Explainable Prediction and Diagnosis of Dementia Based on Biomedical Markers: A Case Study

Kalliopi Klelia Lykothanasi*, Anastasios Giannaros*, Dimitrios Tsoukalos*
Dimitrios Koutsomitropoulos*, Dimitrios Tsolis†
*Computer Engineering and Informatics Department, University of Patras, Patras, Greece
†Department of History and Archaeology, University of Patras, Patras, Greece
k.lykothanasi@ac.upatras.gr; giannaros@ceid.upatras.gr; dimitristsoukalos@gmail.com;
koutsomi@ceid.upatras.gr; dtsolis@upatras.gr

***Abstract*—The need for explainability of AI results and recommendations is shifting from beneficial insight to a strict requirement, especially in critical domains such as biomedicine. In this paper, we introduce an explainable artificial intelligence (XAI) framework designed to support transparent prediction and diagnosis of dementia using high-dimensional biomedical, behavioural, and demographic data. Our approach comes as a part of a holistic framework for digitally enabled healthcare that supports enhanced prevention, risk identification, and relief for dementia and frailty within the COMFORTage EU project. We present an end-to-end pipeline that incorporates data preprocessing, model training, and multi-method explainability using SHAP, DeepSHAP, LIME, Anchor, and DiCE, unified under a modular architecture. The system can accommodate different explanation tools depending on the model and task required, and produces local and global interpretations as well as counterfactual analyses. Two real-world clinical datasets on dementia biomarkers are examined, representing longitudinal studies focusing on preclinical stages of Alzheimer's disease. A qualitative evaluation with clinicians is conducted that highlights strong preferences for precise, graph-based explanations and emphasises the importance of what-if analysis for actionable clinical insight.**

***Index Terms*—Explainable Artificial Intelligence (XAI), Artificial Intelligence (AI), dementia, frailty, digital health**

I. Introduction

The recent growth of AI and Machine Learning (ML) has enabled major advancements in various application domains. Medical informatics is a field that traditionally benefits from the analytical, predictive, and decision support capabilities of current learning algorithms. For AI-aided decisions, however, to be taken into account, corresponding systems need to be able to produce concrete, rational proof or paths of thought that explain the proposed results [1].

This is especially important in critical domains such as healthcare, where traditional human expertise often places AI out-of-the-box suggestions under scepticism and scrutiny. The cost of misleading decisions and diagnoses appears to account for over 17.5% of the total healthcare expenditure [2]. As a result, ethical AI deployment is subject to even more strict monitoring [3].

Explainable AI is therefore a valuable set of methods and processes that allow understanding and trust of the results produced by machine learning algorithms. It transforms "black box" models into transparent systems, revealing how AI reaches specific decisions.

In this paper, we propose a methodology and framework to achieve and support XAI for the interpretable diagnosis of dementia. This methodology forms an integral component of a holistic approach for a digitally enabled healthcare framework for enhanced prevention, risk identification, and relief for dementia and frailty. This is the approach currently carried out within the COMFORTage EU project [4]. The project aims to develop a Virtualised AI-Based Healthcare Platform (VHP) for personalised risk factor analysis and early diagnosis, leveraging cutting-edge technologies like Artificial Intelligence, the Internet of Things (IoT) and Digital Twins. COMFORTage integrates medical, social, and technical data to enhance the quality of life and independence of the ageing population, while simultaneously driving the digital transformation and sustainability of healthcare systems across Europe [5].

In addition to the contributing methodology, we focus on a concrete case study for end-to-end dementia diagnosis and interpretation. This pipeline includes the collection of raw data corresponding to related biomarkers, a library of ML models to support prediction and diagnosis tasks, and user interfaces to explain results. Most importantly, it provides an adaptive toolkit supporting a range of XAI methods and techniques that are to be dynamically applied on a case-by-case basis, depending on the query and model used.

II. Background and Related Work

A. XAI Methods and Tools

In the COMFORTage project, AI models were trained on data comprising over 500 variables per instance. Many of these models, such as deep neural networks, are considered black boxes due to the complexity of the underlying modern ML techniques.

Explainability is therefore a prerequisite functionality of the system, and extensive research was conducted to evaluate available XAI frameworks. The main technical characteristics considered were as follows:

- Whether a tool/framework provides **local interpretation**, meaning whether it can offer explainability for the model's decision(s) regarding the data of a single patient.
- Whether a tool/framework provides **global interpretation**, meaning whether it can identify the general features that have the greatest influence across an entire dataset.

Based on the above, the following explainability methodologies and tools were examined.

SHAP leverages game theory to assign importance values to features by considering their contribution to a model's prediction [6]. This ensures a consistent and comprehensive measure of feature importance. In addition to the base, exact method (`ExactExplainer`), the SHAP package bundles approximate methods both model-agnostic (`PermutationExplainer`) and tailored to specific model families (e.g. `DeepExplainer`, `TreeExplainer`, `LinearExplainer`). `DeepExplainer`, also referred to as **DeepSHAP**, is a combination of the SHAP and DeepLIFT (see below) methods for deep neural network architectures.

LIME builds local surrogate models around individual predictions, offering an interpretable approximation of any classifier's behaviour within a specific neighbourhood of the data [7].

DeepLIFT (Deep Learning Important FeaTures) compares neuron activations to reference activations, attributing contributions to each feature based on the observed differences [8].

ELI5 is a Python package that helps to debug machine learning classifiers and explain their predictions. It provides support for specific machine learning frameworks and packages [9].

InterpretML exposes two types of interpretability: *glass-box*, covering interpretable models designed to be trained from scratch (e.g., linear models, rule lists, generalised additive models), and *black-box*, covering post-hoc explainability techniques for analysing already-trained opaque models [10].

Anchor generates human-readable 'if-then' rules to explain model predictions by identifying conditions that significantly alter predictions [11].

DiCE (Diverse Counterfactual Explanations for ML) generates a set of counterfactual examples that show how a model's prediction would change under minimal perturbations to the input features, providing insights into what would need to be different for an alternative outcome [12].

Dalex [13] is a Python package that implements a model-agnostic interface for interactive explainability and fairness.

Grad-CAM (Gradient-weighted Class Activation Mapping) [14] highlights the image regions that most influence class predictions in convolutional neural networks (CNNs).

While sufficient explainability is one of the key requirements for every ML/AI predictive method used in this project, the "explainability-performance trade-off" must be considered. It is widely accepted that model complexity and predictive accuracy tend to increase in tandem, at the cost of interpretability [15], [16].

B. XAI in Healthcare and Medical Informatics

A substantial amount of research has been conducted in the field of Explainable Artificial Intelligence in healthcare and beyond.

Wani et al. conducted an analysis on research publications dated between 2004 and 2024, focusing on Explainable AI models in healthcare [17]. The research concludes that the combination of XAI and the Internet of Medical Things (IoMT) adds transparency and trust, leading to faster sickness detection and lower medical expenses for both the patient and the healthcare system.

Eke et al. conducted a similar literature review focusing on 69 publications dated between 2015 and 2023 [18]. Throughout their research, they noticed that various XAI methods yield different results when it comes to helping users interpret AI outputs. Another interesting conclusion they reached was that there is no evaluation metric to measure the contribution of XAI. Therefore, assessing the effectiveness of various methodologies in explaining observations relies on subjective interpretability.

Sadeghi et al. [19] conducted a systematic review of the challenges of XAI in healthcare, focusing on the significance of XAI and the challenges that accompany the technology in the sector. Their research points out that transparency often comes at the expense of model efficiency. To successfully embed XAI methodologies in the medical field, self-explanatory techniques need to be employed in an effort to evade the need for post-hoc methodologies. In addition to Sadeghi et al., both Zhang et al. and Prentzas et al. conclude that there are some works of XAI being applied in the medical field, but very few of them include the participation of medical experts [19]–[21].

Wani et al. proposed an XAI-driven method for lung cancer detection. Following an evaluation, their model reached a high score of accuracy, sensitivity, and F1-score (97.43%, 98.71%, and 98.08%, respectively) [22]. Their approach is based on post-hoc explainability based on SHAP.

Singh et al. proposed an XAI-driven approach for Type-2 diabetes diagnosis [23]. Their approach is a hybrid model that combines XGBoost for classification and CNNs for deep feature extraction. The XAI feature is provided by SHAP. Their results suggest that explainability does not have to negate the high performance of the model and point out the need for bias monitoring by removing uncorrelated data.

III. Design and Architecture

A. The COMFORTage Framework and Workflow

The use of artificial intelligence and its explainability is a key pillar of the COMFORTage project. The aim is to develop a platform where doctors and other medical professionals can enter longitudinal patient data, including medical tests, behavioural factors, patient-reported outcomes, medical and family history, environmental factors, and demographics. With the help of AI, physicians can then obtain estimates of the probability of each patient developing dementia or frailty, as well as an indication of which factors influenced the prediction.

Based on these results, the clinician is supported in understanding and evaluating the output of the AI model. The system can immediately display the variables with the greatest influence on a given prediction, indicate whether each influence is positive or negative, and provide the patient with more targeted guidance. For example, if a patient has a 25% chance of developing dementia, the doctor can inspect which factors

led the model to that estimate. If a particular modifiable factor increases the probability by 8%, the doctor can advise the patient to address it and re-run the model to assess the impact.

The development of an XAI component is therefore of critical importance to the project, as each examination may involve hundreds of variables and AI results are often difficult for clinicians to interpret without guidance. Thus, with the use of the XAI component, the probabilistic assessment of the model about a patient developing dementia or frailty can be explained, strengthening clinician confidence and positioning AI as an important tool in patient care.

The data flow process starts and ends with the doctor. Patient data entered by the clinician is stored in the project's Data Lake, which serves as the primary repository for both the Integrated Care Model Library (ICML) and the XAI Framework. The ICML hosts a model trained on the available patient data. When the doctor requests an assessment from a patient, the communication is forwarded via the Data Lake to the ICML, which receives the data and returns the results to the doctor via the Data Lake. As discussed in subsequent sections, these results comprise the serialised model assessment and the influence scores of the top-10 most influential variables.

Results are visualised in the Patient Digital Twin (PDT), the project's dedicated frontend. This data flow, as far as the explainability component is concerned, is depicted in Fig. 1.

B. XAI Component Design

At the conceptual level, the XAI framework serves as an abstraction layer that encapsulates both third-party XAI tools/algorithms and custom, integrated implementations, providing a unified interface for explanation generation. The framework automates the selection, instantiation and parametrisation of applicable explanation tools for every given combination of ML/AI model and dataset. This design treats the model as a black-box prediction generator and requires minimal metadata (such as the model backend/library) to orchestrate the explanation generation process.

While Fig. 1 presents a component diagram of the XAI Framework in relation to the COMFORTage VHP, Fig. 2 details the internals of the component itself. The Explanation Orchestrator is the entry point to the system and controls the explanation generation pipeline that the framework provides. It is responsible for decoupling the XAI Framework API from the specific requirements of any method the framework incorporates. The Explainer Creator is a factory method for all supported explanation generators ("model explainer") instances. The Adapter layer consists of thin wrappers corresponding to each of these model explainers, internal or external to the framework. This enables the higher logic layers to treat explanation tools/algorithms as interchangeable modules that expose a standardised interface for instantiation and explanation generation. Under the adapters, the Core Logic layer houses any internally-defined XAI method. The current version of the framework contains optimised versions of several XAI tools, further customised for maximum compatibility with the AI/ML model types supported by the XAI framework. All model explainers by default define their own "explanation" class. The `MetaExplanation` class is a lightweight container that serves as an explanation "summary", a unified collection of these disparate explanation objects corresponding to a single explanation generation request. Finally, the Data Transformers layer contains (JSON) serialisation logic for all produced explanation types.

The flow starts with the parent module creating an `ExplanationOrchestrator` instance for a combination of an ML/AI model and a training (or "background") dataset. Subsequently, the parent component initiates an explanation generation request for one or multiple inference-time samples. The orchestrator utilises a selection mechanism based on a compatibility matrix structure to determine which XAI tools are compatible with the given setup. This information is forwarded to the `ExplainerCreator`, which instantiates the appropriate adapter for each XAI method specified. The orchestrator queries each registered adapter sequentially. When all XAI methods have produced their corresponding explanations, the orchestrator bundles them into a new `MetaExplanation` object that is returned to the invoking module. Optionally, the parent component can request a JSON-serialised version of the Meta Explanation. In this case, the `MetaExplanation` instance uses the corresponding serializer function for each explanation type to generate individual JSON representations and finally combines them into a single JSON object. The parent can request that the serialised explanation object be either returned as an in-memory string variable or saved to a local file, the name of which is returned to the calling component.

IV. Datasets

The datasets we processed were ALBION [24] and HELIAD [25], which were provided by Aiginiteio University Hospital. Both datasets were anonymised, using a unique identifier for each patient. Each file is accompanied by a dictionary that explains the codes of each variable.

More specifically, the *Albion Dataset* [24] had 545 variables, which were divided into 20 categories such as Demographics, Medical history, etc. In detail, Albion contains a total of 470 examination records from 179 patients. The Albion Dataset continues to be enriched with new patient records.

The *Heliad Dataset* [25] consisted of two phases, phase A and phase B. In total, Heliad had 3227 records, with 2001 unique records.

A. Data Structure and Preprocessing Methodology of the Albion Dataset

The dataset utilised in this study (ALBION dataset) underwent a rigorous structural organisation and preprocessing phase to ensure compatibility with explainable machine learning frameworks. The features were categorised and encoded according to a standardised hierarchical protocol:

Feature Categorisation: Data variables were organised into logical thematic modules (e.g., Demographics, Medical History, Biomarkers). Each module is identified by a unique

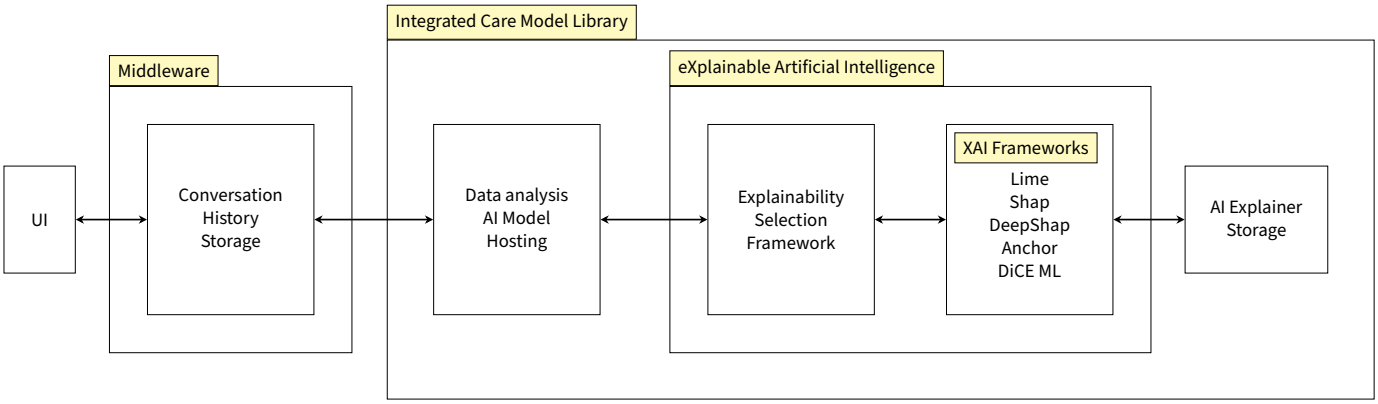

Fig. 1. System components overview.

group code (e.g., G01, G02, ... G20), facilitating a structured approach to multi-modal data analysis.

Feature Encoding and Naming Convention. To maintain data integrity and automated type recognition, a specialised naming convention was implemented for all variables using the format `_X_Y_Z`:

1) **Type Identifier (X)**: Defines the mathematical nature of the variable:
   - **N (Nominal)**: Categorical data without intrinsic ordering (e.g., Gender).
   - **O (Ordinal)**: Categorical data with a defined rank or scale (e.g., Education levels).
   - **S (Scale/Numerical)**: Continuous or discrete quantitative measurements.
2) **Group/Role Identifier (Y)**: Indicates the thematic group code or designates the variable as a Target (`_TRG`) for the prediction models.
3) **Auxiliary Suffix (Z)**: Features identified as non-contributory to the predictive output (e.g., Patient IDs) were flagged with an `_AUX` tag to be excluded from the model training phase while remaining available for metadata reference.

B. Data Cleaning and Quality Control

A systematic data cleaning process was applied to handle noise and missing information:

Missing Value Handling: Specific numeric placeholders for null or missing entries (e.g., 99 or 999) were identified and removed.

Dictionary Validation: Feature descriptions were cross-referenced and validated against the original study's data dictionary to ensure semantic accuracy in the final dataset used for the explainability framework.

V. Implementation

A. XAI Tool Selection

The primary objectives behind the selection of XAI techniques were to maximise coverage across diverse model architectures and task types, and to provide a broader variety of explanation formats than traditional quantitative feature attributions. The framework currently integrates SHAP, DeepSHAP, LIME, Anchor, and DiCE. SHAP provides both local and global feature attributions, facilitating the explanation of predictions ranging from individual samples to entire cohorts.

DeepSHAP was specifically selected to target neural network architectures. It offers a high-fidelity alternative to its parent techniques, SHAP and DeepLIFT, by leveraging deep learning-specific optimisations to provide superior explanation quality with lower computational overhead [6].

LIME was incorporated for its lightweight, model-agnostic nature, allowing for the rapid computation of local feature attributions. Anchor provides a complementary qualitative approach by identifying high-precision rules.

Although limited to classification tasks, Anchor identifies feature values that constitute "sufficient conditions" for a prediction, offering valuable context beyond the additive contributions of SHAP and LIME.

Finally, DiCE provides a counterfactual perspective, explaining the necessary feature perturbations required to definitively steer a model output towards an alternative class or value.

B. XAI Framework Component Implementation

The implementation strategy for the framework was primarily driven by the goal of integrating disparate, research-oriented XAI libraries into a cohesive, production-ready system without the prohibitive overhead of rewriting every underlying algorithm from scratch. Our approach was structured across three axes: library and resource management optimisation, XAI tool customisations for wider compatibility with AI/ML models and output data serialisation and bundling for enhanced interoperability with the PDT and ICML components.

Library Optimisation and Resource Management: Package initialisation can introduce significant overhead in interpreted environments, especially in the case of AI and XAI workflows, which can unnecessarily trigger the loading of heavy libraries such as PyTorch and SciPy. We ad-

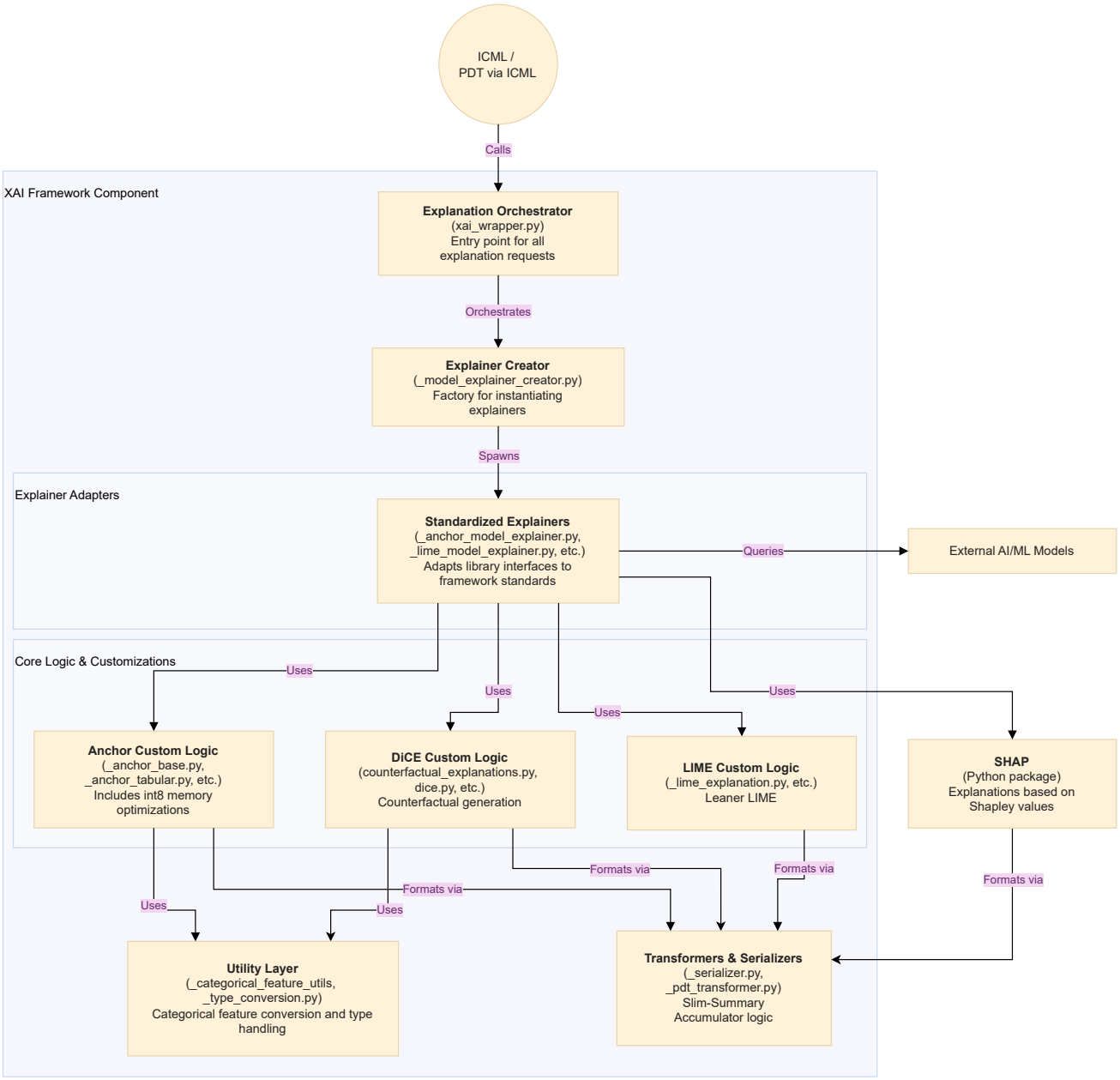

Fig. 2. C4 component diagram of the XAI Framework component.

dressed this by refactoring the import logic of the internal LIME, Anchor and DiCE libraries to enforce lazy package loading and granular rather than broad imports (e.g., `from scipy.sparse import csr_matrix` instead of `import scipy`). Postponing heavy backend loading until such a module is needed by an instantiated adapter or by an XAI tool, we reduced the import and initialisation time of `ExplanationOrchestrator` from approximately 16 seconds to under 5 seconds. Additionally, we streamlined the internal logic of the libraries mentioned above by pruning sub-modules not required by, or incompatible with, our framework, such as `SP-LIME`, `AnchorImage` and `FeasibleModelApprox` (DiCE). This pruning facilitated dependency reduction and, in some instances, decreased the overall memory footprint of the framework.

XAI Tool Customisations and Compatibility: The project requirement to support high-dimensional data and diverse model backbones prompted several targeted XAI tool customisations in the Adapter and Core Logic layers. As LIME natively operates on NumPy array objects only, we added support for PyTorch model explanation by implementing bidirectional conversion logic between `torch.Tensor` and

`numpy.array` objects within the LIME adapter. Furthermore, we modified the execution logic to produce explanations for the model's actual predicted class rather than the library default of the class with label '1', ensuring relevance in both binary and multi-class settings.

Regarding DiCE, we extended the base implementation with native support for XGBoost and CatBoost models. A further salient enhancement was the addition of support for non-neural-network classifiers that return one-hot encoded predictions. Originally, DiCE required wrapping the entire data preparation and model prediction pipeline inside a transformer function to strictly enforce label-encoded predictions. Since the XAI framework receives pre-processed data, we isolated the output encoding conversion to the automated application of an `argmax` function wrapper, integrated into the relevant model interface classes.

Finally, we addressed an original implementation bottleneck in the Anchor core logic that caused severe memory bloat when working with high-dimensional datasets. Specifically, the internal state-tracking arrays used the standard Python 64-bit integer data type (`int`) to represent binary variables. Substituting them with 8-bit unsigned integers (`numpy.uint8`) achieved an 8-fold reduction in state array memory footprint without requiring significant refactoring of the original logic.

Data Transformation and Serialisation: Facilitating information exchange with the internal ICML and database components, as well as the user-facing PDT necessitated the adoption of a standardised and portable data format. Custom classes implemented for LIME, Anchor and DiCE explanations extract only the essential, user-presentable attributes from the native explanation objects, discarding cumbersome metadata such as internal state representations. These slimmed-down explanations are explicitly cast to native Python objects (e.g., `int` instead of `np.int64`) before serialisation, ensuring compliance with the JSON data exchange standard. Fig. 3 illustrates a snippet of a JSON-serialised prediction explanation generated with data from the HELIAD dataset.


```
{
  "shap_explanation": {
    "base_values": [[0.9754, 0.02464]],
    "shap_values": [[
         [7.8678e-08, -8.0331e-08],
         "..."
    ]],
    "data": [[0.0, -0.0070, "..."]],
    "feature_names": ["A2", "A16", "..."]
  },
  "lime_explanation": {
    "explanations_list": [
      {
        "No Dementia Diagnosis": {
          "local_feature_names": [
            "mmper23=0",
            "L42c1 <= 0.00",
            "..."
          ],
          "local_feature_importances": [
            -0.0204, 0.0184, "..."
          ]
        }
      }
    ]
  },
  "anchor_explanation": {
    "anchor_rules": ["A2 = 0.00"],
    "precisions": [1.0],
    "coverages": [0.2057]
  },
  "top_k_shap_features": {
    "k": 10,
    "top_shap_feature_names": [
      "K31b",
      "MET_min_work_wk",
      "K32a",
      "..."
    ],
    "top_feature_contributions": [
      0.0084, 0.0074, 0.0070, "..."
    ]
  }
}
```


Fig. 3. Snippet of a JSON-serialised prediction explanation (data source: HELIAD).

C. Evaluation

It is important to have a measure of the perceived usefulness and expectations users may have from any XAI system. To this end, we conducted a qualitative survey among 32 clinical partners about XAI requirements and preferred explanation modalities. The results are summarised in Table I.

The vast majority of practitioners opted for visual representations and graphs as a means to convey explainability results (70% of respondents). At the same time, precision is critical to respondents (63%), even at the expense of simplicity. Finally, what-if sensitivity analysis of biomarkers and local interpretations on a per-case basis were identified as the most useful functionalities for clinical practice, cited by over half of respondents (56%).

Finally, to the open question "What else would you want or expect from XAI after the recent implementation?", the participants answered that they would expect the XAI implementation to provide clearer, more accessible explanations of how the model calculates risk, using simple language that both healthcare professionals and participants can understand. They emphasised the need for human-language interpretations of graphs and statistics, multiple graph formats, and optional text descriptions to support better understanding of results. They also highlighted the importance of transparency around data sources, the amount of data used, model accuracy, validation methods, and the limitations of the explanations.

VI. Conclusions and Future Work

In this paper, we have proposed and demonstrated an approach and framework for XAI to support early prediction and diagnosis of dementia, based on biomedical markers, behavioural and demographic data. We have designed an end-

TABLE I
QUALITATIVE SURVEY RESULTS

Evaluation Question	**Option 1**	**Option 2**	**Option 3**
What form of explanation would be the most convenient, understandable and effective for the result of a model?	Data table: 3	**Graph: 21**	Description text: 6
What do you consider to be the most important for the XAI?	Simplicity: 9	**Precision: 17**	Speed: 1
Which function of XAI do you consider most useful for clinical practice?	Global interpretation: 3	Local interpretation: 8	**What-if analysis: 14**

to-end XAI pipeline within an integrated healthcare platform, which provides an adaptive toolkit supporting a range of XAI methods and techniques.

We have shown specific optimisations and refactoring to reduce system load and support timely decision making. We have tackled the dimensionality and diversity of biomedical data with certain customisations in the implemented XAI libraries and the execution logic. Finally, we performed a qualitative survey that validates the requirement for user-friendly, timely and accurate decision support that is tailored to specific tasks with varying levels of granularity.

Future research concerning the XAI framework will prioritise the enhancement of its architectural robustness and its empirical validation within operational clinical settings. The inclusion of counterfactual explanations in mixed-encoding scenarios, the extension toward multi-class classification tasks and further memory optimisations need to be considered. Ensuring that explanation presentations in both textual and visual forms are human-readable and user-friendly will foster clinician trust and facilitate the adoption of AI as an assistive technology in the clinical process.

ACKNOWLEDGMENTS

The research leading to the results presented in this paper has received funding from the European Union's Horizon Europe research and innovation program under grant agreement no. 101137301 and is supported by Innovate UK under grant agreement no. 10103541. Views and opinions expressed are however those of the authors only and do not necessarily reflect those of the European Union or European Health and Digital Executive Agency (HaDEA). Neither the European Union nor the granting authority can be held responsible for them.

REFERENCES

[1] 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, Jun. 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253519308103

[2] L. Slawomirski, D. Kelly, K. de Bienassis, K.-A. Kallas, and N. Klazinga, "The economics of diagnostic safety," OECD, Tech. Rep., Mar. 2025. [Online]. Available: https://doi.org/10.1787/fc61057a-en

[3] "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) (Text with EEA relevance)," 2024. [Online]. Available: http://data.europa.eu/eli/reg/2024/1689/oj

[4] "Home – COMFORTage." [Online]. Available: https://comfortage.eu/

[5] G. Manias, S. Likothanassis, E. Alexakis, A. Antoniades, C. Marra, G. M. Giuffrè, E. Charalambous, D. Tsolis, G. Tsirogiannis, D. Koutsomitropoulos, A. Giannaros, D. Tsoukalos, K. K. Lykothanasi, P. Vogazianos, S. Kleftakis, D. Vrachnos, K. Charilaou, J. Lenkowicz, N. Martellacci, A. M. Tudor, N. Borovits, M. Sangiovanni, W.-J. v. d. Heuvel, on behalf of the COMFORTage Consortium, and D. Kyriazis, "Establishing a digitally enabled healthcare framework for enhanced prevention, risk identification, and relief for dementia and frailty," *Journal of Dementia and Alzheimer's Disease*, vol. 2, no. 3, p. 30, Sep. 2025. [Online]. Available: https://www.mdpi.com/3042-4518/2/3/30

[6] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds., vol. 30. Curran Associates, Inc., 2017. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2017/file/8a20a8621978632d76c43dfd28b67767-Paper.pdf

[7] M. T. Ribeiro, S. Singh, and C. Guestrin, ""why should i trust you?": Explaining the predictions of any classifier," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, p. 1135–1144. [Online]. Available: https://doi.org/10.1145/2939672.2939778

[8] A. Shrikumar, P. Greenside, and A. Kundaje, "Learning important features through propagating activation differences," 2017. [Online]. Available: http://arxiv.org/abs/1704.02685

[9] "Overview – ELI5 0.16.0 documentation." [Online]. Available: https://eli5.readthedocs.io/en/stable/overview.html

[10] H. Nori, S. Jenkins, P. Koch, and R. Caruana, "InterpretML: A unified framework for machine learning interpretability," ArXiv, September 2019. [Online]. Available: https://www.microsoft.com/en-us/research/publication/interpretml-a-unified-framework-for-machine-learning-interpretability/

[11] M. T. Ribeiro, S. Singh, and C. Guestrin, "Anchors: high-precision model-agnostic explanations," in *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence*, ser. AAAI'18/IAAI'18/EAAI'18. AAAI Press, 2018. [Online]. Available: https://doi.org/10.1609/aaai.v32i1.11491

[12] R. K. Mothilal, A. Sharma, and C. Tan, "Explaining machine learning classifiers through diverse counterfactual explanations," in *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, ser. FAT* '20. New York, NY, USA: Association for Computing Machinery, 2020, p. 607–617. [Online]. Available: https://doi.org/10.1145/3351095.3372850

[13] H. Baniecki, W. Kretowicz, P. Piątyszek, J. Wiśniewski, and P. Biecek, "dalex: Responsible machine learning with interactive explainability and fairness in Python," *Journal of Machine Learning Research*, vol. 22, no. 214, pp. 1–7, 2021.

[14] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. [Online]. Available: https://doi.org/10.1109/ICCV.2017.74

[15] B. Crook, M. Schlüter, and T. Speith, "Revisiting the performance-explainability trade-off in explainable artificial intelligence (XAI)," in *2023 IEEE 31st International Requirements Engineering Conference Workshops (REW)*, 2023, pp. 316–324. [Online]. Available: https://doi.org/10.1109/REW57809.2023.00060

[16] E. Tjoa and C. Guan, "A survey on Explainable Artificial Intelligence (XAI): Toward medical XAI," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793–4813, 2021. [Online]. Available: https://doi.org/10.1109/TNNLS.2020.3027314

[17] N. A. Wani, R. Kumar, Mamta, J. Bedi, and I. Rida, "Explainable AI-driven IoMT fusion: Unravelling techniques, opportunities, and challenges with explainable AI in healthcare," *Information Fusion*, vol. 110, p. 102472, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253524002501

[18] C. I. Eke and L. Shuib, "The role of explainability and transparency in fostering trust in AI healthcare systems: a systematic literature review, open issues and potential solutions," *Neural Computing and Applications*, vol. 37, no. 4, pp. 1999–2034, Feb. 2025. [Online]. Available: https://doi.org/10.1007/s00521-024-10868-x

[19] Z. Sadeghi, R. Alizadehsani, M. A. CIFCI, S. Kausar, R. Rehman, P. Mahanta, P. K. Bora, A. Almasri, R. S. Alkhawaldeh, S. Hussain, B. Alatas, A. Shoeibi, H. Moosaei, M. Hladík, S. Nahavandi, and P. M. Pardalos, "A review of Explainable Artificial Intelligence in healthcare," *Computers and Electrical Engineering*, vol. 118, p. 109370, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790624002982

[20] Y. Zhang, Y. Weng, and J. Lund, "Applications of Explainable Artificial Intelligence in diagnosis and surgery," *Diagnostics*, vol. 12, no. 2, p. 237, Feb. 2022. [Online]. Available: https://doi.org/10.3390/diagnostics12020237

[21] N. Prentzas, A. Kakas, and C. S. Pattichis, "Explainable AI applications in the medical domain: a systematic review," 2023. [Online]. Available: https://arxiv.org/abs/2308.05411

[22] N. A. Wani, R. Kumar, and J. Bedi, "DeepXplainer: An interpretable deep learning based approach for lung cancer detection using explainable artificial intelligence," *Computer Methods and Programs in Biomedicine*, vol. 243, p. 107879, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S016926072300545X

[23] S. Singh, N. A. Wani, R. Kumar, and J. Bedi, "DiaXplain: A transparent and interpretable artificial intelligence approach for type-2 diabetes diagnosis through deep learning," *Computers and Electrical Engineering*, vol. 126, p. 110470, 2025. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0045790625004136

[24] F. Kalligerou, E. Ntanasi, P. Voskou, G. Velonakis, E. Karavasilis, E. Mamalaki, A. Kyrozis, E. Sigala, N. Economou, K. Patas, M. Yannakoulia, and N. Scarmeas, "Aiginition longitudinal biomarker investigation of neurodegeneration (albion): study design, cohort description, and preliminary data," *Postgraduate Medicine*, vol. 131, no. 7, pp. 501–508, 2019, pMID: 31483196. [Online]. Available: https://doi.org/10.1080/00325481.2019.1663708

[25] E. Dardiotis, M. H. Kosmidis, M. Yannakoulia, G. M. Hadjigeorgiou, and N. Scarmeas, "The hellenic longitudinal investigation of aging and diet (heliad): Rationale, study design, and cohort description," *Neuroepidemiology*, vol. 43, no. 1, pp. 9–14, 07 2014. [Online]. Available: https://doi.org/10.1159/000362723