Challenge Requirements

This document contains the various requirements derived from the challenge document.

Terminology

- **COPD:** Chronic Obstructive Pulmonary Disease.
- **CT Scan:** Computed Tomography Scan.
- **CWE:** Common Weakness Enumeration
- **EHR:** Electronic Health Record.
- **ETL:** Extract, Transform, and Load
- **HIPAA:** Health Insurance Portability and Accountability Act of 1996
- **Laboratory Information:** Data that is generated by a lab for the purpose of patient care.
- **LIS:** Laboratory information Systems, also known as a laboratory information management system (LIMS), is an electronic system that stores Laboratory Information.
- **MRI:** Magnetic resonance imaging.
- **OWASP:** Open Worldwide Application Security Project.
- **PET Scan:** Positron Emission Tomography Scan.

Stakeholders

- Patient
- Client
- Company

Business Objectives

The following objectives serve as business internal user requirements. In a longer term project these business objectives could be written as a or into a project charter.

- BR.01: The project shall create a data architectural presentation.
- BR.02: The project presentation shall contain at least 1 slide with an architecture description.
- BR.03: The project presentation shall present the assumptions and considerations used to drive the design.
- BR.04: The project presentation shall present the benefits and drawbacks of the design.
- BR.05: The project shall take into account data integration.
- BR.06: The project shall take into account data quality.
- BR.07: The project shall take into account scalability.
- BR.08: The project shall take into account performance.
- BR.09: The project shall take into account regulatory compliance.
- BR.10: The project shall take into account user experience.
- BR.11: The project shall identify the data sources to be supported.
- BR.12: The project shall propose a data integration strategy.
- BR.13: The project shall specify a standard format to be used in the data integration strategy.
- BR.14: The project shall define how the data from external data sources will be processed via ETL.

- BR.15: The project data integration strategy shall take into account CT Scan imaging studies, tools, and techniques.
- BR.16: The project data integration strategy shall take into account MRI imaging studies, tools, and techniques.
- BR.17: The project data integration strategy shall take into account PET scan imaging studies, tools, and techniques.
- BR.18: The project shall design a data model that can store and organize patient data.
- BR.19: The project shall design a data model that can store and organize patient data.
- BR.20: The project shall state at a high level what machine learning algorithms might be used to analyze patient data.
- BR.21: The project shall provide a system for a dashboard to view patient risk scores.
- BR.22: The project shall provide a system for a dashboard to provide actionable insights to improve patient care.

User Requirements

- UR.01: The user shall be able to load a dashboard to view patient risk scores.
- UR.02: The user shall be able to load a dashboard to provide actionable insights.
- UR.03: The user shall be able to add patients to the system.
- UR.04: The user shall be able to mark patients as removed from the system.
- UR.05: The user shall be able to upload a patient EHR to the system.
- UR.06: The user shall be able to connect the system to a patient EHR system to the system to allow for the automatic uploading of EHRs.
- UR.07: The user shall be able to upload laboratory information to the system.
- UR.08: The user shall be able to connect the system to a LIS to allow for automatic uploading of Laboratory Information.
- UR.09: The user shall be able to upload insurance claims data to the system.
- UR.10: The user shall be able to connect the system to an insurance claims system for automatic uploading of claims data.
- UR.11: The user shall be able to upload imaging study data to the system.
- UR.12: The user shall be able to connect the system to an imaging study system for automatic uploading of imaging study data.
- UR.13: The user shall be able to find the system's dashboards to provide a good user experience.

Regulatory Requirements

- RR.01: The system shall be compliant with 21 CFR Part 11 as required.
- RR.02: The system shall be compliant with HIPAA as required.
- RR.03: The system shall be compliant with EudraLex Volume 4 Annex 11 as required.
- RR.04: The system shall be compliant with the GDPR as required.
- RR.05: The system data model shall support patient privacy best practices and regulatory requirements as stated by the FDA.
- RR.06: The system data model shall support patient privacy best practices and regulatory requirements as stated by the EU.

Security Requirements

- SR.01: The system shall be designed to minimize the OWASP Top 10 API Security Risks.
- SR.02: The system shall be designed to minimize the OWASP Top 10 Security Risks.
- SR.03: The system shall be designed to minimize the CWE Top 25 Security Risks.
- SR.04: The company shall review, take into consideration efforts, and implement selected efforts to minimize the security risks from the next top 100 CWEs.
- SR.05: The build and code versioning system will be statically analyzed to ensure that keys are not leaked.
- SR.06: The company shall hire a security audit team on a regular basis to minimize unforeseen security risks.
- SR.07: The system shall use modern encryption algorithms.
- SR.08: The system shall encrypt data at rest.
- SR.09: The system shall encrypt data in transit.
- SR.10: The system shall, where reasonable, encrypt data in memory.
- SR.11: The system shall scan uploaded data for viruses and malware.
- SR.12: The system shall use known valid versions of external libraries and not use wild card releases.
- SR.13: The system's build server shall, where possible and reasonable, download libraries and external tools and store them versus using a package manager.
- SR.14: The system's build server shall, where possible and reasonable, when downloading new libraries and external tools, scan the libraries and tools for viruses and malware.
- SR.15: The system's containers shall be installed with antivirus and anti-malware software.
- SR.16: The system's containers shall, where possible, use a white listed based anti-malware and antivirus approach.
- SR.17: The system's users shall have unique credentials
- SR.18: The system's account password strategy will follow industry standard best practices.
- SR.19: The system will have an access and usage log.
- SR.20: The system's design shall take into account and, preferably select where possible and reasonable, technologies built around security and safety patterns: including operating system choices, programming languages, library selection, protocols, data formats, etc.

Product Requirements

- PR.01: The system shall identify patients at risk of disease progression.
- PR.02: The system shall identify at-risk patients in a timely manner to allow for early intervention, prevent worsening of symptoms, or the development of complications
- PR.03: The system shall identify the stage of progression of the disease progression.
- PR.04: The system shall be able to identify lung cancer.
- PR.05: The system shall be able to identify COPD.
- PR.06: The system data model shall support queries for patient EHRs.
- PR.07: The system data model shall support queries for Laboratory Information.
- PR.08: The system data model shall support queries for customer claims data.
- PR.09: The system data model shall support queries for imaging studies.
- PR.10: The system data model shall support reports for patient EHRs.
- PR.11: The system data model shall support reports for Laboratory Information.
- PR.12: The system data model shall support reports for customer claims data.

- PR.13: The system data model shall support reports for imaging studies.
- PR.14: The system shall ensure its data architecture is ACID (Atomic, Consistent, Isolated & Durable) compliant.
- PR.15: The system shall ensure its data model is backed up on a regular basis.
- PR.16: The system shall ensure its data model can be recovered on a per customer basis.
- PR.17: The system shall run regular checks to ensure its data model's integrity is valid.
- PR.18: The system data model shall be designed to isolate access to relevant data by role.
- PR.19: The system data model shall be designed to support upgrades and modifications while minimizing downtime to the end user.
- PR.20: The system shall support a set of data pipelines that can collect, process, and transform the data from data source.
- PR.21: The system shall ensure that data processed by its pipelines is up to date.
- PR.22: The system shall ensure that, when data is updated, it is reprocessed by its pipelines as needed.
- PR.23: The system shall support the necessary algorithms and hardware to extract relevant features for Lung Cancer from CT Scans.
- PR.24: The system shall support the necessary algorithms and hardware to extract relevant features for COPD from CT Scans.
- PR.25: The system shall support the necessary algorithms and hardware to extract relevant features for Lung Cancer from MRI Scans.
- PR.26: The system shall support the necessary algorithms and hardware to extract relevant features for COPD from MRI Scans.
- PR.27: The system shall support the necessary algorithms and hardware to extract relevant features for Lung Cancer from PET Scans.
- PR.28: The system shall support the necessary algorithms and hardware to extract relevant features for COPD from PET Scans.
- PR.29: The system shall support the necessary algorithms and hardware to extract relevant features for Lung Cancer from EHR data.
- PR.30: The system shall support the necessary algorithms and hardware to extract relevant features for COPD from EHR data.
- PR.31: The system shall support the necessary algorithms and hardware to extract relevant features for Lung Cancer from Laboratory Information data.
- PR.32: The system shall support the necessary algorithms and hardware to extract relevant features for COPD from Laboratory Information data.
- PR.33: The system shall consider EHRs as a data source.
- PR.34: The system shall consider Laboratory Information as a data source.
- PR.35: The system shall consider CT Scans as a data source.
- PR.36: The system shall consider MRI Scans as a data source.
- PR.37: The system shall consider PET Scans as a data source.
- PR.38: The system shall supply the ability to support a reasonably broad range of current machine learning algorithms, technologies, practices, patterns, and techniques beyond the specific algorithms, hardware, or tools needed for Lung Cancer and COPD.
- PR.39: The system shall support the ability to upgrade or migrate to new algorithms and hardware that supports the integration of advances in technologies, practices, patterns, and techniques for machine learning.