

INTERNATIONAL JOURNAL OF SCIENTIFIC INFORMATION

Research | ISSN: 2583-8512

A Novel Mathematical Framework for Trustworthy Artificial Intelligence Using Fuzzy Logic, Topology, and Uncertainty Modelling

Parul and Dr. Vishal Saxena

Department of Mathematics, Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India

Jakharparul19@gmail.com

Abstract

Artificial Intelligence (AI) has achieved remarkable progress in recent years, enabling advanced decision-making in healthcare, autonomous systems, cybersecurity, natural language processing, and scientific computing. However, the increasing complexity of modern AI models, particularly deep learning architectures, has introduced significant challenges related to interpretability, transparency, robustness, and uncertainty management. Many advanced AI systems operate as black-box models, where the underlying reasoning process behind predictions remains difficult to understand. This limitation restricts their adoption in applications where reliability and accountability are essential. This study proposes a novel mathematical framework for developing trustworthy artificial intelligence by integrating three complementary mathematical approaches: fuzzy logic, topology, and uncertainty modelling. Fuzzy logic provides an interpretable mechanism for representing ambiguous and incomplete information through membership functions and reasoning rules. Topological Data Analysis (TDA) enables the identification of hidden structural relationships and persistent patterns within complex high-dimensional datasets. Uncertainty modelling enhances reliability by estimating prediction confidence and distinguishing between reliable and uncertain outcomes. The proposed fuzzy-topological uncertainty framework aims to move beyond conventional accuracy-oriented AI models by embedding trust-related characteristics directly into the mathematical architecture. The framework combines interpretable reasoning, structural data analysis, and uncertainty-aware prediction within a unified model. A mathematical formulation is developed, and the framework is evaluated conceptually based on interpretability, robustness, and reliability considerations. The proposed approach provides a foundation for designing AI systems that are not only computationally effective but also transparent, explainable, and trustworthy. The framework has potential applications in safety-critical domains including healthcare, autonomous technologies, cybersecurity, and scientific decision-support systems.

Keywords: Trustworthy Artificial Intelligence , Fuzzy Logic , Topological Data Analysis , Uncertainty Modelling , Explainable Artificial Intelligence , Mathematical Modelling

Address for Correspondence

Dr. Vishal Saxena

Professor

Department of Mathematics

Jayoti Vidyapeeth Women's University, Jaipur, Rajasthan, India



Article Received on: 15.07.26

Revised on: 27.07.26

Approved for publication: 23.08.26

How to Cite this Article: **Parul and Saxena V. A Novel Mathematical Framework for Trustworthy Artificial Intelligence Using Fuzzy Logic, Topology, and Uncertainty Modelling** Int., J. Sci. Info. 2026; 4 (5): 1-13

1. Introduction

Artificial Intelligence (AI) has emerged as one of the most transformative technologies of the modern era, enabling computational systems to perform complex tasks involving learning, reasoning, prediction, and decision-making. Recent advances in machine learning and deep learning have significantly improved AI capabilities in diverse domains, including healthcare, autonomous systems, natural language processing, cybersecurity, finance, and scientific computing (Goodfellow et al., 2016). Deep neural networks have achieved remarkable success because of their ability to automatically extract complex patterns from large-scale datasets. However, despite their superior predictive performance, many advanced AI models function as black-box systems, where the internal reasoning mechanisms behind predictions remain difficult for humans to understand. This limitation has raised concerns regarding transparency, accountability, reliability, and safe deployment, particularly in applications where AI decisions directly influence human lives.

The increasing use of AI in critical domains has created a growing demand for trustworthy artificial intelligence systems. Trustworthy AI focuses not only on predictive accuracy but also on developing systems that are explainable, robust, fair, secure, and accountable throughout their operational lifecycle (Li et al., 2021). Explainable Artificial Intelligence (XAI) has therefore become an important research area aimed at improving human understanding of AI decision-making processes. Ribeiro et al. (2016) introduced the Local Interpretable Model-Agnostic Explanations (LIME) framework to generate interpretable explanations for complex machine learning predictions. However, most existing explanation techniques are post-hoc approaches that attempt to interpret models after training rather than incorporating transparency directly into the mathematical structure of AI systems. Therefore, developing inherently interpretable and reliable AI architectures remains a major research challenge.

Mathematical modelling provides a promising foundation for addressing these limitations by enabling AI systems to represent uncertainty, ambiguity, and complex relationships more effectively. Fuzzy logic is one of the most significant mathematical approaches for handling imprecise information in intelligent systems. Zadeh (1965) introduced fuzzy set theory, which allows elements to possess partial membership values between zero and one rather than following strict binary classification. This capability enables AI systems to represent human-like reasoning and process uncertain concepts such as “high risk,” “moderate probability,” and “low confidence.” In addition, topology and Topological Data Analysis (TDA) provide mathematical tools for understanding structural relationships within complex datasets. Carlsson (2009) demonstrated that topological approaches can identify hidden patterns, connectivity structures, and persistent features in high-dimensional data that may not be captured by conventional statistical methods. Similarly, uncertainty modelling has become essential for reliable AI because intelligent systems must communicate not only predictions but also confidence levels associated with those predictions (Kendall & Gal, 2017).

Although fuzzy logic, topology, and uncertainty modelling have individually contributed to improving AI reliability, their combined integration into a unified mathematical framework remains insufficiently explored.

Existing approaches generally address specific challenges separately, where fuzzy systems focus on reasoning under ambiguity, topology analyses structural complexity, and uncertainty models estimate prediction confidence. This research proposes a novel fuzzy-topological uncertainty framework that integrates these three mathematical perspectives to develop trustworthy AI systems. By embedding interpretable reasoning, structural data understanding, and uncertainty estimation directly into AI architectures, the proposed approach aims to overcome limitations of conventional black-box models and establish a mathematical foundation for developing AI systems that are accurate, transparent, robust, and reliable

2. Literature Review

Artificial Intelligence has evolved rapidly from traditional rule-based systems toward highly sophisticated learning architectures capable of extracting complex patterns from large-scale data. However, the increasing complexity of modern AI models has created significant challenges regarding transparency, interpretability, and reliability. The concept of trustworthy AI has emerged to address these concerns by emphasizing that AI systems should not only achieve high predictive performance but should also demonstrate explainability, robustness, fairness, privacy, and accountability. According to Li et al. (2021), trustworthy AI requires a comprehensive approach that considers the complete lifecycle of AI development, including data collection, model training, deployment, and evaluation. Similarly, Gunning and Aha (2019) highlighted that explainability is essential for increasing human confidence in AI systems, particularly in high-risk applications where users need to understand the reasoning behind automated decisions. Although explainable AI methods have improved transparency, many existing techniques provide explanations only after model training and do not fundamentally address the black-box characteristics of complex learning architectures.

Fuzzy logic has been recognized as an important mathematical approach for developing interpretable and uncertainty-aware AI systems. Zadeh (1965) introduced fuzzy set theory to overcome the limitations of classical binary logic by allowing partial degrees of membership between complete truth and falsehood. This mathematical principle enables computational systems to represent vague concepts and uncertain information similar to human reasoning processes. Fuzzy systems have been widely applied in intelligent control, classification, medical diagnosis, and decision-support applications because their reasoning mechanisms can be expressed through understandable linguistic rules. Dubois and Prade (2000) further contributed to fuzzy mathematics by developing theoretical foundations for uncertainty representation and possibility theory, demonstrating the potential of fuzzy approaches for knowledge representation and intelligent decision-making. However, conventional fuzzy systems often depend on expert-defined rules and membership functions, which can become difficult to design for complex high-dimensional datasets. To overcome these limitations, researchers have investigated hybrid approaches such as neuro-fuzzy systems that combine the learning capability of neural networks with the interpretability of fuzzy reasoning.

Topology and Topological Data Analysis (TDA) have recently gained attention as powerful mathematical approaches for analysing complex AI data structures. Traditional machine learning techniques generally rely

on statistical relationships between features; however, high-dimensional datasets often contain hidden geometric structures that may not be effectively captured using conventional methods. Carlsson (2009) introduced TDA as a framework for extracting meaningful information from complex datasets by analysing topological properties such as connectivity, clusters, and persistent structures. One of the key concepts in TDA is persistent homology, which enables researchers to identify important features that remain stable across different scales. These capabilities make topology particularly valuable for understanding the internal representations learned by deep neural networks. Recent studies have demonstrated that topological methods can support applications including image analysis, anomaly detection, representation learning, and neural network interpretation. Nevertheless, topology-based approaches mainly focus on structural understanding and do not independently provide mechanisms for uncertainty representation or human-interpretable reasoning.

Uncertainty modelling represents another essential component of trustworthy artificial intelligence because real-world data frequently contain noise, incomplete observations, and unpredictable variations. AI systems operating in practical environments must therefore estimate not only the predicted outcome but also the confidence associated with that prediction. Kendall and Gal (2017) investigated uncertainty estimation in deep learning and demonstrated the importance of distinguishing between different forms of uncertainty, including aleatoric uncertainty arising from data variability and epistemic uncertainty related to model limitations. Similarly, Guo et al. (2017) highlighted that many modern neural networks produce poorly calibrated confidence estimates, where highly confident predictions may not always correspond to reliable outcomes. Various approaches, including Bayesian neural networks, Monte Carlo dropout, and probabilistic models, have been developed to address uncertainty estimation. However, these techniques often involve increased computational complexity and may provide limited interpretability, creating a need for alternative mathematical approaches capable of combining reliability estimation with transparent reasoning.

Although fuzzy logic, topology, and uncertainty modelling have individually contributed to improving AI reliability, their integration into a unified mathematical framework remains a relatively unexplored research area. Existing studies typically focus on solving specific challenges independently: fuzzy systems address ambiguity, topological methods analyse structural complexity, and uncertainty models estimate prediction confidence. A comprehensive framework combining these three mathematical perspectives could provide a more complete solution for trustworthy AI by integrating interpretable reasoning, structural awareness, and uncertainty quantification. The proposed research addresses this gap by developing a fuzzy-topological uncertainty framework in which mathematical principles are embedded directly into AI system design. Such an approach has the potential to move beyond conventional black-box models and support the development of AI systems that are not only accurate but also transparent, explainable, and reliable for real-world applications.

3. Research Methodology

3.1 Research Design and Framework Development

This study adopts a mathematical modelling and computational validation approach to develop a trustworthy

artificial intelligence framework by integrating fuzzy logic, topology, and uncertainty modelling. The proposed methodology is designed to address three major limitations of conventional AI systems: lack of interpretability, inability to capture complex structural relationships, and insufficient uncertainty estimation. The framework follows the principle that trust should be embedded within the AI model architecture rather than introduced only through post-hoc explanation methods (Doshi-Velez & Kim, 2017; Li et al., 2021).

The proposed framework consists of three interconnected mathematical components: (i) fuzzy logic-based reasoning for representing uncertainty and interpretability, (ii) topological data analysis for extracting structural information from complex datasets, and (iii) uncertainty modelling for estimating prediction reliability. The overall objective is to formulate an AI model that produces accurate predictions while simultaneously providing interpretable reasoning and confidence-aware outputs.

3.2 Development of the Proposed Fuzzy–Topological Uncertainty Framework

3.2.1 Fuzzy Logic-Based Mathematical Representation

Fuzzy logic is incorporated as the reasoning component of the proposed framework. Unlike classical binary logic, fuzzy set theory allows partial membership representation, enabling AI systems to model uncertain and imprecise information (Zadeh, 1965).

For an input variable x , a fuzzy set A is represented as:

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

where $\mu_A(x)$ represents the membership degree of input x belonging to fuzzy set A , with:

$$0 \leq \mu_A(x) \leq 1$$

The membership function transforms raw input variables into linguistic categories such as low, medium, and high confidence. These fuzzy representations are processed through fuzzy inference rules:

$$IF \ x_1 \text{ is } A_1 \text{ AND } x_2 \text{ is } A_2, \text{ THEN } y \text{ is } B$$

where A_1, A_2 represent input fuzzy sets and B represents the output decision category. This mechanism improves model interpretability by allowing predictions to be associated with understandable reasoning patterns.

3.2.2 Topological Data Representation

To capture complex relationships within high-dimensional datasets, the proposed framework incorporates Topological Data Analysis (TDA). Unlike conventional feature-based approaches, topology analyses the structural properties of data, including connectivity and persistent patterns (Carlsson, 2009).

Given a dataset:

$$X = \{x_1, x_2, \dots, x_n\}$$

where each x_i represents a data point, TDA constructs geometric representations through simplicial complexes.

Persistent homology is then applied to identify structural features that remain stable across different scales.

The persistence of topological features can be represented as:

$$PH_k(X) = \{(b_i, d_i)\}$$

where b_i and d_i represent the birth and death points of a topological feature, and k represents the homological dimension.

These extracted topological features provide additional structural information to the AI model and enhance robustness by identifying meaningful patterns within complex datasets.

3.2.3 Uncertainty Modelling and Confidence Estimation

The third component of the framework focuses on uncertainty estimation. A trustworthy AI system should provide both prediction outcomes and associated confidence levels. Following uncertainty modelling approaches in deep learning, the framework considers two major uncertainty sources: aleatoric uncertainty caused by data variability and epistemic uncertainty caused by model limitations (Kendall & Gal, 2017).

The total predictive uncertainty can be expressed as:

$$U_{total} = U_{aleatoric} + U_{epistemic}$$

The confidence score of a prediction is represented as:

$$C = 1 - U$$

where C indicates prediction confidence and U represents estimated uncertainty.

If uncertainty exceeds a predefined threshold:

$$U > \tau$$

the system identifies the prediction as unreliable and recommends further validation. This mechanism prevents overconfident decisions and improves the reliability of AI outputs.

3.2.4 Integrated Fuzzy–Topological Uncertainty Model

The final stage integrates the three mathematical components into a unified trustworthy AI framework. The proposed model can be expressed as:

$$T_{AI} = F(F_z, T_p, U_m)$$

where:

F_z =	fuzzy	reasoning	component,
T_p =	topological	representation	component,
U_m =	uncertainty modelling component.		

The workflow of the proposed framework is:

*Input Data → Fuzzy Representation → Topological Feature Extraction → Uncertainty Estimation
→ Trustworthy Prediction*

The fuzzy module provides interpretable reasoning, the topological module captures hidden structural relationships, and the uncertainty module evaluates prediction reliability. Through this integration, the proposed framework aims to overcome limitations of conventional black-box AI models by embedding transparency and confidence estimation within the learning process.

3.3 Computational Validation and Evaluation Strategy

The developed framework will be computationally implemented and evaluated using benchmark machine learning datasets. The performance will be compared with conventional AI approaches to determine improvements in prediction reliability, interpretability, and uncertainty awareness.

Model performance will be evaluated using standard classification metrics:

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ \text{Precision} &= \frac{TP}{TP + FP} \\ \text{Recall} &= \frac{TP}{TP + FN} \end{aligned}$$

and:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

where TP, TN, FP, and FN represent true positive, true negative, false positive, and false negative predictions. In addition to predictive performance, trustworthiness will be assessed through uncertainty calibration and robustness analysis. Calibration performance will evaluate whether predicted confidence values accurately correspond to actual model reliability (Guo et al., 2017). Robustness testing will examine model behaviour under noisy and uncertain input conditions.

4. Mathematical Analysis and Framework Evaluation

4.1 Structural Integration of the Proposed Framework

The proposed fuzzy-topological uncertainty framework establishes a unified mathematical structure by integrating three independent mathematical components: fuzzy reasoning, topological representation, and uncertainty estimation. The resulting framework provides a multi-dimensional representation of AI decision-making, where prediction generation is associated with both interpretability and reliability assessment.

The mathematical formulation demonstrates that the proposed framework extends conventional AI models by incorporating additional trust-related parameters. The final AI output can be represented as:

$$Y = f(X, F, T, U)$$

where X represents the original input space, F represents fuzzy reasoning characteristics, T represents topological features, and U represents uncertainty estimation. Unlike conventional models where:

$$Y = f(X)$$

the proposed formulation introduces additional mathematical information that contributes to understanding the prediction process.

The integration of these components provides three fundamental improvements. First, fuzzy modelling enables representation of ambiguous information through continuous membership values. Second, topological analysis captures structural characteristics of complex datasets beyond conventional feature relationships. Third, uncertainty modelling provides confidence estimation associated with generated predictions. Therefore, the proposed framework establishes a mathematical basis for developing AI systems that are both predictive and trustworthy.

4.2 Mathematical Behaviour of the Fuzzy Reasoning Component

The fuzzy component of the proposed framework provides interpretable decision-making by transforming numerical inputs into linguistic representations. The membership function:

$$\mu_A(x): X \rightarrow [0, 1]$$

allows each input observation to possess a degree of belonging to a particular fuzzy category.

The advantage of this representation is that decision boundaries are no longer restricted to rigid classifications. Instead, the model can represent transitional states between categories. For example, an observation may simultaneously possess characteristics of both "medium" and "high" risk categories with different membership values.

The fuzzy inference mechanism therefore improves transparency because model decisions can be traced through mathematical rules rather than hidden parameter relationships. This property addresses a major limitation of deep learning systems, where internal decision processes are often difficult to interpret (Ribeiro et al., 2016).

4.3 Topological Feature Analysis

The integration of topological analysis provides a structural representation of complex data relationships. Traditional machine learning models generally analyse data through numerical feature spaces; however, such approaches may overlook important geometric properties of high-dimensional datasets.

The proposed framework represents data structures through topological features obtained using persistent homology:

$$PH_k(X) = \{(b_i, d_i)\}$$

where b_i and d_i represent the appearance and disappearance points of topological structures.

The resulting persistence information provides an additional representation layer that identifies stable patterns within datasets. Features that remain persistent across different scales are considered structurally meaningful, while short-lived features are treated as potential noise.

This mathematical property improves robustness because the model focuses on underlying data structures rather than individual variations. Therefore, the topological component complements fuzzy reasoning by providing structural awareness alongside interpretability.

4.4 Uncertainty-Aware Prediction Analysis

A major outcome of the proposed framework is the integration of uncertainty estimation directly into the prediction process. The uncertainty component enables the system to evaluate not only the predicted outcome but also the reliability of that outcome.

The total uncertainty is represented as:

$$U_{total} = U_a + U_e$$

where U_a represents uncertainty caused by inherent data variability and U_e represents uncertainty caused by limitations in model knowledge.

The confidence associated with prediction can therefore be expressed as:

$$C = 1 - U_{total}$$

A lower uncertainty value corresponds to higher prediction confidence, whereas increased uncertainty indicates that additional verification may be required.

This approach addresses a critical limitation of conventional AI models, where predictions may appear highly confident even when the input data are unfamiliar or ambiguous. Previous research has demonstrated that uncertainty estimation and calibration are essential components of reliable AI systems (Kendall & Gal, 2017; Guo et al., 2017).

4.5 Overall Framework Evaluation

The mathematical analysis demonstrates that the proposed framework provides a more comprehensive approach to trustworthy AI compared with conventional accuracy-oriented models. The integration of fuzzy logic, topology, and uncertainty modelling creates a system capable of producing predictions together with interpretable reasoning, structural understanding, and confidence estimation.

Table 1: The framework can be evaluated according to three primary trust dimensions

Component	Mathematical Contribution	Trustworthiness Outcome
Fuzzy Logic	Membership functions and inference rules	Improved interpretability

topological Analysis	Structural feature extraction	Enhanced pattern understanding
Uncertainty Modelling	Confidence estimation	Improved reliability

The proposed framework therefore provides a theoretical foundation for developing AI systems that move beyond prediction accuracy toward transparent and reliable decision-making. Unlike conventional black-box models, the proposed approach embeds trust-related characteristics within the mathematical formulation itself.

5. Discussion

The proposed fuzzy–topological uncertainty framework provides a mathematical approach for addressing major challenges associated with current artificial intelligence systems, particularly limited interpretability, insufficient uncertainty awareness, and difficulty in analysing complex data structures. The findings from the framework formulation indicate that trustworthy AI cannot be achieved through predictive accuracy alone but requires the integration of reasoning transparency, structural understanding, and reliability estimation. This observation is consistent with the broader trustworthy AI literature, which emphasizes that AI systems deployed in real-world environments should incorporate explainability, robustness, fairness, and accountability throughout their lifecycle (Li et al., 2021; Gunning & Aha, 2019).

A significant contribution of the proposed framework is the integration of fuzzy reasoning into AI decision-making. Traditional deep learning models often achieve high predictive performance but provide limited insight into how decisions are generated. The incorporation of fuzzy logic addresses this limitation by representing uncertain information through interpretable membership functions and linguistic rules. This finding extends the original concept of fuzzy sets introduced by Zadeh (1965), where uncertainty and partial truth were mathematically represented rather than restricted to binary classifications. Previous studies have demonstrated that fuzzy-based approaches improve interpretability in intelligent systems by allowing decision processes to be expressed through understandable rules (Dubois & Prade, 2000). Therefore, the proposed framework provides a pathway for embedding explainability directly into AI architectures rather than relying exclusively on external explanation methods.

The integration of topological analysis represents another important contribution of this research. Conventional machine learning methods generally depend on feature-based representations; however, complex datasets often contain structural relationships that cannot be fully captured through traditional statistical approaches. The inclusion of Topological Data Analysis enables the identification of persistent structures and geometric relationships within high-dimensional data spaces. This approach is supported by Carlsson (2009), who demonstrated that topology provides additional information about the organization and structure of complex datasets. Recent developments in topological machine learning further indicate that geometric and structural information can enhance representation learning and improve understanding of complex AI models.

The uncertainty modelling component strengthens the reliability of the proposed framework by allowing AI systems to communicate confidence levels associated with their predictions. Existing AI systems frequently generate predictions without adequately indicating whether those predictions are reliable under uncertain conditions. Kendall and Gal (2017) demonstrated that uncertainty estimation is essential for understanding model reliability by separating uncertainty caused by data variability from uncertainty caused by limited model knowledge. Similarly, Guo et al. (2017) showed that modern neural networks often produce poorly calibrated confidence estimates. The proposed framework addresses this limitation by incorporating uncertainty estimation as an integral part of the prediction process, enabling users to distinguish between reliable and uncertain outputs.

The proposed approach also differs from conventional Explainable AI methods that primarily rely on post-hoc interpretation techniques. Approaches such as LIME introduced by Ribeiro et al. (2016) provide valuable insights into model predictions; however, explanations are generated after the underlying model has already been trained. In contrast, the proposed framework attempts to incorporate interpretability within the mathematical formulation itself through fuzzy reasoning and topological representation. This shift from explanation after prediction toward explanation during prediction represents an important direction in trustworthy AI research (Doshi-Velez & Kim, 2017).

Despite its potential contributions, several limitations should be acknowledged. First, the integration of multiple mathematical components may increase computational complexity compared with conventional AI models. Second, the selection of fuzzy membership functions and topological representations may influence framework performance and requires further optimization. Third, although the current study provides a mathematical foundation, empirical validation using benchmark datasets and real-world applications is necessary to evaluate predictive performance, scalability, and practical effectiveness.

Future research should focus on implementing the proposed framework with modern deep learning architectures and evaluating its performance across different application domains. Integration with neural networks, reinforcement learning, and domain-specific AI systems may further improve its practical applicability. Additionally, adaptive mechanisms could be developed to allow fuzzy rules, topological structures, and uncertainty estimation methods to evolve dynamically with changing data environments.

Overall, the proposed framework contributes to the development of trustworthy AI by integrating three essential dimensions: interpretable reasoning, structural data understanding, and uncertainty-aware decision-making. By embedding these principles into a unified mathematical architecture, the research provides a foundation for developing AI systems that are not only accurate but also transparent, reliable, and suitable for human-centred decision environments.

6. Conclusion

This research presents a novel mathematical framework for developing trustworthy artificial intelligence by

integrating fuzzy logic, topology, and uncertainty modelling. The proposed approach addresses key limitations of conventional black-box AI systems by incorporating interpretability, structural understanding, and confidence estimation into the model architecture. Fuzzy reasoning improves transparency, topological analysis captures complex data relationships, and uncertainty modelling enhances prediction reliability. Although further computational validation is required using benchmark datasets and real-world applications, the proposed framework provides a strong mathematical foundation for future trustworthy AI development. This approach may support safer, more explainable, and human-centred AI systems across diverse domains.

Conflict of Interest

The authors declare that there are no conflicts of interest associated with this research. The authors have no financial, personal, or professional relationships that could have influenced the design, analysis, interpretation, or reporting of the findings presented in this manuscript.

Author Contributions

Conceptualization, methodology development, data analysis, and machine learning framework design was performed by the Parul. Manuscript preparation, visualization, and interpretation of results was carried out by the Parul and Dr. Vishal. Both authors reviewed, revised, and approved the final version of the manuscript.

References

- Carlsson, G. (2009). Topology and data. *Bulletin of the American Mathematical Society*, 46(2), 255–308. <https://doi.org/10.1090/S0273-0979-09-01249-X>
- Doshi-Velez, F., & Kim, B. (2017). Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608*. <https://arxiv.org/abs/1702.08608>
- Dubois, D., & Prade, H. (2000). *Fundamentals of fuzzy sets*. Springer. <https://doi.org/10.1007/978-1-4615-4429-6>
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. <https://www.deeplearningbook.org/>
- Gunning, D., & Aha, D. W. (2019). DARPA's Explainable Artificial Intelligence (XAI) program. *AI Magazine*, 40(2), 44–58. <https://doi.org/10.1609/aimag.v40i2.2850>
- Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 1321–1330.
- Kendall, A., & Gal, Y. (2017). What uncertainties do we need in Bayesian deep learning for computer vision? *Advances in Neural Information Processing Systems*, 30, 5574–5584.
- Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, X., & Zhou, B. (2021). Trustworthy AI: From principles to practices. *ACM Computing Surveys*. <https://doi.org/10.1145/3502698>

- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). “Why should I trust you?” Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. <https://doi.org/10.1145/2939672.2939778>
- Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338–353. [https://doi.org/10.1016/S0019-9958\(65\)90241-X](https://doi.org/10.1016/S0019-9958(65)90241-X)
- Zadeh, L. A. (1978). Fuzzy sets as a basis for a theory of possibility. *Fuzzy Sets and Systems*, 1(1), 3–28. [https://doi.org/10.1016/0165-0114\(78\)90029-5](https://doi.org/10.1016/0165-0114(78)90029-5)