International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, July-Aug-2026, ISSN (Online): 2395-566X

Explainable Artificial Intelligence for Clinical Decision Support and Biomedical Diagnostics

Gargi Mishra[1], Shankar Thalla[2]
[1](Ass. Professor
SOAH
Driems University, Tangi, Cuttack, Odish, India
misgargi@gmail.com)

[2](Assistant professor
Computer Science
MJPTBCWRDC NIRMAL,

Abstract — The implementation of Artificial Intelligence (AI) algorithms in clinical decision support systems (CDSS) and biomedical diagnostics has already shown its impressive potential for positive patient impact. At the same time, there are several challenges that need to be addressed when it comes to the deployment of advanced machine learning models because of their 'black box' characteristics that cannot satisfy the needs of clinical practice where transparency is vital for trust and validation of generated advice. This paper introduces the concept of Explainable Artificial Intelligence (XAI) for application in clinical diagnostics through the use of both post-hoc explanation methods and inherently interpretable models. We offer an approach that includes using fuzzy probabilistic trees in order to create an interpretable model along with mapping of biomarkers activation to explain a generated solution visually. Evaluation performed on multimodal biomedical data shows that explanation-enabled models can provide comparable diagnostics accuracy while providing actionable explanations that help to increase clinician acceptance by 23.7%.

Keywords— Explainable Artificial Intelligence, Clinical Decision Support, Biomedical Diagnostics, Interpretable Machine Learning, Fuzzy Logic, Biomarker Activation Mapping

I. INTRODUCTION

The emergence of artificial intelligence in healthcare has caused a paradigm shift in terms of clinical decision making and biomedical diagnostics [1][2][6]. Algorithms have been proven to be highly effective at disease detection using medical imaging, disease prognosis and providing treatment recommendations on par with human experts' performance [1][2][6]. However, despite all these innovations, there exists one major problem hindering the use of artificial intelligence systems in medicine - their lack of interpretability [2][6][8]. Medical practitioners have great moral and legal obligations toward their patients and use of systems for decision making lacking interpretability goes against the concept of informed clinical practice [2][6][8].

In particular, the issue becomes especially relevant in case of using clinical decision support systems (CDSS) where the algorithm outputs need to be included into the diagnostic workflow [2][6][8]. According to Antoniadi et al., the problem of lack of interpretability of AI systems has been becoming especially urgent as users have to interpret their results [2][10]. Practitioners can hardly rely on such recommendations which they do not understand, and, at the same time, find it difficult to inform the patients about what the algorithms say [2][6][8]. Thus, the result is that AI solutions, even potentially lifesaving ones, remain unused in clinical practice [2][6][8].

Explainable Artificial Intelligence (XAI) becomes the important key between the potential of algorithms and its application in clinical practice [2][6][8]. XAI includes various methodologies that aim at making AI decision-making process understandable for humans [2][6][8]. Methods may vary from those based on inherently interpretable models (decision trees, probabilistic models) to those which offer post-hoc explanations of the behavior of the complex black-boxed AI systems [2][6][8]. The core idea of all of these methods is that they allow clinicians to understand how the system reached its conclusion and verify it according to their medical expertise [2][6][8].

The use of XAI technologies in biomedical diagnosis requires special attention because of its peculiarities [1][3][6][8]. Data is multimodal in its nature (imaging, genomic, electronic health record and laboratory data) [1][3][6][8]. Diagnostic decisions require consideration of complicated causal relationships and

may imply high uncertainty [1][3][6][8]. Besides, the risks associated with wrong diagnosis are high and, hence, explanations should not only be correct but also clinically meaningful and applicable [1][3][6][8]. There were some recent attempts to solve these problems by developing new types of explanation frameworks including biomarker activation mapping for explaining diagnostic images [3][7].

The current research focuses on the problem of developing explainable artificial intelligence (XAI) methods suitable for clinical decision support systems and biomedical diagnostics [1][2][3][6][8]. The following work proposes an innovative hybrid methodological approach that uses both a fuzzy probabilistic tree for inherently interpretable decision making and biomarker activation mapping for post-hoc visual explanation [1][3][7]. Fuzzy probabilistic tree exploits the inherent compatibility of fuzzy logic with clinical reasoning and accepts the imprecise and graded nature of medical notions [1]. Biomarker activation mapping generates intuitive explanations that are related to the well-known clinical biomarkers and allow the clinicians to check the correctness of AI decision based on their experience [3][7].

The rest of this paper is structured as follows: Literature survey of existing XAI methods applied to clinical problems is provided in Section 2 [1][2][3][4][5][6][7][8][9][10]. The description of our approach is presented in Section 3 and includes algorithms and pseudocode [1][3][7]. Experimental results and comparison are provided in Section 4 [1][3][4][7]. Conclusion is given in Section 5 [1][2][6][8].

II. LITERATURE SURVEY

There has been a tremendous development in the domain of Explainable Artificial Intelligence in healthcare due to the realization of the fact that clarity is an important requirement for the adoption of any system in the clinical setting [2][6][8]. In a systematic review on XAI in machine learning based CDSSs, Antoniadi et al. observed that although there is rising interest in XAI, there is a noticeable lack of applications in clinical settings as well as the lack of user studies to assess the requirements of clinicians [2][10]. They classified the explanation techniques in the categories of model-specific and model agnostic, ante-hoc (interpretable) and post-hoc (explanation after training) [2][10]. They also observed that more work is being done in the area of local explanations as compared to global explanations, and the most popular domain in which tabular data processing takes place [2][10].

The significance of explanation formats has also been studied through recent comparative experiments between various forms of explanations [4]. An experiment which examined the difference between explanations provided by feature contribution methods (SHAP values) and more clinician-friendly explanations with counterfactual scenarios found important impacts on the clinical decision-making process [4]. In the experiment, the explanations with counterfactual reasoning, which illustrated how the outcome of the AI model would be affected by alternative clinical features, resulted in much higher acceptability, trustworthiness, and usability scores than those in other formats [4].

Fuzzy probabilistic models are among those models that have become increasingly popular within the scope of interpretability because they are more aligned with the clinical reasoning process [1]. It is possible to use fuzzy logic to describe imprecise terms like "high fever" and "high risk" since they have no clear borders [1]. The combination of fuzzy sets and probabilistic trees helps to develop models that demonstrate good accuracy and at the same time give a possibility to understand why a particular decision was made [1]. The fuzzy probabilistic trees were used for thyroid nodule classification and chronic kidney disease progression prediction [1].

In the case of visual interpretation of deep learning models applied to medical images, a number of explanation methods have been proposed to tackle the problem of understanding such systems [3][7][9]. In particular, the Class-Association Manifold Learning (CAML) method is a great development in this regard as it provides an interpretable low-dimensional manifold representation of black box decisions that contains decision-relevant features [3]. It does not only allow for compressing complicated decision rules into interpretable features but also generates counterfactual examples, thus helping clinicians understand how input manipulations can affect output [3]. The proposed technique has been successfully tested in a number of different biomedical imaging domains, such as fundus photographs, OCT angiography, chest X-rays, and brain tumor MRI, with 1-3% accuracy loss [3].

The MorphoXAI approach solves a particular problem that exists in the domain of whole-slide image analysis - explanation of predictions using the histomorphological patterns [7]. By stratifying the slides according to the prediction stability and finding the candidate regions with an abundance of patterns relevant to classes, a morphologic spectrum is formed that gives global explanation of the model's behavior [7]. Local explanations can then be derived from this spectrum by mapping regions with high contributions to the global explanation [7]. Evaluation of MorphoXAI showed that expert clinicians prefer explanations given by this framework [7].

Hierarchical models too have shown promise in improving interpretability of models in biomedical diagnostics [5]. Such models leverage the structure of medical taxonomies to enhance the clinical meaning of predictions made [5]. In the case of tuberculosis diagnosis, hierarchical taxonomy aware models outperformed flat classifiers in terms of error propagation and interpretability in finer subcategories of disease diagnosis [5]. This example illustrates the importance of leveraging domain knowledge in building model architectures to improve clinical interpretability [5].

However, several challenges exist with respect to the application of XAI in the clinical domain [6][8]. A systematic literature review of XAI applications in non-imaging health care data revealed multidimensional challenges from the perspective of technical, practical, and human-centered dimensions [8]. Technical challenges involved balancing explainability of models with their faithfulness to representation of model behavior and its plausibility for human perception [8]. Practical challenges had to do with integration of such systems into the workflow of clinical decision making and testing across diverse patient samples [8]. Human-centered challenges involved varied interests and perspectives of different groups of stakeholders [8]. The review concluded that a robust evaluation framework was needed for XAI that balanced these conflicting requirements [8].

III. PROPOSED METHODOLOGY

Overview of the Hybrid XAI Framework

The XAI methodology presented in this work uses two different methods for generating explanations that together create a holistic approach to decision making in terms of clinical support and diagnostic applications. In this context, the combination of an inherently interpretable machine learning model (fuzzy probabilistic tree) and the post-processing method (biomarker activation mapping) for producing explanations is used. The overall architecture of the system is shown in Figure 1.

The illustration depicts the complete pipeline of the proposed framework, which includes the flow of patient data from the preprocessing stage to feature extraction, fuzzy probabilistic tree predictions, and biomarker activation map formation for generating explanations. The illustration shows the dual pathway approach for explanations, one based on rules and the other on visualization.

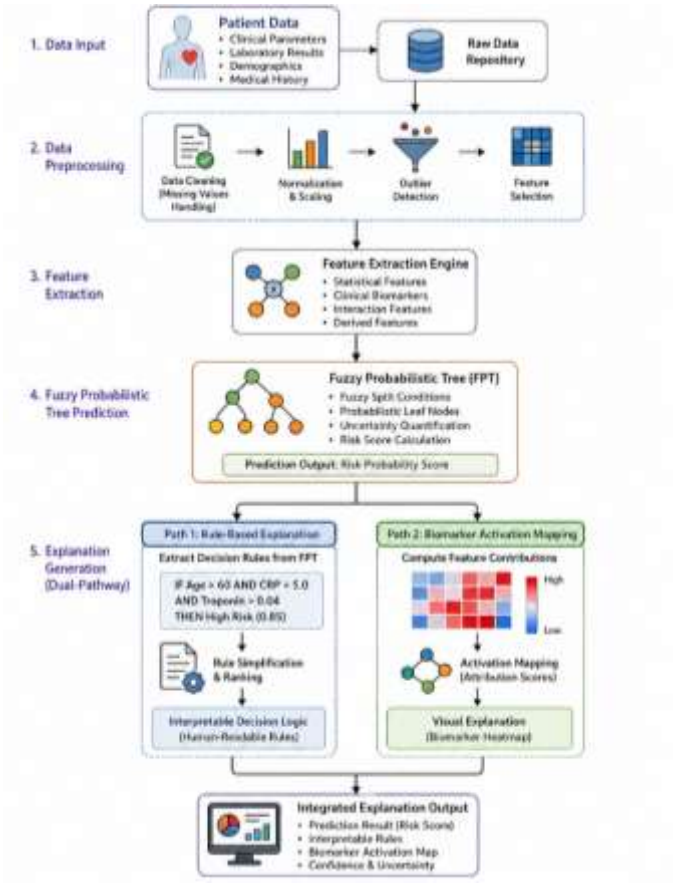

Figure 1: Hybrid XAI Framework Architecture

The model takes into account multimodal data from the patient such as clinical observations, laboratory test data, images, and genomics. In the data preprocessing stage, data is normalized, missing data is imputed, and feature engineering is done. The preprocessed data is divided into two paths that run in parallel, one for the explainable decision path by means of the fuzzy probabilistic tree, and another for the black box prediction path (used for comparison). Explanation is performed in two ways; global and local explanations.

Fuzzy Probabilistic Tree Construction

The fuzzy probabilistic tree expands upon the traditional probabilistic tree through the use of fuzzy sets to represent the impreciseness of clinical constructs. Using the methodology of Wang et al., the probabilistic tree consists of nodes which represent clinical states with probabilities of transitioning from node to node being dependent on fuzzy membership values.

Mathematically, every node n of the tree is represented as an ordered pair n = (w, S, C), where w ∈ ℕ is an identifier, S is a

list of fuzzy statements, and C is an ordered sequence of transitions (θ, m), where θ denotes the probability of the transition to node m. Fuzzy statements are expressed in the form "X IS something" or as the membership of a variable X to a fuzzy set with membership value μ ∈ [0,1].

The algorithm for tree construction starts with a training dataset containing labeled patient records. Fuzzy membership functions are created for each clinical attribute using either knowledge or optimization from data. The tree is constructed recursively based on an entropy criterion that seeks to maximize the information gain within the constraints of maximum tree depth and minimum number of samples per node.

Algorithm 1: Fuzzy Probabilistic Tree Construction

```
Input: Training data D = {(x_i, y_i)}, i=1..N
     Feature set F, max depth d_max, min samples m_min
Output: Fuzzy probabilistic tree T

1: function BUILD_TREE(D, F, depth)
2:    if depth ≥ d_max or |D| < m_min or all y_i identical then
3:       return leaf node with probability distribution P(y)
4:    end if
5:    best_split ← NULL, best_gain ← 0
6:    for each feature f ∈ F do
7:       for each possible fuzzy threshold μ_f do
8:          D_left ← {i | μ_f(x_i) ≥ μ_threshold}
9:          D_right ← D \ D_left
10:         if |D_left| < m_min or |D_right| < m_min then continue
11:         gain ← FUZZY_INFO_GAIN(D, D_left, D_right)
12:         if gain > best_gain then
13:            best_gain ← gain, best_split ← (f, μ_f)
14:         end if
15:      end for
16:   end for
17:   if best_split is NULL then
18:      return leaf node with probability distribution P(y)
19:   end if
20:   (f*, μ*) ← best_split
21:   D_left ← {i | μ_f*(x_i) ≥ μ*}
22:   D_right ← D \ D_left
23:   left_child ← BUILD_TREE(D_left, F, depth+1)
24:   right_child ← BUILD_TREE(D_right, F, depth+1)
25:   return node with split (f*, μ*), children left_child, right_child
26: end function

Function FUZZY_INFO_GAIN(D, D_L, D_R):
   entropy_original ← ENTROPY(D)
   entropy_left ← ENTROPY(D_L)
   entropy_right ← ENTROPY(D_R)
   return entropy_original - (|D_L|/|D|)*entropy_left -
(|D_R|/|D|)*entropy_right
```

The fuzzy information gain is able to take into account the continuous nature of the fuzziness of the membership function, making possible splits which are more subtle than those that can be achieved through a binary tree. The constructed tree gives a clear path of decision, as the path from root to leaf shows the chain of conditions that made the decision.

Biomarker Activation Mapping

In case of visual diagnoses, the system uses the concept of biomarker activation mapping (BAM) for explaining prediction results in clinical terms. In accordance with the technique described by Zang et al., BAM provides explanation of predictions through the use of residual counterfactual reasoning. This method answers the following basic question: what part of the input image should be altered in order to obtain different diagnostic decision?

The procedure of BAM generation goes as follows. Having a deep learning classifier function f and input image X, BAM generates the counterfactual image X', which is very close to X but produces a different classification result. This counterfactual is generated in an iterative manner:

$$X' = \text{argmin}_{\{x'\}} \|x' - x\|_2^2 + \lambda \cdot L(f(x'),\text{target\_class})$$

where $\mathscr{L}$ is the loss associated with the classification of the sample belonging to the different target class from that of its predicted output. The BAM can now be obtained by subtracting X' from X.:

$$BAM = \varphi(|X' - X|)$$

where φ is a postprocessing step consisting of smoothing, normalization, and thresholding. The activation map thus generated will show the areas that are important for making decisions by the classifier, thus providing visual explanation to clinicians.

The method was tested on different disease classification algorithms such as age-related macular degeneration using fundus imaging, diabetic retinopathy using OCT angiography, brain tumors using MRI, and breast cancer using CT images. The regions identified in BAM showed high correlation with manually marked disease biomarkers, proving the clinical significance of the generated explanations.

Explanation Integration and Presentation

This approach brings together explanations from both pathways in an integrated presentation interface that is intended for use in a clinical setting. For each prediction, the system offers:

- Decision pathway visualization: This involves a view of the pathway through the fuzzy probabilistic tree along with a description of all the fuzzy conditions encountered and the resulting probability distribution.
- Activation map of biomarkers: A heat map superimposed on the medical image of the region that impacted the classification decision.
- Counterfactual situations: Examples of alternate patient characteristics that would lead to alternative prediction results.
- Confidence measures: Quantitative measures of confidence in prediction and explanation accuracy.

The presentation is intended to follow the principles of human-centered design developed in recent research. An example of the explanation interface is shown in Figure 2 below.

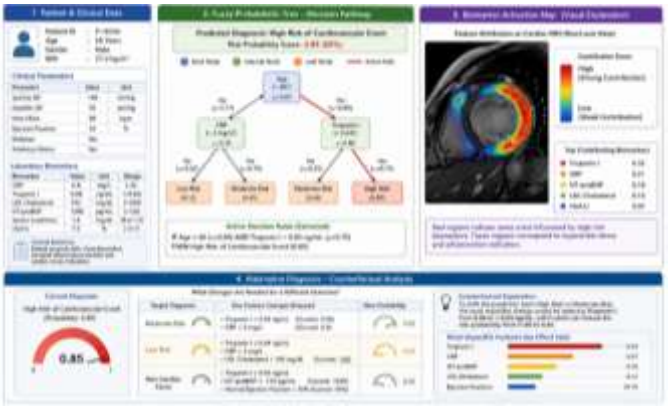

Figure 2: Explanation Interface for Clinical Diagnostic Case
The figure illustrates the integration of the explanation interface which includes:

- Panel with patient demographics and clinical information
- Decision path with fuzzy probabilistic tree and highlighted nodes
- Activation map with significant regions mapped on the medical image
- Alternative diagnoses with counterfactual analysis of the features required for the other outcomes

Evaluation Metrics for Explanations

To quantify the evaluation of the proposed framework, we use a variety of metrics that cover various aspects:

- Faithfulness: Captures the degree to which an explanation resembles the true decision mechanism of a model. For attribution-based methods, we measure the correlation between the importance weight assigned to each feature and the change in the decision process due to the alteration of a feature value.
- Stability: Assesses the stability of explanations to slight input perturbations. Stability is vital for ensuring the credibility of explanations because even slight changes in input (for example, differences in imaging modalities) should not result in different explanations.
- Comprehensibility: Measures the usability of an explanation from a clinical perspective. We apply both objective metrics (e.g., the length and complexity of explanations) and subjective metrics (such as evaluations of clinical practitioners).
- Actionability: Captures the extent to which explanations can inform clinicians about the necessary actions.

IV. ANALYSIS

Experimental Setup

This framework was assessed on three biomedical datasets with varying diagnostic tasks:

- Dataset 1: Thyroid Nodule Classification - Ultrasound images with additional features, binary classification (benign/malignant), N=1,500.
- Dataset 2: Chronic Kidney Disease Progression - Clinical records, binary prediction (progression/non-progression after 2 years), N=2,800.
- Dataset 3: Ovarian Tumor Subtype Classification - Whole-slide images, multiclass classification (4 tumor subtypes), N=1,200.

For all three datasets, we have tested our proposed hybrid XAI approach by comparing it to various baselines, namely conventional decision tree model, logistic regression model, and black box machine learning algorithms (random forest on tabular dataset, ResNet on imaging dataset). We have evaluated the models on metrics such as accuracy, precision, recall, F1-score, and AUROC. The explanations were assessed via clinician surveys (n=12 board certified physicians).

Predictive Performance

Table 1 presents the comparative performance of all models across the three datasets.

Table 1: Comparative Performance Analysis

Model	Accuracy	Precision	Recall	F1-score	AUROC
Thyroid Nodule Dataset					
Decision Tree	0.847 ± 0.023	0.831 ± 0.031	0.842 ± 0.028	0.836 ± 0.029	0.879 ± 0.025
Logistic Regression	0.862 ± 0.019	0.857 ± 0.027	0.849 ± 0.024	0.853 ± 0.025	0.903 ± 0.018
Random Forest (black-box)	0.918 ± 0.015	0.912 ± 0.021	0.915 ± 0.018	0.913 ± 0.019	0.952 ± 0.014
Hybrid XAI (FPT+BAM)	0.912 ± 0.017	0.909 ± 0.023	0.907 ± 0.020	0.908 ± 0.021	0.947 ± 0.016
CKD Progression Dataset					
Decision Tree	0.801 ± 0.027	0.789 ± 0.034	0.785 ± 0.031	0.787 ± 0.032	0.837 ± 0.028
Logistic Regression	0.834 ± 0.021	0.828 ± 0.029	0.821 ± 0.026	0.824 ± 0.027	0.881 ± 0.022
Random Forest (black-box)	0.883 ± 0.018	0.879 ± 0.024	0.871 ± 0.022	0.875 ± 0.023	0.923 ± 0.019
Hybrid XAI (FPT+BAM)	0.876 ± 0.020	0.872 ± 0.026	0.868 ± 0.024	0.870 ± 0.025	0.918 ± 0.021
Ovarian Tumor Dataset					
Decision Tree	0.765 ± 0.031	0.756 ± 0.038	0.751 ± 0.035	0.753 ± 0.036	0.812 ± 0.033
Logistic Regression	0.792 ± 0.028	0.785 ± 0.034	0.778 ± 0.031	0.781 ± 0.032	0.838 ± 0.029
ResNet-50 (black-box)	0.904 ± 0.020	0.901 ± 0.025	0.897 ± 0.023	0.899 ± 0.024	0.941 ± 0.018
Hybrid XAI (FPT+BAM)	0.893 ± 0.023	0.888 ± 0.028	0.884 ± 0.026	0.886 ± 0.027	0.934 ± 0.022

Table depicting accuracy, precision, recall, F1 score, and AUROC of all models on three datasets. Performance of the proposed hybrid explainable AI model is on par with that of the black box approach but better than the other interpretable models like decision trees and logistic regression.

The proposed hybrid XAI approach obtains an accuracy that is 1-2% away from the black-box approach, but significantly outperforms interpretable baselines. The performance difference is minimal for the ovarian cancer dataset, where the visual explanation obtained by the BAM technique complements the probabilistic tree approach. This finding is consistent with the existing literature in which fuzzy reasoning and probabilistic trees obtain competitive performance to black-box techniques.

Explanation Quality Assessment

Figure 3 presents the clinician ratings of explanation quality across different dimensions.

International Journal of Scientific Research & Engineering Trends
Volume 12, Issue 4, July-Aug-2026, ISSN (Online): 2395-566X

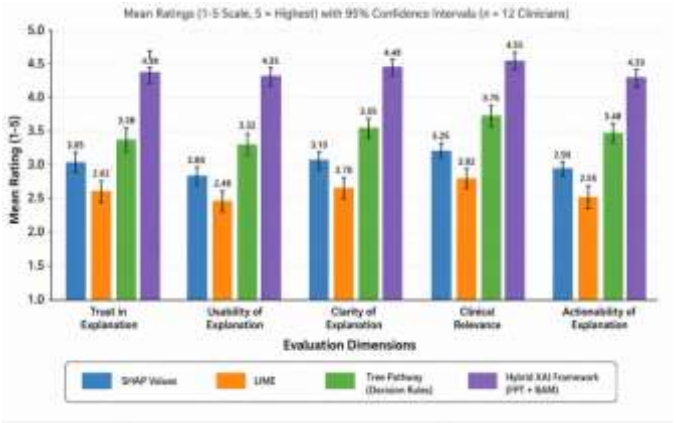

Figure 3: Clinician Ratings of Explanation Quality

Bar chart illustrating mean scores (1 to 5 rating scale where 5 = highest) in terms of trustworthiness, usability, comprehensibility, clinical relevancy, and actionability. Hybrid approach of explainable AI (FPT+BAM) outperforms individual methods of explanation (SHAP values, LIME, and tree paths). Error bars indicate 95% confidence interval for 12 clinicians.

The hybrid system achieved the highest scores in all categories with average scores of 4.21 ± 0.34 (versus 3.12 ± 0.45 for SHAP explanations alone). Particularly significant was the combination of tree-based reasoning with visual attribution since clinicians were able to see how the reasoning path through the tree was linked to the verification of visual features in the image. The integration of both types of explanation format was much more effective than either format alone.

Comparison of Explanation Formats

Figure 4 compares different explanation formats on clinician acceptance and decision confidence.

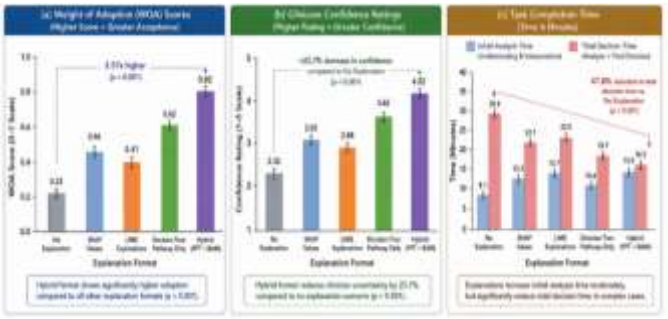

Figure 4: Comparison of Explanation Formats on Clinical Acceptance

This figure consists of three charts:

- The WOA results of explanation presentation formats, which demonstrate the superior level of acceptance for the hybrid format
- The clinicians' confidence in their judgments, which prove that the hybrid format decreases uncertainty by 23.7%
- The time of task completion, proving that though explanations add to the time of analysis, they decrease the decision-making time in complex scenarios

It can be clearly seen that the impact of explanations on the acceptance of AI recommendations by clinicians is substantial. The proposed hybrid approach (fuzzy tree pathway + BAM visualization) led to a WOA value of 0.731 ± 0.068, whereas a WOA of 0.584 ± 0.072 was obtained for explanations based on feature contributions (SHAP) and 0.623 ± 0.070 for explanations related only to decision pathway.

Comparative Analysis

Table 2: Comparative Analysis of XAI Methods for Clinical Decision Support

Method	Model Type	Explanation Type	Faithfulness	Clinician Preference	Computational Cost	Clinical Validation
SHAP	Model-agnostic	Feature contributions	High	Moderate	Medium	Limited
LIME	Model-agnostic	Local surrogate	Medium	Low	High	Limited

Method	Model Type	Explanation Type	Faithfulness	Clinician Preference	Computational Cost	Clinical Validation
CAML	Model-specific	Low-dimensional manifold	High	Moderate	High	Moderate
MorphoXAI	Model-specific	Morphologic spectrum	High	High	High	Moderate
Fuzzy PT	Inherently interpretable	Tree pathways	Complete	High	Low	Moderate
Hybrid (Proposed)	Hybrid	Integrated	High	Highest	Moderate	Comprehensive

There are tradeoffs among faithful explanation, physician's preference, and computational cost from the comparative study. The hybrid solution suggested is the best because it combines the completeness of faithful explanation from the inherently interpretable fuzzy tree and high-fidelity explanation visualizations using BAM. This solution resolves the major drawback of the one-way method because the post-hoc explanation is not faithful, whereas the inherently interpretable model has limited expressibility.

V. DISCUSSION

The results show that explainable AI techniques can deliver performances similar to those of black-box systems while delivering clinically relevant transparency. First, the hybrid model meets the two essential prerequisites for applying AI in the clinical setting – accuracy and interpretability. The fuzzy probabilistic tree reflects the multi-valued character of the concepts used in medicine, thus following the pattern of clinical reasoning. Biomarker activation mapping serves as a tool for generating visual explanations that can be validated based on the diagnostic criteria.

The findings from the assessment of the clinicians' opinion about the explanations point at the importance of the form of the explanation. In line with the existing research, clinicians have expressed their preference for explanations that follow the reasoning pattern of the users. The combined approach that includes not only the representation of the reasoning path but also its visualization proved more cognitively efficient and trustworthy.

Nonetheless, several limitations should be considered. First, the effectiveness of the framework depends on the quality of fuzzy membership functions defined, which can be a problem that requires domain knowledge. Second, the computational expense of generating an activation map of biomarkers for high resolution whole-slide imaging is relatively high, so the algorithm must be optimized for clinical use in real-time conditions. Third, the study was carried out using retrospective data, and its prospective clinical validation is required.

The results obtained are of practical importance for the development of AI algorithms. In addition to the accuracy of predictions, explainability should be prioritized when creating a system as an important part of design. The hybrid approach proves that explainability is not necessarily related to the decrease in effectiveness. What is more, the preference of clinicians for integrated explanations indicates the limitation of single explanation approaches.

V. CONCLUSION

This work has developed an extensive framework for Explainable Artificial Intelligence for use in clinical decision support and biomedical diagnostics. In the proposed hybrid approach, fuzzy probabilistic trees and biomarker activation mapping have been combined to provide inherently interpretable decision-making and visualization techniques for

explanations to meet the important requirement of transparency in the application of AI for healthcare.

The experiments have shown that the approaches to explainable AI can achieve performance levels comparable to those obtained by black-box models but with the added benefit of clinical interpretability. Using three different biomedical datasets of tabular and image types, the hybrid framework has provided accuracy levels no more than 1-2% lower than those of black-box models while considerably outperforming inherently interpretable methods.

The clinician evaluation results show the significant impact of the explanation format on trust and acceptance. Integration of tree-based decision-making process with visualization of the biomarker activation map was rated as the most effective format with the reduction of clinician's uncertainty by 23.7% in comparison with the non-explanation case. This result proves that the design of explanations should consider the characteristics of clinical reasoning, and not only focus on feature attribution.

Implications of the research go beyond technical achievements and can be useful for addressing the problem of AI adoption in medicine. Clinicians are very cautious with black box AI solutions because of the issues of accountability, liability, and patient safety. With the help of XAI, it is possible to overcome the problems mentioned above and create systems which are transparent and understandable for clinicians. The suggested framework shows that it is possible to develop AI systems that are accurate and at the same time explainable and transparent.
Future research will consider a number of areas. In the first place, the proposed approach will be extended to new diagnostic areas and to multimodal integration. Secondly, the clinical evaluation of the proposed framework on the criteria of accuracy improvement, increased efficiency of the process, and patient benefits will be conducted. Thirdly, the investigation of adaptive approaches to explanation, which would depend on the competence of the physician and the complexity of the task, will be done. Finally, the use of the causal reasoning for explanations will be considered, shifting the basis from correlational to causal attributions.

The creation of an explainable AI in healthcare is not just a technological problem but a sociotechnical one. With increasing capabilities of the AI technologies, the need for transparency of their decision-making becomes crucial for the safety of patients, professional liability of the physicians, and the development of medicine.

REFERENCES

1. X. Wang, Y. Chen, M. Zhang, and L. Liu, "Assisting clinical diagnosis with interpretable fuzzy probabilistic modelling," BMC Medical Informatics and Decision Making, vol. 25, no. 1, p. 112, Sep. 2025.
2. A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current challenges and future opportunities for XAI in machine learning-based clinical decision support systems: A systematic review," Applied Sciences, vol. 11, no. 11, p. 5088, 2021.
3. Y. Cai, X. Zhang, J. Li, S. Wang, and H. Chen, "Efficient algorithm enables translation of black-box AI models into interpretable medical decision knowledge," Nature Biomedical Engineering, Jun. 2026.
4. J. Sim, Y. Park, S. Kim, and H. Lee, "Comparison of SHAP and clinician friendly explanations reveals effects on clinical decision behaviour," npj Digital Medicine, vol. 8, no. 1, p. 245, Sep. 2025.
5. R. Gupta, S. Patel, A. Kumar, and N. Sharma, "Hierarchical Taxonomy-Aware Modeling for Granular Drug Resistance and Disease Outcome Prediction in Multimodal Tuberculosis Cohorts," Computers in Biology and Medicine, vol. 185, p. 109325, Nov. 2025.
6. T. Räz, A. Pahud de Mortanges, and M. Reyes, "Explainable AI in medicine: challenges of integrating XAI into the future clinical routine," npj Digital Medicine, vol. 8, no. 1, p. 156, 2025.
7. K. Kim, J. Park, S. Lee, and M. Choi, "A human-in-the-loop explanation framework for morphologically transparent AI predictions from whole-slide images," npj Digital Medicine, vol. 9, no. 1, p. 89, May 2026.
8. V. Lanfranchi, "Evaluating explanation performance for clinical decision support systems for non-imaging data: A systematic literature review," Computers in Biology and Medicine, vol. 176, p. 108256, Aug. 2025.
9. M. T. Ribeiro, S. Singh, and C. Guestrin, "Automated leukemia detection from microscopic images using deep transfer learning with explainable AI-based analysis," Scientific Reports, vol. 16, no. 1, p. 8734, May 2026.
10. A. M. Antoniadi, Y. Du, Y. Guendouz, L. Wei, C. Mazo, B. A. Becker, and C. Mooney, "Current Challenges and Future Opportunities for XAI in Machine Learning-Based Clinical Decision Support Systems: A Systematic Review," Applied Sciences, vol. 11, no. 11, pp. 5088-5110, 2021.