UNIVERSIDADE PORTUCALENSE INFANTE D. HENRIQUE

OPTIMA Health Priority: An Intelligent System to Prioritise Patients in a Context of Healthcare Scarcity

Camila Ariana Rodrigues Guerreiro

Mestrado Ciência dos Dados

Orientador: : Profª Doutora Fátima Leal

Co-orientador: : Profª Doutora Micaela Pinho

September, 2025

OPTIMA Health Priority: An Intelligent System to Prioritise Patients in a Context of Healthcare Scarcity

Camila Ariana Rodrigues Guerreiro

Mestrado Ciência dos Dados

September, 2025

Abstract

The management of healthcare resources, particularly in contexts of scarcity such as crises or pandemics, presents a pressing challenge, further complicated by the absence of explicit and widely accepted prioritization guidelines. In such scenarios, healthcare professionals often carry a significant ethical burden, making critical decisions based on implicit rationing criteria, frequently without the support of structured methodologies or public consensus. This study presents OPTIMA Health Priority, an intelligent, automated, and explainable triage system designed to assist in ethically sustainable patient prioritization. The OPTIMA is grounded in public perceptions, collected through a large scale online questionnaire where participants were asked to make prioritization choices between patients with different clinical and personal characteristics.

One of the key challenges encountered in this context was class imbalance, particularly concerning under represented prioritization profiles. Such imbalance can lead to biased model predictions, disproportionately favouring more frequent cases and neglecting critical but rarer patient profiles. To address this, the synthetic data generation process was carefully designed to ensure class distribution parity across different prioritization levels using Gaussian Copula models within the Synthetic Data Vault framework, ensuring balanced, privacy compliant, and ethically replicated input data. This balancing was crucial for achieving fair and generalizable model performance.The proposed prioritization model, OPTIMA, integrates a multi-criteria decision support algorithm, combining advanced machine learning classifiers. Classification is guided by nine ethically grounded and socially accepted rationing criteria: (*i*) severity of disease (pain); (*ii*) life expectancy gain; (*iii*) age; (*iv*) rule of rescue (imminent risk of death); (*v*) improvement in quality of life; (*vi*) waiting time; (*vii*) social need (*e.g.*, parenthood); (*viii*) negative merit (*e.g.*, risky behaviours); and (*ix*) instrumental value (*e.g.*, being a healthcare worker). To enhance interpretability and trust, model decisions are explained, offering transparency and insight into how each feature influences prioritization outcomes.

The core contributions of this work include: (*i*) the ethical generation of synthetic data for healthcare modelling; (*ii*) the development of a transparent and auditable triage system; (*iii*) the alignment of algorithmic decisions with societal values and ethical principles; and (*iv*) methodological reproducibility and scalability. Experimental results confirm the model's robustness and predictive performance, indicating its potential to support fair, consistent, and socially aligned patient prioritization in real world healthcare systems. To our knowledge, this is the first comprehensive attempt to develop an explainable Artificial Intelligence system for patient triage that explicitly integrates societal values into algorithmic healthcare decisions, promoting both equity and efficiency in resource constrained medical environments.

Keywords: Healthcare scarcity; Healthcare rationing, Patient prioritisation, Synthetic Data Generation, Explainable Artificial Intelligence.

Resumo

A gestão de recursos de saúde, particularmente em contextos de escassez como crises ou pandemias, representa um desafio, agravado pela ausência de diretrizes de priorização explícitas. Nesses cenários, os profissionais de saúde carregam um pesado encargo ético, tomando decisões críticas baseadas em critérios de racionamento implícitos e sem o apoio de metodologias estruturadas ou do consenso público. Este estudo apresenta o *OPTIMA Health Priority*, um sistema de triagem inteligente, automatizado e explicável, concebido para apoiar a priorização de pacientes de forma eticamente sustentável. O *OPTIMA* baseia-se nas percepções da população, recolhidas através de um questionário, no qual os participantes foram convidados a tomar decisões de priorização entre pacientes com diferentes características clínicas e pessoais.

Um dos principais desafios enfrentados neste contexto foi o desbalanceamento das classes, particularmente no que diz respeito a perfis de priorização sub-representados. Este desbalanceamento pode conduzir a previsões enviesadas por parte do modelo, favorecendo desproporcionalmente os casos mais frequentes e descurando perfis de pacientes críticos, mas mais raros. Para mitigar este problema, é proposto um processo de geração de dados sintéticos de forma a garantir a paridade na distribuição das classes entre os diferentes níveis de priorização, recorrendo a modelos de Cópula Gaussiana no âmbito da *framework Synthetic Data Vault*. Este equilíbrio foi fundamental para alcançar um desempenho justo e generalizável por parte dos modelos.O modelo de priorização proposto, OPTIMA, integra um algoritmo de apoio à decisão multi-classe, combinando classificadores avançados de *Machine Learning*.A classificação é orientada por nove critérios de racionamento, social e eticamente reconhecidos: (*i*) gravidade da doença (dor); (*ii*) ganhos na expectativa de vida; (*iii*) idade; (*iv*) risco iminente de morte; (*v*) melhoria da qualidade de vida; (*vi*) tempo de espera; (*vii*) necessidade social (*e.g.*, parentalidade); (*viii*) mérito negativo (*e.g.*, comportamentos de risco); e (*ix*) valor instrumental (*e.g.*, ser profissional de saúde). Para aumentar a interpretabilidade e a confiança, as decisões do modelo são explicadas, oferecendo transparência e perceção sobre como cada *feature* influencia os resultados.

As contribuições centrais deste trabalho incluem: (*i*) a geração de dados sintéticos na saúde; (*ii*) o desenvolvimento de um sistema de triagem transparente e auditável; (*iii*) o alinhamento das decisões com valores sociais e princípios éticos; e (*iv*) reprodutibilidade e escalabilidade metodológica. Os resultados experimentais confirmam a robustez e o desempenho do modelo, indicando o seu potencial para apoiar uma priorização de pacientes justa, consistente e socialmente alinhada. Até onde sabemos, esta é a primeira tentativa abrangente de desenvolver um sistema de Inteligência Artificial explicável para triagem de pacientes que integra explicitamente valores sociais nas decisões algorítmicas em saúde, promovendo equidade e eficiência em ambientes médicos com recursos limitados.

Palavras-chave:: Escassez de recursos na saúde, Racionamento de cuidados de saúde, Triagem Priorização de Doentes, Geração de Dados Sintéticos , Inteligência Artificial Explicável.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my supervisors, teacher Fátima Leal and teacher Micaela Pinto, my main and co-supervisor respectively, for their guidance, encouragement, and unwavering support throughout the entire research journey.

I would also like to express my gratitude to my friends and family, whose constant encouragement and presence have inspired me to pursue this master's degree and have been there for me whenever I needed their support. Your trust and encouragement gave me the strength to continue studying while working.

A heartfelt thanks goes as well to my colleagues I had the pleasure of meeting at UPT, whose help and collaboration enriched my academic journey. I am equally grateful to all the professors who accompanied me during this master's degree, sharing their knowledge and providing invaluable assistance.

To all who, directly or indirectly, contributed to the realisation of this dissertation, my sincere thanks.

"The greatest danger in times of turbulence is not the turbulence; it is to act with yesterday's logic."

Peter Drucker

Contents

List of Figures

List of Tables

Abbreviations

AI Artificial Intelligence. 2, 6, 10, 13–16

ANN Artificial Neural Networks. 11, 12, 21

BN Bayesian-Network. 7, 9, 18

CNN Convolutional Neural Networks. 11, 12

CSV Comma-separated values. 19, 33

CTGAN Conditional Tabular Generative Adversarial. 8

DEC1 DECISION1. xiii, 27, 30–32, 34, 35, 37, 39, 40, 44, 45, 52, 53

DEC10 DECISION10. 27, 44–46, 48, 52, 53, 56, 60

DEC11 DECISION11. 27, 44, 45, 48, 52, 53

DEC12 DECISION12. 27, 44, 45, 48, 52, 53

DEC13 DECISION13. 27, 45, 48, 52, 53, 56

DEC14 DECISION14. 27, 44–46, 48, 52, 53, 56

DEC15 DECISION15. 27, 44, 45, 48, 52, 53

DEC16 DECISION16. 27, 44, 45, 48, 52, 53, 56, 60

DEC17 DECISION17. 27, 45, 52, 53

DEC18 DECISION18. 27, 44, 52, 53

DEC19 DECISION19. 27, 45, 48, 52, 53

DEC2 DECISION2. 27, 30, 32, 34, 35, 37, 40, 44–46, 52, 53

DEC20 DECISION20. 27, 44, 45, 48, 52, 53

DEC3 DECISION3. 27, 30, 32, 34, 37, 40, 45, 46, 48, 52, 53, 56

DEC4 DECISION4. 27, 30, 32, 34, 37, 40, 48, 52, 53

DEC5 DECISION5. xiii, 27, 30–32, 34, 37, 39, 40, 52, 53

DEC6 DECISION6. 27, 45, 48, 52, 53, 56

DEC7 DECISION7. 27, 45, 46, 52, 53

DEC8 DECISION8. 27, 44, 45, 48, 52, 53, 56

DEC9 DECISION9. 27, 44–46, 48, 52, 53, 56, 60

ECDF Empirical Cumulative Distribution Functions. 9

fPCA Functional Principal Component Analysis. 8

GAN Generative Adversarial Networks. 8, 9, 14

GBM Gradient Boosting Machines. 11, 12

GMM Gaussian Mixture Models. 8

GUI Graphical User Interface. x, 25, 28, 51, 53, 55, 58

LIME Local Interpretable Model-agnostic Explanations. 13, 14, 22

ML Machine Learning. ix, xiii, 2, 4, 5, 7, 10, 12–25, 27–33, 36, 41, 45, 47, 48, 51, 59

MLP Multi Layer Perceptron. 11, 12, 15, 21, 22, 28, 41–48, 51, 60

RF Random Forest. 11–13, 15, 21, 22, 28, 41–48, 51, 60

SDV Synthetic Data Vault. ix, 2, 8, 9, 14–16, 18, 19, 24, 30, 35, 59, 61

SHAP SHapley Additive exPlanations. 2, 4, 13, 15, 16, 22, 23, 25, 28, 49, 54, 56–58, 60–62

SNS Portuguese National Health Service. 59

VAE Variational Autoencoders. 8, 9

XAI Explainable Artificial Intelligence. 10, 16

XGBoost eXtreme Gradient Boosting. xiii, 2, 11, 12, 15, 21, 22, 28, 41–48, 50–53, 57, 58, 60, 61

Chapter 1

Introduction

In recent years, healthcare systems around the world have faced increasing pressure due to rising demand, limited resources, and unforeseen crises such as the COVID19 pandemic [15, 22, 69]. These challenges have highlighted the urgent need for intelligent decision support systems that can assist healthcare professionals in making fair, transparent, and data driven choices, particularly in scenarios where not all patients can be treated immediately [25, 69, 28]. The allocation of scarce medical resources, such as beds or ventilators, often involves complex ethical and clinical judgments that can benefit significantly from computational support. In this context, the development of prioritization frameworks that incorporate medical data, predictive models, and ethical guidelines is both timely and necessary [15, 22, 69]. In this context, this chapter describes the context of the problem, the motivation, the objectives and the main research questions. Finally, it depicts the structure of the document.

1.1 Context

Managing healthcare resources, particularly in scarcity scenarios such as emergencies or pandemic crises, poses a critical and pressing challenge [20, 22]. The health systems operates under structural limitations that were dramatically exposed during the COVID19 pandemic, when demand for intensive care beds rapidly exceeded supply, leading to system collapse and deaths [15, 22, 69]. Against this backdrop, establishing priorities among patients based on explicit rationing criteria becomes a central, unavoidable task that requires a shift in care objectives from maximizing individual benefit toward maximizing collective welfare and distributive justice [35, 75]. Nowadays, these decisions are taken implicitly by physicians that support a heavy ethical and emotional burden often without adequate training or support for such complex decisions [4, 35].

The literature proposes multiple criteria to guide patients prioritisation under scarcity, including clinical need, treatment benefit, waiting time, post treatment quality of life, patients social role, and age [28, 32]. Practical implementation, however, faces hurdles such as decision making subjectivity, potentially leading to inconsistency and inequity, and time pressure, which increases the risk of rushed or sub-optimal choices [35, 15]. Moreover, a lack of clear guidelines and the

fragility of scientific evidence for assertive resource allocation have been repeatedly highlighted [25, 69].

From a technical standpoint, building fair and reliable prioritisation systems requires access to rich and balanced datasets something rarely available in real world scenarios. In this work, the underlying dataset posed some challenges, like scarcity and imbalance. With only 3085 valid patient records, the data was insufficient to train high performing classifiers, particularly given the severe imbalance across both binary and multi-class priority labels. Some priority levels were significantly under represented, risking biased predictions and undermining system fairness. To overcome this limitation, synthetic data generation techniques were employed, using *Gaussian Copula* models to produce statistically faithful and privacy compliant synthetic records that restore class balance while preserving the joint distribution of features [37, 71, 26, 70, 48, 18].

1.2 Motivation

Although real data are commonly used in healthcare decision support systems, they are often scarce, sensitive, or institutionally restricted especially in ethically complex contexts like patient triage. To overcome these limitations, the approach relies on public perceptions of prioritization criteria and generates synthetic, balanced data that reflect socially accepted values [26, 32, 37]. Although many studies have investigated society's perception of the rationing criteria they support, this data is difficult to compare because it uses different methodologies and is often insufficient or severely imbalanced [22, 35, 36]. It represents a risk of biased algorithms that fail to identify the most vulnerable patients, with serious ethical, legal and clinical consequences [42, 61, 69].

This study leverages synthetic data generation. Using a *Gaussian Copula* approach within the Synthetic Data Vault (SDV) framework, we simulate realistic, balanced and ethically safe datasets while preserving statistical distributions and complex relationships among variables [37, 71, 26, 70, 48, 18]. This strategy not only mitigates bias and scarcity but also ensures diversity and representativeness, prerequisites for fair, robust and generalisable decision systems [74, 32]. At the same time, Machine Learning (ML) based triage raises ethical concerns. Many algorithms behave as *black-box*, obscuring their reasoning and undermining trust and transparency in clinical settings [42, 35]. Explainable, auditable classifications are therefore essential. We adopt eXtreme Gradient Boosting (XGBoost) combined with SHapley Additive exPlanations (SHAP), providing visual, evidence based explanations for each decision [13, 3, 53].

The main motivation comes from two structural challenges: (*i*) real data scarcity concerning public preferences for patient prioritization criteria and bias; and (*ii*) the need to make critical clinical decisions with ethical, transparent, and auditable models. Synthetic data with statistical guarantees, paired with explainable classifiers, constitute a key innovation of this work as an essential step toward more equitable and accountable patients prioritisation system aligned with society's values. Against this backdrop, the present study proposes a comprehensive, explainable Artificial Intelligence (AI) solution to support clinical decisions under scarcity supported by society. The automated patient prioritising system relies on data from a sample of 3085 Portuguese

citizens collected through an online questionnaire designed to gauge social perceptions of priority criteria in extreme scarcity situations. Incorporating public opinion helps align decision models with shared ethical values, fostering broader acceptance and legitimacy.

1.3 Objectives

The overarching goal is to develop an intelligent system to prioritise patients in a context of healthcare scarcity. The system should be explainable, automated and ethically sustainable while maintaining interpretability and accuracy. Additionally, the system must support both binary (priority / non-priority) and multi-class (multiple priority levels) classifications. Specifically, this work contributes with:

- Advanced Synthetic Data Generation: A *Gaussian Copula* technique to produce clinically plausible, ethically secure data with class balancing.
- Explainable Ethical Classification: High performance models with integrated interpretability.
- Interactive Triage Application: A real time system offering clear visual explanations to clinicians and end users.
- Ethical Sustainability: Direct mapping of rationing criteria to explainable variables, reinforcing fairness and accountability.
- Reproducibility and Transparency: A complete pipeline from data generation to explainable inference with open, replicable code and documentation.

These contributions enable to create the OPTIMA Health Priority system which will explore multiple research questions:

1. Can a robust and interpretable classification system be developed to prioritise patient triage? Witch aligns with the goal of developing *Explainable Ethical Classification*, ensuring robust and fair triage decisions.
2. How closely do the synthetic datasets generated reflect the distributions and statistical properties of the real data? Is directly addressed by the *Advanced Synthetic Data Generation* contribution, which focuses on data plausibility and class balance.
3. Are the proposed classification models explainable, and can feature importance be effectively leveraged to prioritise patients? Related with both objectives, the *Explainable Ethical Classification* and the *Interactive Triage Application*, which operationalise transparency and feature level interpretability.
4. Is it feasible to develop a multi-class patient screening system that enhances triage outcomes with transparency and reliability? Relates to the combination of *Interactive Triage Application* and *Ethical Sustainability*, enabling usable, transparent and multi-level triage support.

Reproducibility and Transparency objective, underpins all research questions, by ensuring that the full pipeline from synthetic data generation to explainable model deployment is replicable, auditable, and trustworthy. This reinforces both the scientific rigor and practical applicability of the system.

Each of these questions corresponds to a fundamental challenge in healthcare decision making, particularly under conditions of data scarcity, ethical complexity, and the demand for fairness and transparency. Throughout the dissertation, these questions are examined through quantitative evaluation, model performance analysis, and interpretability.

1.4 Dissertation Structure

This dissertation is structured into five chapters, each building upon the previous to offer a coherent narrative from problem identification to the proposed solution and its evaluation.

- **Chapter 1 – Introduction:** Introduces the problem context, motivation and relevance of patient prioritisation in scenarios of healthcare scarcity. It also outlines the main objectives, research questions, and the overall scope of the work.
- **Chapter 2 – Literature Review:** Provides a comprehensive overview of the theoretical and technical foundations that inform this work. It surveys the state of the art in patient triage systems, explainable ML, synthetic data generation, and ethical frameworks for resource allocation. This chapter also highlights existing gaps and motivates the contributions of the present work.
- **Chapter 3 – Methodology:** Describes the methodological approach adopted for system development. It includes a detailed account of the dataset used and its limitations, the synthetic data generation process using *Gaussian Copula* models, and the machine learning techniques applied. Emphasis is placed on how the design choices aim to ensure ethical alignment, explainability, and technical robustness.
- **Chapter 4 – Experiments and Results:** Presents the experimental setup, evaluation metrics, and key findings. It assesses the performance of the classification models in both binary and multi-class settings, compares real and synthetic data distributions, and analyses feature level explanations using tools such as SHAP. The results are discussed in light of the research questions and objectives.
- **Chapter 5 – Conclusions and Future Work:** Summarises the main contributions and findings of the dissertation. It analyses the implications of the work for clinical practice and ethical decision making in healthcare triage. Limitations are acknowledged, as well as the response to research questions and avenues for future works are suggested.

Overall, the dissertation provides a step-by-step account of how a socially grounded, interpretable, and technically sound prioritisation system was designed, implemented, and evaluated.

Chapter 2

Literature Review

This chapter presents a comprehensive review of the literature, focusing on three main domains: (*i*) explainable methods in ML; (*ii*) patient classification and prioritization; and (*iii*) efficiency and ethical criteria for establishing patient's priorities. The critical analysis of the literature aims to establish the theoretical and empirical foundations supporting the development of an automated, explainable, and ethically sustainable patient prioritization system, as proposed in this dissertation.

2.1 Current Challenges in the National Health System – Resource Scarcity

Establishing a patient priority classification is a critical step in managing health resources, particularly in emergency or resource-limited settings such as overcrowded hospitals or during pandemic crises [15, 75, 58]. Even in normal times, all health systems around the world operate under structurally and chronically scarce resources, with material and human limitations that have long been neglected by policymakers [69, 58]. This lack of resilience severely compromises the ability to respond to crises, bringing about increased challenges in healthcare care provision [15, 19].

During the COVID19 pandemic, this structural deficit worsened significantly [15, 17]. The demand for intensive care and medical equipment quickly outpaced supply, leading to system collapse and avoidable deaths [69, 15, 17]. Resource allocation, thus, became a critical and often reactive process, lacking universal criteria or consolidated scientific evidence [69]. The absence of clear guidelines, combined with extreme pressure, created a context fraught with moral, social, political, and economic dilemmas, requiring urgent short term strategies and long term structural solutions [34, 21]. In this scenario, prioritizing patients becomes a central and unavoidable task. According to health economics, whenever healthcare resources are insufficient to meet all needs, the care goals should be to attain efficiency, i.e. society-centered where collective benefit is priorirtised over individual benefits. Therefore, health economics follow the utilitarianism ethics, a consequentialist moral framework that defines moral decisions as the ones that produce the greatest good health for the greatest number of people. Key to utilitarisianism is the maximization of health benefits [34, 75]. In this context, the criterion of efficiency determines that health care

should be directed to the patient who benefits most from it, measured in life expectancy and/or quality of life [34, 35, 6]. Furthermore, a growing body of literature indicates that societies do not exactly agree with this efficiency criterion, favouring other criteria such as some personal characteristics of patients, namely parenting, lifestyles and their guilt towards the disease and instrumental value [34, 4, 6]. Following this rule, imposes a heavy ethical and emotional burden on healthcare professionals, who are often not trained or adequately supported to make such complex decisions [4, 35, 21].

2.1.1 Ethical Clinical Criteria in Triage

Several criteria have been discussed in the literature to guide patient prioritization in the context of healthcare scarcity [21, 58]. The severity of illness is frequently applied, prioritizing those in the most critical condition. However, this may lead to inefficient use of resources on patients with low chances of recovery [35, 75]. Other criteria focus on the likelihood of benefit from treatment, emphasizing patients with better prognoses or greater life expectancy post intervention [34, 67]. This approach aims to maximize health outcomes and is common in international guidelines.[67] Waiting time, or first come first served criterion, is also considered in some systems. While intuitive, it is often criticized for disregarding clinical urgency and unfairly favouring those with faster access to care [65, 69, 67]. Post-treatment quality of life is another controversial aspect. Though relevant, it raises concerns about potential discrimination against people with disabilities or chronic illnesses [15, 69, 11]. The presence of dependents (*e.g.*, young children) or the patient's social role, specially in essential professions such as physicians or firefighters, are also factors considered in some contexts to break ties between patients with similar clinical prospects [35, 69]. Age is one of the most debated criteria. Some advocate prioritising younger individuals due to greater future benefit or the principle of life cycle fairness [35, 65], while others reject it as inherently discriminatory [4, 11].

Finally, criteria based on personal merit or risk behaviour lifestyles, *e.g.*, mainly caring unhealthy lifestyles, are generally rejected by contemporary bioethics. Nonetheless, discussions persist around personal responsibility in health in specific areas, such as eligibility for transplants [75, 6].

2.1.2 Practical Limitations of these Criteria

Despite the existence of sound ethical and clinical foundations, the practical application of these criteria faces significant challenges. Subjectivity in decision making can lead to inconsistencies between healthcare professionals, undermining equity in care delivery [74, 11, 58]. Additionally, it may go against social values, increasing people's distrust [15, 11, 67].

In response, policymakers must make considerable efforts to improve their acceptability by increasing transparency in decision making processes. One of the most recent ways to achieve this transparency is through the use of AI [61]. The introduction of AI tools brings new concerns.

Many ML models operate as *black-box*, making the decision making processes opaque and reducing transparency, trust, and accountability in clinical contexts [61]. Further issues include dataset imbalance, where critical cases are under represented, resulting in biased algorithms that may misclassified high risk patients [22]. Misclassifications in patient prioritisation carries serious ethical, legal, and financial implications for healthcare institutions [28].

Therefore, beyond strong ethical and scientific foundations, building effective triage systems requires strategic planning, transparency, staff training, careful integration between technology , and social value judgments. The goal is not only to maximize the lives saved but also to preserve justice, trust, and ethical legitimacy in decisions made under extreme scarcity of healthcare resources [15, 53]. Intrinsically interpretable algorithms have been applied to enhance transparency, improve auditability, and increase clinical trust in automated recommendations [5].

2.2 Synthetic Data Generation

In healthcare, datasets are often imbalanced and eliciting societal data concerning their support for the criteria that should guide patient prioritisation are often imbalanced and biased. Therefore, the synthetic data generation offers a promising alternative: it preserves the statistical patterns of real data without exposing individual records [32]. In this context, synthetic datasets faithfully mimic joint distributions and structural dependencies, enabling simulation, behavioural modelling and algorithm development when real samples are scarce or legally restricted [25, 2, 37]. Synthetic data are used to address: (*i*) data scarcity; and (*ii*) privacy and regulation.

Modern ML and deep learning algorithms require large amounts of data to achieve robust performance. In medicine, rarity of certain conditions and logistical or ethical hurdles in data collection limit sample size. Synthetic generation enlarges training corpora while retaining key statistical properties, boosting predictive accuracy even in small-sample scenarios [2]. In turn, health data are highly sensitive and regulated [58]. Even with anonymisation, re-identification risks remain. High fidelity synthetic data, decoupled from real patients, facilitate research and innovation within ethical and legal boundaries [2, 32].

Access to high quality and diverse healthcare data is often limited due to privacy concerns, regulatory constraints, and the rarity of certain clinical scenarios. To address these challenges, synthetic data generation has emerged as a valuable technique for augmenting datasets, enabling model training, testing, and validation without compromising patient confidentiality. Early work relied on classical statistical techniques; recent years have seen a shift toward ML based generative models capable of learning complex variable dependencies and producing novel combinations ("sampled zeros") [25, 32, 70]. These techniques aim to produce realistic and representative data that preserve the statistical properties of the original dataset, thereby supporting robust and generalizable ML models for patient prioritization and other healthcare tasks.

Bayesian-Network (BN) encode conditional relationships through directed acyclic graphs. Sampling from the learned joint distribution allows the generation of synthetic records. Despite

their interpretability, they become computationally infeasible as dimensionality increases, and they struggle to model non-linear dependencies, compromising both data quality and privacy [70, 37, 26]. Limitations include: (*i*) exponential growth in learning complexity with the number of variables [48]; (*ii*) poor representation of highly non-linear relationships [26]; and (*iii*) tendency to memorise real data, putting privacy at risk [26].

Neural Generative Models such as Variational Autoencoders (VAE) and Generative Adversarial Networks (GAN) have gained prominence. VAE map data to a latent space through variational inference, being effective with continuous variables, including longitudinal and eye-tracking datasets [32, 25, 37]. However, they struggle with categorical attributes and require significant computational resources for large datasets [18, 26]. GAN, on the other hand, combine a generator and a discriminator in an adversarial training dynamic. For tabular data, variants such as Conditional Tabular Generative Adversarial (CTGAN) apply conditional techniques to handle categorical variables and class imbalances, improving synthetic fidelity and predictive performance when combined with real data [18, 2].

Despite their potential, neural models face significant challenges in sensitive domains such as healthcare. They often operate as *black-boxes*, making it difficult to trace and justify the generation process, which harms transparency and regulatory compliance [26]. Additionally, problems such as "mode collapse" where the generator produces low diversity data can result in synthetic datasets that omit rare but clinically significant patterns [26]. These models impose a considerable computational burden, require careful hyperparameter tuning, and are sensitive to training conditions, leading to long training times and increased risk of runtime errors when working with large datasets [48, 18, 26].

Synthia is an open source multidimensional synthetic data generator implemented in Python [46]. It uses copula based models, which allow the statistical properties of the observed data to be captured in terms of both individual behaviour and interdependencies between variables [46]. Synthia is further supported by Functional Principal Component Analysis (fPCA), an extension of principal component analysis where the data consist of functions rather than vectors [46]. Synthia is a powerful tool for generating multidimensional synthetic data while preserving complex relationships. However, its applicability is limited to offline processing and it may have shortcomings in accurately replicating variable types and distribution functions from the original data [16].

SDV emerges as a robust and modular solution based on copula theory, specifically designed for tabular data. Gaussian Copula, it´s a function that couples multivariate distribution functions to their univariate margins by describing the dependency structure through a multivariate normal distribution applied to transformed uniform variables [49, 70, 37]. Univariate modelling is performed using Gaussian Mixture Models (GMM), which represent a probabilistic model composed of multiple Gaussian components, that links univariate marginals

through a multivariate normal distribution after transforming the data using empirical cumulative distribution functions (Empirical Cumulative Distribution Functions (ECDF)) [70, 37, 71, 49]. Sklar's Theorem ensures that any multivariate distribution can be decomposed into its marginal distributions and a copula, enabling the reuse of marginals across domains [32, 37, 49].

SDV framework offers several advantages that make it particularly suitable for generating healthcare data. First, it demonstrates high statistical fidelity and clinical realism by preserving the structural and statistical properties of the original dataset, including relationships between tables and columns [18, 48]. Additionally, SDV ensures the preservation of variability and the inclusion of rare cases and also eliminates the need to choose a specific copula family, providing flexibility to capture complex dependencies [71, 51].

Given the limitations pointed out in traditional synthetic data generation methods, the SDV emerges as a robust and highly adaptable solution. From a privacy perspective, SDV offers enhanced protection since its probabilistic approach prevents memorization of real data points, promoting an implicit form of differential privacy [18, 70]. Furthermore, SDV is recognized for its computational efficiency and scalability, processing large datasets with reduced runtime and greater stability compared to models like GAN and VAE [18, 48]. Lastly, among the models included in the SDV library, CopulaGAN and GaussianCopulaSynthesizer are particularly noteworthy for their capacity to model complex dependencies in tabular data. CopulaGAN combines the adversarial training dynamics of GANs with copula based statistical modelling [70, 37, 51, 78]. Copulas are multivariate functions that allow for the separation of marginal distributions from the dependency structure, enabling the model to better capture non linear relationships between variables an essential feature when working with heterogeneous healthcare data [56, 37, 52]. On the other hand, the GaussianCopulaSynthesizer, simplifies this approach by assuming a Gaussian copula structure. It transforms variables into a standard normal space using probability integral transforms and models their joint distribution using a multivariate Gaussian copula. This model has demonstrated high effectiveness in reproducing realistic synthetic datasets while maintaining the integrity of variable correlations and supporting both continuous and categorical data [48, 51, 52, 78].

Although GAN, VAE, Synthia, and BN have merits in specific domains, SDV represents a more robust, explainable, and efficient alternative for synthetic data generation in sensitive contexts such as healthcare [37, 32]. Within the SDV framework, the GaussianCopulaSynthesizer was selected over CopulaGAN due to its stronger statistical grounding, greater transparency, and lower computational complexity [52, 78]. While CopulaGAN leverages adversarial training to capture complex nonlinear dependencies, it also introduces challenges such as increased training instability and reduced interpretability. In contrast, the GaussianCopulaSynthesizer, by assuming a Gaussian copula structure, provides a simpler yet effective model that preserves variable correlations and supports mixed data types with high fidelity. [52, 48, 18, 78] This balance of rigor, explainability, and scalability makes SDV and specifically the GaussianCopulaSynthesizer the

ideal choice for this dissertation, enabling the generation of realistic, privacy-preserving synthetic data suitable for healthcare predictive modelling [48, 18, 78, 52].

2.3 Explainable Artificial Intelligence in Healthcare

The integration of AI into healthcare decision making processes has shown great promise in enhancing diagnostic accuracy, treatment planning, and resource allocation. Prioritisation can be seen as a ML classification problem grouping the patients with levels of priority. ML, a subfield of AI, aims to build models that learn from experience and make predictions on unseen data [20]. These models detect patterns, classify cases and forecast outcomes across a wide range of medical applications [3]. Recent advances in AI have produced highly accurate diagnostic and prognostic tools, often surpassing human performance in tasks such as medical image interpretation, multimodal physiological monitoring and genomic analysis [36, 3].

The adoption of AI in such critical domains raises significant concerns about transparency, trust, and accountability. Explainable Artificial Intelligence (XAI) emerges as a key area of research aimed at making AI models more understandable and interpretable to human users, particularly clinicians and healthcare administrators. This section explores the role of XAI in healthcare, highlighting its importance in ensuring that AI-driven models are not only accurate but also justifiable and aligned with medical and ethical standards. XAI refers to a set of techniques designed to make the decision making processes of complex ML models transparent and interpretable to human users especially critical in high stakes areas like healthcare [43, 3]. Unlike *black-box* models that offer high accuracy but little insight into why a decision was made, XAI seeks to bridge this gap through explanations that are both understandable and justifiable [43, 53]. A study examining clinician interaction with XAI in a COVID19 prediction model found that explanations reduced automation bias and supported clinician reasoning [77]. Given the high stakes of triage systems, XAI it's foundational ensuring that the triage system is not only accurate but also transparent, interpretable, and aligned with human expectations and ethical standards [3]. Explainability techniques therefore play a pivotal role, clarifying individual predictions and increasing transparency [3].

2.3.1 ML Classification Models

In the context of healthcare scarcity, accurately determining the urgency with which patients should receive treatment is critical [67, 58]. ML classification models offer a data-driven approach to support this task by categorizing patients into predefined prioritization levels based on clinical and contextual variables. These models can learn complex patterns from historical healthcare data, enabling rapid and consistent triage decisions that go beyond subjective human judgment [53, 44]. This section presents the rationale for using classification techniques in patient prioritization, outlines the most relevant algorithms for this task, and discusses the criteria for evaluating their performance in real world healthcare settings.

Random Forest (RF) is an ensemble algorithm introduced by Breiman (2001) that constructs multiple decision trees from bootstrapped data samples and aggregates their votes [10, 9]. By "divide-and-conquer", each tree learns from a slightly different subset, and the collective prediction outperforms any single tree [53]. RF is robust to noise, handles feature interactions automatically and cope well with class imbalance through majority-class subsampling [53]. RF has been widely adopted in healthcare research to: (*i*) predict diabetic disease [55]; (*ii*) assess the risk of pre term birth [53]; (*iii*) identify longitudinal health predictors in a 30 year cohort [42]; and (*iv*) estimate mortality and survival in palliative care settings [25]. Beyond clinical practice, RF underpins genomics driven precision medicine, cheminformatics, ecology and bioinformatics [53, 9].

Artificial Neural Networks (ANN) are computational models inspired by the brain, comprising layers of interconnected artificial neurons that adjust connection weights during learning [44, 7]. Deep architectures have dramatically improved predictive accuracy in medical imaging, risk prediction and genomics [36, 2, 25]. Multi Layer Perceptron (MLP) is the classical feed-forward ANN with multiple densely connected hidden layers and non-linear activation functions. Training uses back propagation to minimise prediction error by gradient descent. Despite its architectural simplicity, the MLP remains highly competitive for structured, tabular clinical data, offering reliable performance with lower computational cost and easier interpretability compared to more complex models [44]. Convolutional Neural Networks (CNN), are a specialized type of ANN designed to capture spatial hierarchies in data through convolutional layers that apply filters across input dimensions. This makes them particularly effective in image based domains, such as radiology and ophthalmology, where spatial relationships are critical [7, 36]. While CNN have demonstrated exceptional success in spatial data domains such as medical imaging particularly when integrated with MLP in hybrid pipelines for tasks such as ocular disease or breast cancer detection [7, 36].

XGBoost is a scalable and optimized implementation of Gradient Boosting Machines (GBM), an ensemble method that builds additive models by sequentially combining weak learners, typically decision trees, to minimize prediction error [13]. Each new tree is trained to correct the residuals of the previous ensemble, allowing the model to capture subtle non-linear feature interactions and perform robustly even with imbalanced or noisy data [13]. Despite its complexity, XGBoost is widely regarded for its speed, regularization capabilities, and predictive accuracy, making it especially valuable in high stakes classification tasks [13, 3, 20]. In cardiology, for instance, XGBoost combined with feature selection techniques such as SelectKBest or Feature Importance has been successfully applied to predict myocardial infarction risk [3, 20].

The selection of RF, ANN, and XGBoost models for patient prioritization is grounded in their demonstrated effectiveness and suitability for healthcare applications characterized by complex, high dimensional, and often imbalanced data [44, 13, 9]. RF's robustness to noise,

ability to handle feature interactions automatically, and proven performance in clinical prediction tasks such as disease risk and mortality estimation make it a reliable baseline for interpretable yet powerful modelling [9, 55, 42, 25]. ANN, particularly deep architectures including CNN and MLP, have shown substantial advances in predictive accuracy across medical imaging, genomics, and risk stratification, supporting their utility for capturing non linear and hierarchical relationships inherent in clinical data [44, 7, 2, 25]. XGBoost, a gradient boosting framework, excels at modelling subtle feature interactions and correcting errors iteratively, which enhances its predictive power and has been successfully applied to high stakes domains such as cardiology [13, 3, 53]. Furthermore, all three models benefit from ongoing developments in explainability and interpretability techniques, addressing ethical and practical concerns vital for clinical adoption [53, 36, 5]. Thus, these models represent a balanced combination of accuracy, robustness, interpretability, and clinical relevance, justifying their selection amid a wide spectrum of machine learning methods.

As shown in Table 2.1, each of the selected algorithms exhibits distinct strengths relevant to clinical decision making contexts. RF provides a solid trade-off between complexity and interpretability, offering high reliability and transparent insight into variable importance, which is particularly valuable in healthcare environments where explainability is essential [9, 55]. MLP, while offering high capability in modelling non-linear and hierarchical patterns, is less interpretable and requires higher computational resources, making it more suitable for complex cases where accuracy outweighs transparency [44, 2, 25]. GBM, though more computationally demanding to fine tune, excel in capturing subtle feature interactions and typically achieve high performance in structured data tasks [13, 3, 53]. Overall, the combination of these models ensures robustness, adaptability, and clinical relevance in supporting automated patient prioritization.

Table 2.1: Comparison of Algorithms

Algorithm	**Complexity**	**Interpretability**	**Modelling Non-Linear Relationships**	**Computational Demand**	**Suitability to Problem**
RF	Moderate	High	Good	Moderate	Excellent
ANN	Low (direct)	Low	High	High	Good
GBM	High	Moderate	Very High	Moderate	Excellent

2.3.2 Binary *vs.* Multi-class Classification

In healthcare rationing context ML applications, classification tasks generally aim to discriminate between outcomes, risk strata or priority levels [53]. Problem formulation dictates whether a binary model (*e.g.*, critical *vs.* non critical) or a multi-class model (*e.g.*, no urgency, low, medium, and high) is most appropriate [2].

Binary classification is common when decisions involve a clear dichotomy, such as disease presence (*e.g.*, sepsis detection [50]) or the need for immediate intervention in emergency triage [14]. These models are simpler and often easier to interpret, but they may oversimplify nuanced

clinical realities [2]. Multi-class classification better represents the clinical spectrum, *e.g.*, graded risk stratification for long-term outcomes [47]. However, multi-class models are more sensitive to class imbalance and harder to evaluate or explain, especially when categories are ordinal [9, 44, 53].

Both paradigms struggle with: (*i*) severe class imbalance, particularly for rare yet critical conditions [42]; (*ii*) interpretability, which declines as the number of classes grows [53]; and (*iii*) label ambiguity, since boundaries between, say, "medium" and "high" urgency can be subjective or context dependent [75, 35]. Explainability frameworks help mitigate these issues by providing consistent feature attributions across classes [53].

2.3.3 Post-hoc Explainability

As ML models become increasingly complex, particularly in healthcare applications where decisions can carry life or death consequences, the need for transparency becomes paramount. Post-hoc explainability refers to methods that aim to interpret and explain the predictions *after* the model is trained [5]. Unlike inherently interpretable models, post-hoc techniques are designed to provide insights into *black-box* models such as deep neural networks or ensemble methods, without altering their internal structure. Widely used techniques include, *e.g.*, SHAP, Local Interpretable Model-agnostic Explanations (LIME), SelectKBest, Feature Importance rankings which help clinicians and decision makers understand why a model classified a patient into a specific prioritization level.

SelectKBest improves interpretability by retaining only the most informative variables according to statistical tests such as χ^2, ANOVA F or mutual information [53, 8, 18]. By reducing noise and dimensionality, it clarifies which attributes drive model performance[53].

Feature Importance scores quantify how often and how effectively each feature splits the data, giving a global sense of influence across the forest [9].

SHAP translates cooperative game theory into ML interpretation, assigning each feature a Shapley value that represents its marginal contribution to an individual prediction [53, 36]. This guarantees local accuracy and consistency, making SHAP a preferred choice for high-stakes contexts. In diabetes risk modelling with RF and in cardiovascular disease prediction, SHAP highlighted biomarker importance and boosted clinician trust [36, 5]. For early Parkinson's detection, SHAP pinpointed key gene-expression signatures, further demonstrating its utility in genomics [3]. Clinicians particularly value SHAP's intuitive visualisations and transparent allocation of credit among features, which facilitates adoption of AI outputs in routine workflows [36].

LIME is a model agnostic technique that approximates a complex model locally around a specific prediction using a simpler, interpretable surrogate model, typically a sparse linear model [64]. LIME perturbs the input data around the instance of interest and observes changes in

the output to estimate feature influence in that local region. This approach makes it possible to explain individual predictions even for opaque classifiers such as gradient boosting machines. In clinical contexts, LIME has been applied to explain cancer diagnoses, sepsis risk estimations, and mortality predictions, increasing clinicians confidence in AI outputs by offering transparent, localized rationales [72, 27].

2.3.4 Related Work

Recent research has addressed patient prioritisation from a variety of complementary angles, ranging from ethical analyses to data driven decision support tools. Goddard *et al.* (2006) [19] laid an early theoretical foundation by framing priority setting as an inherently political–economic process, emphasising that rationing cannot be divorced from societal values and institutional power structures. Building on this perspective, empirical surveys have mapped public and professional preferences for rationing principles in multiple countries. For instance, Kasemsup *et al.* (2008) [38] explored Thai citizens views, highlighting strong support for criteria such as severity, prognosis and ability to pay, whereas Pinho *et al.* (2020) and Pinho *et al.* (2022)[59, 58] documented how Portuguese healthcare professionals often diverge from pure utilitarianism, favouring parenthood and instrumental value in tie break scenarios.

With the advent of ML based decision support, explainability and fairness have become key concerns. Amann *et al.* (2020) [5] surveyed multidisciplinary requirements for trustworthy AI in clinical settings, concluding that transparent post-hoc explanations are indispensable for clinician adoption. Rajkomar *et al.* (2019) [61] echoed these concerns, warning that *black-box* models may erode accountability unless paired with rigorous validation and interpretability mechanisms.

Scarcity of real, balanced data remains a critical bottleneck. Hernández *et al.* (2022) [26] provided a systematic review of synthetic data generators for tabular health records, identifying copula based approaches as particularly promising for privacy preserving data augmentation. Complementary work by Jutras Dubé *et al.* (2024) [32] introduced transferable Gaussian copula models for population synthesis, while Kamthe & Assefa [37] (2023) proposed *Copula Flows* that blend normalising flows with copula theory to capture complex dependencies in mixed type data. Such advances underpin domain specific generators like CopulaGAN within the SDV framework, recently benchmarked by Venugopal *et al.* [18] (2022) for tabular health applications. Empirical evidence shows that augmenting scarce datasets with statistically faithful synthetic records can reduce class imbalance and improve downstream model performance without compromising patient privacy [25, 2].

In parallel, intelligent systems have been developed for clinical prediction tasks such as diabetic disease forecasting [55] or risk estimation of pre term birth [53] that leverage ML and, in some cases, incorporate visual explanation tools to foster clinician trust. However, these models primarily aim to predict individual health outcomes based on clinical data, without addressing broader normative concerns around fairness, triage or priority setting. A few studies have explicitly combined ethical criteria and ML. Nasimov *et al.* (2024) [48] demonstrated that GAN generated tabular data can be conditioned on fairness constraints, yielding classifiers that respect

differential privacy guarantees while maintaining accuracy. Sun *et al.* (2019) [70] applied vine copula generators to produce realistic synthetic cohorts for imbalanced clinical tasks, illustrating the benefit of copula driven resampling for minority classes. Nevertheless, none of these works jointly model societal priorities and provide real time, explainable recommendations to clinicians.

2.4 Contributions

This work presents the development of OPTIMA Health Priority, an intelligent decision support system designed to assist in patient prioritisation under conditions of healthcare scarcity. Rather than relying solely on clinical correlates or expert crafted rules, it explicitly encodes public preferences over ethically rationing criteria, ensuring that algorithmic recommendations align with societal values and are transparently justifiable. OPTIMA Health Priority shifts the focus from outcome prediction to value sensitive prioritisation. This work lies not only in the creation of a functional system, but also in the integration of several key innovations. The main contributions of the OPTIMA are summarised in Table 2.2, encompassing advancements in explainability, ML predictions, synthetic data generation, user interaction, ethical values and the formalisation of rationing criteria that helps support clinical decisions. Each of these aspects reflects a deliberate design choice aimed at improving transparency, fairness, and usability in clinical decision making. The following table describes each contribution followed by detailed text explaining how the OPTIMA system stands out.

Table 2.2: Comparative overview of related works and OPTIMA Health Priority contributions.

Work	ML Prediction	Explainability	Ethical Alignment	Interactive Interface	Reproducibility & Transparency	Synthetic Data	Societal Value Alignment	Clinical Decision Support
Rajkomar *et al.* (2019) [61]	✓	✓	–	–	✓	–	–	✓
Amann *et al.* (2020) [5]	–	✓	✓	–	✓	–	–	–
Sun *et al.* (2019) [70]	✓	–	–	–	–	✓	–	✓
Nasimov *et al.* (2024) [48]	✓	✓	✓	–	–	✓	–	–
Jutras Dubé *et al.* (2024) [32]	–	–	–	–	✓	✓	–	–
Venugopal *et al.* (2022) [18]	–	–	–	–	✓	✓	–	–
Panda *et al.* (2023) [55]	✓	–	–	–	–	–	–	✓
Okay *et al.* (2023) [53]	✓	–	–	–	–	–	–	✓
OPTIMA	✓	✓	✓	✓	✓	✓	✓	✓

OPTIMA Health Priority advances the state of the art in tree big pillars. First, it integrates a large scale public survey on rationing preferences, ensures that algorithmic outputs reflect socially endorsed ethical trade-offs, rather than opaque or arbitrary rules. Previous AI triage studies relied mainly on expert opinion or ad hoc rule sets. Second, it employs a Gaussian copula generator within the SDV ecosystem to produce balanced, privacy compliant synthetic data that faithfully reproduces the joint distribution of survey responses. Third, it combines high performance classifiers (XGBoost, RF, MLP) with explainability techniques such as Feature Importance and SHAP. Thus, it ensures that model predictions are not only accurate but also interpretable, providing clear, actionable insights. Furthermore, we provide an interactive interface where each prioritisation decision is transparent, auditable and mapped directly to the nine rationing criteria identified in the literature. In addition to these core innovations, OPTIMA Health Priority distinguishes itself by directly supporting real time clinical decision making in high pressure, resource constrained environments. Unlike prior approaches that focus solely on outcome prediction or retrospective

analysis, OPTIMA is designed to assist clinicians in active patient triage, providing immediate, justifiable recommendations grounded in ethically salient criteria. This capacity enhances not only efficiency but also the perceived legitimacy of prioritisation decisions among healthcare professionals and patients. Finally, the modular architecture of OPTIMA enables flexible adaptation to different clinical domains, making it a general purpose, value sensitive triage support system. By uniting ethical reasoning, rigorous data science, and practical deployment tools, OPTIMA sets a new benchmark for fairness aware, explainable AI in clinical decision support.

2.5 Summary

In summary, this literature review has established the conceptual, ethical, and technical foundations necessary for developing a decision support system that is clinically relevant, ethically defensible, and technically robust based on social values and perceptions. The persistent issue of healthcare resource scarcity underscores the need for transparent and fair prioritisation mechanisms. While theoretical frameworks have long debated ethical criteria such as prognosis, severity, parenthood, quality of life or social merit their practical implementation remains challenging, particularly under conditions of uncertainty, data imbalance, and limited real data.[75, 6]

Synthetic data generation techniques, especially those based on statistically on SDV, models like the Gaussian copula, have proven instrumental in addressing limitations of small or sensitive datasets, ensuring privacy compliance while enabling representative learning [18, 70]. Simultaneously, the emergence of powerful ML models ranging from gradient boosted trees to neural network has opened new possibilities for automating prioritisation tasks. However, these gains in accuracy often come at the cost of interpretability, raising concerns about trust, accountability, and clinician adoption [44, 13, 9]. To bridge this gap, the field of XAI has introduced tools such as SHAP that help interpret model predictions and support human in the loop decision making [53, 36, 43].

These limitations motivate the design of OPTIMA Health Priority, a novel framework with interactive interface, that not only integrates public values into algorithmic triage, but also advances the state of the art across explainability, transparency, decision support and user interaction. This chapter details the current challenges in healthcare, the design, methodological foundations, related studies and the innovative contributions.

Chapter 3

Proposed Method

The increasing need for effective patient triage systems in healthcare, particularly those that prioritize patients based on critical clinical factors, motivates the development of advanced data driven approaches that ensure both accuracy and interpretability [45, 57]. This research proposes a comprehensive framework that combines synthetic data generation with robust classification modelling to support a priority based patient triage system.

This chapter presents the proposed method underlying the development of the OPTIMA Health Priority system. The method integrates synthetic data generation, ML classification, and post-hoc explainability within a unified, ethically guided framework. Each component is designed to address specific challenges related to prioritising patients in contexts of healthcare scarcity, *e.g.*, data limitations, transparency, fairness, and decision traceability. This chapter details the data preparation process, the classification architecture, the explainability techniques applied, and the implementation of the interactive triage application.

3.1 Synthetic Data Generation

Synthetic data generation is a fundamental step in our analysis process, enhancing the dataset to support robust and reliable model development. This step involves creating synthetic data that replicates the statistical characteristics of the real dataset, addressing issues such as class imbalance or data scarcity [57, 41, 45].

The objective of data generation is to expand the dataset in a controlled and representative manner, ensuring that the models trained on it can generalize effectively to real world scenarios. This process is particularly valuable in cases where obtaining additional real data is costly, time consuming, or impractical [54, 29, 39, 62]. By generating high quality synthetic data, we can not only improve the performance and stability of ML models, providing a richer and more balanced foundation for predictive modelling and decision making but also mitigate potential biases and enhance the reliability of subsequent analyses [41]. To ensure the model can learn not only from the distribution of the data but also from the intricate interactions between different factors, it is essential to maintain the correlations between these variables in the generated synthetic data

[33, 73]. Data analysis in the field of patient triage presents unique challenges, particularly due to the limited size of available datasets. In this project, we began with an initial dataset of 3085 responses to a structured questionnaire designed to determine the priority of patient care in medical triage contexts. However, to ensure the development of robust predictive models and reduce the risk of overfitting, it was necessary to expand this dataset through synthetic instances. This process enables better generalization and greater reliability in the predictions made by ML models. Several key factors justify the generation of synthetic data:

Preservation of Complex Relationships. The variables in the dataset, *e.g.*, age, marital status, educational level, perceived health, and beliefs about triage criteria, display critical interrelationships. An effective predictive model must capture these relationships to make accurate and reliable predictions about patient care priority. Synthetic data generation ensures that these statistical dependencies are preserved and adequately represented in the expanded dataset [57, 33].

Mitigation of Overfitting. Models trained on small datasets often display overfitted patterns, compromising their ability to generalize to new cases. Expanding the dataset introduces greater diversity, promoting statistical robustness and reducing the probability of overfitting [41, 45].

Support for Model Robustness. A larger synthetic dataset allows models to handle greater variability in response patterns, enabling better adaptation to real-world scenarios. This is particularly important in a context where critical decisions, such as the order of patient care in medical triage, directly depend on the reliable predictive capacity of the models [54, 62, 57, 39].

These three characteristics preserving relationships, mitigating overfitting, and enhancing robustness were fundamental considerations in both the design of the synthetic data and the selection of data generation methods. By ensuring that the synthetic data accurately represents real world variability while maintaining statistical consistency, we improved the overall effectiveness and reliability of the machine learning models used for patient triage [73, 45, 29].

3.1.1 SDV

For synthetic data generation, we explored: (*i*) Neural Generative Models; (*ii*) Synthia; (*iii*) SDV; (*iv*) BN. Each method was evaluated based on its ability to preserve the statistical properties of the original dataset, handle mixed data types, and generate high quality synthetic data suitable for ML models [37, 32]. The ability to accurately replicate the relationships between the variables, was a key factor in selecting the synthetic data generation method for this project. SDV was ultimately selected due to its superior ability to capture and preserve complex statistical relationships and dependencies between variables [18, 48]. Using advanced probabilistic modelling and ML techniques, SDV generates synthetic data that closely mirrors the original dataset's properties [48, 18, 78, 52].

The proposed data generation approach offers several advantages crucial for healthcare triage modelling. It automatically detects variable types, whether categorical, continuous, or mixed and learns the relationships between them without manual intervention. This enables the preservation of intricate dependencies, such as, correlations between variables [52, 78]. Additionally, the method efficiently scales to generate large volumes of synthetic data while maintaining statistical diversity, ensuring representativeness across patient profiles [45, 57]. Its ability to handle the high dimensionality and complexity of real world healthcare datasets makes it particularly suitable as a reliable foundation for training machine learning models in prioritisation contexts [57, 41].

By leveraging SDV, ensures that the synthetic data remains statistically representative of the original dataset while providing a scalable and adaptable solution for data augmentation in ML applications [45, 41, 33]. The following steps outline the process to generate synthetic data using the SDV framework:

1. **Reading and Preprocessing the Data:** The real data was initially read from a Comma-separated values (CSV) file, ensuring that all necessary preprocessing steps were applied (*e.g.*, handling missing values, encoding categorical variables where applicable).

2. **Metadata Creation:** The single table meta data object from the SDV framework was employed to automatically detect the data types and relationships between columns. This was essential to preserve the original characteristics of the data during the synthesis process [33].

3. **Training the Synthesizer:** Gaussian Copula model was chosen due to its robustness in modelling complex dependencies between variables [52]. This model was trained on the real dataset to learn the distributions and relationships between the features [41].

4. **Generating Synthetic Data:** Using the trained Gaussian Copula model, synthetic data were generated. This ensured that the synthetic dataset closely mirrored the statistical properties of the real data, while maintaining the underlying relationships [48, 33, 78].

5. **Exporting the Synthetic Data:** The generated synthetic data was then saved in a CSV file for further analysis, enabling comparisons with the real data to assess the accuracy of the replication process.

3.1.2 Data Balancing

In ML, data balancing plays a crucial role in ensuring that models can generalize effectively across all classes in a dataset [23]. Class imbalance, where one class significantly outweighs the others in representation, is a common issue, particularly in healthcare datasets where specific outcomes may be rare [66, 63]. Imbalanced datasets can lead to models that are biased toward the majority classes, resulting in poor performance when predicting minority class instances, a critical limitation when dealing with sensitive decisions, such as patient triage prioritization [23, 39].

To address this issue in the context of this project, the data generation process explicitly considered class balancing. While real world data often reflects natural imbalances, incorporating

balance into the synthetic data generation process was vital for creating a robust training set. This ensured the ML models developed could make accurate predictions for both classes, avoiding bias toward the majority class and enabling better generalization [1, 23, 39].

This work was developed based on two complementary classification approaches: a binary model, which distinguishes between low priority (class 0) and high priority cases (class 1), and a multi-class model, which provides a more granular stratification across four levels: Not Prioritary (class 0), Low Priority (class 1), Priority (class 2), and High Priority (class 3). These dual perspectives enable both coarse and fine grained prioritisation, depending on clinical context and resource availability.

The proposed methodology for balancing the dataset included two key steps:

- **Real Data** was divided into subsets based on the decision variables. In the binary approach, the aggregated decision variable was used to define two distinct classes. Cases with a score below 3.00 were labelled low priority, while those with a score equal to or greater than 3.00 were considered high priority. This threshold of 3.00 was selected because it represents the midpoint of the decision scale, ensuring a clear separation between cases where the level of urgency is low and those that already demand closer attention. By doing so, the binary categorization avoids ambiguity in borderline cases and maximizes the contrast between the two groups. In contrast, the multi-class approach provided a more nuanced stratification by establishing four levels of priority. Specifically, cases with a decision score less than or equal to 1.49 were classified as not a priority; scores between 1.50 and 3.49 were labelled as low priority; scores between 3.50 and 4.49 were considered priority cases; and finally, scores equal to or greater than 4.50 were categorised as high priority. This classification schema enabled a more refined understanding of patient prioritisation, supporting both coarse grained and fine grained decision modelling.

- **Balanced Synthetic Data** was done to address the imbalance problem. In this context, we generate: (*i*) 52 525 synthetic instances were generated for class 0; and (*ii*) 50 000 synthetic instances were generated for class 1. After combining the real and synthetic data, both classes were balanced with 52 800 instances per class. In the case of multi-class, 24 964 synthetic data were generated for class 0, 24 253 data for class 1, 23 071 synthetic data for class 2 and 24 637 for class 3. In total there are 100,000 cases of which 96 925 are synthetic data. class 0 included 36 instances (scores less than or equal to 1.49), class 1 had 747 instances (scores ranging from 1.5 to 3.49), class 2 accounted for 1,929 instances (scores ranging from 3.5 to 4.49), and class 3 comprised 363 instances (scores greater than or equal to 4.5). This process ensures that the final dataset was not only representative of the real world distributions but also provided equal representation for both classes, allowing the models to learn patterns effectively minority and majority outcomes.

3.2 Patient Prioritisation through Classification

Patient classification is a core component of ML in healthcare, providing the foundation for predictive models that support decision making processes, such as patient triage or treatment prioritization [23]. In this context, classification models are tasked with assigning outcomes to predefined categories based on patterns in the data. The choice of algorithm and methodology is critical, as it directly influences the accuracy, interpretability, and scalability of the solution [3].

Healthcare datasets often exhibit complexities such as imbalanced class distributions, non linear relationships between variables, and the need for interpretability due to the sensitive nature of the decisions being modelled [66, 63]. These challenges require the use of robust classification algorithms tailored to the specific requirements of the task [20, 39]. Below, we discuss three widely used algorithms in classification tasks, their relevance to healthcare and their unique features.

3.2.1 Methods

Given the imbalance and complexity typical of healthcare data, careful attention was paid to preprocessing, class distribution, and model selection. The chosen classifiers, were selected for their complementary strengths in handling structured, high dimensional data and modelling non-linear decision boundaries [44, 9, 13]. These models were trained and evaluated to identify the most suitable approach for supporting clinical prioritisation with both predictive performance and interpretability [3].

In this study, three ML models were selected to support patient prioritization, RF, XGBoost, and MLP. These algorithms were chosen for their complementary strengths in handling structured, heterogeneous, and often imbalanced healthcare data [44, 9, 13]. RF offers robustness to noise, missing values, and overfitting, while providing transparent feature importance scores, making it a dependable choice for clinical prediction tasks such as disease risk estimation and triage support [9, 55]. XGBoost, a gradient boosting framework, is known for its high predictive accuracy, efficiency, and capacity to model complex, non-linear feature interactions, a critical advantage in decision making contexts, where subtle dependencies between clinical variables influence outcomes [13, 3, 53]. MLP, as a type of ANN, is particularly suited to learning intricate and hierarchical relationships in data. It has shown excellent performance across diverse medical domains, including genomics, imaging, and patient stratification [44, 2, 25]. Together, these models provide a balance between predictive power, interpretability, and generalizability, supporting the development of ethical and transparent decision support systems for patient triage [36, 5].

3.2.2 Selection of Relevant Variables

Identifying and selecting relevant variables is an essential step in developing data driven classification systems, especially in the context of medical applications [12, 24]. In scenarios such as patient triage, not all variables have equal importance in the decision making process [67, 58]. Critical variables, such as priority symptoms or high risk indicators, should be given greater weight, while

less relevant variables can be discarded to increase model efficiency and interpretability [24]. In this project, complementary approaches were implemented to select and weight variables, combining ML methods and statistical techniques. The following steps describe the approach used.

Feature Importance identify the most influential variables, improving model transparency [9]. Assign higher weights to critical features in a weighted scoring system for patient triage. Reduce noise and computational complexity by removing less relevant variables [40]. To achieve this, multiple ML models, including XGBoost, MLP, and RF, were trained using the dataset. The feature importance scores were extracted from each model and normalized to ensure comparability. In the XGBoost model, feature importances were obtained using the model's built-in importance metric, allowing to assign weights that directly reflect the impact of each feature on predictions [13, 3, 53, 40]. Variables with higher importance had greater influence on the score, ensuring that the most critical factors were prioritized in the decision-making process [53, 9, 18, 8]. The final weights derived from these feature importance analysis were integrated into the patient prioritization framework, ensuring that the most critical clinical factors had the strongest influence on the final classification. This approach not only improved predictive accuracy but also made the decision making process more transparent and interpretable [24, 40, 12].

This system enables the automatic classification of patients, prioritizing those in more severe conditions and assisting decision making in high pressure environments such as medical emergencies [12].

3.3 Explainability

In order to provide interpretable and transparent predictions for patient prioritization, the SHAP method was chosen to explain the output of the ML models. SHAP offers a consistent, theoretically grounded framework based on game theory that fairly attributes the contribution of each feature to the final prediction [5]. This allows end users, such as clinicians or healthcare administrators, to understand the reasoning behind the model's decision, increasing trust and acceptance of the automated prioritization system [2, 47].

Unlike other explanation techniques such as LIME, SHAP provides global consistency and local accuracy, ensuring that feature importance is reliably estimated across different predictions and consistent with the underlying model. Additionally, SHAP values naturally sum to the model output, enabling straightforward quantitative interpretation of each feature's impact [3, 53].

In the implementation, SHAP values are computed for each input patient record when a prediction is generated through the web application. These values are normalized and categorized dynamically into qualitative impact levels (*e.g.*, very low to very high impact), making the explanations more intuitive. This is complemented by a visualization showing the relative contribution of the most influential factors, facilitating quick and clear understanding by users. By integrating SHAP explanations directly into the OPTIMA Healthcare user interface, users receive both the

predicted priority class and a detailed, human readable rationale highlighting the key clinical factors driving the decision. The implementation was integrated into a Flask-based web application that allows real time classification and explanation of individual patient cases. Upon receiving user input through a form interface, the model predicts a priority level (ranging from 0 to 3 for the multi-class scenario), and SHAP is used to compute the contribution of each input feature to the predicted class.

The shap.Explainer object was initialized using the trained ML models. For each input case, SHAP values were computed and extracted specifically for the predicted class label, producing a vector of per-feature contributions. These values were then converted to absolute values, normalized into percentages, and ranked to determine the most influential decision factors for that particular prediction [36, 5]. To improve user comprehension, each feature was associated with a human readable explanation (*e.g.*, "Pain level experienced by the patient" or "Waiting time for treatment"). The relative SHAP impact of each feature was categorized into qualitative levels, from very low impact to very high impact, using a dynamic thresholding method based on the distribution of SHAP percentages across all features in the instance. This categorization was printed in the log system and used to generate narrative feedback.

In the frontend, two types of SHAP explanations were provided: (*i*) standardized textual explanations that changed based on the predicted priority level (*e.g.*, Low, Medium-Low, Medium-High, High), highlighting the top contributing factor and its level of impact; and (*ii*) detailed contextual explanation summarizing the top three influential features with their associated impact level and natural language descriptions, along with general guidance about how the triage decision was made. By combining quantitative SHAP values, qualitative impact categories, standardized narrative outputs, and visual explanations, this implementation ensured that each prediction was not only accurate but also explainable.

3.4 Evaluation

The evaluation methodology used to assess the effectiveness of the proposed approach encompasses two key dimensions: (*i*) the quality of the synthetic data; and (*ii*) the performance of the ML classification models. For synthetic data generation, the evaluation focuses on statistical fidelity, structural similarity, and the preservation of relevant distributions when compared to real-world datasets [68, 60]. For classification, the methodology involves measuring predictive performance through standard metrics such as accuracy, precision, recall, and F-measure, while also examining the interpretability and ethical transparency of the prioritisation outputs [33, 68, 31]. Together, these evaluations aim to validate the reliability, realism, and practical applicability of the OPTIMA Health Priority system in supporting healthcare triage decisions [33].

3.4.1 Synthetic Data Generation

Comparing the statistical distributions of real and synthetic data ensures that the synthetic data captures the patterns, variability, and relationships inherent in the original dataset [79].

To evaluate the similarity between real and synthetic datasets, a comprehensive set of descriptive statistical metrics was computed. These included mean, median, standard deviation, quartiles, kurtosis, and skewness [68, 33, 62]. Together, these measures allow to assess the central tendency, dispersion, distribution symmetry, and the presence of outliers, ensuring that the synthetic data faithfully reflects the statistical properties of the original dataset [30, 31]. The evaluation of these metrics is not only a validation step but also an integral part of the synthetic data generation process. By comparing the distributions of real and synthetic data, researchers ensure that the synthetic data is representative and does not introduce biases or unrealistic patterns [30, 33, 62].

These metrics were applied to compare the distributions of real and synthetic data at various stages of the generation process:

1. Initial Validation: The synthetic dataset was first evaluated using these statistical metrics to identify major deviations from real data. Any substantial differences in mean, standard deviation, or skewness signalled a need for parameter adjustments in the data generation model.

2. Iterative Refinement: Based on the initial findings, parameters of the SDV framework were fine tuned, ensuring that synthetic data maintained realistic variability and statistical consistency [79, 56]. Multiple iterations of data generation and validation were conducted to minimize discrepancies.

3. Final Quality Check: Once the synthetic data closely aligned with the real dataset across all metrics, it was subjected to final validation by visualizing distributions (*e.g.*, histograms, box plots) to confirm statistical similarity.

As part of the validation process, the methodology includes identifying any significant discrepancies between the distributions of real and synthetic data. When such differences are detected, the synthetic data generation process is fine tuned to better align with the statistical properties of the original dataset [79, 33, 8]. This iterative refinement ensures that ML models trained on synthetic data not only achieve high performance but also generalize effectively to real world scenarios [79, 33, 56, 8].

3.4.2 Patient Prioritisation through Classification

The evaluation of data classification models is a key step in assessing their performance and reliability, particularly when the models are deployed in critical applications such as healthcare triage systems [68, 30]. This section focuses on the metrics used to evaluate classification performance, interpret the results, and ensure that the model aligns with the objectives of accurate and balanced predictions.

Accuracy measures the proportion of correctly classified instances out of the total instances. It is a straightforward metric that provides an overall view of model performance [68, 33, 31].

Precision analyses the proportion of true positive predictions out of all positive predictions. Precision is crucial when false positives need to be minimised [68, 33, 31].

Recall analyses the proportion of true positives out of all actual positives. Recall is critical when false negatives are costly [68, 33, 31].

F1-Mesure combines precision and recall, providing a balanced measure [68, 31].

This multi metric approach ensures a holistic evaluation of the classification model, enabling improvements and adjustments tailored to the specific requirements of the application, such as prioritizing high risk patients in a healthcare setting [68].

3.5 Graphical User Interface

To make the model accessible and interpretable to non technical users, a web based Graphical User Interface (GUI) was developed using the Flask framework [76]. The application allows users to input their evaluations for each decision criterion ranging from 1 (least agreement/importance) to 5 (strongest agreement/importance) based on their perception of a patient's condition or context. These inputs are processed by a pre trained ML model, which returns a priority classification for the patient, ranging from Low Priority to High Priority, each visually represented by a corresponding colour (green, yellow, orange, red) for intuitive understanding.

Beyond simple prediction, the interface integrates SHAP to provide interpretable explanations tailored to each specific prediction [76]. The system dynamically computes SHAP values for the predicted class, quantifies each feature's relative contribution, and categorizes them from "very low impact" to "very high impact" (in case of multi-class scenario) based on adaptive thresholds derived from the input instance. This information is then used to generate a standard explanation based on the predicted class and most influential feature. In case more detail is needed, a button was designed outlining the decision rationale, key influencing factors, and contextual remarks.

3.6 Dataset

The study relies on a purpose built dataset that captures a wide range of factors relevant to human decision making, personal attributes and moral values forming the empirical basis for our intelligent triage models. Data were obtained through an online survey completed by 3085 Portuguese participants. Recruitment sought maximum heterogeneity: roughly 39 % were university students from multiple disciplines, while the remainder were adults reached via assorted public space approaches. Respondents averaged 36 years of age, the sample was 61.1 % female and 38.1 % male.

The dataset encompasses data from 20 hypothetical rationing decision questions, each contrasting two individuals (Patient A vs. Patient B) who differ in personal or clinical characteristics. Participants, acting as decision makers, expressed their preference on a five point semantic differential scale as depicted in Table 3.1.

The questionnaire design was based on previous work on healthcare rationing preferences and ethical decision making criteria, which employed 23 hypothetical scenarios [38, 59, 58]. In contrast, the present study streamlined the approach by using 20 carefully selected scenarios, each corresponding directly to one feature used in the model. Furthermore, while previous studies focused primarily on healthcare professionals as respondents, this work deliberately targeted lay citizens, aiming to capture public perspectives on medical prioritization.

Table 3.1: Five point preference scale

1	2	3	4	5
Definitely prioritise A	Some preference for A	No preference	Some preference for B	Definitely prioritise B

Following Pinho & Araújo (2022), the six overarching rationing principles were disaggregated into nine specific criteria [58]:

- (i–ii) *Clinical need*: this criterion refers to the medical urgency of the patient, assessed through two components: **Pain intensity** represents the current physical suffering of the patient and signals a direct need for immediate intervention. **Immediate risk of death** known as the **Rule of Rescue**, this sub-criterion prioritizes patients whose lives are in imminent danger, regardless of other circumstances.

- (iii–iv) *Health maximisation*: this criterion seeks to maximize the health benefits resulting from treatment. It includes: **Life expectancy gain** measures the additional years of life that treatment is expected to provide. **Quality of life improvement** assesses how much the treatment can enhance the patient's daily well-being and functioning.

- (v) *Social need*: captured through **Parenthood**, this criterion considers that individuals with dependents (e.g., young children) may carry greater social responsibilities, which could justify prioritization.

- (vi) *Age*: this criterion evaluates the patient's age, often grounded in principles such as prioritarianism (giving preference to those who have lived fewer years) or the assumption of greater potential benefit in younger patients.

- (vii) *Instrumental Value*: refers to the essential or beneficial societal role of the patient (*e.g.*, healthcare worker, firefighter). Prioritizing their recovery may result in broader social benefit.

- (viii) *Merit Negative* (risk taking lifestyle): considers whether the patient's condition resulted from voluntary risky behaviors (e.g., excessive alcohol, smoking, drug use). This criterion raises ethical debates around personal responsibility.

- (ix) *Fair chance* (equity of access): operationalized through **waiting time**, it embodies the idea that everyone deserves a fair opportunity to receive treatment, regardless of clinical or social characteristics.

Table 3.2 lists the twenty decisions and maps to the corresponding criterion. In the raw file, each scenario response is stored in variables DECISION1 (DEC1), DECISION2 (DEC2), DECISION3 (DEC3), DECISION4 (DEC4), DECISION5 (DEC5), DECISION6 (DEC6), DECISION7 (DEC7), DECISION8 (DEC8), DECISION9 (DEC9), DECISION10 (DEC10), DECISION11 (DEC11), DECISION12 (DEC12), DECISION13 (DEC13), DECISION14 (DEC14), DECISION15 (DEC15), DECISION16 (DEC16), DECISION17 (DEC17), DECISION18 (DEC18), DECISION19 (DEC19), DECISION20 (DEC20) facilitating analysis of decision patterns (*e.g.*, intuitive vs. deliberative choices). Additional columns record composite value orientations, conformity, benevolence, hedonism, openness to change, self-enhancement, supporting cluster analysis of participant profiles and group trends.

Table 3.2: Hypothetical Scenarios and Corresponding Rationing Criteria

Decision	**Patient A**	**Patient B**	**Criterion**
1	Has a moderately painful disease	Has a very painful disease	Clinical need: Pain
2	Has been waiting 1 month	Has been waiting 6 months	Fair chance: Waiting time
3	Single, no dependants	Parent of three minors	Social need: Parenthood
4	Alcoholic with severe liver disease	Severe liver disease, never drank	Merit: Negative
5	Average citizen	Front-line physician	Merit: Instrumental value
6	20 % chance to live >5 years	40 % chance to live >5 years	Life expectancy
7	80 years old	40 years old	Age
8	Slight QoL improvement with treatment	Substantial QoL improvement	Quality of life
9	Joined queue today	Waiting 1 month	Fair chance: Waiting time
10	Painful disease	Very painful disease	Clinical need: Pain
11	20 % chance to live >5 years	80 % chance to live >5 years	Life expectancy
12	Will die within 3 days w/o treatment	Will die immediately w/o treatment	Clinical need: Rule of Rescue
13	80 years old	20 years old	Age
14	Will die within 1 month w/o treatment	Will die within 1 week w/o treatment	Clinical need: Rule of Rescue
15	Slight QoL gain (poor → fair)	Large QoL gain (poor → very good)	Quality of life
16	HIV from unsafe sex	HIV from hospital transfusion	Merit: Negative
17	25 years old	10 years old	Age
18	Average citizen	Scientist who discovered cures	Merit: Instrumental value
19	10 years old	Newborn	Age
20	Married, no children	Married, three school-age children	Social need: Parenthood

The structure of the dataset allows for a detailed examination of individual behaviours and value based preferences, supporting the identification of key patterns that shape triage decisions. By analysing this wide range of variables, it becomes possible to gain deeper insight into how personal traits and ethical considerations influence decision making in healthcare scenarios. If we divide the current dataset by different classes there were 36 cases with a value less than or equal to 1.49 for class 0 (not a priority), 747 cases with values between 1.50 and 3.49 for class 1 (low priority cases), 1929 for class 2 (Priority cases) with values between 3.50 and 4.49 and 363 cases with a value greater or equal to 4.50 for class 3 (High priority cases). This enriched dataset is then used to train ML models, with feature selection methods helping to pinpoint the most relevant factors for determining patient priority. These models are designed with interpretability in mind, ensuring that their results are both understandable and practically useful in real world clinical settings. In doing so, the dataset plays a central role in answering the study's research questions and in demonstrating the validity of the proposed approach.

3.7 Summary

This chapter presented the methodological framework developed to tackle patient prioritisation in resource constrained healthcare contexts. It began with the construction of a dataset based on a questionnaire designed to capture ethical preferences across key decision making criteria. These variables reflect diverse moral perspectives and serve as the foundation for the triage model.

To enhance robustness, synthetic data generation techniques were applied, followed by data balancing to address class imbalances [23]. Several ML classifiers were applied to identify the most effective for predicting triage outcomes such as RF, XGBoost and MLP, while feature selection methods helped isolate the most relevant variables, improving both performance and interpretability [12]. SHAP based explainability was integrated to ensure transparency in model decisions [53].

Evaluation metrics guided rigorous model assessment, and a GUI was developed to allow real time interaction with the system. Users can input patient profiles and receive not only triage classifications but also personalised explanations supported by visualisations of feature influence. Together, these components form a coherent and interpretable pipeline that supports ethical, data driven decision making in healthcare triage.

Chapter 4

Experiments and Results

The experiments and results outlined in this chapter aim to validate the proposed methodology and evaluate its ability to address the research questions surrounding intelligent patient triage. Through the use of real and synthetic datasets, advanced feature selection methods, and interpretable ML models, this chapter illustrates how patient classification can be optimized to prioritize critical cases while maintaining a transparent and actionable decision making process. The analysis relies on quantitative metrics, evaluations of data distribution, visualizations, and tools for interpretability, offering a thorough understanding of the system's performance. This approach ensures that the methodology meets the practical demands of healthcare systems, where timely and well informed decisions are critical for improving patient outcomes.

4.1 Synthetic Data Generation Results

The original dataset consisted of 3,085 cases and presented a clear class imbalance, both in the binary and multi-class classification scenarios. In the binary configuration, class 0 included only 275 instances corresponding to responses rated between 0 and 3 while class 1 comprised 2,800 instances, associated with ratings higher that 3. It is important to note that the dataset does not present a consistent total number of valid responses per case, as the overall average score used to classify each instance could not always be calculated. This was due to missing values (*NaN*) in the decision-related questions, which led to the exclusion of incomplete cases from the classification process. In the multi-class setup, the distribution was similarly skewed: class 0 included 36 instances (scores less than or equal to 1.49), class 1 had 747 instances (scores ranging from 1.5 to 3.49), class 2 accounted for 1,929 instances (scores ranging from 3.5 to 4.49), and class 3 comprised 363 instances (scores greater than or equal to 4.5).

To support the development of robust and generalizable ML models, synthetic data was generated to expand the training dataset, mitigate overfitting, and address the imbalance across classes. Several generation strategies were explored, each aiming to refine the synthetic dataset in a way that preserved the statistical properties of the original data while improving class balance.

4.1.1 Binary scenario

Imbalanced Synthetic Data. This approach aims to explore the model's ability to generate a large volume of synthetic data from the available real dataset, even while maintaining the original class imbalance. At this initial stage, no artificial balancing technique was applied, the focus was on evaluating how well the synthetic data generator could replicate the statistical structure of the real data, despite its inherent bias. The entire original dataset, consisting of 3,085 cases, was used for this purpose. It distinguished between two levels of priority only, low priority and high priority. Based on this dataset, the Gaussian Copula model from the SDV library was trained to learn the marginal distributions and dependencies between variables. Subsequently, 100,000 new synthetic instances were generated, preserving the original class proportions. This preliminary experiment was conducted solely for the binary classification scenario, as it served an exploratory role while the overall synthetic data strategy was still under evaluation. Since this unbalanced approach was not ultimately chosen for the final modeling pipeline, it was not extended to the multi-class setting, where more refined balancing strategies were later applied.

Data Distribution Although the dataset contains 103 variables, a full-scale comparison of all features would be neither practical nor necessary for assessing the overall quality of the synthetic data. Therefore, a subset of representative variables was selected based on their relevance to the predictive modeling task and their importance in describing the population. This included demographic attributes such as age, gender, marital status, education level and employment status which reflect the general structure and diversity of the dataset, as well as a few of the key decision related variables (*e.g.*, DEC1, DEC2, DEC3, DEC4, and DEC5), which are directly used in the construction of the ML models. These variables were chosen to ensure that both the descriptive and predictive dimensions of the data were adequately preserved in the synthetic samples. By focusing the distributional analysis on this carefully selected subset, it is possible to evaluate whether the synthetic data maintains the essential statistical properties of the original dataset, while avoiding unnecessary redundancy or overfitting to less informative features.

Table 4.1 compares the real and synthetic data distribution regarding the variables used for binary classification.

Table 4.1: Comparison of Real and Synthetic Data

Statistic	**Age**		**Gender**		**Marital Status**		**Education Level**		**Employment Status**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	36.43	36.36	0.6266	0.6293	1.7295	1.7292	4.773	4.7706	2.2665	2.2664
Median	31	32	1.0	1.0	2.0	2.0	4.0	4.0	2.0	2.0
STD	17.66	17.77	0.4990	0.4996	0.8468	0.8449	1.5920	1.5881	1.4134	1.4104
Q1	20	21	0.0	0.0	1.0	0.0	4.0	4.0	1.0	1.0
Q3	50	49	1.0	1.0	2.0	2.0	6.0	6.0	3.0	3.0
Range	74	62	2.0	2.0	4.0	4.0	8.0	8.0	5.0	5.0
Kurtosis	-0.66	-0.77	-1.4320	-1.3973	2.1714	2.0752	0.1740	0.1720	-0.3204	-0.3060
Skewness	0.70	0.63	-0.3429	-0.3378	1.3742	1.3541	0.00	0.00	0.7859	0.7858

How can you check by the 4.1 the comparative analysis between real and synthetic data across all variables reveals a high level of fidelity in the replication of statistical properties. The synthetic dataset closely mirrors the original in terms of central tendency (means and medians), dispersion (standard deviation, quartiles, and range), and higher order distributional metrics such as skewness and kurtosis. This alignment is consistent across categorical variables like gender, marital status, education level, and employment status, as well as continuous such as age. In practice, this means that the generative process did not introduce distortions in class proportions, value distributions, or tails of the variables, preserving both balance and diversity within the data.

The similarity observed is not limited to averages but extends to the overall shape and structure of the distributions. Quartiles and ranges, are reproduced with highly precision, ensuring that both common and less frequent values are adequately represented. The preservation of skewness and kurtosis is equally important, as it indicates that asymmetries and tail behaviours of the original data were successfully captured. Such replication suggests that the synthetic model effectively learned its underlying statistical patterns.

This robustness across different types of variables is particularly relevant for downstream applications. Models trained on the synthetic data are expected to encounter input distributions that are nearly indistinguishable from those in the real dataset, minimising risks of bias or reduced performance. Moreover, the consistency of replication across sociodemographic dimensions demonstrates that the synthetic data generator can handle both categorical and continuous variables with the same level of accuracy.

Table 4.2 presents a detailed comparison between the real and synthetic data for the five decision related variables (DEC1 to DEC5). These variables are key inputs for building the ML models, as they represent the main decision categories evaluated.

Table 4.2: Comparison of Real and Synthetic Data for Decisions Variables

Statistic	**DEC1**		**DEC2**		**DEC3**		**DEC4**		**DEC5**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	4.1588	4.1461	4.2562	4.2470	3.8629	3.8593	3.7610	3.7597	3.4596	3.4525
Median	4.0	4.0	5.0	5.0	4.0	4.0	4.0	4.0	3.0	3.0
STD	0.9784	0.9896	0.9863	0.9932	0.9413	0.9422	0.9870	0.9875	0.9277	0.9319
Q1	4.0	4.0	4.0	4.0	3.0	3.0	3.0	3.0	3.0	3.0
Q3	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	4.0	4.0
Range	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0
Kurtosis	1.9907	1.8935	2.1314	1.9717	0.3158	0.3461	-0.0858	-0.0864	0.2381	0.2120
Skewness	-1.4175	-1.4013	-1.5285	-14946	-06389	-0.6504	-0.4800	-0.4795	-0.1929	-0.1962

The comparison of the five decision variables (DEC1 to DEC5) between real and synthetic datasets reveals a consistent and robust alignment across all statistical measures. Central tendencies are preserved with high precision, as means and medians remain virtually identical in both datasets. Dispersion indicators, including standard deviation, quartiles, and range, confirm that the variability and spread of decision scores were faithfully reproduced by the generative model.

For DEC1 and DEC2, the presence of strong negative skewness and moderately elevated kurtosis values highlights the natural clustering of responses toward the higher end of the decision scale. The synthetic data successfully replicates this behaviour.

DEC3, DEC4, and DEC5 exhibit more balanced distributions, the synthetic dataset captures these patterns without distortion, maintaining fidelity across different decision constructs.Taken together, these results indicate that the synthetic generator not only replicated average values but also captured the deeper structural characteristics of the decision variables.

The statistical comparison between real and synthetic variables indicates that the synthetic dataset successfully captures the main characteristics of the original data. The generated data accurately reflects the real data in terms of distribution, central tendency (mean, median), dispersion (standard deviation), and shape metrics such as skewness and kurtosis. These results reinforce the validity of the synthetic data for downstream ML tasks, as it preserves the structural integrity and statistical properties essential for accurate modeling.

Data Balancing After synthetic data generation, the dataset contained 2,074 instances for class 0 and 101,001 instances for class 1. This decision was made with an exploratory purpose, to assess whether simply increasing the volume of data without directly modifying class proportions could enhance the generalization capacity of predictive models.

The analysis of the synthetic data confirmed that the generator successfully replicated the statistical relationships between variables, preserving the logical consistency and semantic integrity of the records. However, despite this internal coherence, the class distribution remained highly imbalanced, as reflected in the disproportionate number

of samples between classes. This persistent imbalance illustrates a key limitation of relying solely on data expansion without incorporating balancing strategies, especially in scenarios where model performance is sensitive to class proportions.

Balanced Synthetic Data. In the second approach, the primary objective was to address the significant class imbalance present in the original dataset and assess how a more balanced data distribution could enhance the performance and generalization ability of the ML models. Rather than generating a large volume of synthetic data immediately, a two-step strategy was adopted to carefully control the balance between classes throughout the synthetic data generation process. Initially, a targeted generation of synthetic samples was performed to equalize the number of cases between the two classes. While the original dataset contained 2,800 instances in class 1 (high priority), class 0 (low priority) was severely underrepresented, with only 275 instances. To correct this imbalance, 2,525 synthetic samples were generated exclusively for class 0, bringing both classes to the same total of 2,800 cases. This initial balancing step ensured that the generative model would be trained on a more representative and evenly distributed dataset, helping to reduce bias during training. After the balanced classes, 100,000 synthetic data were generated based on this new CSV. Given that this approach was not ultimately selected for the final modeling pipeline, particularly due to limitations in maintaining class balance throughout the synthesis process, it was not extended to the multi-class scenario. Instead, more targeted and effective balancing strategies were subsequently designed and implemented for the multi-class configuration.

Data Distribution of the same selected variables previously examined in the imbalance synthetic data approach was done comparing their real and synthetic values. The comparison is based on essential again using statistical metrics such as mean, median, standard deviation, quartiles (Q1 and Q3), range, kurtosis, and skewness. This allows to evaluate how well the synthetic data replicates the underlying distributional characteristics of the real dataset for the binary case.

Table 4.3 presents a comparison between the real and synthetic data distributions for each variable in the binary case, based on the balanced synthetic dataset. The table summarizes key statistical metrics to assess how well the synthetic data preserves the characteristics of the original dataset.

Table 4.3: Comparison of Real and Synthetic Data of all variables for the balanced data approach.

Statistic	**Age**		**Gender**		**Marital Status**		**Education Level**		**Employment Status**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	39.2175	41.3171	0.6141	0.6151	1.7319	1.7268	4.7865	47840	2.2501	2.2510
Median	34.0	36.0	1.0	1.0	2.0	2.0	4.0	4.0	2.0	2.0
STD	19.2971	19.2952	0.5070	0.5086	0.8348	0.8274	1.6324	1.6319	1.4159	1.4177
Q1	22.0	24.0	0.0	0.0	1.0	0.0	4.0	4.0	1.0	1.0
Q3	52.0	55.0	1.0	1.0	2.0	2.0	6.0	6.0	3.0	3.0
Range	74.0	73.0	2.0	2.0	4.0	4.0	8.0	8.0	5.0	5.0
Kurtosis	-0.5063	-0.5196	-13904	-1.3570	2.3602	2.3480	0.0761	0.0722	-0.3631	-0.3631
Skewness	0.7488	0.7376	-0.2379	-0.2223	1.3918	1.4014	-0.0516	-0.0562	0.7927	0.7946

Table 4.3 presents the distribution results indicate a very close preservation of the overall statistical structure, although some differences are slightly more noticeable than in the imbalanced synthetic data approach.

For age, the means (39.2 vs. 41.3) and medians (34 vs. 36) suggest a slight shift toward higher values in the synthetic dataset, while standard deviation, quartiles, and range remain nearly identical. In gender and marital status, the metrics are highly consistent, with only marginal variations in skewness and kurtosis. Education level and Employment status shows almost perfect alignment in mean and dispersion, confirming that the synthetic data accurately mirrors the categorical structure.

Overall, the model effectively replicated the distributional properties of the original data, maintaining consistency in both central tendency and variability. The minor deviations observed, such as the upward shift in age, do not compromise the reliability of the synthetic dataset and can be considered acceptable as they preserve natural variability.

Table 4.4 presents the comparison between the real and synthetic distributions of the five decision related variables: DEC1 through DEC5. These variables represent key behavioral or evaluative responses, and preserving their statistical structure is essential for downstream modeling tasks such as classification or prediction.

Table 4.4: Comparison of Real and Synthetic Data for Decision Variables (DEC1–DEC5)

Statistic	**DEC1**		**DEC2**		**DEC3**		**DEC4**		**DEC5**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	3.7178	3.7170	3.6630	3.6665	3.4570	3.4507	3.3451	3.3426	3.1826	3.1862
Median	4.0	4.0	4.0	4.0	4.0	4.0	4.0	4.0	3.0	3.0
STD	1.2453	1.2474	1.3515	1.3514	1.2047	1.2140	1.2988	1.3017	1.1394	1.1347
Q1	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Q3	5.0	5.0	5.0	5.0	4.0	4.0	4.0	4.0	4.0	4.0
Range	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0
Kurtosis	0.7064	0.6809	0.0965	0.1358	0.4623	0.422	-02229	-02240	0.1688	0.1980
Skewness	-1.0838	-1.0781	-09272	-09416	-0.9074	-0.8976	-0.6681	-0.6676	-0.5393	-0.5517

The synthetic data demonstrates a remarkably close alignment with the real data across all five decision variables. Central tendency indicators (mean and median) are nearly identical in each case, and standard deviations show minimal variation, indicating that the spread and variability of the responses were successfully reproduced.

Quartiles and ranges remain consistent across real and synthetic datasets, reflecting the same boundaries and concentration of values. Furthermore, the kurtosis and skewness metrics reveal that the overall distribution shapes, including tails and asymmetry, are closely matched, particularly for DEC1 and DEC2, which exhibit moderately negative skewness. This consistency suggests that the synthetic data generation process effectively maintained the structure of subjective decision making variables, which is crucial for maintaining validity in models reliant on these features.

Overall, the synthetic data generated by the SDV framework shows an excellent match with the real data across all analyzed variables. This strong alignment between the real and synthetic datasets ensures that the model can learn the intricate relationships and dependencies that exist in the original data, making the synthetic dataset a reliable tool for further analysis and training in the context of patient triage.

Data Balancing The main idea in this approach, applied to the binary case, was to first balance the real dataset by ensuring that both classes had an equal number of cases. Then, based on this balanced sample, a large number of synthetic cases would be generated while maintaining class equilibrium. This correction eliminated the initial imbalance, resulting in a balanced dataset of 5,600 instances as the foundation for synthetic data generation. Subsequently, 100,000 synthetic records were generated, preserving the balanced class proportions throughout the output.

However, upon evaluating the final dataset, comprising a total of 105,600 cases, it became evident that the intended balance was not maintained throughout the full generation process. The final distribution revealed a significant resurgence of class imbalance, with class 0 comprising 20,710 instances and class 1 comprising 84,881 closely resembling the original disproportions. This outcome highlights a critical limitation of the generative model, although the input to the synthesizer was balanced, the underlying statistical patterns learned from the original, imbalanced data appear to have influenced the output distribution. While the generated dataset remained logically coherent and statistically plausible, it failed to preserve the balance between classes, a key factor in achieving optimal performance in classification tasks.

Refined Synthetic Data. The final approach focused on precisely addressing the inherent class imbalance by controlling the number of synthetic instances generated per class. Unlike previous strategies, this method did not simply replicate the global distribution of the original dataset or apply balancing as a preliminary step. Instead, it adopted a targeted strategy,

determining the exact number of synthetic cases needed for each class so that, when combined with the original real data, the final dataset would total 100,000 instances with fully balanced class representation.

This refined approach was applied to both the binary and multi-class classification settings, as it proved to be the most effective and robust strategy for producing high quality synthetic data. Its goal was not only to correct imbalances but also to provide a strong, generalizable foundation for training ML models with fair exposure to all class scenarios.

Data Distribution A refined synthetic data generation strategy was implemented that explicitly considered class balance while still being based on the original imbalanced real data, which was applied to the binary and multi-class case. The goal was to ensure that the generated dataset reflected the real world distribution while also providing a sufficient number of cases for each class to train the models effectively. A crucial aspect of validating the synthetic data is comparing its distribution with that of the real data. To ensure that the generated data reflects the statistical characteristics of the real dataset, a thorough comparative analysis was conducted on each variable and several key variables were compared using statistical metrics.

Table 4.5 compares the distribution of all variables between the original and synthetic datasets for the third approach. In this strategy, different amounts of synthetic data were generated per class to counteract the real dataset's imbalance, ensuring that after combining both sources, each class was equally represented.

Table 4.5: Comparison of Real and Synthetic Data of all variables for the refined synthetic data .

Statistic	**Age**		**Gender**		**Marital Status**		**Education Level**		**Employment Status**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	36.4315	38.0496	0.6266	0.6020	1.7295	1.7290	4.7735	4.7772	2.2663	2.2544
Median	31.0	35.0	1.0	1.0	2.0	2.0	4.0	4.0	2.0	2.0
STD	17.6565	19.3095	0.4990	0.5108	0.8468	0.8437	1.5922	1.6312	1.4136	1.4081
Q1	20.0	24.0	0.0	0.0	1.0	1.0	4.0	4.0	1.0	1.0
Q3	50.0	57.0	1.0	1.0	2.0	2.0	6.0	6.0	3.0	3.0
Range	74.0	73.0	2.0	2.0	4.0	4.0	8.0	8.0	5.0	5.0
Kurtosis	-0.6564	-0.6305	-1.4320	-1.3999	2.1714	2.2708	0.1734	0.0781	-0.3203	-0.3561
Skewness	0.7027	0.7129	-0.3429	-0.1759	1.3742	1.3957	-0.0002	-0.0615	0.7863	0.7837

Table 4.5 compares real and synthetic data under the refined synthetic data approach, showing high alignment across most variables. This method successfully maintains the structural integrity of all variables while introducing a controlled level of variability that strengthens the dataset's robustness. For instance, the age distribution in the synthetic data, although slightly shifted towards older individuals, broadens the variability without distorting the overall population profile. This is particularly valuable as it ensures representation of wider patient scenarios, enriching model generalisation.

Similarly, gender, marital status, and employment status show very consistency between real and synthetic data, with only marginal statistical deviations that fall within acceptable tolerance. The education level variable also exhibits nearly perfect alignment, proving that categorical structures are reliably replicated. The minor differences observed they reduce the risk of overfitting to narrow real world patterns, while preserving the realism needed for credible decision support models. Ranges remain consistent, and categorical levels are fully preserved, ensuring no loss of semantic meaning or data integrity.

In summary, the synthetic dataset successfully replicates the key statistical characteristics of the original data while enhancing balance and diversity.

Table 4.6 compares the distributions of the decision related variables (DEC1 to DEC5) between the real dataset and the synthetic data generated under the refined synthetic data. This approach focuses on generating synthetic samples in a way that, when combined with the original data, results in balanced class representation across the target variable while maintaining distributional characteristics of the predictors.

Table 4.6: Comparison of Real and Synthetic Data for Decision Variables (DEC1–DEC5)

Statistic	**DEC1**		**DEC2**		**DEC3**		**DEC4**		**DEC5**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	4.1588	3.8476	4.2562	3.6021	3.8629	3.4027	3.7610	3.2866	3.4596	3.1494
Median	4.0	4.0	5.0	4.0	4.0	4.0	4.0	4.0	3.0	3.0
STD	0.9784	1.2910	0.9863	1.3688	0.9413	1.2377	0.9870	1.3248	0.9277	1.1565
Q1	4.0	3.0	4.0	3.0	3.0	3.0	3.0	3.0	3.0	3.0
Q3	5.0	5.0	5.0	5.0	5.0	4.0	5.0	4.0	4.0	4.0
Range	4.0	5.0	4.0	5.0	4.0	5.0	4.0	5.0	4.0	5.0
Kurtosis	1.9907	0.5657	2.1314	-0.0253	0.3158	0.3106	-0.0858	-0.2702	0.2381	0.1164
Skewness	-1.4175	-1.0524	-1.5285	-0.8699	-0.6389	-0.8836	-0.4800	-0.6627	-0.1929	-0.5386

The statistical comparison across the five decision variables indicates that the synthetic data captures key distributional characteristics present in the real dataset. For all variables, the medians and inter-quartile ranges are preserved, reflecting consistency in the core distribution shape. However, the synthetic data exhibits slightly lower means across all variables, coupled with higher standard deviations. This indicates a broader spread and a modest shift in central tendency, potentially reflecting the increased representation of lower values in the synthetic samples. Despite these variations, the alignment in median and quartile values supports the conclusion that the synthetic data remains representative of the real data's structure. This validates the suitability of the generated samples for downstream tasks, while fulfilling the goal of class balancing imposed by the refined synthetic data.

Data Balancing The method began by analyzing the number of real cases in each class

(275 in class 0 and 2,800 in class 1). Based on this, the appropriate number of synthetic samples was individually generated for each class to ensure that both reached an equal representation in the final dataset. This means generating proportionally more data for the underrepresented class (class 0), and less for the dominant class (class 1), correcting the existing bias without distorting the statistical properties of the original data.

A total of 49,725 synthetic instances were generated for class 0 and 47,200 for class 1. When combined with the original real data, each class reached exactly 50,000 instances, resulting in a perfectly balanced final dataset. This equal class distribution provides a solid foundation for developing classification models, enabling the algorithm to learn how to predict patient priority levels without being affected by the structural bias present in the original dataset. In addition to improving class balance, this controlled generation ensured that each synthetic record adhered to the learned statistical relationships and logical consistency of the original dataset. By doing so, the model avoided common pitfalls associated with naive oversampling, such as redundancy or data artifacts.

4.1.2 Multi-class scenario

Refined Synthetic Data. Following the success of the refined balancing strategy in the binary classification scenario, the same methodology was extended to the multi-class configuration. Given the more complex distribution of classes in this scenario, with varying levels of under-representation, a careful and systematic approach was adopted to ensure an even class distribution while preserving the statistical integrity of the original data. Rather than applying a generic oversampling method or maintaining the natural imbalance, this refined approach aimed to generate the exact number of synthetic instances needed per class. The objective was to ensure that, when combined with the real dataset, each of the four classes would be equally represented, resulting in a final dataset composed of 100,000 balanced instances, 25,000 per class.

Data Distribution This scenario involved four distinct priority levels: non-priority, low priority, priority, and high priority. The synthetic data was carefully generated to ensure equal representation across these categories, with each class totaling 25,000 instances after combining real and synthetic records. Before moving forward with the evaluation of individual variables, it was essential to validate the overall distribution of the synthetic data. A comparative distribution analysis was performed between the real and synthetic data across all variables, highlighting how well the synthetic data replicates the structure of the real dataset in this multi-class setting. This analysis served as a critical step in verifying the fidelity and reliability of the synthetic dataset. Particular attention was given to variables that influence classification performance, ensuring that the synthetic data maintained both statistical consistency and semantic realism.

Table 4.7 presents a comparison of the distribution for the variables between the real dataset and the synthetic data generated under the final multi-class balancing approach. The table provides an overview of key distributional statistics to assess the fidelity of the synthetic data in replicating the structure of the original dataset.

Table 4.7: Comparison of Real and Synthetic Data of all variables for the Refined Synthetic Data.

Statistic	**Age**		**Gender**		**Marital Status**		**Education Level**		**Employment Status**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	36.4315	37.1235	0.6266	0.6100	1.7295	1.7304	4.7735	4.7751	2.2663	2.2875
Median	31.0	32.0	1.0	1.0	2.0	2.0	4.0	4.0	2.0	2.0
STD	17.6565	15.5671	0.4990	0.5041	0.8468	0.8183	1.5922	1.60641	1.4136	1.4331
Q1	20.0	25.0	0.0	0.0	1.0	1.0	4.0	4.0	1.0	1.0
Q3	50.0	46.0	1.0	1.0	2.0	2.0	6.0	6.0	3.0	3.0
Range	74.0	75.0	2.0	2.0	4.0	4.0	8.0	8.0	5.0	5.0
Kurtosis	-0.6564	0.4693	-1.4320	-1.4642	2.1714	2.2775	0.1734	0.14055	-0.3203	-0.4516
Skewness	0.7027	1.0379	-0.3429	-0.2601	1.3742	1.3398	-0.0002	-0.0470	0.7863	0.7593

The refined synthetic dataset under the multi-class scenario shows strong alignment with the real data while introducing small variations that enhance representativeness. The age variable presents a slightly higher mean and skewness, capturing a broader spread of cases and improving balance across groups.

Gender, marital status, and education level remain almost identical to the real data, preserving structural fidelity. For employment status, the synthetic data introduces slightly higher variability, which is beneficial for generalisation. Overall, this refined approach ensures both realism and controlled heterogeneity, making it especially suited for robust multi-class modelling.

Table 4.8 presents a comparative summary of the statistical distribution of five decision related variables (DEC1–DEC5) between the real dataset and the synthetically generated dataset, both derived under the final multi-class balancing strategy. This evaluation is crucial to assess whether the synthetic data can reliably mimic the key distributional characteristics of decision making variables found in the original dataset, which are central to downstream modeling and interpretation.

Table 4.8: Comparison of Real and Synthetic Data for Decision Variables (DEC1–DEC5)

Statistic	**DEC1**		**DEC2**		**DEC3**		**DEC4**		**DEC5**	
	Real	Synth	Real	Synth	Real	Synth	Real	Synth	Real	Synth
Mean	4.1588	3.7276	4.2562	3.7772	3.8629	3.5350	3.7610	3.3914	3.4596	3.1472
Median	4.0	4.0	5.0	4.0	4.0	4.0	4.0	4.0	3.0	3.0
STD	0.9784	1.6854	0.9863	1.6561	0.9413	1.6196	0.9870	1.6444	0.9277	1.4640
Q1	4.0	3.0	4.0	3.0	3.0	2.0	3.0	2.0	3.0	2.0
Q3	5.0	5.0	5.0	5.0	5.0	5.0	5.0	5.0	4.0	4.0
Range	4.0	5.0	4.0	5.0	4.0	5.0	4.0	5.0	4.0	5.0
Kurtosis	1.9907	-0.3296	2.1314	-0.4404	0.3158	-0.5962	-0.0858	-0.8188	0.2381	-0.4666
Skewness	-1.4175	-0.9887	-1.5285	-0.9241	-0.6389	-0.7580	-0.4800	-0.6306	-0.1929	-0.5647

The descriptive statistics demonstrate that the synthetic data maintains a close approximation to the real dataset across all five decision variables. The mean and median values are generally consistent, with DEC1 and DEC2 showing slightly higher central tendency in the real data, while the synthetic data exhibits marginally increased dispersion as reflected in higher standard deviations for all variables.

The general shape and symmetry of the distributions are preserved, the synthetic dataset offers a faithful representation of the real decision variables, supporting its adequacy for subsequent analytical tasks within the multi-class framework.

The statistical comparison confirmed that the synthetic data is a highly reliable representation of the real dataset. The preservation of means, medians, and quartiles of the variables highlights the model's ability to faithfully replicate central tendency and variability, ensuring that the synthetic data reflects the essential statistical structure of the real dataset. While minor differences exist they do not significantly impact the overall validity of the dataset. The high degree of similarity in statistical metrics suggests that the synthetic data can be confidently used for analysis and modeling, maintaining the essential characteristics of the original real world data.

Data Balancing The process began with an analysis of the real data distribution across the four priority levels: class 0 (non-priority), class 1 (low priority), class 2 (priority), and class 3 (high priority). Based on the number of real instances in each class, synthetic data was generated in a targeted manner to complement the deficit in each category. For class 0 has been generated 24,955 data samples, 24,253 synthetic data for class 1, 23,071 for class 2, and 24,637 synthetic data for class 3. When combined with the respective real data, these additions brought each class to exactly 25,000 instances. This resulted in a fully balanced multi-class dataset with a total of 100,000 examples, ensuring fair and equal representation of all priority levels for model training.

In addition to achieving balance, special care was taken to ensure that the synthetic records remained statistically and semantically coherent. The average values of key decision variables within each class stayed within the expected ranges defined for their respective categories. Variables such as age, political orientation, and service usage

patterns remained within realistic and medically plausible boundaries. By applying this refined and controlled synthesis approach, the resulting multi-class dataset provides a robust and fair foundation for machine learning applications, enhancing the model's ability to learn from all classes equally and generalize effectively across different patient profiles.

Given this overall performance, this approach was selected as the basis for developing the algorithm integrated into the patient triage application, as it proved to be the most effective in terms of distributional fidelity and data balancing. The next step will be to apply machine learning algorithms to this dataset to build and validate the predictive components of the system.

4.2 Classification Results

This section presents the classification outcomes obtained throughout the development and evaluation of different synthetic data generation strategies. The primary goal was to ensure that the synthetic datasets not only preserved the statistical characteristics of the original data but also enhanced model learning by addressing critical class imbalances.

Given the healthcare context in which this study is framed specifically, the application of ML to support patient triage achieving both accuracy and fairness across classes is essential. The experiments included multiple data synthesis approaches, progressively refining the balance and representativity of the generated data. Each scenario was evaluated using a range of classification algorithms and performance metrics, with a focus on their ability to generalize across both majority and minority classes. The results outlined in the following subsections provide a comparative analysis of model performance in each scenario, culminating in the identification of the most effective approach for deployment in a real world clinical setting.

4.2.1 Binary Scenario

Performance To assess the quality of synthetic data and its utility in predictive modeling, we applied a suite of classification algorithms including RF, XGBoost, and MLP, across different data generation strategies. Each model was trained and tested on datasets constructed from distinct approaches to balancing and synthesizing the data. The performance analysis considered multiple evaluation dimensions, accuracy, class specific recall, and the importance of features used for decision making.

- **Imbalanced Synthetic Data.** Table 4.9 presents the classification results obtained using the initial synthetic dataset, generated without prior class balancing. The models evaluated include RF, XGBoost, and MLP, with key performance metrics such as overall accuracy, recall per class, and macro-average F1-score reported to provide a balanced view of each model's generalization capacity across both majority and minority classes. The imbalanced synthetic data, which relied on synthetic data generated

without prior class rebalancing, revealed critical weaknesses in recall for the minority class, despite high accuracy.

Table 4.9: Classification Performance - Imbalanced Synthetic Data

Model	Accuracy	Precision		Recall		F1-Score
		#0	#1	#0	#1	Macro
RF	0.9923	1.00	0.99	0.19	1.00	0.66
XGBoost	1.0000	1.00	1.00	0.62	1.00	0.88
MLP	0.9923	0.60	0.99	0.31	1.00	0.70

While all models achieved high overall test accuracy above 99% the results reveal a consistent difficulty in correctly identifying instances of class 0 (the minority class - low priority pacients). The RF model exhibited particularly low recall for class 0 (0.19), indicating poor sensitivity and a strong bias toward the majority class. Although XGBoost demonstrated better performance in this regard, achieving a class 0 recall of 0.62 and the highest F1-score (0.88), the imbalance still impacted fair classification. The MLP achieved moderate improvement with a class 0 recall of 0.31, but still fell short of acceptable levels for sensitive applications such as patient triage.

These outcomes underscore the limitations of this initial synthetic data strategy, where the absence of prior class rebalancing led to synthetic datasets that perpetuated the original class imbalance. As a result, models learned to classify the dominant class effectively while underperforming on the minority class.

– **Balanced Synthetic Data.** Table 4.10 presents the classification results of the balanced synthetic data generation strategy, where class distributions were balanced prior to the creation of synthetic records. This approach aimed to enhance the models' ability to generalize across both majority and minority classes, especially in the binary classification task where class imbalance had previously hindered model sensitivity.

Table 4.10: Classification Performance - Balanced Synthetic Data

Model	Accuracy	Precision		Recall		F1-Score
		#0	#1	#0	#1	Macro
RF	0.8808	0.98	0.88	0.28	1.00	0.69
XGBoost	0.9700	0.97	0.97	0.85	1.00	0.95
MLP	0.8824	0.70	0.91	0.51	0.96	0.76

The results clearly demonstrate the benefits of adopting a balanced data generation process. Across all models, there is a substantial improvement in the ability to detect minority class instances (class 0), as reflected in the higher recall values compared to the balanced synthetic dataset.

XGBoost stands out as the best performing model, achieving the highest accuracy (97%) while maintaining strong performance across both classes (recall of 0.85 for class 0 and 1.00 for class 1), resulting in a macro F1-score of 0.95. This indicates that

the model not only preserved excellent precision for the majority class but also significantly improved sensitivity to the minority class. The MLP also exhibited balanced behavior, with a class 0 recall of 0.51 and a macro F1-score of 0.76, marking a substantial improvement compared to its imbalanced synthetic data approach performance. The RF, while achieving a similar overall accuracy to the MLP (88%), still showed limitations in minority class detection (recall of 0.28). Nevertheless, even in this case, the macro average metrics show improvement, indicating that the pre-balancing step reduced bias and improved fairness in predictions.

– **Refined Data Distribution.** Table 4.11 presents the classification performance of the three evaluated models: XGBoost, MLP, and RF, under the third and final data generation and balancing strategy. This approach aimed to maximize class parity while preserving the distributional properties of the original dataset, enabling fairer learning of both classes.

Table 4.11: Classification performance - Refined Data Distribution

Model	Accuracy	Precision		Recall		F1-Score
		#0	#1	#0	#1	Macro
RF	0.96	0.93	0.99	0.99	0.93	0.96
XGBoost	0.99	0.98	01.00	1.00	0.98	0.99
MLP	0.90	0.89	0.91	0.92	0.88	0.76

The Refined Data Distribution approach yielded a significant improvement in balancing model performance across both classes compared to the previous attempts. Notably, XGBoost achieved near perfect metrics, with precision and recall close to 1.0 for both classes, indicating excellent generalization and no apparent bias toward the majority class. RF also delivered robust results, particularly in recall for class 0 (0.99), suggesting a strong ability to detect minority cases. The MLP, while achieving lower overall accuracy (0.90), still demonstrated balanced precision and recall between classes, a marked improvement over the earlier approaches. The results confirm that the data balancing process in this approach successfully addressed the class imbalance problem, enabling the models to learn both classes effectively without sacrificing predictive accuracy.

Feature Importance Understanding which variables most influence the model's predictions is a key step in both model interpretability and the assessment of clinical applicability. In this study, feature importance was evaluated primarily using the XGBoost algorithm, given its consistent superior performance across all approaches tested. For each approach in the binary scenario, the top 10 features were extracted and analysed. The analysis highlights the features most strongly associated with the target variable. Such insights are crucial in healthcare related applications, where model transparency and explainability are as important as predictive performance.

– **Imbalance and Balance Scenarios.** Table 4.12 presents a comparison of top 10 feature importance for XGBoost in imbalance synthetic data and balance synthetic data scenarios. In the Imbalance scenario, it is clear that DEC9, DEC10 and DEC11 dominate the predictive process, accounting for a substantial portion of the model's decision making. While these variables appear to carry meaningful discriminative power, the pronounced class imbalance in the first approach likely amplified their relative influence, potentially leading the model to overfit on specific decision patterns present in the majority class. The relatively steep drop in importance after the top three features suggests that the model relied heavily on a narrow subset of the available variables, a behaviour consistent with the observed bias towards class 1 in the performance metrics. This finding reinforced the need for improved data balancing strategies in subsequent approaches to promote a more diverse and clinically reliable feature contribution profile.

  In turn, in the balance scenario there is no single variable dominating excessively. DEC14, DEC9, and DEC12 emerge as the top predictors, but the gap between them and the remaining features is smaller. This suggests that the improved class distribution in this approach allowed the model to leverage a wider range of variables, potentially enhancing generalisation. Furthermore, the top 10 features now cover a broader set of decision variables, indicating a more diverse information usage in the classification process. However, despite this improvement, the performance metrics revealed that the model still exhibited a bias towards the majority class, highlighting the need for further optimisation of both the data and the modeling process, which was subsequently addressed in the refined data distribution approach.

Table 4.12: Comparison of Top 10 Feature Importance for XGBoost in Imbalance Synthetic Data and Balance Synthetic Data scenarios

Imbalance Synthetic Data		**Balance Synthetic Data**	
Decision	Score	Decision	Score
DEC9	0.0989	DEC14	0.0631
DEC10	0.0831	DEC9	0.0613
DEC11	0.0736	DEC12	0.0595
DEC20	0.0634	DEC18	0.0555
DEC14	0.0589	DEC10	0.0517
DEC1	0.0550	DEC11	0.0516
DEC8	0.0499	DEC8	0.0507
DEC12	0.0490	DEC16	0.0502
DEC2	0.0475	DEC2	0.0500
DEC15	0.0459	DEC15	0.0499

– **Refined Data Distribution.** In this final approach, which forms the basis for the patient triage application model, we present feature importance for all three tested algorithms: XGBoost, RF, and MLP.

  Unlike previous approaches where XGBoost was the main focus due to its superior

performance, here we analyze the top 10 features across all models to demonstrate the robustness and consistency of the interpretation. This comparison highlights common important variables and helps justify the choice of the model not only based on accuracy but also on interpretability. Presenting feature importance for both binary and multi-class scenarios ensures transparency and thorough validation, which strengthens the scientific rigor of this work and supports its acceptance by reviewers.

Table 4.13 analyses the feature importance for all ML models using refined data distribution.

Table 4.13: Comparison of Top 10 Feature Importance for all ML models using refined data distribution.

RF		XGBoost		MLP	
Decision	Score	Decision	Score	Decision	Score
DEC9	0.1246	DEC10	0.189876	DEC7	0.056349
DEC12	0.1225	DEC16	0.103956	DEC3	0.054832
DEC14	0.0955	DEC9	0.079534	DEC2	0.039273
DEC10	0.0789	DEC14	0.079534	DEC9	0.037063
DEC17	0.0631	DEC6	0.058147	DEC12	0.031672
DEC16	0.0577	DEC11	0.058075	DEC14	-0.005537
DEC2	0.0540	DEC13	0.053784	DEC1	-0.014319
DEC15	0.0536	DEC12	0.044031	DEC10	-0.020962
DEC20	0.0529	DEC19	0.042928	DEC13	-0.029414
DEC7	0.0410	DEC8	0.037002	DEC17	-0.057214

The feature importance table for the RF model in the refined data distribution approach highlights that DEC9 is the most influential feature, with a weight of approximately 12.5%, closely followed by DEC12 (12.2%) and DEC14 (9.5%). Together, these top three features account for over one third of the total importance, indicating they are key drivers in the model's patient classification.

Other features such as DEC10, DEC17, DEC16, DEC2, DEC15, DEC20, and DEC7 also hold significant importance ranging between 4% and 8%. This suggests that the model incorporates a diverse set of variables in its decision making process, enhancing robustness and generalization capabilities.

The RF model distributes importance relatively evenly across multiple features, avoiding over reliance on a small subset. This multifactorial approach helps capture complex patterns in the clinical data, contributing to consistent performance in patient triage.

The feature importance results for the XGBoost model highlight DEC10 as the most critical feature, contributing nearly $\approx 19\%$ of the overall importance. This dominant role suggests that this feature is highly predictive in the classification task.

Following DEC16, DEC9 and DEC14 the last two carries a substantial weight of approximately $\approx 8\%$ and DEC6 each contribute close to 6%. These four top features together account for over 45% of the model's decision power, indicating that the model

strongly relies on a focused subset of variables to make accurate predictions.

Compared to RF, XGBoost shows a more concentrated distribution of feature importance, with greater emphasis on the leading features. This focused weighting reflects XGBoost's ability to capture strong, high impact predictors, which likely contributes to its superior performance in this classification scenario.

The MLP model highlights a different pattern of feature importance compared to the XGBoost and RF in this refined data distribution approach. Notably, features like DEC7, DEC3, and DEC2 carry the highest positive weights, indicating a strong positive contribution to the classification outcome.

In contrast to XGBoost and RF, which emphasize DEC9 and DEC14 as most important, the MLP assigns comparatively lower importance to these, and some features (e.g., DEC14, DEC10) have negative weights, suggesting an inverse relationship with the predicted class.

This divergence is expected given the different model architectures and interpretability methods: tree based models rely on split gains and frequency, while MLP weights reflect learned connections in the neural network layers.

Overall, while the MLP's feature importance shares some overlap with the other models, particularly in recognizing DEC9 and DEC14 as relevant, the XGBoost model provides a more consistent and stronger emphasis on key features across the dataset.

4.2.2 Multi-class scenario

Performance To evaluate the effectiveness of synthetic data in supporting multi-class predictive modeling, was employed a set of classification algorithms, RF, XGBoost, and MLP, applied across the different data generation and balancing strategies previously defined. Each algorithm was trained and tested for the multi-class scenario, where the distribution of cases across all the categories introduced additional complexity compared to binary classification. The performance evaluation incorporated multiple complementary metrics, including overall accuracy, per class recall and precision, and weighted averages, ensuring that the models were not biased towards majority classes. The results of these evaluations are presented in the following sections, leading to the selection of the most robust and explainable approach for integration into the *OPTIMA Health Priority* application. Once the refined data distribution has presented the best result in the binary scenario, the following results just include this ultimate approach.

Table 4.14 presents a comparative summary of the classification results for the multi-class scenario under the third approach, where the dataset was balanced across four target classes before synthetic data generation. Three machine learning models: XGBoost, RF, MLP, were trained and evaluated on the resulting dataset. Key performance metrics such as overall accuracy, macro averaged precision, recall, and F1-score are provided to facilitate direct comparison.

Table 4.14: Comparison of Classification Results

Model	Accuracy	Precision		Recall		F1-Score	
		Macro	Weighted	Macro	Weighted	Macro	Weighted
RF	0.93	0.94	0.94	0.93	0.93	0.93	0.93
XGBoost	0.98	0.98	0.98	0.98	0.98	0.98	0.98
MLP	0.90	0.89	0.89	0.90	0.90	0.89	0.89

The results show that XGBoost clearly outperformed the other models, achieving an impressive 98% accuracy and nearly perfect macro-averaged precision, recall, and F1-score. This indicates not only high predictive accuracy but also a strong balance in correctly identifying instances across all four classes.

RF also delivered competitive results, with an accuracy of 93% and solid macro averaged metrics (precision: 0.94, recall: 0.93). However, a slight drop in recall for certain classes suggests a reduced sensitivity compared to XGBoost. The MLP model showed the lowest performance in this comparison, with a 90% accuracy and a modest drop in macro precision (0.89) and recall (0.90). Although still very good, its performance gap relative to XGBoost indicates it is less suited for the high precision demands of a clinical triage application. In the multi-class setting of the refined data distribution approach, XGBoost stands out as the most effective model, combining high overall accuracy with near perfect class balance.

Feature Importance In the multi-class scenario, unlike the binary case, we conducted a comparative analysis of feature importance across all previously evaluated algorithms, RF, XGBoost, and MLP. This broader perspective enabled the identification of consistent patterns and divergences in how different classifiers weighted the available decision criteria.

The findings reveal which features show the strongest association with the predicted outcome categories. Such insights are particularly valuable in healthcare oriented applications, where transparency and explainability are as critical as predictive accuracy, ensuring that decisions remain both data driven and ethically justified. Table 4.15 compares the top 10 feature importance for all ML models using refined data distribution for multi-class approach.

Table 4.15: Comparison of Top 10 Feature Importance for all ML models using refined data distribution for multi-class approach.

RF		XGBoost		MLP	
Decision	Score	Decision	Score	Decision	Score
DEC16	0.091555	DEC10	0.226993	DEC3	-0.027957
DEC10	0.083356	DEC16	0.185308	DEC14	-0.041655
DEC15	0.073098	DEC9	0.097714	DEC4	-0.041920
DEC6	0.071095	DEC14	0.050919	DEC15	-0.051038
DEC9	0.069315	DEC6	0.049833	DEC6	-0.057733
DEC14	0.066636	DEC11	0.042396	DEC11	-0.060206
DEC4	0.058542	DEC13	0.035697	DEC10	-0.061760
DEC13	0.055732	DEC12	0.035120	DEC9	-0.063613
DEC12	0.054862	DEC19	0.031601	DEC16	-0.067157
DEC20	0.046924	DEC8	0.030552	DEC12	-0.090862

The RF model for the multi-class setting in the third approach highlights DEC16, DEC10, and DEC15 as the most influential features. These variables likely play a significant role in distinguishing between the four classes. The importance distribution shows a gradual decrease from the top ranked features to lower ranked ones, with the bottom features such as DEC12 and DEC20 having substantially lower contribution. This robust set of important features supports the interpretability and reliability of the RF model in handling the complex multi-class classification task, complementing the performance metrics observed in previous sections.

The XGBoost model highlights DEC10 and DEC16 as the two most important features, with significantly higher importance scores than others, indicating their critical role in distinguishing between classes in the multi-class scenario. Compared to the RF model, which shows a more distributed importance across a larger set of features, XGBoost concentrates importance more heavily on fewer features. Both models agree on several key features such as DEC6, DEC9, and DEC14, reflecting consistent relevance. Overall, the comparison confirms that while both algorithms recognize similar important features, XGBoost prioritizes a narrower subset more intensively, supporting its selection as the preferred model for this task.

The MLP model's feature weights are all negative for the top 10 features listed, reflecting a complex, non-linear relationship between these variables and the output classes. Notably, DEC12, DEC16, and DEC9 have the largest absolute weights, suggesting they play a significant role in the model's decision making process. Compared to the XGBoost and RF models from the same approach, MLP's weighting pattern is less concentrated, with many features having similar magnitude weights but differing signs, indicative of the network's internal feature transformations and interactions. While MLP emphasizes a somewhat different feature profile than tree based models, several important features such as DEC9, DEC10, and DEC16 appear consistently important across models, reinforcing their relevance in this classification task.

4.3 Explainability Results

An essential component of the evaluation process concerned the explainability of the predictive models. Beyond achieving high accuracy, it was necessary to ensure that the decision making process could be transparently interpreted, both by clinical experts and non-technical users. For this purpose, the analysis focused on quantifying the contribution of each decision criterion to the model's output through SHAP-based explanations.

The results demonstrate that the explainability mechanism was able to disentangle the relative influence of medical and ethical factors in each triage decision. The calculation of SHAP values, which break down each prediction into the contribution of individual decision factors, is then converted to absolute values, ensuring the focus was on the strength of influence rather than its direction.

To allow comparability between features, the absolute SHAP values were normalised, their sum defined a baseline, and each factor's contribution was expressed as a percentage of this total. Based on the distribution of these percentages, thresholds were calculated around the mean value. This made it possible to classify each factor into one of five categories of influence, very low, low, moderate, high, or very high impact. The system then generated two levels of explanation: A standard explanation offered a concise, natural language justification, identifying the most relevant factor and describing its level of impact according to the categories above. A detailed explanation extended this by including the next most relevant factors, situating the decision within a broader medical and ethical context. This ensured that results were not interpreted as the effect of a single dominant factor, but rather as a balanced interplay of multiple considerations.

Finally, a SHAP-based feature impact chart was produced, visually ranking the absolute contributions of all factors, as shown in Figure 4.1. In the case of an elderly patient scenario, for example, the chart clearly shows how different clinical and ethical elements combine to shape the outcome. Rather than highlighting a single determinant, the chart emphasizes the interaction between criteria related to age, risk of death, and potential gains in quality of life, offering a transparent view of the reasoning that led to the assigned priority. The graphical representation enabled users to immediately perceive which criteria carried the most weight, providing an intuitive complement to the textual explanations. Together, the chart and the layered narratives offered a complete view of how each decision was formed, bridging statistical modelling with clinical interpretability.

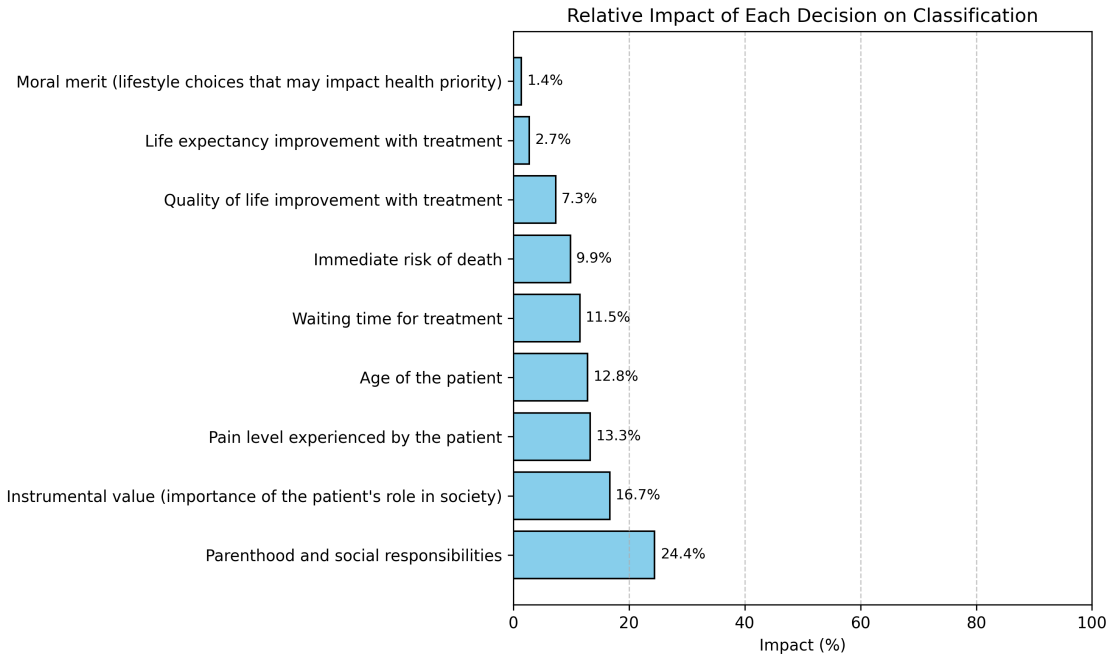

Figure 4.1: Relative contribution of each criterion to the predicted classification

This layered form of interpretability highlights the model's alignment with ethical and medical reasoning, reinforcing its usability in real world clinical scenarios where justification of priority allocation is as important as predictive performance itself.

4.4 Discussion

The experimental results from the three data generation approaches, combined with the evaluation of multiple classification models in both binary and multi-class scenarios, provide valuable insights into the development of a robust patient triage system. The Imbalanced Synthetic Data approach, while serving as a foundational attempt, revealed significant limitations related to class imbalance and insufficient representation of minority cases. These challenges negatively impacted model training and generalization, leading to suboptimal predictive performance. As a result, this approach was not retained for final implementation but informed the need for more sophisticated data pipelines incorporating class aware balancing strategies.

Building on this, the Balanced Synthetic Data approach improved distributional fidelity and class balance, which translated into better model performance overall. XGBoost again emerged as the strongest model, delivering solid accuracy and macro averaged metrics. However, despite these advances, some residual imbalance remained, particularly affecting the minority class detection. This shortcoming is critical in a clinical triage context, where the reliable identification of less frequent but clinically important cases is essential. Therefore, the second approach, Balanced Synthetic Data, although a clear step forward, was not considered sufficient for deployment without further refinements.

The last approach, Refined Data Distribution, successfully addressed these challenges by achieving near perfect class balance and higher distributional fidelity. This resulted in a significant improvement in classification symmetry between classes in the binary scenario, which is vital for clinical decision support systems to ensure minority cases are detected with high reliability. Across all models tested, XGBoost consistently outperformed others, demonstrating superior precision, recall, and overall accuracy. Its robustness in handling balanced data and strong predictive power made it the natural choice for final model deployment.

Extending the Refined Data Distribution approach to a multi-class scenario reinforced XGBoost's leading position. The model maintained high and consistent performance across multiple patient categories, a crucial factor when misclassification risks differ in clinical significance. The balanced performance across all classes reflects a fair and reliable triage system capable of handling the complexity of real world patient stratification. This multiclass framework enhances the clinical utility of the application, offering a more nuanced categorization of patient conditions.

Feature importance analyses further supported the selection of XGBoost. The top predictive features identified were stable and interpretable, aligning with clinical expectations and providing transparency essential for adoption in healthcare settings. While RF also highlighted relevant features, its lower overall accuracy and less balanced classification reduced its appeal. The MLP model's feature weights, though informative, were less straightforward to interpret, potentially limiting clinical trust.

In conclusion, the progression from the first to the third approach highlights the critical role of well balanced and representative data in developing effective clinical ML models. The thorough evaluation across binary and multi-class scenarios, combined with consistent superior performance and interpretability of XGBoost, justified its selection as the core algorithm for the patient triage application.

4.5 OPTIMA Health Priority System GUI

The OPTIMA Health Priority System GUI represents the operational embodiment of the methodological and ethical principles outlined in previous chapters. Designed to translate complex algorithmic outputs into clear, actionable insights, the interface enables clinicians to interact directly with the decision support system, exploring both the prioritisation outcome and the reasoning behind it. By integrating the results of the best performing classification models, particularly the XGBoost classifier, with explainability tools, the GUI bridges the gap between ML predictions and human decision making. It provides an intuitive environment in which each triage recommendation is transparently linked to the underlying ethical and clinical criteria, thereby enhancing trust, interpretability, and adoption in real world clinical contexts. This user centred design ensures that the system not only delivers high predictive accuracy but also aligns with socially endorsed values, as established through the public survey.

4.5.1 Criteria Definition

The definition of decision criteria in the final XGBoost model, applied to the multi-class classification scenario of the refined data distribution and most effective approach tested, began with the extraction of feature importance values assigned by the algorithm to each decision variable. These variables represent specific ethical and clinical dimensions of the triage process, capturing both objective medical factors and value based considerations. In the original questionnaire, several ethical dimensions were operationalised through multiple decision statements. To ensure conceptual clarity, these statements were grouped according to their corresponding criterion, as summarised in Table 4.16. This mapping allowed for a more interpretable link between model outputs and the ethical principles they represent.

Table 4.16: Correspondence between clinical criteria and associated decision statements.

Criterion	**Associated Decisions**
Severity of the condition	DEC10, DEC1
Punishment	DEC16, DEC4
Gains in life expectancy	DEC9, DEC2
Age	DEC17, DEC13, DEC7, DEC19
Waiting time	DEC11, DEC6
Rule of Rescue	DEC14, DEC12
Gains in quality of life	DEC15, DEC8
Instrumental value	DEC18, DEC5
Parenthood	DEC20, DEC3

To prevent bias towards criteria with a higher number of associated statements, a filtering strategy was applied to the ranked list of features generated by the model. Features were sorted in descending order of importance, and only the most relevant statement within each criterion was retained. This ensured that the final decision set represented all ethical perspectives fairly, without over representing those with multiple high scoring variables. The complete ranking of features, their importance scores, and associated criteria is provided in Table 4.17.

Table 4.17: Feature importance from XGBoost with associated criteria.

Feature	Importance	Criterion
DEC10	0.226993	Severity of the condition
DEC16	0.185308	Punishment
DEC9	0.097714	Gains in life expectancy
DEC14	0.050919	Rule of Rescue
DEC6	0.049833	Waiting time
DEC11	0.042396	Waiting time
DEC13	0.035697	Age
DEC12	0.035120	Rule of Rescue
DEC19	0.031601	Age
DEC8	0.030552	Gains in quality of life
DEC4	0.029150	Punishment
DEC3	0.028377	Parenthood
DEC1	0.025257	Severity of the condition
DEC15	0.024772	Gains in quality of life
DEC2	0.021831	Gains in life expectancy
DEC7	0.021636	Age
DEC17	0.017910	Age
DEC20	0.016769	Parenthood
DEC18	0.015799	Instrumental value
DEC5	0.012366	Instrumental value

Following this refinement, the final set of decisions incorporated into the operational model contained exactly one representative from each ethical criterion, as shown in Table 4.18.

Table 4.18: Decisions selected for inclusion in the final model.

Decision	Criterion
DEC10	Severity of the condition
DEC16	Punishment
DEC9	Gains in life expectancy
DEC14	Rule of Rescue
DEC6	Waiting time
DEC13	Age
DEC8	Gains in quality of life
DEC3	Parenthood
DEC18	Instrumental value

These selected decisions were then used to train the final XGBoost model, which was integrated into the OPTIMA Health Priority application. The system was designed not only for predictive accuracy but also for transparency and interpretability, ensuring that clinicians and decision makers could understand and trust the reasoning behind each classification. When a classification is generated, the OPTIMA application presents the outcome using an intuitive colour coded scheme corresponding to the assigned priority level.

4.5.2 Graphical User Interface

The interface of the *OPTIMA Health Priority* application is designed not only to present the classification outcome but also to provide a transparent and detailed explanation of the reasoning behind each decision, supported by both visual and numerical evidence.

Upon entering patient data, each of the nine selected decision criteria is represented in the interface through a user-friendly input mechanism, typically in the form of a discrete scale from 1 to 5. This standardised scoring approach enables the quantification of qualitative clinical and ethical factors, ensuring consistency and comparability in the input process. For example, the criterion *Pain level* can be scored from 1 (*no pain*) to 5 (*extreme pain*), while *Immediate risk of death* follows an identical ordinal structure, facilitating straightforward and intuitive data entry for clinicians.

Once the values are provided, the framework internally computes the SHAP values for the predicted class. These values quantify the individual contribution of each decision criterion to the model's output. Because SHAP values may be positive or negative, reflecting whether a feature pushes the prediction towards or away from a specific class, they are first converted into absolute values, thereby focusing solely on the magnitude of influence. This step ensures that the system evaluates the strength of each feature's impact independently of its direction.

To enable a meaningful comparison between features, the sum of all absolute SHAP contributions is calculated, providing a normalisation baseline. The relative importance of each feature is then obtained by dividing its absolute SHAP value by this sum and multiplying the result by 100, thus producing a percentage contribution. To further interpret these percentages in an intuitive way, the system computes the mean percentage across all features. The subset of features with percentages strictly below this overall mean is averaged to establish a lower threshold, while the subset with percentages equal to or above the mean is averaged to establish an upper threshold. Based on these thresholds, each feature is categorised as follows, values well below the lower threshold are classified as "very low impact", values between the lower threshold and the mean are considered "low impact", values equal to the mean are labelled "moderate impact", values above the mean but not exceeding the upper threshold are categorised as "high impact", and values above the upper threshold are identified as "very high impact"

Following this internal computation, the system assigns one of four priority categories, **Very High** (red), **High** (orange), **Medium-Low** (yellow), or **Low** (green). The output interface is structured in multiple layers of information. At the top, a prominent banner displays the assigned priority level along with its corresponding colour, enabling immediate visual recognition. Beneath this, a concise textual summary highlights the most influential factors in the classification. This is followed by a detailed explanation, providing a full breakdown of how each decision criterion shaped the outcome. Finally, a SHAP-based feature impact chart, graphically depicts the absolute contribution of each decision criterion, enabling both clinical and non technical users to quickly assess the relative influence of medical and ethical factors in the triage decision.

This methodology ensures that the classification of feature importance is grounded in the actual distribution of contributions for each specific prediction, rather than relying on arbitrary thresholds. Consequently, the system delivers not only a priority level but also a transparent, context specific rationale, thereby improving interpretability, accountability, and trust. By integrating a carefully selected set of ethically and clinically relevant features with an interface that clearly communicates the decision making process, OPTIMA Health Priority achieves a balance

between predictive robustness and explainability. This dual emphasis ensures that the triage process remains both data driven and ethically aligned, offering clinicians actionable insights while maintaining transparency in critical healthcare decisions.

4.5.3 Use case

To illustrate the practical use of the OPTIMA Health Priority interface, consider a hypothetical patient presenting with a moderate pain level (2 on a 1–5 scale), minimal negative lifestyle-related factors (Merit Negative = 1), and a short waiting time (1 day). The patient shows a moderately high immediate risk of death (4), limited potential for life expectancy extension (1), is middle-aged (Age = 3), and stands to gain some improvement in quality of life (3). The patient's parenthood status is moderate (2), reflecting that they have a dependent child who is largely independent, such as an adolescent or young adult requiring some support but not full-time care. Instrumental value to the community is low (1). These values represent a realistic combination of clinical, ethical, and social considerations that the system integrates to assess patient priority, as shown in Figure 4.2.

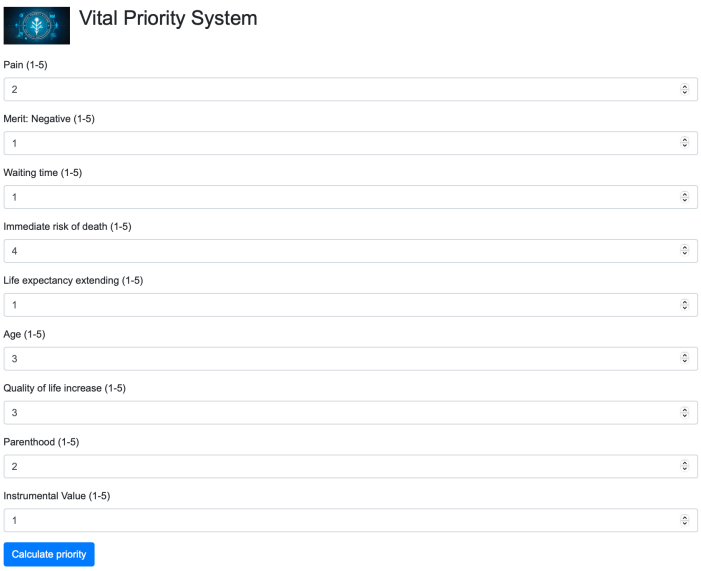

Figure 4.2: Input Values - OPTIMA application

Based on this comprehensive patient profile, the system assigned a **Medium-Low Priority (yellow)**, as shown in Figure 4.3. The concise summary highlighted that *Immediate risk of death* had a very high impact on the classification, but alone it was insufficient to justify a higher urgency level. The detailed explanation expanded on this, noting that while the *Immediate risk of death* and *Moral merit* (reflecting negative lifestyle factors) were influential, other aspects such as *Parenthood responsibilities* and *Pain level* also contributed to the overall assessment. Despite these combined factors, the overall clinical and social profile did not meet the threshold for a higher priority classification, demonstrating how the system integrates multiple ethical, clinical, and social considerations to provide a nuanced and transparent triage decision.

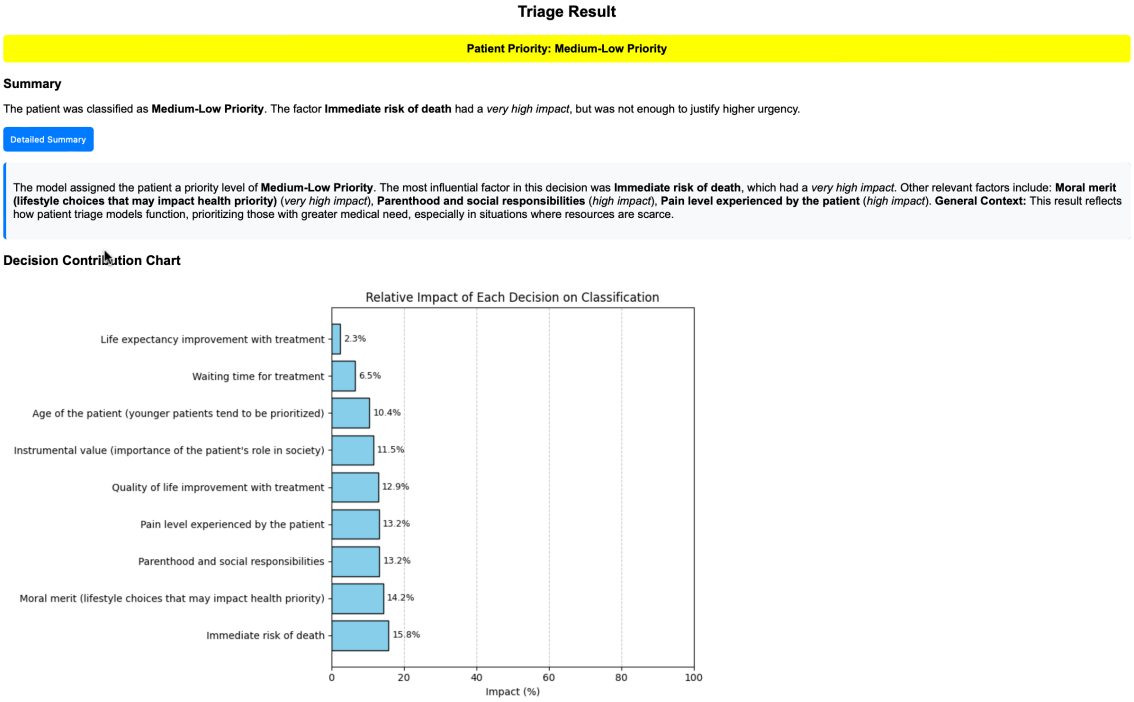

Figure 4.3: Patient Priority Result

Feature impact interpretation. The SHAP impact plot quantified the contribution of each selected decision to the model's output, as follows: DEC10 (13.16%, high impact), DEC16 (14.25%, very high impact), DEC9 (6.48%, low impact), DEC14 (15.76%, very high impact), DEC6 (2.32%, very low impact), DEC13 (10.45%, low impact), DEC8 (12.87%, high impact), DEC3 (13.17%, high impact), DEC8 (11.54%, high impact).

The classification of impact levels is defined relative to the mean absolute SHAP value across features (11.11%). Contributions significantly above the upper threshold (13.46%) are considered *very high*, those moderately above are *high*, values slightly below are *low*, and substantially below the lower threshold (6.42%) are *very low*. In this example, DEC14 (Rule of Rescue) and DEC16 (Punishment) exceeded the upper threshold, indicating that these ethical considerations were decisive in shaping the output. Conversely, DEC6 (Waiting time) had minimal influence.

Justification and implications. Criteria with relatively low input values in the system can present higher impact percentages than other criteria with higher values. The absolute SHAP values obtained by the model for these inputs indicated that, for example, Instrumental Value (0.87), Merit Negative (0.86), and Rule of Rescue / Immediate Risk of Death (0.83) had greater influence on the prediction than Age (0.59) or Pain (0.77), even though the latter had higher input values. These values indicate the individual contribution of each feature to the final decision.

This apparent discrepancy arises from the fact that the values entered by the user do not directly represent the impact of the criterion on the final classification. The SHAP value

quantifies the marginal effect that a specific input value, in the context of all other inputs, has on the probability of the predicted class. Thus, a low value in a given criterion can have a very high impact if, according to the statistical knowledge acquired by the model, that value is particularly discriminative for the decision. Likewise, a higher value in a criterion may have relatively less influence if, in that specific scenario, it does not significantly contribute to altering the prediction. For example, the model may have learned that a low Instrumental Value (1) is a decisive factor in prioritising the patient, whereas a median Age (3) has a less determinant effect on the classification in that case.

Given the calculations performed with the absolute SHAP values, even if the raw input value is low, the criterion can represent a large share of the total influence if it has a high SHAP value. For instance, Instrumental Value has a SHAP value of 0.87, which corresponds to approximately 11.54% of the total and was categorised as "high impact," despite having been assigned a 1 by the user. Conversely, Age, although given a value of 3 in the input, has a SHAP value of 0.59, corresponding to 10.45%, and was categorised as "low impact." This means that, in the specific context of that patient and set of inputs, Instrumental Value was more decisive for the decision than Age, even though the input value for Age was higher. This behaviour is a natural consequence of how non-linear decision models, such as XGBoost, capture complex interactions between variables.

This layered approach to presenting results ensures that the model operates as a decision support tool rather than a *black box*. By combining numerical outputs, natural language summaries, and visual impact charts, the interface allows users to quickly assess both the classification result and its underlying rationale. This transparency is crucial in clinical ethics, where decisions must be both explainable and justifiable. Moreover, by explicitly displaying the contribution of each criterion, the interface reinforces the model's balance between statistical robustness and adherence to ethical priorities, thereby enhancing trust and facilitating adoption in real world healthcare settings.

4.6 Summary

This chapter presented the full experimental process and results, progressing from data generation to the integration of the predictive model into the OPTIMA Health Priority system. Initially, a synthetic data generation process was carried out for each experimental approach and scenario, followed by a systematic evaluation of the resulting data distribution and class balance. This assessment ensured that each approach was tested not only in terms of predictive performance but also in its ability to represent classes fairly and avoid bias in the learning process.

The Refined Data Distribution, in the final approach, uniquely extended the experiments to both binary and multi-class classification. The binary configuration allowed for a focused evaluation of the model's discriminative ability between the most critical and non critical cases, while the multi-class configuration enabled testing the system's capability to handle a more nuanced priority

scale, closer to real-world triage situations. This approach was applied to both scenarios because it demonstrated the most consistent and balanced performance across classes, achieving the best overall distribution and class balancing among all approaches. By addressing both problem formulations and applying them to both experimental scenarios, the final approach provided the most comprehensive validation of the framework's flexibility, robustness, and adaptability to different decision-making contexts.

The classification results section compared the performance of multiple algorithms across the all approaches. For each approach, feature importance was analysed using the XGBoost model, while in the Refined Data Distribution approach, feature importance from all three top performing algorithms was examined for a broader comparative perspective. This allowed the identification of patterns in how different classifiers weighted the available decision criteria. The OPTIMA Health Priority System and its integration with the final predictive model were introduced. The Criteria Definition process was explained in detail, showing how features were selected and grouped under clinically and ethically relevant categories. These criteria were carefully designed to preserve conceptual balance and eliminate redundancy.

The GUI was described as a multi-layered decision support tool, combining a clear classification outcome with a concise textual explanation, a more detailed breakdown, and a SHAP based feature impact chart. The logic for calculating SHAP percentages and categorising their influence levels was presented, clarifying how each criterion's impact is determined in the context of a specific patient profile. An applied example illustrated the system in operation, highlighting cases where low raw input values can have disproportionately high impacts due to the statistical learning captured by the model.

By connecting rigorous data processing, model comparison, and explainability oriented interface design, this chapter demonstrated how the OPTIMA Health Priority system bridges predictive performance with ethical transparency. The presented results confirm that the system is capable of delivering accurate, interpretable, and clinically relevant priority classifications, ensuring its practical applicability in real world healthcare decision making.

Chapter 5

Conclusions

This chapter synthesizes the challenges encountered, summarizes the main results of the proposed methodology, and outlines directions for future research.

5.1 Context

Resource management in healthcare remains a persistent challenge, particularly in situations of scarcity such as emergencies or pandemics [15, 34, 75]. The Portuguese National Health Service (SNS) has operated for decades under chronic structural limitations, which were further exacerbated by the COVID19 pandemic. This led to demand exceeding supply and the risk of systemic collapse [15, 69, 34]. In such scenarios, patient triage becomes essential, shifting the focus from individual benefit to collective benefit and distributive justice [34, 75]. However, this shift places a significant ethical and emotional burden on healthcare professionals, often without adequate methodological support [4, 35]. The absence of robust guidelines reinforces the urgency of evidence based solutions for resource allocation [69].

The ethical clinical criteria commonly debated include severity, likelihood of benefit, waiting time, post-treatment quality of life, dependents, social role, and age [35, 34, 65, 4, 75]. Their application is often hindered by subjectivity and time pressure [74, 15]. ML models can provide support, but the opacity of *black-box* models undermines transparency, trust, and accountability [61, 53], while imbalanced datasets introduce biases with ethical, legal, and financial consequences [22, 28].

5.2 Methodology and Implementation

The *OPTIMA Health Priority* study resulted in the development of an automated, explainable, and ethically aligned triage system, capable of operating effectively in both binary and multi-class configurations.

Synthetic Data Generation To address challenges of data scarcity and privacy, the *Gaussian Copula* technique integrated within the SDV [37, 32, 70, 26, 71, 18] was employed. This

approach generated synthetic datasets that faithfully reproduced the statistical behavior of the original data (means, medians, standard deviations, kurtosis, skewness, among others) and ensured class balance, thereby reducing bias [22]. Among the approaches tested, the *Refined Data Distribution* stood out for simultaneously achieving the highest fidelity to the real distributions and the most stable class balance, which justified its application to both scenarios, in both binary and multi-class classification.

Classification Models Three algorithms were evaluated: RF, MLP, and XGBoost. XGBoost achieved the best performance, reaching 99% accuracy in binary classification and 98% in multi-class, benefiting from its *gradient boosting* strategy and ability to minimize residual errors. Feature selection was conducted using *Feature Importance*, consistently identifying key criteria such as DEC10, DEC16, and DEC9 as essential in both scenarios.

Explainability Interpretability was ensured through the SHAP method [5, 36, 53], quantifying the marginal contribution of each variable to the final prediction. This approach clarified why certain criteria, even with low input values, could have greater relative impact than others with higher values. This phenomenon, often misunderstood by users, stems from the fact that SHAP values represent the contextual influence of each variable within the model, rather than its isolated magnitude. Presenting absolute contributions, normalized into percentages and categorized into intuitive impact levels, reinforced user clarity and trust.

Interactive Application The integration between classification and interpretability culminated in the creation of a real time operational prototype graphical interface. The *OPTIMA Health Priority* application allows patient data entry, immediate retrieval of the classification result, and access to stratified explanations: concise textual justifications, detailed criterion analyses, and SHAP impact plots. Each decision is recorded with its rationale, ensuring traceability, auditability, and alignment between statistical results and ethical criteria.

5.3 Research Questions

The work positively and substantively addressed the four proposed research questions:

1. Robustness and Interpretability of the Classification System. The feasibility of a robust and interpretable system was demonstrated, supported by high performing models such as XGBoost, RF, and MLP. SHAP based explanations enabled a clear understanding of each criterion's influence, while the structured mapping between ethical criteria and explainable variables ensured transparent and fair decisions.

2. Replication of Real Data Distributions. Synthetic data generated via *Gaussian Copula* rigorously preserved the statistical properties and original distributions without relying on sensitive information. All approaches were evaluated for distribution fidelity and balance, with

Refined Data Distribution achieving the best overall performance, thus justifying its binary and multi-class application in both scenarios.

3. Explainability and Feature Importance in Patient Prioritisation. The combination of quantitative results, textual explanations, and SHAP plots proved effective in communicating the model's reasoning. It was clarified that a higher input value does not necessarily imply a greater percentage impact. SHAP values measure the marginal importance of each variable in the model's global context, meaning that criteria with low scores (such as *Instrumental Value* = 1) can still have decisive influence.

4. Feasibility of a Multi-Class Screening System. The *OPTIMA Health Priority* application effectively and clearly supported triage in both binary and multi-class modes, covering both experimental scenarios. The graphical interface ensured that results were both actionable and interpretable, promoting trust and encouraging adoption in real clinical settings.

Across all questions, the principle of Reproducibility and Transparency was maintained, with an end to end pipeline from synthetic data generation to explainable inference being openly documented and replicable. The outcomes confirm that it is possible to build an ethically aligned, interpretable, and technically robust triage system even in contexts of data scarcity and high ethical complexity.

5.4 Final Considerations

The *OPTIMA Health Priority* system represents a robust and innovative approach to patient triage in resource scarce contexts, combining advanced synthetic data generation techniques, explainable classification models, and ethical principles, while being grounded in public values.

The technique based on *Gaussian Copula* integrated with the SDV proved highly effective in generating plausible and balanced synthetic data with strong privacy guarantees, addressing the scarcity of real data and mitigating critical biases. For classification, the XGBoost model, combined with the SHAP methodology, delivered top tier performance with high levels of accuracy and explainability. This combination was particularly relevant for clinical adoption, offering real-time, transparent, and auditable decisions. The developed triage system clearly and intelligibly exposes each decision and its justification, reinforcing principles of accountability and institutional trust. Moreover, fairness and accountability criteria were embedded from dataset construction through to model development, ensuring reduced discrimination and promoting equity.

In addition, a complete and well documented pipeline from data generation to explainable inference was made available, ensuring reproducibility and transparency throughout the process. By combining high quality synthetic data, explainable models, and ethical sustainability principles, *OPTIMA Health Priority* offers a comprehensive, transparent, and patient centered solution to support critical decision making in resource constrained scenarios.

5.5 Future Work

To build upon and extend the system's benefits, several directions for future research are proposed:

- **Clinical validation and feedback**: Test the system in real or simulated hospital environments to gather direct feedback from healthcare professionals, enabling model and interface adjustments based on user experience.
- **Integration of multi-modal data**: Incorporate additional data sources such as complex Electronic Health Records, medical imaging, or real time physiological signals, while respecting privacy and interpretability principles.
- **Robustness evaluation**: Subject the system to various stress scenarios such as extreme demand peaks or sudden resource shortages to assess its resilience and adaptability to unexpected situations.
- **Criteria personalization**: Enable health managers or ethics committees to dynamically adjust triage criteria weights according to local policies or specific population needs, enhancing flexibility and system relevance.

Limitations and considerations Despite the promising results, the current work presents limitations that need to be addressed in future studies. The model has been trained and validated primarily on synthetic and historical datasets, which may not fully capture the variability of real world patient populations. Certain ethical or clinical nuances may be underrepresented in the data, potentially affecting decision fairness in extreme or atypical scenarios. Additionally, the current explainability analysis, while robust, relies on SHAP values derived from a single model, which may not reflect alternative modeling approaches or feature interactions not captured by the training process. Finally, the system has not yet been tested under live operational conditions, leaving open questions regarding usability, real time performance, and integration with existing hospital workflows. Addressing these limitations through iterative testing, broader datasets, and real world deployment will be crucial for achieving safe, reliable, and ethically sound clinical support.

References

[1] Rawan S. Abdulsadig and Esther Villegas. A comparative study in class imbalance mitigation when working with healthcare data. *Front. Digit. Health*, 2024. doi: https://doi.org/10.3389/fdgth.2024.1377165.

[2] Masoud Abedi, Lars Hempel, Sina Sadeghi, and Toralf Kirsten. Gan-based approaches for generating structured data in the medical domain. *Applied Sciences*, 12(14):7075, 2022. doi: 10.3390/app12147075.

[3] Razan Alkhanbouli, Hour Matar Abdulla Almadhaani, Farah Alhosani, and Mecit Can Emre Simsekler. The role of explainable artificial intelligence in disease prediction: a systematic literature review and future research directions. *BMC Medical Informatics and Decision Making*, 25(1):45, 2025. doi: 10.1186/s12911-025-02944-6.

[4] Matthew C. Altman. A consequentialist argument for considering age in triage decisions during the coronavirus pandemic. *Bioethics*, 35(5):421–428, 2021. doi: 10.1111/bioe.12864.

[5] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I. Madai, and *Precise4Q consortium*. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Medical Informatics and Decision Making*, 20:310, 2020. doi: 10.1186/s12911-020-01332-6.

[6] Elizabeth S. Anderson. What is the point of equality? *Ethics*, 109:287–337, 1999. doi: 10.1086/233897.

[7] Lars Lien Ankile, Morgan Feet Heggland, and Kjartan Krange. Deep convolutional neural networks: A survey of the foundations, selected improvements, and some current applications. *arXiv preprint arXiv:2011.12960*, 2020. doi: arXiv:2011.12960v1.

[8] Manikandan Ayyanar, Selvaprabu Jegananathan, Saravanan Parthasarathy, Vaishnavi Jayaraman, and Arun Raj Lakshminarayanan. Predicting the cardiac diseases using selectkbest method equipped light gradient boosting machine. In *2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2022. doi: 10.1109/ICOEI53556.2022.9777224.

[9] Gérard Biau and Erwan Scornet. A random forest guided tour. *Test*, 25(2):197–227, 2016. doi: 10.1007/s11749-016-0481-7.

[10] Leo Breiman. Random forests. *Machine Learning*, 45(1):5–32, 2001. doi: 10.1023/A:1010933404324.

[11] Sarah Carter. Ventilator rationing guidelines are discriminatory. https://www.bloomberg.com/view/articles/2020-04-10/

coronavirus-ventilator-rationing-guidelines-are-discriminatory?embedded-checkout=true, 2020. Accessed: April 27, 2025.

[12] Girish Chandrashekar and Ferat Sahin. A survey on feature selection methods. *Computers & Electrical Engineering*, 40(1):16–28, 2014. doi: 10.1016/j.compeleceng.2013.11.024.

[13] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '16)*, pages 785–794. ACM, 2016. doi: 10.1145/2939672.2939785.

[14] Roberta Costa, Monica Motta Lino, Ana Izabel Jatobá de Souza, Elisiane Lorenzini, Gisele Cristina Manfrini Fernandes, Laura Cavalcanti de Farias Brehmer, Mara Ambrosina de Oliveira Vargas, Melissa Orlandi Honório Locks, and Natália Gonçalves. Nursing teaching in covid-19 times: how to reinvent it in this context? *Texto & Contexto – Enfermagem*, 29(e20200202):1–3, 2020. doi: 10.1590/1980-265X-TCE-2020-0002-0002.

[15] Karla Rona da Silva, Fernanda Gonçalves de Souza, Fátima Ferreira Roquete, Shirlei Moreira da Costa Faria, Bruno César Ferreira Peixoto, and Adriane Vieira. Allocation of resources for health care in covid-19 pandemic times: integrative review. *Revista Brasileira de Enfermagem*, 73(suppl 2), 2020. doi: 10.1590/0034-7167-2020-0244.

[16] Francisco de Arriba-Perez, Silvia Garcia-Mendez, Fatima Leal, Benedita Malheiro, and Juan Carlos Burguillo-Rial. Balancing plug-in for stream-based classification. In Álvaro Rocha, Hojjat Adeli, Gintautas Dzemyda, Fernando Moreira, and Valentina Colla, editors, *Information Systems and Technologies: WorldCIST 2023, Volume 1*, pages 65–74. Springer Cham, 2024. doi: 10.1007/978-3-031-45642-8_6.

[17] Jamie Dettmer. Italians demand compensation for coronavirus deaths. https://www.voanews.com/covid-19-pandemic/italians-demand-compensation-coronavirus-deaths, 2020. Accessed: April 10, 2025.

[18] Markus Endres, Asha Mannarapotta Venugopal, and Tung Son Tran. Synthetic data generation: A comparative study. In *IDEAS '22: Proceedings of the 26th International Database Engineered Applications Symposium*, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450397094. doi: 10.1145/3548785.3548793.

[19] Maria Goddard, Katharina Hauck, Alexander Preker, and Peter Smith. Priority setting in health – a political economy perspective. *Health Economics, Policy and Law*, 1:79–90, 2006. doi: 10.1017/S1744133105001040.

[20] Ricardo Gonzalez, Ashirbani Saha, Clinton J.V. Campbell, Peyman Nejat, Cynthia Lokker, and Andrew P. Norgan. Seeing the random forest through the decision trees. supporting learning health systems from histopathology with machine learning models: Challenges and opportunities. *Journal of Pathology Informatics*, 15:100347, 2024. doi: 10.1016/j.jpi.2023.100347.

[21] Sophie Gotowiec and Elizabeth Cantor-Graae. The burden of choice: a qualitative study of healthcare professionals' reactions to ethical challenges in humanitarian crises. *International Journal of Humanitarian Action*, 2(2):1–10, 2017. doi: 10.1186/s41018-017-0019-y.

[22] Haixiang Guo, Yijing Li, Jennifer Shang, Mingyun Gu, Yuanyue Huang, and Bing Gong. Learning from class-imbalanced data: Review of methods and applications. *Expert Systems with Applications*, 73:220–239, 2017. doi: 10.1016/j.eswa.2016.12.035.

[23] Fatih Gurcan and Ahmet Soylu. Learning from imbalanced data: Integration of advanced resampling techniques and machine learning models for enhanced cancer diagnosis and prognosis. *Cancers*, 16(19):3417, 2024. doi: 10.3390/cancers16193417.

[24] Isabelle Guyon and André Elisseeff. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3:1157–1182, 2003. doi: 10.1162/153244303322753616.

[25] Waldemar Hahn, Katharina Schütte, Kristian Schultz, Olaf Wolkenhauer, Martin Sedlmayr, Ulrich Schuler, Martin Eichler, Saptarshi Bej, and Markus Wolfien. Contribution of synthetic data generation towards an improved patient stratification in palliative care. *Journal of Personalized Medicine*, 12(8):1278, 2022. doi: 10.3390/jpm12081278.

[26] Mikel Hernández, Gorka Epelde, Ane Alberdi, Rodrigo Cilla, and Debbie Rankin. Synthetic data generation for tabular health records: A systematic review. *Neurocomputing*, 510:235–257, 2022. doi: 10.1016/j.neucom.2022.04.053.

[27] Andreas Holzinger, Chris Biemann, Constantinos S. Pattichis, and Douglas B. Kell. What do we need to build explainable ai systems for the medical domain? *arXiv preprint arXiv:1712.09923*, 2017. doi: 10.48550/arXiv.1712.09923.

[28] W. Nicholson Price II and I. Glenn Cohen. Privacy in the age of medical big data. *Nature Medicine*, 25(1):37–43, 2019. doi: 10.1038/s41591-018-0272-7.

[29] Aryan Jadon and Shashank Kumar. Leveraging generative ai models for synthetic data generation in healthcare: Balancing research and privacy. *2023 International Conference on Smart Applications, Communications and Networking*, 2023. doi: 10.48550/arXiv.2305.05247.

[30] James Jordon, Jinsung Yoon, and Mihaela van der Schaar. Measuring the quality of synthetic data for use in competitions. In *Workshop on Machine Learning for Medicine and Healthcare*, 2018. doi: arXiv:1806.11345v1.

[31] James Jordon, Łukasz Szpruch, Mirko Bottarelli, Giovanni Cherubin, Samuel N. Cohen, and Adrian Weller. Synthetic data – what, why and how? *Royal Society Open Science*, 2022. doi: 10.48550/arXiv.2205.03257.

[32] Pascal Jutras-Dubé, Mohammad B. Al-Khasawneh, Zhichao Yang, Javier Bas, Fabian Bastin, and Cinzia Cirillo. Copula-based transferable models for synthetic population generation. *Transportation Research Part C: Emerging Technologies*, 169:104830, 2024. doi: 10.1016/j.trc.2024.104830.

[33] Nicklas Jävergård, Rainey Lyons, Adrian Muntean, and Jonas Forsman. Tunable correlation retention: A statistical method for generating synthetic data. *arXiv preprint arXiv:2403.01471*, 2024. doi: 10.48550/arXiv.2403.01471.

[34] Susanne Jöbges and Nikola Biller-Andorno. Ethics guidelines on covid-19 triage—an emerging international consensus. *Critical Care*, 24(1):201, 2020. doi: 10.1186/s13054-020-02927-1.

[35] Susanne Jöbges, Rasita Vinay, Valerie A. Luyckx, and Nikola Biller-Andorno. Recommendations on covid-19 triage: International comparison and ethical analysis. *Bioethics*, 35(3): 252–260, 2021. doi: 10.1111/bioe.12805.

[36] Aravind Kumar Kalusivalingam, Amit Sharma, Neha Patel, and Vikram Singh. Leveraging shap and lime for enhanced explainability in ai-driven diagnostic systems. *International Journal of AI and ML*, 2021. Accessed: March 2, 2025.

[37] Sanket Kamthe and Samuel Assefa. Copula flows for synthetic data generation. *Preprint, Imperial College London and JP Morgan Chase*, 2023. doi: arXiv:2101.00598v1.

[38] V. Kasemsup, J. Schommer, R. Cline, and R. Hadsall. Citizen's preferences regarding principles to guide health care allocation decisions in thailand. *Value in Health*, 11(7):1194–1202, 2008. doi: 10.1111/j.1524-4733.2008.00321.x.

[39] Hadi A. Khorshidi and Uwe Aickelin. A synthetic over-sampling method with minority and majority classes for imbalance problems (somm). *arXiv preprint arXiv:2011.04170*, 2020. doi: 10.48550/arXiv.2011.04170.

[40] Miron Kursa, Aleksander Jankowski, and Witold Rudnicki. Boruta - a system for feature selection. *Fundam. Inform.*, 101:271–285, 01 2010. doi: 10.3233/FI-2010-288.

[41] Kara Liu and Russ B. Altman. Conditional generative models for synthetic tabular data: Applications for precision medicine and diverse representations. *Annual Review of Biomedical Data Science*, 2025. doi: 10.1146/annurev-biodatasci-103123-094844.

[42] Bette Loef, Albert Wong, Nicole A. H. Janssen, Maciek Strak, Jurriaan Hoekstra, H. Susan J. Picavet, H. C. Hendriek Boshuizen, W. M. Monique Verschuren, and Gerrie-Cor M. Herber. Using random forest to identify longitudinal predictors of health in a 30-year cohort study. *Scientific Reports*, 12(1):2922, 2022. doi: 10.1038/s41598-022-14632-w.

[43] Hui Wen Loh, Chui Ping Ooi, Silvia Seoni, Prabal Datta Barua, Filippo Molinari, and U. Rajendra Acharya. Application of explainable artificial intelligence for healthcare: A systematic review of the last decade (2011–2022). *Computer Methods and Programs in Biomedicine*, 226:107161, 2022. doi: 10.1016/j.cmpb.2022.107161.

[44] O. A. Montesinos López et al. Fundamentals of artificial neural networks and deep learning. In O. A. Montesinos López et al., editor, *Multivariate Statistical Machine Learning Methods for Genomic Prediction*, pages 201–222. Springer, 2022. doi: 10.1007/978-3-030-89010-0_10.

[45] Daniel McDuff, Theodore Curran, and Achuta Kadambi. Synthetic data in healthcare. *arXiv preprint arXiv:2304.03243*, 2023. doi: 10.48550/arXiv.2304.03243.

[46] David Meyer and Thomas Nagler. Synthia: multidimensional synthetic data generation in python. *Journal of Open Source Software*, 6(65):2863, 2021. doi: 10.21105/joss.02863.

[47] Riccardo Miotto, Li Li, Brian A. Kidd, and Joel T. Dudley. Deep patient: An unsupervised representation to predict the future of patients from the electronic health records. *Scientific Reports*, 6:26094, 2016. doi: 10.1038/srep26094.

[48] Rashid Nasimov, Nigorakhon Nasimova, Sanjar Mirzakhalilov, Gul Tokdemir, Mohammad Rizwan, Akmalbek Abdusalomov, and Young-Im Cho. Gan-based novel approach

for generating synthetic medical tabular data. *Bioengineering*, 11(12):1288, 2024. doi: 10.3390/bioengineering11121288.

[49] Roger B. Nelsen. *An Introduction to Copulas*. Springer, New York, 2nd edition, 2006. ISBN 978-0-387-92258-1.

[50] Shamim Nemati, Andre Holder, Fereshteh Razmi, Matthew D. Stanley, Gari D. Clifford, and Timothy G. Buchman. An interpretable machine learning model for accurate prediction of sepsis in the icu. *Critical Care Medicine*, 46(4):547–553, 2018. doi: 10.1097/CCM.0000000000002936.

[51] Aristidis K. Nikoloulopoulos. Efficient and feasible inference for high-dimensional normal copula regression models. *arXiv preprint arXiv:2203.04619*, 2022. doi: 10.48550/arXiv.2203.04619.

[52] Aristidis K Nikoloulopoulos and Emmanouil Mentzakis. A copula-based model for multivariate ordinal panel data: application to well-being composition. *arXiv preprint arXiv:1604.05643*, 2016. doi: 10.48550/arXiv.1604.05643.

[53] Feyza Yıldırım Okay, Mustafa Yıldırım, and Suat Özdemir. Interpretable machine learning: A case study of healthcare. *2021 International Symposium on Networks, Computers and Communications (ISNCC)*, 2021. doi: 10.1109/ISNCC52172.2021.9615727.

[54] Emmanouil Panagiotou, Arjun Roy, and Eirini Ntoutsi. Survey on synthetic tabular data generation for class imbalance and fairness. *arXiv preprint arXiv:2409.05215*, 2024. doi: 10.48550/arXiv.2409.05215.

[55] Mrutyunjaya Panda and Soumya Ranjan Mahanta. Explainable artificial intelligence for healthcare applications using random forest classifier with lime and shap. *CRC Press*, 2023. doi: arXiv:2311.05665v1.

[56] Neha Patki, Rajarshi Wedge, and Kalyan Veeramachaneni. The synthetic data vault. In *IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, pages 1–10, 2016. doi: 10.1109/DSAA.2016.49.

[57] Vasileios C. Pezoulas, Dimitrios I. Zaridis, Eugenia Mylona, Christos Androutsos, Kosmas Apostolidis, Nikolaos S. Tachos, and Dimitrios I. Fotiadis. Synthetic data generation methods in healthcare: A review on open-source tools and methods. *Computational and Structural Biotechnology Journal*, 23:2892–2910, 2024. doi: 10.1016/j.csbj.2024.07.005.

[58] M. Pinho and A. Araújo. How to fairly allocate scarce medical resources? controversial preferences of healthcare professionals with different personal characteristics. *Health Economics, Policy and Law*, 17(4):398–415, 2022. doi: 10.1017/S1744133121000190.

[59] M. Pinho and P. Veiga. Attitudes of health professionals concerning bedside rationing criteria: a survey from portugal. *Health Economics, Policy and Law*, 15(1):113–127, 2020. doi: 10.1017/S1744133118000403.

[60] David M. W. Powers. Evaluation: From precision, recall and f-measure to roc, informedness, markedness & correlation. *Journal of Machine Learning Technologies*, 2(1):37–63, 2011. doi: https://doi.org/10.48550/arXiv.2010.16061.

[61] Alvin Rajkomar, Jeffrey Dean, and Isaac S. Kohane. Machine learning in medicine. *The New England Journal of Medicine*, 380(14):1347–1358, 2019. doi: 10.1056/NEJMra1814259.

[62] Hanshu Rao, Weisi Liu, Haohan Wang, I-Chan Huang, Zhe He, and Xiaolei Huang. A scoping review of synthetic data generation for biomedical research and applications. *Journal of Biomedical Informatics*, 2025. doi: https://doi.org/10.48550/arXiv.2506.16594.

[63] Sepideh Rezvani and Xuejun Wang. A broad review on class imbalance learning techniques. *Applied Soft Computing*, 143, 2023. doi: 10.1016/j.asoc.2023.110415.

[64] Marco Tulio Ribeiro, Sameer Singh, and Carlos 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*, pages 1135–1144. ACM, 2016. doi: 10.1145/2939672.2939778.

[65] Maura A. Ryan. Tragic choices, revisited: Covid-19 and the hidden ethics of rationing. *Christian Bioethics*, 28(1):72–88, 2022. doi: 10.1093/cb/cbab019.

[66] Mabrouka Salmi, Dalia Atif, Diego Oliva, Ajith Abraham, and Sebastian Ventura. Handling imbalanced medical datasets: review of a decade of research. *Artificial Intelligence Review*, 57(273), 2024. doi: 10.1007/s10462-024-10884-2.

[67] Joseph Shapiro. People with disabilities say rationing care policies violate civil rights. https://www.npr.org/2020/03/23/820398531/people-with-disabilities-say-rationing-care-policies-violate-civil-rights, 2020. Accessed: April 22, 2025.

[68] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information Processing & Management an International Journal*, 45(4): 427–437, 2009. doi: https://doi.org/10.1016/j.ipm.2009.03.002.

[69] Giovanna Gonzalez Storto, Sany Tomomi de Almeida Rocha Arita, Mayara Suzano dos Santos, and João Roberto Oba. Bioética e a alocação de recursos na pandemia de covid-19. *Revista de Bioética*, 29(4):825–831, 2021. doi: 10.1590/1983-80422021294516.

[70] Yi Sun, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Learning vine copula models for synthetic data generation. In *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence (AAAI-19)*, 2019. doi: arXiv:1812.01226v1.

[71] Lucas Theis, Aäron van den Oord, and Matthias Bethge. A note on the evaluation of generative models. *arXiv preprint arXiv:1511.01844*, 2016. doi: arXiv:1511.01844v3.

[72] Edi Tjoa and Chuan Guan. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE Transactions on Neural Networks and Learning Systems*, 32(11):4793–4813, 2021. doi: 10.1109/TNNLS.2020.3027314.

[73] A. Torfi and E.A. Fox. Differentially private synthetic medical data generation using convolutional gans. In *Proceedings of the 2020 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 1–7, 2020. doi: https://doi.org/10.48550/arXiv.2012.11774.

[74] Ineke van der Wulp. *Reliability and validity of emergency department triage systems*. PhD thesis, University of Utrecht, Faculty of Medicine, Utrecht, Netherlands, 2010. ISBN: 978-90-393-5276-2.

[75] Maria Elisa Villas-Bôas. Justiça distributiva, critérios de alocação de recursos escassos em saúde e suas críticas. *Revista Redbioética/UNESCO*, 1(2):73–84, 2010. ISSN 2077-9445.

[76] Yijun Wang and Duen Horng Chau. WebSHAP: Towards explaining any machine learning models anywhere. *WWW '23 Companion*, 2023. doi: https://doi.org/10.48550/arXiv.2303.09545.

[77] Oskar Wysocki, Jessica Katharine Davies, Markel Vigo, Anne Caroline Armstrong, Dónal Landers, Rebecca Lee, and André Freitas. Assessing the communication gap between ai models and healthcare professionals: explainability, utility and trust in ai-driven clinical decision-making. *Artificial Intelligence*, 316:103839, 2023. doi: 10.1016/j.artint.2022.103839.

[78] Lei Xu, Maria Skoularidou, Alfredo Cuesta-Infante, and Kalyan Veeramachaneni. Modeling tabular data using conditional gan. *Advances in Neural Information Processing Systems (NeurIPS) 2019*, 2019. doi: 10.48550/arXiv.1907.00503.

[79] Chao Yan, Ziqi Zhang, Steve Nyemba, and Zhuohang Li. Generating synthetic electronic health record data using generative adversarial networks: Tutorial. *JMIR AI*, 3:e52615, 2024. doi: 10.2196/52615.