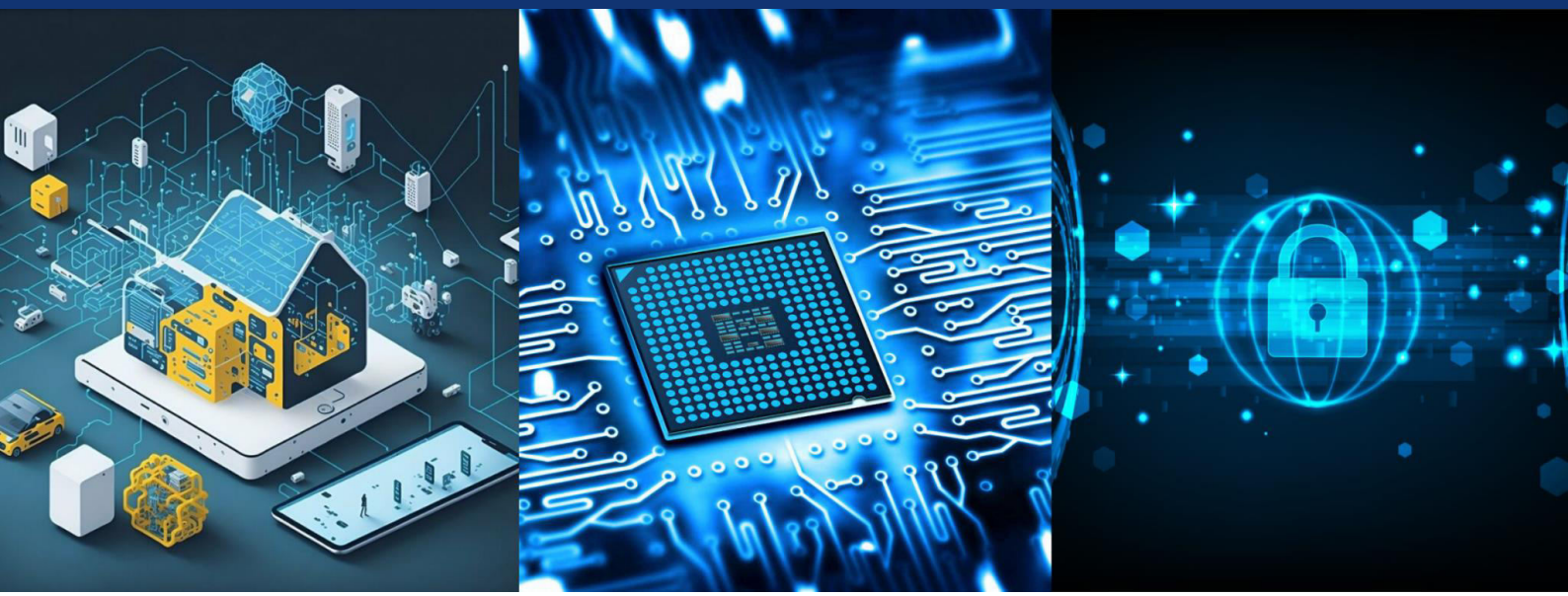
IJIRCCE

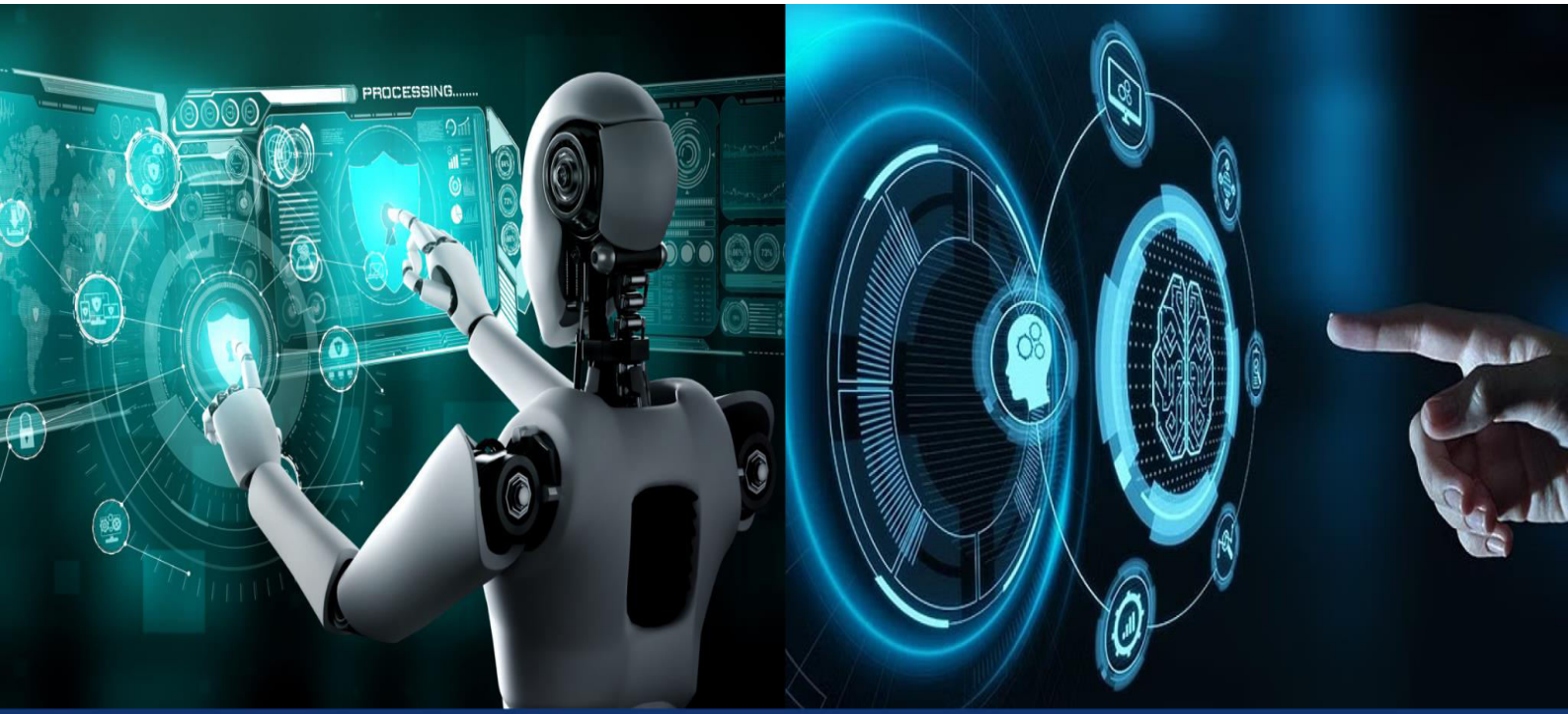
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 14, Issue 4, April 2026

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404058

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Design and Evaluation of Explainable Machine Learning Model for Predictive Decision Making in Real World Systems

N. Prasad[1],K.Hema Sri [2],D.Sai Manikanta[3],A. Jithin Reddy [4], R. Chantan Suryakumar [5]

Assistant Professor, Department of Information Technology, Sir C R Reddy College of Engineering, Eluru, India[1]

Final Year Students, Department of Information Technology, Sir C R Reddy College of Engineering, Eluru, India[2, 3, 4, 5]

ABSTRACT: Machine learning models are commonly employed for predictive decision-making in practical systems; however, numerous models operate as black boxes, rendering their decisions challenging to decipher. This project is all about making a machine learning model that can explain itself and give correct predictions and a clear explanation of why it worked. The suggested system uses a gradient-boosted model and explainability methods to find the most important factors that affect decisions. The model is tested with data that is similar to what would be used in real life, like in healthcare, finance, and traffic management, to make sure it works well even when things aren't clear. The system makes machine learning models more reliable and understandable decision-support tools by giving users useful explanations and predictions. This makes the models more transparent, builds user trust, and helps people make better decisions.

I. INTRODUCTION

This paper addresses Machine Learning (ML) is one of the most important technologies in modern computing. It lets systems look at huge amounts of data and make predictions with little help from people. ML is used in many fields, including healthcare, finance, transportation, and manufacturing .models are being used more and more to help make important decisions. These systems help businesses find patterns, guess what will happen, and make operations run more smoothly, which boosts productivity and efficiency. Even though these models have these benefits, many advanced machine learning models work like "black boxes." This means that they make accurate predictions but don't explain how they do it. In real life, where it's important to know why decisions are made, this lack of transparency makes things .It is important to know why decisions are made. Decision-makers can't just look at outputs in sensitive situations like medical diagnosis or financial risk assessment without knowing what factors are affecting them. This project is focused on creating an explainable machine learning model that can make both accurate predictions and useful explanations in order to solve this problem. The proposed approach aims to improve transparency, build user trust, and support responsible use of AI-driven solutions by adding interpretability to predictive systems .As AI becomes more common in everyday decisions, the need for openness has grown. When users know how decisions are made, they are more likely to trust AI-driven systems. Even very accurate models may not be accepted by stakeholders if they don't have clear explanations. Explainability is also very important for making sure that things are fair and that people are held accountable. Organisations can find and fix biases by knowing which features affect predictions. This is especially important in fields like health care and finance, where making biased decisions Also, interpretable models help developers make systems work better by pointing out important factors and showing possible data problems. The goal of this project is to make a machine learning system that strikes a balance between accuracy and interpretability .This project seeks to tackle this issue by creating a predictive model that performs well and provides clear explanations. The goal is to make sure that people who make decisions can trust and use machine learning predictions well.

It aims to make the system more transparent by highlighting which features influence decisions and by explaining how those decisions are made. Because of this, it works well in areas like healthcare, finance, and operational analytics, where understanding the reasoning behind outcomes is important.

The system is designed to be more transparent by showing which features matter most and by clearly explaining how decisions are reached. This makes it especially useful in fields like healthcare, finance, and operational analytics, where

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

it's important to understand the reasoning behind outcomes.

Although it currently works with structured data, the framework has the potential to be adapted to other areas over time. Ultimately, the goal of the project is to create a strong foundation for building AI systems that people can trust—systems that not only improve efficiency but also promote accountability.

II. LITERATURE REVIEW

Machine learning has become an essential tool for making predictions and supporting decisions in real- world areas like healthcare, finance, manufacturing, transportation, and smart infrastructure. Earlier models—such as linear regression, logistic regression, and decision trees—were easier to understand because their structure clearly showed how input factors influenced the results.

As the need for higher accuracy grew, more advanced models like Random Forests, Gradient Boosting Machines, Support Vector Machines, and deep neural networks were developed. While these models often deliver better performance, they come with a major drawback: they are much harder to interpret. Many of them function as "black boxes," meaning it's difficult to understand how they arrive at their decisions. This lack of transparency can be a serious concern, especially in critical applications where decisions need to be fair, explainable, and accountable.

To address this issue, the field of Explainable Artificial Intelligence (XAI) has gained increasing attention. The goal of XAI is to make machine learning models more understandable to humans without significantly reducing their accuracy. In general, interpretability methods fall into two main categories: intrinsic interpretability, where models are designed to be understandable from the start, and post-hoc explanation techniques, which are used to explain complex models after they have been trained.

Models like decision trees, rule-based systems, and generalized additive models are designed to be transparent from the start, making it easy for users to follow how decisions are made step by step. However, they can sometimes struggle when it comes to capturing very complex, nonlinear patterns in large datasets .To deal with this, post- hoc explanation methods are used to interpret more complex "black-box" models after they've been trained. Two of the most popular techniques are LIME and SHAP. LIME explains individual predictions by approximating the complex model with a simpler one in a small, local area around a specific case SHAP, which is based on ideas from game theory, assigns a contribution value to each feature, helping ensure consistent and reliable explanations. Another useful approach is counterfactual explanations, which show the smallest changes needed in the input to get a different outcome—making it easier to explore "what-if" scenarios.

Explainable machine learning is especially important in areas where decisions have a direct impact on people's lives. In healthcare, doctors use interpretable models for diagnosing diseases, assessing patient risks, and planning treatments, and clear explanations help them trust and verify the results. In finance, explainability is crucial for tasks like credit scoring, fraud detection, and loan approvals, particularly because regulations often require clear reasoning behind automated decisions. In industrial settings, explainable models are used in predictive maintenance to identify the causes of equipment failures, helping reduce downtime and improve overall efficiency

Authors	Year	Title
Ribeiro er al.	**2022**	**Explainibility in machine learning model**
Zhang &Liu	**2023**	**Interpretable Machine Learning for Decision systems**
Ali et al.	**2024**	**Explainable AI using SHAP and LIME**
Shobaki et al.	**2025**	**Adaptive Explainable ai for dynamic system**
K.Q.		

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404058

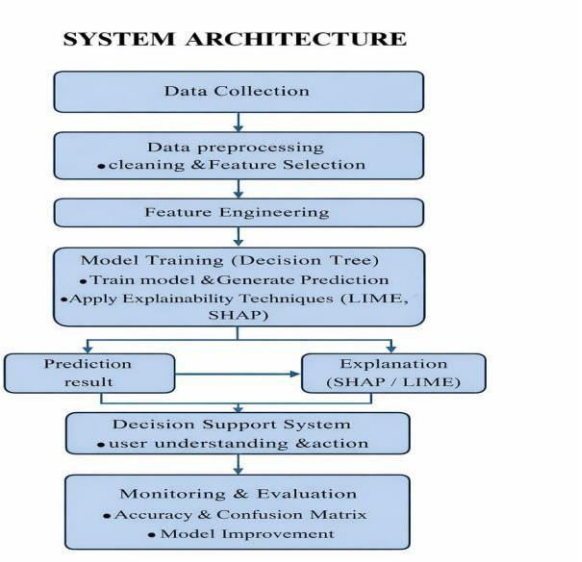
www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. PROBLEM STATEMENT

Traditional machine learning models are often very good at making accurate predictions, but many of them operate like "black boxes." In simple terms, they give you an answer without clearly showing how they arrived at it. This makes it difficult for users to understand the reasoning behind decisions, especially when the models are complex .because of this lack of transparency, people may hesitate to trust these systems, particularly in situations where decisions have serious consequences. For example, in areas like healthcare or finance, it's not enough to just get the right prediction—users also need to know *why* a certain decision was made. Without clear explanations, it becomes harder to verify results, detect errors, or ensure that the system is fair and unbiased .This challenge highlights the growing need for machine learning models that are not only accurate but also explainable.



IV. SYSTEM ARCHITECTURE

The system architecture is designed as a clear,step-by-step workflow that makes the entire explainable AI process easy to follow and understand. It begins with data generation, where synthetic data is created to simulate real-world scenarios and establish a known reference point. This is followed by data loading and inspection, ensuring that the dataset is accurate and ready for further processing. In the preprocessing stage, the data is cleaned, scaled, and prepared so that it can be effectively used by the model. Once the data is ready,the model training phase builds a predictive model—such as a decision tree— that not only performs well but is also easier to interpret.

After training, the model is evaluated to check its performance and reliability using appropriate metrics. The system then moves to inference and prediction, where new data is fed into the trained model to generate results. These predictions are not just outputs; they are accompanied by explanations that help users understand how the model arrived at its decisions. then saved and deployed, making the system images. All stages are integrated into a unified end-to-end framework, providing an efficient and scalable solution for real-world applications including surveillance, autonomous systems, and industrial automation.

V. EXISTING SYSTEM

In many real-world applications, existing systems for predictive decision-making rely on traditional machine learning models or rule-based approaches. These systems are designed to handle structured data, using statistical techniques or machine learning algorithms to analyze past information and generate predictions or classifications. They are widely used in areas such as healthcare for risk prediction, finance for credit scoring and fraud detection, and industrial settings for identifying equipment faults. Common models include logistic regression, decision trees, random forests, support vector machines, and deep neural networks.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

The workflow of these systems typically follows a structured sequence. It begins with data collection, where relevant historical data is gathered. This is followed by preprocessing, where the data is cleaned and prepared for analysis, and feature selection, where the most important variables are identified. Next is model training, where the system learns patterns from labeled data using supervised learning methods. After training, the model is evaluated to measure its performance and accuracy before being deployed into a real-world environment .Once deployed, the system processes new data, either in real time or in batches, and produces outputs such as risk scores, probability values, or classification results to support decision- making processes. While these systems are highly effective and focus strongly on achieving accuracy, they often prioritize performance over transparency, making it difficult for users to fully understand how decisions are made.

VI. PROPOSED SYSTEM

The proposed system is an Explainable Machine Learning (XML) framework designed to provide both accurate predictions and clear, understandable explanations. Unlike traditional black-box models, this system is built with interpretability at its core, ensuring that every prediction can be explained in a meaningful way. The overall architecture follows a modular structure that includes data collection, preprocessing, feature engineering predictive modelling, explanation generation, and monitoring. This organized design makes the system easier to scale, maintain, and deploy in real-world applications. The process begins with collecting raw data from trusted sources such as databases, sensors, or enterprise systems. This data is then stored and used as the foundation for training and validating the model. Next, preprocessing steps like cleaning the data, normalizing values, encoding categorical variables, and selecting important features are applied to improve data quality and consistency. A feature store is also used to ensure that the same transformations are applied during both training and real-time predictions, which helps maintain stability and accuracy.

At the heart of the system, the predictive model generates results along with confidence scores to show how certain it is about each prediction. After that, the explainability module steps in to make the decision-making process more transparent by providing both local and global explanations. It highlights which features influenced a specific prediction and how the model behaves overall. In addition, performance monitoring tools are used to track accuracy and detect issues like model drift. With continuous feedback and updates, the system stays reliable, adaptable, and transparent even in changing real-world conditions.

VII. RESULT

- The project titled **"Design and Evaluation of Explainable Machine Learning Model for Predictive Decision Making in Real World Systems"** focuses on building a model that provides both accurate predictions and clear explanations.
- The system collects and preprocesses data before using it to train a machine learning model for predictive analysis.
- The trained model is evaluated using performance metrics such as:
  - Accuracy
  - Classification report
  - Confusion matrix
- These evaluation methods help demonstrate the system's ability to analyse data effectively and produce reliable predictions.
- A key feature of the project is the integration of **Explainable Artificial Intelligence (XAI)** techniques.
- XAI techniques help users understand how different input features influence the model's decisions.
- The system provides feature importance and visual explanations to improve transparency.
- These explanations make it easier for users to interpret how and why a particular prediction was made.
- This level of interpretability is especially important in real-world applications where understanding the reasoning behind decisions is critical.

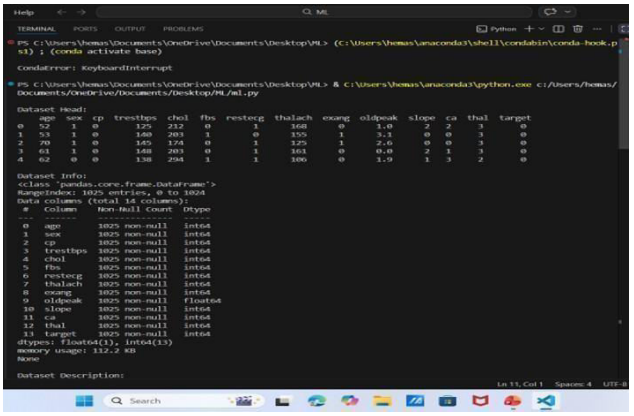
Output screen:

© 2026 IJIRCCE | Volume 14, Issue 4, April 2026 DOI: 10.15680/IJIRCCE.2026.1404058

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

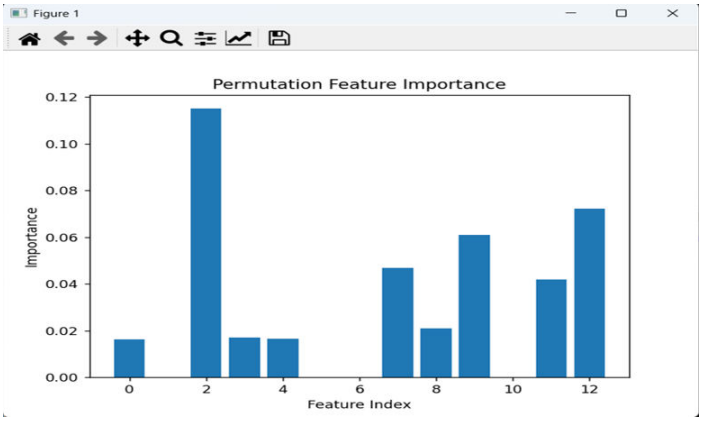
International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Overall, the results highlight that the proposed model strikes a strong balance between detection accuracy and computational efficiency, making it well-suited for practical applications such as surveillance, healthcare, and intelligent automation systems.

Output screens:



VIII. CONCLUSION

This paper presented the **"Design and Evaluation of Explainable Machine Learning Model for Predictive Decision Making in Real World Systems"** is focused on building a system that not only makes accurate predictions but also clearly explains how those predictions are made. In this approach, data is first collected and preprocessed so that it is clean and suitable for training a machine learning model used for predictive analysis.

Once the model is trained, it is evaluated using standard performance measures such as accuracy, a classification report, and a confusion matrix. These metrics help show how well the system is performing and whether it can make reliable predictions on new data.A major part of this project is the use of Explainable Artificial Intelligence (XAI) techniques. These techniques make the model more transparent by showing how different featurs influence its decisions. With tools like feature importance and visual explanations, users can better understand the reasoning behind each prediction.

IX. FUTURE WORK

The proposed explainable machine learning system can be significantly enhanced in several important directions to improve its performance, scalability, and real-world applicability. One key improvement is to extend the system to handle more complex and diverse types of datasets, including unstructured data such as images, text, audio, and time-series data. This would allow the framework to be applied in broader domains like medical imaging, natural language processing, financial forecasting, and smart surveillance systems .Another important area for future development is the

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

integration of more advanced deep learning models, such as convolutional neural networks and transformer-based architectures, along with stronger explainability techniques. This combination can help achieve higher predictive accuracy while still maintaining transparency in decision-making, which is often a major challenge in deep learning systems .The system can also be improved by enabling real- time deployment in cloud-based and edge computing environments. This would support faster processing, scalability, and continuous decision-making in dynamic real-world applications such as IoT systems, autonomous systems, and live monitoring platforms.

REFERENCES

1. **Ribeiro et al. (2016):** Introduced **LIME** for explaining individual predictions using local surrogate models.
2. **Lundberg & Lee (2017):** Introduced **SHAP** for fair feature attribution using game theory.
3. **Molnar (2020):** Comprehensive guide to **interpretable machine learning methods**.
4. **Samek et al. (2017):** Focus on **visualizing and interpreting deep neural networks**.
5. **Wachter et al. (2017):** Proposed **counterfactual explanations** for model decisions.
6. **Kim et al. (2018):** Introduced **TCAV**, explaining models using human concepts.
7. **Breiman (2001):** Developed **Random Forests** and feature importance for interpretability.
8. **Friedman (2001):** Proposed **Partial Dependence Plots (PDP)** for understanding feature effects.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details