

Powering responsible artificial intelligence with high-quality real-world data: the S-RACE platform for scalable, multi-specialty clinical research

Received: 1 August 2025

Accepted: 30 October 2025

Cite this article as: Traverso, A., Tiano, D., Corvaglia, A. *et al.* Powering responsible artificial intelligence with high-quality real-world data: the S-RACE platform for scalable, multi-specialty clinical research. *npj Digit. Med.* (2025). <https://doi.org/10.1038/s41746-025-02132-w>

Alberto Traverso, Donato Tiano, Andrea Corvaglia, Alessio Dimonte, Edoardo Luigi Draetta, Bruno Fabiani, Patrick Scuri, Simone Barbieri, Marcello Agazzi, Muhammad Arslan, Daniele Celada, Filippo Chiabrando, Lorenzo Cibrario, Giulio Cielo, Alberto Colombo, Stefano Contini, Marta Liberotti, Marco Montagna, Francesca Rita Ogliari, Anna Palmisano, Francesco Pisu, Davide Serra, Diego Varani, Davide Vignale, Andrea Luigi Vitali, Alan Zambello, Chiara Chiapponi, Marco Denti, Antonio Esposito & Carlo Tacchetti

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

TITLE PAGE

Powering Responsible Artificial Intelligence with High-Quality Real-World Data: The S-RACE platform for Scalable, Multi-Specialty Clinical Research

Alberto Traverso¹, Donato Tiano¹, Andrea Corvaglia¹, Alessio Dimonte¹, Edoardo Luigi Draetta¹, Bruno Fabiani¹, Patrick Scuri¹, Simone Barbieri¹, Marcello Agazzi¹, Muhammad Arslan¹, Daniele Celada¹, Filippo Chiabrando², Lorenzo Cibrario¹, Giulio Cielo¹, Alberto Colombo¹, Stefano Contini¹, Marta Liberotti¹, Marco Montagna², Francesca Ogliari², Anna Palmisano^{2,3}, Francesco Pisu¹, Davide Serra¹, Diego Varani¹, Davide Vignale^{2,3}, Andrea Luigi Vitali¹, Alan Zambello¹, Chiara Chiapponi⁵, Marco Denti¹, Antonio Esposito^{1,2,3,+}, Carlo Tacchetti^{1,2,3,+}.

¹ S-RACE (San-Raffaele Ai Center), University Vita-Salute San Raffaele, Milan, Italy

² School of Medicine, University Vita-Salute San Raffaele, Milan, Italy

³ Experimental Imaging Center, IRCCS San Raffaele Hospital, Milan, Italy

⁴ Advanced Imaging for Personalised Medicine Unit, IRCCS San Raffaele Hospital, Milan, Italy

⁵ Microsoft Italia, Milan, Italy

+ Equally contributing senior authors

Corresponding authors: Dr. Alberto Traverso, S-RACE (San-Raffaele Ai Center), University Vita-Salute San Raffaele, Via Olgettina 60, 20132, Milan, Italy, traverso.alberto@hsr.it, Prof. Antonio Esposito, S-RACE (San-Raffaele Ai Center), University Vita-Salute San Raffaele, Via Olgettina 60, 20132, Milan, Italy, esposito.antonio@hsr.it

ABSTRACT

The translation of Artificial Intelligence (AI) into clinical practice demands high-quality Real-World Data (RWD), yet unstructured healthcare information poses a significant barrier. To address this, we developed S-RACE, a secure, cloud-based platform designed to systematically transform raw hospital data into high-quality, research-grade evidence. S-RACE features an end-to-end pipeline, starting with on-premises anonymisation for data privacy, followed by Natural Language Processing (NLP) to extract and standardise clinical information into the FHIR format. This curated data foundation is essential for building robust AI models. The platform's integrated "Data Science Lab" supports responsible AI development, incorporating explainability techniques and adhering to governance standards like ISO 42001 and the EU AI Act. Currently, S-RACE is populated with data from 31,276 patients, powering 19 research projects across fields including oncology, cardiology, and diabetes. We demonstrate its utility in kidney cancer and aortic stenosis, where models trained on S-RACE's automatically processed RWE showed performance comparable to those trained on manually curated data. S-RACE provides a scalable, governed environment for RWD curation, offering a trustworthy foundation to accelerate the clinical adoption of responsible AI.

INTRODUCTION

The convergence of Real-World Evidence (RWE) and Artificial Intelligence (AI) is reshaping modern medicine, offering the potential to develop advanced Clinical Decision Support Systems (CDSS) that are essential for personalised patient care. AI technologies are uniquely capable of analysing vast and complex datasets to uncover clinical insights that would otherwise remain hidden, promising to improve patient outcomes by optimising treatments and accelerating drug discovery¹.

However, the transformative potential of AI in healthcare is fundamentally constrained by the quality of the Real-World Data (RWD) it relies on. RWD is typically collected from disparate hospital IT systems and is often unstructured, sparse, and lacking in standardisation. This inherent variability creates a significant barrier to transforming raw data into the 'regulatory grade' evidence needed for robust clinical applications². Here, 'regulatory grade' refers to Real-World Data with sufficiently high levels of accuracy, completeness, and traceability to be considered reliable for supporting regulatory submissions regarding a medical product's effectiveness or safety.

Without first addressing these foundational data quality issues, AI models are prone to generating unreliable or erroneous outcomes, rendering them unusable in clinical settings and hindering their adoption. In the absence of foundational quality of RWD for AI applications, AI models can produce 'hallucinations' or outcomes that do not reflect actual data, making these applications unusable³.

While the traditional approach of tedious, manual data curation can produce high-quality datasets, it is not a scalable solution for processing the vast amount of data required for modern AI development. However, such manually curated datasets remain an invaluable 'gold standard' for benchmarking the performance and accuracy of automated data processing algorithms.

To unlock the full potential of RWD, the healthcare sector requires sophisticated, end-to-end clinical data science platforms. These platforms are specifically designed to systematically ingest, process, and harmonise complex RWD, streamlining the creation of high-quality, analysis-ready datasets that can support the development of reliable AI-driven systems.

Furthermore, the process of ingesting and managing RWD must be aligned with a complex and evolving regulatory landscape, as highlighted in our previous publications^{4,5}. The responsible integration of AI into healthcare demands robust governance frameworks to ensure safety, ethics, and trustworthiness. Key international standards and regulations, such as ISO/IEC 42001:2023 and the EU AI Act, mandate stringent requirements for data quality, risk management, and human oversight for high-risk AI systems⁶⁻⁸. These frameworks underscore the global commitment to ensure the ethical and safe development and deployment of AI in healthcare. Therefore, any platform aiming to generate research-grade RWD must be built upon a foundation of strong governance and regulatory compliance from the outset.

In response to this critical need, we developed together with Microsoft (Redmond, Washington, USA) and Porini (Milan, Italy) the S-RACE (San Raffaele Ai Center) platform. S-RACE platform is a novel, cloud-based solution engineered to directly address the challenges of data quality and governance in healthcare AI. Its core function is a comprehensive data science pipeline that begins with secure, on-premises data anonymisation, followed by NLP-driven data extraction and standardisation into the FHIR format, creating a structured and high-quality data foundation for research.

This paper presents a comprehensive overview of the S-RACE platform, detailing its architecture, functionalities, and its systematic approach to transforming raw clinical data into research-grade RWD. We demonstrate how S-RACE serves as a collaborative environment where clinicians and data scientists can jointly develop and validate responsible AI-driven decision support systems, built upon the high-quality data foundation the platform provides. Through practical examples from our ongoing clinical research, we will illustrate the platform's capabilities and show how a dedicated focus on data quality and responsible AI governance accelerates the clinical translation of AI, ultimately enhancing patient care. Finally, we will report details on the type of RWD data in the platform available for researchers.

RESULTS

The S-RACE platform is underpinned by a robust governance model designed to ensure the generation of high-quality, research-grade data, in full alignment with Responsible AI principles and key regulatory frameworks like the EU AI Act and ISO 42001:2023. Central to this model is a meticulous data quality assessment process. The platform employs a hybrid data quality model by combining an expert driven evaluation with an automated pre-processing workflow.

Before model development begins, project proposals presented by the clinical PIs are rigorously assessed using a Data Quality Checklist, which contains 39 questions across five categories: Summary, Collection, Pre-Processing, Metadata, and Data. The PI of the study and an S-RACE contact person jointly complete the questionnaire. For each question, they provide a textual response and a score from 0 (worst) to 3 (best). This four-level Likert scale ('Useless' to 'Valuable') evaluates five quality dimensions: Accessibility, Accuracy, Completeness, Consistency, and Relevancy. The team compiles the weighted and normalised evaluations into a Summary Report, which then undergoes a peer review by the S-RACE project management and IT team to catch

omissions or inconsistencies. If needed, a follow-up meeting is held with the clinical team to finalise the questionnaire before the final report is generated.

After the project is approved by the steering committee and the data has been transferred to the platform, the manual review is complemented by the Preliminary Exploratory Data Analysis (PExDA) framework—an automated pipeline. PExDA performs baseline quality checks, such as identifying and flagging patients with missing outcome data.

This entire process ensures data quality and consistency, focusing on a key principle: assessing whether the data is "fit for purpose" for a specific research question. This approach is intentionally dynamic: a dataset unsuitable for one basic model may be perfectly acceptable for a more advanced algorithm capable of handling its specific limitations, allowing the platform to continuously adapt to evolving AI techniques.

The S-RACE platform is built on three architectural pillars that create an end-to-end pipeline for transforming raw clinical data into a foundation for trustworthy AI. More details on the building blocks highlighted below are provided in Supplementary Figure 1. The platform's architecture is fundamentally shaped by a 'privacy by design' approach, in full compliance with the EU's General Data Protection Regulation (GDPR). The data pipeline begins with an on-premises engine that performs pseudonymisation before any data is transferred to the cloud. This process targets direct identifiers (e.g., name, medical record number, social security number) and replaces them with a unique, irreversible cryptographic hash. The mapping key linking the pseudonym to the original identifier is stored exclusively within the hospital's secure on-premises infrastructure and is never exposed to the cloud environment. This segregation is a critical security control that minimises the risk of re-identification. All platform activities are governed by principles of data minimisation and purpose limitation, ensuring researchers can only access the specific data necessary for their approved study protocols. This de-identification process is tailored to each data modality. For medical images in DICOM format, the on-premises engine first scrubs personally identifiable information from all metadata tags. Similarly, for unstructured data like clinical notes, Natural Language Processing (NLP) models are applied to redact personal identifiers before the records are pseudonymised and transferred to the cloud for further processing.

The process begins with the Universal Data Platform, which uses a hybrid-cloud approach to ensure security and privacy. Raw data is first processed by an on-premises engine for pseudonymisation before being securely transferred to the cloud. There, AI-powered services, including Natural Language Processing (NLP) and medical ontologies, parse unstructured text from clinical reports. This information is then transformed and structured according to the FHIR (Fast Healthcare Interoperability Resources) standard, creating a high-quality, harmonised, and analysis-ready dataset.

The Clinician AI Hub provides an interactive environment for clinicians and researchers to explore the curated data. Using data visualisation tools, they can conduct preliminary analyses to assess the quality and suitability of the dataset for a given research question. As a crucial step we ensure that the data is fit for purpose before proceeding to complex modelling using our developed

Preliminary Exploratory Data Analysis (PExDA) framework, reinforcing the platform's commitment to data quality (Supplementary Table S2).

The Data Science Lab offers a comprehensive environment within Microsoft Azure ML Studio for building and validating machine learning models on the high-quality data. To ensure the development of responsible AI, the lab integrates tools that support rigorous traceability and reproducibility (MLflow) and model transparency. Explainable AI (XAI) techniques, such as SHAP (SHapley Additive exPlanations)⁹, are employed to make model predictions interpretable. The platform also incorporates the Microsoft Responsible AI Toolbox (<https://github.com/microsoft/responsible-ai-toolbox>) to assess fairness, evaluate model performance across different patient cohorts, and mitigate potential biases, ensuring that the resulting AI systems are not only accurate but also trustworthy and equitable.

S-RACE is designed to foster multi-institutional collaboration through a flexible, but governed environment. For a given research project, multiple researchers can be granted secure access to develop and test models on high-quality datasets. This collaborative development is enhanced using Microsoft Azure pipelines, which allow for the reuse of code and expertise, promoting efficiency and standardisation across projects. However, development is decoupled from deployment. Models with robust validation, approved by a governance committee, are promoted to a central "Model Registry." This registry facilitates their translation into the clinical world for trials measuring model benefit and monitoring their safety and effectiveness in a pragmatic real-world clinical setting¹⁰. Each registered model is further enriched with comprehensive metadata (Supplementary Table S3), crucial for ensuring transparency, traceability, and responsible governance throughout its lifecycle, which are based on the AIME registry for AI in biomedical research¹¹. AIME provides a standardised framework for documenting AI models—like how clinical trials are registered—to ensure transparency and reproducibility. By adopting this structure, our Model Registry captures essential information like intended use, development data, performance metrics, and validation strategies, which is crucial for traceability and responsible governance throughout the model's lifecycle. This ensures that only robust and trustworthy AI is considered for translation. To facilitate prospective validation or use in clinical trials by collaborators, these approved models can also be deployed as user-friendly web applications, a practice detailed later in our kidney cancer research project. Furthermore, the platform supports two distinct collaboration models: for centralised studies, external researchers can work directly within the secure S-RACE environment, while for scenarios where data cannot be shared, S-RACE supports privacy-preserving Federated Learning based on the NVFlare framework, enabling decentralised model training across institutions¹².

As of September 2025, there are 19 clinical research projects on-going (13 projects with data loaded on the platform, 6 projects to be loaded on the platform after IRB approval) which allowed us to integrate within the AI platform our 5 major IT data sources: EHRs, pathology, lab tests, PACS, eCRF (e.g., RedCap), and disease-specific internal databases for a total of 31276 patients (Figure 1 panel A). The projects span the following domains: oncology (8), cardiovascular disease (6), neuroendocrine disorders (3), neurosciences (2). Examples of type of data imported for some research projects are shown in Figures 1 b-c. A synthesis of the investigated clinical research

questions, the number of included patients, and the type of imported data for each of the project is shown in Table 1.

Research Project	Cohort Details	Endpoints	Design	Data Types
Oncology				
mNSCLC	Stage IV patients with NSCLC receiving first line (chemo)immunotherapy treatment. Size (estimated): 1000	Early treatment failure (PD within 3 months) Long survivors (OS > 30 months)	Retrospective Multicentric (N=18)	Clinical Demographics Histopathology Laboratory Radiology (CT/PET)
ccRCC	Non metastatic ccRCC patients undergoing nephrectomy Size: 3210	Kidney cancer specific survival	Retrospective Multicentric (N=2)	Clinical Demographics Histopathology Laboratory
mCRC	Patients with liver metastases from CRC undergoing hepatic resection. Size (estimated): 800	Early recurrence (within 3 months) after surgery Risk stratification of OS, PFS Response to neo-adjuvant chemotherapy	Retrospective Multicentric (N=2)	Clinical Demographics Histopathology Laboratory

				Radiology (CT)
Liver and bile duct cancer	Patients with locally/advanced liver and bile duct cancer receiving first line chemo-immunotherapy treatment Size: 735	Early recurrence (within 3 months) after treatment Risk stratification of OS, PFS	Retrospective Multicentric (N=46)	Clinical Demographics Genomics Histopathology Laboratory
Prostate Cancer Detection	Patients undergoing magnetic resonance imaging and biopsy for the presence of prostate cancer Size: 2143	Detection rate of clinically significant prostate cancer	Retrospective Monocentric	Clinical Demographics Histopathology Laboratory Radiology (MRI)
Prostate Cancer AKI	Patients undergoing radical prostatectomy Size: 3351	Post-surgery AKI Post-surgery kidney failure	Retrospective Monocentric	Clinical Demographics Histopathology Laboratory
Lung Cancer Screening	Heavy smoker individuals undergoing low-dose chest CT for lung cancer screening	Non-invasive classification of pulmonary nodules.	Retrospective / Prospective Monocentric	Clinical Demographics

	Size: 290 + 1550	Assessment of cardiovascular disease risk		Histopathology Laboratory Radiology (low dose CT)
Pancreatic Cancer	Patients with intraductal papillary mucinous neoplasm (IPMN) of the pancreas Size: 2000 + 700	Development of a multi-omics panel for malignant IPMN identification Prognostic stratification for OS	Prospective Monocentric	Clinical Demographics Genomics Histopathology Laboratory Radiology (CT)
Cardiovascular Disease				
Aortic Stenosis TAVI	Patients with aortic stenosis undergoing TAVI. Size: 440	All-cause mortality/hospitalisation for heart failure one/three years after TAVI	Retrospective Multicentric (N=2)	Clinical Demographics Radiology (cardiac CT)
Aortic Stenosis TAVI	Patients with aortic stenosis undergoing TAVI. Size: 610	Estimate aortic stenosis severity using parameters derived from CT images.	Retrospective Monocentric	Clinical Demographics Radiology (cardiac CT)
Aortic Stenosis TAVI	Patients with aortic stenosis undergoing TAVI.	Automatic CT-based quantification of myocardial extracellular volume	Retrospective Monocentric	Clinical Demographics

	Size:460	All-cause mortality/heart failure one/three years after TAVI		Radiology (cardiac CT)
Cardiovascular Disease	Patients undergoing coronary-CT angiography for suspected coronary artery disease Size: 2250	Identifying novel quantitative CT-based biomarkers for risk stratification Prognostic model for a composite outcome of major adverse events at 4 years	Multicentric (N=2)	Clinical Demographics Radiology (cardiac CT)
Cardiovascular Disease	Patients undergoing coronary-CT angiography for suspected coronary artery disease. Size: 4000	Develop a novel computational model that integrates multi-scale and multi-modal health data Prognostic model for a composite outcome of major adverse events at 3 years	Multicentric (N=2)	Clinical Demographics Genomics Radiology (cardiac CT)
Endocrine Disorders				
T1 DM	Paediatric patients with type 1 diabetes mellitus Size: 688	Partial clinical remission, defined as insulin-dose adjusted haemoglobin A1c ≤ 9	Retrospective Monocentric	Clinical Demographics Laboratory
T2 DM	Patients with type 2 diabetes mellitus who have been prescribed	Patients most likely to benefit from second-line antidiabetic treatments beyond metformin	Retrospective Monocentric	Clinical Demographics Laboratory

	metformin as initial therapy Size: 250			
Internal Medicine	Patients admitted to the internal medicine departments, either outpatient or inpatient services Size (estimated): 6000	Death during hospitalisation Discharge to palliative care Hospital stays greater than 25 days	Prospective Monocentric	Clinical Demographics Laboratory
Neurosciences				
Multiple Sclerosis	Patients who started therapy for MS with a medium efficacy treatment Size: 4114	NEDA loss after medium efficacy treatment	Retrospective Monocentric	Clinical Demographics Radiology (MRI)
Neural Critical Care	Patients with spontaneous intracranial haemorrhage in the intensive care unit Size: 2000	Patient's condition, assessed with GOSE (Glasgow Outcome Scale-Extended) one year after discharge	Retrospective/Prospective Monocentric	Clinical data Demographics Genomics Histopathology Laboratory Radiology (CT/MRI)

To validate the platform's core capability of generating high-quality data, we developed a pre-operative AI model to predict cancer-specific mortality in patients with non-metastatic clear cell renal cell carcinoma (ccRCC)¹³. In the initial 'Business Understanding' phase for this project, we conducted a systematic literature review to define the clinical problem and identify key prognostic variables. The cited review was instrumental for this purpose, providing a comprehensive list of established factors that we subsequently used to validate the successful feature extraction by our RWD processing pipeline. This project served as a direct test of S-RACE's data processing pipeline. We utilised two distinct datasets from over 2,000 patients: a manually curated clinical dataset (eCRF), representing the traditional 'gold standard' for research but with a limited number of variables, and a dataset of raw, unstructured RWD (more than 200 variables) automatically ingested and processed by the S-RACE platform (Figure 2, top panel). The central experiment was to compare the performance of AI models developed on these two data sources. Following automated data processing, models were developed in the Data Science Lab using a hybrid strategy that balanced predictive power with clinical interpretability: a Random Survival Forest model was used for feature selection, and a Cox Proportional Hazard model was compared to end-to-end ML algorithms such as survival trees. The key finding was that models trained on the automatically processed RWD performed comparably to those trained on the manually curated dataset. Furthermore, by applying Explainable AI (XAI) techniques, we not only confirmed the importance of known clinical predictors but also identified novel prognostic variables present only in the raw RWD (Figure 2, medium panel). This result provides strong evidence that the S-RACE platform can successfully transform complex, raw clinical data into reliable, research-grade evidence, thereby overcoming a primary bottleneck in the development of scalable and trustworthy AI for healthcare. The model was then implemented as a Web-based app to further speed up ease of use within the clinic and by external collaborators (Figure 2, bottom panel).

The second example demonstrates how the S-RACE platform's responsible AI capabilities can enhance model development, particularly when working with smaller, more specialised cohorts. Following the CRISP-DM methodology, the 'Business Understanding' phase of the TAVI project involved a systematic review of the literature. The cited review was critical for defining the clinical challenge of predicting treatment futility and identifying existing risk stratification models. This step informed our project objectives and provided the necessary benchmarks for the subsequent 'Evaluation' phase. The project aimed to identify patients with severe aortic stenosis who were unlikely to benefit from Transcatheter Aortic Valve Implantation (TAVI)¹⁴, using a high-quality dataset of approximately 500 patients. Given the limited cohort size, ensuring model robustness was paramount. The S-RACE Data Science Lab enabled the implementation of a sophisticated stratified nested cross-validation strategy, which is critical for generating reliable and unbiased performance estimates from smaller datasets (Figure 3, top panel). More importantly, this project leveraged the platform's integrated tools for responsible AI to move beyond standard accuracy metrics. A decision tree-based error analysis was conducted to automatically identify specific patient subgroups where the model was most likely to make erroneous classifications. By pinpointing these areas of underperformance, this methodology allows for targeted model refinement to improve fairness and clinical utility (Figure 3, bottom panel). To prevent any risk of

data leakage, our error analysis workflow is strictly partitioned. During the iterative development cycle, the analysis is performed exclusively on the validation set to guide model debugging and refinement, while the held-out test set remains untouched. Once all development is complete, the analysis is then applied a single time to the test set in a purely descriptive capacity. In this final stage, it does not guide any model changes but instead serves to complement aggregate metrics, enhancing transparency by documenting the final model's performance across key subgroups. This analysis also serves as a guide for targeted data acquisition; by understanding the characteristics of patient profiles where the model is weakest, we can leverage the S-RACE platform's data ingestion capabilities to automatically retrieve additional, relevant data from the hospital's IT systems. This creates a powerful feedback loop for continuous model improvement, demonstrating a significant step towards developing more precise and equitable AI models for clinical decision support.

A key strength of the S-RACE platform, demonstrated in this research project, is its ability to agnostically ingest all available data for a patient, including large-format data like medical images, and use them to create a “deep patient phenotype”. This enables the extraction of “opportunistic biomarkers” from imaging studies, such as the pre-TAVI planning total body CT scans, that were performed for routine clinical care. The platform automates the analysis of these images, moving beyond standard cardiac measures like annular dimensions, ejection fraction and chamber size. Using deep learning-based segmentation and radiomics, it quantifies features such as the volume and signal distribution of abdominal fat, muscle, and bone, as well as organs like the liver and kidney. This holistic characterisation allows for the identification of subclinical comorbidities and vulnerabilities not captured in standard clinical reports, providing a richer dataset to enhance prognostic models and improve their accuracy. As shown in Table 1, the deep learning image analysis solutions developed within one specific project will be applied to the images of the other cohorts.

DISCUSSION

The S-RACE platform has been developed to address the fundamental challenge in clinical AI: the need for a scalable and governed process to transform raw, heterogeneous hospital data into high-quality, research-grade RWD. The transformation of raw RWD into trustworthy evidence requires more than just technical data cleaning; it demands a rigorous and principled approach to the entire research lifecycle, as outlined in frameworks like the PRINCIPLED checklist for RWD re-use¹⁵. The S-RACE platform was designed to be a comprehensive ecosystem that provides researchers with the functionalities to operationalise such a principled approach. For instance, each study on S-RACE starts with a clear definition of a research question and a systematic literature review within the CRISP-DM framework to support a robust study design. The S-RACE platform provides clinicians with interactive tools in the Clinician AI Hub to define endpoints and cohorts. To address the critical challenge of confounding, S-RACE focuses on creating a holistic view of a patient by integrating multi-modal data, thereby providing a richer set of covariates for adjustment in statistical models. For bias remediation, our 'Responsible AI Development' pillar integrates tools like the Microsoft Responsible AI Toolbox to systematically assess fairness and

identify subgroup underperformance. Finally, transparency and reproducibility are enforced through the mandatory use of MLflow for experiment tracking and a comprehensive AIME-based Model Registry for transparent documentation from development through to deployment.

While other notable platforms and frameworks such as N3C¹⁶, i2b2 transMART¹⁷, MSK-CHORD¹⁸, and Ehrapy¹⁹ share the goal of advancing RWE, S-RACE is distinguished by several key architectural and philosophical choices that prioritise data quality, security, and regulatory readiness, as summarised in Table 2.

Feature	S-RACE	N3C Enclave	i2b2 transMART	MSK-CHORD	Ehrapy
Architecture	Hybrid-cloud with mandatory on-premises pseudonymisation.	Centralised, secure cloud enclave.	Typically, on-premises, with federated query capabilities.	Cloud-based, focused on automated data integration.	Open-source Python framework for EHR data analysis.
Data Privacy Model	Privacy by Design': pseudonymisation occurs before data leaves the hospital environment.	Centralised de-identified patient data.	Data remains at local institutions; federated queries on aggregate counts.	Automated de-identification and extraction pipelines.	Software library; infrastructure-agnostic.
Collaboration Model	Dual support for centralised analysis and federated learning (NVFlare).	Centralised analysis within the secure enclave.	Federated queries and local analysis.	Centralised analysis of integrated data.	User-implemented; not an inherent feature.
Core Ecosystem	Deeply integrated into a single vendor ecosystem (Microsoft Azure).	Mix of tools within the NCATS secure environment.	Open-source ecosystem with multiple plugins and extensions.	Custom-built, disease-specific (oncology).	N/A (software library).
Regulatory Alignment	Proactively designed to align with ISO 42001:2023 and the EU AI Act.	Focused on data use agreements and IRB oversight.	Primarily focused on data standardisation (e.g., OMOP).	Focus on research-grade data curation for clinical outcomes.	N/A (software library).
Common data model (supported ontologies)	FHIR (UMLS)	Custom (N/A)	Custom (N/A)	N/A	N/A

A primary differentiator is our strategic emphasis on data quality as the foundational output. The platform is engineered first and foremost as an engine for data curation. This is exemplified by our hybrid-cloud architecture, which features a mandatory on-premises anonymisation step. This 'privacy by design' approach ensures sensitive patient data never leaves the hospital's secure environment before being pseudonymised. This is not merely a technical choice but a core governance principle that builds institutional, clinical, and patient trust, and it contrasts with models that may transfer raw identifiable data to the cloud, increasing the attack surface and complicating regulatory compliance.

The S-RACE platform's design aligns with key frameworks for trustworthy research. It operationalizes the FAIR principles by ensuring data are Findable and Accessible via a governed hub, Interoperable through the mandatory FHIR standard, and Reusable thanks to comprehensive documentation in the AIME-based Model Registry. The platform's emphasis on detailed metadata in the Data Quality Checklist and Model Registry also adheres to the MINERVA framework²⁰. Finally, S-RACE supports the PRINCIPLED¹⁵ checklist by providing an ecosystem for robust study design, bias remediation using integrated tools, and better handling of confounding through multi-modal data integration.

Furthermore, S-RACE is deeply integrated within the Microsoft Azure ecosystem. This deliberate choice provides a cohesive, enterprise-grade environment that leverages a suite of interoperable tools for every stage of the pipeline. Using a single, secure cloud environment for data processing (Cognitive Health Services), model development (Azure ML Studio), and collaboration simplifies security and identity management, streamlines workflows, and facilitates easier auditing compared to assembling a solution from multiple, disparate vendors. This tight integration ensures both robust performance and a clear chain of custody for data and models.

A crucial aspect of the S-RACE vision is its role as a catalyst for collaborative research. The platform is not merely a technical tool, but a managed ecosystem designed to bring together clinicians and data scientists from multiple institutions. Governance is embedded into the collaborative workflow: each project operates within a segregated workspace with role-based access controls, ensuring that researchers only see the data relevant to their approved study. This structure supports two powerful modes of collaboration. First, it allows for centralised analysis, where external partners can securely access and work with curated, high-quality datasets. Second, it is equipped for privacy-preserving federated learning, enabling the development of more generalisable models by training algorithms across decentralised datasets without ever moving sensitive patient data. This dual capability makes S-RACE a flexible and powerful hub for multi-centre studies, accelerating scientific discovery by creating larger, more diverse virtual cohorts while upholding the highest standards of data protection and project governance.

The versatility of the S-RACE platform is another key strength. Unlike more specialised platforms focused on a single disease area, it is currently populated with disease-specific data for a total of 31,276 patients, powering 19 distinct clinical research projects across diverse domains including oncology, cardiology, and diabetes. This demonstrates the platform's technical scalability and, more importantly, the successful implementation of a standardised, repeatable data curation

pipeline. This proves its value as a central institutional asset that can break down data silos, foster cross-disciplinary research, and maximise the return on investment in data infrastructure.

The clinical examples presented in this paper serve to illustrate these strengths in practice. The kidney cancer research project provides direct validation for our primary mission: by showing that models built on automatically ingested RWD can perform as well as those built on manually curated data, we demonstrate the platform's success as a scalable data curation engine. The aortic stenosis research project highlights the next layer of the platform's capabilities, showing how this high-quality data foundation enables more advanced and responsible AI development. It showcases how S-RACE facilitates the creation of deep patient phenotypes through the extraction of opportunistic biomarkers from imaging data, and how its integrated tools can be used to analyse model fairness and guide a continuous feedback loop of improvement.

Finally, the entire S-RACE framework was built with proactive alignment to the evolving regulatory landscape. Its core features directly address the requirements of standards like ISO 42001:2023 and the EU AI Act. For instance, the centralised model registry provides the versioning and detailed documentation essential for traceability, while the integrated responsible AI tools for error analysis and fairness assessment directly support the risk management and bias mitigation mandates of these regulations. By prioritising the generation of high-quality, reliable data within a responsible and collaborative framework, S-RACE provides a robust, future-proofed solution to accelerate the development and translation of trustworthy AI in medicine.

Beyond its technical capabilities, S-RACE is fundamentally a collaborative ecosystem. It is engineered to unite clinicians and data scientists from multiple institutions, supporting both centralised analysis within its secure environment and privacy-preserving federated learning for studies where data cannot be shared. This collaborative union is achieved by providing two distinct, purpose-built environments that operate on the same governed data foundation: The Clinician AI Hub a no-code, interactive environment, allows clinicians to use data visualisation tools to explore cohorts and assess data quality without requiring programming knowledge; The Data Science Lab is a parallel environment providing data scientists with a comprehensive suite of tools in Microsoft Azure ML Studio for advanced model development and validation. This dual-environment structure effectively bridges the expertise gap. Clinicians can define clinically meaningful problems using accessible tools, while data scientists can apply rigorous computational methods to the exact same curated data. This dual capability establishes S-RACE as a powerful hub for multi-centre research, accelerating the creation of more generalisable and robust AI models. The importance of this work lies in its demonstration of a robust, governed, and scalable environment for RWD curation. S-RACE provides a trustworthy foundation to accelerate the development and clinical adoption of responsible AI. The deep integration with Microsoft Azure raises important considerations regarding data privacy and vendor interoperability, which we address through specific governance and technical choices. First, we state unequivocally that Microsoft, as the cloud provider, has no technical or legal access to any patient data hosted within the S-RACE platform; all data remains under the exclusive control of our institution. Second, to mitigate vendor lock-in and ensure interoperability, the platform relies on open standards. All curated data are structured in the Fast Healthcare Interoperability

Resources (FHIR) format, ensuring it can be exported and used in other systems. Furthermore, our support for the open-source NVFlare framework enables privacy-preserving federated learning, allowing for direct collaboration with institutions regardless of their underlying infrastructure, thus promoting a vendor-neutral research ecosystem.

A key limitation of the current S-RACE platform is its primary focus on predictive / prognostic modelling rather than formal causal inference. While the integrated Microsoft Responsible AI libraries provide tools for related tasks—such as generating individualised counterfactual explanations with DiCE or estimating population-level treatment effects with EconML—the platform does not yet automate the rigorous design required for robust causal claims, such as systematic confounder selection or formal "prediction under intervention" analyses. Establishing a full causal inference framework remains a valuable direction for future work. Furthermore, a related challenge for any deployed model that requires further investigations is the potential for performance degradation due to distributional changes over time. The S-RACE platform is designed to mitigate this risk through its governance structure. Every model promoted to the central Model Registry is registered with metadata establishing a baseline reference for its training data distribution and performance. Our post-deployment protocol includes the continuous monitoring of outcome distributions and model calibration against this baseline. Any significant deviation is flagged to a governance committee, which can trigger model recalibration or retraining to ensure its continued safety and efficacy in a real-world clinical setting. S-RACE uses pseudonymisation for secure, longitudinal clinical research data linkage, a capability precluded by full anonymisation. The platform ensures GDPR compliance via a 'privacy by design' architecture, storing the vital mapping key securely on-premises, separate from the cloud-processed data. However, research on full anonymisation techniques (e.g. k-anonymisation and synthetic data) is currently ongoing but not yet released. Another limitation is that curating Real-World Data (RWD) by excluding records with poor quality or missing information introduces a risk of selection bias, as the excluded patients may systematically differ from the final study cohort. To mitigate this, we employ the platform's Exploratory Data Analysis (EDA) tools to thoroughly characterize and compare both the excluded and included populations, thereby ensuring transparency regarding any potential bias. Furthermore, we minimize exclusions—typically limiting them only to records missing a primary outcome—by using imputation techniques. We also address the risk that researchers might unintentionally build models that merely confirm pre-existing hypotheses. This is mitigated through two primary safeguards: a structured research process guided by the CRISP-DM framework, and the mandatory application of Explainable AI (XAI) techniques, such as SHAP. XAI significantly enhances transparency by revealing whether a model is relying on spurious correlations or on the hypothesized clinical factors, which directly challenges and validates researchers' underlying assumptions.

METHODS

The platform's robust AI capabilities are powered by a suite of Microsoft technologies, forming a comprehensive ecosystem for RWD processing and AI development²¹:

- Microsoft Cognitive Health Services: Utilised for advanced Natural Language Processing (NLP) and the application of medical ontologies, these services are crucial for extracting structured, clinically relevant information from unstructured clinical text, such as physician notes and reports. The NLP pipeline is based on the commercial product by Microsoft Text Analytics for Health (TA4H, <https://tinyurl.com/mxnysfav>) which performs anonymisation of the clinical notes and then extracts medical concepts and relations among these concepts using standard ontologies such as UMLS (Unified Medical Language System). The common data model is the FHIR (Fast Healthcare Interoperability Resources). Additional data models such as OMOP can be obtained, for example, using conversion tools (e.g., FHIR to OMOP, <https://build.fhir.org/ig/HL7/fhir-omop-ig/>).
- Microsoft Power BI (Business Intelligence): Integrated within the Clinician AI Hub, Power BI enables intuitive data visualisation and preliminary analysis, allowing clinicians to explore insights from RWE in an accessible format.
- Microsoft Azure ML Studio: This comprehensive and scalable environment supports the entire ML model development and deployment lifecycle within the Data Science Lab, providing data scientists with the tools needed for robust model creation. Azure Machine Learning Studio allows to model any type of clinical questions from regression to classification and survival analysis by using both a “white box” modelling approach (e.g., logistic regression, Cox models) and “black box” approaches (e.g., random forests). In this last scenario, explainability tools are used for clinical understanding of the results.
- Standardised Explainability Techniques: A core component of the Data Science Lab, these techniques are employed to enhance the transparency and interpretability of AI models, addressing the 'black box' challenge and building trust among clinical users.

The S-RACE platform integrates data from five primary institutional IT systems to build a comprehensive, multi-modal data foundation for research. The core data types include: (i) Electronic Health Records (EHRs), providing clinical and demographic information; (ii) Pathology and Laboratory systems, providing histopathology and lab test results; (iii) the Picture Archiving and Communication System (PACS), for medical imaging such as CT, PET, and MRI; (iv) Genomics data for multi-omics studies; and (v) research-specific sources like electronic Case Report Forms (eCRFs) and internal databases. This integration enables the creation of a multimodal view of patients' data for everyone in the research cohort. The specific data types utilised for each of the 19 ongoing clinical research projects are detailed in Table 1.

The platform's architecture is fundamentally shaped by a 'privacy by design' approach, in full compliance with the EU's General Data Protection Regulation (GDPR). To correct an inconsistency in the original manuscript, we clarify that the process employed is strictly pseudonymisation, as legally defined under GDPR. The on-premises data ingestion engine automatically removes all direct patient identifiers (e.g., name, national health service number, medical record number) and replaces them with a unique, irreversible cryptographic hash. The mapping key that links this pseudonym back to the original patient identifier is stored exclusively within the hospital's secure

on-premises infrastructure and is never exposed to the cloud environment. This technical segregation is the primary control that substantiates our 'privacy by design' claim, serving as an auditable safeguard that ensures sensitive patient data is protected before leaving the hospital's trusted domain. This approach is the specific technical safeguard that enables our legal basis for processing under GDPR: scientific research that allows for longitudinal patient follow-up, which requires re-linkability. While our primary compliance framework is GDPR, these technical and organisational measures also align with the core principles of other international standards, such as the security and privacy rules within HIPAA. Furthermore, our use of Microsoft Azure services for data processing is governed by a formal Data Protection Addendum (DPA). This contractual agreement legally obligates Microsoft to adhere to its responsibilities as a data processor under GDPR, ensuring that all data handling meets the required compliance standards.

Although the implementation presented in this paper relies on Microsoft Azure services, the underlying methods and workflows (e.g., data ingestion, storage, processing, and analysis pipelines) are cloud-agnostic. Equivalent services exist in other cloud providers (e.g., AWS S3 + EMR, Google Cloud Storage + Dataproc) or can be deployed on-premises using containerized solutions (e.g., Kubernetes + Spark). Therefore, the approach described is portable and not limited to Microsoft Azure.

DATA AVAILABILITY

Data can be made available under request. We can grant access to the S-RACE platform to the to access the data within the secure environment of our Cloud.

CODE AVAILABILITY

For each research project which will be submitted as original publication, the developed source code will be shared with the editors and reviewers as a private repository on the institutional GitHub and shared during the review process. Under request, we can grant access to specific research projects to their Azure ML studio environments for collaboration purposes.

ACKNOWLEDGEMENTS

This work was supported by the Supported by the Italian Ministry of Research, Complementary Actions to the NRRP "D34health - Digital Driven Diagnostics, prognostics and therapeutics for sustainable Health care" Grant (# PNC0000001).

AUTHOR CONTRIBUTIONS

AT, CT, AE: DESIGN, WRITING, SUPERVISION, REVIEW. DT, AC, AD, ELD, BF, PS, SB, MA, MUA, DC, FC, LC, GC, AC, SC, ML, MM, FO, AP, FP, DS, DV, DVI, ALV, AZ, CC, MD: WRITING, REVIEW. All the authors have read and approved the manuscript.

COMPETING INTERESTS

The authors declare no competing interests.

REFERENCES

1. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nat Med* **28**, 31–38 (2022).
2. Purpura, C. A., Garry, E. M., Honig, N., Case, A. & Rassen, J. A. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. *Clin Pharmacol Ther* **111**, 135–144 (2022).
3. Sun, Y., Sheng, D., Zhou, Z. & Wu, Y. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. *Humanit Soc Sci Commun* **11**, 1278 (2024).
4. Embí, P. J., Rhew, D. C., Peterson, E. D. & Pencina, Michael. J. Launching the Trustworthy and Responsible AI Network (TRAIN): A Consortium to Facilitate Safe and Effective AI Adoption. *JAMA* **333**, 1481 (2025).
5. Van Genderen, M. E., Kant, I. M. J., Tacchetti, C. & Jovinge, S. Moving Toward Implementation of Responsible Artificial Intelligence in Health Care: The European TRAIN Initiative. *JAMA* **333**, 1483 (2025).
6. *Ethics and Governance of Artificial Intelligence for Health: Large Multi-Modal Models. WHO Guidance.* (World Health Organization, Geneva, 2024).
7. Golpayegani, D., Pandit, H. J. & Lewis, D. Comparison and Analysis of 3 Key AI Documents: EU's Proposed AI Act, Assessment List for Trustworthy AI (ALTAI), and ISO/IEC 42001 AI Management System. in *Artificial Intelligence and Cognitive Science* (eds. Longo, L. & O'Reilly, R.) vol. 1662 189–200 (Springer Nature Switzerland, Cham, 2023).
8. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance.* (World Health Organization, Geneva, 2021).
9. Loyola-Gonzalez, O. Black-Box vs. White-Box: Understanding Their Advantages and Weaknesses From a Practical Point of View. *IEEE Access* **7**, 154096–154113 (2019).
10. You, J. G., Hernandez-Boussard, T., Pfeffer, M. A., Landman, A. & Mishuris, R. G. Clinical trials informed framework for real world clinical implementation and deployment of artificial intelligence applications. *NPJ Digit Med* **8**, 107 (2025).
11. Matschinske, J. *et al.* The AIME registry for artificial intelligence in biomedical research. *Nat Methods* **18**, 1128–1131 (2021).
12. Roth, H. R. *et al.* NVIDIA FLARE: Federated Learning from Simulation to Real-World. <https://doi.org/10.48550/ARXIV.2210.13291> (2022) doi:10.48550/ARXIV.2210.13291.
13. Usher-Smith, J. A. *et al.* Risk models for recurrence and survival after kidney cancer: a systematic review. *BJU International* **130**, 562–579 (2022).
14. Webb, J. G. *et al.* TAVI in 2022: Remaining issues and future direction. *Archives of Cardiovascular Diseases* **115**, 235–242 (2022).
15. Desai, R. J. *et al.* Process guide for inferential studies using healthcare data from routine clinical practice to evaluate causal effects of drugs (PRINCIPLED): considerations from the FDA Sentinel Innovation Center. *BMJ* e076460 (2024) doi:10.1136/bmj-2023-076460.
16. Haendel, M. A. *et al.* The National COVID Cohort Collaborative (N3C): Rationale, design, infrastructure, and deployment. *Journal of the American Medical Informatics Association* **28**, 427–443 (2021).
17. Firnkorn, D., Merker, S., Ganzinger, M., Muley, T. & Knaup, P. Unlocking Data for

Statistical Analyses and Data Mining: Generic Case Extraction of Clinical Items from i2b2 and tranSMART. *Stud Health Technol Inform* **228**, 567–571 (2016).

18. Jee, J. *et al.* Automated real-world data integration improves cancer outcome prediction. *Nature* **636**, 728–736 (2024).
19. Heumos, L. *et al.* An open-source framework for end-to-end analysis of electronic health record data. *Nat Med* **30**, 3369–3380 (2024).
20. Gini, R. *et al.* Metadata for Data Discoverability and Study Replicability in observational Studies (MINERVA): Lessons Learnt From the MINERVA Project in Europe. *Pharmacoepidemiol Drug Saf* **33**, e5884 (2024).
21. The Transformative Role of Microsoft Azure AI in Healthcare. *IJETER* **12**, 108–113 (2024).

FIGURE LEGENDS

Figure 1. Panel A (top): Overview of the total number of unique patients automatically imported in the S-RACE platform for each clinical research projects (research projects can share the same patients for similar inclusion criteria). Each project contributed to improve the automated ingestion of related RWD from our hospital's data warehouse. There are 31276 unique patients loaded in the platform, corresponding to 13 research projects. Multiple research projects can share the same type of patients but tackle different research questions. Follow up data, when available, are automatically updated with a night batch routine. Panel B (bottom-left): examples of PowerBI dashboards of four oncology research projects (metastatic lung cancer [mNSCLC], kidney cancer [ccRCC], metastatic colorectal cancer [mCRC], liver and bile duct cancer), multiple sclerosis, type 1 diabetes mellitus [T1DM] and the detailed type of data available for a specific research project. The number of patients indicated represents the final size of the original cohort used to develop the ML models, obtained after updating the inclusion criteria based on the refinement clinical research questions and the automated pre-processing pipeline for quality check of the original data (e.g., removal of patients with missing outcome data). Panel C (bottom-right): examples of PowerBI dashboards for the research projects spanning the medical imaging domain. These projects were fundamental in supporting the automated integration of the hospital's PACS in the S-RACE platform. Abbreviations: AKI (Acute Kidney Injury), TAVI (Transcatheter Aortic Valve Implantation), CAD (Cardiovascular Disease), ccRCC (Clear Cell Renal Cell Carcinoma), mCRC (Metastatic Colorectal Cancer), mNSCLC (Metastatic Non-Small Cell Lung Cancer), T1DM (Type-1 Diabetes Mellitus), T2DM (Type-2 Diabetes Mellitus).

Figure 2 schematic summary of the ccRCC research project. **Top panel:** overview of the ML design process. Different ML models were employed considering a combination of “black box” models as preliminary feature selectors (e.g., Survival Random Forest). Selected features are then combined using “white box” models such as Cox Proportional Hazard. Developed models have then been externally validated (TRIPOD type III) and benchmarked against state-of-the art clinical models (e.g., GRANT) endorsed by clinical societies. **Medium panel:** all the models, during training and validation, are always challenged with explainability protocols. Here we show the use of feature importance and SHAP values to rank the eight prognostic pre-operative features. **Bottom panel:** the final model was released as a Web-based app to further promote prospective or additional external validation. Data can be inserted manually using the insertion form (panel A). For each individual patients, results are reported: risk assignment and individual KM curve compared to the KM of the cohort used to train the model (panel A), SHAP individual explanations (panel B), and distribution of single patient’s values with respect to the training cohort (panel C).

Figure 3: schematic summary of the severe aortic stenosis research project. **Top left panel:** overview of the ML analysis pipeline. The ML models are developed on the internal cohort using a nested cross validation. Nested cross validation was chosen because it provides an unbiased and more reliable estimate of a model's true performance by separating hyperparameter tuning from final evaluation, preventing data leakage and promoting robust generalisation compared to standard cross validation. Following a similar approach of the kidney cancer research project, the best performing model is then re-trained on all the internal cohort and then externally validated. **Top right panel:** application of the responsible AI error decision tree. Based on the features included in the final model, a decision tree is trained to optimise the split among subpopulations with the highest / lowest number of erroneous classifications.

TABLE LEGEND

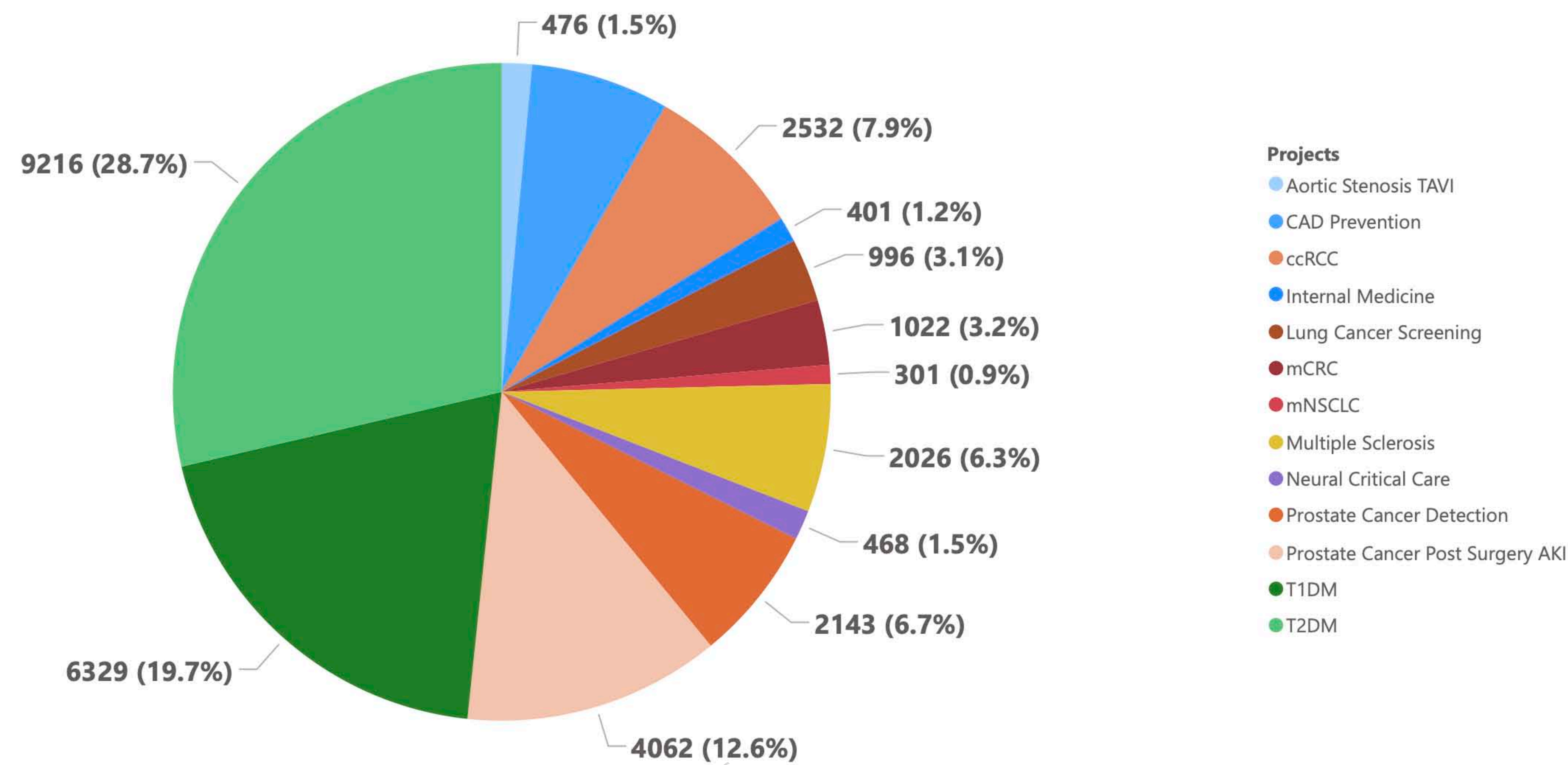
Table 1: overview of the on-going clinical research projects and the type of data used. For some projects, the sample size would likely be increased due to the additional inclusion of external centres. For monocentric projects, external cohorts will be collected for models’ validation (TRIPOD type III models) or the validation will be performed by sharing the Web-based version of the model or using a federated learning approach (TRIPOD type IV models). Abbreviations: AKI (Acute Kidney Injury), TAVI (Transcatheter Aortic Valve Implantation), CAD (Cardiovascular Disease), ccRCC (Clear Cell Renal Cell Carcinoma), mCRC (Metastatic Colorectal Cancer), mNSCLC (Metastatic Non-Small Cell Lung Cancer), T1DM (Type-1 Diabetes Mellitus), T2DM (Type-2 Diabetes Mellitus), CT (Computed Tomography), PET (Positron Emission Tomography), MRI (Magnetic Resonance Imaging)

Table 2: comparison of various platforms and frameworks designed for the ingestion and use of RWD.

ARTICLE IN PRESS

Total number of unique patients on S-RACE Platform is **31.276**

Number of patients by project



Projects

- Aortic Stenosis TAVI
- CAD Prevention
- ccRCC
- Internal Medicine
- Lung Cancer Screening
- mCRC
- mNSCLC
- Multiple Sclerosis
- Neural Critical Care
- Prostate Cancer Detection
- Prostate Cancer Post Surgery AKI
- T1DM
- T2DM

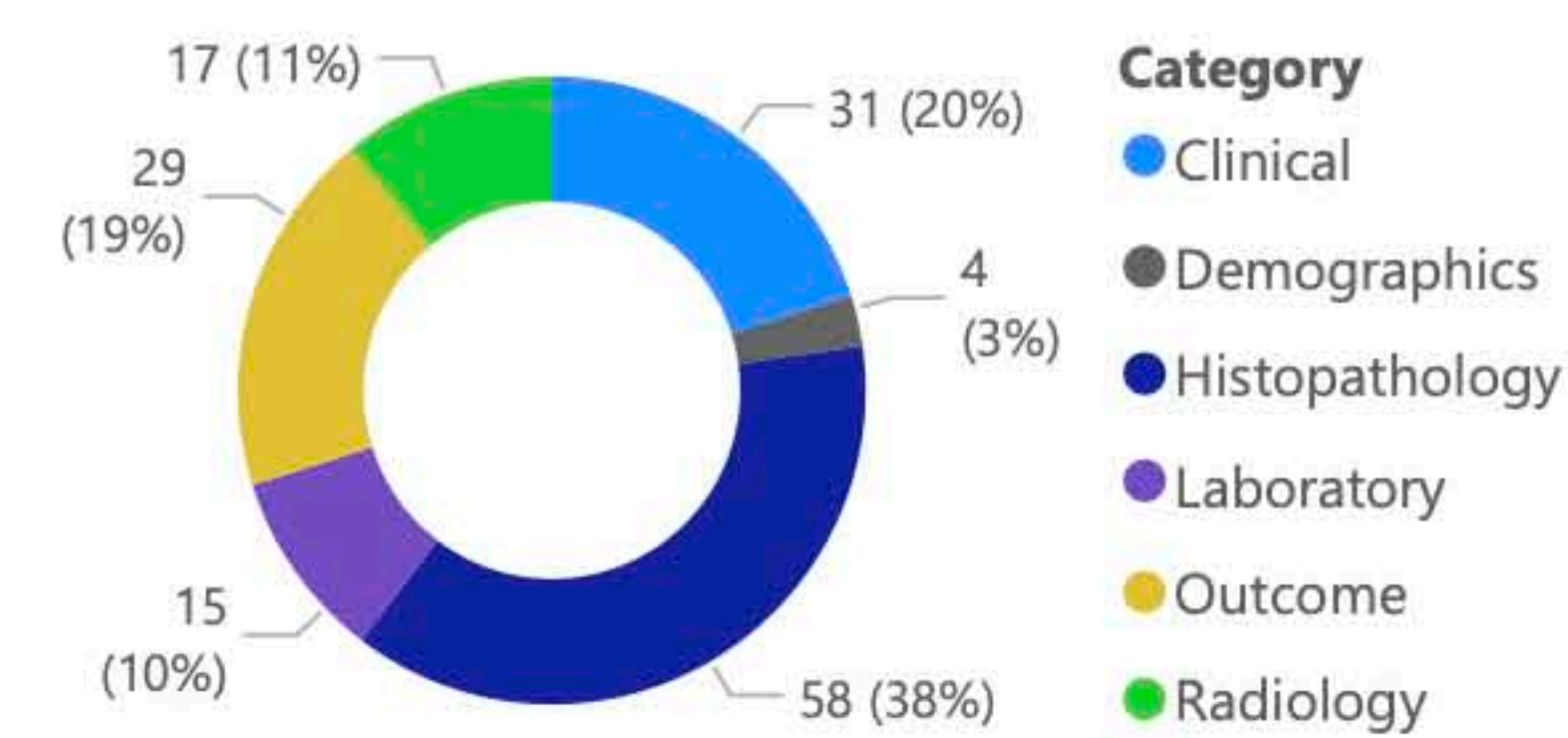
A

Total number of patients on S-RACE Platform is **31.276**

This dashboard focuses on the **6 ML studies** present of the Platform out of 19 in general

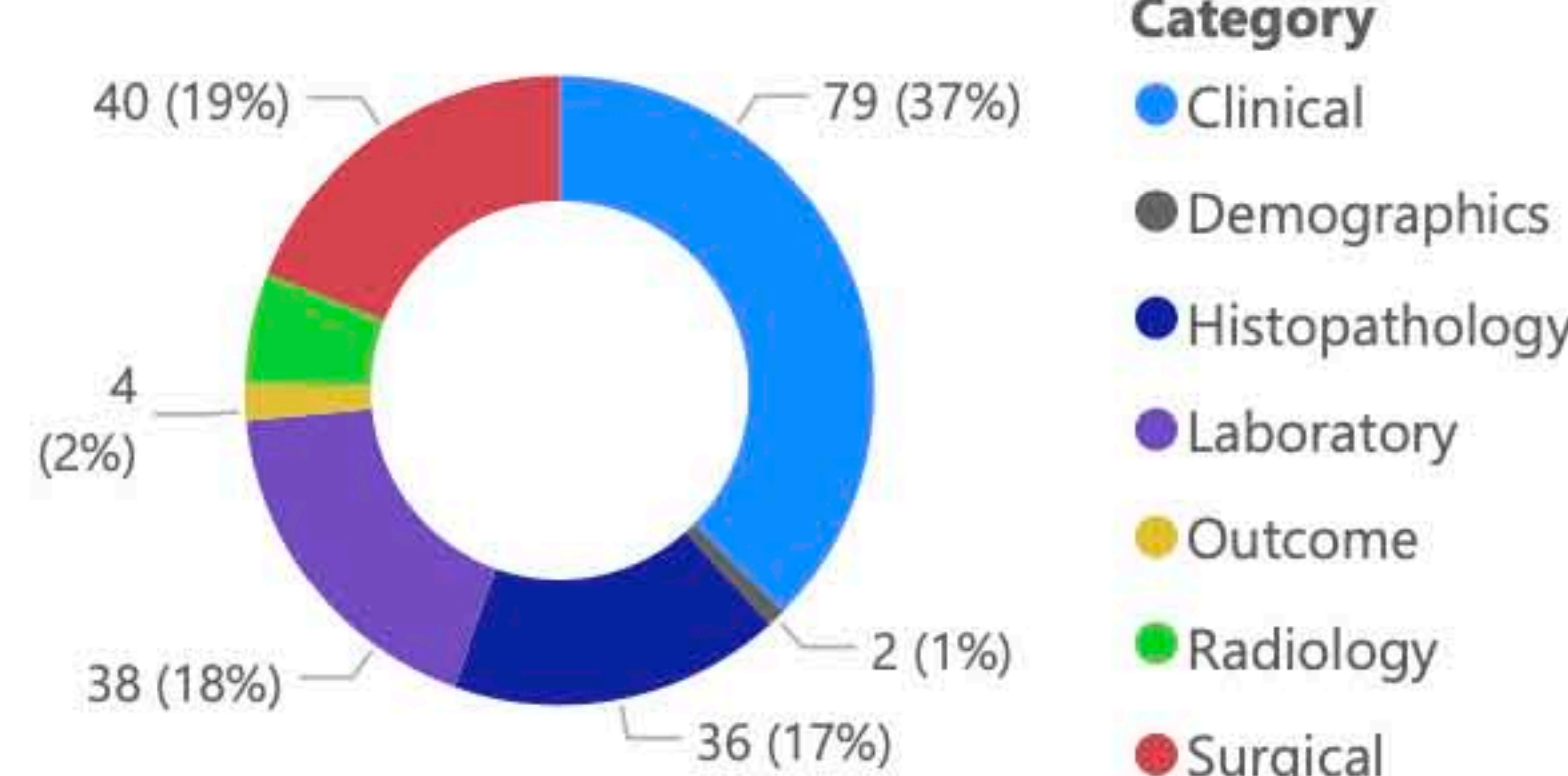
Total Patients - mNSCLC **274**

Variables category



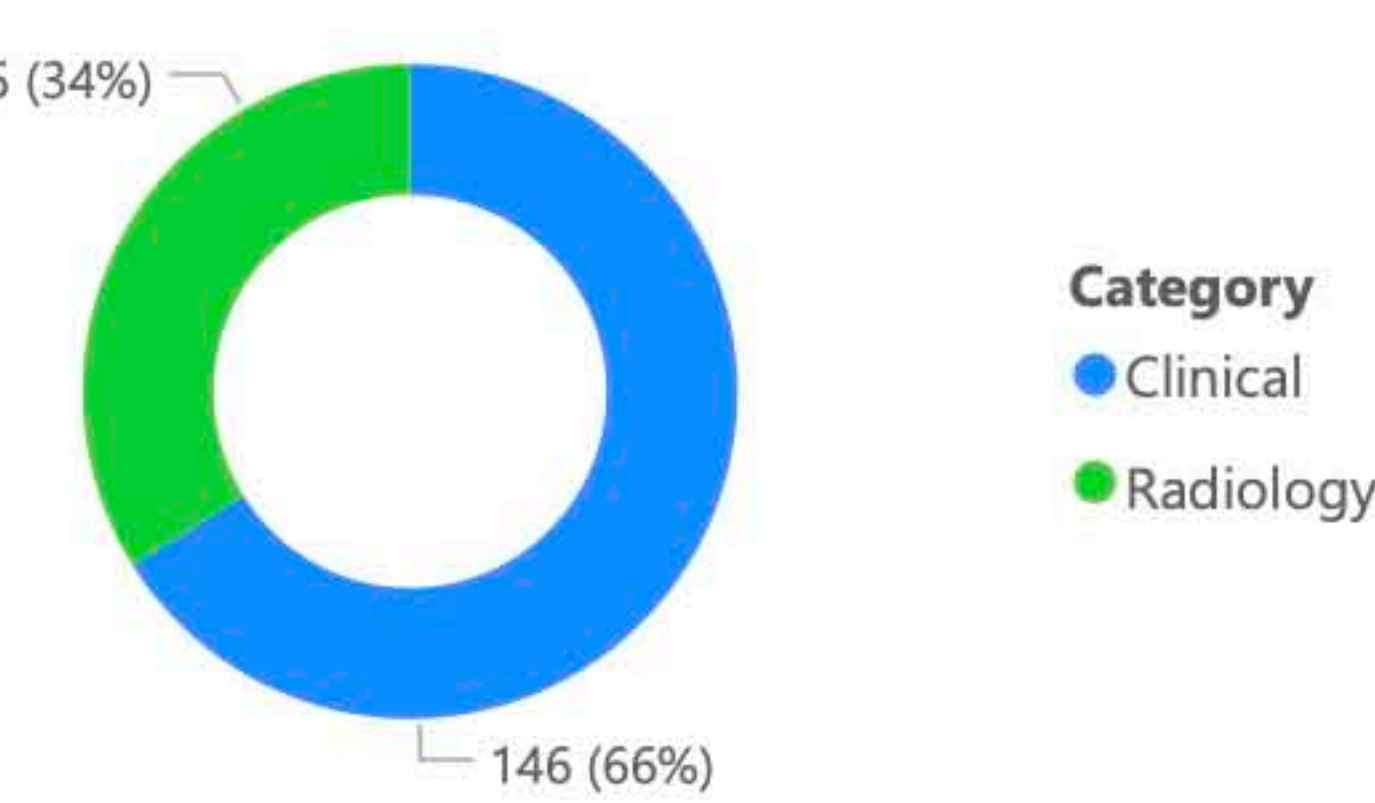
Total Patients - ccRCC **2,536**

Variables category



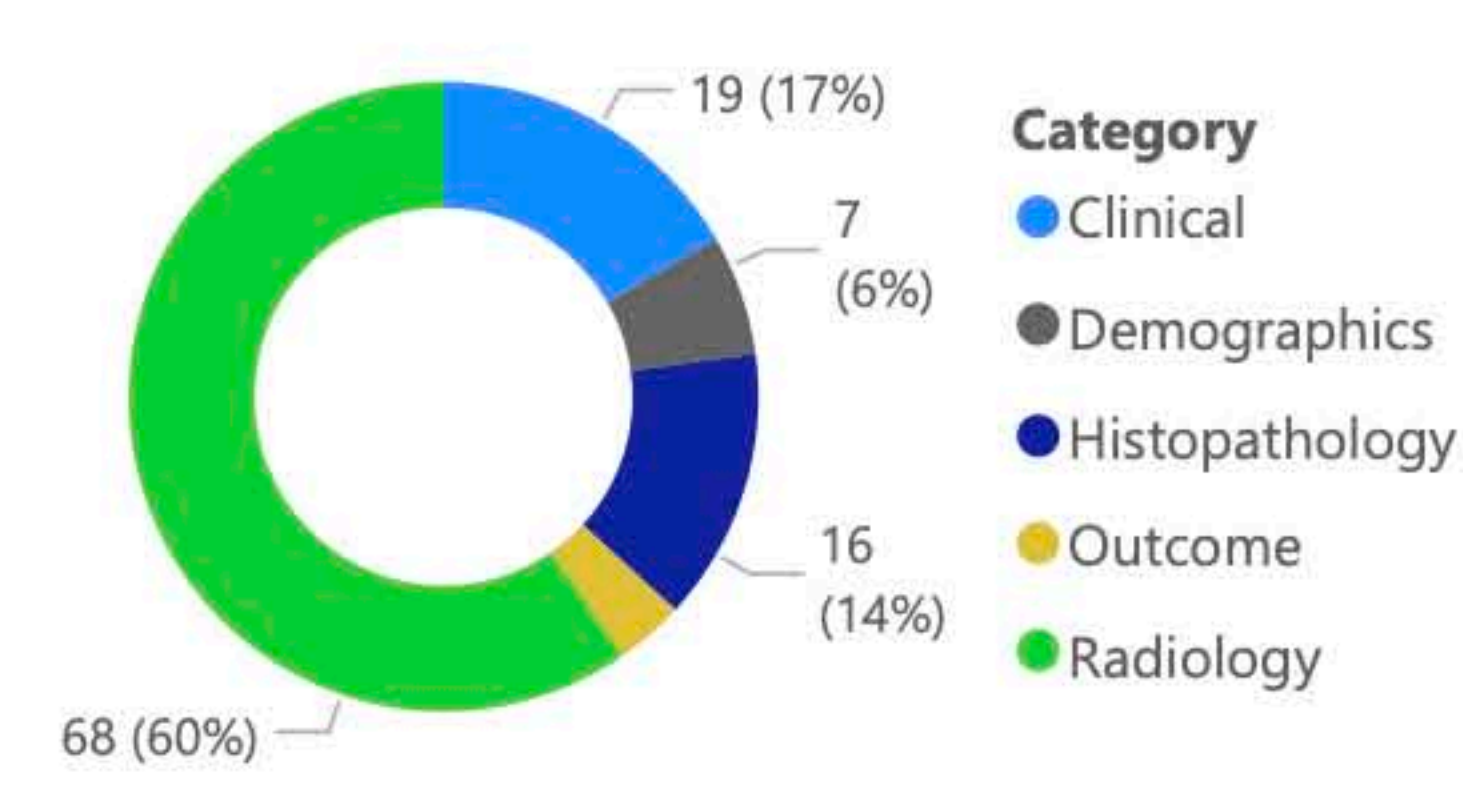
Total Patients - Multiple Sclerosis **4,114**

Variables category



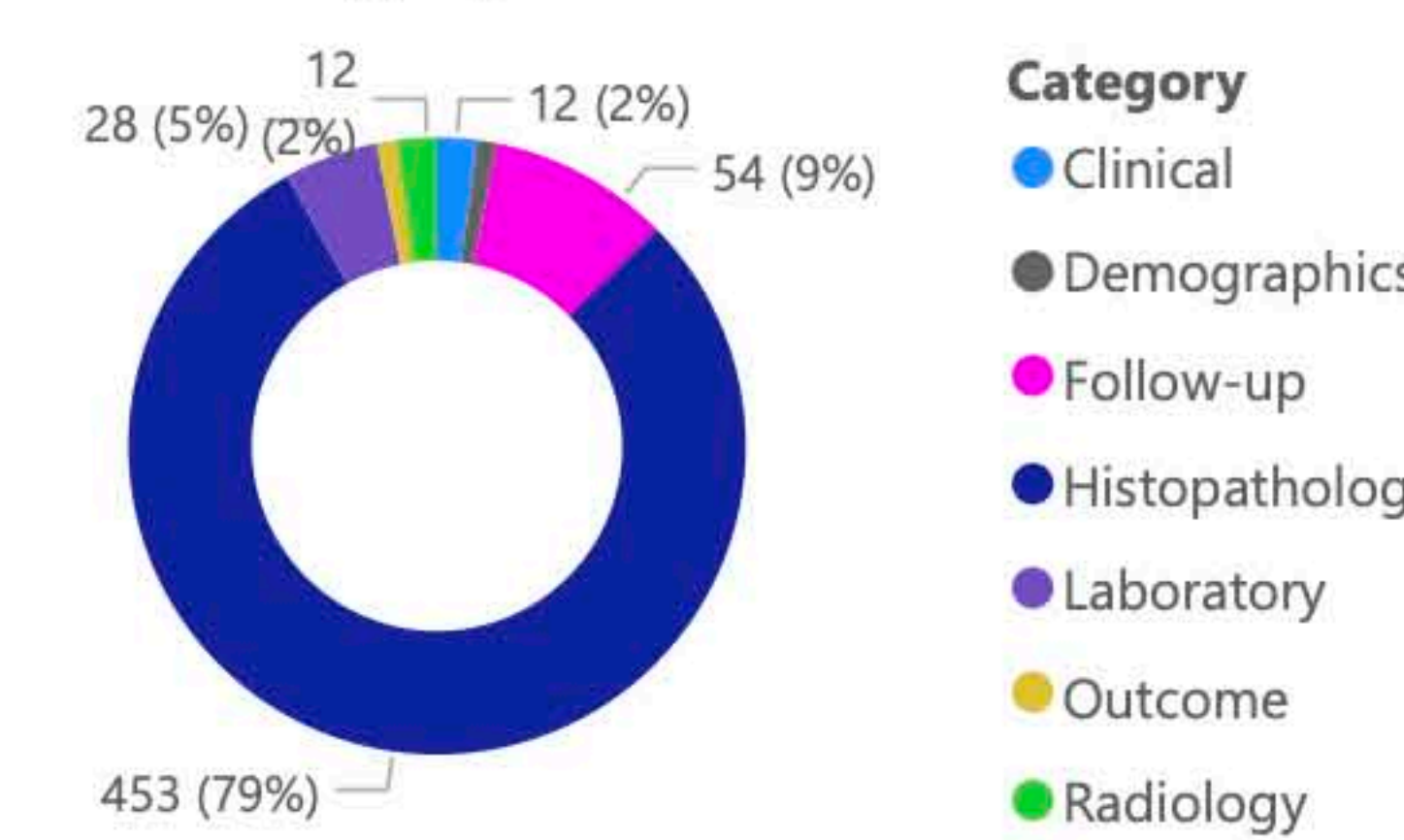
Total Patients - mCRC **446**

Variables category



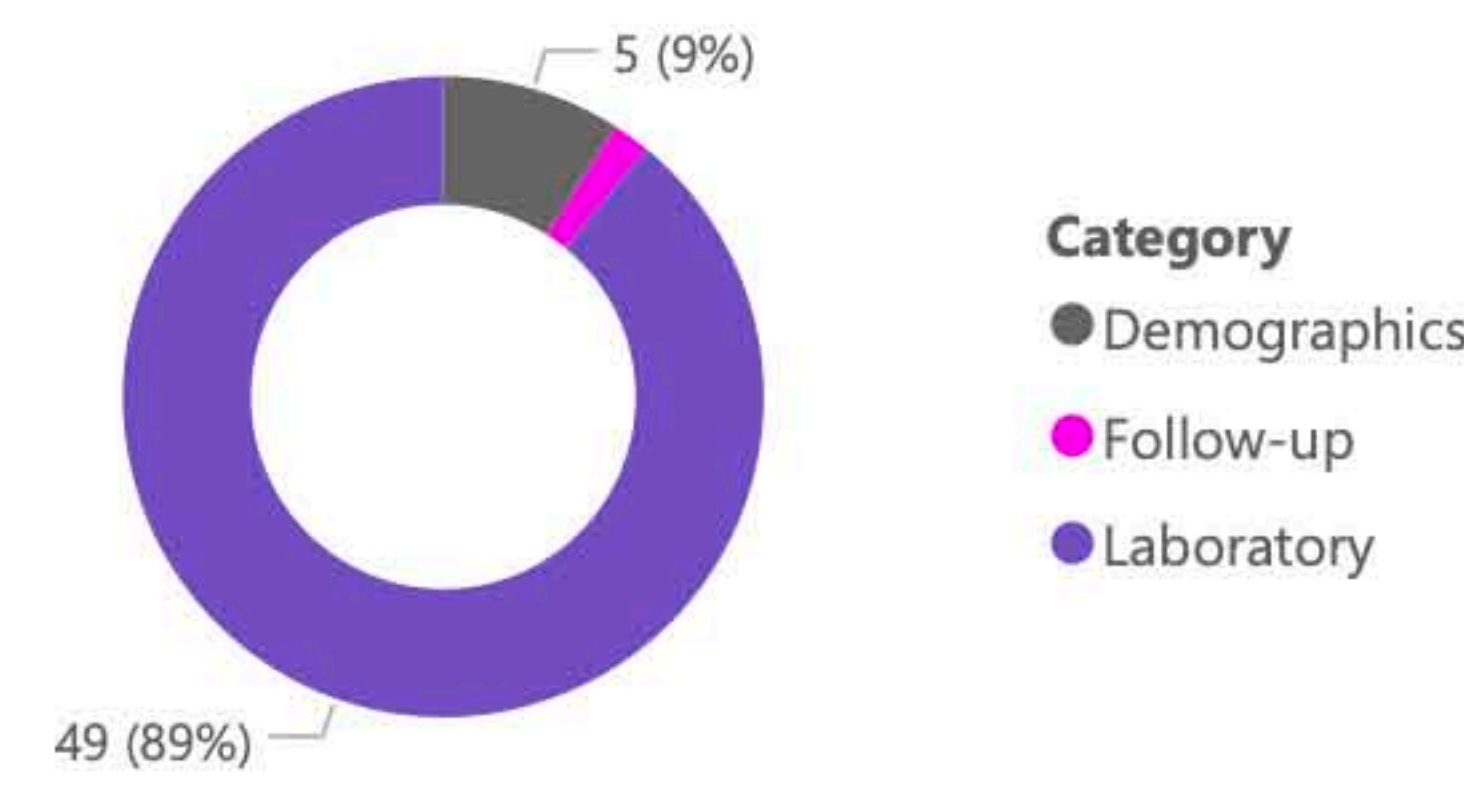
Total Patients - Liver & Bile Duct Cancer **735**

Variables category



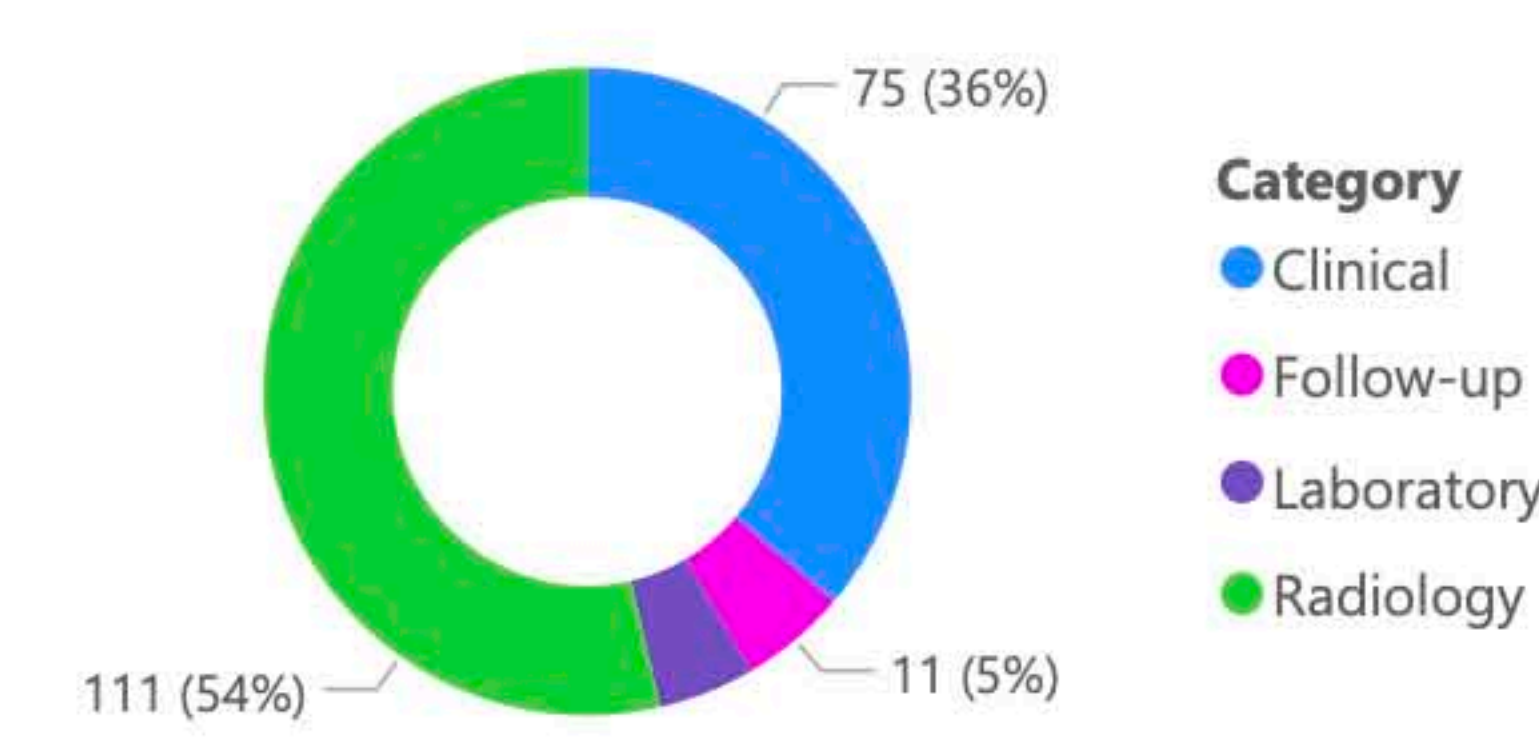
Total Patients - T1DM **688**

Variables category



Total Patients - Aortic Stenosis TAVI **440**

Variables category

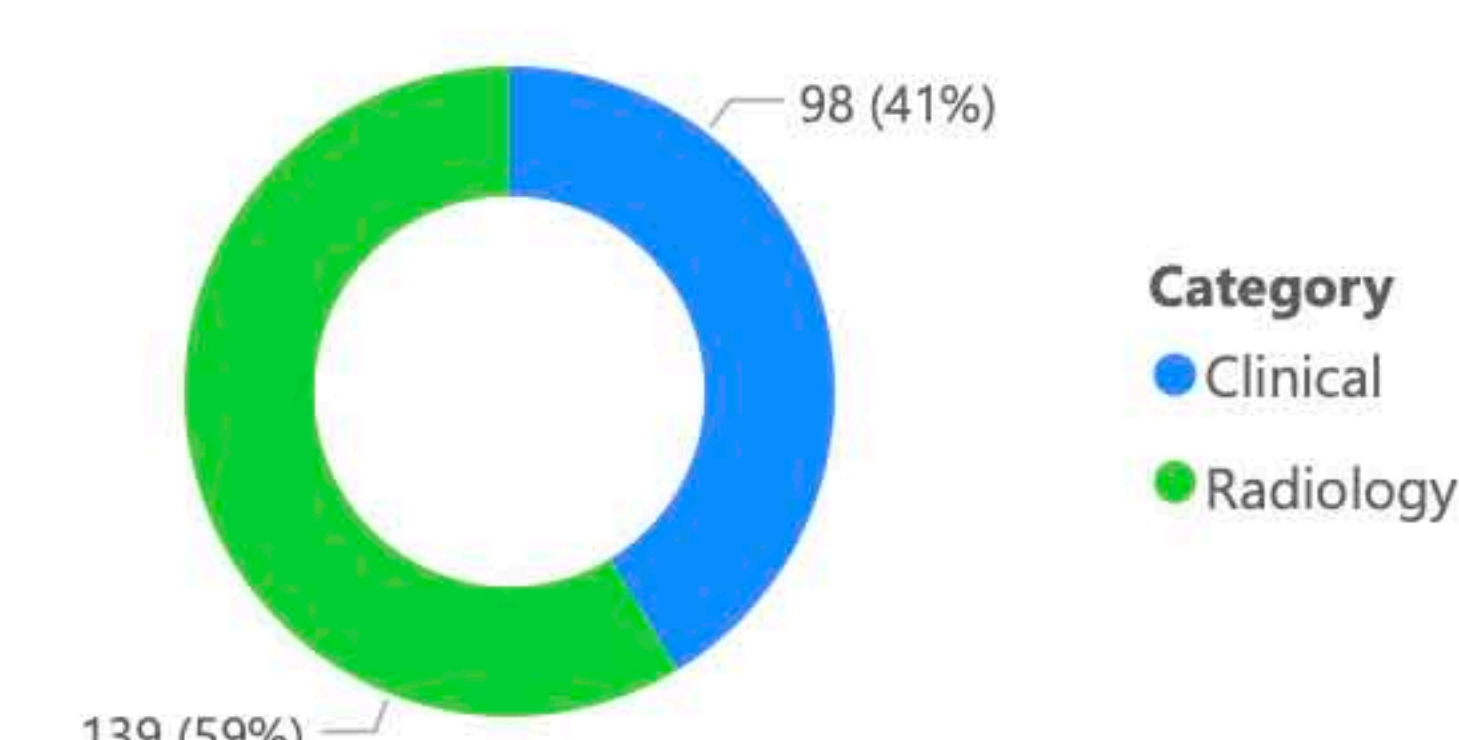


Total number of patients on S-RACE Platform is **31.276**

This dashboard focuses on **6 imaging studies** present of the Platform out of 19 in general

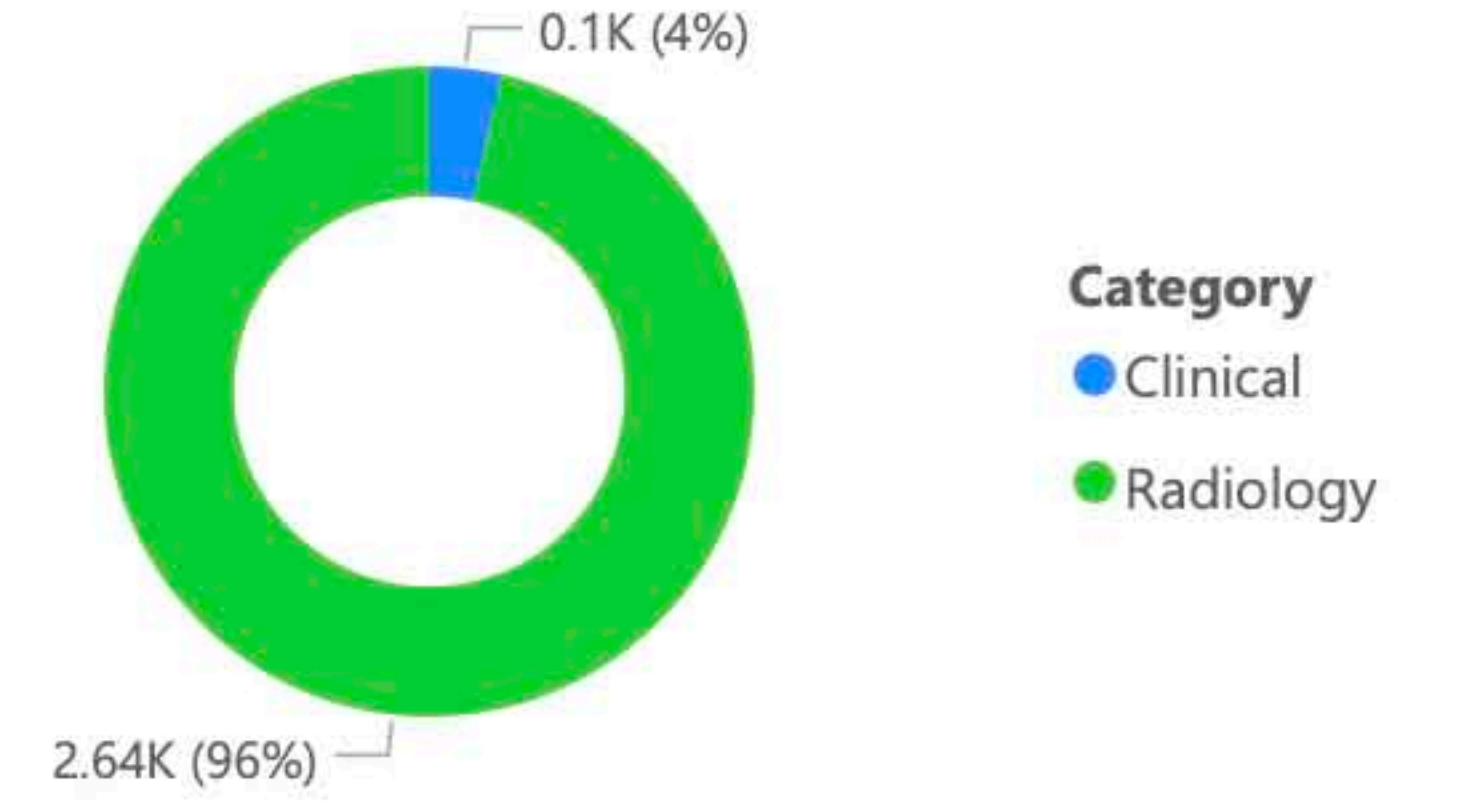
Total Patients - CAD Prevention **2,700**

Variables category



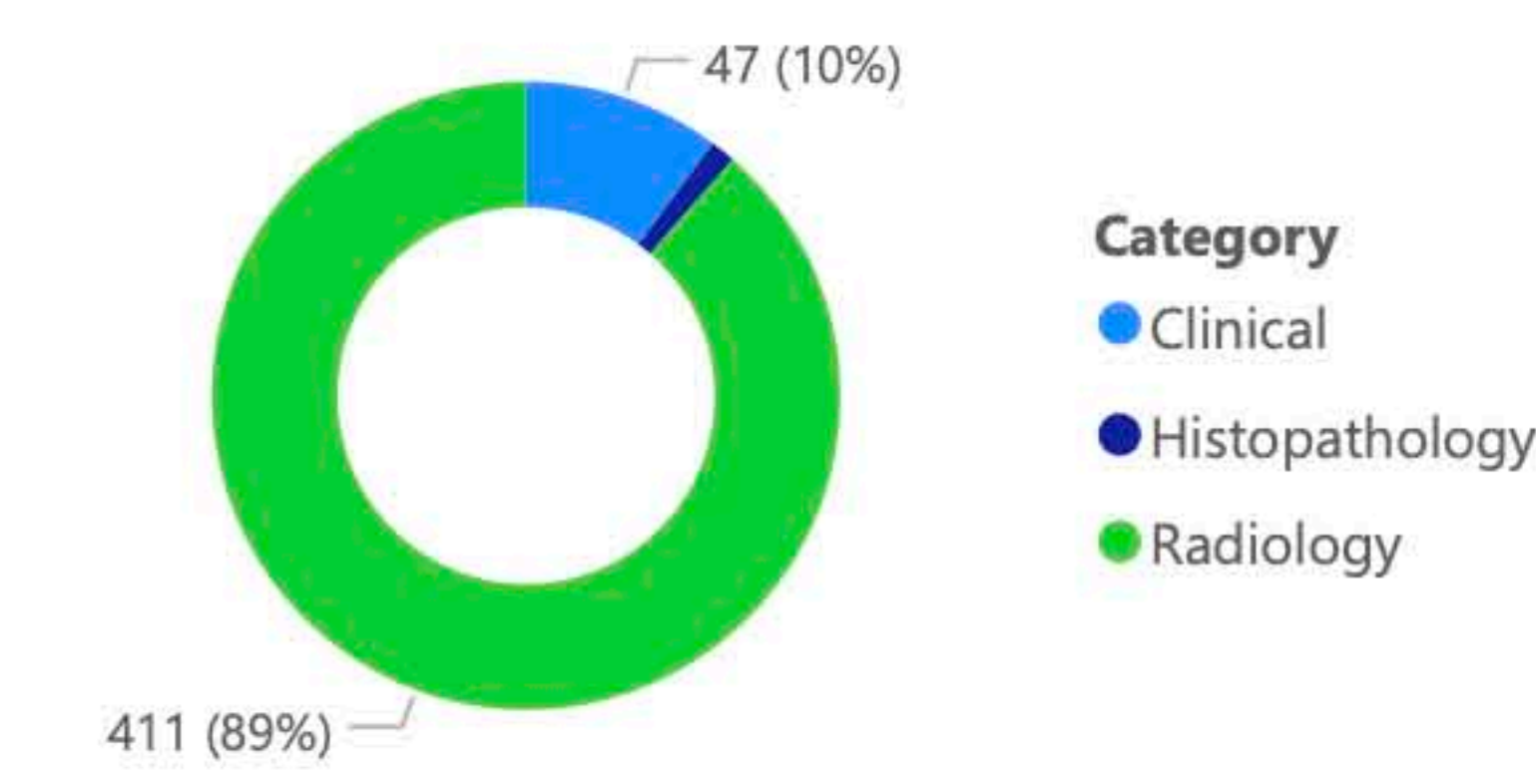
Total Patients - Cardiovascular Disease **2,250**

Variables category



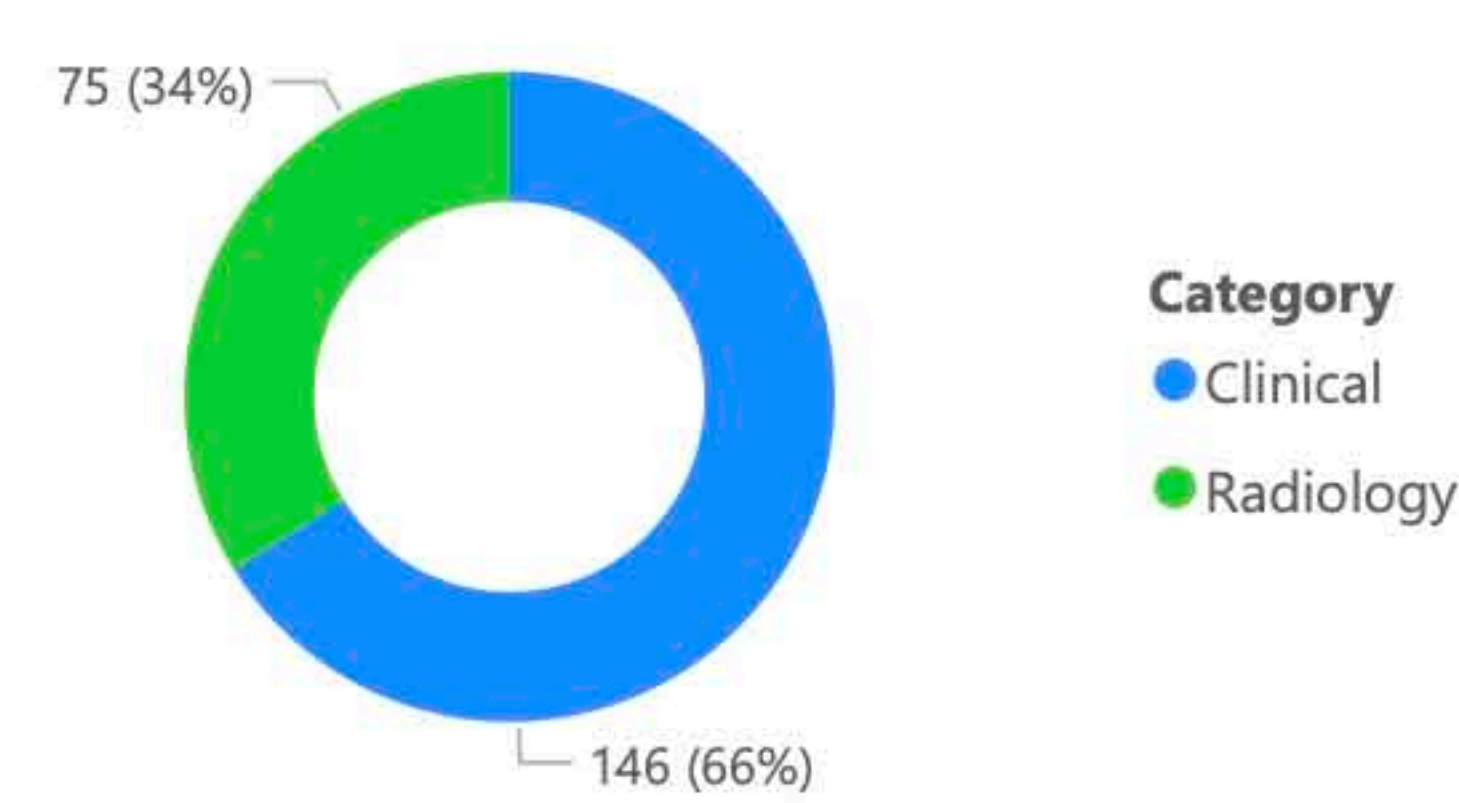
Total Patients - CAD Radiomics **4,000**

Variables category



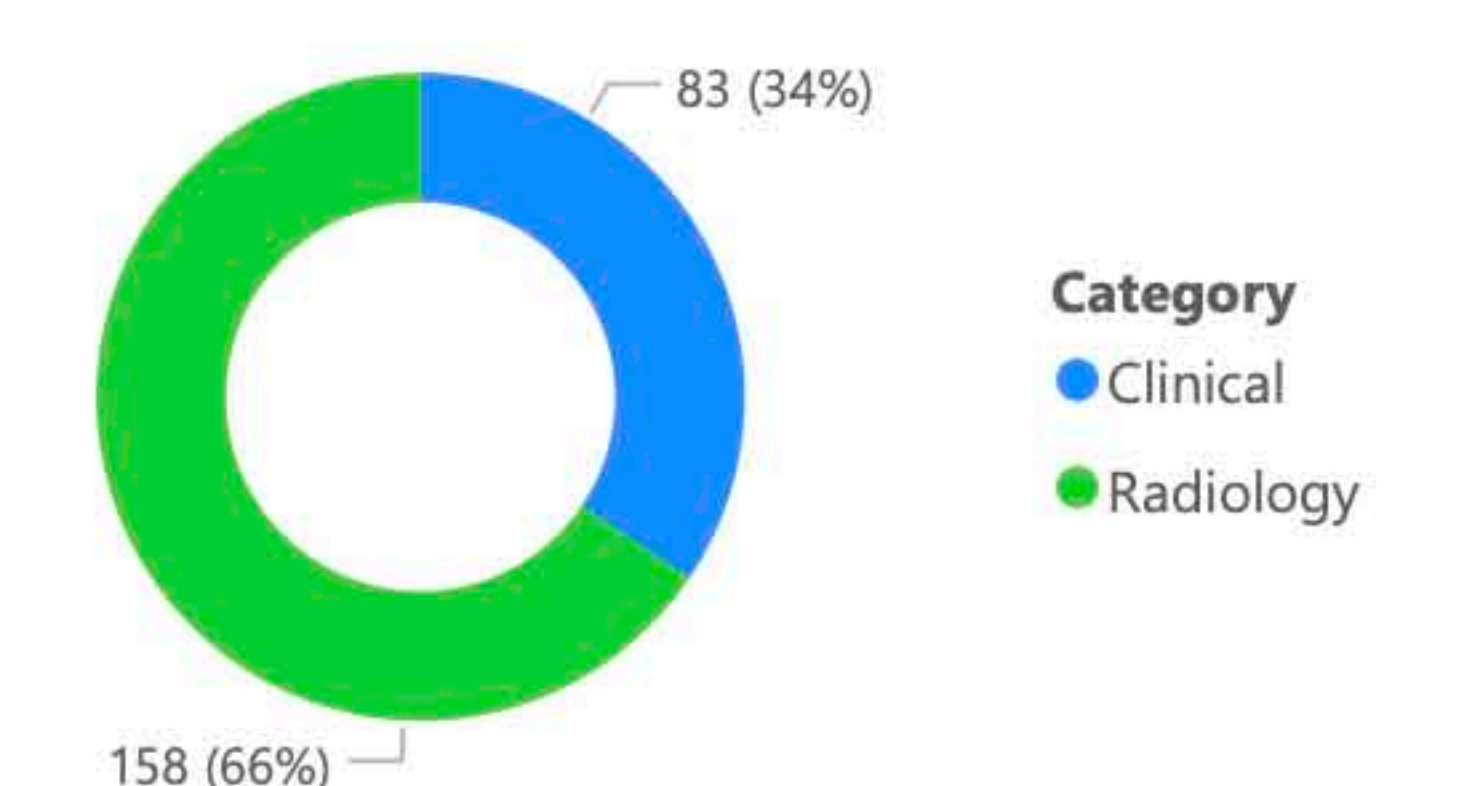
Total Patients - Prostate Cancer Detection **2,143**

Variables category



Total Patients - Myocardial Fibrosis **2,250**

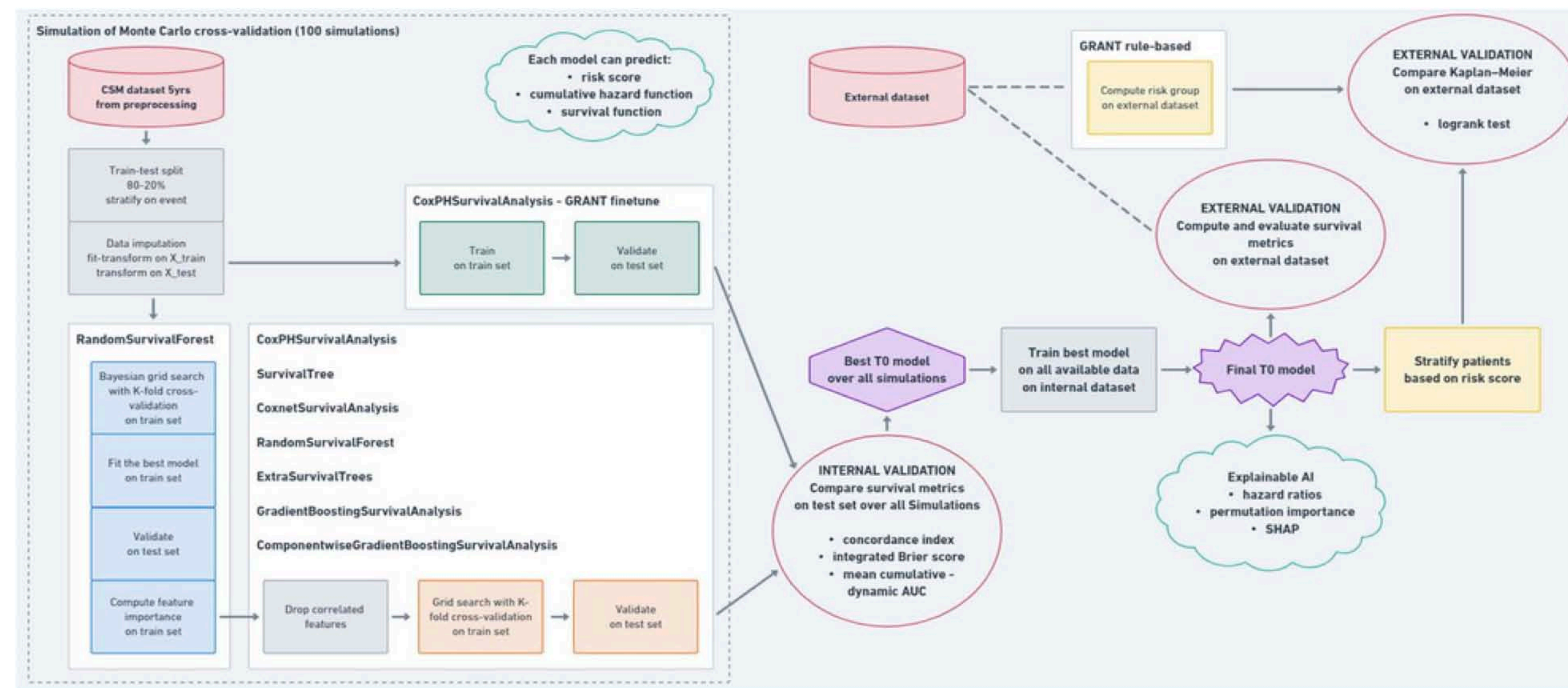
Variables category



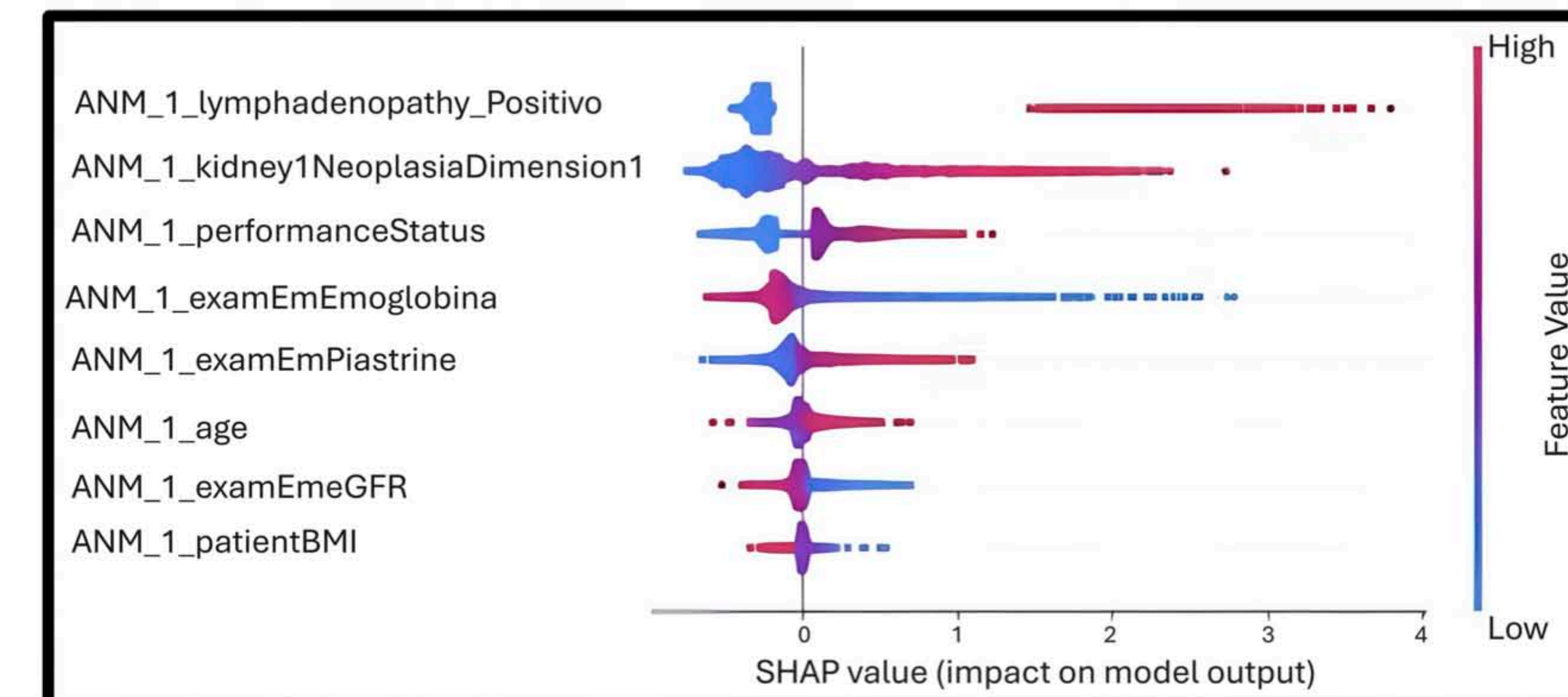
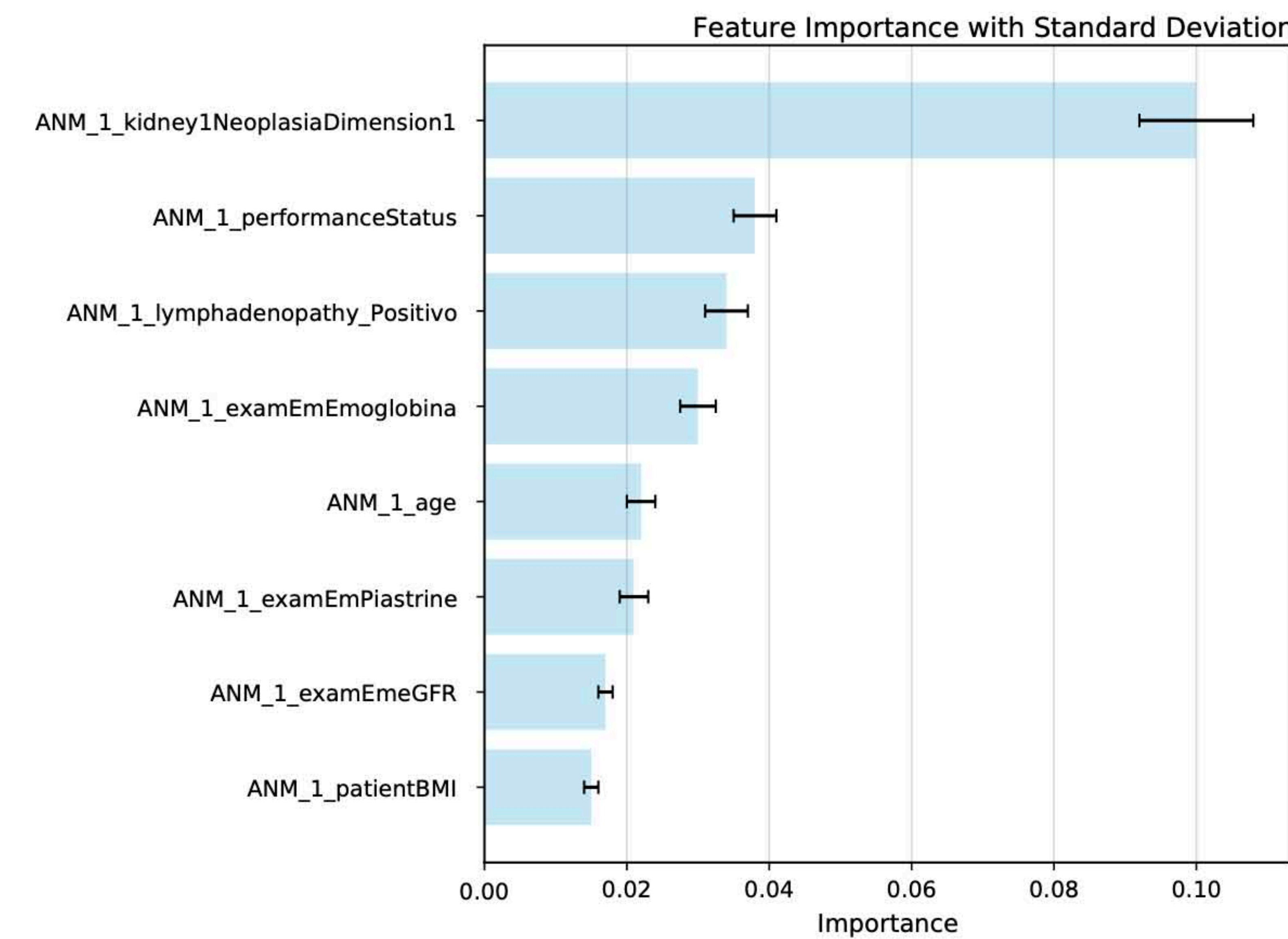
B

C

1. Kidney Cancer Specific Survival Model – ML Design



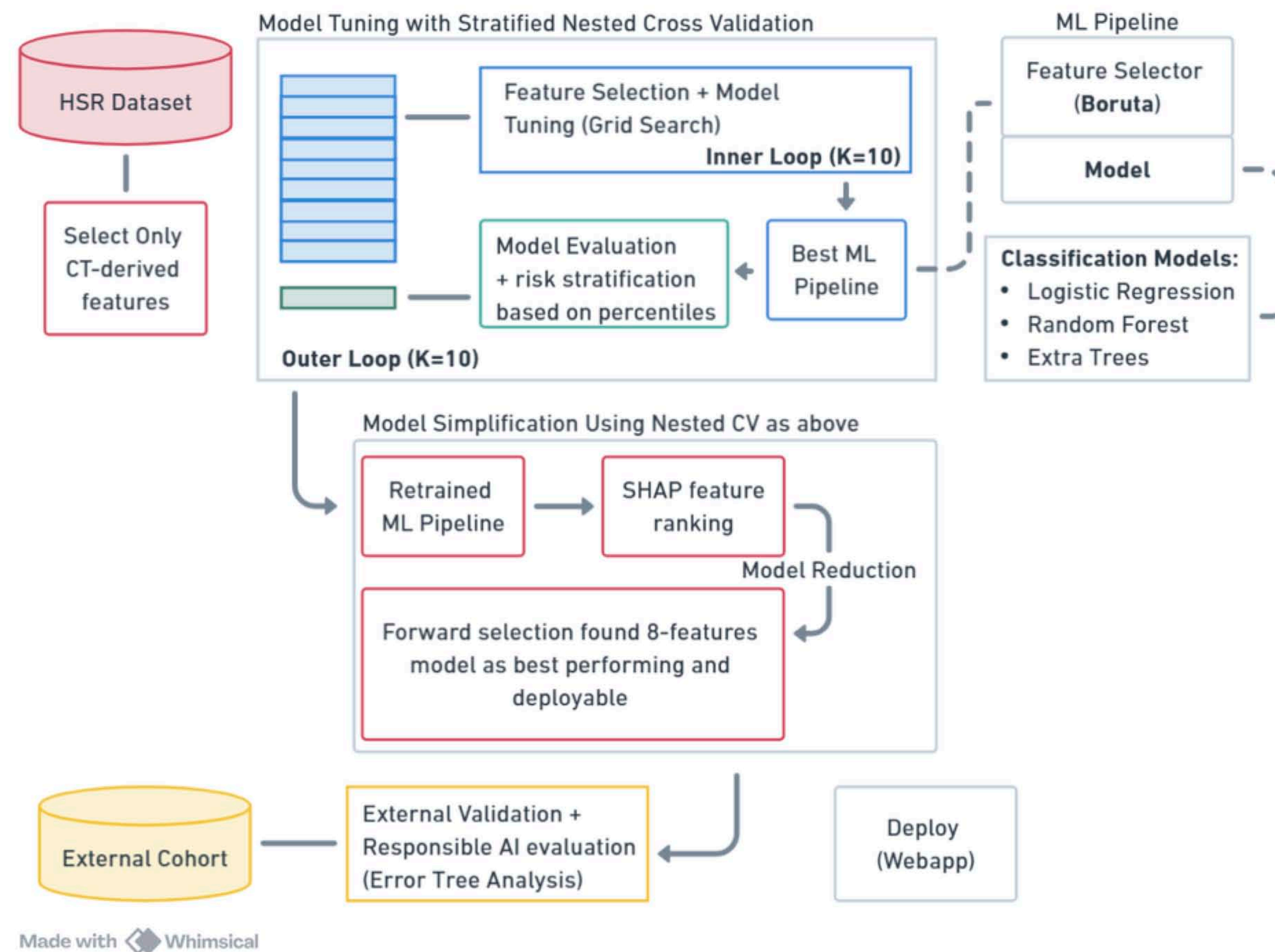
2. Kidney Cancer Specific Survival Model – ML Design



3. Kidney Cancer Specific Survival Model – ML Design



1. Aortic Stenosis TAVI Model – ML Methodology



2. Aortic Stenosis TAVI Model – Error Analysis

