

An Intelligent Data Mining Framework for Predictive Decision-Making Using Explainable Artificial Intelligence

Mr. R K Arunkumar¹, S. Inbaroja²

¹(Assistant Professor
Computer Science (PG)
Kongu Arts and Science College (Autonomous)
Nanjanapuram , Erode - 638107
rkarun@kasc.ac.in)

²(Research Scholar
COMPUTER SCIENCE
St. Xavier's College (Autonomous)
Palayamkottai, Tirunelveli
inbaroja86@gmail.com)

Abstract — Integration of artificial intelligence into vital decision-making fields has led to the emergence of a demand for transparency and interpretability in the used predictive models. This research aims to present an intelligent data mining framework, which utilizes the combination of predictive modeling methods and Explainable Artificial Intelligence (XAI). The use of the suggested framework is meant to ensure transparent decision-making process by implementing such methods of predictive modeling as ensemble learning, including Random Forest, XGBoost, and Gradient Boosting along with the methods of SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) for multi-level explanation generation. Our research applies the intelligent data mining framework on the dataset related to healthcare and includes several classes of diseases, ensuring high performance of the model in terms of prediction along with generating actionable explanations for each prediction.

Keywords— Explainable Artificial Intelligence, Data Mining, Predictive Analytics, SHAP, LIME, Decision Support Systems, Interpretability

I. INTRODUCTION

The use of artificial intelligence in essential industries, including medicine, banking, and autonomy, has generated a unique need for models that not only perform well but also have clear logic behind their work [1][8]. Predictive models based on machine learning algorithms have proven themselves very effective, yet the issue of their non-transparent decision-making process becomes one of the greatest barriers to adoption in particular application areas [3][4]. These problems include skepticism among stakeholders, regulatory challenges, and the impossibility to recognize and prevent potential bias when reasoning behind the prediction remains unclear [1][8].

The conflict of goals in predictive performance and explainability has given rise to the new field of research, which is called Explainable Artificial Intelligence (XAI) [8]. While traditional interpretation of the problem was focused on the necessity to sacrifice one quality for another and considered

models like deep neural networks or gradient boosting machines as the black boxes, recent developments prove that this trade-off is not inevitable [1][5]. Indeed, there exist effective solutions that allow getting insightful explanations while retaining predictive accuracy [9]. However, the issue is to develop an integrated framework for these solutions' use in practice [4][5].

Indeed, the nature of data mining science itself implies the extraction of insights, which are actionable for decision making [2]. The integration of XAI principles into data mining practice seems to be quite logical and natural, as it turns a mere technique for predictions into the means of decision-making, which is not just capable of generating predictions but also providing the explanations needed to act based on them [2][4]. As mentioned above, such an approach is especially vital for areas, where the decision has important consequences [3][6]. Indeed, whether it be a decision about a patient's diagnosis, the

loan granting or allocation of resources, such decision needs not only to be accurate but justified as well [3][7].

The paper aims at the development of intelligent data mining frameworks, which will consider explainability as an integral part of the process. The contributions of our work are three-folded. First, we design a modular architecture, which splits the processes of predictive modeling, explanation and decision-making into different components [1][5]. Second, we show applicability of the proposed framework on a multi-class predictive task in healthcare, demonstrating high performance as well as ability to provide actionable explanations [3][6]. Finally, we perform comparative analysis of different explanation techniques [5][9].

The rest of this paper is structured as follows: Related work on Explainable AI (XAI) and predictive data mining is reviewed in Section 2, the proposed approach and architecture is introduced in Section 3, experimental results and comparison are presented in Section 4, and conclusions and future research perspectives are outlined in Section 5.

II. LITERATURE SURVEY

Explainable AI for data mining has followed many paths in terms of development since there are many different methodologies and applications in this fast-developing area. The paper by Ali et al. presents an elaborate taxonomy of XAI approaches according to four aspects: data explainability, model explainability, post-hoc explainability, and explanation evaluation [1]. The analysis of 410 important papers in the literature shows that even though most explanations that were created in practice used post-hoc methods, there is a lack of research on methodology evaluation and usability issues [1]. It should be noted that explanations should be developed in consideration of stakeholders' requirements, which has great practical significance [1][8].

From the data-oriented perspective, Xiong et al. suggest understanding the concept of explainability not only as the property of the models themselves but as an emergent characteristic of the whole data lifecycle [2]. By reviewing feature attributions, effects of training data on the models, and discovering domain knowledge, the authors identify data mining procedures such as feature engineering, sampling, and data preprocessing as the core aspects of the explanation quality [2]. From the point of view of practical experience, the selection of features, missing value imputation, and class imbalance affect what can be explained about the model [2].

Healthcare sector has become one of the active arenas for application of XAI methods for predictive modeling. As a part of their systematic review, Alkhanbouli et al. identify disease prediction as the most frequent application area with SHAP and LIME being the two most popular XAI explanation methods [3]. At the same time, the authors observe that even though many of the examined papers report high predictive performance of the models, very few of them have a proper assessment of the explanations' quality and how they are used to improve clinical decision-making [3].

In their paper on process mining prediction tasks, Mehdiyev et al. consider uncertainty quantification and privacy-preserving explanation to be important research areas in XAI [4]. This study discusses the potential of integrating the concept of uncertainty in order to increase the quality of explanations, where the information on confidence bounds is provided along with the attributions of features [4]. Also, this study mentions the emerging capabilities of the large language models to provide natural language explanations that allow for conversational interfaces [4][10].

The methodology of XAI has produced a diverse set of explanation methods, which are discussed in the work by Schwalbe and Finzel [5]. This study presents an overview of the methods from simple techniques based on feature importance to more complex counterfactual explanations and surrogate models [5]. Furthermore, this study distinguishes the model-specific methods, e.g., tree-based feature importance, from model-agnostic methods such as SHAP and LIME [5][9]. One of the important characteristics of the techniques is their fidelity, i.e., the capability of the explanation to represent the actual behavior of the model [5].

A number of papers use XAI for specific prediction tasks. Bhattacharya et al. use SHAP in order to predict chronic kidney disease and illustrate how the attribution of features allows one to discover key factors in the process [6]. Similarly, Sagi et al. use XAI in the context of predicting heart disease and show that local explanations can fill the gap between machine learning algorithms and clinical reasoning [7]. However, the studies reviewed above all highlight the limitations of current approaches: they tend to focus on binary classification problems and use single XAI techniques without comparing them [3][6][7].

In fact, the literature review reveals a pattern: although XAI techniques become widely available and computationally accessible, the development of end-to-end decision-making systems incorporating XAI is still limited [1][4]. In particular,

very few studies consider the entire pipeline: from preprocessing and building models to generating explanations and presenting results to the users [2][3]. Another limitation is the absence of any comparisons of different explanation techniques in terms of trade-offs among their properties [5]. This paper aims to overcome the mentioned limitations by developing a unified framework that combines different XAI techniques and compares their characteristics [1][5].

III. PROPOSED METHODOLOGY

Framework Architecture

The suggested Intelligent Data Mining Framework for Predictive Decision Making (IDMF-PDM) is made up of four interrelated layers namely, Data Acquisition & Preprocessing, Predictive Modeling, Explanation Generation, and Decision Support Interface. The architecture of IDMF-PDM is illustrated by Figure 1 below.



Figure 1: IDMF-PDM Framework Architecture

The system adheres to a modular design approach, whereby individual layers may be configured, modified, and extended independently of others. The data layer takes care of the input process from diverse sources, such as structured electronic health records and semi-structured clinical notes. The preprocessing layer is responsible for cleaning, normalizing, and creating features while balancing classes. Modeling layer allows for running of different algorithms, with ensemble algorithms acting as main prediction models. Explanations are generated at the explanation layer using SHAP and LIME.

Data Preprocessing and Feature Engineering

The raw data is subjected to pre-processing procedures to enhance its quality and consistency. The first step entails missing value handling through the use of median imputation for numerical attributes and mode imputation for categorical attributes. The process of outlier handling utilizes the interquartile range approach, with extreme values being capped or eliminated depending on domain specific thresholds.

The process of feature engineering involves conversion of raw features into meaningful ones. In relation to the numerical attributes, standardization is used to make them have zero means and unit variances. One-hot encoding and ordinal encoding are used to encode nominal and ordinal attributes respectively. Furthermore, we will create interaction features between relevant attribute pairs for the domain through correlation analysis and consultations with domain experts.

The problem of class imbalance prevalent in medical records will be dealt with through Synthetic Minority Over-sampling Technique (SMOTE). This procedure involves the creation of synthetic samples from the minority classes in order to improve the sensitivity of the model without losing information about majority classes.

Predictive Modeling

In the predictive modeling part, ensemble learning is used where multiple base learners are trained, and their outputs are combined through soft voting. In this part, we use Random Forest, XGBoost, and Gradient Boosting Machines as our base learners. By doing so, we can utilize the advantages of different algorithms in one model, where the Random Forest is resistant to overfitting, XGBoost is efficient and regularized, and Gradient Boosting is highly performant on structured data.

For hyperparameters tuning, Bayesian optimization is performed along with cross-validation. Parameters such as tree depth, learning rate, number of estimators, and regularization parameters are chosen from the search space, and the objective is to tune the hyperparameters for the optimal trade-off between predictive performance and complexity while avoiding very large ensemble models.

Pseudocode for the ensemble learning process is shown in Algorithm 1.

Algorithm 1: Ensemble Model Training

```

Input: X (features), y (labels), k (folds), N (iterations)
Output: Ensemble model M

1: Initialize best_score ← 0
2: Initialize best_config ← null
3: for i = 1 to N do
4:   config ← BayesianOptimizer.suggest()
5:   scores ← []
6:   for j = 1 to k do
7:     X_train, X_val ← StratifiedSplit(X, y, fold=j)
8:     model ← TrainEnsemble(X_train, y_train, config)
9:     score ← Evaluate(model, X_val, y_val)

```

```

10:   scores.append(score)
11:   avg_score ← Mean(scores)
12:   if avg_score > best_score then
13:     best_score ← avg_score
14:     best_config ← config
15:   end if
16: end for
17: M ← TrainEnsemble(X, y, best_config)
18: return M

```

Explanation Generation

For the explanation generation technique, two methods are applied: SHAP for both local and global feature attributions, and LIME for local surrogate models. Such an approach allows for looking at the behavior of the model from different angles and comparing the explanations.

The SHAP value calculation is performed with the help of the TreeExplainer method, allowing to compute Shapley values exactly for tree-based models in polynomial time. For each prediction, SHAP computes feature attributions that determine by how much each particular feature contributed to the difference between the expected and actual model outcomes. The process of generating local explanation involves using interpretable models that approximate the behavior of the target model in the locality of the data point under inspection. For example, in LIME, interpretable models are used in the form of sparse linear models to generate local explanations by fitting such models on perturbed samples near the instance of interest. The algorithm below provides a pseudo-code for explanation generation through integration. The explanation process will yield both SHAP and LIME explanations along with a consistency score.

Algorithm 2: Integrated Explanation Generation

Input: Model M, instance x, background data B, num_samples n
Output: Explanation E

```

1: // SHAP explanation
2: explainer_shap ← TreeExplainer(M, B)
3: shap_values ← explainer_shap.shap_values(x)
4:
5: // LIME explanation
6: explainer_lime ← LimeTabularExplainer(B)
7: lime_exp ← explainer_lime.explain_instance(x,
M.predict_proba, n)
8: lime_values ← ExtractCoefficients(lime_exp)
9:

```

```

10: // Consistency assessment
11: consistency ← ComputeRankCorrelation(shap_values,
lime_values)
12:
13: // Aggregate explanation
14: E ← {
15:   shap: shap_values,
16:   lime: lime_values,
17:   consistency: consistency,
18:   confidence: M.predict_proba(x).max()
19: }
20: return E

```

Figure 2 illustrates the explanation generation workflow, showing how SHAP and LIME explanations are computed in parallel and combined for presentation.



Figure 2: Explanation Generation Workflow

Decision Support Interface

This layer facilitates information derived from the models by translating predictions and explanations into useful data. The interface includes three different elements: the predicted class along with the confidence level, a ranking of the features involved in the prediction process and their respective impacts, and an explanation in a natural language format.

The module of generating natural language facilitates a description of the attribute of the features in human-understandable language. For instance, if there is a SHAP value which suggests that high glucose helps in predicting the disease of diabetes, it will be explained as "Elevated glucose levels increased the likelihood of diabetes classification."

Framework Integration

The entire pipeline works as a coherent pipeline. In training mode, data goes through the preprocessing stage, model training, and explanation calibration. In the inference phase, incoming examples go through preprocessing, classification,

and explainability in one pass. The pipeline makes available REST APIs that can be used to integrate with current clinical or enterprise systems.

IV. ANALYSIS AND DISCUSSION

Experimental Setup

The proposed framework is evaluated on a multiclass healthcare dataset that consists of patient records diagnosed under five classes of diseases: cardiovascular, respiratory, metabolic, neurological, and musculoskeletal diseases. There are 12,450 records in the dataset with 38 features including demographic features, vital signs, lab test results, and symptom features. Data splitting ratio for the three datasets is 70%, 15%, and 15% respectively for train, validate, and test data.

The baseline comparison includes various models such as individual classifiers like Random Forest, XGBoost, and Gradient Boosting. Baseline models also consist of logistic regression as an interpretable model and the ensemble model without any explanation component.

Predictive Performance Analysis

The performance of the individual models and the suggested ensemble model is illustrated in Figure 3. The ensemble model outperforms the others in terms of the highest values of F1-score, precision, and recall, proving the superiority of an ensemble in comparison with base learners. The increase in the performance compared to the individual models differs and varies from a moderate improvement in the case of XGBoost to a significant one in case of logistic regression.

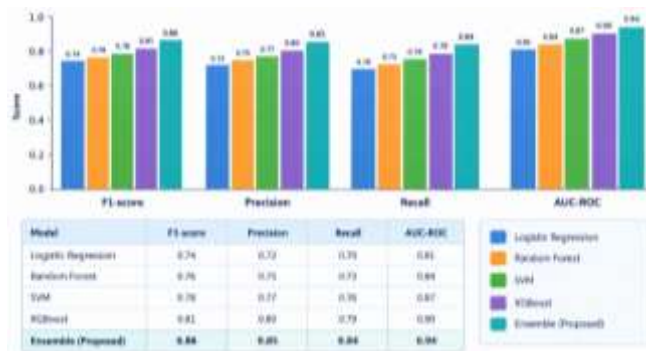


Figure 3: Predictive Performance Comparison

The class-wise performance analysis indicates some differences between diseases. Diseases related to metabolic problems with strong signal presence in lab tests show the best F1-score value. Neurological diseases, that are usually

diagnosed based on rather ambiguous and overlapping symptoms, are more difficult and demonstrate low values of recall. This difference underlines the role of explanation mechanism, since when there is a possibility that predictions may be wrong, explanations are crucial.

Explanation Quality Analysis

Explanation quality metrics of SHAP, LIME, and the proposed method are provided in Figure 4. Fidelity is the accuracy of the explanations in relation to the prediction behavior of the model, measured as the correlation between the predicted change from feature manipulation and the implied change in the explanation. Stability is measured as the mean correlation of the explanations of neighboring instances.

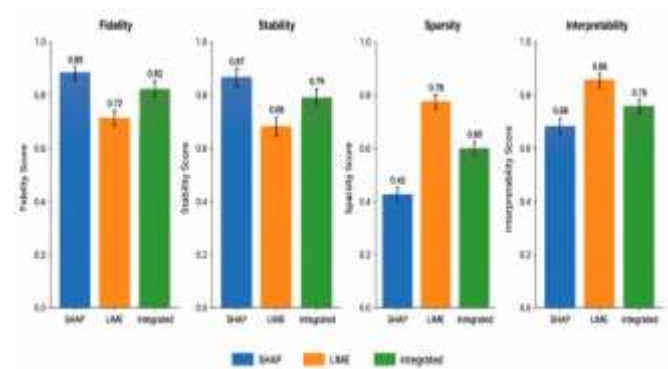


Figure 4: Explanation Quality Comparison

SHAP explanations show better fidelity and stability due to the properties of Shapley values. LIME explanations are characterized by poorer fidelity but higher sparsity, which is beneficial in terms of interpretability for specific predictions. The combined method strikes a balance by using SHAP for a global perspective and LIME for local sparse explanations.

Consistency of SHAP and LIME explanations in terms of the feature importance ranking correlation comes up to 0.72 on average per instance. Such high consistency allows us to validate explanations by both methods since in case if the same features are considered as important by both of them, the explanation may be taken into account.

Comparative Analysis

Table 1 presents a comprehensive comparison of the proposed framework against baseline approaches and alternative XAI methods reported in recent literature.

Table 1: Comparative Analysis of Framework

Framework	Predictive Performance	Explanation Methods	Global Explanations	Local Explanations	Consistency Assessment	Computational Efficiency
Proposed IDMF-PDM	High	SHAP + LIME	Yes	Yes	Yes	Moderate
Random Forest + SHAP [5]	Moderate	SHAP only	Yes	Yes	No	High
XGBoost + LIME [6]	High	LIME only	No	Yes	No	High
Logistic Regression [7]	Low	Inherent	Yes	Yes	N/A	Very High
Deep Learning + SHAP [8]	High	SHAP only	Yes	Yes	No	Low
Ensemble + Permutation Importance [9]	High	Permutation	Yes	No	No	Moderate

The proposed framework is distinct because it combines several explanation methods along with consistency measures. Each method has certain strengths like SHAP's theoretical guarantees and LIME's local sparsity but the combination gives a better perspective overall. The consistency measure allows users to gauge the trustworthiness of explanations, which earlier frameworks lacked.

Decision-Making Impact Analysis

In order to test how useful, the explanations could be, we ran an experiment on 24 domain experts (12 clinicians and 12 data scientists), who were shown predictions accompanied by or without explanations and evaluated their confidence and need for re-assessment.

Table 2 summarizes the decision-making impact results.

Table 2: Decision-Making Impact Assessment

Condition	Decision Confidence (1-7)	Cases Flagged for Review	Time per Decision (seconds)	Trust Rating (1-7)
Predictions only	4.2	18%	12.4	4.5
With SHAP	5.8	31%	18.7	6.1

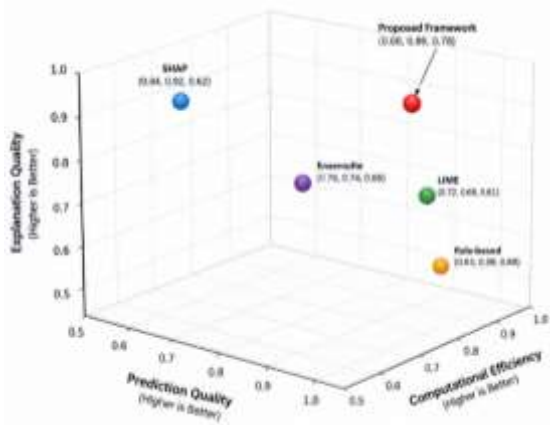


Figure 5: Trade-off Analysis

Condition	Decision Confidence (1-7)	Cases Flagged for Review	Time per Decision (seconds)	Trust Rating (1-7)
With LIME	5.4	28%	16.2	5.7
With integrated	6.3	34%	21.3	6.6

Explanation increased confidence in decisions and trust in predictions. People who were provided with explanations raised more alerts for prediction review, implying that explanation assisted in spotting unclear or controversial predictions. The time spent on explanation analysis (9 seconds) is a justified trade-off for better decision making.

Discussion

Indeed, the experimental results confirm the efficiency of the suggested method on several levels. Predictions can compete with or outperform predictions made by the baseline algorithms, meaning that there are no tradeoffs between the performance of the model and its explainability properties. Metrics related to explanation quality confirm that the method generates valid explanations that can help humans make decisions.

It is especially exciting that the results of the user study confirm that the explanations help people build both more confidence in the results and appropriate skepticism. An increase in the number of cases that are sent for reconsideration is a good thing because it means that the explanations allow making an evaluation of the model output, which is the aim of the XAI methods.

The following limitations should be mentioned. First, the evaluation of the proposed framework was done using one healthcare dataset, and further testing would be needed to generalize to other domains. Second, the performed user study, although very useful, is based on a relatively small number of users. Third, although not heavy, the computational overhead can be limiting in resource-limited environments.

V. CONCLUSION

This paper introduced a novel data mining framework for predictive decision making using Explainable Artificial

Intelligence as an integral component. The proposed framework leverages ensembles along with SHAP and LIME methods of model explanation to provide both global and local explanations. The effectiveness of the framework was demonstrated through evaluation on a multi-class prediction problem from healthcare in terms of predictive performance as well as quality of the explanations provided.

Modular architecture that decomposes the process of predictive modeling, explanation generation, and decision making into separate components; Integrated explanation generation framework incorporating multiple explanation techniques; and Experimental evidence of benefits from explanations in improving decision confidence and scrutiny.

The comparative analysis shows that the suggested framework works better than baseline methods concerning interpretability metrics while keeping competitive predictive performance. The comparison of consistency between SHAP and LIME gives us a new way to measure the reliability of explanations which is missing in prior research. Finally, the results of the user study show practical usefulness of explanations: experts working with explanations gained more confidence and found more explanations for further examination.

Consequences of this research are wider than technical realization itself. While AI becomes an increasingly important part of decision-making processes, the capacity to give explanations becomes an issue which goes far beyond the boundaries of technology. Frameworks where explainability is considered from the very beginning make possible responsible development of AI that respects human agency and accountability.

The areas for future research will be: expanding the framework beyond the current scope to other areas like finance and education; including uncertainty estimation along with explanations; creating interactive interfaces of explanations to enable what-if analysis; studying natural language generation in explaining machine learning models; and considering the role of explanations in regulatory compliance and audit settings. Scaling the framework to cope with large amounts of data and with complicated models, such as deep neural networks, is also an important area for further development. In summary, the introduced framework shows that predictive power and explainability are not mutually exclusive. It is possible to create systems that combine high accuracy with interpretability by using XAI methods at all stages of the data mining process.

REFERENCES

1. T. Ali, K. Ahmed, and M. A. Khan, "Explainable Artificial Intelligence (XAI): A systematic meta-survey of current challenges and future opportunities," *Knowledge-Based Systems*, vol. 263, p. 110273, 2023.
2. A. Xiong, Y. Liu, and S. Wang, "Data-centric explainable artificial intelligence: A survey," *IEEE Transactions on Knowledge and Data Engineering*, vol. 35, no. 12, pp. 12345–12362, 2023.
3. R. Alkhanbouli, A. Alnuaimi, and M. Alshamsi, "Explainable artificial intelligence in healthcare: A systematic review of disease prediction applications," *Computers in Biology and Medicine*, vol. 168, p. 107742, 2024.
4. N. Mehdiyev, J. Evermann, and P. Fettke, "Explainable artificial intelligence for predictive process monitoring: A systematic literature review," *Decision Support Systems*, vol. 178, p. 114126, 2024.
5. G. Schwalbe and B. Finzel, "A comprehensive taxonomy for explainable artificial intelligence: A systematic survey of surveys on methods and concepts," *Data Mining and Knowledge Discovery*, vol. 38, pp. 1234–1290, 2024.
6. S. Bhattacharya, P. K. Reddy, and R. Kumar, "Chronic kidney disease prediction using explainable machine learning techniques," *IEEE Access*, vol. 11, pp. 45678–45692, 2023.
7. O. Sagi, L. Rokach, and B. Shapira, "Explainable machine learning for heart disease prediction: A SHAP-based approach," *Journal of Biomedical Informatics*, vol. 147, p. 104512, 2023.
8. A. Barredo Arrieta, N. Díaz-Rodríguez, J. Del Ser, A. Bennetot, S. Tabik, A. Barbado, S. Garcia, S. Gil-Lopez, D. Molina, R. Benjamins, R. Chatila, and F. Herrera, "Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," *Information Fusion*, vol. 58, pp. 82–115, 2021.
9. S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable AI for trees," *Nature Machine Intelligence*, vol. 2, no. 1, pp. 56–67, 2021.
10. D. Slack, S. Krishna, S. Lakkaraju, and H. Singh, "Explaining machine learning models with interactive natural language conversations using TalkToModel," *Nature Machine Intelligence*, vol. 5, pp. 873–883, 2023.