IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Universal AI : An Interface for Implementing Explainable Artificial Intelligence Techniques on Diverse Datasets

Shruti Rasne[1], Kajal Lokhande[2], Nimisha Jadhav[3],
Computer Engineering,
Keystone School of Engineering, Pune, India

Abstract: In the rapidly evolving landscape of artificial intelligence (AI), achieving transparency and interpretability in machine learning models has become paramount. This paper introduces UniversalAI, an innovative and adaptive interface designed to facilitate the seamless implementation of explainable artificial intelligence (XAI) techniques across a spectrum of diverse datasets. As AI systems are increasingly integrated into critical decision-making processes, the demand for model interpretability and trustworthiness has grown significantly. UniversalAI responds to this imperative by providing a comprehensive solution that transcends the limitations of existing XAI interfaces.

Motivated by the need for a flexible and universally applicable tool, UniversalAI is conceived to accommodate the intricacies associated with various types of datasets, ranging from structured to unstructured and heterogeneous data. The interface's architecture is meticulously designed to provide users with an intuitive and customizable platform, ensuring that XAI techniques can be effectively applied irrespective of the dataset's inherent complexities.

Index Terms - Interpretability, Explainablity, Robustness, User-Friendly Interface, Transparency

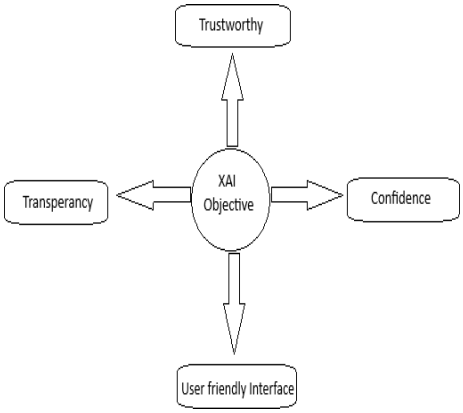
I. INTRODUCTION

The growing significance of explainability in AI systems stems from an increasing awareness of the ethical, societal, and practical implications associated with the deployment of sophisticated machine learning models. As AI applications become integral to diverse domains, ranging from healthcare and finance to autonomous systems, the demand for transparency and interpretability in the decision making processes of these systems has become more pronounced.

In ethical terms, the use of AI raises concerns about accountability and the alignment of algorithmic decisions with human values. Explainability is crucial to ensure that the reasoning behind AI-generated outcomes is understandable to individuals affected by these decisions. This emphasis on ethical considerations is particularly pertinent in applications where AI plays a role in critical domains, such as healthcare diagnoses or financial transactions.

Building user trust is another compelling reason for prioritizing explainability in AI systems. Users are more likely to adopt and rely on AI technologies when they can comprehend the rationale behind the decisions made by the algorithms. This trust is pivotal, especially in contexts where AI influences sensitive areas of human life. Addressing concerns related to bias and promoting fairness in AI systems is also a key aspect of the increasing emphasis on explainability. The opaqueness of certain AI models can contribute to

biased outcomes. Explainable AI tools enable stakeholders to identify and rectify biases, ensuring fair and equitable decision-making. Explainability is not only crucial for understanding and validating AI decisions but also for improving and refining models. When AI systems produce unexpected or incorrect results, having insights into the model's decision-making process facilitates debugging and enhancement. This iterative process contributes to the ongoing development of more accurate and reliable AI models. Furthermore, as AI becomes increasingly integrated into society, there is a growing need for social acceptance and understanding of AI technologies. Explainable AI plays a role in demystifying complex algorithms, making them more accessible to a broader audience. This, in turn, contributes to a positive perception of AI applications and facilitates broader societal acceptance.



objectives of ai

II. CHALLENGES:

Implementing Explainable Artificial Intelligence(XAI) techniques across diverse datasets introduces a set of intricate challenges rooted in the multifaceted nature of data. The varied sources, structures, and characteristics of datasets contribute to the complexities faced by researchers and practitioners seeking to enhance transparency and interpretability in AI systems.

One of the primary challenges lies in the heterogeneity of data types. Datasets often comprise a mix of structured, unstructured, and semi-structured data, necessitating adaptable XAI techniques capable of handling diverse formats. The challenge extends to interpreting relationships and contributions within high-dimensional feature spaces, common in datasets like images or genomics.

Temporal dynamics present another layer of complexity. Datasets with time-series components demand specialized approaches to understand how XAI techniques operate over time, capturing the evolution of patterns and relationships effectively.

Imbalanced datasets, where certain classes are underrepresented, pose challenges in ensuring that XAI explanations are not biased towards the majority class. Striking a balance and accurately representing the nuances of minority classes require careful consideration.

The availability of labeled data is crucial for XAI, yet it is not always abundant. Lack of ground truth or uncertainty in labeling can compromise the reliability of explanations, particularly when assessing the accuracy of interpretability results.

Domain-specific challenges further complicate the implementation of XAI techniques. Different fields, such as healthcare or finance, have unique characteristics and intricacies. Adapting XAI to address these domain-specific challenges requires a nuanced understanding of the specificities within each field.

Capturing interactions between features is another hurdle. In datasets with non-linear relationships or intricate feature dependencies, identifying and interpreting these interactions becomes challenging, yet it is crucial for understanding the decision making process of complex models.

Scalability is a persistent concern as datasets grow in size. XAI techniques must be computationally efficient and scalable to handle large volumes of data without compromising the quality of explanations.The presence

of noise or outliers in the dataset can impact the robustness of XAI models. Ensuring resilience to noisy data and outliers is vital to maintaining the accuracy of explanations.

Additionally, the subjectivity of user interpretationadds another layer of complexity. Different users mayhave diverse perspectives and expectations regarding interpretability. Adapting XAI techniques to align with varied user needs and domain knowledge introduces challenges in defining interpretability that satisfies a broad spectrum of users.

Addressing these challenges requires a holistic approach that combines expertise in machine learning,domain-specific knowledge, and continuousinnovation. Researchers and practitioners must navigate these complexities to develop XAI techniques that are adaptable, robust, and capable of providing meaningful insights across a diverse range of datasets.

III. MOTIVATION

The motivation behind the development of UniversalAI is grounded in the growing significance of artificial intelligence (AI) across various industriesand applications. As AI systems become integral to decision-making processes in critical domains such as healthcare, finance, and autonomous systems, the demand for transparency, interpretability, and accountability has intensified. UniversalAI emerges asa response to this imperative, driven by a commitmentto providing a comprehensive solution that addresses the ethical, societal, and practical challengesassociated with AI.

One fundamental motivation is the rising importance of AI technologies in shaping and influencing human experiences. As AI becomes more embedded in our daily lives, there is a compelling need to understand how these systems make decisions. UniversalAI is motivated by the recognition that transparency in AI models is not only a technical requirement but also a societal necessity to ensure responsible and ethical use.

Ethical considerations play a pivotal role in motivatingthe development of UniversalAI. As AI applications impact various aspects of human life, there is a heightened awareness of the ethical implications of algorithmic decision-making. The interface seeks to address these ethical considerations by providing a tool that enhances the interpretability of AI models, fostering accountability, and promoting ethical use.

Building user trust is another key motivation. Trust is a critical factor influencing the acceptance and adoption of AI technologies. Users are more likely to trust and embrace AI systems when they can comprehend and trust the decision-making processes. UniversalAI aims to build this trust by offering a flexible interface that makes AI models more understandable and accessible to a broader audience.

Moreover, UniversalAI is motivated by the real worldimpact of AI on various applications. In domains such as healthcare and finance, where AI decisions can have direct consequences on individuals' lives, the need for interpretable models is not just a technical necessity but a means to empower users with insights for more informed decision making.

Emphasizing the need for a flexible interface underscores the acknowledgment of the diverse natureof datasets and the dynamic landscapes in which AI systems operate. UniversalAI recognizes that datasets vary in types, structures, and characteristics. The interface is designed to be adaptable, accommodating different data formats, and ensuring that users can apply Eplainable AI (XAI) techniques seamlessly across a broad spectrum of data types.

The flexibility of the interface is also driven by the dynamic nature of data. Datasets evolve over time, anda flexible interface must be capable of adapting to

Moreover, UniversalAI is motivated by the real worldimpact of AI on various applications. In domains such as healthcare and finance, where AI decisions can have direct consequences on individuals' lives, the need for interpretable models is not just a technical necessity but a means to empower users with insights for more informed decision making.

Emphasizing the need for a flexible interface underscores the acknowledgment of the diverse natureof datasets and the dynamic landscapes in which AI systems operate. UniversalAI recognizes that datasets

vary in types, structures, and characteristics. The interface is designed to be adaptable, accommodating different data formats, and ensuring that users can apply Eplainable AI (XAI) techniques seamlessly across a broad spectrum of data types.

The flexibility of the interface is also driven by the dynamic nature of data. Datasets evolve over time, and a flexible interface must be capable of adapting to changing data landscapes, accommodating new features, and adjusting to variations in data distributions. UniversalAI provides a platform thatremains effective across different iterations of datasets, ensuring its relevance in dynamic data environments.

Furthermore, the need for a flexible interface is highlighted by the fact that different users may have distinct requirements and preferences when it comes to implementing XAI techniques. A one-size-fits-all solution is not sufficient. UniversalAI acknowledges this by offering a customizable tool that caters todiverse user requirements, allowing users to tailor theapplication of XAI methods based on their specific needs.

IV. LITERATURE SURVEY

Paper 1: Explainable Artificial Intelligence (XAI):Concepts and Challenges in Healthcare
Author: Tim Hulsen
Publication year: 10 August 2023

Artificial Intelligence (AI) describes computersystems able to perform tasks that normally requirehuman intelligence, such as visual perception, speech recognition, decision-making, and language translation. Examples of AI techniques are machinelearning, neural networks, and deep learning. AI can be applied in many different areas, such aseconometrics, biometry, e-commerce, and the automotive industry. In recent years, AI has found its way into healthcare as well, helping doctorsmake better decisions ("clinical decision support"), localizing tumors in magnetic resonance images, reading and analyzing reports written by radiologists and pathologists, and much more However, AI has one big risk: it can be perceived as a "black box", limiting trust in its reliability, whichis a very big issue in an area in which a decision canmean life or death. As a result, the term Explainable Artificial Intelligence (XAI) has been gaining momentum. XAI tries to ensure that AI algorithms (and the resulting decisions) can be understood by humans.

Artificial Intelligence (AI) encompasses computer systems capable of human-like tasks, from visual perception to language translation. Techniques likemachine learning and neural networks drive AI applications in diverse fields, including healthcare. However, AI's "black box" nature raises concerns about trust, especially in critical decision-making contexts. The concept of Explainable Artificial Intelligence (XAI) addresses this issue by making AI algorithms and decisions understandable to humans. This narrative review explores central XAI concepts, outlines challenges in its application to healthcare, and assesses its potential to enhanceunderstanding and trust in the field. The paper also discusses alternatives for building trust in AI and suggests future research directions in XAI. Keywords: XAI, AI, artificial intelligence, explainable, explainability, machine learning, deep learning, data science, big data, healthcare, medicine.

Paper 2: Fundamental Misconceptions in Current XAI Research
Author: Timo Freiesleben and Gunnar Konig
Publication year: 7 June 2023

Despite progress in the field, significant parts of current XAI research are still not on solid conceptual, ethical, or methodological grounds. Unfortunately, these unfounded parts are not on thedecline but continue to grow. Many explanationtechniques are still proposed with out clarifyingtheir purpose. Instead, they are advertised with evermore fancy-looking heatmaps or only seemingly relevant benchmarks. Moreover, explanation techniques are motivated with questionable goals, such as building trust, or rely on strong assumptionsabout the 'concepts' that deep learning algorithms learn. In this paper, we highlight and discuss these and other misconceptions in current XAI research. We also suggest steps to make XAI a more substantive area of research. Furthermore, these techniques are sometimes justified by questionable goals like building trust or relying on assumptions about the 'concepts' deep learning algorithms learn.This paper aims to highlight and

discuss these misconceptions in current XAI research while proposing steps to establish XAI as a more substative research.

Paper 3: Implementation of explainable artificial intelligence in commercial communication systemsusing micro systems
Author: Hariprasath Manoharan, Teekaraman Yuvaraja,Ramya Kuppusamy and Arun Radhakrishnan
Publication year: 2023

In this study, we explore the integration of Explainable Artificial Intelligence (XAI) in commercial communication systems, specificallyfocusing on the banking and finance sector (BFS). Acknowledging the virtues of Artificial Intelligence(AI) and decision-making systems, we address challenges in application changes by implementingXAI design systems. Additionally, we tackle the issue of the substantial size of standard sensingmodels by introducing micro electro-mechanicalsystems (MEMS). The developed model undergoes testing with five scenarios, including multiple parametric arrangements, incorporating interpretability processes. Comparative analysis with existing models reveals that the integration of XAI and MEMS yields significant improvements, achieving an impressive average transparency of 96% and mitigating data impairments in BFS applications.
The collaborative symphony emanates from diverse minds converging, hailing from esteemed institutions across India, Kuwait, and Ethiopia. Theresult is a holistic solution, a harmonious marriage of XAI and MEMS poised to revolutionize BFS applications. Beyond mere functionality, our integration aspires to elevate transparency, allay data security concerns through conviction management, and introduce avant-garde mathematical designs. In this crucible of examination, our model doesn't just compete; itexcels. Comparative analysis against existing models unequivocally highlights the transformative power of the harmonious integration of XAI and MEMS. The results are not just commendable; they're extraordinary, revealing an impressive average transparency of 96 percent.
Beyond therealm of system clarity, our model emerges as a beacon mitigating data impairments within BFS applications, signaling a monumental shift in the trajectory of technological solutions for the financial sector. This isn't just research; it's a manifesto for a new era of adaptive, technologically-driven excellence in the BFS landscape.

Paper 4: Explainable Artificial Intelligence of Multi-Level Stacking Ensemble for Detection of Alzheimer's Disease Based on Particle Swarm Optimization and the Sub-Scores of Cognitive Biomarkers
Author: ABDULAZIZ ALMOHIMEED , REDHWAN M. A. SAAD, SHERIF MOSTAFA,NORA MAHMOUDEL-RASHIDY.
Publication year: 30 october 2023

In this study, we address Alzheimer's disease (AD),a progressive neurological disorder impacting millions globally with its hallmark symptoms of memory loss and cognitive decline. Early detectionis crucial for effective intervention and improved quality of life. Leveraging machine learning, we introduce a novel multi-level stacking model integrating diverse models and modalities to predictdistinct AD classes. Our approach includes cognitive sub-scores from the Alzheimer's Disease Neuroimaging Initiative dataset. At Level 1, six basemodels (RF, DT, SVM, LR, KNN, NB) train each modality, and stacking further combines their outputs. At Level 2, stacking models based on RF, LR, DT, SVM, KNN, and NB are utilized for each modality. Level 3 merges the stacking model outputs, creating a new dataset for final prediction. Feature selection optimization with Particle Swarm Optimization enhances model efficiency. Our model outperforms single-modality approaches and achieves superior results (92.08 % accuracy, 92.07% precision, 92.08% recall, and 92.01% F1- score for two classes; 90.03% accuracy, 90.19% precision, 90.03% recall, and 90.05% F1-score for three classes). This comprehensive approach shows promise for enhancing early AD diagnosis, emphasizing efficiency, effectiveness, and trust through explainable artificial intelligence (XAI) the results of our meticulously crafted model surpass those of single-modality approaches, showcasing superior performance

metrics. For two classes, we achieve an outstanding 92.08 percent accuracy, 92.07 percent precision, 92.08 percent recall, and a remarkable 92.01 percent F1-score. In the context ofthree classes, the model demonstrates exceptional performance with 90.03 percent accuracy, 90.19precision, 90.03 recall, and a notable 90.05 percent F1-score. This comprehensive approach not only demonstrates promise in enhancing early AD diagnosis but also underscores our commitment to efficiency, effectiveness, and trust through the integration of explainable artificial intelligence (XAI). In the pursuit of advancing diagnostic methodologies, our model stands as a testament to the potential of a holistic and technologically advanced approach in the realm of neurodegenerative diseases.

Paper 5: Applications of Explainable Artificial Intelligence in Diagnosis and Surgery
Author: Yiming Zhang , Ying Weng and Jonathan Lund
Publication year: 19 January 2022

In recent years, the potential of artificial intelligence(AI) in medicine has become evident, yet the challenge of explainability hinders its clinical applications. Addressing this limitation, research on explainable artificial intelligence (XAI) hasemerged, offering both decision-making and interpretability of AI models. This review explores the current trends in medical diagnosis and surgical applications using XAI, covering articles publishedbetween 2019 and 2021 from reputable sources. The unfurling narrative identifies a recurring pattern where techniques are at times justified by questionable objectives, such as building trust, or grounded in assumptions about the 'concepts'ingrained within deep learning algorithms. The overarching goal of this paper is to not only shed light on these misconceptions but to proactivelypropose tangible steps towards establishing XAI as a more substantive and grounded research domain. The introduction sets the tone by explicitly addressing those already immersed in the intricate world of XAI, offering constructive feedback with the ultimate aim of enhancing transparency within the broader field of Machine Learning (ML).

V. SYSTEM ARCHITECTURE:

Data Input Later:

The architecture begins with a data inputlayer that accepts diverse datasets in different formats. This layer is equippedto handle structured, unstructured, and semistructured data, recognizing theheterogeneity present in various domains.

Preprocessing Module:

A preprocessing module follows the datainput layer to clean, standardize, and transform the incoming data. This module addresses challenges such as missing values, outliers, or data inconsistencies to ensure that the subsequent XAI techniques operate on high-quality, preprocessed data.

Feature Extraction and Selection:

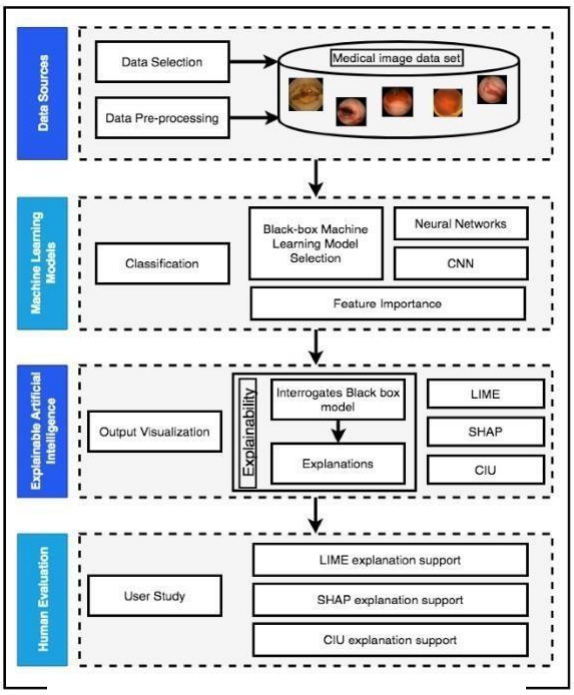
Depending on the nature of the dataset, afeature extraction and selection component may be integrated. This module identifies relevant features and, if necessary, reduces the dimensionality of the data to enhance the efficiency of subsequent XAI techniques

XAI Technique Integration Layer:

The core of Universal AI's architecture involves an XAI technique integration layer. This layer incorporates various XAI methods, such as rule-basedsystems, LIME (Local Interpretable Model-agnostic Explanations), SHAP(SHapley Additive exPlanations), or model-agnostic techniques. Each XAI method is integrated in a modular fashion, allowing users to select and apply the most suitable technique fortheir specific dataset and objectives.

Visualization and Interpretability Layer:

The results of the XAI techniques are visualized through an interpretation layer. This layer provides meaningfulvisual representations of the model's decision-making process, making it comprehensible to users. Visualization may include feature importance plots,decision trees, or other relevant graphics,depending on the XAI techniqueemployed.



system architecture

VI. IMPLEMENTATION:

LIME

LIME, or Local Interpretable Model-agnostic Explanations, is a technique in the field of explainableartificial intelligence (XAI) designed to provide local and interpretable explanations for the predictions of machine learning models. The primary objective of LIME is to shed light on the decisionmaking process of complex, often black-box models by creating alocal, interpretable surrogate model around a specific instance of interest.

In the LIME methodology, a crucial concept is the development of a local surrogate model. This model serves as a simplified and interpretable approximationof the behavior of the complex model in the vicinity ofthe instance being explained. Typically, this surrogatemodel is chosen to be a simple and understandable model, such as a linear model or a decision tree.

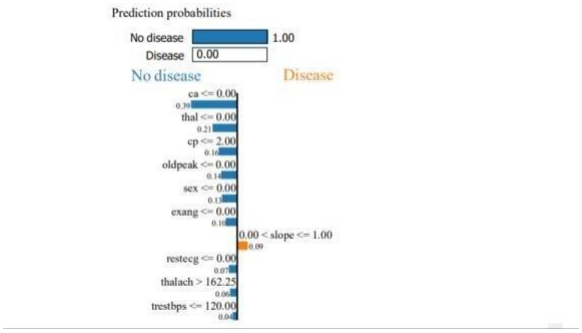
The process begins by selecting an instance for which an explanation is sought. Perturbations are then introduced to the features of this instance, creating a set of perturbed samples. The complex model is used to obtain predictions for each of these perturbed samples. Importantly, the perturbed samples are weighted based on their proximity to the original instance, assigning more weight to samples that are closer.

The weighted samples are then used to train the local surrogate model, with the weights indicating the importance of each perturbed sample in the training process. This local surrogate model provides a comprehensible representation of how the complex model behaves locally, offering insights into the significance of different features for the specific prediction of interest.

LIME is applicable to a wide range of machine learning models, irrespective of their complexity or type. It is particularly useful when dealing with models that are considered black boxes, such as deep neural networks, where understanding the decision making rationale is challenging.

The technique finds applications in various domains, including healthcare, finance, and image classification, where the interpretability of individual predictions is crucial. It excels in providing explanations for individual predictions on a local level, contributing to the transparency and trustworthiness ofmachine learning models.

While LIME offers versatility and model-agnostic interpretability, it is not without its challenges. The explanations generated by LIME may be sensitive to the choice of perturbations, and the inherently simplified nature of the local surrogate model may not capture all intricacies of the complex model. Despite these limitations, LIME remains a valuable tool in thetoolkit of XAI, offering a means to demystify the decision-making processes of sophisticated machine learning models. It is implemented as a Python library,providing practical functionality for researchers and practitioners working on interpretability in machine learning.



lime implementation

SHAP

SHAP, or SHapley Additive exPlanations, stands out as a prominent technique in the domain of explainable artificial intelligence (XAI). Developed based on cooperative game theory, SHAP values are rooted in the concept of Shapley values, which provides a fair and comprehensive approach to understanding the contribution of each feature in a machine learning model's prediction.

The fundamental idea behind SHAP values lies infairly attributing the contribution of each feature to the prediction made by a model. In essence, SHAP valuesaim to quantify the impact of individual features on the output of a machine learning model, fostering interpretability and transparency.

At the core of SHAP is the concept of Shapley values from cooperative game theory. These values provide a fair distribution of a "payout" among a coalition of participants. In the context of machine learning, SHAP values extend this concept to features, considering all possible combinations (coalitions) of features and determining the fair contribution of each feature to themodel's prediction.

One of the key strengths of SHAP values is theiradherence to the principle of fairness. By considering all possible coalitions of features, SHAP values ensurea fair allocation of the contribution of each feature. This fairness is achieved by calculating the average marginal contribution of a feature across all possible coalitions. The "additivity" property of SHAP values is noteworthy, ensuring that the sum of SHAP values forall features equals the difference between the model's prediction for a specific instance and the average prediction across all instances. This property enhancesthe interpretability and coherence of the explanations provided by SHAP.

The calculation process involves establishing a reference or baseline prediction, permuting featurevalues to create different instances, and calculating theShapley values by considering all possible permutations and coalitions of features. The resulting SHAP values offer insights into the impact of each feature on a specific prediction, with positive values indicating a positive contribution and negative values indicating a negative impact.

SHAP values find applications across diverse machinelearning models, including tree-based models, linear models, and neural networks. They are employed for both classification and regression tasks, contributing to the interpretability of model predictions.

While SHAP values offer a theoretically grounded and fair method for feature attribution, they come with certain limitations. The computational intensity of calculating SHAP values for all possible permutationscan be a challenge, particularly for complex models and large datasets. Additionally, the interpretability of SHAP values is dependent on the interpretability of theunderlying model.

Implemented as a Python library, SHAP provides a practical and accessible tool for researchers and practitioners in the field of XAI. Overall, SHAP values serve as a powerful and consistent framework for interpreting the output of machine learning models, advancing the understanding of feature contributions in complex systems.

VII. REFERENCES

[1] "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable"-Author: Christoph Molnar, 2023.

[2] "Explanatory Model Analysis" - Authors: David S. Watson, Bruce G. Buchanan January 2020.

[3] "Explainable Artificial Intelligence: Understanding, Visualizing, and Interpreting Deep Learning Models - Authors: Christoph Heindl, Andreas Holzinger, 2020.

[4] "Explainable Artificial Intelligence: Understanding, Visualizing, and Interpreting Deep Learning Models , Author: Sameer Singh 2020

[5] "Explainable AI in Healthcare: Empowering Humans in Decision Making", Authors: Fong-Li Chong, Saman Halgamuge, Published: 2020

[6] "Explainable AI: Foundations, Architectures, and Applications", Editors: Kacprzyk, Janusz, Filipe, Joaquim, 2019

[7] "Explanations in Artificial Intelligence: Insights from the Social Sciences" - Authors: Tim Miller, Piers Howe, Anna R. Cox, 2019.

[8] "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities, and Challenges toward Responsible AI" Authors: A. Holzinger, C. Biemann, et al, 2019.

[9] "Explainability of Machine Learning Models: A Survey",Authors: Ahmad Alabduljabbar, Xavier Ferrer, et al,2017.

[10] "Explainable AI: Beware of Inmates Running the Asylum Or: How I Learnt to Stop Worrying and Love the Social and Behavioural Sciences" Authors: L. Biran, C. B. Cooper,2017.

[11] "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable" Authors: Christoph Molnar,2017.

[12] "Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models" - Authors: K. S. Pedersen, A. Johansen, et al, 2017.

[13] "A Survey of Methods for Explaining Black Box Models" - Authors: Riccardo Guidotti, Anna Monreale, etal, 2015.

[14] "Towards A Rigorous Science of Interpretable Machine Learning" - Authors: Finale Doshi-Velez, Been Kim, 2015.

[15] "Explainable Machine Learning for Scientific Insights and Discoveries" - Authors: Been Kim, Finale Doshi-Velez, 2015

[16] "Counterfactual Explanations without Opening the Black Box: Automated Decisions and the GDPR"- Authors: Sandra Wachter, Brent Mittelstadt, Chris Russell, 2015.

[17] "An Empirical Study on the Impact of Design Choices on the Performance of XAI Methods" Authors: Sarah Tan, Wentao Wang, et al, 2015.