Intelligent Systems Reference Library 232

Mayuri Mehta
Vasile Palade
Indranath Chatterjee *Editors*

Explainable AI: Foundations, Methodologies and Applications

Springer

Intelligent Systems Reference Library

Volume 232

The aim of this series is to publish a Reference Library, including novel advances and developments in all aspects of Intelligent Systems in an easily accessible and well structured form. The series includes reference works, handbooks, compendia, textbooks, well-structured monographs, dictionaries, and encyclopedias. It contains well integrated knowledge and current information in the field of Intelligent Systems. The series covers the theory, applications, and design methods of Intelligent Systems. Virtually all disciplines such as engineering, computer science, avionics, business, e-commerce, environment, healthcare, physics and life science are included. The list of topics spans all the areas of modern intelligent systems such as: Ambient intelligence, Computational intelligence, Social intelligence, Computational neuroscience, Artificial life, Virtual society, Cognitive systems, DNA and immunity-based systems, e-Learning and teaching, Human-centred computing and Machine ethics, Intelligent control, Intelligent data analysis, Knowledge-based paradigms, Knowledge management, Intelligent agents, Intelligent decision making, Intelligent network security, Interactive entertainment, Learning paradigms, Recommender systems, Robotics and Mechatronics including human-machine teaming, Self-organizing and adaptive systems, Soft computing including Neural systems, Fuzzy systems, Evolutionary computing and the Fusion of these paradigms, Perception and Vision, Web intelligence and Multimedia.

Indexed by SCOPUS, DBLP, zbMATH, SCImago.

All books published in the series are submitted for consideration in Web of Science.

Mayuri Mehta · Vasile Palade · Indranath Chatterjee
Editors

Explainable AI: Foundations, Methodologies and Applications

Springer

Editors
Mayuri Mehta
Department of Computer Engineering
Sarvajanik College of Engineering
and Technology
Surat, Gujarat, India

Vasile Palade
Centre for Computational Science
and Mathematical Modelling
Coventry University
Coventry, UK

Indranath Chatterjee
Department of Computer Engineering
Tongmyong University
Busan, Korea (Republic of)

ISSN 1868-4394 ISSN 1868-4408 (electronic)
Intelligent Systems Reference Library
ISBN 978-3-031-12806-6 ISBN 978-3-031-12807-3 (eBook)
https://doi.org/10.1007/978-3-031-12807-3

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Artificial Intelligence (AI) has brought about a revolution in many real-world sectors and has become an integral part of our everyday lives. While AI-enabled systems are undoubtedly benefiting real-world sectors, there is still a risk in blindly trusting the recommendations, insights, or predictions provided by them. Many of these systems are often complex and opaque. They operate as a black box, meaning that users do not understand how decisions are being made by such systems. Thus, the key limitation of today's intelligent systems is their inability to explain their decisions and actions to human users. This issue is especially important for risk-sensitive applications, such as security, clinical decision support, or autonomous driving. A lack of explainability hampers our capacity to fully trust AI systems.

It is for this reason that AI techniques need to have explanatory capabilities, for users to understand why certain decisions are made. The methods developed to provide such capabilities have come to be known as explainable AI (XAI). Explainable AI contrasts with the so-called 'black box' machine learning. XAI helps present decisions being made with additional information about how and why the AI system arrived at a particular decision, including an interface to explain which features influenced its decision.

In this book, readers will learn about Explainable AI, including what it is, what the fundamentals of this area are, why it is needed, and how it is to be developed. Explainable AI offers a way to make decision-making more transparent and trustworthy. In other words, XAI aims to remove the so-called black box from the AI models being developed and explain the model decisions in an understandable form. It refers to an AI system's capacity to explain the logic behind its action to a human person. It can take two forms: explaining it to a computer scientist in a specialized language or explaining it to the system user in a human understandable form. It is critical because it is intimately related to human confidence in the AI system's usage and, more formally, whether that faith is well-placed by verifying things about the machine's behavior.

This book covers concepts related to model transparency, interpretable machine learning and explanations, various methods for Explainable AI, evaluation methods and metrics for XAI, ethical, legal, and social issues related to AI and XAI, as well as a range of applications and examples of XAI in different real-life sectors, such as healthcare, autonomous driving, and law enforcement. The editors are thankful to the authors who submitted their research work to this book, as well as to all the anonymous reviewers for their insightful remarks and significant suggestions that helped enhance the quality of this book. We hope that readers will find the book useful.

Surat, India	Mayuri Mehta
Coventry, UK	Vasile Palade
Busan, Korea (Republic of)	Indranath Chatterjee
June 2022	

Contents

Contributors

Aluvalu Rajanikanth CBIT, Hyderabad, India

Apostolopoulou Eleni School of Science, Technology and Health, York St John University, York, UK

Banerjee Puja Academy of Scientific and Innovative Research, Ghaziabad, India

Barnwal Rajesh P. AI & IoT Lab, IT Group, CSIR-Central Mechanical Engineering Research Institute, Durgapur, India

Bhattacharya Tanmay Department of IT, Techno Main, Kolkata, India

Brcic Luka Medical University Graz, Graz, Austria

Carvalho Rita Medical University Graz, Graz, Austria

Chennam Krishna Keerthi Vasavi College of Engineering, Hyderabad, India

Chopra Mayank Department of Computer Science and Informatics, Central University of Himachal Pradesh (H.P.), Dharamshala, India

Deshmukh Rashmi Department of Technology, Shivaji University, Kolhapur, India

Evans Theodore Medical University Graz, Graz, Austria

Eyo Eyo Umo Faculty of Environment and Technology, Civil Engineering Cluster, University of the West of England, Bristol, UK

Geißler Christian Medical University Graz, Graz, Austria

Goel Amit Kumar Delhi, India

Holzinger Andreas Medical University Graz, Graz, Austria

Hore Sirshendu Department of CSE, Hooghly Engineering and Technology College, Pipulpati, Hooghly, West Bengal, India

Jansen Christoph Medical University Graz, Graz, Austria

Kanarachos Stratis Faculty of Engineering and Computing, Coventry University, Coventry, UK

Kargl Michaela Medical University Graz, Graz, Austria

Klyushin D. A. Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Kumar Ajay Department of Computer Science and Informatics, Central University of Himachal Pradesh (H.P.), Dharamshala, India

Kumar Dheeraj Department of Information Technology, Parul Institute of Engineering and Technology, Parul University, Vadodara, India

Lu Yang School of Science, Technology and Health, York St John University, York, UK

Maheswari V. Uma KG Reddy College of Engineering, Hyderabad, India

Mehta Mayuri A. Department of Computer Engineering, Sarvajanik College of Engineering and Technology, Sarvajanik University, Surat, India

Mudrakola Swapna Vasavi College of Engineering, Hyderabad, India;
Matrusri Engineering College, Hyderabad, India

Müller Heimo Medical University Graz, Graz, Austria

O. Chergykalo D. Faculty of Computer Science and Cybernetics, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

Onyekpe Uche School of Science, Technology and Health, York St John University, York, UK;
Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry, UK

Palade Vasile Centre for Computational Science and Mathematical Modelling, Coventry University, Coventry, UK

Plass Markus Medical University Graz, Graz, Austria

Rao K. Gangadhara CBIT, Hyderabad, India

Regitnig Peter Medical University Graz, Graz, Austria

Sarkar Arjun Leibniz Institute for Natural Product Research and Infection Biology, Hans Knöll Institute, Jena, Germany

Sarnayak Anshul Department of Technology, Shivaji University, Kolhapur, Maharashtra, India

Sharma Mukta Delhi, India

Shringare Yash Department of Technology, Shivaji University, Kolhapur, Maharashtra, India

Singhal Priyank Moradabad, India

Zerbe Norman Medical University Graz, Graz, Austria

Abbreviations

AAR	After-Action Review
ABS	Anti-lock Braking System
AdaBoost	Adaptive Boosting
AEPS	Average Error Per Second
AI	Artificial Intelligence
ANFIS	Adaptive Neuro Fuzzy Inference System
ANN	Artificial Neural Network
ANOVA	Analysis of Variance
API	Application Programming Interface
Ar	Anger
ARDI	Actors, Resources, Dynamics, and Interactions
ASV	Asymmetric Shapley Values
AV	Autonomous Vehicle
Bm	Boredom
CAM	Class Activation Map
CBIR	Content Based Image Retrieval
CBR	Case-Based Reasoning
CDSS	Clinical Decision Support System
CEM	Contrastive Explanation Methods
CGLES	Common Ground Learning and Explanation System
Cm	Clam
CNN	Convolutional Neural Network
CNV	Choroidal NeoVascularization
CoD	EMODB + RAVDESS
COVID-19	Coronavirus Disease 2019
CRSE	Cumulative Root Squared Error
CSL	Classical
CT	Computed Tomography
CT Scan	Computed Tomography Scan
DARPA	Defense Advanced Research Projects Agency
DC	Direct Current

DeepLIFT	Deep Learning Important FeaTures
DeepSHAP	Deep Shapley Additive Explanations
DICOM	Digital Imaging and Communications in Medicine (standard)
DL	Deep Learning
DME	Diabetic Macular Edema
DNN	Deep Neural Networks
DR	Diabetic Retinopathy
Dt	Disgust
EC	European Commission
ECU	Electronic Control Unit
EG	Expressive Gradients
EHR	Electronic Health Record
ELI5	Explain Like I'm 5
EU	European Union
FAST	Fourier Amplitude Sensitivity Test
FCNN	Fully Convolutional Neural Networks
FDA	United States Food and Drug Administration
FFPE	Formalin-Fixed Paraffin-Embedded
FIS	Fuzzy Inference Systems
FMEA	Failure Mode and Effect Analysis
Fr	Fear
GBP	Guided BackPropagation
GLM	Generalized Linear Rule Models
GNN	Graph Neural Network
GNSS	Global Navigation Satellite System
GPIO	General Purpose Input/Output
Grad-CAM	Gradient weighted Class Activation Mapping
GSM	Global System for Mobile communication
GUI	Graphical User Interface
HAII	Human-AI Interaction
HCI	Human-Computer Interaction
HD	High Definition
HOG	Histogram of Oriented Gradients
HTML	HyperText Markup Language
HTTP	Hyper Text Transfer Protocol
Hy	Happy
I/O	Input-Output
ICE	Individual Conditional Expectation
IDNN	Input Delay Neural Network
IDS	Intrusion Detection Systems
IG	Integrated Gradient
IMP	Important
IMU	Inertial Measurement Unit
IoT	Internet of Things
IO-VNBD	Inertial Odometry Vehicle Navigation Benchmark Dataset

ISO	International Organization for Standardization
IT	Information Technology
IVDR	European In-Vitro Devices Regulation
KNN	K-Nearest Neighbors
LAN	Local Area Network
LIME	Local Interpretable Model-Agnostic Explanations
LIMS	Laboratory Information System
LRP	Layer-wise Relevance Propagation
LSTM	Long Short-Term Memory
LT	Latest
MDR	European Medical Devices Regulation
MFNN	Multi Feedforward Neural Network
MIABIS	Minimum Information About Biobank Data Sharing (standard)
MIS	Minimally Invasive Surgery
ML	Machine Learning
MLP	Multi-Layer Perceptron
MRI	Magnetic Resonance Imaging
NCD	Non-Communicable Diseases
NIDS	Network-based Intrusion Detection System
Nl	Neutral
NLP	Natural Language Processing
NN	Neural network
NumPy	Numerical Python
OAT	One-Step-At-A-Time
OCT	Optical Coherence Tomography
OEM	Original Equipment Manufacturer
OpenCV	Open-Source Computer Vision Library
PCA	Principal Component Analysis
PD	Partial Dependence
PDP	Partial Dependence Plots
Perm	Permutation
PIMP	Permutation Importance
PRRC	Person Responsible for Regulatory Compliance
QII	Quantitative Input Influence
RBD-FAST	Random Balance Designs-Fourier Amplitude Sensitivity
RBFNN	Radial Basis Function Neural Network
RBIA	Risk-based Internal Auditor
R-CNN	Region-Based Convolutional Neural Networks
RELU	Rectified Linear Unit
RGB	Red Green Blue
RNN	Recurrent Neural Network
ROI	Region of Interest
RT-PCR	Reverse Transcription-Polymerase Chain Reaction
sci-fi	Science fiction
SCS	System Causability Scale

Sd	Sad
SHAP	SHapley Additive exPlanations
SIM	Subscriber Identification Module
SQL	Structured Query Language
Su	Surprise
SUS	System Usability Scale
SVM	Support Vector Machine
TCP	Transmission Control Protocol
TED	Teaching Explanations for Decisions
t-SNE	t-Distributed Stochastic Neighbor Embedding
UAV	Unmanned Aerial Vehicle
UI	User Interface
UI / UX	User Interface/User Experience
WhONet	Wheel Odometry neural Network
WKNN	Weighted K-Nearest Neighbors
WSI	Whole Slide Image
XAI	Explainable Artificial Intelligence
YOLO	You Only Look Once

Chapter 1
Black Box Models for eXplainable Artificial Intelligence

Krishna Keerthi Chennam, Swapna Mudrakola, V. Uma Maheswari, Rajanikanth Aluvalu, and K. Gangadhara Rao

Abstract Machine learning algorithms are becoming popular nowadays in cyber security applications like Intrusion Detection Systems (IDS). Most of these models are anticipated as a Black Box. Previously black box was a model where the user cannot see the internal logic. To reach the goal of overwhelming the crucial weakness, the cost may vary. This is related to both ethical and practical problems. Explainable Artificial Intelligence (XAI) is crucial to converting the machine learning algorithms to appreciate the management by accepting the human experts to understand the data evidence. Important role of trust management is to accept the impact of malicious data to identify the intrusions. This chapter addresses the XAI method to appreciate trust management using the decision tree models. Basic decision tree models are used to simulate a human contact to decision making by dividing the options into multiple small options for the IDS area. This chapter aims to implement the arrangement of issues labeled in the various black box methods. This survey helps the researcher to understand the classification of various black box models.

Keywords Black box · Cyber security · Decision trees · Intrusion detection system · Artificial intelligence

K. K. Chennam (✉) · S. Mudrakola
Vasavi College of Engineering, Hyderabad, India
e-mail: krishnakeerthich@gmail.com

S. Mudrakola
Matrusri Engineering College, Hyderabad, India

V. U. Maheswari
KG Reddy College of Engineering, Hyderabad, India

R. Aluvalu · K. G. Rao
CBIT, Hyderabad, India

M. Mehta et al. (eds.), *Explainable AI: Foundations, Methodologies and Applications*, Intelligent Systems Reference Library 232,
https://doi.org/10.1007/978-3-031-12807-3_1

1.1 Introduction to Machine Learning

There was a huge increase in artificial intelligence (AI) in a glimpse. Machine learning is a subset of AI. The main importance of machine learning is identifying the structure of data or format suitable data models used by the users. However, Machine learning is related to computer science and varies from former computational methods. Previously, Algorithms were written exclusively programmed instructions for computers to solve problems. Now machine learning (Othman et al. 2018) algorithms are used to educate the computers on data inputs and data statistics, analysis is used to produce output values within a range. Automatically decision is taken based on the sample data with the help of models and inputs. Many technologies are using machine learning (Gilpin et al. 2018) algorithms and get benefited. Facial recognition is one of the technologies which permit social media platforms like Facebook and Instagram's to help the users tag and share friends' photos (Logas et al. 2022). Movies or television shows using optical character recognition technology help to change images to text into movable (Jiang et al. 2022). Self-driving cars also depend on machine learning to map the routes (Saha and De 2022). Machine learning is consistently improving technology, which requires continuously improving methodologies for analyzing may affect the machine learning process (Pazzani et al. 2001). Supervised and unsupervised learning are two basic machine learning methods. Along with these two methods k-nearest neighbor algorithm, decision tree learning methods and deep learning are other important concepts in machine learning.

Firstly, supervised learning purpose is to learn by similar outputs by identifying errors and changing the models depending on the output (Cai et al. 2022). This model also uses the patterns to identify the labeled values and unlabeled data also. Supervised learning algorithms will make sure to identify the images and produce labels to the particular image by seeing the cat image, supervised learning will be able to identify and label it as an animal. Unsupervised learning is to identify the secret patterns in the data and automatically identify the classification of raw data. This is used for transactional data and complex data is more expansive and unrelated to organize properly (Kotenko et al. 2022). Example like unsupervised learning will be able to tag all cat images and group it.

Machine learning is based on statistics with basic knowledge by understanding and supporting machine learning algorithms. Correlation is used to identify the relation among two dependent or independent variables. Regression was used for identifying the relation among dependent and independent variable. When an independent variable is given and needs to identify the dependent variable, the regression statistics used to identify it is called regression enables prediction capabilities. To identify the pattern k-nearest neighbor algorithm is used for regression and classification. Small and positive integer is k value. Example of separating the square and circle shapes into two different classes, this classification is used.

Decision tree is a predictive algorithm based on the models, observations, analysis and gives target data values. This model is created to predict the target based input values. The data attributes identified based on the observation are branches the conclusion of data target values is nothing but leaves. Deep learning is introduced based on neural networks with multiple layers in artificial neural networks based on hardware. The output is connected to an input to the next layer in the deep learning process. Computer vision and speech recognition have realized significant advances in deep learning approaches (Li et al. 2022). Humans can give biased decisions that lead to negative results, machine learning helps to overcome such issues and give unbiased decisions. Black box (Guo 2020; Perarasi et al. 2020a) systems exploit sophisticated machine learning models to identify separated secure data. Medical status, risk of insurances, eligibility score for credit cards acknowledge using machine learning algorithms construct predictive models and map the features into class in the learning phase (Svenmarck et al. 2018). The learning process is formed by the digital trances that are left after operating daily activities like social media activities, purchases, etc. Huge data may handle human biases and prejudices. Decision models are accomplished by inheriting biases, wrong decisions and illegal activities. Various scientific communities studied the issues of discussing machine learning decision models. Even though illustratable machine learning is the important case and accepted newly considering the situation, many ad-hoc distributed results.

The rest of the chapter is organized as follows. The First section discusses the importance of cyber security in XAI. Next section discusses Deep learning using XAI which follows the Intrusion Detection System (IDS). Section 1.5 is about applications of cyber security in XAI. Section 1.6 discusses the comparison of XAI using black box methods and finally about the conclusion.

1.1.1 Motivation

The unique aim of the chapter is to reach the novelty in research work using machine learning. AI understands different technologies under the same umbrella like machine learning to predict the results. Machine learning ultimately reaches the goal to reach for accurate results with training the model.

1.1.2 Scope of the Paper

Machine learning is one of the best options in career applications for smart systems to handle business attacks. Target is to calculate human intelligence and be able to make decisions more precisely under any situation. AI handles the different technologies that come under the same domain like pattern recognition, big data, machine learning, artificial intelligence and various other technologies. This is the reason AI is having much future scope in many applications.

1.2 Importance of Cyber Security in eXplainable Artificial Intelligence

Industries progressively improved with a better complex cyber security (Pienta et al. 2020) ecosystem depending on various types like users, technology and processes to functional roles. Cyber security is dependent on relations between users and groups, users, organizations and technology, technology and users. From the above trusting peers, cyber security prevents separately to defend against cyber attacks. AI models cite the knowledge from the gathered data. Actually, no human will believe the AI system for the possible and desirable quality of data, difficult methods and accountability, trained AI engineer. AI is trust related software that gives solutions to cyber-attacks. You may ask how to trust the AI models in cyber security, which are developed based on data analysis and predict the solutions from the data. The simple answer for this question is that XAI (Guo 2020; Arrieta et al. 2020) will justify reliability, ability, and trustworthiness. Main challenge for AI is the inability to understand and compare between transition models. A simple example is Autonomous vehicles (Perarasi et al. 2020b). Trustworthy AI should explain its decisions to allow the human expert to understand the underlying data evidence and causal reasoning.

Complex black box models study from machine learning and deep learning parameters. Based on the black boxes models, AI engineers identify direct models to make decisions and identify the behavior of models. Cyber security is liable for attacks and targets the trusted security in critical systems. Therefore XAI from AI plays an important role in developing the solution based AI with interpretability. Interpretability further assures uniformly in decision-making to detect the imbalanced dataset. Interpretability strengthens the powerful solution based AI using highlighting hidden could change the prediction. The decision tree model is developed based on the Intrusion detection system attacks (Svenmarck et al. 2018; Stampar and Fertalj 2015). The intrusion detection system developed fast in study and organization research in exchange for increasing cyber attacks on government and commercial enterprises internationally and action on cost is increased consistently (Lee et al. 2001). The main harmful cyber crimes are from vicious associates, denial of servers, web attacks, and organizations may lose the intellectual property related to vicious attacks in the system. Organizations install various firewalls, software like antivirus and intrusion detection systems against those attacks. Intrusion detection is a crucial role in cyber security, grants to determine vicious network activities previously compromises data connection, availability and opportunity. It is a method to identify security breaches by interrogating models in the data system.

Day-to-day, the digital system is adopted by the world. The network access leads to a lack of security issues that the Internet of Things devices (Lee et al. 2001; Chennam et al. 2022). Intrusion attacks with high possibilities on Internet of Things devices connected to the internet lead to network devices safely from intrusion. An IDS was developed to avoid important data from vicious acts. Important data with network access needs to be permanently protected from all pursuit to consume, expose, alter, disable, steal or gain unauthorized access. Traditional intrusion detection systems, mainly signature-based, identify only popular attacks and may not identify new attacks. Machine learning is the best approach which is exclusively developed to maintain detection accuracy.

Artificial Intelligence (AI) has helped all the industries with effective results in deploying various applications to monitoring, Decision Making, Solving Complex problems, creative approaches, observation analysis, Language Recognition and Learning. Artificial intelligence has collaborated with additional technology like Machine Learning, Neural networks and Deep Learning. Artificial intelligence is used to compute the programs and prepare the system to behave like a human brain (Uma Maheswari et al. 2021; Deshpande et al. 2020). The AI has excelled in thinking, retrieving and taking decisions sometimes faster than the human brain. AI applications are used in medical Care, Teaching and Learning, Law, Commerce and public Departments etc. The above applications are intended to say that algorithms rule the world by AI, which is inevitable (Swapna et al. 2022).

XAI advantages are mainly concerned with ethics and continuous improvements. XAI required enough trust to handle the AI. For decades various AI models gave biased results or not perfect results which lead to ensuring the safety in AI decisions without any faults. To justify the final decisions taken by the AI required logical reasoning in decision making. AI helps to identify the malware weekly updated and all possibilities of pattern recognition, behavioral attacks of ransomware able to identify before entering into the system. Bots help to clear maximum chunks in internet networks. Stolen login details can create false account details, tampering data; bots can be the correct menace. Handling automatic threats is not possible alone. AI and machine learning will heal to construct the good bots to identify the engine crawlers, bad bots etc. AI starts to identify the data and accepts to provide cybersecurity to understand the strategy consistently.

1.2.1 Importance of Trustworthiness

The importance of Trustworthiness is an essential aspect to measure the safety, performance and reliability. The qualities requirements to say as trustworthy are the system must be accountable, fair, reliable behavior, reasonable and acceptable. The author Stephen Hawking says "AI can spread faster and can be violent if it is not controlled properly". The AI systems need to be authorized and validated in each design and implementation phase. The AI systems need to be authorized and validated in each phase of the design and implementation. There are different algorithms used to predict the risk of the system, and it occur due to low-quality data training, narrow perception of the problem, technical issue management etc. can lead to unrecoverable loss of people, properties and loss the trust on AI practices. AI applications are used in important applications like facial recognition software, Tagging picture in television media, Health care practices and self-driving car are the high-risk applications, wrong decision may cause life. The author Davinder Kaur has raised some questions

Table 1.1 Questioner table states the importance of Trustworthy in AI

Research questionnaire	Proposed solution
Purpose of proving AI is trustworthy	Decisions taken by the AI system should be ethical practice, robust in nature, Lawful and acceptable
What protocols are used to work AI systems?	We can empower and help to maintain the AI system lawful practices
Why human control involved	AI systems need to collaborate with human intervention and machines in cognitive decision making
Reasons for AI acceptable	AI systems have proven to be trustworthy, fast and usable

to understand the requirements needed to conclude the AI system as worthy (Kaur et al. 2022) (Table 1.1).

The Black Box Model uses AI methods, the results are obtained, but its design will not help to justify the result. The explanations are required to extract the output function. We need to apply some techniques to find the reason to conclude (Zhang et al. 2022). The Post-Hoc Explainable is a reverse engineering process that starts to reach the initial state from the destination. Explainable algorithms like Support Vector Machine (SVM), Multi-Layer Neural Network, Convolution Neural Network and Recurrent Neural Network (Hermansa et al. 2022). XAI uses machine learning techniques to justify the results. The reasonable techniques are explained by simplifying the problem, Feature Connectivity, Local Reasoning, Visible Reasoning and Multi Classifier. The importance of AI is used to make better decisions, explain deep learning, Model Debugging, and build the latest model (Brito et al. 2022) (Fig. 1.1).

Machine Learning (ML) methods Contribution for XAI—The Machine Learning method works for limited data. The ML required defined features to the drive result. The complex problems will simplify and solve phase-wise, network designs are

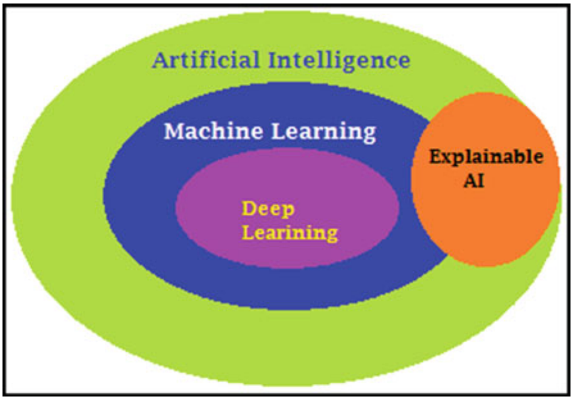

Fig. 1.1 Representation of AI, DL, ML, XAI Association