UNIVERSITÀ DEGLI STUDI DELL'AQUILA

Department of Information Engineering, Computer Science and Mathematics

Master Degree in Applied Data Science

AI-Assisted Medical Triage for Resource-Limited Settings: Developing a Prototype for Nigeria

Thesis Advisor

Prof. Pasquale Caianiello

Candidate

Hassanat Oluwatobi Awodipe

280020

Academic Year 2024-2025

Dedication

To Peace and Segun. And Lucia and Luca. And of course, baby Agata😊. Without you guys, my time in L'Aquila would have been terrible, and I would have abandoned my studies at some point.

Acknowledgements

It feels quite surreal to have made it to this stage; nothing has been more challenging. I couldn’t have done this without the support I got from friends, family, professors, classmates and even strangers.

Thanks to my parents, siblings and my in-laws.

Thanks to my family members who prayed for me or advised me: Dr Adebambo, Aunt Yinka, Abolore, Khadijah, Azeezah, Auntimi, mummy Ibadan,

Thanks to my friends who studied with me: Tamsir, Abdullateef, Tabatha, Francesca, Leonardo, Lamin, Esther, Lamin, Krishna, Michael,

Thanks to my friends who asked about my studies: Agnes, Sodiq, Jai, Halima, Faheedat, Peace, Linda, Temi, Bola, Fatima, Hanif and Kafilah, Ruqoyyah and AbdulAfeez, Tesleemah and Tijani, Abdulsalam and Memunah, Muinat and Olatunji, Khadijah, Oluwagajuwa,

Thanks to my husband, Idris and my daughter, Oreofe, for your unconditional love.

Thanks to my thesis supervisor, Prof. Caianiello for his guidance and time.

Finally, thank you, God, for it is by your infinite wisdom that you specially placed these people in my life. *AlhamduliLlah*!

Abstract

Background: Emergency departments in Nigeria face persistent challenges such as overcrowding, limited staff, and poor infrastructure, making efficient triage essential for improving patient outcomes.

Aim: This study explores the potential of artificial intelligence (AI) to support triage decision-making in resource-limited settings by developing an AI-assisted triage prototype.

Method: A synthetic dataset of 600 patient records was generated to reflect clinically plausible presentations, and a Random Forest classifier was trained and evaluated. Global model behaviour was examined using feature-importance analysis, while SHAP decision plots provided local interpretability for individual predictions.

Result: To demonstrate practical usability, a lightweight web-based prototype was implemented using Streamlit, enabling real-time triage predictions and explanation visualisations. System testing showed stable performance and user-friendly interaction.

Conclusion: The results highlight the feasibility of deploying interpretable, low-compute AI tools in Nigerian healthcare settings and provide a foundation for future integration using real clinical data.

Keywords: Artificial Intelligence, Medical Triage, Resource-Limited Settings, Nigeria, Random Forest, Explainable AI (XAI), SHAP (Shapley Additive Explanations), Feature Importance, Synthetic Data, Streamlit Prototype, Emergency Care.

Contents

Dedication ii
Acknowledgements iii
Abstract iv
Contents vi
List of Figures viii
List of Tables ix
1 Introduction 1
1.1 Background & Motivation 1
1.2 Problem Statement and Objectives 4
1.3 Scope & Limitation 5
1.4 Thesis Structure 5
2 Literature Review 6
2.1 Evolution of Triage Systems 6
2.2 AI in Healthcare for RLS 12
2.3 Issues with AI Adoption in Healthcare for RLS 14
2.4 Current State of Nigeria's Digital Health 18
3 Methodology 22
3.1 Technical Approach 22
3.1.1 Overview 22
3.1.2 Synthetic Data Design 23
3.1.3 Data Generation Process 24
3.1.4 Data Preprocessing 26
3.1.5 Algorithm Selection 26
3.1.6 Model Training 28
3.1.7 Model Evaluation 28
3.1.8 Model Explainability 31
3.2 Practical Approach 32
3.2.1 Overview 32
3.2.2 Prototype Architecture 32
3.2.3 User Interaction 34
4 Results & Discussion 37

4.1 Exploratory Data Analysis (EDA) Results 37
4.2 Model Comparison 40
4.3 Model Training Outcomes 42
4.4 Explainability Analysis 45
4.4.1 Feature Importance Analysis 45
4.4.2 SHAP-Based Model Interpretation 46
4.5 Implementation Results 50
4.5.1 Deployment Considerations 56
5 Conclusion 58
5.1 Limitations and Future Work 59
6 Bibliography 61

List of Figures

Figure 2.1: MTS decision-flow diagram illustrating the sequential assessment of clinical urgency. 7
Figure 2.2: ATS categories, response times, and associated clinical indicators 8
Figure 2.3: ESI decision algorithm illustrating the stepwise triage process used to categorize patients. 9
Figure 2.4: The framework of the Knowledge-based Uncertainty-inspired Triage System (KUTS). 10
Figure 2.5: Overview of AI-based Healthcare Applications for RLS 16
Figure 2.6: Recommended Pyramid for the implementation of AI in Healthcare for RLS 20
Figure 3.1: The decision-making process for the IITT 23
Figure 3.2: Working mechanism of the Random Forest model. 27
Figure 3.3: Confusion matrix 30
Figure 3.4: ROC - AUC 31
Figure 3.5: Illustration of the system architecture. 34
Figure 3.6: Workflow of the Triage Prototype 35
Figure 4.1: Distribution of the numeric variables across triage categories. 39
Figure 4.2: Distribution of categorical variables across triage categories 39
Figure 4.3: ROC Curve showing Trade-off between TP and FP 44
Figure 4.4: Normalised feature importances derived from the Random Forest classifier 45
Figure 4.5: SHAP decision plot showing the contribution of variables to the Urgent triage classification 47
Figure 4.6: SHAP Waterfall Plot for Predicted Class – Urgent. 48
Figure 4.7: SHAP waterfall plot showing the model's reasoning for a misclassified observation 49
Figure 4.8: SHAP waterfall plots illustrating model reasoning for a misclassified case 50
Figure 4.9: Landing page of the AI-Assisted Medical Triage web application 52
Figure 4.10: Input Validation 53
Figure 4.11: Prediction Outcome 54
Figure 4.12: Patient Summary and Recent Classifications. 55
Figure 4.13: Generated Summary. 56

List of Tables

Table 4.1: Descriptive statistics for the categorical data. 37

Table 4.2: Descriptive Statistics for the Numerical Data 38

Table 4.3: Performance evaluation of the MLP classifier on the test dataset 40

Table 4.4: Performance evaluation of the XGB classifier on the testing set.**Error! Bookmark not defined.**

Table 4.5: Performance evaluation of the RF classifier on the testing set. . **Error! Bookmark not defined.**

Table 4.6: Evaluation of the tuned RF classifier on the testing dataset 42

Table 4.7: Confusion matrix of the tuned model 43

Table 4.4: Performance evaluation of the XGB classifier on the testing set.**Error! Bookmark not defined.**

Table 4.5: Performance evaluation of the RF classifier on the testing set. . **Error! Bookmark not defined.**

1 Introduction

1.1 Background & Motivation

Nigeria, Africa's most populous nation, with over 215 million inhabitants [1] possesses one of the continent's largest and most complex healthcare systems [2]. The system operates through a three-tier structure comprising primary, secondary, and tertiary healthcare services. The Primary Healthcare (PHC) level is community-based and provides essential services such as immunisation, maternal and child health, and disease prevention. The Secondary Healthcare level consists of state and general hospitals that manage more complex medical cases and minor surgical procedures. The Tertiary Healthcare level includes teaching hospitals and specialised centres offering advanced treatments, research, and surgical interventions. In addition to these public institutions, Nigeria's healthcare landscape also includes private hospitals and facilities owned by faith-based organisations (FBOs), all operating under national regulations and standards [3]. As of 2021, the country had an estimated 40,821 health facilities distributed across these tiers [3].

Despite this extensive healthcare network, access to quality care remains highly uneven, with the majority of hospitals concentrated in urban areas, particularly Lagos and other major cities. Rural and underserved regions face persistent challenges in reaching timely and adequate medical services. Moreover, Nigeria faces a critical shortage of healthcare personnel, largely driven by the emigration of skilled professionals seeking better opportunities abroad [2], [4], [5]. The World Health Organisation (WHO) recommends a doctor-to-patient ratio of 1:1,000, whereas Nigeria's ratio stands at approximately 1:2,500, or about 4 doctors per 10,000 people [1]. This severe human resource gap, combined with the uneven distribution of facilities, places immense strain on the healthcare system, particularly on Emergency Departments (EDs), where patient overcrowding and long waiting times are common [6].

In Nigeria, Emergency Medicine (EM) is still evolving as a medical speciality despite its formal introduction in 1998 [7] and the establishment of the first postgraduate training curriculum in 2020 [8], [9]. In many hospitals, the ED is also referred to as the Accident and Emergency Department (AED) [8], [9], [10]. In the latest assessment of emergency care services using the Emergency Care Assessment Tool

(ECAT) - a validated framework developed by the African Federation of Emergency Medicine (AFEM) – it was revealed that Nigeria's emergency care capacity remains suboptimal [8]. The study, which evaluated seven tertiary institutions (one per geopolitical zone) and the National Hospital Abuja, highlighted persistent deficiencies in providing medical interventions for six common, life-threatening conditions. These challenges have been attributed to limited governmental commitment and institutional gaps in implementing Emergency Medicine within postgraduate training programs [11].

Beyond the broader reforms required in Nigeria's Emergency Medicine, a particularly critical area is the adoption and consistent use of standardised triage tools. In a multicountry study involving 192 medical facilities across 139 Low- and Middle-Income Countries (LMICs), including Nigeria, it was found that only about half of the facilities demonstrated any formal triage capability [12]. Triage tools are designed to prioritise patients based on the severity of their condition and the availability of treatment resources [13], [14], [15]. When effectively implemented, these tools have been shown to improve efficiency, accelerate patient treatment, and reduce morbidity and mortality rates [16]. Triage processes are generally expected to last between 2-5mins [17] and yet be thoroughly done, which is why only experienced medical practitioners are expected to triage [13]. Triage can be carried out anywhere, including the scene of an accident, outpatient queue, emergency room and at multiple stages of patient management to ensure continuous reassessment and appropriate care prioritisation [14].

In most Nigerian hospitals, patients are attended to on a first-come, first-served basis, except in obvious emergency cases such as severe trauma or accidents. This practice highlights the limited use of standardised triage systems across the country. A recent study involving nurses in the emergency departments of four tertiary hospitals in Lagos State, in southwestern Nigeria, revealed that only 30% had ever used a formal triage tool such as the Emergency Severity Index (ESI) or the WHO Interagency Integrated Triage Tool [13]. The majority reported relying primarily on professional experience and subjective judgment to assess patient urgency. Similarly, a study conducted among nurses in three tertiary hospitals in Enugu, southeastern Nigeria, found that 71.4% lacked knowledge of proper triage procedures, and only 28.4% had received any form of emergency care training [10]. The study did not clarify whether these gaps were due to the absence of triage tools in the facilities or the non-utilisation of available ones, underscoring the broader systemic and training deficiencies in emergency care delivery.

Triage tools are widely utilised in emergency departments globally to guide the prioritisation of patients based on clinical urgency[14]. While there is no single universal triage system, existing models are typically adapted to fit local healthcare contexts and resource capacities [16]. Among the most widely adopted systems are the Australasian Triage Scale (ATS), Emergency Severity Index (ESI), Manchester Triage System (MTS), and Canadian Triage and Acuity Scale (CTAS) [10], [18], [19]. In the United States, the

Emergency Severity Index (ESI) is the predominant framework, while the Simple Triage and Rapid Treatment (START) system is also used for mass-casualty scenarios [14]. Australia employs the ATS, which has served as the foundation for several other international adaptations, including the CTAS in Canada, the Chinese Four-Level and Three-District Triage Standard (CHT), and the MTS, commonly used across Europe [14], [19], [20]. Additional region-specific scales include the Korean Triage and Acuity Scale (KTAS) and the Taiwan Triage Acuity Scale (TTAS). With the exception of the CHT, which is based on four categories, these systems typically follow a five-level classification, where Level 1 indicates the need for immediate life-saving intervention, and Level 5 denotes non-urgent conditions that can safely wait for medical attention [18]. The CHT is based on four categories. Category one is for a critically ill patient, while category four is considered non-emergent.

In Africa, few standardised triage scales have been developed for local use. The most prominent among them is the South African Triage Scale (SATS), which was developed and widely adopted in South Africa [21]. SATS is a five-level, colour-coded triage system that has demonstrated validity and reliability across several LMICs, including Rwanda, Afghanistan, Haiti, and Pakistan, and has even been adapted for use in Norway, a High-Income Country (HIC) [16], [21]. However, there is no documented evidence of its use in Nigeria [10], [13], [22]. Another notable system, the Interagency Integrated Triage Tool (IITT), was developed specifically for LMIC contexts by the World Health Organisation (WHO), Médecins Sans Frontières (MSF), and the International Committee of the Red Cross (ICRC). The IITT is a three-level triage tool that combines features from both the SATS and the Emergency Triage Assessment and Treatment (ETAT) framework [23], [24]. Despite its relevance, evidence of its adoption remains limited to Papua New Guinea, Bangladesh, South Sudan, and Vanuatu.

During the COVID-19 pandemic in 2020, Nigeria developed its first digital triage tool called the Wellvis COVID Triage Tool. It was a public-facing online platform designed to assess users' need for COVID testing and to optimize the use of limited healthcare resources [25]. While the majority of users were based in Nigeria, the tool also recorded engagement from Ghana, the United States, the United Kingdom, Canada, and South Africa. The system categorised users into low-, medium-, or high-risk levels of COVID-19 exposure, demonstrating the potential of locally developed digital triage tools for broader health system strengthening.

This thesis is motivated by the need to provide a basic, national triage tool that can be used by first responders in an emergency department or an outpatient waiting room in order to prioritise patients and shorten waiting times. In developing the prototype, specific attention was given to the practical needs of triage staff, including varying levels of technical proficiency, limited internet reliability, and ease of interface navigation. Consequently, the web application was designed to be lightweight, intuitive, and resilient to common operational constraints in clinical settings. This research is happening at a time when the interest and investment in AI-driven solutions in the healthcare sector are starting to grow, both in Nigeria and globally. This study hopes to lay a foundation for the development and adoption of local triage tools across the country.

1.2 Problem Statement and Objectives

Resource-limited settings refer to a condition where the resources required to achieve a goal are limited in availability. In many LMIC regions, medical facilities face challenges such as lack of infrastructure, limited healthcare personnel, and inconsistent access to expert diagnosis, resulting in overcrowded waiting areas, delayed treatment, and elevated morbidity and mortality rates [8], [26].

Current AI triage tools are trained and validated in HICs [27], [28], [29], [30], [31]. Most AI medical tools trained in the same manner do not reflect local epidemiological contexts and may lead to poor performance [4], [5]. Therefore, this study aims to answer the question: **How can AI-assisted triage be studied, adapted, and potentially implemented in Nigeria's healthcare system to improve patient prioritisation, outcomes and efficiency, given local constraints?**

To address this question, the study aims to:

1. Investigate the feasibility of AI-assisted triage in Nigerian healthcare settings, considering local epidemiology, workflow, and resource constraints.
2. Develop a machine learning–based triage model using synthetic data to simulate realistic patient scenarios.
3. Evaluate the model's predictive performance and interpretability using established metrics and eXplainable AI (XAI) techniques
4. Demonstrate a practical application of the model through a lightweight web-based interface that can be used by medical personnel.

While this work represents a preliminary step, it establishes a foundation for further research and eventual integration of AI-assisted triage tools into Nigeria's healthcare system.

1.3 Scope & Limitation

This study focuses on the development and preliminary evaluation of an AI-assisted triage tool tailored to the Nigerian healthcare context, with comparisons to other Low- and Middle-Income Countries (LMICs) where relevant. A state-of-the-art literature review informed the design of the tool, and multiple iterations of development were undertaken to produce an initial prototype suitable for pilot testing in clinical settings. The model specifically simulates patient triage in emergency department (ED) waiting areas, guiding patients either to the emergency ward for urgent cases or to the outpatient ward for non-urgent cases.

The primary limitation of this study is the lack of publicly available, digitised patient data in Nigeria. Most hospitals maintain patient records in paper-based formats, which are manually updated and stored [32]. Efforts to collect primary data through questionnaires with medical personnel were challenging due to administrative and bureaucratic constraints. To address this, the study utilised synthetic data generated to approximate the statistical characteristics of real patient populations, including arrival patterns and triage priorities. While this approach allows preliminary model development and evaluation, it does not fully capture the variability and complexity of real-world patient data, which should be addressed in future research through pilot testing and integration with actual hospital datasets.

1.4 Thesis Structure

- Chapter 2 reviews context-relevant literature under multiple subtopics. It explores global traditional triage models and compares them with AI triage models. It then reviews AI in healthcare for LMICs before delving into the Nigeria-specific context.
- Chapter 3 describes the data used to build the ML model and explains the models used for classification.
- Chapter 4 discusses the results of applying the model to new data. It also reviews the explainability of the model, feature importances and false predictions.
- Chapter 5 provides the conclusion of the thesis and recommends areas for further research.

2 Literature Review

2.1 Evolution of Triage Systems

“Emergency care can broadly be defined as the delivery of health services for conditions that require rapid intervention to avert death or disability (such as shock or respiratory failure), or for which delays of hours can worsen prognosis or render care less effective (such as management of an asthma exacerbation, or suturing of wounds)” [33]. These conditions vary widely, and as many patients may present to the hospital at the same time, in need of emergency care, the process of triage becomes a necessity.

Triage is the initial assessment of patients to determine the necessity, urgency, and type of medical intervention required [34]. It is a fundamental process in EDs, as patients may arrive unpredictably, spanning all ages and presenting with diverse medical, surgical, trauma, or obstetric conditions of varying acuity [16], [19], [33]. In most EDs, the demand for care frequently exceeds available resources, including both personnel (e.g., cardiologists, neurologists) and equipment (e.g., electrocardiograms, radiographic imaging) as well as basic procedural capacity (e.g., laceration repair). The availability and distribution of these resources vary across facilities, and these limitations often guide triage decisions [14]. Ultimately, the goal of triage is to prioritise patients efficiently, ensuring that those in most urgent need receive prompt treatment while optimising the use of limited emergency care resources [14], [19].

Traditional triage systems typically rely on a combination of clinical algorithms, flowcharts, and sometimes scoring systems to guide decision-making. The MTS, developed between 1994 and 1996 by Kevin Mackway-Jones, consists of a series of flowcharts tailored to 52 distinct presenting complaints[14], [20], [35]. During triage, the nurse interviews the patient to determine the chief complaint and selects the corresponding flowchart. Each flowchart contains a hierarchy of discriminators, which are key clinical features associated with the presenting complaint, such as airway compromise, circulatory compromise, pain severity, or level of consciousness. These discriminators are ranked according to urgency, from immediate (red) to non-urgent (blue), and are associated with maximum recommended waiting times; for example, non-urgent cases are expected to be seen within 120 minutes [14]. The MTS is symptom-based and pre-defined, meaning that it includes all possible symptom groups that the system can classify. Figure

2.1 illustrates the hierarchy of discriminators that is applied once a presenting complaint has been selected [36] .

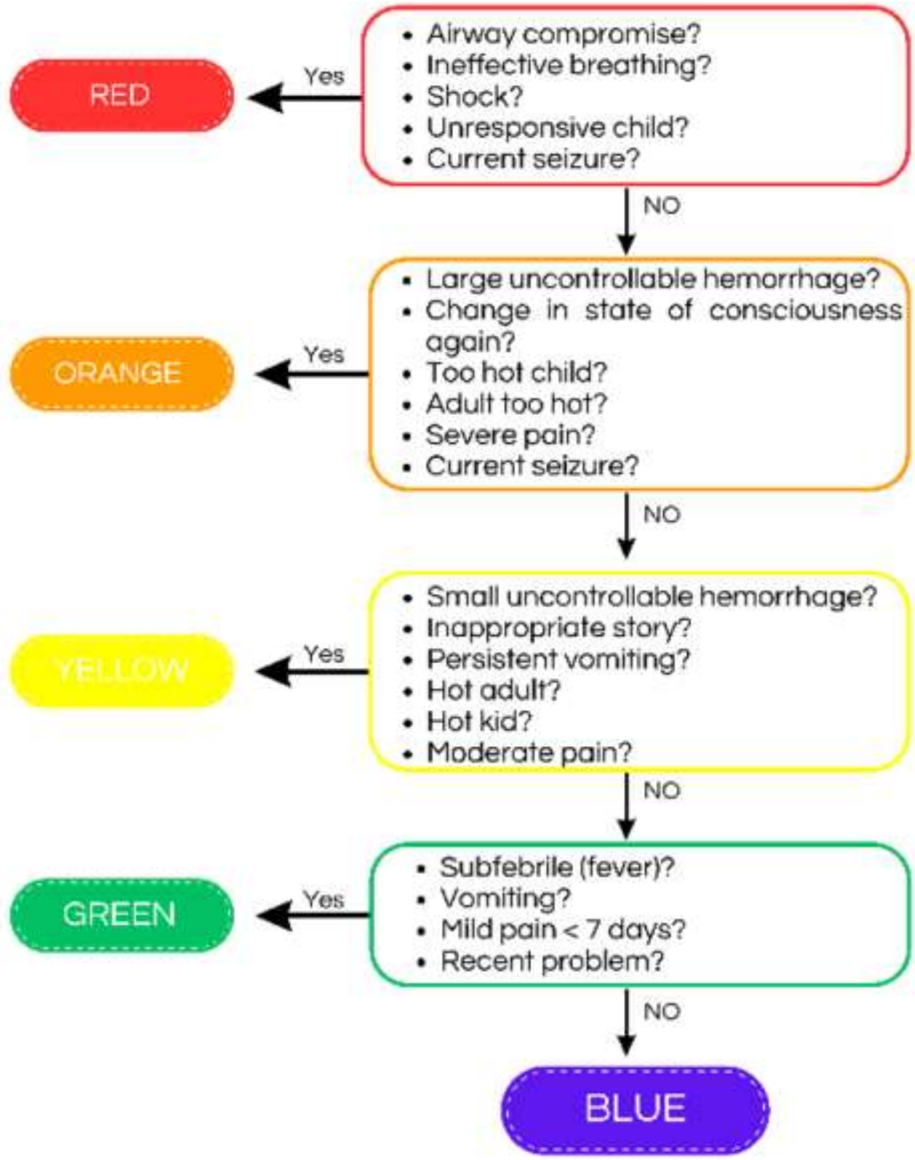


Figure 2.1: MTS decision-flow diagram illustrating the sequential assessment of clinical urgency.

The ATS was initially introduced as the National Triage Scale (NTS) in 1994 and, following multiple reviews and validation studies, was formally launched as the ATS in 2000 [14], [20]. The system assigns a triage classification based on the patient's presenting problem, general appearance, and key physiological findings [14]. The ATS incorporates 79 clinical descriptors, which triage nurses match against the patient's condition to determine the recommended maximum waiting time before physician assessment. The waiting times in the ATS are equivalent to those used in the MTS. Figure 2.2 illustrates how patient clinical indicators are used to justify the assigned triage category [14]. The figure outlines the five ATS acuity levels (Categories 1–5), specifying the required maximum waiting time for medical assessment, a concise description of the clinical urgency associated with each category, and key physiological or symptomatic indicators guiding category assignment.

Australasian Triage Scale Category	Response Time	Category Description	Clinical Indicators
Category 1 (RED)	Seen Immediately	Life Threatening Conditions	Cardiac/Respiratory Arrest Immediate risk of airway, respiratory rate < 10/min, Extreme Respiratory Distress. BP less than 80 in adult. Severe shock in child/infant GCS scale less than 9 Prolonged seizure IV overdose Severe behavioral disorder
Category 2 (ORANGE)	Seen within 10 minutes	Imminently life threatening, time sensitive treatment needed, or Severe pain.	Airway risk (stridor) Circulatory Compromise (HR less than 50 or greater than 150, Hypotension, severe blood loss, poor perfusion). Chest pain likely cardiac related Suspected sepsis, Febrile Neutropenia, Fever with lethargy Acute Stroke GCS less than 13 Suspected Testicular Torsion High Risk History (toxic ingestion, venomous bite, pain suggesting PE, AAA, ectopic pregnancy.
Category 3 (GREEN)	Seen within 30 minutes	Potentially life threatening, situational urgency, or severe pain	Severe Hypertension, Moderate blood loss Moderate Shortness of breath Vomiting Dehydration Seizure (post ictal), Head Injury with LOC (now alert) Physiologically stable suspected sepsis Severe pain Limb injury consisting of limb deformity or severe laceration, altered sensation, absent pulse. Potential child abuse Behavioral/Psychiatric patient very distressed, risk of self-harm, potentially aggressive.
Category 4 (BLUE)	Seen within 60 minutes	Potentially serious condition, situational urgency or complex case	Mild Hemorrhage Foreign Body Aspiration without respiratory distress Chest injury without rib pain or respiratory distress Minor head injury without LOC Moderate pain Vomiting or diarrhea without dehydration Inflammation or foreign body in eye without vision changes Minor limb trauma (ankle sprain, fracture, uncomplicated laceration with normal vital signs) Swollen, erythematous joint Semi Urgent mental health problems with no immediate risk to personnel.
Category 5 (white)	Seen within 120 minutes	Less urgent or Clinical-Administrative problems	Minimal pain with no risk factors Low risk history Minor symptoms of illness Minor symptoms of low risk condition Abrasions or minor laceration Scheduled revisit Immunizations Patient with chronic psychiatric symptoms in social crisis.

Figure 2.2: ATS categories, response times, and associated clinical indicators

Another widely used triage tool is the ESI. Also, 5-level like the ATS and MTS, the ESI was developed by Richard Wuerz, Judith R. Brill and David Eital, and first implemented in 1999 [20], [35]. Unlike ATS and MTS, which rely primarily on physiological findings or clinical discriminators, the ESI incorporates both the severity of the patient's condition and the anticipated level of resource utilisation to determine triage assignment [14], [18]. Figure 2.3 below illustrates the decision-making process of the ESI [18]. The flowchart begins with the assessment of immediate life-threatening conditions (ESI Level 1), followed by the evaluation of whether the patient can safely wait and the anticipated number of resources required for care. By assessing resource needs for patients who can safely wait for treatment, the ESI provides a structured approach that balances urgency with operational efficiency.

Although the traditional triage systems discussed above are well-established and validated, they are not without limitations [37]. These systems are prone to under-triage - when a patient is incorrectly assigned a lower urgency level, potentially worsening their condition - and over-triage, where patients are assigned higher urgency levels than necessary, leading to inefficient use of limited resources. For instance, the ESI and SATS have been reported to misclassify approximately 34% [30] and 31.7% [21] of cases, respectively, while the MTS has an estimated 11–25% likelihood of under-triage [38]. The ATS, though often considered more reliable than other five-level triage scales, remains inadequate for assessing psychiatric presentations [37]. Moreover, studies have also identified systemic inequities in ESI triage outcomes, with Black, Hispanic, and elderly patients more frequently assigned to lower acuity categories, raising concerns about bias and fairness in clinical decision-making [27].

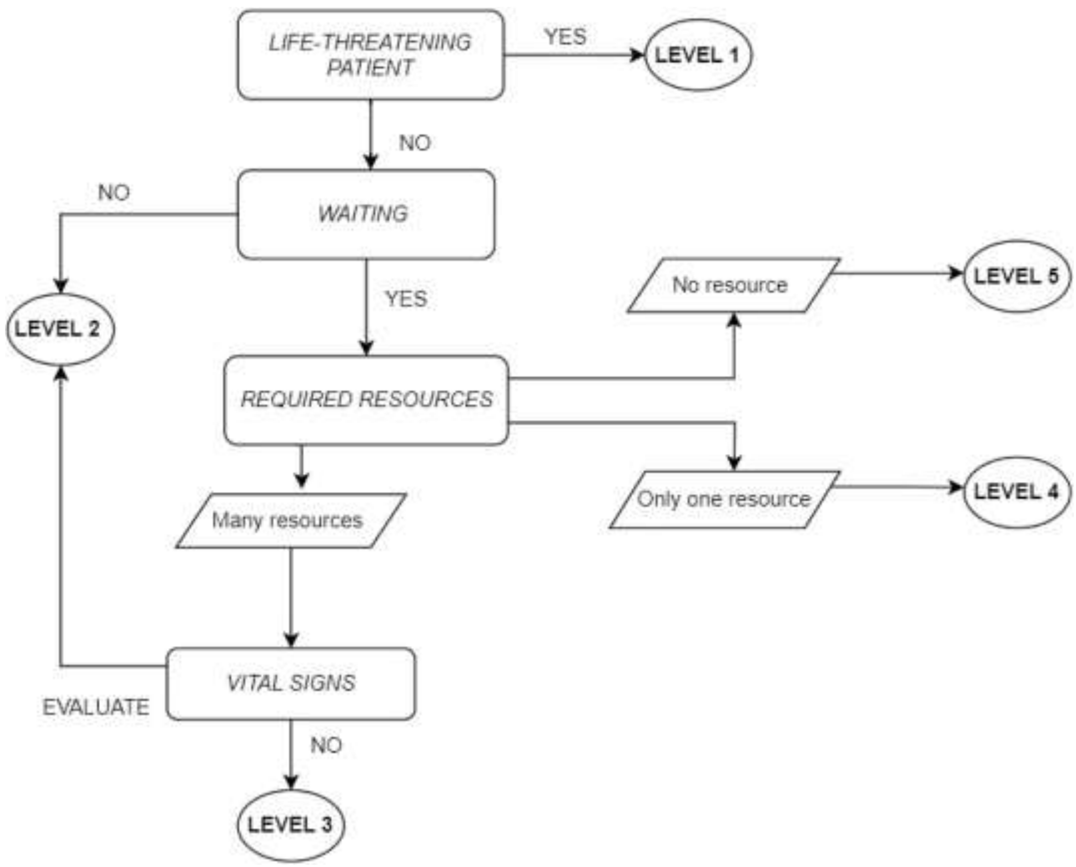


Figure 2.3: ESI decision algorithm illustrating the stepwise triage process used to categorize patients.

As a result of the limitations of traditional triage systems, computational and machine learning approaches have increasingly been explored in emergency triage research[18]. Models such as Decision Trees (DT), Support Vector Machines (SVM), Naïve Bayes, Random Forests, and eXtreme Gradient Boosting (XGBoost) have demonstrated higher accuracy than the Emergency Severity Index (ESI), even when implemented as relatively shallow architectures[30].

A notable advancement in this area is the Knowledge-based Uncertainty-inspired Triage System (KUTS), developed by researchers in Guangzhou, China, to address the issue of mistriage in moderate-acuity cases, specifically those corresponding to ESI level 3 [30]. The KUTS framework integrates patient information and expert knowledge as prompts for a pre-trained language model (PLM), which is subsequently fine-tuned to extract relevant patient features. An uncertainty-based classifier, utilising *Softplus activation* and a *Dirichlet* distribution, is then applied to determine the final triage category.

The model was trained and tested on the MIMIC-IC-ED database, with ESI categories serving as the basis for expert annotation. When compared with other AI triage models, including Decision Trees, Multi-Layer Perceptrons (MLP), and Interpretable Energy Prediction for Transformers (IrEne), the KUTS achieved an average improvement of approximately 5% in predictive accuracy. A key strength of this model is its ability to explicitly incorporate expert domain knowledge, thereby enhancing both interpretability and reliability in decision-making. Figure 2.4 illustrates the KUTS model framework.

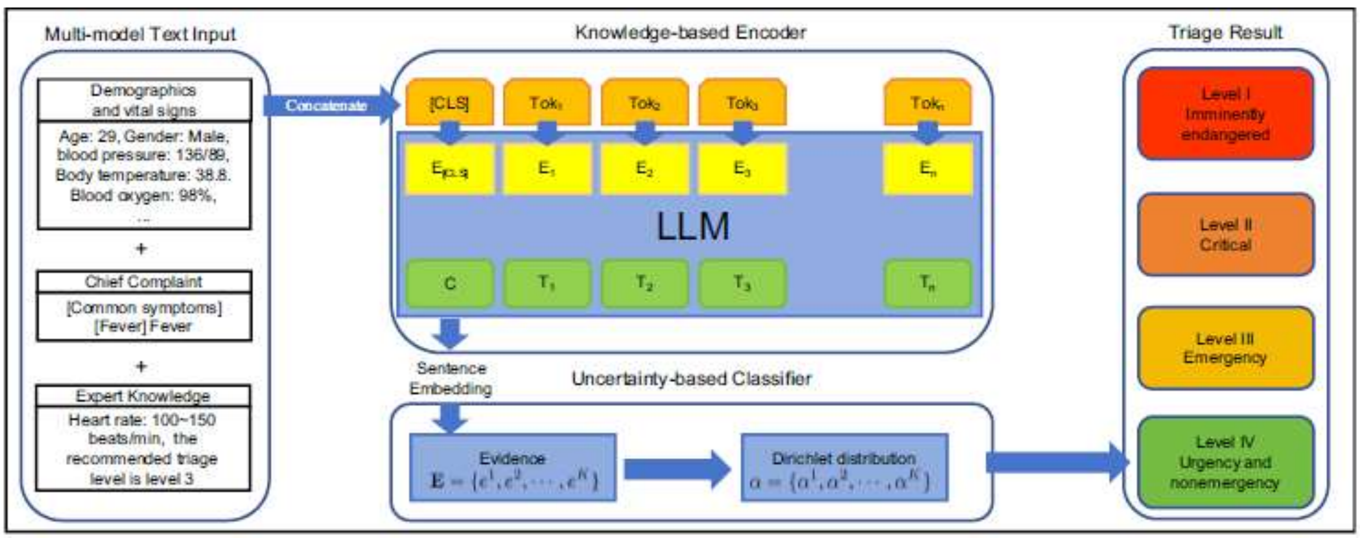


Figure 2.4: The framework of the Knowledge-based Uncertainty-inspired Triage System (KUTS).

At the emergency department of Marburg University Hospital, an artificial intelligence-based symptom assessment application was evaluated against the Manchester Triage System (MTS) to assess its clinical safety and accuracy [38]. The AI app recorded an undertriage rate of 8.9% and an overtriage rate of 57.1%, yet 94.7% of its triage decisions were judged to be clinically safe by physicians. The high incidence of overtriage was attributed to the app's "safety-first" design philosophy, whereby patients are intentionally directed to the ED more often than necessary to minimise the risk of missing a true emergency case.

The application in question, Ada app, was developed in Germany and is certified as a Class IIa medical device under European regulations. It is currently recognised as one of the most extensively validated clinical AI systems worldwide [39]. The app employs an adaptive question flow, supported by a large-scale medical knowledge base and Bayesian network reasoning, to process user-reported symptoms. Upon completion, it generates five possible conditions with associated probability scores, along with an urgency level that advises the user on the appropriate next steps. Although Ada was originally designed for self-assessment by patients in non-emergency settings, the Marburg study explored its potential applicability within emergency department waiting areas, highlighting the growing interest in leveraging AI-driven triage tools to enhance clinical decision-making and patient flow management [38].

The application of Large Language Models (LLMs) in assessing clinical acuity has also been explored in recent studies conducted in the United States. Researchers at the University of California, San Francisco evaluated the performance of OpenAI's GPT-4 and GPT-3.5 Turbo in identifying patients with higher-acuity presentations based solely on clinical history data [40]. A dataset comprising 10,000 matched pairs of ED visits with differing ESI scores was analysed, without incorporating vital signs or physical examination findings typically collected in the ED setting.

In addition, a subsample of 500 visit pairs was independently classified by a physician to serve as a reference standard for comparison with the LLM outputs. GPT-4 achieved an overall accuracy of 89% in identifying higher-acuity cases, outperforming GPT-3.5 Turbo, which achieved 84% accuracy. Within the physician-reviewed subset, GPT-4's performance (88% accuracy) was comparable to that of the human expert (86%). These findings underscore the growing potential of LLMs as decision-support tools in emergency medicine, particularly for rapid patient prioritisation based on textual clinical data.

In a related study conducted in Turkey, researchers evaluated GPT-4 (ChatGPT) for its ability to predict ED triage outcomes according to locally established triage protocols [41]. The study took place in the ED of a tertiary hospital, where patient data, including age, oxygen saturation, and other clinical parameters, were collected upon arrival. Patients were first triaged by the hospital's triage team following standard procedures. The same data were then provided to ChatGPT for automated triage classification, and an EM physician's assessment served as the gold standard.

The results demonstrated near-perfect agreement between both human and AI triage decisions: the triage team achieved a Cohen's Kappa of 0.893 relative to the gold standard, while ChatGPT achieved 0.899. These findings indicate that LLMs can approach expert-level performance in structured triage environments, provided that clear local rules and well-defined input variables are available.

Although the LLMs discussed above were general-purpose language models not specifically trained for emergency medicine or triage, they nonetheless demonstrated remarkably high performance in classifying patient acuity levels [40]. Building on this progress, numerous studies have explored the use of Machine Learning (ML) and Deep Learning (DL) techniques, including Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Natural Language Processing (NLP), to develop specialised models aimed at improving triage accuracy and efficiency in emergency departments [18] [42].

Across multiple comparative analyses, ML-based triage models consistently outperformed traditional systems such as the ESI and MTS in predictive capacity and reliability. In particular, algorithms like XGBoost, Random Forest, and Deep Neural Networks (DNNs) achieved superior classification performance. Moreover, when ML models were combined with NLP components capable of interpreting unstructured clinical notes, triage outcomes were significantly enhanced [29].

These findings underscore a broader shift toward data-driven, adaptive triage systems, highlighting the transformative potential of AI in emergency care, especially for RLS that face personnel shortages and lack standardised triage tools.

2.2 AI in Healthcare for RLS

Despite the numerous challenges faced by Low- Middle- Income Countries (LMICs), they are increasingly engaging in the adoption of AI to enhance healthcare delivery. Resource-limited settings (RLS) broadly include LMICs, rural or underserved regions in HICs, disaster-affected zones, and refugee camps [43]. These settings are typically characterised by political instability, limited human and financial resources, inadequate infrastructure, and weak health governance systems. Such conditions often result in poor healthcare quality, rising disease burdens, and under-resourced medical facilities [44].

Afghanistan, similar to Somalia, Yemen, Sudan, the Central African Republic, and Palestine, has experienced prolonged political instability, foreign interference in governance, and recurrent natural disasters such as floods and earthquakes [45]. Consequently, the Afghan healthcare system has remained fragmented, underfunded, and poorly coordinated for decades, with conditions continuing to deteriorate [46]. In 2022, the Ministry of Public Health reported that one in two Afghans suffers from depression or anxiety [47]. In response, researchers developed a prototype AI-based chatbot designed to provide preliminary diagnostic support and treatment guidance for mental health disorders among patients in Kabul hospitals. The chatbot enabled users to discuss sensitive issues anonymously and automatically forwarded reports to doctors for follow-up. Although the system has not yet been deployed at scale, most respondents expressed confidence that such a tool could help reduce the prevalence of mental health disorders in the country.

In Ethiopia, researchers examined the perceptions of medical professionals toward AI in healthcare, aiming to inform effective adoption strategies and policymaking [48] . The study revealed that fewer than half of the respondents were familiar with AI applications, with awareness notably higher among physicians than among nurses and auxiliary staff. Despite this limited knowledge, most participants expressed optimism, believing that AI could enhance diagnostic accuracy, treatment outcomes, and overall healthcare efficiency, though they also advocated for a cautious, well-regulated adoption. This cautious optimism mirrors the sentiments of many healthcare professionals across RLS [4][49], who recognise AI's transformative potential but remain concerned about job displacement, ethical implications, algorithmic bias, and data privacy.

Unlike Ethiopia, India has a booming tech scene, contributing actively to the development of AI medical tools. Amid the debate on AI's accuracy and privacy issues, there exist AI healthcare companies developing locally relevant and affordable tools [50]. Some of the innovative ones include **Niramai** developing a device for the diagnosis of early-stage breast cancer, **Qure.ai** developing an AI app for radiology image analysis and **Artelus and Remidio** offering telemedicine services for ophthalmology screening [50]. In addition, AI is already being used in India in robotics-assisted surgery, IBM Watson for oncology, and AI chatbots/voice assistants for patient engagement [51]. These tools enhance diagnosis, empower patients, and support medical staff, but they require improved policies, infrastructure, and trust frameworks.

The Swahili version of the Ada app, an AI-based symptom-assessment app (SAA), was made available in East Africa in 2019. Swahili is one of the 10 most widely spoken languages in the world [52]. As at 2021, it was reported that the app had over 92,000 users and 94,000+ assessments in Tanzania [53]. The app was developed as part of Ada's Global Health Initiative (GHI) to mitigate the problems arising from the shortage of health workers [54]. Ada Health partnered with a Swiss foundation, Fondation Botnar, and the Muhimbili University of Health and Applied Sciences in Tanzania to localise the app which is available in iOS and Android versions. A study protocol on the app was scheduled to be carried out by July 2021 in Tanzania, but results are yet to be published [53] [55].

Latin America countries like Argentina, Brazil, Mexico, Chile and Colombia also share in the burden of limited resources on healthcare quality. The challenges, deeply rooted in underinvestment in healthcare and education, lead to poor adherence to clinical guidelines and weak patient outcomes, especially in public/philanthropic care [44]. In a systematic review of diagnostic and screening AI tools for RLS in Brazil, 25 papers were identified that discussed how AI tools are being developed, tested and adopted in Brazillian health systems [56]. These tools addressed conditions such as ophthalmology and infectious diseases, diabetic retinopathy research and glaucoma diagnosis. An interesting finding was the predominance of public funding for the identified papers, indicating the potential for use by the Brazilian

Unified Health System (UHS) at no cost to Brazilian citizens. The review also demonstrated the diversity of AI technology used and is optimistic about the advancement of AI applications in Brazil's healthcare.

Other examples of AI in RLS include **Gukiza**, which is a remote patient monitoring platform designed to enhance cancer treatment and clinical trial processes [43]. It was developed by *Hurone AI* and is currently available for use by doctors in Nigeria, Kenya and Rwanda [57]. A computer-aided tuberculosis (TB) diagnosis system was trialled in Peru [43]. The system can rapidly identify markers of TB based on patient details and chest X-ray readings. In South Africa, a fuzzy-logic-based AI model for early detection of cholera outbreaks was developed. The model analyses environmental and epidemiological data to accurately make a forecast. *Afya Rekod* is a digital health platform launched in Kenya. It manages patient health records and provides AI-based analysis and reports, prompting patients to be more engaged in their health [58].

While these developments may not be fully integrated across the country where they are developed, they reflect the possibilities of AI in the RLS healthcare system. These tools, each tailored to the peculiarities of the RLS, demonstrate that context-aware innovation is both feasible and beneficial in such environments. Although AI-assisted triage tools are not yet widely deployed in RLS, their introduction could be a welcome advancement given the positive reception and adaptability of other AI medical applications. Nevertheless, realising this potential requires addressing persistent barriers that continue to limit the large-scale implementation of AI in LMICs.

2.3 Issues with AI Adoption in Healthcare for RLS

Beyond the many benefits associated with the use of AI in healthcare, several challenges persist. These include inadequate funding, limited workforce training, unreliable data systems, poor infrastructure, weak policy and regulatory frameworks, and socio-cultural factors that affect technology adoption [5], [59]. Developers often attempt to adapt to these constraints by creating lightweight or mobile-based tools in local languages. Still, these solutions rarely overcome the deeper structural limitations that impede scalability and sustainability.

Financial constraints and infrastructure challenges are complementary issues hindering AI adoption. On one hand, AI solutions require a high initial investment that may not yield comparable returns in the first few years, and insufficient budget allocation for national healthcare further pushes AI down the priority list, making it appear as a luxury. On the other hand, poor infrastructures like unreliable internet connections, unstable power supply, and outdated computing systems complicate the development and large-scale deployment of these AI tools [5], [60]. Upgrading infrastructures makes for easy development and scalability of AI tools, albeit requiring a monetary commitment that is already difficult for RLS.

The performance and reliability of any AI model depend heavily on the quality and representativeness of the data and algorithms on which it is trained [5], [43]. In many RLS, however, the scarcity of local data and the reliance on external datasets introduce significant bias, as imported data may not reflect the local population's health conditions or context. Moreover, because many algorithms learn from statistically dominant patterns, they risk reinforcing existing health inequalities. Data collection itself presents major obstacles: few medical facilities utilise Electronic Health Record (EHR) systems, and those that do often struggle with issues of data inconsistency, incompleteness, and lack of standardisation, rendering much of the data unsuitable for analysis. Most healthcare institutions still rely on paper-based records, which are vulnerable to loss, damage, and fragmentation. Beyond technical limitations, ethical and security concerns further complicate data management. Patients may not always give informed consent for data use or fully understand how their information will be processed. Data protection practices also remain weak - paper files are often stored in unsecured rooms, while digital records may be hosted on cloud servers located outside the country, raising concerns about privacy and sovereignty.

Human resource limitations present another significant barrier to AI adoption in resource-limited settings. Skilled IT professionals are often scarce, as many experts migrate to high-income countries in search of better opportunities [2], [5]. Developing, deploying, and maintaining effective AI systems requires collaboration among data analysts, engineers, scientists, software developers, and domain experts. When these roles are poorly defined or understaffed, the available personnel may experience work overload, and the resulting systems risk being clinically ineffective. Furthermore, most healthcare workers receive little or no training in the use of AI technologies, leading to potential misuse, mistrust, or outright resistance. Poorly developed or inadequately validated AI models can produce inaccurate recommendations that may harm patients [43]. Studies have also shown that some healthcare workers remain reluctant to adopt digital tools, as observed among nurses using triage systems in tertiary hospitals in Lagos [13].

The success of AI implementation ultimately depends on the people who will interact with it, including the health workers and the patients. In many RLS, low levels of digital and general literacy make even simple technologies, such as mobile health applications, difficult to use [61]. In Nigeria, a significant portion of the population continues to seek care from religious or traditional healing centres rather than formal medical facilities. Patients living with sensitive conditions such as HIV or mental illness may be reluctant to disclose personal information through digital platforms, while older adults often exhibit resistance due to unfamiliarity with technology. Additionally, many patients still prefer direct, face-to-face interactions with healthcare providers over digital consultations. These socio-cultural factors shape how technology is perceived and adopted, regardless of the specific AI tool involved. Ensuring cultural compatibility during development is therefore essential [5]. This can be achieved through early engagement with potential users, incorporating their feedback into the design process to improve acceptance and usability

Medical applications of AI in resource-limited settings have been examined across 42 unique studies and grouped into four broad categories, although some overlap exists among them [43]. As illustrated in Figure 2.5, these categories enabled researchers to conduct an ethical evaluation across six key dimensions: bias and fairness, non-maleficence, privacy and security, autonomy, and transparency. The studies also identified several ethical dilemmas. For instance, questions were raised about whether AI systems should be permitted to make final clinical decisions in contexts where human oversight is scarce, given the existing shortage of healthcare workers, or whether such delegation compromises patient safety. Another dilemma concerns the morality of investing in AI infrastructure when basic healthcare systems remain underdeveloped. To address these concerns, the researchers proposed strategies for ethical implementation, including the development of context-specific ethical and regulatory frameworks tailored to the realities of resource-limited settings.

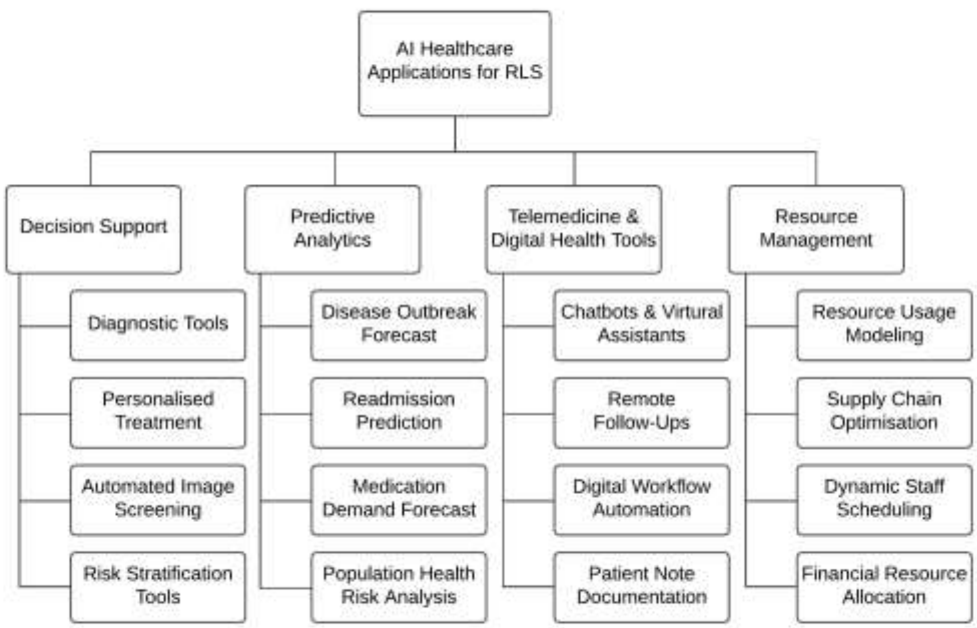


Figure 2.5: Overview of AI-based Healthcare Applications for RLS

Policy gaps within legal and institutional frameworks further complicate the adoption and acceptance of AI in healthcare. Unclear definitions of roles, responsibilities, and accountability create confusion among key stakeholders. For instance, in the event of an erroneous AI-generated clinical decision, it remains uncertain who should be held responsible - the developers, the hospital management, or the healthcare professional involved [5]. The absence of a standardised regulatory process also raises concerns about how new AI tools should be tested, validated, and approved before integration into the health system, and which authorities should oversee these processes.

In Nigeria, several agencies play roles that touch on aspects of AI governance. The **Federal Ministry of Communication, Innovation and Digital Economy (FMCIDE)** is primarily responsible for driving technological advancement through its agencies, including the **National Information Technology Development Agency (NITDA)**, which coordinates national IT development in line with the National Information Technology Policy, and the **National Data Protection Commission (NDPC)**, which enforces the **Nigeria Data Protection Act (NDP Act)**. Additionally, the **National Agency for Food and Drug Administration and Control (NAFDAC)** regulates medical devices; the **Medical and Dental Council of Nigeria (MDCN)** determines professional standards and competencies for medical personnel; and the **Federal Ministry of Health and Social Welfare (FMOH)** oversees national health policy and infrastructure development. Yet medical professionals express concern that existing frameworks remain fragmented and insufficiently robust, particularly due to the absence of a centralised regulatory body dedicated to AI governance in healthcare [4], [60].

Finally, the introduction of AI into resource-limited settings should be phased and gradual, accompanied by regular audits, ethical reviews, and continuous policy updates. Active participation of local stakeholders including community leaders, clinicians, and patients, should be prioritised to prevent "technological colonisation," where externally developed AI solutions fail to align with local values, cultures, and traditions [5]. International collaboration and capacity-building initiatives can help bridge gaps in funding and technical expertise; however, such partnerships must clearly define intellectual property rights and data governance structures. Addressing concerns about data colonialism and exploitation, where data from vulnerable communities is used for commercial or academic gain by external actors, is essential to fostering mutual trust and equitable collaboration. In addition, stronger partnerships between the government and the private sector are needed to drive sustained investment in infrastructure and ensure the long-term viability of AI integration in healthcare [4], [5], [56], [60].

2.4 Current State of Nigeria’s Digital Health

Recently inaugurated by the federal government in March 2024 is the Nigeria Digital Health Initiative (NDHI) Implementation Committee [32]. The committee is tasked with developing and overseeing the establishment of a national digital health architecture aimed at transforming healthcare delivery through technology. The NDHI aims to enhance healthcare services by enabling efficient data exchange, implementing a national Electronic Medical Record (EMR) system, improving health indicators, and strengthening quality assurance mechanisms. Since data collection and accessibility remain major challenges for AI development in Nigeria, the successful implementation of a nationwide EMR system is expected to significantly improve data availability and reliability for future AI-driven healthcare solutions.

Complementing the government’s efforts under the NDHI, *Dobic Health*, a leading technology-driven healthcare firm in Nigeria, launched **SmartMRS** in August 2025 - the country’s first AI-enabled Electronic Medical Records (EMR) platform [62]. Developed in collaboration with Nigerian doctors, nurses, and laboratory scientists, SmartMRS exemplifies the growing local innovation in digital health infrastructure. The platform integrates two key AI components: **Diagnocare AI**, which analyses symptoms, patient history, and contextual information to suggest possible diagnoses and guide treatment decisions; and the **Clinical Vision Model**, which interprets X-ray images, generates diagnostic reports, and provides clinical insights. Such innovations align with the NDHI’s objective of strengthening digital data systems and demonstrate the potential of public–private collaboration in accelerating the adoption of AI-assisted healthcare solutions in Nigeria.

In addition to government and private-sector initiatives, researchers and organisations have also contributed to advancing digital health policy in Nigeria. A *Digital Health Transformation Roadmap* was developed to guide the integration of digital technologies into the healthcare system, with a particular focus on improving Maternal, Newborn, Adolescent, and Child Health (MNACH) outcomes [63]. The roadmap outlines a three-phase implementation plan spanning 2025–2030, which provides clear timelines, stakeholder responsibilities, and measurable performance indicators. Phase 1 focuses on foundational preparation, addressing the existing gaps in infrastructure and policy required for scalable digital health integration. Phase 2 emphasises validation and expansion, supporting multi-stakeholder pilot projects to test, refine, and scale context-appropriate innovations. Phase 3 centres on national integration and full interoperability across digital health systems. At present, Phase 1 is approximately 45% complete, and Phase 2 has commenced at 15%. By clearly defining responsibilities, key actions, and success metrics, this roadmap enhances accountability and provides a structured pathway for achieving digital health maturity.

Furthermore, it creates an enabling environment that could support the future adoption and scaling of AI-assisted triage tools such as the one developed in this study.

In a related development, the Science for Africa Foundation (SFA Foundation), in collaboration with mDoc and Nigeria Health Watch, convened an event titled *The Nigerian AI Healthcare Horizon* to examine the opportunities, challenges, and policy considerations surrounding the ethical use of AI in healthcare. The discussions culminated in a set of recommendations aimed at guiding the successful and responsible integration of AI technologies into Nigeria's health system[64]. However, as with the digital health roadmap mentioned earlier, this initiative was not directly organised or endorsed by the federal government, raising concerns about the extent to which these recommendations will be formally adopted or implemented within national policy frameworks.

Despite infrastructural and policy constraints, most Nigerian healthcare professionals are aware of and optimistic about the potential of AI to enhance diagnostics, reduce medical errors, and improve overall quality of care. However, the absence of formal AI training, limited institutional support, and inadequate infrastructure continue to hinder implementation.

A 2023 cross-sectional study involving 404 healthcare professionals across Nigeria's six geopolitical zones found that more than half of the respondents possessed good knowledge of AI applications and believed that AI would complement rather than replace human intelligence. Only 42.2% expressed concern that AI adoption could lead to job losses [65]. Similarly, a separate online cross-sectional survey of 214 medical practitioners across various Nigerian regions reported that 75% of participants held positive perceptions of AI, and 87% supported the inclusion of AI and coding courses in medical curricula. Interestingly, the study also noted that professionals in private hospitals exhibited more favourable attitudes toward AI than their counterparts in public institutions [49].

A study examining the perspectives of 15 oncologists across nine major tertiary hospitals in Nigeria found that, while oncologists possessed a basic theoretical understanding of AI applications, most lacked practical experience in using such technologies [4]. Despite this, they recognised AI as a potential solution to key challenges in cancer care delivery, emphasising the importance of context-sensitive adoption, south–south collaboration, and technology partnerships rather than uncritical use of externally developed systems. The study was limited by institutional bias, as it focused exclusively on large, urban teaching hospitals, likely underestimating barriers faced in primary health centres and rural facilities where most Nigerians access care. To guide AI integration in resource-constrained settings, the researchers subsequently developed a readiness pyramid, shown below, outlining the prerequisites for successful implementation.

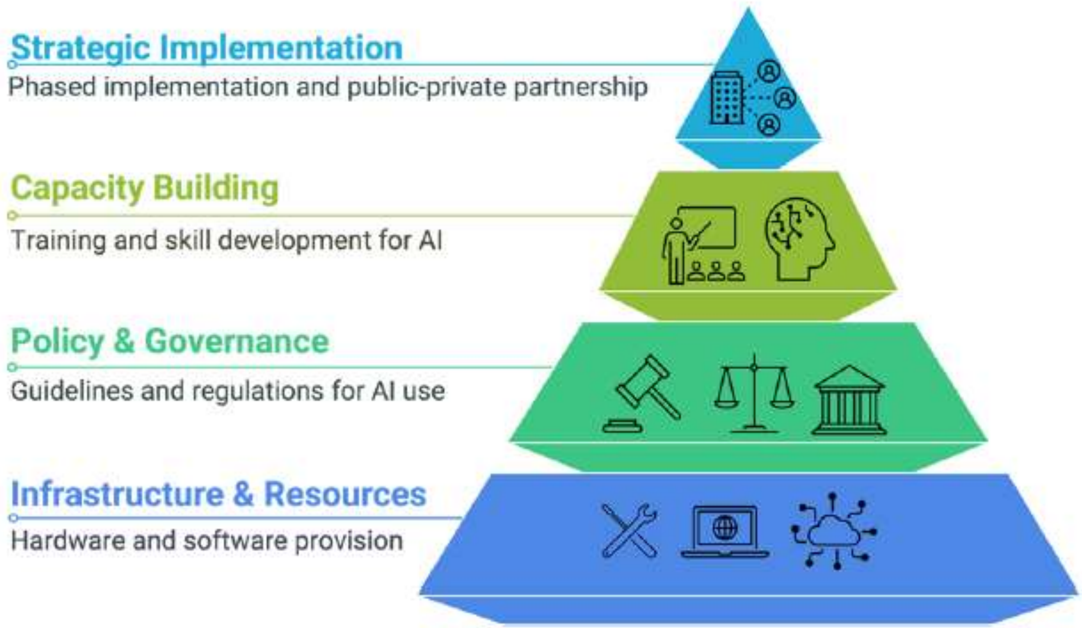


Figure 2.6: Recommended Pyramid for the implementation of AI in Healthcare for RLS

Collectively, these findings suggest that while awareness and acceptance of AI are growing among Nigerian healthcare professionals, systematic investment in AI education, infrastructure, and policy is necessary to translate optimism into practical adoption.

The adoption of AI in Nigeria's healthcare sector is still evolving, but several locally developed tools demonstrate its growing potential. In diagnostics, *XOLANI Health* offers multiple AI-enabled solutions to support radiology and clinical pathology workflows, including **Xolani Bridge**, a web-based teleradiology platform connecting users to a global network of radiologists; **DICOM-X**, an AI-assisted collaborative DICOM viewer that analyses medical images in real time; and **Xolani AIR**, a cloud-based SaaS platform available via web and mobile applications that assists with radiology diagnoses [66], [67]. In maternal and infant care, *Dobic Health* established **HelpMum**, which includes digital health tools such as **ADVISER** (an AI-driven vaccination optimisation framework), **VaxBot** for immunisation scheduling, **MamaBot** for pregnancy guidance, and **StratifyAI** for pregnancy risk assessment, alongside an e-learning platform for traditional birth attendants [68], [69]. In pharmacy and retail healthcare, *ProNOV* Technology supports digital transformation by tracking drug expiry, managing inventory, monitoring sales, and providing real-time analytics [59], [70]. In chronic disease management, *Hub Care Health* runs an Insights Lab that integrates data and people to provide personalized care, and improved health outcome [59] [71]. Finally, telemedicine platforms such as *Mobihealth International, iWello, cribMD, Tremendoc,* and *DRO Health* expand access to affordable healthcare by connecting patients to health providers via mobile applications

[72]. These initiatives highlight the diverse applications of AI in Nigeria, ranging from diagnostics and pharmacy management to chronic care and virtual healthcare delivery.

3 Methodology

The aim of the thesis is to explore how AI-assisted triage could be adapted and implemented in Nigeria's healthcare system. The study adopts a prototype-based, data-driven approach to evaluate the feasibility of an AI-assisted triage system. Developing a prototype allows for practical exploration of how clinicians might interact with an AI-triage tool in real-world conditions, while using synthetic data ensures ethical and logistical feasibility in a resource-limited context. This approach bridges theoretical modelling and practical application, enabling the assessment of system performance, usability and adaptability without relying on patient data. It thus provides a controlled, low-risk pathway toward understanding how AI-assisted triage could improve efficiency and access in Nigerian healthcare settings.

The research approach (technical + practical) is justified in the following ways:

- Safety: synthetic data avoids patient risk.
- Feasibility: enables proof-of-concept without requiring full clinical deployment.
- Adaptability: allows evaluation of potential in resource-limited settings.

3.1 Technical Approach

3.1.1 Overview

This section outlines the technical approach adopted for data generation and model development. Due to the unavailability of comprehensive triage datasets from Nigerian healthcare institutions and given that conducting a clinical survey would have required extensive administrative approval processes beyond the project's timeframe, synthetic data was generated to simulate realistic triage scenarios. The generated dataset was used to train a Random Forest classifier, selected for its robustness and interpretability in handling heterogeneous medical data. Hyperparameter tuning was performed systematically using a cross-validation approach to optimise model performance. Evaluation metrics included Accuracy, Precision, Recall and AUC Score to assess predictive reliability across triage categories. Furthermore, model explainability was explored through SHAP analysis and feature importance ranking, enabling interpretation of key variables influencing triage decisions.

Data generation, preprocessing, and model development were conducted in the *Google Colab* environment. This cloud-based platform provided an accessible workspace with GPU acceleration and enabled the visualisation of intermediate outputs throughout model training and evaluation.

3.1.2 Synthetic Data Design

The synthetic dataset was designed to represent common triage decision-making contexts encountered in Nigerian healthcare facilities, particularly in emergency departments and primary care settings [73], [74], [75] . The simulated data included variables typically used by clinicians to determine triage urgency levels. These features were selected through a review of existing triage protocols, such as the Emergency Triage Assessment and Treatment (ETAT) and the Interagency Integrated Triage Tool (IITT), which share similar decision criteria relevant to resource-limited environments [24]. The ETAT utilises a three-scale urgency level while the IITT relies on specific physiological parameters to assign urgency levels. Based on these findings, the data was generated and labelled. The figure below shows the structure of the IITT [24].

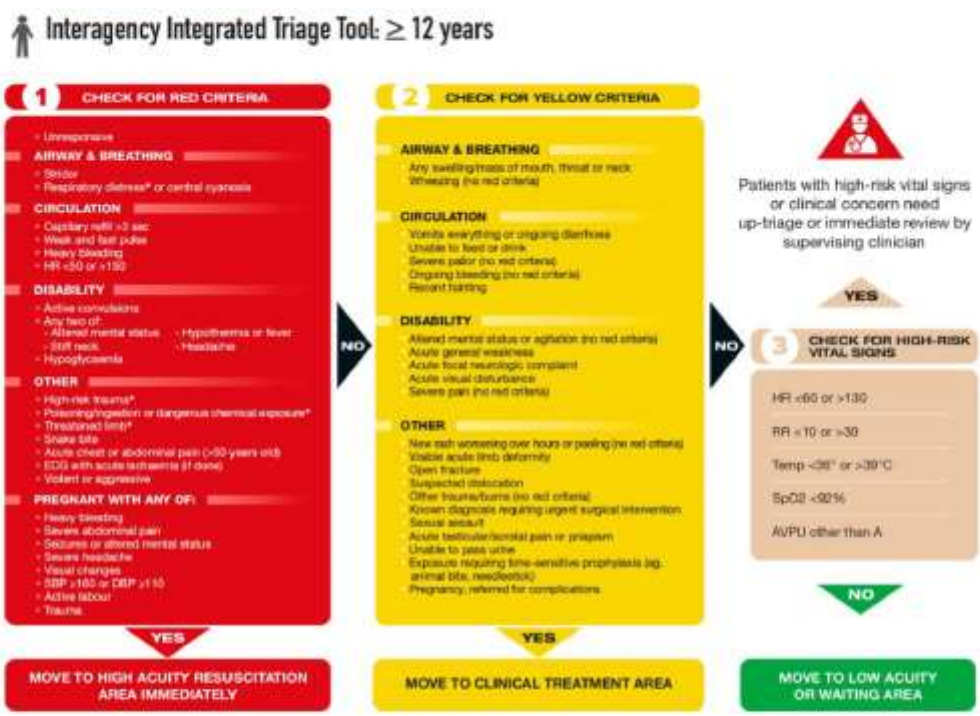


Figure 3.1: The decision-making process for the IITT

The dataset comprised the following categories of features:

- **Demographic attributes:** age and sex
- **Vital Signs:** heart rate, respiratory rate, systolic and diastolic blood pressure, body temperature, and oxygen saturation.
- **Arrival Information:** mode of arrival, chief complaint
- **Medical Status:** Consciousness state, active bleeding, pregnancy status
- **Triage label:** the target variable representing the triage category - *Emergency*, *Urgent*, or *Non-Urgent*.

3.1.3 Data Generation Process

Data generation was implemented programmatically using Python's data manipulation and simulation libraries, such as *NumPy, Pandas and Random*. Conditional logic was applied to ensure internal consistency among variables; for example, patients with extremely high heart rates or critically low oxygen saturation were more likely to be assigned to the *Emergency* category. This probabilistic rule-based generation ensured that the dataset preserved clinically coherent relationships among variables.

Each feature in the dataset was designed to reflect clinically plausible values consistent with those typically observed in the emergency departments (ED) of tertiary hospitals in Nigeria [6], [22], [74]. The dataset comprised 14 attributes (13 predictor variables and one target variable) with a total of 600 observations. To ensure balanced class representation, the target labels were evenly distributed across the observations. While this approach does not fully reflect real-world distributions, where non-emergent cases are typically more frequent [6], it was necessary to manage potential model bias. The dataset was stored in a structured tabular format (CSV) and version-controlled to ensure reproducibility. The descriptions of the data features are explained below:

1. **Triage:** This is the target feature. It includes *Emergency, Urgency and Non-urgent* outcomes. Specific attributes such as *Vital Signs, AVPU Scale,* and *Mode of Arrival* were individually generated for each target.
2. **Respiratory Rate:** The number of breaths a person takes per minute. For a healthy adult at rest, the normal range is typically between 12 and 20 breaths per minute.
3. **Heart Rate:** The number of times the heart beats per minute. A normal resting heart rate for adults typically falls between 60 and 100 beats per minute.

4. **Systolic Blood Pressure (SBP):** The first (top/upper) number in a blood pressure reading. It measures the pressure the blood pushes against the artery walls when the heart beats. A normal SBP reading is <130.
5. **Diastolic Blood Pressure (DBP):** The second (bottom/lower) number in a blood pressure reading. It measures the pressure the blood pushes against the artery walls while the heart muscle rests between beats. A normal DBP reading is <85
6. **Temperature**: The degree of hotness or coldness of the human body. Normal body temperature is 37 °Celsius
7. **Oxygen Saturation:** The percentage of haemoglobin in the blood that is carrying oxygen. For example, an SpO2 of 97% means 97% of the haemoglobin is carrying oxygen and 3% is not. A healthy SpO2 level for most people is typically 95-100%.
8. **Chief Complaint:** This refers to the main complaint that brings the patient to the ED or the hospital in general. The options are *Chest pain, Difficulty breathing, Fever, Seizure or loss of consciousness, Abdominal pain, Injury, Headache, Vomiting or diarrhoea, Weakness or fatigue, Pregnancy-related complication, Psychiatric/behavioural emergency,* or *Other*. The assignment of a complaint to an observation is not entirely random, but rather depends on other variables, such as **Active Bleeding, Age**, and **Sex**. This allowed for consistency, such as only *females* below the age of *50* being *pregnant* or *actively bleeding* during an *injury*.
9. **Mode of Arrival:** This states how the patient arrives at the facility. The options are *Walk-in, Private Vehicle* or *Ambulance*. Private vehicle includes all modes of transportation (privately owned or public commute, such as cars, bikes, tricycles) that do not constitute walking to the hospital or using an ambulance. Each triage category had a different weight for the distribution of this attribute to an observation, for example, in the *emergency* category, the weight was (0.2, 0.3, 0.5)
10. **Active Bleeding:** Yes or No. This records whether the patient is actively bleeding on arrival. It was generated as a random choice whose weights varied according to *Injury* and *Pregnancy-related* chief complaints. Otherwise, the weight for distribution was (*0.1, 0.9*)
11. **Pregnancy:** Yes or No.
12. **AVPU Scale**: This describes the state of consciousness of the patient. A: Alert, V: Responds to voice, P: Responds to pain or U: Unresponsive. The distribution of this feature was weighted; for example, the weighted distribution for the *non-urgent* category was [0.6, 0.2, 0.15, 0.05]
13. **Age**: This is the age of the patient. The minimum value is 18 while the maximum is 80.
14. **Sex**: Male or female. It was generated as a random choice with no weights attached.

3.1.4 Data Preprocessing

Prior to model training, several preprocessing steps were performed to prepare the data for machine learning.

1. **Data Cleaning:** Synthetic datasets, while structured, may still contain noise or implausible combinations. Outlier detection rules were implemented to identify and adjust extreme values that fell outside clinical boundaries.
2. **Feature Encoding:** Categorical binary variables such as *sex* and *pregnancy* were encoded using binary encoding, while multinomial variables such as *mode of arrival* and *chief complaint* were encoded using one-hot encoding. This process converted the variables into a numerical format suitable for the ML model.
3. **Normalisation:** Although Random Forest classifiers are generally insensitive to feature scaling, MLP classifiers are not. Normalisation was applied to certain continuous variables to ensure comparability across plots and interpretability during later feature importance analysis.
4. **Data Splitting:** The cleaned dataset was partitioned into **training (70%)**, **validation (10%)**, and **testing (20%)** subsets. This ensured robust model evaluation and reduced overfitting risks.
5. **Exploratory Data Analysis**: Descriptive statistics and visualisations (*boxplots,* and *countplots*) were produced to confirm realistic data behaviour and to verify that no single variable disproportionately influenced the overall data distribution.

3.1.5 Algorithm Selection

The triage problem in this study is formulated as a supervised multi-class classification task, where the model predicts patient urgency levels - *Emergency*, *Urgent*, or *Non-Urgent* - based on a combination of clinical and demographic features. Given the critical nature of triage decisions, the selected algorithm needed to provide not only strong predictive performance but also interpretability and robustness to noise. Furthermore, the model had to be computationally efficient enough to run on resource-limited systems typical of Nigerian healthcare facilities.

Several studies have reviewed the ML and AI models that perform better than traditional triage models. Of these AI models, the high-performance ones are XGBoost, DNN, MLP and SVM [29] [76]. In this thesis, XGBoost, Random Forest and MLP classifier algorithms were compared, and the algorithm with the highest accuracy was further enhanced to be used for development. The three algorithms performed well on the test set, with accuracy levels well above 90%. However, the rationale behind selecting the RF classifier is that it can be easily explained and is less computationally expensive.

The Random Forest algorithm is known for its strong performance in handling heterogeneous datasets that combine categorical and continuous variables, its resilience to overfitting, and its ability to handle missing values and non-linear relationships [77]. This makes them particularly well-suited for medical triage, where decision boundaries between urgency levels are complex and not easily separable by linear models.

Bootstrap Aggregation, also known as bagging is the ensemble technique used in RF. A Random Forest is made up of multiple decision trees. Each tree is fitted to a sample of the data or the entire dataset (*Bootstrap*). The output from each tree is then combined and *aggregated* to give the output of the RF. The figure below depicts the graphical representation of the random forest model [78]. For continuous targets, the average of single trees' predictions is calculated, but for discrete targets (as is the case of the triage category), the mode is calculated. Additionally, Random Forests require minimal parameter tuning and provide built-in measures of feature importance, which facilitate model explainability - an essential consideration for building trust amongst patients and healthcare workers.

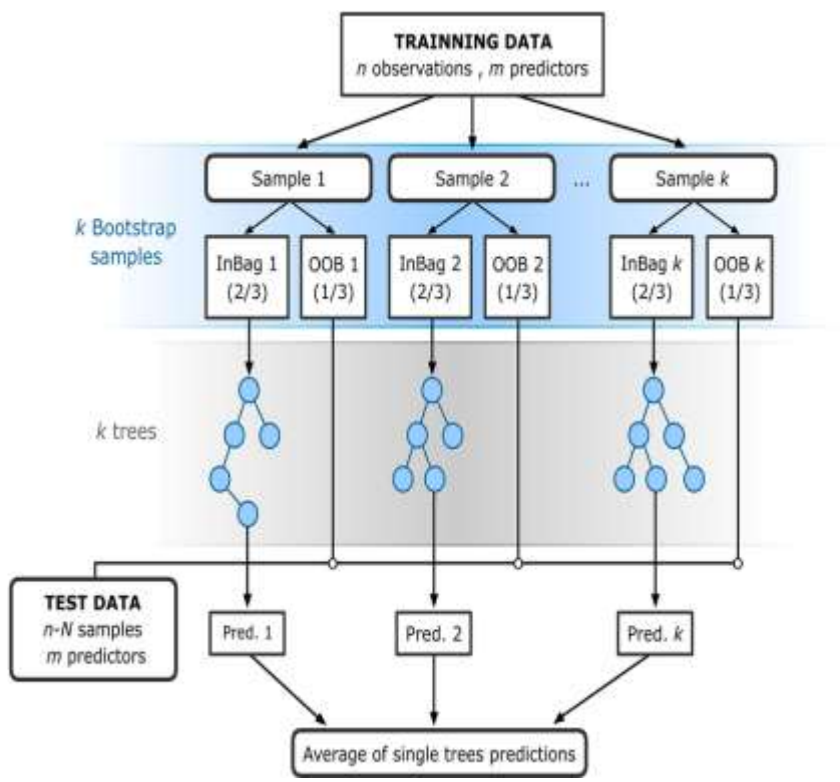


Figure 3.2: Working mechanism of the Random Forest model.

From an implementation standpoint, the Random Forest algorithm offers practical advantages for low-resource healthcare environments. It can be deployed on standard computing hardware without the need for specialised GPU acceleration [79] and model retraining or updates can be performed efficiently as more data becomes available.

3.1.6 Model Training

'A random forest is a meta estimator that fits a number of decision tree classifiers on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.' [80]. The Random Forest classifier was implemented using the Scikit-learn library (*version 1.7.2*) in Python. The dataset was split into training and testing subsets using a stratified 80:20 ratio to preserve class proportions. This ensured that all triage levels were proportionally represented in both subsets, thereby minimising class imbalance effects.

Hyperparameter optimisation for the Random Forest (RF) model was conducted using Randomised Search Cross-Validation (*RandomizedSearchCV*), which samples a fixed number of parameter combinations from a predefined search space. Unlike Grid Search CV, which evaluates all possible combinations, randomised search provides a more computationally efficient approach by randomly sampling configurations and testing a subset of them. In this study, 30 iterations (*n_iter = 30*) were performed, enabling a broad yet efficient exploration of the parameter space while minimising computational cost. This approach enabled the identification of a high-performing parameter combination without the exhaustive evaluation required by grid search.

Key parameters explored included the number of decision trees (*n_estimators*), maximum tree depth (*max_depth*), and the minimum number of samples required for node splits (*min_samples_split*). Five-fold cross-validation was conducted on the training data to estimate model generalizability and reduce overfitting risk. Each model configuration was evaluated on the mean accuracy, and the best-performing combination was selected for final training. To ensure experimental reproducibility, random seeds were fixed (*random_state=42*), and model versions were tracked using a consistent software environment defined in the *requirements.txt* file.

3.1.7 Model Evaluation

Model evaluation was performed on the 20% test subset, which was withheld from training to simulate unseen clinical data. The evaluation focused on both overall predictive performance and class-wise discrimination capability, reflecting the practical need to accurately identify high-risk (Emergency) cases without excessive false positives.

Model performance was assessed using standard classification metrics: **Accuracy**, **Precision**, **Recall**, and **F1-score**. These were computed for each triage category (*Emergency*, *Urgent*, and *Non-Urgent*) to evaluate the model's ability to classify patient urgency levels correctly. For a more comprehensive evaluation, the **Receiver Operating Characteristic (ROC)** curves and corresponding **Area Under the Curve (AUC)** scores were also computed using a one-vs-rest (OvR) approach. Each triage class (*Emergency*, *Urgent*, and *Non-Urgent*) was compared against the others, and macro-averaged AUC values were reported to assess the overall discriminative ability of the Random Forest classifier. The AUC scores provided an aggregate measure of separability between triage categories based on predicted probabilities.

The following definitions explain the terminology associated with understanding the model's performance [81]

- **Accuracy** measures the proportion of correctly classified instances among all cases, providing an overall indicator of model performance.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision** (or Positive Predictive Value) indicates the proportion of true positives among all cases predicted as positive. In a medical triage context, this measures how often a case predicted as *Emergency* is truly an emergency.

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall** (or Sensitivity) quantifies the model's ability to identify all actual positive cases. For triage, it shows how effectively the system detects all *Emergency* or *urgent* cases.

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-score** is the harmonic mean of Precision and Recall, providing a balanced measure when both false positives and false negatives carry significant cost.

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

In these equations:

- **TP (True Positive):** The number of cases correctly predicted as belonging to a given class (e.g., actual *Emergency* predicted as *Emergency*).
- **FP (False Positive):** The number of cases incorrectly predicted as belonging to a class (e.g., *Urgent* or *Non-Urgent* cases wrongly predicted as *Emergency*).
- **FN (False Negative):** The number of cases that actually belong to a class but were not predicted as such (e.g., *Emergency* cases missed by the model).
- **TN (True Negative):** The number of cases correctly predicted as *not* belonging to a given class.
- **Confusion Matrix:** This is a table that compares the predicted labels to the true labels by showing the number of TP, FP, FN and TN. The figure below shows a confusion matrix [82]

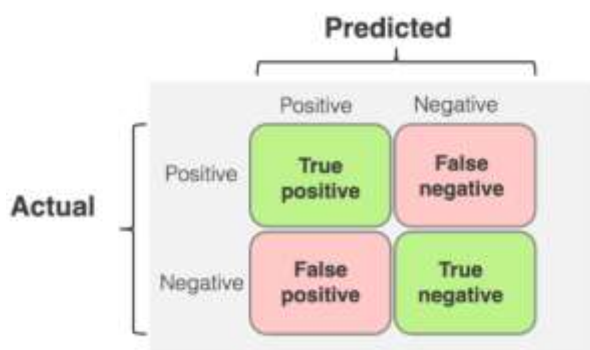


Figure 3.3: Confusion matrix

- **ROC (Receiver Operating Characteristic (ROC) Curve**: This illustrates the trade-off between sensitivity (True Positive Rate) and the False Positive Rate at various decision thresholds. It provides insight into the model's ability to distinguish between different triage levels based on predicted probabilities.
- **AUC (Area Under the Curve) Score**: This quantifies the discriminative capacity of ROC, with higher values indicating stronger separability between classes. Figure 3.4 below shows an ROC curve [82].

  Using the OvR approach, each class is compared against all other classes combined:
  - Emergency vs. {Urgent + Non-Urgent}
  - Urgent vs. {Emergency + Non-Urgent}
  - Non-Urgent vs. {Emergency + Urgent}

  This yields one ROC curve per class and an AUC value for each class. Then the macro-average AUC is calculated. Macro means all classes are treated equally.

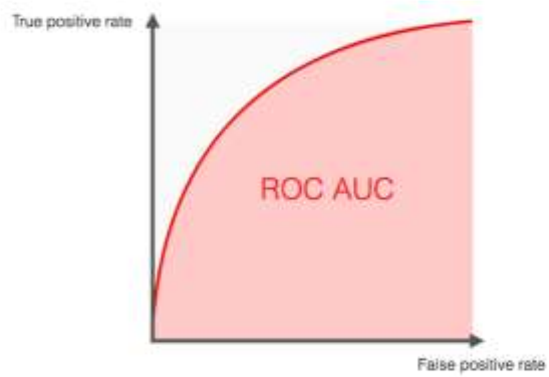


Figure 3.4: ROC - AUC

3.1.8 Model Explainability

Global explainability of the Random Forest model was examined using the built-in feature importance scores. In this context, feature importance represents the average contribution of each variable to the model's decision-making process across all trees and all predictions. The RF algorithm computes these scores based on the total reduction in Gini impurity achieved by each feature when used to split the data. Gini impurity quantifies the degree of class mixing within a node and is defined as:

$$G = 1 - \sum_{i=1}^{K} p_i^2$$

where p_i represents the proportion of samples belonging to class i. A lower Gini value indicates a purer node, containing predominantly one class [83]. Features that result in larger impurity reductions are assigned higher importance values, indicating a stronger influence on the overall triage classification outcomes. To facilitate interpretation, the feature importance values were normalised and visualised as percentages, providing an aggregated view of the most influential clinical indicators used by the model. This analysis aids interpretability by indicating which patient characteristics the model considered most relevant in assigning triage categories.

Local Explainability was achieved using SHAP (SHapley Additive exPlanations) decision and waterfall plots, which trace how each feature influences the predicted triage level for individual cases. SHAP, which is frequently used for explaining AI triage models [29], provides a unified framework for interpreting complex models by assigning each feature a contribution value indicating how it influences an individual prediction relative to the model's baseline expectation.

SHAP is a model-agnostic explainability method derived from cooperative game theory that attributes an importance value (SHAP value) to each input feature, representing its contribution to an individual model prediction [84]. For tree-based models, such as the Random Forest, SHAP provides a specialised implementation known as *TreeExplainer*. The *TreeExplainer* algorithm efficiently computes SHAP values for tree ensembles by leveraging the model's internal structure and accounting for feature dependencies.

The *TreeExplainer* was applied to the trained model to create a multioutput decision plot. This plot was then used to illustrate how the model arrived at a specific triage decision for a single observation. Each feature's SHAP value shows its directional contribution to the prediction, starting from the model's baseline output (the average prediction across all triage categories) and moving toward the final predicted class. This allowed the model's reasoning to be visualised step by step, thereby supporting transparent clinical decision-making

3.2 Practical Approach

3.2.1 Overview

The practical component of this study focused on developing a prototype system to demonstrate the potential real-world application of the AI-assisted triage model. This subsection outlines the overall architecture of the triage system and describes how its components interact to support medical decision-making. The developed prototype is an interactive, web-based application that allows medical personnel to input patient information, obtain real-time triage recommendations, and view explanations for each prediction. The design considerations, interface layout, and user interaction workflow are also described to illustrate the system's usability and potential integration into existing medical workflows.

The web-based prototype was implemented using the *Streamlit* framework within the *PyCharm* Integrated Development Environment (IDE). *PyCharm's* debugging and live preview features supported iterative development and testing of the application on a local server (*localhost*) before preparing it for potential public deployment.

3.2.2 Prototype Architecture

The prototype was designed to be lightweight, intuitive, and accessible in low-resource settings. It was developed using *Streamlit*, an open-source Python framework that supports the rapid deployment of machine learning models as web applications compatible with both mobile and desktop devices [85].

The system was organised into several functional modules to ensure scalability, maintainability, and clarity of implementation. The architecture follows a client–server model that connects the user interface with the machine learning inference engine, supported by auxiliary utility modules. The key modules and their functions are summarised below:

- **app.py** - This serves as the main entry point of the application. It defines the layout and structure of the web interface, manages user interactions, and connects all functional components. Through *app.py*, the user can input patient data, trigger predictions, and view both triage outcomes and SHAP-based explanations.
- **utils.py** - This module provides utility functions that support the triage workflow. It contains routines for validating and cleaning user inputs (*validate_inputs*), formatting patient information for display (*format_patient_summary*), and exporting results to a structured file (*export_results_to_csv*). These functions ensure data consistency and improve usability by managing routine operations within the application.
- **model_interface.py** - This module handles the integration between the trained Random Forest model and the *Streamlit* front end. It contains methods for loading the serialised model (*load_model*), preprocessing patient data (*_get_expected_features* and *_prepare_features*), predicting triage probabilities (*predict*), retrieving model metadata (*get_model_info*), computing global feature importances (*get_feature_importance*), and generating explanations for predictions (*explain*). This separation of tasks allows the model to be easily replaced or retrained without altering the user interface.
- **components** - This is a folder made up of two modules: ***input_forms.py*** and ***result_display.py***. The former deals with accepting inputs from users. It contains a method*, render_patient_input_form,* that organises the information to be completed by the triage officer. The latter deals with how the triage result is presented. It includes a method called *render_prediction_result* that organises the triage result, result explanation, recommendations, patients' summary and the footer with past triage classifications.
- **requirements.txt** - This file specifies the Python package dependencies required for the system to function properly, *including streamlit, scikit-learn, shap, pandas,* and related libraries. This ensures the application can be replicated and deployed consistently across environments.

Together, these components form an end-to-end architecture that connects the model's predictive capabilities to a user-friendly interface. This modular design enables future extensions such as connecting to a remote database or integrating additional triage criteria. Figure 3.5 illustrates the overall system architecture and module interactions.

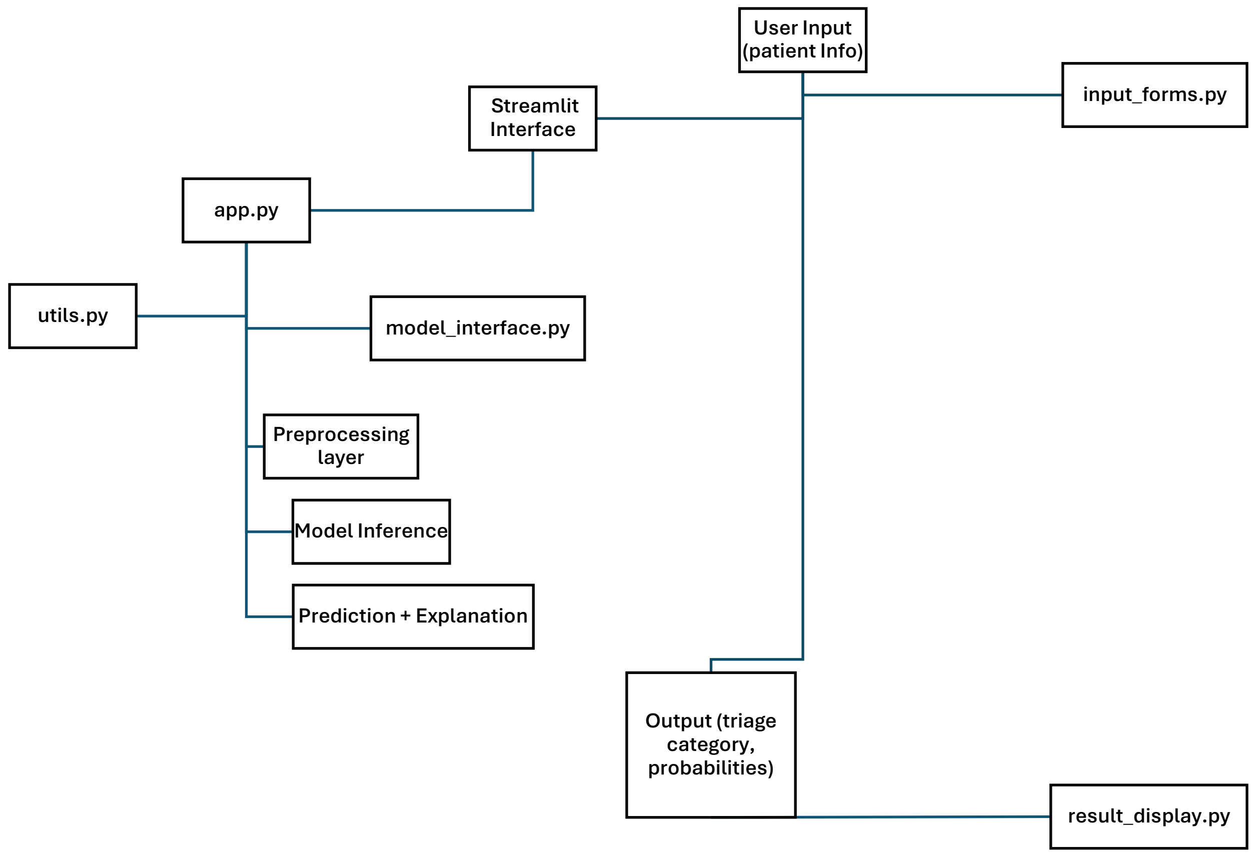


Figure 3.5: Illustration of the system architecture.

3.2.3 User Interaction

The User Interface (UI) design emphasised simplicity, interpretability, and responsiveness, enabling users to visualise model outputs (such as predicted triage levels and SHAP-based explanations) directly through a web interface. The system's workflow begins with user interaction for patient data entry, proceeds through input validation and preprocessing, and culminates in the model's triage prediction and local explanation generation.

When a triage officer accesses the application, the *app.py* module initialises the UI and handles all front-end logic. The webpage is split into two columns; the first column shows the input form for patient data, while the second column is for the display of the triage result. While no result is displayed, a visualisation for the feature importance is made available.

Upon submission of patient data (e.g., vital signs and complaint), the application calls validation functions defined in *utils.py* to ensure that entries are complete, correctly formatted, and within clinically valid ranges. Once the inputs are validated, they are transformed into the appropriate numerical format expected by the Random Forest model.

The processed data are then passed to the model inference layer implemented in *model_interface.py*. This module first loads the trained model (serialised in *.pkl* format) using the *joblib* library. When the user requests a triage prediction, the *predict()* method returns the predicted triage category, and a column chart showing the three categories alongside their probabilities.

Then the *explain()* method computes the associated SHAP values using the *TreeExplainer* algorithm. The SHAP multioutput decision plot is rendered in real time, illustrating how specific features contributed positively or negatively to the predicted triage category. This interpretability feature allows medical personnel to understand the rationale behind each recommendation.

The result, which includes the patient summary and predicted triage level, can be exported using the *export_results_to_csv()* utility for documentation or audit purposes.

On the sidebar of the web page, the method *get_model_info()* displays essential metadata about the model, such as the name and training dataset characteristics. Short information on 'how to use the system' is also provided here.

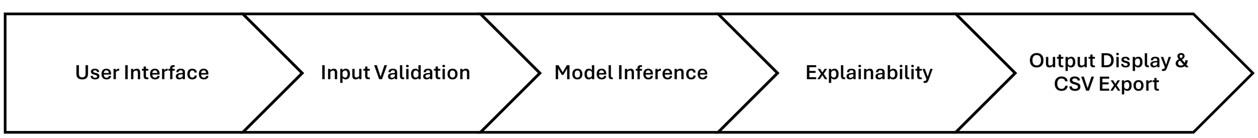


Figure 3.6: Workflow of the Triage Prototype

This practical implementation served to validate the model's usability and highlight how AI-assisted triage could integrate into existing clinical workflows. The *Streamlit*-based web interface was selected for its lightweight architecture and fast loading times, characteristics that align well with the limited internet bandwidth typical of LRS. Furthermore, the interface's simple navigation and minimal data input requirements facilitate ease of use and rapid triage processing, which can contribute to reducing patient waiting times and improving the overall efficiency of care delivery. Figure 3.6 above illustrates the logical workflow of the system from data input to triage output.

4 Results & Discussion

Building on the methodology described in Chapter 3, this chapter presents the results of developing and evaluating an AI-driven triage system for healthcare settings in Nigeria. The analyses follow the structured approach previously outlined, beginning with an exploratory data analysis to summarise key characteristics and patterns in the dataset. This is followed by a comparison of machine learning models, an evaluation of training outcomes for the selected model, and an interpretability assessment to understand its decision-making process. The chapter concludes with the implementation phase, demonstrating how the model can be deployed within real-world healthcare workflows in resource-limited settings.

4.1 Exploratory Data Analysis (EDA) Results

As outlined in Section 3.1.2, the dataset used for this study was generated synthetically, ensuring consistency and uniformity with Nigerian EDs. This section presents the EDA conducted to assess data quality and identify key trends relevant to triage prediction.

	sex	mode_of_arrival	chief_complaint	active_bleeding	AVPU_scale	pregnancy	Triage_Category
count	600	600	600	600	600	600	600
unique	2	3	12	2	4	2	3
top	Male	Private vehicle	Difficulty breathing	No	Alert	No	emergency
freq	307	249	68	499	351	539	200

Table 4.1: Descriptive statistics for the categorical data.

The summary for the categorical variables provided counts, unique category counts, and mode frequencies for features such as *sex*, *mode of arrival*, and *AVPU scale*. The results indicated that most patients were *alert* and arrived in a private vehicle, consistent with typical clinical distributions in RLS.

The numeric summary provided measures of central tendency (mean, median) and dispersion (standard deviation, minimum, and maximum) for key physiological indicators such as age, heart rate, respiratory rate, systolic and diastolic blood pressure, temperature, and oxygen saturation.

	age	resp_rate	heart_rate	systolic_bp	diastolic_bp	temperature	oxygen_sat
count	600.000000	600.000000	600.000000	600.00000	600.000000	600.00000	600.000000
mean	50.545000	18.680000	85.471667	121.98500	81.600000	37.52750	94.131667
std	17.471885	3.917462	13.207192	9.13147	8.107376	0.54671	3.853017
min	18.000000	12.000000	60.000000	104.00000	62.000000	36.40000	81.000000
25%	36.000000	16.000000	75.000000	115.00000	75.000000	37.10000	92.000000
50%	50.000000	18.000000	84.000000	120.00000	81.000000	37.50000	95.000000
75%	65.000000	21.000000	95.000000	128.00000	87.000000	37.90000	97.000000
max	80.000000	27.000000	116.000000	154.00000	103.000000	39.20000	100.000000

Table 4.2: Descriptive Statistics for the Numerical Data

The generated dataset displayed realistic clinical ranges: for example, heart rate values typically ranged between 60 and 116 beats per minute, and systolic blood pressure values between 104 and 154 mmHg, reflecting plausible physiological variation across triage levels. The mean temperature (approximately 37.5°C) and oxygen saturation (around 94%) also aligned with expected normal values, while the wider spread in respiratory rate and blood pressure indicated diversity among simulated patient conditions. These patterns suggest that the data generation process effectively captured the heterogeneity expected in real-world triage scenarios, ensuring that subsequent model training and evaluation were based on clinically coherent distributions.

The boxplots in Figure 4.1 below illustrate the variation in key clinical indicators, showing that synthetic patient data maintains realistic physiological patterns across triage levels. Higher heart rate, respiratory rate, and temperature values correspond with the emergency group, while oxygen saturation decreases, reflecting plausible triage dynamics

The plots in Figure 4.2 below illustrate the relative frequency of categorical variables. Overall, the distributions align with realistic triage patterns - emergency cases were more likely to arrive by ambulance or present as unresponsive, while non-urgent cases were more often walk-ins or alert. The balanced spread of counts across variables further supports the internal validity of the synthetic dataset.

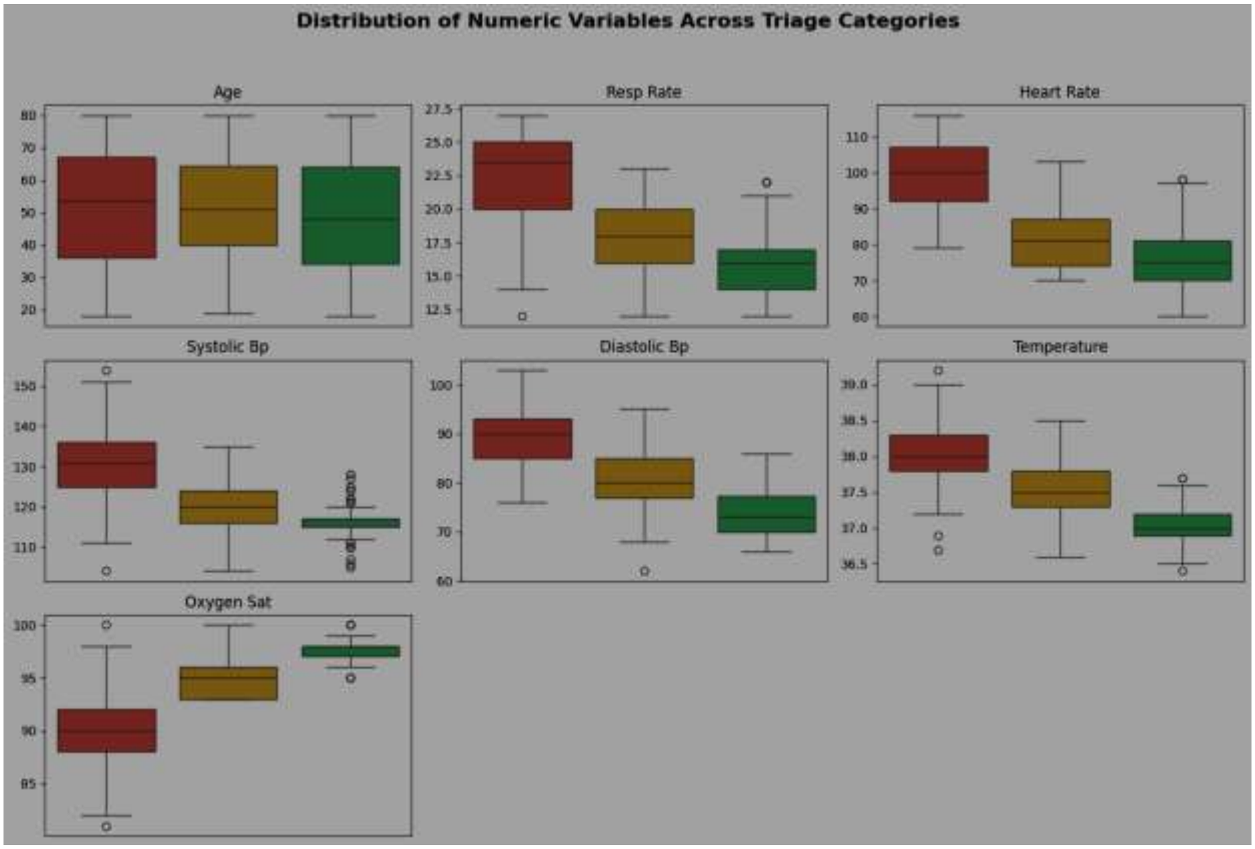


Figure 4.1: Distribution of numeric variables across triage categories

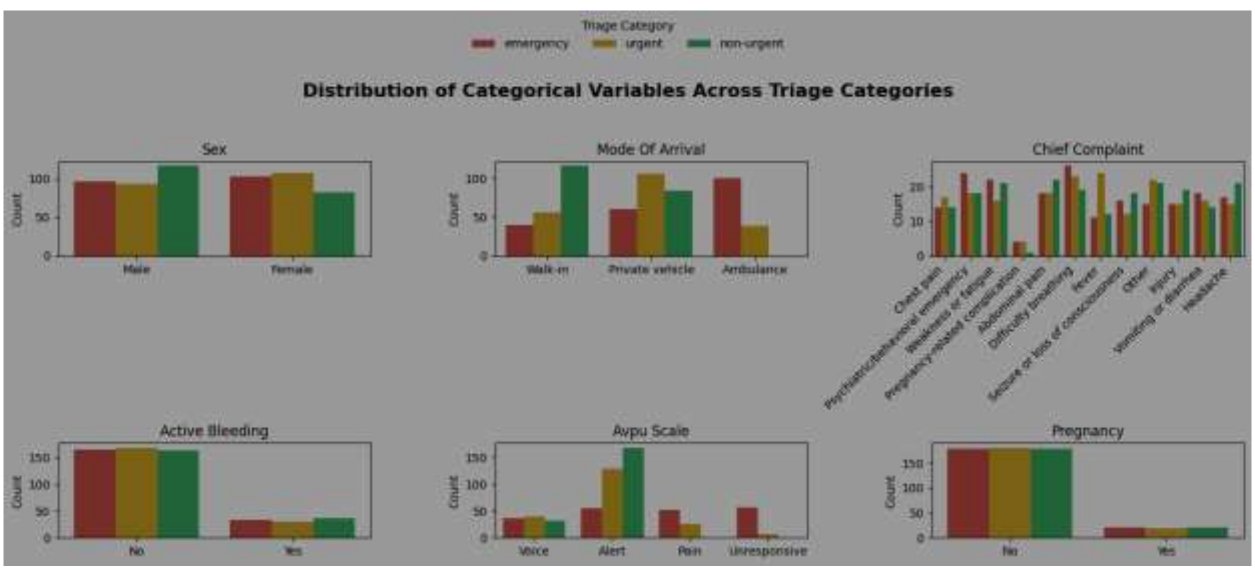


Figure 4.2: Distribution of the categorical variables across triage categories.

4.2 Model Comparison

Following the modeling framework discussed in Chapter 3, multiple algorithms were trained and evaluated to determine the most effective model for triage classification. The following tables describe the performance of the algorithms.

```
Multi Layer Perceptron (MLP) Classifier: Model Evaluation on Testing Set
Classification Report:
              precision    recall  f1-score   support

           0       0.97      0.95      0.96        40
           1       0.89      0.82      0.86        40
           2       0.86      0.95      0.90        40

    accuracy                           0.91       120
   macro avg       0.91      0.91      0.91       120
weighted avg       0.91      0.91      0.91       120

Confusion Matrix:
[[38  2  0]
 [ 1 33  6]
 [ 0  2 38]]
```

Table 4.3: Performance evaluation of the MLP classifier on the test dataset

The first algorithm, Multi Layer Perceptron (MLP), had a high accuracy of 91%. Based on the confusion matrix, the model correctly classified 38 out of 40 *Emergency(0)* and *Non-Urgent(2)* cases, and 33 out of 40 *Urgent (1)* cases. The *support* (40) indicates the number of true instances for each category. The numbers of misclassified observations are recorded as the off-diagonal values; these are the False Positives and False Negatives (FN). Most misclassifications occurred within the *Urgent* category, where six instances were predicted as *Non-Urgent* and one as *Emergency*. This pattern suggests that urgent cases exhibit overlapping clinical characteristics with both higher and lower severity levels, making them more challenging to distinguish [30]. Importantly, only two Emergency cases were under-classified, suggesting the model maintains a conservative bias toward over-triage rather than under-triage, which is preferable in healthcare settings [38].

To validate the choice of Random Forest, a Gradient Boosting model (XGB) was also trained using identical training data and evaluation metrics. While the XGB classifier achieved an equal overall accuracy (93%) with the RF, it requires substantially more computational time and offers less interpretability [79]. Given the emphasis on transparency and real-time decision support in clinical settings, Random Forest was selected as the final model for prototype implementation.

```
eXtreme Gradient Boosting (XGB) Classifier: Model Evaluation on Testing Set
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99        40
           1       0.92      0.85      0.88        40
           2       0.86      0.95      0.90        40

    accuracy                           0.93       120
   macro avg       0.93      0.92      0.93       120
weighted avg       0.93      0.93      0.93       120

Confusion Matrix:
[[39  1  0]
 [ 0 34  6]
 [ 0  2 38]]
```

Table 4.4: Performance evaluation of the XGB classifier on the testing set.

```
Random Forest (RF) Classifier: Model Evaluation on Testing Set
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99        40
           1       0.92      0.88      0.90        40
           2       0.88      0.95      0.92        40

    accuracy                           0.93       120
   macro avg       0.93      0.93      0.93       120
weighted avg       0.93      0.93      0.93       120

Confusion Matrix:
[[39  1  0]
 [ 0 35  5]
 [ 0  2 38]]
```

Table 4.5: Performance evaluation of the RF classifier on the testing set.

4.3 Model Training Outcomes

Model performance was initially summarised using scikit-learn's classification_report, which provides per-class Precision, Recall, and F1-scores. To further assess the classifier's discriminative ability, Receiver Operating Characteristic (ROC) curves and corresponding Area Under the Curve (AUC) values were computed separately using the *roc_auc_score*() function, adopting a one-vs-rest (OvR) approach for the multi-class triage problem. The table below shows the results of tuning the RF model.

```
RF Classifier: Model Evaluation on Testing Set
Best parameters:
{'n_estimators': 500, 'min_samples_split': 2, 'min_samples_leaf': 1, 'max_features': 'log2', 'max_depth': 10}
Classification Report:
              precision    recall  f1-score   support

           0       1.00      0.97      0.99        40
           1       0.95      0.88      0.91        40
           2       0.89      0.97      0.93        40

    accuracy                           0.94       120
   macro avg       0.94      0.94      0.94       120
weighted avg       0.94      0.94      0.94       120

Confusion Matrix:
[[39  1  0]
 [ 0 35  5]
 [ 0  1 39]]
```

Table 4.4: Evaluation of the tuned RF classifier on the testing dataset

The Random Forest (RF) classifier was optimised using a randomised search approach, which identified the following best parameters: *500 estimators, a maximum tree depth of 10, a minimum sample split of 2, a minimum of 1 sample per leaf*, and the use of the *log2 feature selection* strategy for node splits. These parameters provided a strong balance between model complexity and generalisation, allowing the classifier to achieve robust performance without overfitting.

The classification report demonstrates excellent predictive performance across all triage categories, with an overall accuracy of 94%. Specifically, the model achieved precision and recall scores above 0.88 for all classes, indicating both high reliability in predicting *True Positives* and a low rate of *False Negatives*. The *Emergency* class (0) achieved perfect precision (1.00) and a high recall (0.97), suggesting the model rarely misclassified non-emergency cases as emergency, while successfully identifying nearly all true emergencies. The *Urgent* class (1) had slightly lower recall (0.88), reflecting occasional confusion with the neighbouring classes, whereas the *Non-Urgent* class (2) maintained high precision (0.89) and recall (0.97), showing consistent identification of less critical cases. The macro and weighted averages (both 0.94) further confirm balanced performance across the three categories.

The *F1-score* provides a harmonic balance between *Precision* and *Recall*, offering a single measure of predictive consistency. In this study, the dataset was balanced across triage categories; therefore, the F1-score closely paralleled the overall *Accuracy* metric. However, it remained valuable for identifying whether the classifier maintained comparable performance in detecting and correctly predicting each triage level. This ensured that high accuracy was not achieved at the expense of misclassifying clinically critical cases.

The confusion matrix provides additional insight into the classification behaviour of the model:

True \ Pred	**Emergency (0)**	**Urgent (1)**	**Non-Urgent (2)**
Emergency (0)	39	1	0
Urgent (1)	0	35	5
Non-Urgent (2)	0	1	39

Table 4.5: Confusion matrix of the tuned model

The RF classifier correctly identified nearly all *Emergency (39/40)* and *Non-urgent (39/40)* cases. Misclassifications were minimal and primarily occurred within the *Urgent* category, where five urgent cases were incorrectly labelled as non-urgent and one non-urgent case was labeled as urgent.

The performance of the tuned RF model was compared with the untuned baseline model (discussed in the previous section), in which all parameters were set to their default values, except the random state fixed at 42. The tuned model achieved only a marginal improvement, correctly identifying one additional Non-Urgent case. Given that the misclassification of a Non-Urgent case as Urgent carries minimal clinical consequence, this gain is not considered practically significant. Consequently, the results suggest that the default RF model already provides near-optimal performance, and extensive hyperparameter tuning may offer limited benefit relative to its computational cost in this context.

Furthermore, the Figure 4.3 below presents the ROC curves for the RF classifier, illustrating its ability to distinguish between the three triage categories. The ROC curve plots the true positive rate (recall) against the false positive rate at varying classification thresholds, providing a graphical representation of the model's diagnostic performance. The AUC quantifies this performance, where a value of 1.0 represents perfect classification and 0.5 indicates no discriminative power.

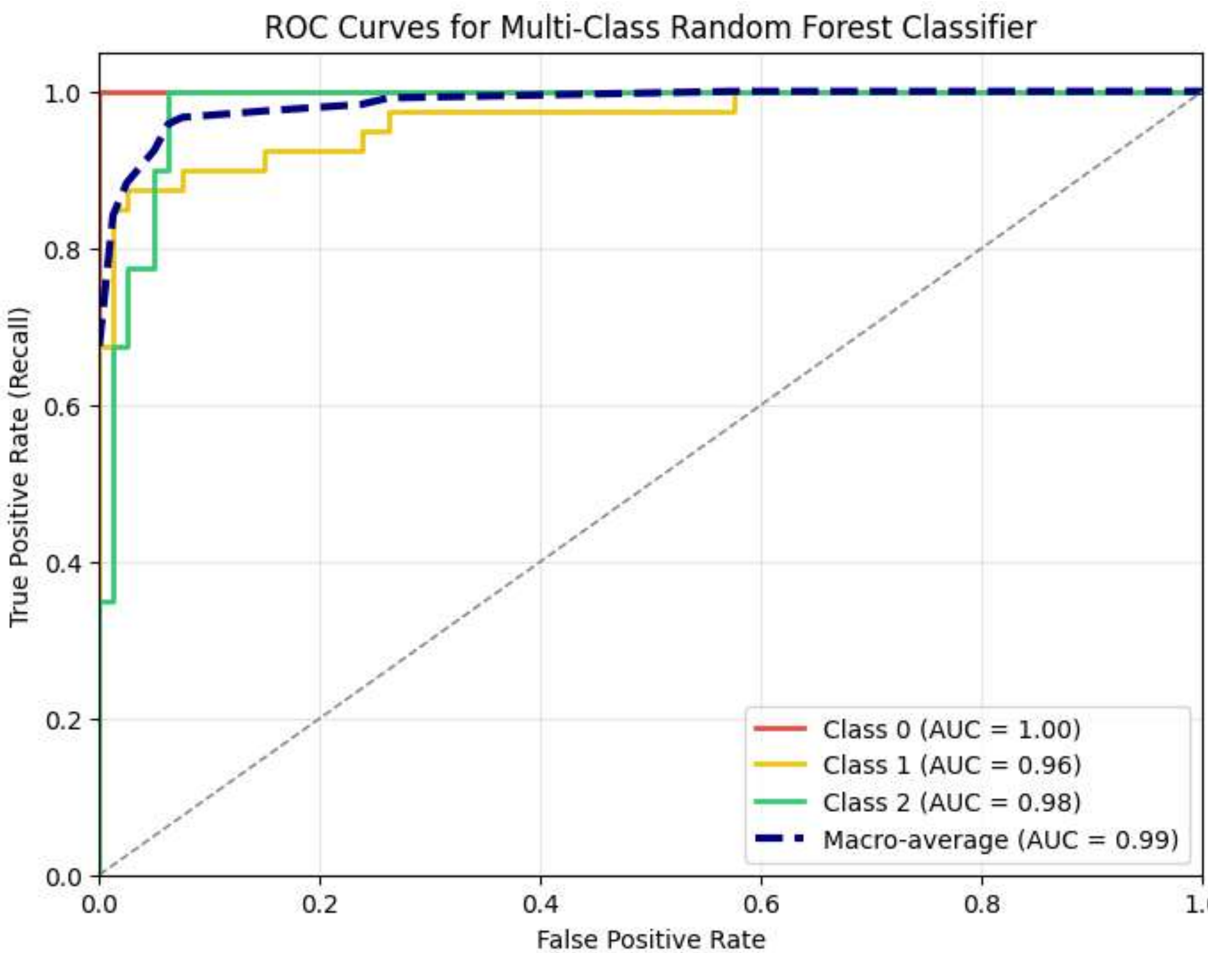


Figure 4.1: ROC Curve showing Trade-off between TP and FP

As shown in the figure, the model achieved *AUC values* of *1.00* for *Emergency (Class 0), 0.96* for *Urgent (Class 1),* and *0.98 (Class 2)* for *Non-Urgent* cases, yielding a *macro-average AUC of 0.99.* These results demonstrate that the RF classifier effectively distinguishes between all triage categories with minimal overlap. The perfect AUC for the *Emergency* class highlights the model's exceptional ability to identify critical cases - a vital property in clinical triage applications where under-classifying emergencies could have severe consequences. The slightly lower AUC for the Urgent class (0.96) suggests minor overlap between urgent and non-urgent presentations, which is consistent with the confusion matrix findings. Overall, the near-perfect macro-average AUC indicates that the model achieved strong and balanced discriminative capacity across all classes, reinforcing its suitability for AI-assisted triage in resource-limited healthcare settings such as Nigeria.

4.4 Explainability Analysis

Following model evaluation, an interpretability assessment was conducted to gain insights into how the Random Forest classifier derived its triage predictions. This step is essential for ensuring that the model's decision-making process aligns with established clinical reasoning and can be trusted for deployment in real-world emergency care settings.

4.4.1 Feature Importance Analysis

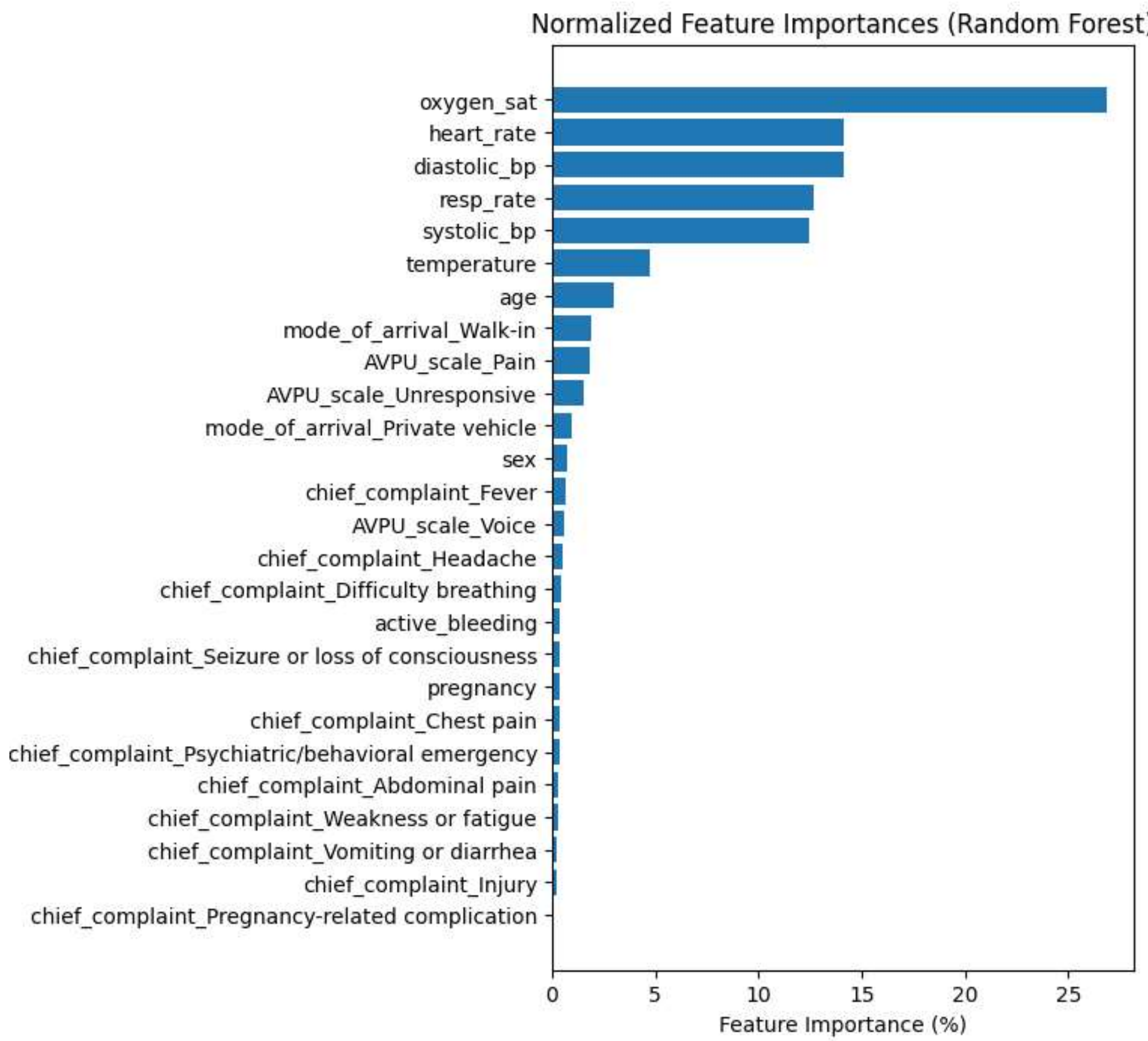


Figure 4.2: Normalised feature importances derived from the Random Forest classifier

Figure 4.4 above presents the normalised feature importances computed from the RF model. All variables used for model training are included. As shown, physiological variables such as oxygen saturation, heart rate, blood pressure and respiratory rate emerged as the most influential predictors of triage category. Among non-vital features, temperature and age contributed moderately to classification performance, while features such as mode of arrival and the AVPU scale (representing patient consciousness level) had smaller yet non-negligible effects.

The predominance of vital signs in the feature ranking aligns with conventional triage practices, where physiological instability serves as a key determinant of emergency priority. In contrast, chief complaint variables (e.g., fever, headache, chest pain) exhibited minimal importance, suggesting that the model relied more heavily on objective physiological data than on subjective symptom descriptions.

4.4.2 SHAP-Based Model Interpretation

To further interpret model behaviour at class-specific levels, SHAP (SHapley Additive exPlanations) values were generated for the Random Forest model. The SHAP multioutput decision plot in Figure 4.5 illustrates the relative impact of the top features across the three triage classes for a given instance.

In the Figure, the horizontal axis shows the model output value (class probability), and the three lines correspond to the outputs for the three classes (Emergency, Urgent, Non-urgent). The centre grey line represents the expected output value, computed as the average predicted probability for that class across the training set. The features are ordered by descending importance. The importance is calculated over the observations plotted. This is usually different from the importance ordering for the entire dataset [84] . At the top of the plot, each line strikes the x-axis at its corresponding observation's predicted value. This value determines the colour of the line on a spectrum, and for this reason, it was not intuitive to colour the lines according to triage category. Moving from the bottom of the plot to the top, SHAP values for each feature are added to the model's base value. This shows how each feature contributes to the overall prediction. At the bottom of the plot, the observations converge at the expected value.

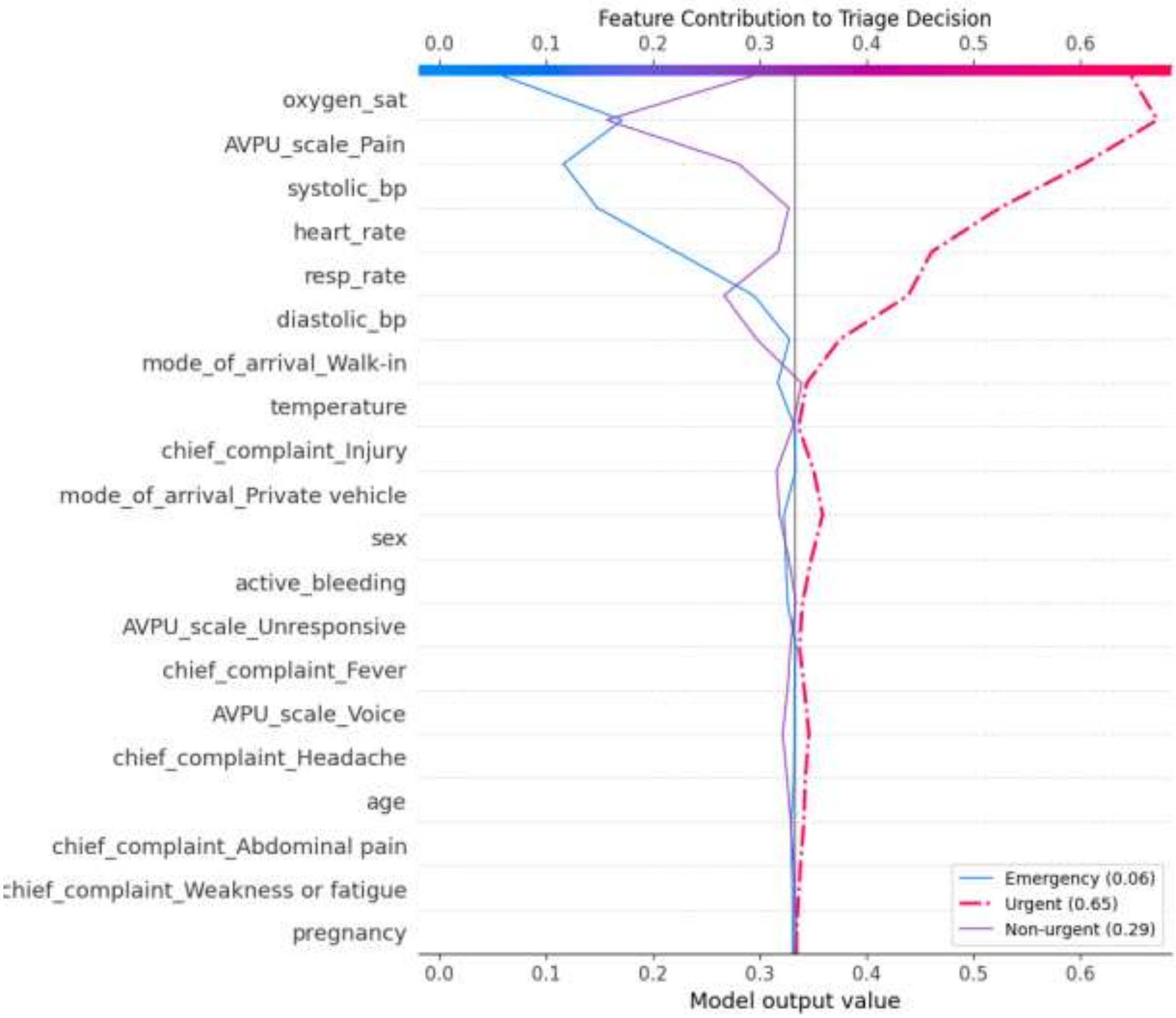


Figure 4.3: SHAP decision plot showing the contribution of variables to the Urgent triage classification

In the above image, a correctly predicted class is examined. The result revealed that variables like *AVPU_scale_pain, systolic_bp, heart_rate* and *resp_rate* contributed positively to the model's decision to classify the patient as *Urgent*. This influence is illustrated more clearly in the waterfall plot presented in Figure 4.6, where each feature's value and corresponding SHAP contribution are displayed.

The plot begins at the model's baseline prediction (the expected value of all predictions, $E[f(X)] = 0.334$) and successively adds or subtracts feature contributions to arrive at the final predicted probability of 0.648 for the *Urgent* class. Red bars indicate features increasing class probability, and blue bars indicate features decreasing it. In this example, the patient's systolic blood pressure (121 mmHg), AVPU scale (Pain response), diastolic blood pressure (80 mmHg), and heart rate (84 bpm) were the most influential features pushing the model toward classifying the case as *Urgent*. These parameters collectively reflect moderate physiological deviation - indicative of a patient requiring timely medical attention but not in immediate life-threatening danger.

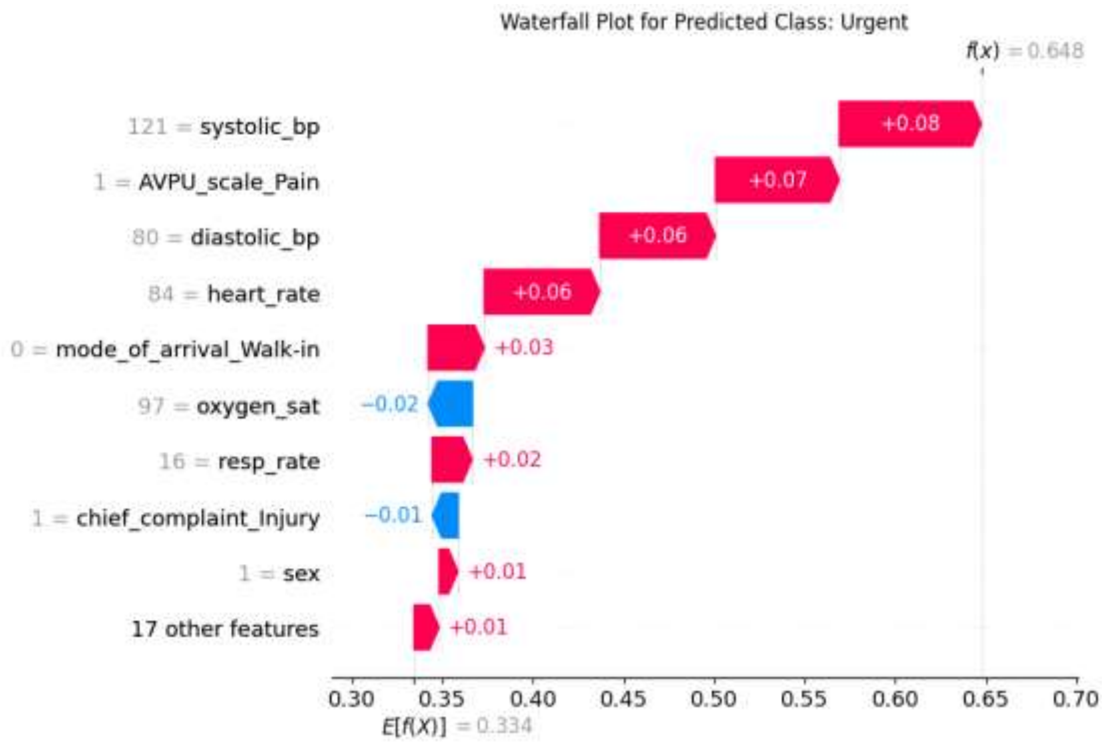


Figure 4.4: SHAP Waterfall Plot for Predicted Class – Urgent.

Conversely, features such as oxygen saturation (97%) and chief complaint (Injury) slightly decreased the probability of an *Urgent* classification, nudging the model marginally toward a less severe category. These negative contributions suggest that normal oxygenation and a relatively stable presenting complaint tempered the model's overall severity assessment.

Going further to examine a misclassified output, Figure 4.10 below illustrates the SHAP multioutput decision plot. In this instance, the true label was *Urgent*, but the model predicted *Non-Urgent* with a probability of 0.73. The plot displays the contribution of each feature to the model's output probabilities for all three classes - *Emergency* (blue), *Urgent* (purple dashed line), and *Non-Urgent* (pink solid line).

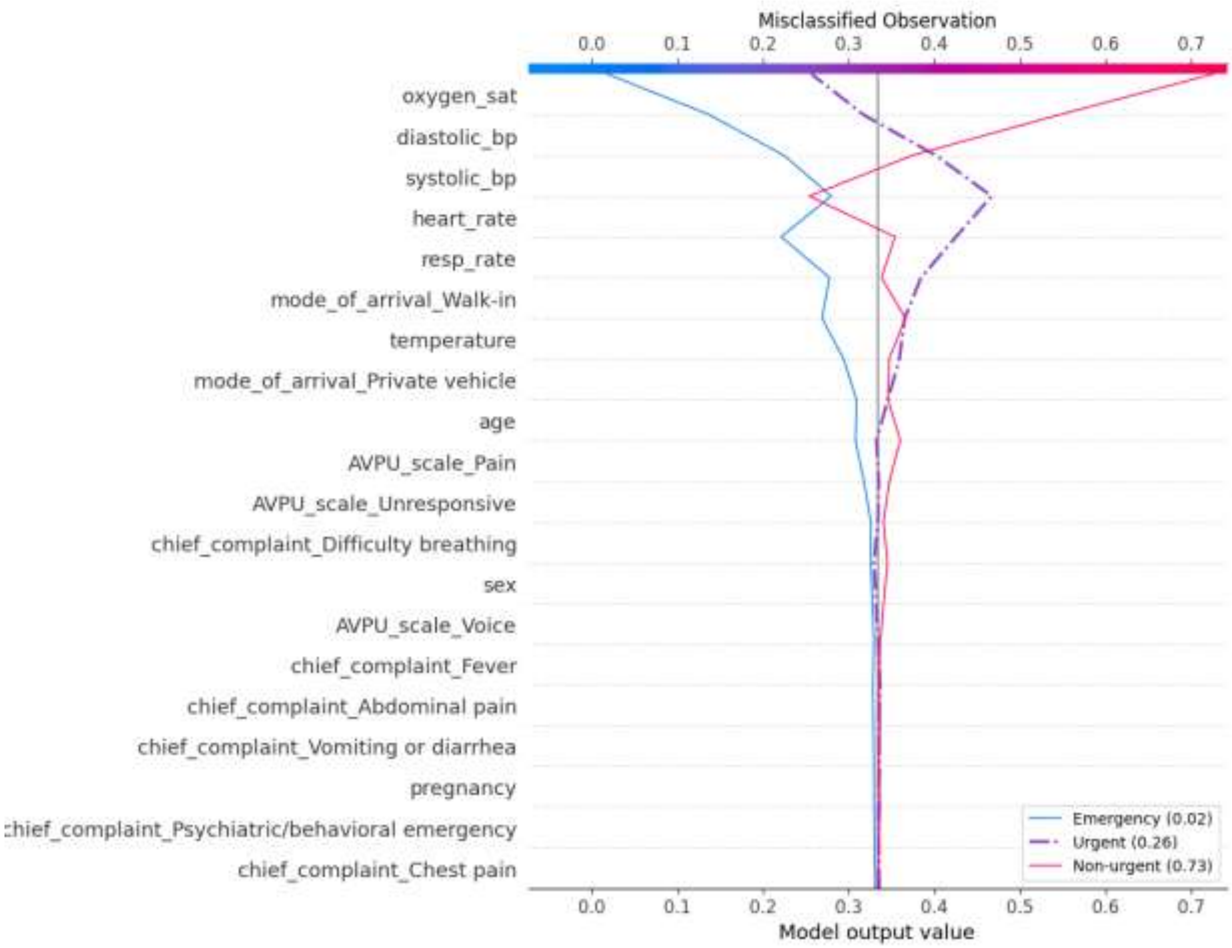


Figure 4.5: SHAP waterfall plot showing the model's reasoning for a misclassified observation

Figures 4.8 below show the waterfall plots for the misclassified observation. Here, we can see both the feature values and the importance values assigned by SHAP. The features with the largest impact on this misclassification include *oxygen saturation, diastolic blood pressure* and *systolic blood pressure*. In this case, relatively normal values for oxygen saturation and blood pressure appear to have strongly influenced the model toward the *Non-Urgent* category, as indicated by their substantial rightward pull of the pink curve. Conversely, mild elevations in heart rate and respiratory rate contributed modestly to the *Urgent* class probability but were insufficient to counterbalance the influence of normal vital signs.

The left plot shows how relatively normal vital signs (e.g., diastolic BP, systolic BP, oxygen saturation) increased the predicted probability of the Non-urgent class, while the right plot shows how these same features reduced the likelihood of the correct Urgent classification. The baseline values for the Urgent and Non-urgent classes differ slightly, reflecting the independent class-specific probability distributions learned by the model.

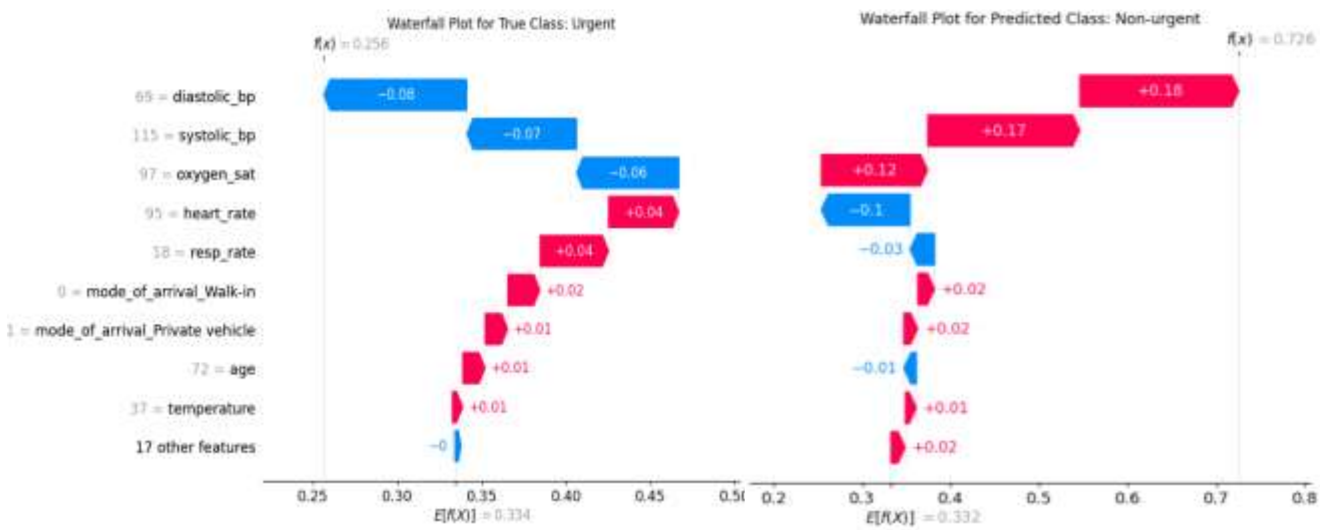


Figure 4.6: SHAP waterfall plots illustrating model reasoning for a misclassified case

This result highlights that the model's reasoning, although data-driven, may at times underestimate urgency when physiological parameters remain within normal ranges, even if other contextual factors such as age or presenting complaint suggest otherwise. In this case, the patient was 72 years old with a primary complaint of difficulty breathing. These features would normally raise concern in clinical triage. However, relatively normal vital signs contributed to a higher predicted probability for the *Non-urgent* class, leading to under-triage. Such cases underscore the need to incorporate clinical oversight or complementary rule-based safeguards when deploying AI triage systems, as the consequences of under-triage generally pose greater clinical risk than over-triage.

This interpretability assessment reinforces the clinical plausibility of the Random Forest model's decision logic, supporting its potential trustworthiness and transparency in resource-limited emergency settings. Both the global feature importance analysis and the SHAP-based local explanations consistently highlight the dominant influence of vital signs in determining triage outcomes. These parameters are objective, rapidly measurable, and strongly associated with patient acuity in clinical practice, aligning the model's internal logic with established triage principles.

4.5 Implementation Results

The triage system is deployed at https://ai-triage-nigeria.streamlit.app/

System testing and validation were conducted to ensure that the AI-assisted triage prototype functioned reliably across its components and produced consistent, interpretable results. This process focused on verifying input handling, model integration, output generation, and User Interface (UI) responsiveness.

Functional testing was carried out to confirm that each module behaved as intended. The tests were designed to assess data validation, model inference, and output consistency. The *utils.py* functions were tested with various edge cases, such as missing, out-of-range, or incorrectly formatted inputs, to confirm that the validation routines correctly identified errors and provided clear feedback to users. The *model_interface.py* module was tested to ensure that the RF model loaded successfully, generated predictions for valid input data, and returned corresponding SHAP explanations without error. The export functionality was also verified to confirm that prediction results and patient summaries were saved correctly in *.csv* format.

The UI (defined in *app.py*) was evaluated interactively to confirm the smooth flow between data entry, model prediction, and explainability visualisation. Tests were performed using representative patient profiles covering each triage category (e.g., *Emergency*, *Urgent*, *Non-Urgent*) to ensure that the visual and textual outputs updated dynamically and accurately reflected model results. The responsiveness of the interface was verified under different devices (desktop and mobile) and browsers (Chrome and Safari) configurations to confirm compatibility.

In addition, system validation was performed to confirm the overall consistency between the predicted triage category and its SHAP-based explanation. Specifically, for each test case, the direction of influence (positive or negative contribution) from relevant features such as *heart rate* or *temperature* was examined to verify that the SHAP interpretation aligned with clinical reasoning.

Overall, the testing process confirmed that the prototype was technically stable, user-friendly, and capable of providing interpretable triage predictions. These findings support its suitability for demonstration and future adaptation in resource-limited healthcare contexts.

The following screenshots show the implementation results.

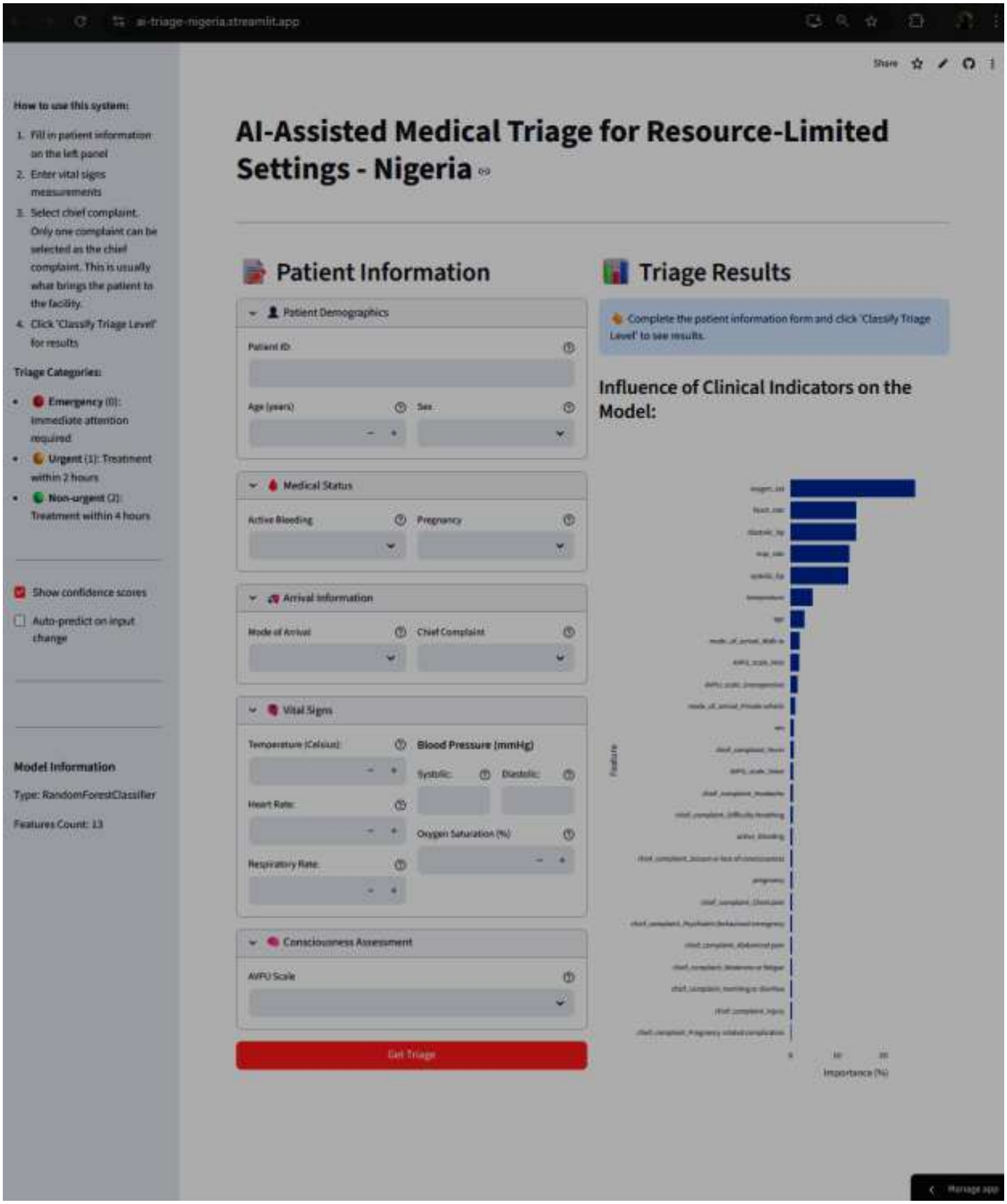


Figure 4.7: Landing page of the AI-Assisted Medical Triage web application

The above figure shows the User Interface for entering patient information and viewing model outputs. The expanded sidebar provides system instructions and triage category definitions. The left column allows triage officers to input demographic details, vital signs, clinical status, and presenting complaints, while the right panel displays the generated triage result. In addition, the feature importance rankings can also be seen, immediately informing clinicians which indicators mostly affect the model's prediction.

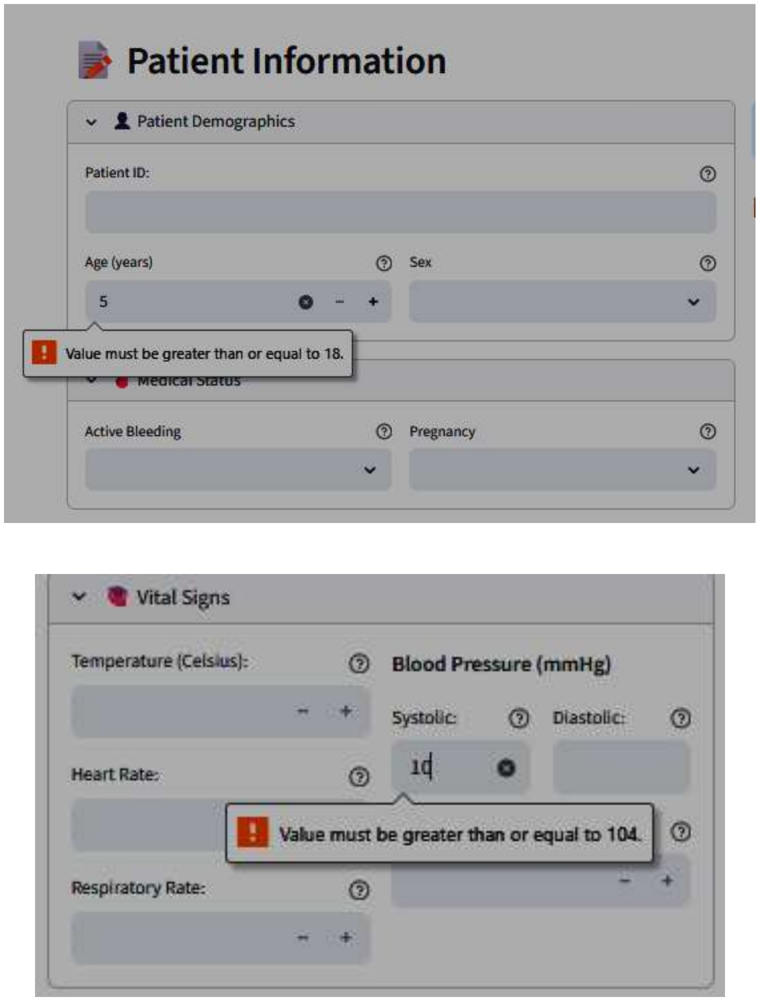


Figure 4.8: Input Validation

The system's handling of out-of-range values is demonstrated to verify that appropriate and informative feedback is provided to users. In the upper image, a paediatric age is entered, prompting the system to indicate that patients must be 18 years or older. In the lower image, an implausible systolic blood pressure value is submitted, and the system responds with a corresponding validation message.

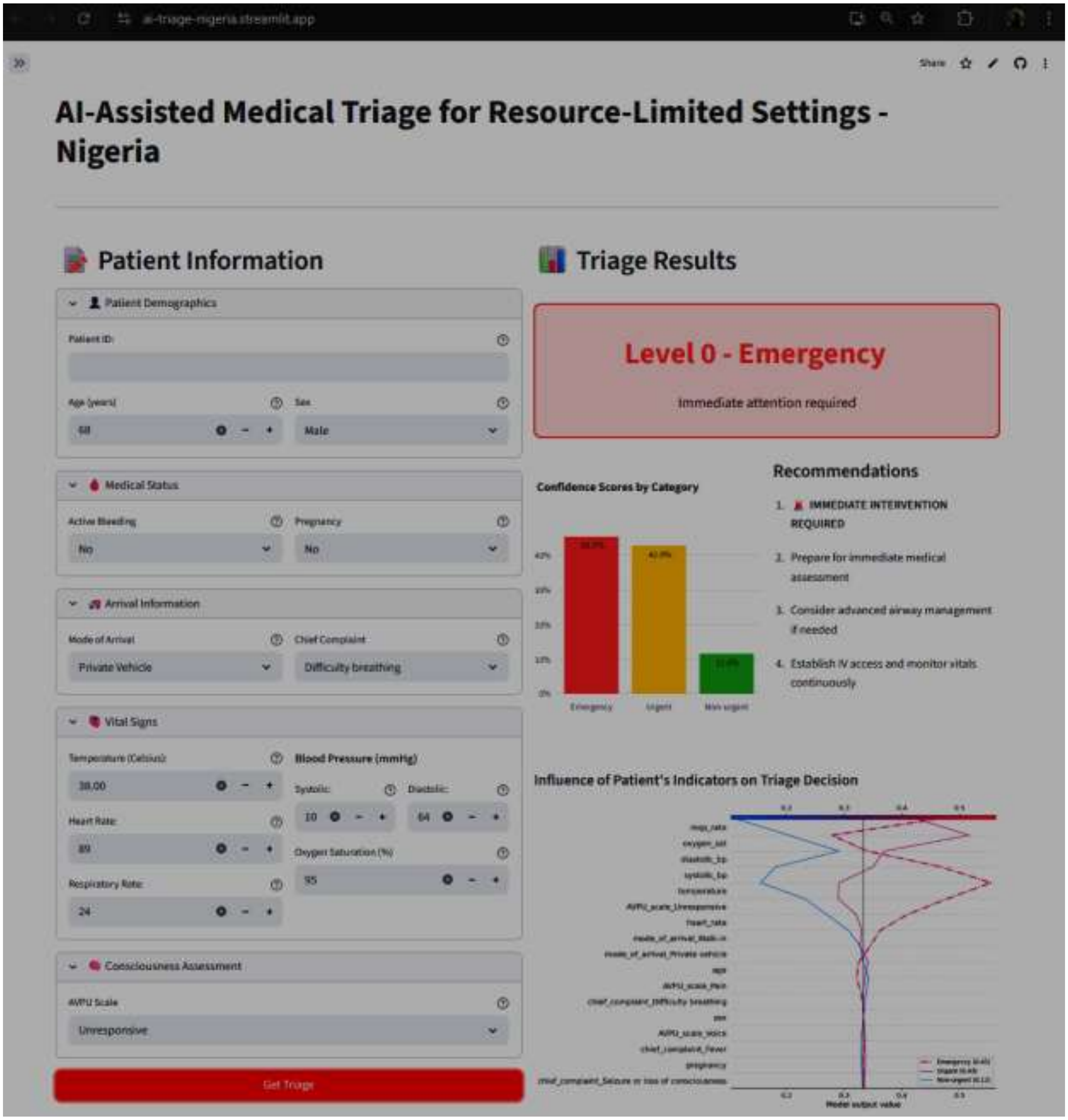


Figure 4.9: Prediction Outcome

The right-hand panel displays the entered patient information, which forms the basis for the predicted triage category shown in the top-left section. The left-hand panel additionally presents the model's confidence scores, recommended triage actions, and an explanation of how key clinical indicators contributed to the prediction.

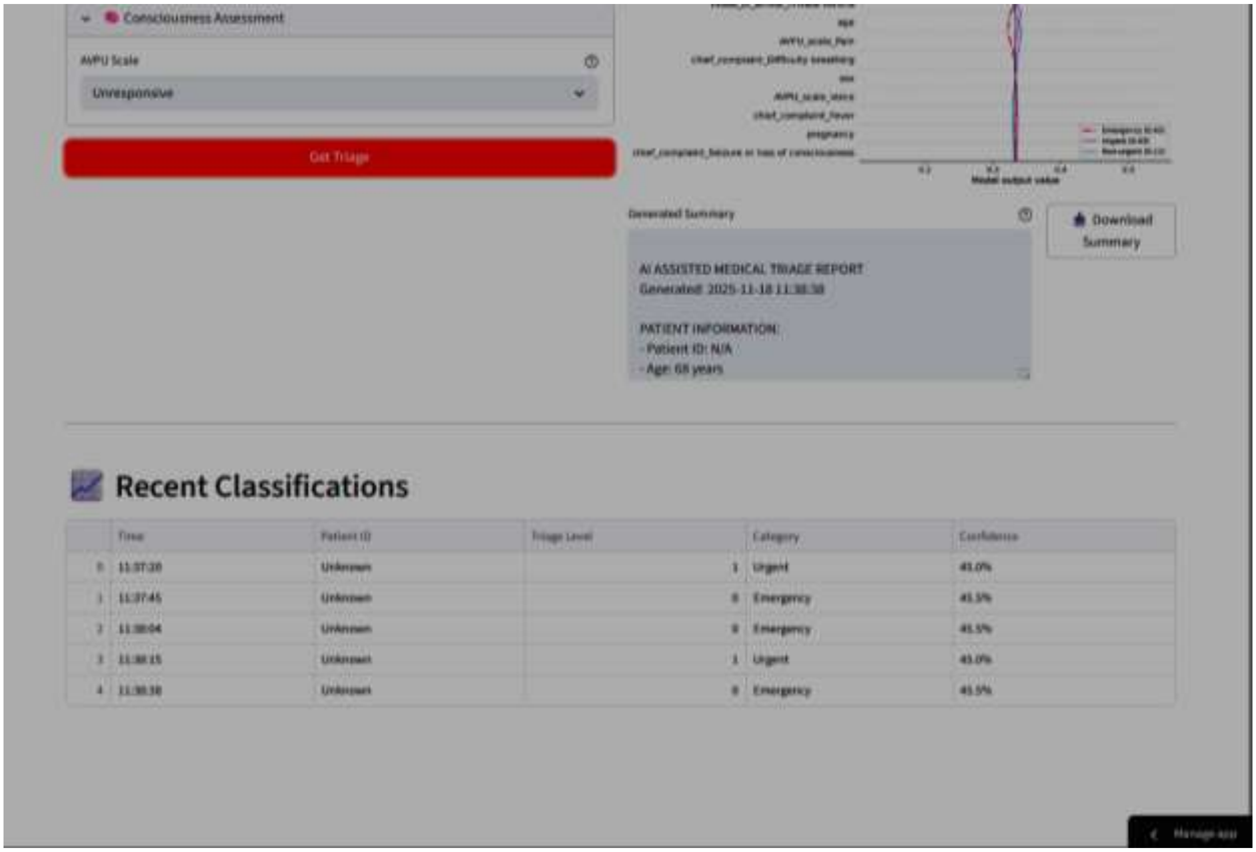
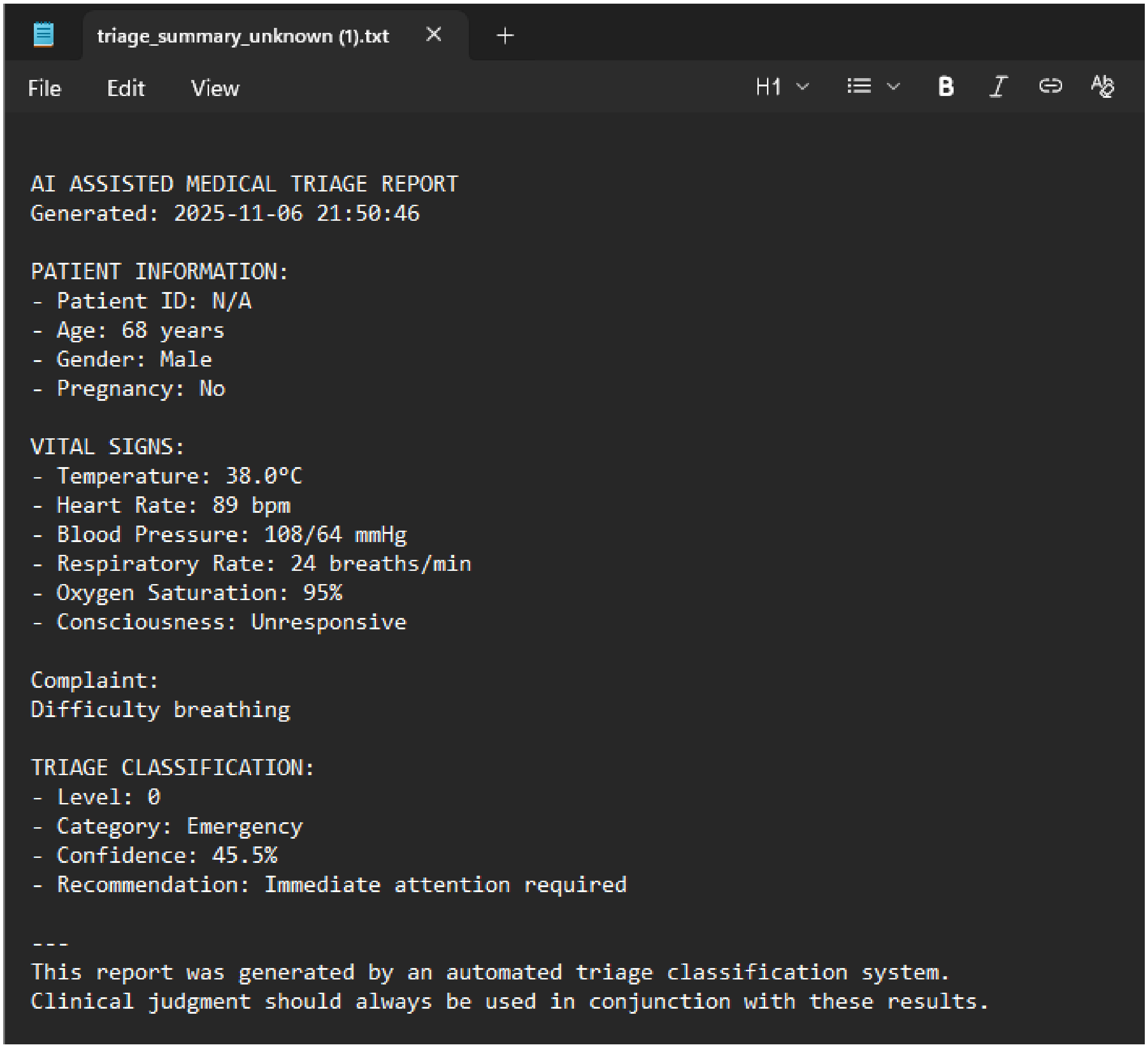


Figure 4.10: Patient Summary and Recent Classifications.

Directly below the SHAP explanation, the interface provides an option to download a patient summary report, which includes the entered clinical information and the corresponding triage outcome. At the bottom of the page, a table displays the five most recent triage classifications, offering users a quick overview of recent system activity.



```
AI ASSISTED MEDICAL TRIAGE REPORT
Generated: 2025-11-06 21:50:46

PATIENT INFORMATION:
- Patient ID: N/A
- Age: 68 years
- Gender: Male
- Pregnancy: No

VITAL SIGNS:
- Temperature: 38.0°C
- Heart Rate: 89 bpm
- Blood Pressure: 108/64 mmHg
- Respiratory Rate: 24 breaths/min
- Oxygen Saturation: 95%
- Consciousness: Unresponsive

Complaint:
Difficulty breathing

TRIAGE CLASSIFICATION:
- Level: 0
- Category: Emergency
- Confidence: 45.5%
- Recommendation: Immediate attention required

---
This report was generated by an automated triage classification system.
Clinical judgment should always be used in conjunction with these results.
```

Figure 4.11: Generated Summary

The automatically generated report includes the date and time the triage assessment was completed, alongside a summary of the patient's input data and predicted triage category. A disclaimer is provided at the bottom of the report to guide appropriate clinical use.

4.5.1 Deployment Considerations

The deployment of the AI-assisted triage system in real healthcare settings requires careful attention to infrastructure, usability, and data governance. While the prototype was developed and tested in a controlled environment, real-world implementation in Nigerian healthcare facilities would depend on factors such as internet reliability, hardware availability, staff training, and compatibility with existing patient record systems.

Streamlit's lightweight, web-based architecture supports flexible deployment options, including hosting on local servers or hospital intranets, thereby reducing dependence on external cloud services. While deployment on the free Community Cloud tier is convenient for development and demonstration, it comes with limitations: applications automatically enter a sleep state after 12 hours of inactivity to conserve resources. Ensuring continuous availability would therefore require configuration as a private app shared with authorised healthcare providers or migration to a paid hosting solution such as *Streamlit in Snowflake*.

Additional considerations include ensuring compliance with patient data privacy regulations, enabling offline or low-connectivity access where possible, and establishing mechanisms for periodic model retraining using locally collected clinical data to maintain performance and relevance.

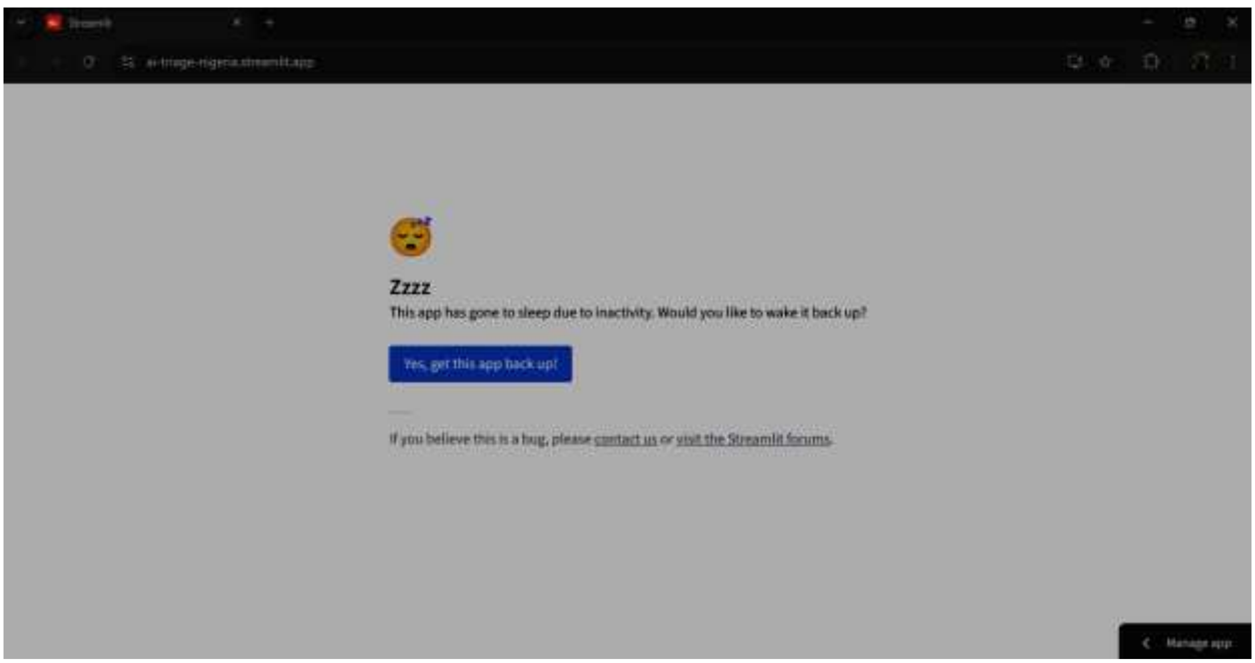


Figure 4.12: Triage application in a sleep state when hosted on the free Streamlit Community Cloud

5 Conclusion

This thesis set out to develop an AI-based triage system tailored for low-resource settings, with a particular focus on the Nigerian healthcare system. Nigeria's health sector continues to face significant challenges, including shortages of trained personnel, recurrent disease outbreaks, and inadequate infrastructure. These constraints contribute to patient overcrowding and elevated morbidity and mortality rates. Triage systems play a critical role in prioritising patients who require urgent care, and while traditional tools such as the Manchester Triage System (MTS) and Emergency Severity Index (ESI) remain widely used in high-income countries, AI-driven approaches are increasingly being adopted, with ongoing research aimed at improving their performance.

The solution proposed in this study draws inspiration from traditional low-resource triage tools such as the Interagency Integrated Triage Tool (IITT) and the South Africa Triage Scale (SATS), as well as the Wellvis COVID-19 triage application, which demonstrated practical value in Nigeria. The prototype developed here is based on a Random Forest model trained on synthetic data, achieving an accuracy of 94% on test cases. It was deployed as a lightweight web application using Streamlit and integrates Explainable AI through SHAP visualisations to enhance transparency and promote trust among healthcare workers.

This work contributes to advancing AI inclusivity in low-resource settings and aligns with the National Digital Health Initiative's objectives for strengthening Nigeria's digital health infrastructure. Overall, the study demonstrates both the necessity and the feasibility of AI-supported triage systems in Nigeria, offering a proof of concept that bridges traditional triage methods with modern AI techniques. It also situates the prototype within the broader global context by comparing traditional and AI-based triage approaches and analysing current applications of AI in healthcare across low-resource environments. Taken together, the findings highlight a significant unmet need for AI-enhanced triage solutions and point toward a promising direction for improving emergency care delivery in Nigeria.

5.1 Limitations and Future Work

This study was constrained primarily by the limited availability of real-world clinical data, which reflects the current state of poor digitisation within the Nigerian healthcare system. Retraining and validating the model using de-identified real patient data would greatly enhance its reliability and clinical relevance. Additionally, benchmarking the system against established low-resource triage tools such as the IITT or SATS would provide a more rigorous assessment of its comparative strengths and potential areas for improvement.

Although the Random Forest model demonstrated strong performance, future research could explore the integration of Natural Language Processing techniques and more advanced clinical AI models. Incorporating unstructured data, such as clinical notes, may enrich the model's decision-making capabilities and improve predictive accuracy. Regardless of the modelling approach, it will be important to maintain transparency through appropriate Explainable AI methods to support clinical trust and accountability.

Prior studies on emergency care in Nigeria indicate that many health workers rely primarily on personal experience rather than formal triage tools. As such, usability evaluations with emergency department staff are essential to assess ease of use, understand how SHAP explanations influence trust, and inform iterative improvements through co-design sessions. Future research should also investigate patient perceptions, including satisfaction with the triage categories assigned to them, to ensure that the system is not only clinically effective but also accepted by its intended users.

6 Bibliography

[1] D. D. Sasu, “Health in Nigeria - statistics & facts _ Statista.” Accessed: Oct. 07, 2025. [Online]. Available: https://www.statista.com/topics/6575/health-in-nigeria/#topicOverview

[2] R. A. Aranmolate, “The State of Healthcare in Nigeria: Challenges and Opportunities.” [Online]. Available: https://gml.com.ng/the-state-of-healthcare-in-nigeria-challenges-and-opportunities/

[3] Christian Connections for International Health, “An Overview of the Health System in Nigeria and the contributions and role of faith-based health providers in the overall health system,” Mar. 2021. Accessed: Oct. 16, 2025. [Online]. Available: https://www.ccih.org/wp-content/uploads/2017/09/HSS-FBO-Report-Nigeria.pdf

[4] D. B. Olawade, I. O. Ojo, E. O. Oisakede, V. I. Joel-Medewase, and O. Z. Wada, “Artificial intelligence in Nigerian oncology practice: A qualitative exploration of oncologists’ perspectives,” *J Cancer Policy*, vol. 45, Sep. 2025, doi: 10.1016/j.jcpo.2025.100626.

[5] I. Gabriel and A. Precious, “Challenges of Implementing AI in Low-Resource Healthcare Settings,” Aug. 2025. [Online]. Available: https://www.researchgate.net/publication/394275718

[6] J. G. Makama, P. Iribhogbe, and E. A. Ameh, “Overcrowding of accident & emergency units: Is it a growing concern in Nigeria?,” *Afr Health Sci*, vol. 15, no. 2, pp. 457–465, 2015, doi: 10.4314/ahs.v15i2.20.

[7] W. Yinusa, “Emergency Medicine as a Medical Speciality in Nigeria: Challenges and Prospects,” *Nigerian Postgraduate Medical Journal*, vol. 31, no. 4, pp. 346–352, 2024, doi: 10.4103/npmj.npmj_139_24.

[8] K. Umoga, M. Muhammad, K. Olawale, O. Mbbs, and C. Ngaruiya, “Assessment of emergency care services in Nigerian hospitals: a cross-sectional study,” Sep. 2022. [Online]. Available: https://ssrn.com/abstract=4218805

[9] O. O. Olaomi, G. O. Okoye, and B. Balarabe-Musa, “Current Trends in Emergency Medicine in Nigeria,” Dec. 01, 2024, *Wolters Kluwer Medknow Publications*. doi: 10.4103/njcp.njcp_136_24.

[10] S. Nneoma aꙍ, C. Anthonia, and D. Kenneth, “Knowledge and Perceived Skills of Triage Practices of Nurses in Emergency Units of Selected Tertiary Hospitals in Enugu State, South-East Nigeria,” 2022. [Online]. Available: https://www.sdiarticle5.com/review-history/87128

[11] A. DaCosta, A. Osonuga, and O. Adesegun, “The Urgent Need for Postgraduate Medical Training in Emergency Medicine in Nigeria,” Mar. 01, 2020, *African Federation for Emergency Medicine*. doi: 10.1016/j.afjem.2019.09.003.

[12] S. Abujaber, C. Y. Chang, T. A. Reynolds, H. Mowafi, and Z. Obermeyer, “Développer des outils de mesure pour la recherche en matière de soins d’urgence dans les pays à faible et moyen revenus,” *African Journal of Emergency Medicine*, vol. 6, no. 3, pp. 116–124, Sep. 2016, doi: 10.1016/j.afjem.2016.06.003.

[13] D. Onwochei, E. E. Nwaomah, and O. Omitogun, “Adoption of Emergency Severity Index among Nurses in the Emergency Department of Selected Hospitals in Lagos State,” *Current Trends in Life Science Research*, vol. 2, Dec. 2023, doi: 10.61867/pcub.v2i2b.058.

[14] C. C. Yancey and M. C. O’Rourke, *Emergency Department Triage*. StatPearls Publishing LLC, 2023. [Online]. Available: https://www.ncbi.nlm.nih.gov/books/NBK557583/

[15] *Emergency triage assessment and treatment (ETAT) : facilitator guide*. World Health Organization, Dept. of Child and Adolescent Health and Development, 2005.

[16] C. Uwamahoro *et al.*, “Evaluation of a modified South African Triage Score as a predictor of patient disposition at a tertiary hospital in Rwanda,” *African Journal of Emergency Medicine*, vol. 10, no. 1, pp. 17–22, Mar. 2020, doi: 10.1016/j.afjem.2019.10.001.

[17] M. Zou, L. Ramjan, K. Thompson, and G. Kabil, “Facilitators and barriers to triage efficiency in adult emergency departments: An integrative literature review,” 2025, *Elsevier Australia*. doi: 10.1016/j.auec.2025.09.006.

[18] A. Defilippo, G. Bertucci, C. Zurzolo, P. Veltri, and P. H. Guzzi, “On the computational approaches for supporting triage systems,” Jul. 01, 2023, *John Wiley & Sons Inc*. doi: 10.1002/INMD.20230015.

[19] A. H. Elbaih, “Triyaj Türleri,” *Arşiv Kaynak Tarama Dergisi*, vol. 26, no. 4, pp. 441–467, Dec. 2017, doi: 10.17827/aktd.326944.

[20] P. Sambade Alonso, “COMPARATIVE ANALYSIS OF GLOBAL TRIAGE SYSTEMS: COMPETENCES OF NURSING AND HOSPITAL AND PREHOSPITAL APPLICATION AUTHOR,” *Journal Nursing Valencia*, vol. 4, 2004.

[21] T. S. Jessop, L. Martin, H. J. Lategan, J. M. Dixon, N. K. Moul-Millman, and E. Steyn, “The Suitability of the South African Triage Scale (SATS) in Triaging Patients With Penetrating Neck Injuries at a High-Level Trauma Center, in the Western Cape, South Africa,” *World J Surg*, vol. 49, no. 5, pp. 1351–1358, May 2025, doi: 10.1002/wjs.12578.

[22] L. S. Wen, J. I. Oshiomogho, G. I. Eluwa, A. P. Steptoe, A. F. Sullivan, and C. A. Camargo, “Characteristics and capabilities of emergency departments in Abuja, Nigeria,” *Emergency Medicine Journal*, vol. 29, no. 10, pp. 798–801, Oct. 2012, doi: 10.1136/emermed-2011-200695.

[23] R. Mitchell, L. White, L. Elton, C. Luke, S. Bornstein, and V. Atua, “Triage implementation in resource-limited emergency departments: sharing tools and experience from the Pacific region,” *Int J Emerg Med*, vol. 17, no. 1, Dec. 2024, doi: 10.1186/s12245-024-00583-8.

[24] R. Mitchell *et al.*, “Triage systems in low-resource emergency care settings,” *Bulletin of the World Health Organization;*, Jan. 30, 2025.

[25] A. Owoyemi *et al.*, “Mobile health approaches to disease surveillance in Africa; Wellvis COVID triage tool,” 2021, *SAGE Publications Inc.* doi: 10.1177/2055207621996876.

[26] J. G. Makama, P. Iribhogbe, and E. A. Ameh, “Overcrowding of accident & emergency units: Is it a growing concern in Nigeria?,” *Afr Health Sci*, vol. 15, no. 2, pp. 457–465, 2015, doi: 10.4314/ahs.v15i2.20.

[27] J. S. Hinson and S. Levin, “Data-Driven Approach Yields New Approach for Emergency Department Triage.” [Online]. Available: https://www.acepnow.com/article/data-driven-approach-yields-new-approach-for-emergency-department-triage/?singlepage=1&theme=print-friendly

[28] R. A. Taylor *et al.*, “Impact of Artificial Intelligence–Based Triage Decision Support on Emergency Department Care,” *NEJM AI*, vol. 2, no. 3, Feb. 2025, doi: 10.1056/aioa2400296.

[29] B. M. Porto, “Improving triage performance in emergency departments using machine learning and natural language processing: a systematic review,” Dec. 01, 2024, *BioMed Central Ltd*. doi: 10.1186/s12873-024-01135-2.

[30] T. Liu *et al.*, “A foundational triage system for improving accuracy in moderate acuity level emergency classifications,” *Communications Medicine*, vol. 5, no. 1, Dec. 2025, doi: 10.1038/s43856-025-01052-w.

[31] "TriageGO - Optimized Triage with Artificial Intelligence." [Online]. Available: https://www.radiometeramerica.com/en-us/products/clinical-intelligence/triagego

[32] Federal Ministry of Health & Social Welfare Nigeria, "FG inaugurates Committee on Digital in Health Initiative." [Online]. Available: https://health.gov.ng/fg-inaugurates-committee-on-digital-in-health-initiative/#:~:text=

[33] H. B. Bergquist, T. W. Burkholder, O. A. Muhammad Ali, Y. Omer, and L. A. Wallis, "Considerations for service delivery for emergency care in low resource settings," Jan. 01, 2020, *African Federation for Emergency Medicine*. doi: 10.1016/j.afjem.2020.07.002.

[34] S. E. Edwards, "A practical approach to outpatient triage," *Paediatr Child Health*, vol. 34, no. 6, pp. 197–199, Jun. 2024, doi: 10.1016.

[35] D. Peta, A. Day, W. S. Lugari, V. Gorman, N. A. Ahayalimudin, and V. M. T. Pajo, "Triage: A Global Perspective," *J Emerg Nurs*, vol. 49, no. 6, pp. 814–825, Nov. 2023, doi: 10.1016/j.jen.2023.08.004.

[36] M. Sousa, M. N. Cunha, and J. Rodrigues, "Patients' Perceptions Resulting from the Contact with Emergency Departments Using the Manchester Triage System Protocol," *Univers J Public Health*, vol. 12, no. 3, pp. 471–480, 2024, doi: 10.13189/ujph.2024.120305.

[37] N. Agency for Clinical Innovation, "Emergency department triage models: Evidence check," Dec. 2024. Accessed: Oct. 16, 2025. [Online]. Available: aci.health.nsw.gov.au

[38] F. Cotte *et al.*, "Safety of Triage Self-assessment Using a Symptom Assessment App for Walk-in Patients in the Emergency Care Setting: Observational Prospective Cross-sectional Study," *JMIR Mhealth Uhealth*, vol. 10, no. 3, Mar. 2022, doi: 10.2196/32340.

[39] "Better health starts with Ada." [Online]. Available: https://about.ada.com

[40] C. Y. K. Williams *et al.*, "Use of a Large Language Model to Assess Clinical Acuity of Adults in the Emergency Department," *JAMA Netw Open*, vol. 7, no. 5, May 2024, doi: 10.1001/jamanetworkopen.2024.8895.

[41] S. Paslı, A. S. Şahin, M. F. Beşer, H. Topçuoğlu, M. Yadigaroğlu, and M. İmamoğlu, "Assessing the precision of artificial intelligence in emergency department triage decisions: Insights from a study with ChatGPT," *American Journal of Emergency Medicine*, vol. 78, pp. 170–175, Apr. 2024, doi: 10.1016/j.ajem.2024.01.037.

[42] S. Tyler *et al.*, "Use of Artificial Intelligence in Triage in Hospital Emergency Departments: A Scoping Review," *Cureus*, May 2024, doi: 10.7759/cureus.59906.

[43] R. C. Zhao and X. Yuan, "AI in Healthcare for Resource Limited Settings: An Exploration and Ethical Evaluation," in *WWW Companion 2025 - Companion Proceedings of the ACM Web Conference 2025*, Association for Computing Machinery, Inc, May 2025, pp. 1953–1960. doi: 10.1145/3701716.3717747.

[44] J. P. Telles, F. Tuon, A. Cunha, and J. L. L. Rocha, "The burden of limited resources in Latin America on healthcare quality and management," Mar. 01, 2025, *Elsevier Ltd*. doi: 10.1016/j.lana.2025.101014.

[45] F. Iacoella, N. Dehingia, M. Almanzar, U. K. Huynh, and Z. Bruckauf, "Healthcare access in humanitarian settings: Mobile services successes and limits in Afghanistan," *Soc Sci Med*, vol. 364, Jan. 2025, doi: 10.1016/j.socscimed.2024.117541.

[46] M. Basij-Rasikh, E. S. Dickey, and A. Sharkey, "Primary healthcare system and provider responses to the Taliban takeover in Afghanistan," *BMJ Glob Health*, vol. 9, no. 2, Feb. 2024, doi: 10.1136/bmjgh-2023-013760.

[47] E. Sabet, S. Shafiullah Sadat, and H. A. Khaliqi, "AI Based Chatbot: A Case Study Afghanistan Healthcare Services Mental Health Disorder," *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS*, vol. 07, no. 05, May 2024, doi: 10.47191/ijmra/v7-i05-28.

[48] M. D. Kasaye, A. G. Getahun, and M. H. Kalayou, "Exploring artificial intelligence for healthcare from the health professionals' perspective: The case of limited resource settings," *Digit Health*, vol. 11, Jan. 2025, doi: 10.1177/20552076251330552.

[49] S. Ojedokun, S. Afolabi, O. Olukoyejo, and T. Alatishe, "Perceptions and Opinions of Medical Professionals on Artificial Intelligence in Optimizing the Healthcare Sector," 2024.

[50] N. Kapoor, S. N. Sanjana, S. B. Davalagi, P. S. Balu, and S. Sethia, "AI Horizons in Indian Healthcare: A Vision for Transformation and Equity," *Indian Journal of Community Medicine*, vol. 49, no. Suppl 2, pp. S210–S216, Dec. 2024, doi: 10.4103/ijcm.ijcm_806_24.

[51] S. K. Das, R. K. Dasgupta, S. D. Roy, and D. Shil, "AI in Indian healthcare: From roadmap to reality," Jun. 01, 2024, *KeAi Publishing Communications Ltd.* doi: 10.1016/j.ipha.2024.02.005.

[52] "Swahili gaining popularity globally _ Africa Renewal".

[53] E. Millen *et al.*, "Study protocol for a pilot prospective, observational study investigating the condition suggestion and urgency advice accuracy of a symptom assessment app in sub-Saharan Africa: the AFYA- € Health' Study," *BMJ Open*, vol. 12, no. 4, Apr. 2022, doi: 10.1136/bmjopen-2021-055915.

[54] Ada Health GmbH, "World's first AI health guidance app in Swahili launched by Ada Health - Ada." Accessed: Sep. 30, 2025. [Online]. Available: https://about.ada.com/press/191119-worlds-first-ai-health-guidance-app-in-swahili/

[55] M. Schmude *et al.*, "Investigating the Potential for Clinical Decision Support in Sub-Saharan Africa With AFYA (Artificial Intelligence-Based Assessment of Health Symptoms in Tanzania): Protocol for a Prospective, Observational Pilot Study," *JMIR Res Protoc*, vol. 11, no. 6, Jun. 2022, doi: 10.2196/34298.

[56] L. Mancini *et al.*, "Artificial Intelligence in Healthcare: Systematic Review of Diagnostic and Screening Tools in Brazil's Resource-Limited Settings (Preprint)," *JMIR AI*, Sep. 2024, doi: 10.2196/69547.

[57] "About Hurone AI - Hurone AI." Accessed: Nov. 02, 2025. [Online]. Available: https://hurone.ai/about-us/

[58] "About Us." [Online]. Available: https://afyarekod.com/about-us

[59] K. V. Odunuga *et al.*, "Artificial Intelligence in Nigeria Health Sector," *Scholar J Computational Science*, pp. 16–19, Aug. 2024, doi: 10.5281/zenodo.13226114.

[60] O. Chidoluo Vitus, "THE IMPACT OF ARTIFICIAL INTELLIGENCE ON HEALTHCARE DELIVERY IN NIGERIA," *Reproductive Health THE IMPACT OF ARTIFICIAL INTELLIGENCE ON HEALTHCARE DELIVERY IN NIGERIA*, vol. 1, no. 2, 2024, doi: 10.9567/ISSN.2024/WSJ.92.

[61] O. O. Bello and A. O. Ajao, "Digital Literacy and Skills Development in Nigeria: Policies, Barriers and Recommendations," *Journal of African Innovation & Advanced Studies (JAIAS)*, vol. 5, no. 2, pp. 133–150, Aug. 2024, [Online]. Available: www.africanscholarpub.com

[62] Techpoint Africa, "SmartMRS integrates two AI features that simplifies patient diagnoses." [Online]. Available: https://techpoint.africa/coverage/smartverge-the-smartmrs-advantage/

[63] H. Bawa-Muhammad, "Digital Health in Nigeria: A Path to SDG 3 for Maternal and Child," Abuja, May 2025. doi: http://dx.doi.org/10.2139/ssrn.5243011.

[64] Science for Africa Foundation, "Leveraging Artificial Intelligence (AI) to Strengthen Health Systems in Nigeria," 2024.

[65] O. P. Adigwe, G. Onavbavba, and S. E. Sanyaolu, "Exploring the matrix: knowledge, perceptions and prospects of artificial intelligence and machine learning in Nigerian healthcare," *Front Artif Intell*, vol. 6, 2023, doi: 10.3389/frai.2023.1293297.

[66] M. Okwukwu, D. Oluwafemi Olofin, and A. Akintayo Taiwo, "Artificial Intelligence in Nigeria Healthcare: A Review of State, Challenges and Opportunities," *EthAIca*, vol. 4, p. 210, Jul. 2025, doi: 10.56294/ai2025210.

[67] Xolani Health, "DICOM-X." [Online]. Available: https://xolanibridge.com

[68] "HelpMum in the News Improving Maternal and Infant Healthcare in Africa with AI and low cost Innovations." [Online]. Available: https://helpmum.org

[69] O. Kehinde *et al.*, "Deploying ADVISER: Impact and Lessons from Using Artificial Intelligence for Child Vaccination Uptake in Nigeria," Dec. 2023, [Online]. Available: http://arxiv.org/abs/2402.00017

[70] ProNOV Health, "ProNOV; disrupting retail community pharmacy in Africa," Feb. 2018. [Online]. Available: https://medium.com/@ProNOV/pronov-disrupting-retail-community-pharmacy-in-africa-bab495e9389

[71] "Insights Lab – HubCare Health".

[72] V. Fakiya, "9 telemedicine startups changing how Nigerians access healthcare." Accessed: Oct. 08, 2025. [Online]. Available: https://techpoint.africa/insight/9-affordable-telemedicine-startups-nigeria/

[73] A. D. Olusegun-Joseph, O. Akande, E. Otrofanowei, E. O. Nwoye, O. B. Olopade, and J. N. Ajuluchukwu, "Medical mortality in an emergency department in Nigeria: The transition is obvious!," *Afr Health Sci*, vol. 21, no. 1, pp. 172–179, Mar. 2021, doi: 10.4314/ahs.v21i1.23.

[74] B. Bello, Y. Raji, C. Amira, R. Braimoh, O. Olowoyo, and B. Akodu, "Blood pressure pattern among adults in Lagos: Analysis of data from public health screenings," *The Nigerian Journal of General Practice*, vol. 16, no. 1, p. 10, 2018, doi: 10.4103/njgp.njgp_20_17.

[75] C. Nee-Kofi Mould-Millman, S. Rominski, and R. Oteng, "Ambulance or taxi? High acuity prehospital transports in the Ashanti region of Ghana," *African Journal of Emergency Medicine*, vol. 4, no. 1, pp. 8–13, 2014, doi: 10.1016/j.afjem.2013.04.004.

[76] N. J. Garrido, F. González-Martínez, S. Losada, A. Plaza, E. del Olmo, and J. Mateo, "Innovation through Artificial Intelligence in Triage Systems for Resource Optimization in Future Pandemics," *Biomimetics*, vol. 9, no. 7, Jul. 2024, doi: 10.3390/biomimetics9070440.

[77] Sruthi, "Random Forest Algorithm in Machine Learning," Analytics Vidhya. [Online]. Available: https://www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/

[78] V. F. Rodriguez-Galiano, M. Sanchez-Castillo, J. Dash, P. M. Atkinson, and J. Ojeda-Zujar, "Modelling interannual variation in the spring and autumn land surface phenology of the European forest," *Biogeosciences*, vol. 13, no. 11, pp. 3305–3317, Jun. 2016, doi: 10.5194/bg-13-3305-2016.

[79] A. Rubinsteyn, J. Li, Y. Liao, and R. Power, “Learning Random Forests on the GPU,” 2013. [Online]. Available: https://www.researchgate.net/publication/259150669

[80] “RandomForestClassifier — scikit-learn 1.7.2 documentation.” Accessed: Nov. 02, 2025. [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html

[81] “How to interpret a confusion matrix for a machine learning model.” Accessed: Nov. 02, 2025. [Online]. Available: https://www.evidentlyai.com/classification-metrics/confusion-matrix

[82] “How to interpret a confusion matrix for a machine learning model”.

[83] “Splitting Decision Trees with Gini Impurity,” Analytics Vidhya. [Online]. Available: https://www.analyticsvidhya.com/articles/gini-impurity/

[84] S. Lundberg, “SHAP Documentation.” Accessed: Nov. 02, 2025. [Online]. Available: https://shap.readthedocs.io/en/latest/index.html

[85] Snowflake Inc., “API Reference - Streamlit Docs.”