zhang *et al.* [118] proposed a k-nearest multi-label feature selection method to predict the side effect of drugs using an optimal feature from the input dataset. Yang *et al.* [119] proposed a non-linear dimensionality reduction approach to investigate the performance of the classification problem of magnetic resonance spectroscopy brain-tumour against the traditional methods. The author used Laplacian Eigen maps followed by a k-means clustering technique to assess tumour grades. Zhao and Bolouri [120] proposed an object-oriented regression model to identify high dimensional omic data in clinical studies and assess their relationship with the prognostic outcome. This will reduce the penalty of using high dimensionality data and retain the interpretability of the stage one cancer patient data. Kim *et al.* [121] proposed an architecture based on DL for the prediction of the human genome that provides advantages like increased accuracy and less prediction time. The author made the scheme interpretable by computing each feature's importance and considering only optimal features for classification. Hao *et al.* [122] proposed a pathway associated with deep neural network-based architecture that predicts complex biological processes in prognoses. It is a DL-based multilayer architecture to predict clinical outcomes. This technique is applied for the long-term prediction of brain cancer that earlier showed poor prognostic performance. Author in [123] proposed a detailed study on discovering Type-2 diabetes using a sparse balance support vector machine. The author computed the experimental results and a clinical use-case scenario with better performance than other traditional approaches.

B. FEATURE EXTRACTION

Feature extraction techniques reduce the number of feature from the dataset by creating more feature out of it that combines the variables and still describe the same data. For better interpretability of AI models, feature extraction and its correlation among different features are important. Eck *et al.* [124] proposed a microbiota based diagnostics for explaining classifiers decision. Microbiota data is challenging, and high dimensional sparse comprises high interpersonal variation. The author used this method to explain the classifier's decision on skin microbiota and inflammatory bowel disease data. This explanation of microbiota diagnostic can increase the confidence in the decision support system. Ge *et al.* [125] proposed a logistic and RNN model that explains the ICU mortality prediction. The ML-based model achieves better accuracy in such a scenario but does not provide explicit interpretability. The model uses sequential features with multiple values and non-sequential features that contribute to the outcome. Zuallaert *et al.* [126] proposed a predictive DL approach that provides better performance as compared to the current traditional approaches. The authors used a CNN that outperformed other traditional approaches in accuracy and efficiency but limited explainability. To provide interpretability, the author used techniques to visualize relevant biological information that recovers the features that are important to form a new genetic combination. Suh *et al.* [127] proposed an EXAI based risk calculator for prostate cancer. The author aims to validate the risk and explain the clinical significance using the AI model. Approximately 2800 patient data are used among 948 as a test set, and a gradient boosting algorithm is applied with the tuning of hyper-parameter and feature selection in the development. It utilizes Shapley value to determine feature importance. Singh *et al.* [128] proposed a DL-based framework for interpreting the decisions of retinal image classification to improve its acceptability in medical applications. The proposed framework is used for interpreting the classification of optical coherence tomography images. The result shows successful attribution of the specific pathological region from the given images. The feature extraction explains the important characteristics of the feature with their relative importance in clinical interpretation.

C. KNOWLEDGE REFINING

Knowledge refining is transferring knowledge from a large ML model into a small model that can be applied in real-world situations without significant loss in performance. Some model consists of approximately 500-600 Gigabyte of training data with 100-200 million parameters; such models helps to improve the performance of the system but cannot be deployed on edge device just because of their complexity and size. Knowledge refining and distillation perform more commonly on the neural network with complex architecture and several layers. With the advent of DL in speech, image and language processing is the area where knowledge refining and distillation techniques can be realized. Model compression and tree regularization are some of the explainable model compression approaches. Caruana *et al.* [129] presented a case study on intelligent models for predicting pneumonia risk using pairwise interaction of additive model in 30-day hospital readmission and concluded that the same method also generated the same efficiency as with a large number of parameters. Authors in [130] have created a predictive model that is accurate and interpretable for human experts. The model consists of a readily interpretable if-else statement. The author claimed that the Bayesian rule list has better accuracy than the current traditional approaches on an experimental basis. The model better predicts stroke in patients with atrial fibrillation with better accuracy and interpretability. Che *et al.* [131] proposed an interpretable mimic learning that uses gradient trees to learn an interpretable model with the same prediction accuracy. The model is used to predict acute lung injury, providing wide interpretability to clinical decision-making. Authors in [132] proposed a visualization technique that is used to inspect and understand the ML model; the method visualizes the reasoning of the model with a systematic explanation of data used to train the model. The author focused on rule-based knowledge-based representation from input and output variables. A rule-based matrix approach is used to explore, understand and validate the model output. Xiao *et al.* [133] explored the DL models that distil complex relationships and accurate prediction. A DL model is utilized to predict the patient's readmission risk in hospitals by the local and global context of electronic health records of 5393 patients. EXAI approaches become stable if they rely on knowledge refining and distillation.

D. PROXY REPRESENTATION

The interpretable surrogate model is trained to know the prediction accuracy of the black-box model. These models are applicable in situations where the outcome of interest is difficult, expensive and time-consuming. The purpose of the surrogate model is to provide interpretability to the outcome while at the same time keeping the accuracy of the underlying model. Disease prediction using a local interpretable model explains any classifier by training the local model instead of the global surrogate model to explain the individual prediction. Pan *et al.* [134] proposed detailed study on a prediction model for central precocious puberty in girl. The central precocious puberty affects girls' mental and physical development in childhood. The traditional stimulation test makes the patient more uncomfortable as it requires multiple blood samples. The study was carried out in a sample size of 1757 girls, and follicle-stimulating and basal serum luteinizing hormones are the main factors to-word the prediction probability. Authors in [135] proposed an artificial neural network-based method for differentiating autism spectrum disorder. The author extracted genomic data from approximately 500 patients and 450 healthy individuals for the study. Kovalev *et al.* [136] proposed a ML survival model to explain the outcome based on combination of input-output. It extends local interpretable model agnostic explanations that approximate any learning model with a local interpretable model to explain an individual's predictions. The main idea is to include the cox proportional hazard model to test it in its local area, and the advantage of using it impacts the prediction. Authors in [137] proposed an algorithm for lung cancer computer-aided design explanation; the important benefit of the algorithm is that it uses natural language for the explanation. The algorithm is comprised of two-part; first, its selection of important features from segmented lung objects. The second is connecting important features and transforming them into an explainable natural language. Panigutti *et al.* [138] proposed a model that takes a patient's clinical history as an input to predict the next visit to the doctor; such technique is also called an agnostic explainability technique. The quality of the explanation is improved by using temporal dimensional data and domain knowledge. Proxy or surrogate representation is widely used in EXAI; if the surrogate model is too complex, the clinical explanation becomes difficult.

E. ATTENTION MECHANISM

The attention mechanism was introduced to pay specific attention to the relevant part of the system that contains useful information. It enhances the efficiency of the encoder-decoder model, where the decoder utilizes the most important part of the input sequence by combining encoded input and letters. Encoder-decoder used it for relation extraction in natural language processing domains where specific words hold significant importance. Zhang *et al.* [139] proposed a framework to learn personalized representation of health record. This model improves the prediction performance and clearly understands disease correlation. This model helps us predict the risk of hospitalization based on electronic health records, and it provides better accuracy with baseline approaches. Authors in [140] proposed a reverse time attention model based on neural attention that detects and analyzes the past visits of patients and identifies the valuable significant clinical variable. It gave more attention to past visits, where the model was tested on around 260k patients' data over eight years with approximate 18 million visits and provided better accuracy than the current solutions. Kaji *et al.* [141] proposed a DL-based model to detect the presence of the microorganism in the blood; such models are generally called attention-based models. The model is trained on approximately 56k patients' data to detect three different blood microorganisms with three different datasets. Authors in [142] proposed an acuity score framework that assesses the severity of the patient in ICU using a temporal and interpretable DL model. The proposed framework is compared with the baseline approach on the same model input and found to have better accuracy than others. These models are important for identifying life-saving interventions during the ICU stage. The model was trained on two standard datasets provided by the academic medical centre of Florida and Israel. Hu *et al.* [143] proposed an DL scheme to provide prediction of type-1 virus such as immuno-deficiency and an explanation of the predicted site. This model outperforms all the baseline approaches by automatically learning genomic context from the primary DNA sequence. The general approach enhances the explainability of the model through a focus on some of the portions from the input sequence which affect the prediction outcome.

VIII. CHALLENGES AND FUTURE RESEARCH DIRECTIONS

This section presents the open challenges and future perspectives in integrating EXAI in AI-enabled applications to address fairness, transparency, and accountability to achieve human-centric AI. The research question RQ 4 is handled by this section as it incorporates the challenges of EXAI in addressing the gaps with AI to counter future operational deployments.

A. INTEROPERABILITY AND VISUALIZATION

Apart from other application verticals, there is an inadequacy in interoperability and visualization, such as human attention's capability to understand the explanation maps and measures to verify the correctness and completeness of the explanation maps generated by the EXAI system. As a result, there is a need to implement better explanations for better visualization and interoperability of the explanation map in mission-critical applications. In such cases, EXAI models like SHAP and LIME can explain the prediction variables(input) by computing feature contribution towards the output.

B. HUMAN MACHINE INTERACTION

The design, development, and deployment of responsible human-centred AI are essential. The interaction between humans and machine is necessary for comprehensive explanations of models to the user. Adaptive explainability modelling provides context-aware explanations based on various human profiles. The combination of social science, human behaviour, and human-machine interaction empirical studies impact EXAI research. The design of a human-enabled feedback mechanism through derived machine explanations (visual or logical) can enhance human-machine interaction by incorporating transparency, ethics, judgment, and social norms. Better human-machine interaction creates an explanation map for the right user to understand the results, such as in a clinical environment. This requires further research to optimize research methods to identify the correct problem.

C. DECISION SUPPORT SYSTEMS

Decision support systems incorporate healthcare tools to enhance decision-making and provide knowledge and person-specific information to all the stakeholders directly involved in a health information system. Developing an ML-based clinical decision support system should incorporate collaborative inputs from all the stakeholders from different domains as it impacts the overall consequence. However, various stakeholders from other fields face the issue of unstructured medical data. Exploring data mining and explainability in the model allows privacy in the exchanged data and enables the extraction of meaningful information to capture critical medical information.

D. SECURITY

Security is one of the major concerns in the explainability of AI, where some techniques try to generate false explainable maps through randomized input and mitigate sensitive and private information. EXAI can be used to identify additional risks associated with healthcare, making the system aware of the impact of risk in healthcare. To safeguard the private data in healthcare, models can be developed that memorize part of the training data and prevents privacy attacks on identified sensitive data. Incorporating improved techniques ensures data availability and heterogeneity to increase the overall system performance and control process design.

E. INTEGRATION WITH AI

AI-based systems and algorithms are data and power-hungry, which poses continuous requirements to computing systems (such as cores and graphical processing units) for efficient working. ML and DL-based modelling provide unknown and unpredictable outputs. To achieve human-level accuracy, AI methods require hyper-parameter optimization, fine-tuning, robust computing capability, enormous datasets, and continuous data training. These data are generated from millions of user IoT devices and are vulnerable to cyber-attacks. The data generated from AI-based systems is also biased; thus, EXAI provides obvious explainability to enable human-level intelligence.

F. INTERPRETABILITY VS. PERFORMANCE TRADE-OFF

Model complexity is defined in terms of the data distribution to infer meaningful accuracy. A trade-off between interpretability and performance becomes then present. Explainability techniques can minimize the trade-off between model complexity and its interpretability. Developing fully transparent models can provide entirely local & global explanations and enhance model accuracy and representations. These methods should support consistency with human reasoning and understanding.

G. MODEL AND EXPLANATIONS UNCERTAINTIES

Communication uncertainty provides information about the model and explanations uncertainties to the user. Explanation methods should provide the definite quantification of uncertainty implied by the explanations. This varies during the life cycle of the ML model. Therefore, a rigorous study is required to quantify and derive the reliability of uncertainties in the model life's phase and explain methods to users for easy response.

H. DATA SHARING

Data sharing refers to the data collection and preparation phase, the raw data available for usage by different entities. Stakeholders can share the data either directly or at the central-server level. Explanations at this stage are not required to secure the privacy of the user data as the models are in the developed setting. FL helps in sharing raw data by employing distributed ML techniques. However, with an increase in the number of collaborators/clients, the model is susceptible to membership inference attacks, collusion attacks, etc. Integration of adaptive learning techniques and explainability can provide a more robust sharing mechanism at the data and model level.

I. FRAMEWORK TO FORMALIZE EXPLAINABLE ARTIFICIAL INTELLIGENCE

The research in [144] fuses different domain knowledge in black-box methods for better explainability. However, challenges such as formalism in explanation, quantification, and comprehensibility are not studied in detail. Developing a generic framework to leverage knowledge from multiple domains is essential for critical mission applications to provide greater confidence and reliability.

J. LEVERAGING ACCOUNTABILITY AND TRANSPARENCY IN AI-BASED MODELS

Responsible AI removes biases generated from data, algorithm modelling, and biases to enhance the model's fairness, accountability, and transparency. Developing an analyzable AI system is required to account for any failure in the decision and provide more comprehensible explanations for mission-critical applications. Adoption of EXAI enables trust, understating, and weeding out potential risks. Therefore, an EXAI model should be developed to adhere to accountability, fairness and transparency conditions.

K. DATA QUALITY

The outcome of the AI system depends on the quality of data used in the models. Any bias, uncertainty, or incompleteness results in quality degradation. This also affects the explainability and decision capability of the AI system. The developed system should be able to transmit the quality of the data, produce disclaimers through derived explanations, and communicate to users about the risk involved in specific explanations. By adopting measurement and quality metrics like completeness, accuracy, and consistency, EXAI systems can ensure end-to-end transmission of quality information across various demographics and users without degradation of the performance.

IX. *ExoCOVID*: A CASE-STUDY OF ECG MONITORING HEALTHCARE ECOSYSTEM WITH FEDERATED TRANSFER LEARNING AND EXAI

This section proposes an end-to-end framework for electrocardiogram (ECG) based healthcare in a federated setup. A deep CNN architecture is presented, where data is locally trained at hospital setups using FL, which addresses the issues of privacy and data availability. The authors proposed a classifier for arrhythmia disease on the benchmark MIT-BIH Arrhythmia database [145]. To improve the interpretability, an EXAI module is presented on top of the classification, improving the prediction results. This section also augments addressing research question RQ 5 by offering EXAI-enabled and AI-FL-assisted applications for heart ECG monitoring systems.

A. CASE-STUDY CONTRIBUTIONS

Following are the contributions of the proposed study.

1) A CNN-autoencoder design in a federated healthcare environment is proposed to denoise raw ECG signals from patients.
2) A federated transfer learning (FTL) approach is designed to use the encoding section of the autoencoder, and a CNN-based classifier model is built. The classifier classifies the ECG data viz. $\{N, S, V, F, Q\}$, where N denotes the non-ecotic (normal) beats, S denotes the supra-ventricular ectopic beats, V denotes the ventricular ectopic beats, F denotes the fusion beats, and Q denotes the unknown beats.
3) The CNN-classifier is integrated with an EXAI module that provides interpretability on the decision process of the classifier. The EXAI module is generic, and its functionality can be integrated into any classifier. Moreover, as the ecosystem is federated, the EXAI model interpretability improves with successive iterations of the global server model, and learning losses are minimized.
4) The MIT-BIH data is initially unsampled to create more data samples, and then $10 - 30\%$ random noise is inserted to improve privacy. The framework predicts a classification accuracy of $\approx 94\%$ on noisy data and 98% on clean data.
5) Experimental validation is performed through proposed evaluation metrics like Precision, Recall, and F1-score, with a comparative discussion of accuracy against similar schemes. EXAI outputs of Grad-CAM for ECG signals are modelled to find local interpretations of the ECG dataset.

B. MIT-BIH DATASET

The MIT-BIH dataset [145] consists of ECG recordings from different subjects from 1975 to 1979. A total of 109, 144 samples are recorded in the dataset, where 23 random recordings are chosen from a set of 24-hour 4000 ambulatory ECG samples over a diverse patient population. In the total population, $\approx$ 60% are inpatients (that are admitted to BIH hospital), and 40% are outpatients (not admitted to BIH hospital). Around 25 recordings are selected from significant arrhythmia classes as the less common sample in the random sample set. The experiment uses ECG at 125 kHz input frequency, and $\approx$ 10-30% is induced in the original dataset to preserve privacy. TABLE 7 presents the heartbeats (ECG signals classes) which are present in the dataset.

TABLE 7. Five classes of ECG signals.

Class description	Symbols associated
Non-ecotic beats (normal beats)	N
Supraventricular ectopic beats	S
Ventricular ectopic beats	V
Fusion beats	F
Unknown beats	Q

TABLE 8. Extracted beats from the MIT-BIH ECG signal dataset [146].

Type	N	S	V	F	Q
Beats	90502	2777	7226	802	8031

1) PREPROCESSING DATASET

MIT-BIH dataset is imbalanced in terms of the number of samples, which might result in class overfitting. The ECG signals also have a noise component from the nearby power supply, body movements, and other factors. Due to this, feature extraction is a difficult task, which limits the extraction of ECG heartbeats from signals. In this case, the ECG signals are normalized using Eq. (16) as follows.

$$S_n = \frac{|S - S_{min}|}{|S_{max} - S_{min}|} \tag{16}$$

where S is the value of the considered sample, and S_{min} and S_{max} are the minimum and maximum among all considered samples, and S_{norm} denotes the normalized sample. Based on Eq. (16), T-waves of the ECG signals are matched with R-peak, and the signals are classified into ECG beats. TABLE 8 denotes the extracted beats (classified according to types) and the frequency of the beats captured from the ECG dataset [146]. Once we obtain the extracted beats, noise-removal techniques based on discrete wavelet transform (DWT) are applied [147]. It resulted in a reduction of ECG heartbeats from 280 to 141 samples.

2) DATASET DISTRIBUTION (ORIGINAL AND RELABANCED)

As discussed above, the dataset used for the analysis has unbalanced classes. This might result in the overfitting of the model. To eradicate the issue, the data is upsampled. TABLE 9 shows the original and resultant upsampled data distribution. In the real-world scenario, the ECG data collected is noisy; thus, to map it with practical information, about 10-30 per cent of noise has been added to the dataset.

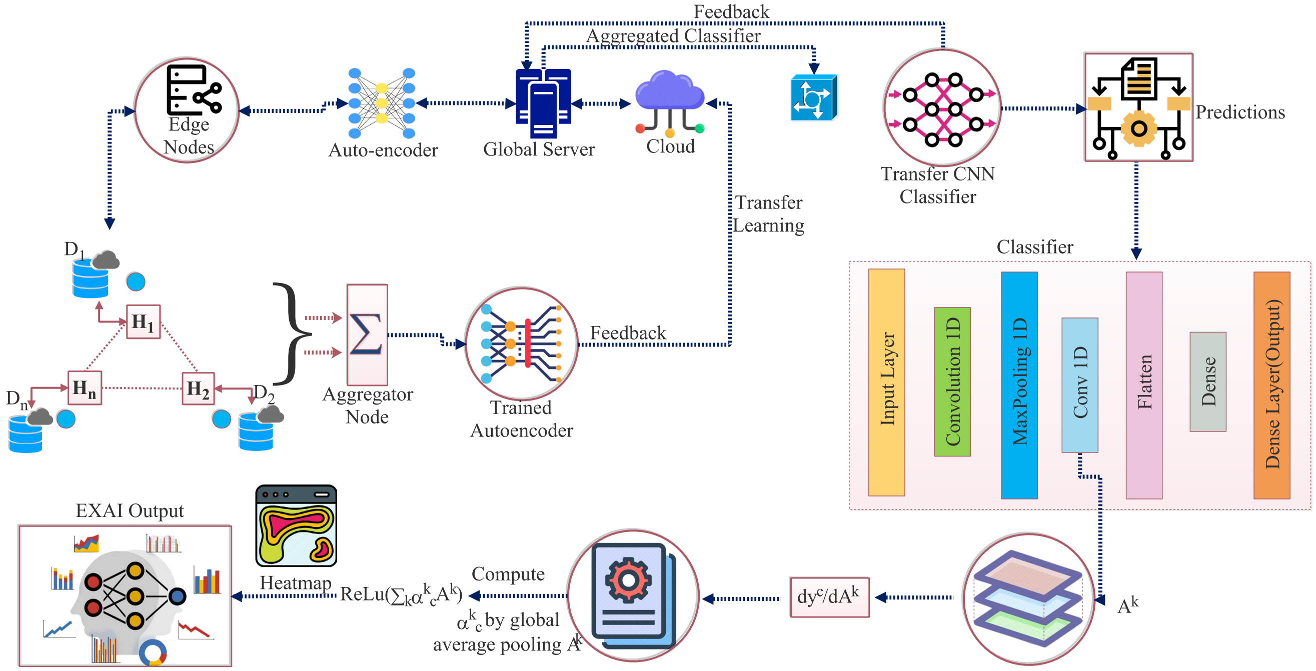

FIGURE 12. ***ExoCOVID*: An EXAI and FTL-based ECG monitoring scheme.**

TABLE 9. Original and upsampled dataset distribution.

	Non ecotic beats (Normal beat)	Supraventricular ectopic beats	Ventricular extopic beats	Fusion beats	Unknown beats
Original	82.8%	7.3%	6.6%	2.5%	0.7%
Upsampled version	20.1%	20.1%	20%	19.9%	19.9%

The model trains the modified dataset (with noise addition) to predict the results.

C. THE PROPOSED FEDERATED HEALTHCARE FRAMEWORK

This section presents the schematics of the proposed framework by Raza *et al.* [113]. FIGURE 12 presents the details. A federated healthcare setup is considered that consists of n local edge (hospitals) nodes $E = \{H_1, E_2, \ldots, H_n\}$. Any i^{th} hospital H_i has the corresponding data D_i, $i \in [1, n]$. TABLE 7 shows the classification of the data into five classes as mentioned earlier. A global server model G_s runs an autoencoder AE, specifically hyper tuned with the predefined parameters. All H_i are required to send request to G_s to download AE to start training the model on local data D_i. Once the AE model is trained on local data, AE weights $\{w_1, w_2, \ldots, w_n\}$ are transferred back to G_s. Upon receiving the weights, G_s waits for locally trained AE models from H_i. Based on the local weights and trained model, G_s computes the aggregation function using Eq. (17) as follows.

$$F(w) = \sum_{i=1}^{n} \frac{n_k}{n_t} w^k_{r+1} \tag{17}$$

where $F_k(w) = \frac{1}{n_k} \sum_{i \in P_k} f_i(w)$. Based on $F(w)$, G_s constructs the classifier C and sets the weight of three convolutional layers for C. As the convolutional layers are untrained, G_s sends C to all H_i, which trains them on local D_i, and the weights are sent back to G_s. The process is iterated until the accuracy of the global model improves, and the loss function is minimized.

D. CNN-BASED AE MODEL

As indicated, local edge nodes are considered local silos, where a single hospital unit is considered operational. All the data is combined to form $D_{total} = \sum_{i=1}^{n} D_i$, and is trained by the federated setup as discussed in section IX-C. The federated setup is adopted so that local hospital data is not shared and G_s is trained. The accuracy for the prediction model trained on aggregate data D is denoted by A_t while A_l denotes the prediction accuracy of the model trained on local D_i. A series of ECG raw input signals $S = \{S_1, S_2, \ldots, S_k\}$ is considered to be passed through the autoencoder AE. The local AE model consists of three layers $L_{AE} = \{I_{AE}, H^{12}_{AE}, O_{AE}\}$, where I_{AE} is the input AE layer, H^{12}_{AE} are 12 hidden layers, and O_{AE} is the output layer. H^{12}_{AE} is further classified as $\{C^6_{AE}, MP^3_{AE}, U^3_{AE}\}$, where C^6_{AE} denotes the six convolutional layers, MP^3_{AE} are the three max-pooling layers, and U^3_{AE} are the three unsampling layers. A varying learning rate η is considered to keep the reconstruction loss at a minimum. η is defined using Eq. 18.

$$\eta = \begin{cases} 0.01 & \text{if } e \leq 40 \\ \eta & \textit{otherwise} \end{cases} \tag{18}$$

Next, we present the description of the CNN-classifier model.

E. CNN CLASSIFIER

A classifier model $M_c = \{L^4_{conv}, L^2_{FC}, L^3_{MP}, L^1_{SM}\}$, where L^4_{conv} represents four convolutional layers, L^3_{FC} represents three maxpooling layers, L^2_{FC} represents two fully connected layers, and L^1_{SM} represents one softmax layer for classification. The AE part is transferred to the classifier model, which allows three L_{conv} layers to be not trained (the layers are kept fixed, and parameters are not updated during propagation). It leverages the local nodes N_l with trained parameters. The last two L^2_{conv}, and the L^2_{FC} layers are at a higher level, and parametric updates are done. The softmax function is defined using Eq. (19) as follows.

$$y_i = \frac{e^{L_c}}{\sum_{i=1}^{C} e^{L_c}} \tag{19}$$

where L_c is the learning probability of class c, and C is the total number of classes.

F. EXAI INTEGRATION

The local regions setup is visualized using the gradient-weight class activation mapping (Grad-CAM) approach for normal and myocardial infection. Knowing whether the heart tissues are getting proper oxygen flow is important, which leads the practitioners to know about artery blockage. For the same, gradient of score value (denoted as y^c) against the feature activation A^k (for k^{th} kernel) is presented. The gradient G_c for c class is presented as in Eq. (20) as follows.

$$G_c = \frac{\partial y^c}{\partial A^k} \tag{20}$$

The value of G_c specifically depends on the ECG signal. Once the derivative is constructed, a global average pooling is applied, and the weighted linear combination of A^k and α^k_c is computed as presented in Eq. (21) as follows.

$$Grad_CAM(c) \; = ReLU(\sum_k \alpha^c_k A^k) \tag{21}$$

A rectified linear unit (ReLU) function is applied to consider positive values only.

G. THE FEDERATED TASK LEARNING

The learning process continues with the newly emerging data. The edges personalize the ECG signal classifier by keeping all the layers of the static convolution network in the final updated classifier. Then the training phase starts from the dense layers for personalization. This is due to the convolution layers' goal of extracting low-level features for activity detection and the highly coupled layers. At a higher level, it concentrates on learning particular aspects of the task. The global aggregator server (GAS) creates an AE with the defined hyperparameters. Once the AE is defined and created, the GAS waits for the signals sent by AE. Also, when the client asks for the GAS, the GAS sends AE to the client; likewise, the rest of the communications take place. The client trains the same on receiving the AE in its local database. After completion of the training, the client will send the AE to the GAS. After receiving the consensus, which occurs and completes when the GAS receives the desired number of rounds, it aggregates the weights and can be depicted using Eq. (22) as follows.

TABLE 10. Implementation details.

Parameters	Values
No. of edges	3
Global server configuration	i-6700HQ CPU 32 GB RAM
Training ratio	80:20
Random Noise added to the Edge1, Edge2 & Edge3 (in percentage)	20:30:10
Batch Size	100
Epochs	150
Learning rate	0.001

$$Agg(w) = \sum_{k=1}^{n} (n_{ks}/n_s) w^{ks}_{r+1} \tag{22}$$

where,

$$Agg_k(w) = 1/n_{ks} \sum_{i \in P_{ks}} agg_i(w)$$

Here the aggregated weights are denoted as $Agg(w)$; the n_s symbolizes data samples of all the clients. The n_{ks} is the k^{th} participants. In every ML problem statement, the $Agg(w)$ is defined for the i^{th} $x, yandw$, this is actually the loss prediction of x_i, y_i and w. For the data partition p_{ks} over the n clients. The p_{ks} is the data point indexes of the ks client. The n denotes the number of partitions for each round, and the global round is denoted as r—the aggregation results in the new CNN-based classifier used for the classification. For transfer learning, the encoder part of the AE is considered, and the weight will be transferred to the AE. After receiving the client, train into its local data and send it to the aggregator using Eq. (22). Clients specify the aggregated weights as new weights for their local data, which may then be utilized for predictions. The EXAI module taps the gradients during forecasts and produces a visual explanation.

H. PERFORMANCE EVALUATION

In this subsection, we discuss the performance evaluation in terms of evaluation metrics, EXAI Grad-CAM results, and the comparative analysis with existing schemes. The details are presented as follows.

1) SIMULATION PARAMETERS

The implemented hardware details contain only three local Raspberry Pi devices, called Edges1, Edge2, and Edge3—which were used to train the autoencoder and classifier locally. These devices are configured through the Pi 3 Model B+ with a 1.4GHz, 64-bit quad-core ArmV8 CPU and 1GB LPDDR2 SDRAM. TABLE 10 indicates the other implementation details of the system configurations and other training parameters.

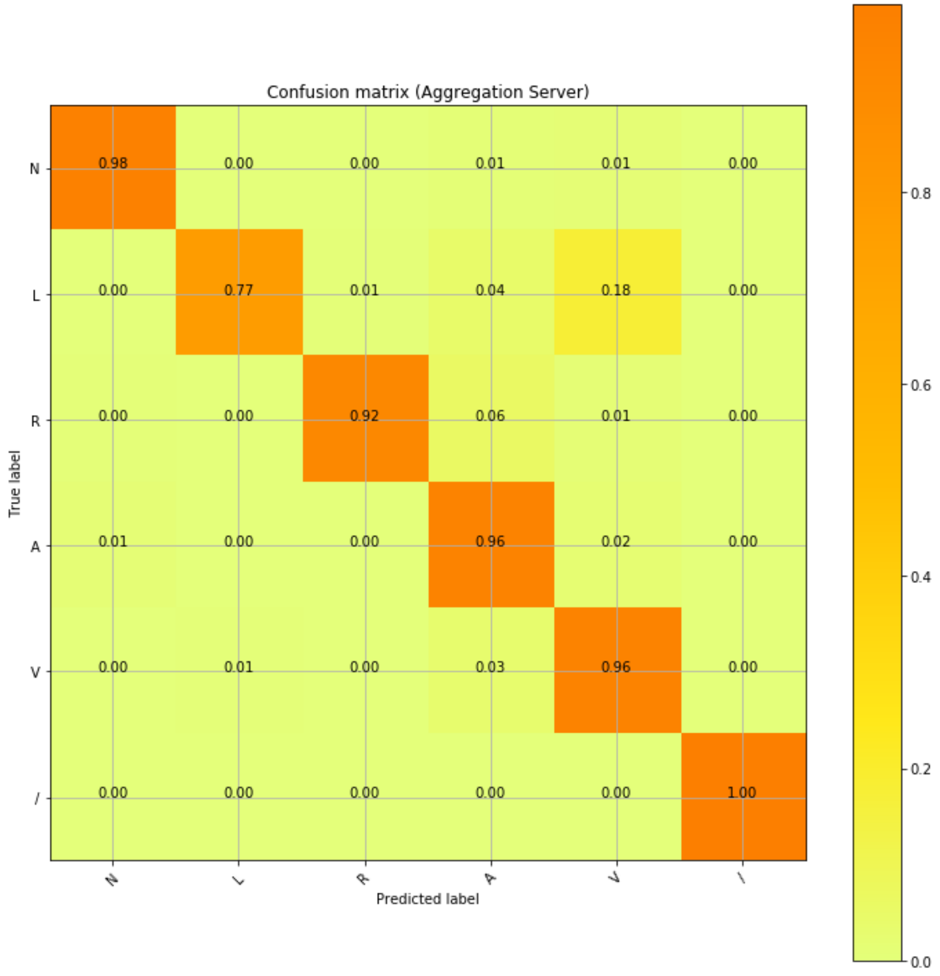

FIGURE 13. **Confusion matrix.**

2) EVALUATION METRICS

- *Precision*: The precision can be defined as
$$T_p/(T_p + F_p). \tag{23}$$
T_p denotes the number of true positives, and F_p denotes the number of false positives. The precision is intuitively the ability of the classifier not to label a negative sample as positive.
- *Recall*: The recall metric can be defined as
$$T_p/(T_p + F_n) \tag{24}$$
were F_n is denoted as False negative.
- *F1-Score*: The F1-score will be calculated using Eq. (25) as
$$\frac{2 * precision * recall}{precision + recall} \tag{25}$$

3) SUFFICIENCY OF THE SELECTED METRICS

In the considered evaluation, precision and recall are two critical model evaluation measures and metrics. While precision refers to the percentage of relevant results, recall refers to the percentage of total relevant results accurately categorized by the algorithm. The harmonic mean of precision and recall is F1. Thus, to predict the accuracy of the underlying model, these three metrics alone are sufficient. In some cases, however a trade-off is required, and a choice must be made between optimizing the precision and recall.

Once the edges and aggregators are formed, the three standard metrics were used to investigate the performance evaluations. These are used to assess the classifier's performance. The precision, recall, and F1-score are calculated to define the classification accuracy of the ECG signals. The classification accuracy is the total number of correct predictions divided by the total number of predictions produced for a dataset. The main reason for this is that the enormous number of examples from the classification model (or classes) will overload the number of samples from the minority class, which leads to the inexperienced models can achieve accuracy scores of 90% or 99%, depending on the severity of the class imbalance. Hence, a solution is needed to solve such situations. The same is solved using precision and recall. The confusion matrix is also presented, which provides additional information into

TABLE 11. Proposed framework performance (noisy version): Edge 1.

	Precision	Recall	F1-score
N	0.890	0.910	0.900
S	0.940	0.890	0.920
V	0.930	0.960	0.940
F	0.950	0.940	0.950
Q	0.990	0.990	0.990
Accuracy	0.940	0.940	0.940
Macro average	0.940	0.940	0.940
Weighted average	0.940	0.940	0.940

TABLE 12. Proposed framework performance (noisy version): Edge 2.

	Precision	Recall	F1-score
N	0.850	0.870	0.860
S	0.910	0.870	0.880
V	0.910	0.940	0.920
F	0.930	0.930	0.930
Q	0.980	0.980	0.980
Accuracy	0.910	0.910	0.910
Macro average	0.910	0.910	0.910
Weighted average	0.920	0.910	0.910

TABLE 13. Proposed framework performance (noisy version): Edge 3.

	Precision	Recall	F1-score
N	0.940	0.980	0.960
S	0.980	0.920	0.950
V	0.950	0.990	0.970
F	0.990	0.940	0.960
Q	0.970	1.00	0.980
Accuracy	0.970	0.970	0.970
Macro average	0.970	0.970	0.970
Weighted average	0.970	0.970	0.970

TABLE 14. Proposed framework performance (noisy version): Edge 4.

	Precision	Recall	F1-score
N	0.900	0.920	0.910
S	0.940	0.890	0.910
V	0.930	0.960	0.940
F	0.950	0.960	0.950
Q	0.990	0.990	0.990
Accuracy	0.940	0.940	0.940
Macro average	0.940	0.940	0.940
Weighted average	0.940	0.940	0.940

not just a predictive model's performance but also which classes are forecasted correctly, which inaccurately, and what types of errors are being made. A two-class classification issue with negative (class 0) and positive (class 1) classes yields the simplest confusion matrix. FIGURE 13 presents the details of the confusion matrix obtained for the GAS, as local edge models (Edge 1, Edge 2, and Edge 3, collaboratively send data to GAS.)

4) PERFORMANCE OF EVALUATION METRICS

TABLE 11, 12, 13, and 14 shows the precision, recall, classification accuracy and F1-score performance of each locally trained C in Edge 1, Edge 2, Edge 3, and aggregated C.

The results obtained were calculated using a noisy version of the previously provided data. TABLE 15 presents the accuracy, precision, recall, and F1-score obtained using the clean (original) data to evaluate the performance of the proposed classifier. The accuracy achieved is 98.95%. The performance is evaluated on the noisy data since the data is expected to be noisy in the real world.

TABLE 15. Proposed framework performance comparison with the denoised and original dataset.

	Precision	Recall	F1-score
N	0.950	0.990	0.970
S	0.980	0.970	0.980
V	0.970	0.990	0.980
F	0.990	0.930	0.960
Q	1.00	1.00	1.00
Accuracy	0.980	0.980	0.980
Macro average	0.980	0.980	0.980
Weighted average	0.980	0.980	0.980

5) GRAD-CAM INTERPRETATION

The CAM and GRAD-CAM platform was used to validate the model interpretation of the ECG dataset. Both CAM and GRAD-CAM enable the creation of 'visual explanations' for how a CNN model based its categorization and aided in interpreting the ECG pattern analysis results.

The ECG signal characteristics require extensive and critical comprehension to validate the explainability of the EXAI module.

The ECG signal consists of different waves, which must be critically examined. Each series includes a P-wave, QRS-wave, and T-wave representing the electrical activity captured during a single heartbeat. Its pattern representation is shown in FIGURE 15. When these waves are distorted, aberrant heartbeat signals result. Thus, the two categories of a regular pulse and a myocardial infarction are created. A shortage of blood flow and oxygen to the heart muscle is referred to as cardiac ischemia.

Ischemia can result in cardiac tissue death and a heart attack (myocardial infarction) if it is severe or persists for an extended period. Hence, it is important to classify these two waves in the inception state so that the correct measures must be taken to eradicate the critical situation. The suggested EXAI-enable framework has been implemented to identify the same, and its results are displayed in FIGURE 14. FIGURE 14 depicts the differences between the Myocardial infraction CAM means and the normal CAM means.

Hence, the suggested EXAI-enabled framework demonstrates that the proposed classifier considers these essential input sample properties and will classify the P-wave, QRS-wave, and T-wave for Myocardial infraction heartbeat and normal heartbeat. These insights can aid clinical practitioners in diagnosing the underlying health problems. However, it is strongly recommended to consult a clinical practitioner before utilizing these findings for medical advice.

The state of the art in TABLE 16 indicates how the proposed case study scheme has proved by providing the accuracy as 98 %, as compared to the other schemes [148], [149], and [150].

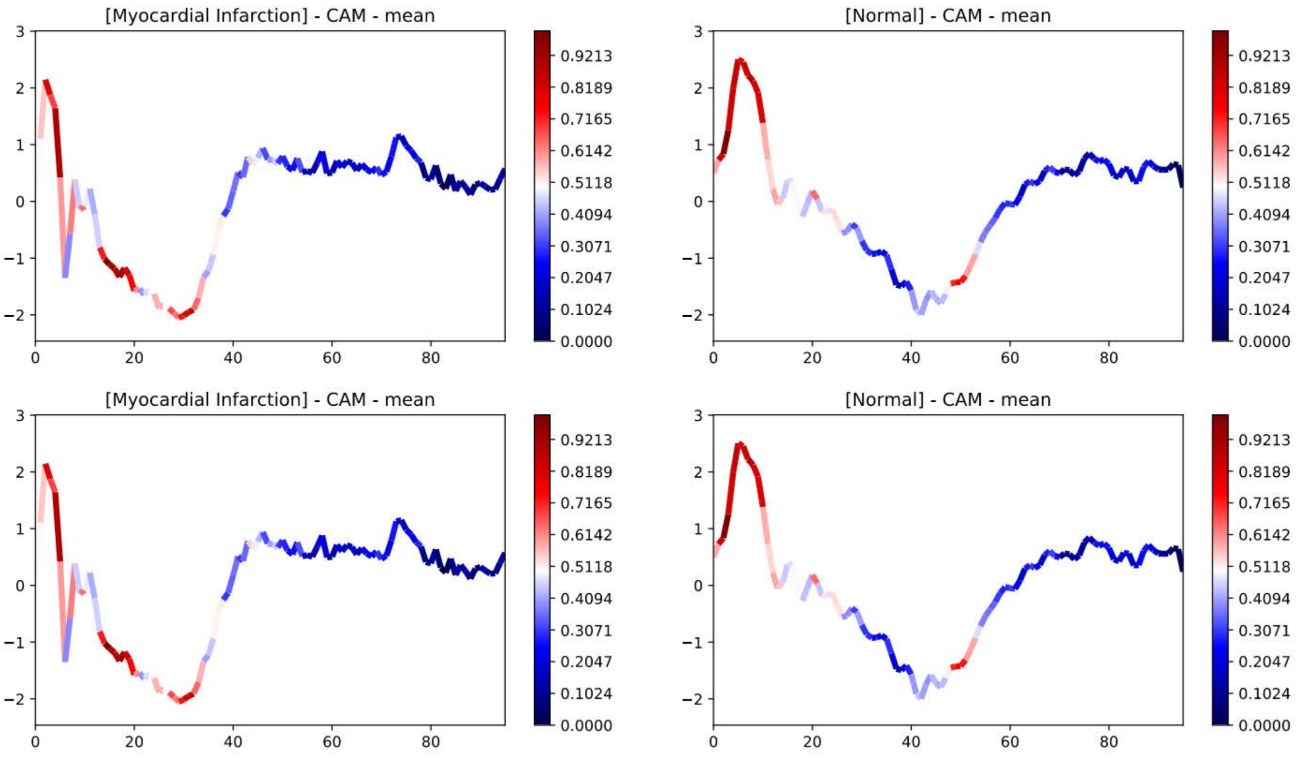

FIGURE 14. The EXAI outputs of the Grad-CAM for ECG signals.

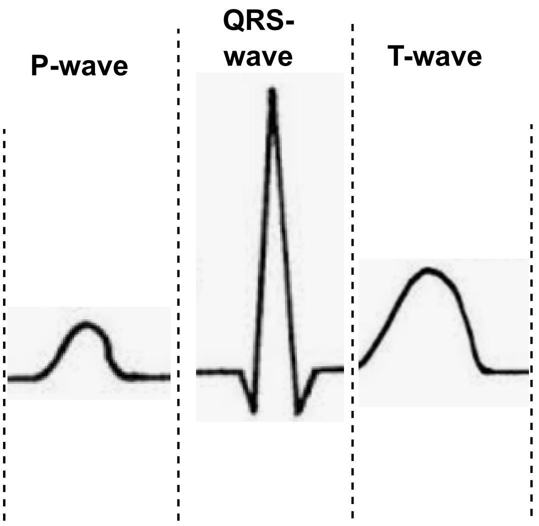

FIGURE 15. ECG pattern: major waves of the single normal pattern.

TABLE 16. State-of-the-art and its comparison with proposed framework.

Scheme	EXAI	Raw Input	Security	Realistic	Accuracy
Chen *et al.* [148]	N	N	Y	N	99%
Liaqat *et al.* [149]	N	N	N	N	86.5%
Atal *et al.* [150]	N	N	N	N	93.2%
Proposed case study	Y	Y	Y	Y	98%

Y denotes the parameter is present, N denotes the parameter is not present

X. LESSONS LEARNED

The authors discussed the utility of the Explainable AI (EXAI) and its applications. The discussion also addresses the opportunities and potential project implementations in healthcare industries and the domain of its associated application like ECG. The authors address the impact of the explainable AI over the AI terminology & its broad applications in the healthcare industry, which the existing surveys did not address with all the specifications and attributes. The authors have reviewed the recent literature and presented a solution taxonomy for EXAI-based medical-assisted programmable strategies. The paper also presents the design, architecture, ecosystems related to healthcare, and end-to-end communication explanations of EXAI that support the base of DL concepts (through implementation and performance analysis of auto-encoders and CNN), 5G communication network and its utilities in healthcare systems. The discussions on open issues and challenges have also been showcased. The authors have also introduced a healthcare-specific case study by taking concepts of CNN encoder and overall optimization to support the clause mentioned in the paper and support the argument.

In the healthcare 5.0 scenario, intelligent devices like wearable IoT devices are interfaced with the human body, which might be prone to false alarm probability, malfunction, time lags etc., thereby decreasing the model computational capability, accuracy, and interpretability for a trustworthy solution. The current ML models are also sensitive to small perturbations and the length of data sets, affecting the explainable model's outcome. Therefore, in real-time applications, these explainability models are in the very early stage of development. Soon, the field of EXAI would rise with extensive research and inculcation of more sophisticated models to enable robustness, better explainability, interpretability and verifiability of machine decisions by human experts in the healthcare context.

XI. CONCLUSION AND FUTURE SCOPE

Healthcare 5.0 ecosystems have shifted towards digital wellness, where decision models are analytics-driven with real-time predictions and informatics support. Thus, white-box analytics is preferred over traditional BB models to support this shift. The rise of interpretability and questions over the validity of AI models has shifted AI models to adapt EXAI decision modules. EXAI builds trust in clinical practices and allows interpretability and model debugging, which enhances performance through bias reduction. The proposed survey highlights the scope of EXAI in Healthcare 5.0 through supported techniques, architectures, and proposed models. The survey outlines the basics of EXAI, associated metrics, and different EXAI use-case applications. An EXAI-driven architecture for classifying and segmentation the COVID-19 patient is proposed. A solution taxonomy of EXAI in Healthcare 5.0 is presented, supported by a case study that integrates FL and EXAI for decentralized healthcare setups. The case study is presented with performance evaluations which validate the benefits of EXAI in health setups. Finally, the open issues, research challenges, and lessons learned from the survey are discussed.

In future scope, the authors intend to explore the integration of security and trust parameters in EXAI-driven healthcare 5.0. Integrating blockchain with EXAI would be explored in IoT-driven medical setups, where patients' critical data would be stored on blockchain ledgers, and through assisted EXAI modules, explainability and verifiability of analytical models would be explained sought. This would assure the dual benefits of trusted and interpretable analytics in decentralized healthcare setups.

REFERENCES

[1] P. Bhattacharya, S. Tanwar, U. Bodkhe, S. Tyagi, and N. Kumar, "BinDaaS: Blockchain-based deep-learning as-a-Service in healthcare 4.0 applications," *IEEE Trans. Netw. Sci. Eng.*, vol. 8, no. 2, pp. 1242–1255, Apr. 2021.

[2] R. Gupta, A. Shukla, P. Mehta, P. Bhattacharya, S. Tanwar, S. Tyagi, and N. Kumar, "VAHAK: A blockchain-based outdoor delivery scheme using UAV for healthcare 4.0 services," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Jul. 2020, pp. 255–260.

[3] R. Gupta, A. Shukla, and S. Tanwar, "AaYusH: A smart contract-based telesurgery system for healthcare 4.0," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Jun. 2020, pp. 1–6.

[4] E. Mbunge, B. Muchemwa, S. Jiyane, and J. Batani, "Sensors and healthcare 5.0: Transformative shift in virtual care through emerging digital health technologies," *Global Health J.*, vol. 5, no. 4, pp. 169–177, Dec. 2021.

[5] C. F. Pasluosta, H. Gassner, J. Winkler, J. Klucken, and B. M. Eskofier, "An emerging era in the management of Parkinson's disease: Wearable technologies and the Internet of Things," *IEEE J. Biomed. Health Inform.*, vol. 19, no. 6, pp. 1873–1881, Nov. 2015.

[6] P. A. Laplante and N. Laplante, "The Internet of Things in healthcare: Potential applications and challenges," *IT Prof.*, vol. 18, no. 3, pp. 2–4, May/Jun. 2016.

[7] B. Mohanta, P. Das, and S. Patnaik, "Healthcare 5.0: A paradigm shift in digital healthcare system using artificial intelligence, IoT and 5G communication," in *Proc. Int. Conf. Appl. Mach. Learn. (ICAML)*, Bhubaneswar, India, May 2019, pp. 191–196.

[8] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Incorporating explainable artificial intelligence (XAI) to aid the understanding of machine learning in the healthcare domain," in *Proc. Irish Conf. Artif. Intell. Cogn. Sci.*, Dublin, Ireland, 2020, pp. 1–12.

[9] A. Verma, P. Bhattacharya, M. Zuhair, S. Tanwar, and N. Kumar, "VaCoChain: Blockchain-based 5G-assisted UAV vaccine distribution scheme for future pandemics," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 5, pp. 1997–2007, May 2022.

[10] J. Amann, A. Blasimme, E. Vayena, D. Frey, and V. I. Madai, "Explainability for artificial intelligence in healthcare: A multidisciplinary perspective," *BMC Med. Informat. Decis. Making*, vol. 20, no. 1, p. 310, Dec. 2020.

[11] *Global Explainable AI Market (2021 to 2030)*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.nextmsc.com/report/explainable-ai-market

[12] A. Shaban-Nejad, M. Michalowski, and D. L. Buckeridge, "Explainable AI in healthcare and medicine: Building a culture of transparency and accountability," in *Proc. AAAI Int. Workshop Health Intell. (W3PHIAI)*, vol. 914. New York, NY, USA: Springer, Feb. 2020. [Online]. Available: https://link.springer.com/book/10.1007/978-3-030-53352-6

[13] S. Mohseni, N. Zarei, and E. D. Ragan, "A multidisciplinary survey and framework for design and evaluation of explainable AI systems," 2018, *arXiv:1811.11839*.

[14] *FWF Project Reference Model of Explainable AI for the Medical Domain*. Accessed: Apr. 24, 2022. [Online]. Available: https://human-centered.ai/project/explainable-ai-fwf-32554/

[15] *From One Year to Six Weeks: Highmark Health Teams With IBM to Accelerate AI in Urgent Times*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.ibm.com/blogs/journey-to-ai/2021/01/highmark-health-teams-with-ibm-advancing-data-science-to-stem-a-growing-tide-of-sepsis-and-covid-19-h

[16] *SAS Analytics and Solutions*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.sas.com/en_us/home.html

[17] *The Antidote Project or Explainable AI*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.inria.fr/en/explainable-ai-algorithm-learning

[18] *Explainable AI in Healthcare*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.mq.edu.au/research/research-centres-groups-and-facilities/healthy-people/centres/australian-institute-of-health-innovation/our-projects/explainable-ai-in-healthcare

[19] *Human-Centered AI Projects*. Accessed: Apr. 24, 2022. [Online]. Available: https://human-centered.ai/projects/

[20] *DoC Researcher Funded to Collaborate With Industry on World-Leading AI Research*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.imperial.ac.uk/news/204114/doc-researcher-funded-collaborate-with-industry/

[21] *Cancer Research UK Project Award: Cancer Risk Algorithms and Their Influence on Clinical Judgement*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.imperial.ac.uk/people/o.kostopoulou

[22] M. R. Islam, M. U. Ahmed, S. Barua, and S. Begum, "A systematic review of explainable artificial intelligence in terms of different application domains and tasks," *Appl. Sci.*, vol. 12, no. 3, p. 1353, Jan. 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/3/1353

[23] *In Latest Facebook Data Exposure, History Repeats Itself*. Accessed: Apr. 24, 2022. [Online]. Available: https://www.wired.com/story/facebook-apps-540-million-records/

[24] *Maze Ransomware Operators Allegedly Breached National Highways Authority of India (NHAI)*. Accessed: Apr. 24, 2022. [Online]. Available: https://securityaffairs.co/wordpress/105467/cyber-crime/maze-ransomware-india-nhai.html

[25] M. A. Ahmad, A. Teredesai, and C. Eckert, "Interpretable machine learning in healthcare," in *Proc. IEEE Int. Conf. Healthcare Informat. (ICHI)*, New York, NY, USA, Jun. 2018, p. 447.

[26] E. H. Shortliffe and M. J. Sepúlveda, "Clinical decision support in the era of artificial intelligence," *J. Amer. Med. Assoc.*, vol. 320, no. 21, pp. 2199–2200, 2018.

[27] D. Kim, J. Chung, J. Choi, M. D. Succi, J. Conklin, M. G. F. Longo, J. B. Ackman, B. P. Little, M. Petranovic, M. K. Kalra, M. H. Lev, and S. Do, "Accurate auto-labeling of chest X-ray images based on quantitative similarity to an explainable AI model," *Nature Commun.*, vol. 13, no. 1, p. 1867, Apr. 2022.

[28] K. M. Abiodun, J. B. Awotunde, D. R. Aremu, and E. A. Adeniyi, "Explainable AI for fighting COVID-19 pandemic: Opportunities, challenges, and future prospects," in *Computational Intelligence for COVID-19 and Future Pandemics*. Singapore: Springer, 2022, pp. 315–332.

[29] A. Anand, T. Kadian, M. K. Shetty, and A. Gupta, "Explainable AI decision model for ECG data of cardiac disorders," *Biomed. Signal Process. Control*, vol. 75, May 2022, Art. no. 103584.

[30] S. Wesolowski, G. Lemmon, E. J. Hernandez, A. Henrie, T. A. Miller, D. Weyhrauch, M. D. Puchalski, B. E. Bray, R. U. Shah, V. G. Deshmukh, R. Delaney, H. J. Yost, K. Eilbeck, M. Tristani-Firouzi, and M. Yandell, "An explainable artificial intelligence approach for predicting cardiovascular outcomes using electronic health records," *PLOS Digit. Health*, vol. 1, no. 1, pp. 1–17, Jan. 2022.

[31] H. Han and X. Liu, "The challenges of explainable AI in biomedical data science," *BMC Bioinf.*, vol. 22, no. S12, p. 443, Jan. 2022.

[32] D. Luckey, H. Fritz, D. Legatiuk, J. J. P. Abadía, C. Walther, and K. Smarsly, "Explainable artificial intelligence to advance structural health monitoring," in *Structural Health Monitoring Based on Data Science Techniques*. Cham, Switzerland: Springer, 2022, pp. 331–346.

[33] G. Antoniou, E. Papadakis, and G. Baryannis, "Mental health diagnosis: A case for explainable artificial intelligence," *Int. J. Artif. Intell. Tools*, vol. 31, no. 3, May 2022, Art. no. 2241003.

[34] E. Vayena, A. Blasimme, and I. G. Cohen, "Machine learning in medicine: Addressing ethical challenges," *PLOS Med.*, vol. 15, no. 11, pp. 1–4, Nov. 2018.

[35] B. Kitchenham, O. P. Brereton, D. Budgen, M. Turner, J. Bailey, and S. Linkman, "Systematic literature reviews in software engineering—A systematic literature review," *Inf. Softw. Technol.*, vol. 51, no. 1, pp. 7–15, 2009.

[36] S. Keele, "Guidelines for performing systematic literature reviews in software engineering," EBSE, Keele University, Keele, U.K., Version 2.3, Tech. Rep. EBSE-2007-01, 2007.

[37] Y. Liu, J.-W. Bi, and Z.-P. Fan, "Ranking products through online reviews: A method based on sentiment analysis technique and intuitionistic fuzzy set theory," *Inf. Fusion*, vol. 36, pp. 149–161, Jun. 2017.

[38] J. Zhang, Z. Yin, P. Chen, and S. Nichele, "Emotion recognition using multi-modal data and machine learning techniques: A tutorial and review," *Inf. Fusion*, vol. 59, pp. 103–126, Jul. 2020.

[39] D. Saraswat, P. Bhattacharya, A. Singh, A. Verma, S. Tanwar, and N. Kumar, "Secure 5G-assisted UAV access scheme in IoBT for region demarcation and surveillance operations," *IEEE Commun. Standards Mag.*, vol. 6, no. 1, pp. 58–66, Mar. 2022.

[40] C. Maathuis, "On explainable AI solutions for targeting in cyber military operations," in *Proc. Int. Conf. Cyber Warfare Secur.*, 2022, vol. 17, no. 1, pp. 166–175.

[41] R. Ureña, G. Kou, Y. Dong, F. Chiclana, and E. Herrera-Viedma, "A review on trust propagation and opinion dynamics in social networks and group decision making frameworks," *Inf. Sci.*, vol. 478, pp. 461–475, Apr. 2019.

[42] Z. Zhang, Y. Gao, and Z. Li, "Consensus reaching for social network group decision making by considering leadership and bounded confidence," *Knowl.-Based Syst.*, vol. 204, Sep. 2020, Art. no. 106240.

[43] Y. Dong, M. Zhan, G. Kou, Z. Ding, and H. Liang, "A survey on the fusion process in opinion dynamics," *Inf. Fusion*, vol. 43, pp. 57–65, Sep. 2018.

[44] R.-X. Nie, Z.-P. Tian, J.-Q. Wang, and K. S. Chin, "Hotel selection driven by online textual reviews: Applying a semantic partitioned sentiment dictionary and evidence theory," *Int. J. Hospitality Manag.*, vol. 88, Jul. 2020, Art. no. 102495.

[45] A. Verma, P. Bhattacharya, Y. Patel, K. Shah, S. Tanwar, and B. Khan, "Data localization and privacy-preserving healthcare for big data applications: Architecture and future directions," in *Emerging Technologies for Computing, Communication and Smart Cities*, P. K. Singh, M. H. Kolekar, S. Tanwar, S. T. Wierzchoń, and R. K. Bhatnagar, Eds. Singapore: Springer, 2022, pp. 233–244.

[46] T. Folke, S. C. Yang, S. Anderson, and P. Shafto, "Explainable AI for medical imaging: Explaining pneumothorax diagnoses with Bayesian teaching," *CoRR*, vol. abs/2106.04684, pp. 1–20, Jun. 2021. [Online]. Available: https://arxiv.org/abs/2106.04684, doi: 10.48550/arxiv.2106.04684.

[47] R.-X. Ding, I. Palomares, X. Wang, G.-R. Yang, B. Liu, Y. Dong, E. Herrera-Viedma, and F. Herrera, "Large-scale decision-making: Characterization, taxonomy, challenges and future directions from an artificial intelligence and applications perspective," *Inf. Fusion*, vol. 59, pp. 84–102, Jul. 2020.

[48] M. McFarland, "Uber shuts down self-driving operations in Arizona," *CNNMoney*, vol. 809, p. 3, May 2018.

[49] M. Bojarski, D. D. Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Müller, J. Zhang, X. Zhang, J. Zhao, and K. Zieba, "End to end learning for self-driving cars," 2016, *arXiv:1604.07316*.

[50] J. Haspiel, N. Du, J. Meyerson, L. P. Robert, Jr., D. Tilbury, X. J. Yang, and A. K. Pradhan, "Explanations and expectations: Trust building in automated vehicles," in *Proc. Companion ACM/IEEE Int. Conf. Hum.-Robot Interact.* New York, NY, USA: Association for Computing Machinery, Mar. 2018, pp. 119–120.

[51] X. Chao, "Behavior monitoring methods for trade-based money laundering integrating macro and micro prudential regulation: A case from China," *Technol. Econ. Develop. Econ.*, vol. 25, no. 6, pp. 1081–1096, 2019.

[52] S. Moradi and F. M. Rafiei, "A dynamic credit risk assessment model with data mining techniques: Evidence from Iranian banks," *Financial Innov.*, vol. 5, no. 1, pp. 1–27, Mar. 2019.

[53] M. Guo, X. Liao, J. Liu, and Q. Zhang, "Consumer preference analysis: A data-driven multiple criteria approach integrating online information," *Omega*, vol. 96, Oct. 2020, Art. no. 102074.

[54] G. Kou, P. Yang, Y. Peng, F. Xiao, Y. Chen, and F. E. Alsaadi, "Evaluation of feature selection methods for text classification with small datasets using multiple criteria decision-making methods," *Appl. Soft Comput.*, vol. 86, Jan. 2020, Art. no. 105836.

[55] S. Tan, R. Caruana, G. Hooker, and Y. Lou, "Distill-and-compare: Auditing black-box models using transparent model distillation," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.* New York, NY, USA: Association for Computing Machinery, Dec. 2018, pp. 303–310.

[56] R. A. Berk and J. Bleich, "Statistical procedures for forecasting criminal behavior: A comparative assessment," *Criminol. Pub. Policy*, vol. 12, no. 1, p. 513, Feb. 2013.

[57] P. Giudici and E. Raffinetti, "Explainable AI methods in cyber risk management," *Qual. Rel. Eng. Int.*, vol. 38, no. 3, pp. 1318–1326, Apr. 2022. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/qre.2939

[58] H. H. Guerrero and J. R. Bradley, "Failure modes and effects analysis: An evaluation of group versus individual performance," *Prod. Oper. Manage.*, vol. 22, no. 6, pp. 1524–1539, Nov. 2013.

[59] J. Huang, Z. Li, and H. C. Liu, "New approach for failure mode and effect analysis using linguistic distribution assessments and TODIM method," *Rel. Eng. Syst. Saf.*, vol. 167, pp. 302–309, Nov. 2017.

[60] M. Aggarwal and A. F. Tehrani, "Modelling human decision behaviour with preference learning," *Informs J. Comput.*, vol. 31, no. 2, pp. 318–334, Apr. 2019.

[61] S. Corrente, S. Greco, M. Kadziński, and R. Słowiński, "Robust ordinal regression in preference learning and ranking," *Mach. Learn.*, vol. 93, nos. 2–3, pp. 381–422, 2013.

[62] V. C. Raykar, R. Duraiswami, and B. Krishnapuram, "A fast algorithm for learning a ranking function from large-scale data sets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1158–1170, Jul. 2008.

[63] Z. Zhao, H. Lu, D. Cai, X. He, and Y. Zhuang, "User preference learning for online social recommendation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 9, pp. 2522–2534, Sep. 2016.

[64] G. Kou, X. Chao, Y. Peng, F. E. Alsaadi, and E. Herrera-Viedma, "Machine learning methods for systemic risk analysis in financial sectors," *Technol. Econ. Develop. Economy*, vol. 25, no. 5, pp. 716–742, 2019.

[65] P. W. Koh and P. Liang, "Understanding black-box predictions via influence functions," in *Int. Conf. Mach. Learn.*, Sydney, NSW, Australia, 2017, pp. 1885–1894.

[66] M. Bojarski, P. Yeres, A. Choromanska, K. Choromanski, B. Firner, L. Jackel, and U. Müller, "Explaining how a deep neural network trained with End-to-End learning steers a car," 2017, *arXiv:1704.07911*.

[67] V. Dignum, "Responsible artificial intelligence: Designing AI for human values," *ITU J., ICT Discoveries*, vol. 1, pp. 1–8, Sep. 2017.

[68] T. Miller, "Explanation in artificial intelligence: Insights from the social sciences," *Artif. Intell.*, vol. 267, pp. 1–38, Feb. 2019.

[69] DARPA. (2017). *Powerful But Limited: A DARPA Perspective on AI*. Accessed: Mar. 10, 2022. [Online]. Available: https://sites.nationalacademies.org/cs/groups/pgasite/documents/webpagepga_177035.pdf

[70] *A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy*, Global Catastrophic Risk, Washington, DC, USA, Inst. Work. Paper 17-1, 2017.

[71] M. van Lent, W. Fisher, and M. Mancuso, "An explainable artificial intelligence system for small-unit tactical behavior," in *Proc. 16th Conf. Innov. Appl. Artif. Intell. (IAAI)*. San Jose, CA, USA: AAAI Press, 2004, pp. 900–907.

[72] H. Mankodiya, D. Jadav, R. Gupta, S. Tanwar, A. Alharbi, A. Tolba, B.-C. Neagu, and M. S. Raboaca, "XAI-fall: Explainable AI for fall detection on wearable devices using sequence models and XAI techniques," *Mathematics*, vol. 10, no. 12, p. 1990, Jun. 2022. [Online]. Available: https://www.mdpi.com/2227-7390/10/12/1990

[73] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, and D. Pedreschi, "A survey of methods for explaining black box models," *ACM Comput. Surv.*, vol. 51, no. 5, pp. 1–42, Sep. 2019.

[74] G. Vilone and L. Longo, "Explainable artificial intelligence: A systematic review," 2020, *arXiv:2006.00093*.

[75] P. Hamet and J. Tremblay, "Artificial intelligence in medicine," *Metabolism*, vol. 69, pp. S36–S40, Apr. 2017.

[76] F. Amisha, P. Malik, M. Pathania, and V. K. Rathaur, "Overview of artificial intelligence in medicine," *J. Family Med. Primary Care*, vol. 8, no. 7, pp. 2328–2331, Jul. 2019.

[77] J. K. Ruffle, A. D. Farmer, and Q. Aziz, "Artificial intelligence-assisted gastroenterology—Promises and pitfalls," *Off. J. Amer. College Gastroenterol.*, vol. 114, no. 3, pp. 422–428, 2019.

[78] M. Bhavin, S. Tanwar, N. Sharma, S. Tyagi, and N. Kumar, "Blockchain and quantum blind signature-based hybrid scheme for healthcare 5.0 applications," *J. Inf. Secur. Appl.*, vol. 56, Feb. 2021, Art. no. 102673.

[79] R. Gupta, P. Bhattacharya, S. Tanwar, N. Kumar, and S. Zeadally, "GaRuDa: A blockchain-based delivery scheme using drones for healthcare 5.0 applications," *IEEE Internet Things Mag.*, vol. 4, no. 4, pp. 60–66, Dec. 2021.

[80] D. Gunning and D. Aha, "DARPA's explainable artificial intelligence (XAI) program," *AI Mag.*, vol. 40, no. 2, pp. 44–58, Jun. 2019.

[81] S. Brocas, S. Friedler, M. Hardt, J. Kroll, S. Venkatasubramanian, and H. Wallach, "The FAT-ML workshop series on fairness, accountability, and transparency in machine learning," Princeton Univ., Princeton, NJ, USA, Tech. Rep. Accessed: Mar. 20, 2022. [Online]. Available: https://www.fatml.org/

[82] C. Mistry, U. Thakker, R. Gupta, M. S. Obaidat, S. Tanwar, N. Kumar, and J. J. P. C. Rodrigues, "MedBlock: An AI-enabled and blockchain-driven medical healthcare system for COVID-19," in *Proc. IEEE Int. Conf. Commun.*, Jun. 2021, pp. 1–6.

[83] H. Mankodiya, M. S. Obaidat, R. Gupta, and S. Tanwar, "XAI-AV: Explainable artificial intelligence for trust management in autonomous vehicles," in *Proc. Int. Conf. Commun., Comput., Cybersec., Informat. (CCCI)*, Oct. 2021, pp. 1–5.

[84] X.-H. Li, C. Chen Cao, Y. Shi, W. Bai, H. Gao, L. Qiu, C. Wang, Y. Gao, S. Zhang, X. Xue, and L. Chen, "A survey of data-driven and knowledge-aware explainable AI," *IEEE Trans. Knowl. Data Eng.*, vol. 34, no. 1, pp. 29–49, Jan. 2022.

[85] U. Pawar, D. O'Shea, S. Rea, and R. O'Reilly, "Explainable AI in healthcare," in *Proc. Int. Conf. Cyber Situational Awareness, Data Anal. Assessment (CyberSA)*, Dublin, Ireland, Jun. 2020, pp. 1–2.

[86] Y. Zhang, Y. Weng, and J. Lund, "Applications of explainable artificial intelligence in diagnosis and surgery," *Diagnostics*, vol. 12, no. 2, p. 237, Jan. 2022.

[87] G. Yang, Q. Ye, and J. Xia, "Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond," *Inf. Fusion*, vol. 77, pp. 29–52, Jan. 2022.

[88] 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.

[89] 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.

[90] 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.

[91] N. Taimoor and S. Rehman, "Reliable and resilient AI and IoT-based personalised healthcare services: A survey," *IEEE Access*, vol. 10, pp. 535–563, 2022.

[92] M. Moradi and M. Samwald, "Deep learning, natural language processing, and explainable artificial intelligence in the biomedical domain," 2022, *arXiv:2202.12678*.

[93] J. C. C. Kwong, A. Khondker, C. Tran, E. Evans, A. I. Cozma, A. Javidan, A. Ali, M. Jamal, T. Short, F. Papanikolaou, J. R. Srigley, B. Fine, and A. Feifer, "Explainable artificial intelligence to predict the risk of side-specific extraprostatic extension in pre-prostatectomy patients," *Can. Urol. Assoc. J.*, vol. 16, no. 6, pp. 213–221, Jan. 2022.

[94] W. Shi, L. Tong, Y. Zhuang, Y. Zhu, and M. D. Wang, "EXAM: An explainable attention-based model for COVID-19 automatic diagnosis," in *Proc. 11th ACM Int. Conf. Bioinf., Comput. Biol. Health Informat.* New York, NY, USA: Association for Computing Machinery, Sep. 2020.

[95] M. R. Karim, T. Döhmen, M. Cochez, O. Beyan, D. Rebholz-Schuhmann, and S. Decker, "DeepCOVIDExplainer: Explainable COVID-19 diagnosis from chest X-ray images," in *Proc. IEEE Int. Conf. Bioinf. Biomed. (BIBM)*, Dec. 2020, pp. 1034–1037.

[96] M. M. Ahsan, K. D. Gupta, M. M. Islam, S. Sen, M. L. Rahman, and M. Shakhawat Hossain, "COVID-19 symptoms detection based on NasNetMobile with explainable AI using various imaging modalities," *Mach. Learn. Knowl. Extraction*, vol. 2, no. 4, pp. 490–504, Oct. 2020.

[97] M. Nazar, M. M. Alam, E. Yafi, and M. M. Su'ud, "A systematic review of human–computer interaction and explainable artificial intelligence in healthcare with artificial intelligence techniques," *IEEE Access*, vol. 9, pp. 153316–153348, 2021.

[98] D.-N. Le, V. S. Parvathy, D. Gupta, A. Khanna, J. J. P. C. Rodrigues, and K. Shankar, "IoT enabled depthwise separable convolution neural network with deep support vector machine for COVID-19 diagnosis and classification," *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 11, pp. 3235–3248, Nov. 2021.

[99] L. Tan, K. Yu, A. K. Bashir, X. Cheng, F. Ming, L. Zhao, and X. Zhou, "Toward real-time and efficient cardiovascular monitoring for COVID-19 patients by 5G-enabled wearable medical devices: A deep learning approach," *Neural Comput. Appl.*, early access, pp. 1–14, Jul. 2021, doi: 10.1007/s00521-021-06219-9.

[100] P. P. Angelov, E. A. Soares, R. Jiang, N. I. Arnold, and P. M. Atkinson, "Explainable artificial intelligence: An analytical review," *WIREs Data Mining Knowl. Discovery*, vol. 11, no. 5, 2021, Art. no. e1424.

[101] J. D. Fuhrman, N. Gorre, Q. Hu, H. Li, I. El Naqa, and M. L. Giger, "A review of explainable and interpretable AI with applications in COVID-19 imaging," *Med. Phys.*, vol. 49, no. 1, pp. 1–14, Jan. 2022.

[102] S. K. Jagatheesaperumal, P. Mishra, N. Moustafa, and R. Chauhan, "A holistic survey on the use of emerging technologies to provision secure healthcare solutions," *Comput. Electr. Eng.*, vol. 99, Apr. 2022, Art. no. 107691.

[103] T. Speith, "A review of taxonomies of explainable artificial intelligence (XAI) methods," in *Proc. ACM Conf. Fairness, Accountability, Transparency*. New York, NY, USA: Association for Computing Machinery, Jun. 2022, pp. 2239–2250, doi: 10.1145/3531146.3534639.

[104] A. M. Oprescu, G. Miró-Amarante, L. García-Díaz, V. E. Rey, A. Chimenea-Toscano, R. Martínez-Martínez, and M. C. Romero-Ternero, "Towards a data collection methodology for responsible artificial intelligence in health: A prospective and qualitative study in pregnancy," *Inf. Fusion*, vols. 83–84, pp. 53–78, Jul. 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253522000355

[105] W. Qi and H. Su, "A cybertwin based multimodal network for ECG patterns monitoring using deep learning," *IEEE Trans. Ind. Informat.*, vol. 18, no. 10, pp. 6663–6670, Oct. 2022.

[106] P. Thakkar, K. Varma, V. Ukani, S. Mankad, and S. Tanwar, "Combining user-based and item-based collaborative filtering using machine learning," in *Information and Communication Technology for Intelligent Systems*, S. C. Satapathy and A. Joshi, Eds. Singapore: Springer, 2019, pp. 173–180.

[107] C. Verma, V. Stoffová, Z. Illés, S. Tanwar, and N. Kumar, "Machine learning-based student's native place identification for real-time," *IEEE Access*, vol. 8, pp. 130840–130854, 2020.

[108] K. Patel, D. Mehta, C. Mistry, R. Gupta, S. Tanwar, N. Kumar, and M. Alazab, "Facial sentiment analysis using AI techniques: State-of-the-art, taxonomies, and challenges," *IEEE Access*, vol. 8, pp. 90495–90519, 2020.

[109] H. Mankodiya, D. Jadav, R. Gupta, S. Tanwar, W.-C. Hong, and R. Sharma, "OD-XAI: Explainable AI-based semantic object detection for autonomous vehicles," *Appl. Sci.*, vol. 12, no. 11, p. 5310, May 2022. [Online]. Available: https://www.mdpi.com/2076-3417/12/11/5310

[110] A. Arun, C. V. Jawahar, and M. P. Kumar, "Weakly supervised instance segmentation by learning annotation consistent instances," 2020, *arXiv:2007.09397*.

[111] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?" 2016, *arXiv:1611.07450*.

[112] J. Frade, T. Pereira, J. Morgado, F. Silva, C. Freitas, J. Mendes, E. Negrão, B. F. de Lima, M. C. D. Silva, A. J. Madureira, I. Ramos, J. L. Costa, V. Hespanhol, A. Cunha, and H. P. Oliveira, ''Multiple instance learning for lung pathophysiological findings detection using CT scans,'' *Med. Biol. Eng. Comput.*, vol. 60, no. 6, pp. 1569–1584, Jun. 2022.

[113] A. Raza, K. P. Tran, L. Koehl, and S. Li, ''Designing ECG monitoring healthcare system with federated transfer learning and explainable AI,'' *Knowl.-Based Syst.*, vol. 236, Jan. 2022, Art. no. 107763.

[114] D. Vale, A. El-Sharif, and M. Ali, ''Explainable artificial intelligence (XAI) post-hoc explainability methods: Risks and limitations in non-discrimination law,'' *AI Ethics*, early access, pp. 1–12, Mar. 2022, doi: 10.1007/s43681-022-00142-y.

[115] F. Milletari, N. Navab, and S.-A. Ahmadi, ''V-Net: Fully convolutional neural networks for volumetric medical image segmentation,'' 2016, *arXiv:1606.04797*.

[116] K. He, X. Zhang, S. Ren, and J. Sun, ''Deep residual learning for image recognition,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 770–778.

[117] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, ''ImageNet: A large-scale hierarchical image database,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Miami, FL, USA, Jun. 2009, pp. 248–255.

[118] W. Zhang, F. Liu, L. Luo, and J. Zhang, ''Predicting drug side effects by multi-label learning and ensemble learning,'' *BMC Bioinf.*, vol. 16, no. 1, pp. 1–11, Dec. 2015.

[119] G. Yang, F. Raschke, T. R. Barrick, and F. A. Howe, ''Manifold learning in MR spectroscopy using nonlinear dimensionality reduction and unsupervised clustering,'' *Magn. Reson. Med.*, vol. 74, no. 3, pp. 868–878, Sep. 2015.

[120] L. P. Zhao and H. Bolouri, ''Object-oriented regression for building predictive models with high dimensional omics data from translational studies,'' *J. Biomed. Informat.*, vol. 60, pp. 431–445, Apr. 2016.

[121] S. G. Kim, N. Theera-Ampornpunt, C.-H. Fang, M. Harwani, A. Grama, and S. Chaterji, ''Opening up the blackbox: An interpretable deep neural network-based classifier for cell-type specific enhancer predictions,'' *BMC Syst. Biol.*, vol. 10, no. S2, pp. 243–258, Aug. 2016.

[122] J. Hao, Y. Kim, T.-K. Kim, and M. Kang, ''PASNet: Pathway-associated sparse deep neural network for prognosis prediction from high-throughput data,'' *BMC Bioinf.*, vol. 19, no. 1, pp. 1–13, Dec. 2018.

[123] M. Bernardini, L. Romeo, P. Misericordia, and E. Frontoni, ''Discovering the type 2 diabetes in electronic health records using the sparse balanced support vector machine,'' *IEEE J. Biomed. Health Informat.*, vol. 24, no. 1, pp. 235–246, Jan. 2020.

[124] A. Eck, L. M. Zintgraf, E. F. J. de Groot, T. G. J. de Meij, T. S. Cohen, P. H. M. Savelkoul, M. Welling, and A. E. Budding, ''Interpretation of microbiota-based diagnostics by explaining individual classifier decisions,'' *BMC Bioinf.*, vol. 18, no. 1, pp. 1–13, Dec. 2017.

[125] W. Ge, J.-W. Huh, Y. R. Park, J.-H. Lee, Y.-H. Kim, and A. Turchin, ''An interpretable ICU mortality prediction model based on logistic regression and recurrent neural networks with LSTM units,'' in *Proc. AMIA Annu. Symp.* Bethesda, MD, USA: American Medical Informatics Association, 2018, p. 460.

[126] J. Zuallaert, F. Godin, M. Kim, A. Soete, Y. Saeys, and W. De Neve, ''SpliceRover: Interpretable convolutional neural networks for improved splice site prediction,'' *Bioinformatics*, vol. 34, no. 24, pp. 4180–4188, Dec. 2018.

[127] J. Suh, S. Yoo, J. Park, S. Y. Cho, M. C. Cho, H. Son, and H. Jeong, ''Development and validation of an explainable artificial intelligence-based decision-supporting tool for prostate biopsy,'' *BJU Int.*, vol. 126, no. 6, pp. 694–703, Aug. 2020.

[128] A. Singh, A. R. Mohammed, J. Zelek, and V. Lakshminarayanan, ''Interpretation of deep learning using attributions: Application to ophthalmic diagnosis,'' *Proc. SPIE*, vol. 11511, pp. 39–49, Aug. 2020.

[129] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, ''Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission,'' in *Proc. 21st ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Sydney, NSW, Australia, Aug. 2015, pp. 1721–1730.

[130] 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, no. 3, pp. 1350–1371, 2015.

[131] Z. Che, S. Purushotham, R. Khemani, and Y. Liu, ''Interpretable deep models for ICU outcome prediction,'' in *Proc. AMIA Annu. Symp.* Bethesda, MD, USA: American Medical Informatics Association, 2016, p. 371.

[132] Y. Ming, H. Qu, and E. Bertini, ''RuleMatrix: Visualizing and understanding classifiers with rules,'' *IEEE Trans. Vis. Comput. Graphics*, vol. 25, no. 1, pp. 342–352, Jan. 2019.

[133] C. Xiao, T. Ma, A. B. Dieng, D. M. Blei, and F. Wang, ''Readmission prediction via deep contextual embedding of clinical concepts,'' *PLoS ONE*, vol. 13, no. 4, Apr. 2018, Art. no. e0195024.

[134] L. Pan, G. Liu, X. Mao, H. Li, J. Zhang, H. Liang, and X. Li, ''Development of prediction models using machine learning algorithms for girls with suspected central precocious puberty: Retrospective study,'' *JMIR Med. Informat.*, vol. 7, no. 1, Feb. 2019, Art. no. e11728.

[135] S. Ghafouri-Fard, M. Taheri, M. D. Omrani, A. Daaee, H. Mohammad-Rahimi, and H. Kazazi, ''Application of single-nucleotide polymorphisms in the diagnosis of autism spectrum disorders: A preliminary study with artificial neural networks,'' *J. Mol. Neurosci.*, vol. 68, no. 4, pp. 515–521, Aug. 2019.

[136] M. S. Kovalev, L. V. Utkin, and E. M. Kasimov, ''SurvLIME: A method for explaining machine learning survival models,'' *Knowl.-Based Syst.*, vol. 203, Sep. 2020, Art. no. 106164.

[137] A. Meldo, L. Utkin, M. Kovalev, and E. Kasimov, ''The natural language explanation algorithms for the lung cancer computer-aided diagnosis system,'' *Artif. Intell. Med.*, vol. 108, Aug. 2020, Art. no. 101952.

[138] C. Panigutti, A. Perotti, and D. Pedreschi, ''Doctor XAI: An ontology-based approach to black-box sequential data classification explanations,'' in *Proc. Conf. Fairness, Accountability, Transparency*. New York, NY, USA: Association for Computing Machinery, Jan. 2020, pp. 629–639.

[139] J. Zhang, K. Kowsari, J. H. Harrison, J. M. Lobo, and L. E. Barnes, ''Patient2Vec: A personalized interpretable deep representation of the longitudinal electronic health record,'' *IEEE Access*, vol. 6, pp. 65333–65346, 2018.

[140] E. Choi, M. T. Bahadori, J. Sun, J. Kulas, A. Schuetz, and W. Stewart, ''RETAIN: An interpretable predictive model for healthcare using reverse time attention mechanism,'' in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 1–9.

[141] D. A. Kaji, J. R. Zech, J. S. Kim, S. K. Cho, N. S. Dangayach, A. B. Costa, and E. K. Oermann, ''An attention based deep learning model of clinical events in the intensive care unit,'' *PLoS ONE*, vol. 14, no. 2, Feb. 2019, Art. no. e0211057.

[142] B. Shickel, T. J. Loftus, L. Adhikari, T. Ozrazgat-Baslanti, A. Bihorac, and P. Rashidi, ''DeepSOFA: A continuous acuity score for critically ill patients using clinically interpretable deep learning,'' *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, Dec. 2019.

[143] H. Hu, A. Xiao, S. Zhang, Y. Li, X. Shi, T. Jiang, L. Zhang, L. Zhang, and J. Zeng, ''DeepHINT: Understanding HIV-1 integration via deep learning with attention,'' *Bioinformatics*, vol. 35, no. 10, pp. 1660–1667, May 2019.

[144] S. Rabiul Islam, W. Eberle, S. Bundy, and S. K. Ghafoor, ''Infusing domain knowledge in AI-based 'black box' models for better explainability with application in bankruptcy prediction,'' 2019, *arXiv:1905.11474*.

[145] G. B. Moody and R. G. Mark, ''The impact of the MIT-BIH arrhythmia database,'' *IEEE Eng. Med. Biol. Mag.*, vol. 20, no. 3, pp. 45–50, May 2001.

[146] A. Shoughi and M. B. Dowlatshahi, ''A practical system based on CNN-BLSTM network for accurate classification of ECG heartbeats of MIT-BIH imbalanced dataset,'' in *Proc. 26th Int. Comput. Conf., Comput. Soc. Iran (CSICC)*, Mar. 2021, pp. 1–6.

[147] A. Mishra, G. Dharahas, S. Gite, K. Kotecha, D. Koundal, A. Zaguia, M. Kaur, and H.-N. Lee, ''ECG data analysis with denoising approach and customized CNNs,'' *Sensors*, vol. 22, no. 5, p. 1928, Mar. 2022.

[148] Y. Chen, X. Qin, J. Wang, C. Yu, and W. Gao, ''FedHealth: A federated transfer learning framework for wearable healthcare,'' *IEEE Intell. Syst.*, vol. 35, no. 4, pp. 83–93, Jul. 2020.

[149] S. Liaqat, K. Dashtipour, A. Zahid, K. Assaleh, K. Arshad, and N. Ramzan, ''Detection of atrial fibrillation using a machine learning approach,'' *Information*, vol. 11, no. 12, p. 549, Nov. 2020.

[150] D. K. Atal and M. Singh, ''Arrhythmia classification with ECG signals based on the optimization-enabled deep convolutional neural network,'' *Comput. Methods Programs Biomed.*, vol. 196, Nov. 2020, Art. no. 105607.

DEEPTI SARASWAT is currently pursuing the Ph.D. degree with the Dhirubhai Ambani Institute of Information and Communication Technology (DA-IICT), Gandhinagar, India. She has more than five years of industrial experience at Samsung Research and Development Research Institute, Bengaluru, India, where she has worked on implementation of web browsers and the Internet of Things (IoT). She is employed as an Assistant Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad, India. She has authored and coauthored articles in leading SCI journals and IEEE conferences. Her research interests include artificial intelligence, data security and privacy, blockchain technology, optimization techniques, and the IoT. She is awarded two Best Paper Awards in IEEE-ICIEM-2021 and 2022, London, U.K.

PRONAYA BHATTACHARYA (Member, IEEE) is currently employed as an Assistant Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad, India. He has over eight years of teaching experience. He has authored or coauthored more than 70 research papers in leading SCI journals and top core IEEE COMSOC A* conferences. Some of his top-notch findings are published in reputed SCI journals, like IEEE Journal of Biomedical and Health Informatics, IEEE Transactions on Vehicular Technology, IEEE Internet of Things Journal, IEEE Transactions on Network Science and Engineering, IEEE Transactions on Computational Social Systems, IEEE Transactions on Network and Service Management, IEEE Access, IEEE Sensors Journal, *IEEE Internet of Things Magazine*, *IEEE Communications Standards Magazine*, *ETT* (Wiley), *Expert Systems* (Wiley), *FGCS* (Elsevier), *OQEL* (Springer), *WPC* (Springer), ACM-MOBICOM, IEEE-INFOCOM, IEEE-ICC, IEEE-CITS, IEEE-ICIEM, IEEE-CCCI, and IEEE-ECAI. He has 1043 citations to his credit with an H-index of 17 and an i10-index of 24. His research interests include artificial intelligence, healthcare analytics, optical switching and networking, blockchain, and the IoT. He has been appointed at the capacity of a keynote speaker, a technical committee member, and the session chair across the globe. He was awarded eight Best Paper Awards in Springer ICRIC-2019, IEEE-ICIEM-2021, IEEE-ECAI-2021, Springer COMS2-2021, and IEEE-ICIEM-2022. He is a Reviewer of 21 reputed SCI journals, like IEEE Internet of Things Journal, IEEE Transactions on Industrial Informatics, IEEE Transactions on Vehicular Technology, IEEE Journal of Biomedical and Health Informatics, IEEE Access, *IEEE Network*, *ETT* (Wiley), *IJCS* (Wiley), *MTAP* (Springer), *OSN* (Elsevier), *WPC* (Springer), and others.

ASHWIN VERMA received the B.Tech. degree in IT from DAVV, Indore, and the M.Tech. degree in CSE from NIT, Jaipur, in 2013. He is currently pursuing the Ph.D. degree in CSE with Amity University, Jaipur, Rajasthan. He is employed as an Assistant Professor with the Computer Science and Engineering Department, Institute of Technology, Nirma University. He has seven years of teaching and academic experience. He has authored and coauthored more than six articles in leading SCI journals and IEEE conferences. Some of his top findings are published in IEEE Journal of Biomedical and Health Informatics, *JISA* (Elsevier), IEEE-ICIEM-2021, and many more. His research interests include healthcare 4.0, federated learning, blockchain technology, and 5G and beyond communications. He is awarded the Best Research Paper Award in IEEE ICIEM-2021, London, U.K.

VIVEK KUMAR PRASAD received the B.Tech. degree in computer science and engineering from MITS Rayagada, Odisha, the M.Tech. degree in computer science and engineering from the MVJ College of Engineering, Bengaluru, and the Ph.D. degree in cloud computing from Nirma University with the following dissertation title "SLAMMP Framework for Efficient Resource Monitoring and Prediction at an IaaS Cloud." He is currently working as an Assistant Professor with the Computer Science and Engineering Department. He has more than 12 years of teaching experience. He has many publications to his credit in international journals and international conferences. He has written seven book chapters in the SCOPUS-based international edited books. He has been actively involved in the organization of various workshops in the cloud computing domain. His research was supported by funding from the Department of Science and Technology, New Delhi, India. His research interests include distributed computing, cloud computing, machine learning, and artificial intelligence. He is a TPC member and a reviewer for many international conferences across the globe. He is a Lifetime Member of professional societies, like ISTE.

SUDEEP TANWAR (Senior Member, IEEE) received the B.Tech. degree from Kurukshetra University, India, in 2002, the M.Tech. degree (Hons.) from Guru Gobind Singh Indraprastha University, Delhi, India, in 2009, and the Ph.D. degree in wireless sensor network, in 2016. He is currently working as a Full Professor at Nirma University, India. He is also a Visiting Professor with Jan Wyzykowski University, Poland, and the University of Pitesti, Romania. He is also leading the ST Research Laboratory, where group members are working on the latest cutting-edge technologies. He has authored four books and edited 20 books and more than 270 technical articles, including top cited journals and conferences, such as IEEE Transactions on Network Science and Engineering, IEEE Transactions on Vehicular Technology, IEEE Transactions on Industrial Informatics, IEEE Transactions on Green Communications and Networking, IEEE TCSC, IEEE Internet of Things Journal, *IEEE Network*, IEEE Wireless Communications, ICC, IWCMC, GLOBECOM, CITS, and INFOCOM. He initiated the research field of blockchain technology adoption in various verticals, in 2017. His H-index is 54. His research interests include blockchain technology, wireless sensor networks, fog computing, smart grid, and the IoT. He is a member of the Technical Committee on Tactile Internet of IEEE Communication Society. He has been awarded the Best Research Paper Awards from IEEE IWCMC-2021, IEEE ICCCA-2021, IEEE GLOBECOM 2018, IEEE ICC 2019, and Springer ICRIC-2019. He has won the Dr. K. W. Wong Annual Best Paper Prize for 2021 sponsored by Elsevier (publishers of *JISA*). He has served many international conferences as a member of the Organizing Committee, such as the Publication Chair for FTNCT-2020, ICCIC 2020, and WiMob2019, and the General Chair for IC4S 2019, 2020, ICCSDF 2020, and FTNCT 2021. He is also serving the editorial boards for *COMCOM* (Elsevier), *IJCS* (Wiley), *Cyber Security and Applications* (Elsevier), *Frontiers in Blockchain*, and *SPY* (Wiley).

GULSHAN SHARMA received the B.Tech., M.Tech., and Ph.D. degrees. He is currently working as a Senior Lecturer with the Department of Electrical Engineering Technology, University of Johannesburg. His research interests include power system operation and control and application of AI techniques to the power systems. He is working as an Academic Editor of the *International Transactions on Electrical Energy Systems* (Wiley) and a Regional Editor of *Recent Advances in E & EE* (Bentham Science). He is a Y Rated Researcher from NRF South Africa.

PITSHOU N. BOKORO received the M.Phil. degree in electrical engineering from the University of Johannesburg, Johannesburg, South Africa, in 2011, and the Ph.D. degree in electrical engineering from the University of the Witwatersrand, in 2016. He is currently an Associate Professor with the University of Johannesburg. His research interests include modeling and reliability prediction of insulating materials and dielectrics, power quality, and renewable energies. He is a Senior Member of the South African Institute of Electrical Engineers.

RAVI SHARMA is currently working as a Professor with the Centre for Inter-Disciplinary Research and Innovation, University of Petroleum and Energy Studies, Dehradun, India. He is passionate in the field of business analytics and worked in various MNCs as a leader of various software development groups. He has contributed various articles in the area of business analytics, prototype building for startup, and artificial intelligence. He is leading academic institutions as a consultant to uplift research activities in inter-disciplinary domains.