**Chapter 1**

**PREAMBLE**

* 1. **Introduction**

Patient readmission is a major challenge faced by hospitals and healthcare systems around the world. It not only has a negative impact on patient outcomes, but it also significantly increases healthcare costs. Predicting which patients are at risk of readmission is a key step in addressing this problem. In recent years, machine learning and artificial intelligence (AI) have shown great promise in this regard, but we are unable to know how it makes certain decisions. This black box nature makes it difficult to understand and trust. In this work, we present a patient readmission prediction system based on explainable AI, which is designed to provide transparent and understandable predictions. This is achieved using interpretable machine learning models and techniques that allow us to understand the factors that contribute to the risk of readmission for individual patients. By providing this level of transparency, our system aims to support healthcare professionals in their decision-making and help them to better understand and address the needs of their patients.

* 1. **Existing System**

There are several existing systems for patient readmission prediction that use machine learning and artificial intelligence (AI) techniques. These systems typically involve the use of large datasets of patient data, which are analysed using machine learning algorithms to identify patterns and trends that may be predictive of readmission risk. One commonly used approach is to build a predictive model using a supervised machine learning algorithm, such as a decision tree or a logistic regression model. These models are trained on a dataset of patient records, with each record containing a set of features and a label indicating whether the patient was readmitted within a certain time period after their initial discharge. The model is then used to predict the readmission risk for new patients based on their feature values. Overall, the effectiveness of these systems varies and depends on several factors, including the quality and complexity of the machine learning algorithms used, the quality and relevance of the data, and the specific context in which the system is being used.

* 1. **Drawback**

One of the key challenges in healthcare is the prevention of patient readmissions, which occur when a patient is discharged from the hospital but then requires readmission within a short period of time due to a worsening of their condition or the emergence of new health issues. To address this problem, healthcare professionals have traditionally relied on their clinical expertise and judgement to identify patients at risk of readmission. However, this approach has its limitations, and there is growing recognition that machine learning and AI can play a valuable role in predicting which patients are at risk of readmission. By analysing large datasets of patient data, machine learning models can identify patterns and trends that may not be immediately apparent to humans and can provide more accurate and reliable predictions. However, the black-box nature of it can be difficult to understand and explain the factors that contribute to their predictions. This lack of transparency can be a barrier to the adoption of machine learning and AI in healthcare, as it can make it difficult for healthcare professionals to trust and rely on these systems.

* 1. **Proposed system**

In this work, we present a patient readmission prediction system based on explainable AI(XAI), which is designed to overcome these limitations. By using interpretable machine learning models and techniques, our system is designed to provide transparent and understandable predictions, allowing healthcare professionals to better understand and address the needs of their patients. Using explainable AI, our system aims to support healthcare professionals in their decision making and help them to better understand and address the needs of their patients.

* 1. **Plan of Implementation**

1. Obtain data from 130 US hospitals [6] in a zipped format.
2. Extract and load the data onto a Python kernel for analysis.
3. Perform data understanding, cleaning, and preprocessing.
4. Appropriately select relevant features, manipulate them, remove outliers, and normalise the data as per the model requirements.
5. Train and test several machine learning models, including Logistic Regression, Boosted Decision Tree, Decision Forest, Support Vector Machine, and Neural Network.
6. Evaluate the models using metrics such as accuracy, recall, precision, F1, and AUC.
7. Introduce and examine using an Explainable AI model.
8. Present the information in a format that is easily understandable for humans and show how a decision was made.
9. Using this information, try improving the model.

**1.6. Problem Statement**

In this project we are going to decrease the opaqueness of this blackbox and use some of

the techniques of explainable deep learning to know what exactly  causes the results. ​

Also we will present explainable classification of a medical data set using suitable deep

learning techniques.

**1.7. Objective of the Project**

* To improve the accuracy and reliability of readmission risk predictions: By using advanced machine learning techniques and interpretable models, the system aims to provide more accurate and reliable predictions of readmission risk, which can help healthcare professionals to identify and intervene with high-risk patients.
* To increase transparency and explainability: The use of explainable AI techniques allows the system to provide transparent and understandable explanations of its predictions, which can help healthcare professionals to better understand the factors that contribute to readmission risk and make more informed decisions.
* To support clinical decision-making: The system aims to provide real-time, actionable insights to healthcare professionals, helping them to identify and address the needs of patients at risk of readmission.

**Chapter 2**

**Literature Survey**

* **Hu, Y. and Sokolova, M., 2020. Explainable Multi-class Classification of Medical Data. arXiv preprint arXiv:2012.13796:** 
  + This paper presents an explainable multi-class classification method for a large medical dataset.
  + The authors discuss various techniques such as knowledge-based feature engineering, dataset balancing, model selection, and parameter tuning.
  + Six algorithms are evaluated in this study: SVM, Naive Bayes, Gradient Boosting, Decision Trees, Random Forest, and Logistic Regression.
    - The results show that Gradient Boosting and Random Forest outperform the other algorithms in terms of three-class classification accuracy.
* **Tjoa, E. and Guan, C., 2020. A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE transactions on neural networks and learning systems, 32(11), pp.4793-4813.**
  + This research investigates the use of explainable artificial intelligence (AI) in the medical field, particularly for the purposes of improving transparency and trust in AI/ML applications.
  + The focus is on data sources such as images, omics data, and text. The goal is to facilitate the adoption of AI/ML in the medical domain.
* **Ashfaq, A., Sant’Anna, A., Lingman, M. and Nowaczyk, S., 2019. Readmission prediction using deep learning on electronic health records. Journal of biomedical informatics, 97, p.103256.:** 
  + This paper presents a survey on the interpretability and explainability of machine learning (ML) algorithms.
  + It categorises different approaches to interpretability, including those that are formalised mathematically, provide visual explanations, or focus on improving task performance with explanations.
  + The survey applies these categories to the medical field.
* **Moradi, M. and Samwald, M., 2021. Post-hoc explanation of black-box classifiers using confident itemsets. Expert Systems with Applications, 165, p.113941:**
  + In this paper, the authors propose CIE, a post-hoc and model-agnostic method for explaining the predictions made by black-box classifiers.
  + CIE discretizes the input space and extracts confident itemsets representing features highly associated with a class label.
  + These itemsets are used to create concise explanations for individual instances and class-wise explanations through optimization of fidelity, interpretability, and coverage objectives.
* **Ashfaq, A., Sant’Anna, A., Lingman, M. and Nowaczyk, S., 2019. Readmission prediction using deep learning on electronic health records. Journal of biomedical informatics, 97, p.103256.**
  + In this study, a deep learning framework was developed to identify patients at risk of readmission at the time of discharge using a cost-sensitive LSTM model and both human and machine derived features.
  + The potential annual cost savings of implementing the model in a hospital setting were also highlighted. The authors plan to improve the model and conduct a feasibility study to implement it in clinical workflow in the future.

**Chapter 3**

**System Requirements Specification**

* 1. **Functional Requirements**
* The AI model should be able to provide clear and understandable explanations for the reasoning behind its decisions so that it is comprehensible by non technical humans.
* The system should be able to determine the model that gives the most accurate and precise result after recalibration.
* The system should be able to present the explanations in a clear and concise manner, using language and visualisations that are easy for a non-expert to understand.
* The system should be able to provide different levels of detail in the explanations, depending on the needs and goals of the user.
* The system should be able to trace the reasoning behind the AI model's decisions, showing how specific input data and rules led to the final outcome.
* The system should be able to identify and highlight the most important factors that contributed to the AI model's decisions.
* The system should be able to evaluate the reliability and trustworthiness of the explanations, and provide warnings if the explanations are based on uncertain or incomplete data.
  1. **Non -Functional Requirements**
* The system should be secure, protecting sensitive data and preventing unauthorised access.
* The system should be reliable, with a high uptime and a low rate of errors or failures.
* The system should be flexible, able to adapt to changing requirements or conditions without requiring major rework.
* The system should be able to handle complex, multi-faceted problems, and provide explanations that take into account the interactions and trade-offs between different factors.
* The system should be able to operate in real-time, providing explanations as decisions are made, or in batch mode, providing explanations for a large number of decisions at once.
* A model that has been designed to consider cost factors and utilises a combination of features generated both by human and machine processes.
* The system should be efficient, with a fast response time and low resource usage.

**3.3 System Configurations**

The recommended system specification would include:

* Processor: A computer processor with at least 4 cores and a frequency of 2.0 GHz or higher.
* Memory: At least 8 GB of RAM to allow for efficient data processing and model training.
* Storage: A solid-state drive with at least 128 GB of storage space to store data and machine learning models.
* Operating System: A 64-bit operating system such as Windows, macOS, or Linux to support machine learning frameworks and libraries.
* Software: A machine learning framework such as TensorFlow, PyTorch, or scikit-learn to build and train machine learning models along with SHAP framework.
* Development Environment: A programming language such as Python or R, and an integrated development environment (IDE) such as PyCharm or RStudio to write and run machine learning code.

**Chapter 4**

**System Design**

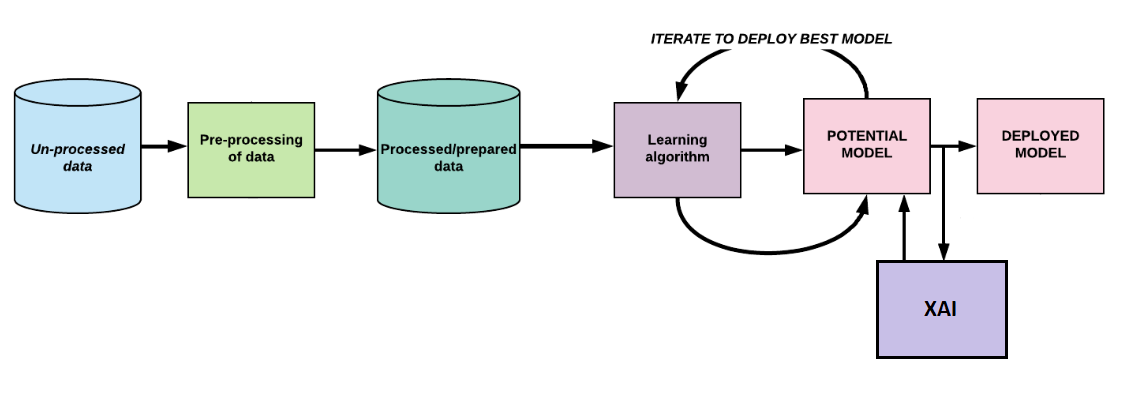
**4.1 System Development Methodology**

Agile software development is a set of methodologies that prioritise collaboration and adaptability in the development process. It is based on iterative development, in which teams work together to continuously improve upon their work. Agile processes generally emphasise disciplined project management with frequent inspection and adaptation, a focus on teamwork and accountability, the use of best practices for rapid delivery of high-quality software, and a business approach that aligns development with customer needs and company goals. The Agile Manifesto, created by leading figures in the software industry, outlines the key principles of Agile development and reflects the approaches that have been successful in the industry. Any development process that aligns with these principles can be considered Agile.

.

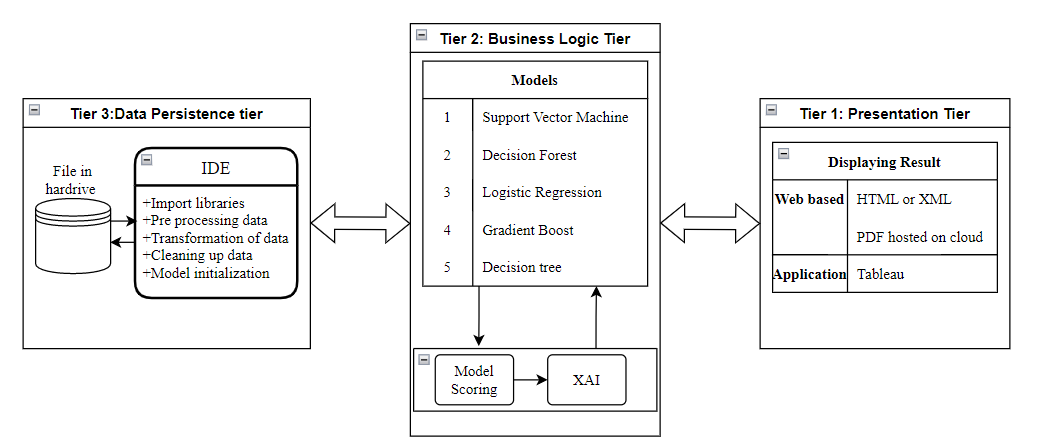
**4.2 Data flow diagram**

A data flow diagram (DFD) illustrates how data is processed by a system in terms of inputs and outputs. As its name indicates its focus is on the flow of information, where data comes from, where it goes and how it gets stored.

****

**4.3 System Architecture**

The architecture of a system outlines the interconnectedness of its components and the flow of data between them. It reflects the structure, functions, and relationships of the system and how it interacts with other systems and the external environment. This architecture can be thought of as the blueprint for the system, and it may evolve as the system is utilised. In the field of architecture, the term "system" typically refers to the architecture of software rather than the physical layout of structures or machinery.



**Bibliography**

1. Hu, Y. and Sokolova, M., 2020. Explainable Multi-class Classification of Medical Data. arXiv preprint arXiv:2012.13796.
2. Ashfaq, A., Sant’Anna, A., Lingman, M. and Nowaczyk, S., 2019. Readmission prediction using deep learning on electronic health records. *Journal of biomedical informatics*, *97*, p.103256.
3. Tjoa, E. and Guan, C., 2020. A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE transactions on neural networks and learning systems, 32(11), pp.4793-4813.
4. Moradi, M. and Samwald, M., 2021. Post-hoc explanation of black-box classifiers using confident itemsets. Expert Systems with Applications, 165, p.113941.
5. Ashfaq, A., Sant’Anna, A., Lingman, M. and Nowaczyk, S., 2019. Readmission prediction using deep learning on electronic health records. Journal of biomedical informatics, 97, p.103256.

* DataSet:

1. https://archive.ics.uci.edu/ml/datasets/diabetes+130-us+hospitals+for+years+1999-2008