FMDB *Transactions on Sustainable Health Science Letters*

Explainable Machine Learning Framework for Organ Transplant Prioritization using SAHP and SHAP

Farrukh Arslan[1], S. Rubin Bose[2], J. Angelin Jeba[3,*], Shahid Ullah[4], Bushra Rehman[5], Kawsher Rahman[6], Mohamed Sameh Mohamed Elhadad[7], Mohamed Ibrahim Hamad Aborakika[8], Lorans Ashraf Botros Ajeep Rasan[9]

[1]Department of Computer Science, Purdue University, West Lafayette, Indiana, United States of America.
[2]School of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram, Chennai, Tamil Nadu, India.
[3]Department of Electronics and Communication Engineering, S.A. Engineering College, Chennai, Tamil Nadu, India.
[4]Department of Bioinformatics and Biochemistry, S Khan Lab, Mardan, Khyber Pakhtunkhwa, Pakistan.
[5]Institute of Pathology and Diagnostic Medicine, Khyber Medical University, Peshawar, Khyber Pakhtunkhwa, Pakistan.
[6]Department of General Medicine, Beanibazar Cancer and General Hospital, Beanibazar, Sylhet, Bangladesh.
[7,8,9]Department of Biomedical Sciences, Zewail City of Science and Technology, October Gardens, Giza Governorate, Egypt.
farslan@purdue.edu[1], rubinbos@srmist.edu.in[2], angelinjeba@saec.ac.in[3], shahid@habdsk.org[4], drbushra.ipdm@kmu.edu.pk[5], drkawsher.rahman@gmail.com[6], s-mohamed.elhadad@zewailcity.edu.eg[7], s-mohamed.aborakika@zewailcity.edu.eg[8], s-lorans.ajeep@zewailcity.edu.eg[9]

Abstract: Organ transplants have become one of the most important parts of health care, primarily because of the imbalance between those who need organs and the available organ donor pool. Due to this disparity, choosing the next organ recipient among other candidates can be quite challenging. This process largely relies on several medical criteria, including urgency, compatibility, severity, waiting period, and survival chances. Taking all these aspects into account and processing them manually could be problematic, even leading to biases during prioritisation. The proposed solution is a Smart Organ Transplant Priority Analyzer that facilitates the prioritization of patients waiting for an organ transplant using Machine Learning, SAHP, and SHAP. The tool evaluates patient information, assigns a priority score and category, and provides recommendations on who should be a top candidate to receive the organ. SAHP assigns weights to medical parameters to make sure that crucial factors play a greater role in determining the transplant recipient. Machine Learning algorithms, such as random forests, are used to estimate patients' transplant priority scores. SHAP provides interpretability of results by explaining why a particular outcome was achieved and the contribution of each criterion to that outcome.

Keywords: Organ Transplant Prioritisation; Machine Learning; SHAP and XGBoost; Random Forest; Support Vector Machine (SVM); Explainable AI; Health Care; Transplant Recipient.

Received on: 28/07/2025, **Revised on:** 05/10/2025, **Accepted on:** 04/12/2025, **Published on:** 18/08/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSHSL

DOI: https://doi.org/10.69888/FTSHSL.2026.000733

Cite as: F. Arslan, S. R. Bose, J. A. Jeba, S. Ullah, B. Rehman, K. Rahman, M. S. M. Elhadad, M. I. H. Aborakika, and L. A. B. A. Rasan, "Explainable Machine Learning Framework for Organ Transplant Prioritization using SAHP and SHAP," *FMDB Transactions on Sustainable Health Science Letters*, vol. 4, no. 3, pp. 224–236, 2026.

[*]Corresponding author.

1. Introduction

Organ transplantation remains one of the most significant therapeutic interventions available in modern medicine, offering a definitive treatment option for patients suffering from end-stage organ failure. For individuals affected by chronic kidney disease, liver cirrhosis, congestive heart failure, or advanced pulmonary disorders, transplantation is frequently not merely a treatment option but the only viable path toward long-term survival and improved quality of life [1]. Despite significant advances in surgical technique, immunosuppressive therapy, and organ preservation technology over the past several decades, the fundamental challenge confronting transplant medicine has remained largely unresolved: the persistent and often widening gap between the number of patients awaiting organ transplantation and the limited supply of viable donor organs. This scarcity necessitates that healthcare systems, transplant centers, and clinical decision-makers engage in a continuous, ethically weighty process of prioritization, determining which patients, among many equally deserving candidates, should receive an available organ first [2]. The process of organ allocation is inherently multidimensional and cannot be reduced to a single clinical metric. Physicians and transplant coordinators must simultaneously weigh a range of interdependent factors, including the severity and progression of the patient's underlying disease, the degree of immunological and physiological compatibility between the donor organ and the recipient, the length of time the patient has spent on the waiting list, and the patient's projected likelihood of post-transplant survival and recovery [11]. Each of these criteria carries independent clinical significance, yet none can be considered in isolation, as trade-offs frequently arise between competing priorities.

A patient with the most severe disease presentation, for instance, may not always have the highest probability of a successful outcome, while a patient who has waited longest may not necessarily be in the most urgent clinical need [3]. Reconciling these competing considerations in a manner that is both medically sound and ethically defensible becomes increasingly difficult as the number of patients on the waiting list grows and the criteria themselves become more numerous and nuanced [10]. Consequently, there is a recognised and ongoing need for decision-support mechanisms that can assist clinicians in navigating this complexity while preserving fairness, consistency, and transparency across the allocation process [4]. The increasing availability of clinical data, coupled with advances in computational methods, has created new opportunities to address these challenges through data-driven approaches. Machine learning, in particular, enables the analysis of large volumes of heterogeneous patient information and the identification of patterns and relationships that may not be immediately apparent through conventional clinical assessment alone. By learning from historical patient data, machine learning models can stratify patients by disease severity and estimated transplant priority, providing a systematic, reproducible complement to clinical judgment [5]. However, adopting such computational tools within a domain as sensitive as organ allocation introduces its own set of challenges. Machine learning models, particularly those based on ensemble methods or kernel-based classifiers, are often characterised by a degree of opacity that limits their interpretability. In a clinical context where decisions directly affect patient survival, it is not sufficient for a model to produce a recommendation; clinicians, patients, and oversight bodies must also be able to understand the reasoning underlying that recommendation.

A prediction that cannot be explained or justified is unlikely to be trusted or adopted in practice, regardless of its underlying statistical accuracy. In response to these considerations, the present work proposes a hybrid decision-support framework that integrates three complementary methodologies: the Saaty-based Analytic Hierarchy Process (SAHP), machine learning classification, and SHAP-based explainability [6]. The SAHP component is employed to formally structure and weight the criteria relevant to transplant prioritization, drawing on established multi-criteria decision-making principles to encode clinical and ethical judgment in a transparent, mathematically consistent manner [7]. This expert-informed weighting scheme provides an interpretable foundation that is not solely dependent on patterns extracted from historical data, thereby helping to mitigate the risk that a purely data-driven model might inadvertently reproduce or amplify biases present in past allocation practices. Building on this foundation, machine learning algorithms are employed to classify patients into priority categories based on their clinical and physiological characteristics, leveraging models such as random forests, gradient boosting methods, and support vector machines to capture complex, nonlinear relationships in the data [8]. Finally, SHAP (Shapley Additive explanations) is incorporated as an explainability layer, enabling quantification of each feature's contribution to a patient's predicted priority score. This allows clinicians to move beyond a simple numerical output and to understand precisely which factors, whether disease severity, organ compatibility, waiting time, or specific laboratory values, most strongly influenced a given prediction.

The overarching objective of this paper is not to replace clinical decision-making but to augment it [15]. The system is conceived as a tool that provides physicians and transplant teams with a more comprehensive, consistent, and evidence-based view of each patient's status, thereby reducing ambiguity and supporting more standardised decision-making across cases and institutions [9]. By combining structured expert knowledge, predictive modelling, and interpretable explanations within a single framework, the proposed approach seeks to enhance both the objectivity and the transparency of the organ allocation process, without diminishing the essential role of clinical expertise and judgment in the final decision [12]. The significance of this work stems directly from the gravity of the decisions it seeks to support. Determining who receives a life-saving organ transplant is among the most consequential decisions made within modern healthcare systems, carrying direct implications for patient

survival. Even incremental improvements in the fairness, consistency, and transparency of this process therefore carry substantial ethical and practical value [13]. By demonstrating how machine learning, structured multi-criteria decision analysis, and explainable artificial intelligence can be integrated into a coherent decision-support system, this paper aims to contribute meaningfully to the broader field of healthcare technology, illustrating a pathway through which computational tools and clinical expertise can operate collaboratively to ensure that organ transplantation is conducted in a manner that is equitable, accountable, and ultimately beneficial to the patients who depend on it most [14].

2. Literature Review

Organ transplant prioritization draws upon several distinct but complementary bodies of research, spanning machine learning, explainable artificial intelligence, multi-criteria decision-making theory, and broader applications of computational systems in clinical and healthcare informatics. Each of these domains contributes essential components to the framework proposed in this study, and a review of the relevant literature reveals both the strengths of existing approaches and the specific gap that this paper seeks to address. Breiman [1] introduced the Random Forest algorithm as an ensemble learning method that constructs a multitude of decision trees during training and aggregates their outputs, typically via majority voting, to produce a final prediction. The strength of this approach lies in its use of bootstrap aggregation (bagging) combined with random feature selection at each split, which collectively reduces the variance associated with individual decision trees and mitigates the risk of overfitting that plagues single-tree models. Because Random Forest can accommodate a large number of heterogeneous input features, handle nonlinear relationships without extensive feature engineering, and provide relatively stable performance across varied datasets, it has become one of the most widely adopted algorithms in medical decision-support systems, where clinical datasets are frequently high-dimensional, noisy, and imbalanced. Its robustness and interpretability relative to more complex deep learning architectures make it particularly well-suited to structured clinical data used in transplant prioritisation, where feature importance and reliability are as critical as raw predictive accuracy. While Random Forest and similar ensemble methods offer strong predictive performance, their internal decision-making process remains largely opaque to end users.

This limitation becomes particularly problematic in high-stakes domains such as healthcare [2]. Lundberg and Lee [2] addressed this limitation by proposing SHAP (Shapley Additive Explanations), a unified, theoretically grounded framework for interpreting the output of any machine learning model. Drawing on the concept of Shapley values from cooperative game theory, SHAP assigns a given prediction to the individual contributions of each input feature, ensuring that these contributions satisfy desirable mathematical properties, such as local accuracy, consistency, and robustness to missing values. This unification of previously disparate interpretability methods under a single, theoretically justified framework represented a substantial advance for explainable AI, enabling practitioners to move beyond post-hoc heuristic explanations toward attributions with formal guarantees. Building on this foundational work, Lundberg et al. [3] developed TreeExplainer, an optimised algorithm specifically designed to compute exact Shapley values efficiently for tree-based ensemble models. This extension addressed a critical computational bottleneck: the naive computation of Shapley values scales exponentially with the number of features, rendering it impractical for models such as Random Forests and XGBoost, which often operate on dozens of clinical variables. By making exact, efficient explanation feasible for tree ensembles specifically, this work directly enabled the pairing of high-performing ensemble classifiers with rigorous, computationally tractable interpretability. This combination underpins the explainability component of the present study. The necessity of such interpretability in clinical contexts is reinforced by Holzinger et al. [4], who examined the role of Explainable AI (XAI) specifically within healthcare applications.

Their work argued that the deployment of black-box models in medicine is fundamentally constrained by the requirement for transparency, not merely as a technical preference but as an ethical and regulatory necessity. Clinicians, they noted, cannot be expected to act on algorithmic recommendations they cannot interrogate, particularly in decisions with irreversible consequences for patient outcomes [9]. Holzinger et al. [4] further emphasised that explainability serves multiple functions beyond clinician trust, including facilitating error detection, supporting regulatory compliance, and enabling meaningful human oversight of automated systems. This perspective directly motivates the integration of an explainability layer within the present system, ensuring that predictions generated by the machine learning component are never presented to clinicians as unqualified outputs but are instead accompanied by transparent, feature-level justification. Complementing the machine learning and explainability literature, the present study draws heavily on multi-criteria decision-making theory, particularly the Analytic Hierarchy Process (AHP) introduced by Saaty [6]. AHP provides a structured mathematical technique for decomposing complex decisions involving multiple, often competing, and criteria into a hierarchy of simpler pairwise comparisons. By eliciting relative judgments between pairs of criteria and organising them into a comparison matrix, AHP derives a consistent set of numerical weights that reflect the relative importance of each criterion, while simultaneously offering a built-in consistency check to detect and correct contradictory judgments.

Saaty's foundational contribution: Saaty [5] extended and formalised this methodology, establishing AHP as one of the most widely used and rigorously validated tools in multi-criteria decision-making across domains ranging from engineering to healthcare resource allocation. The adaptation of this method into a simplified variant, referred to here as SAHP, reflects a

broader trend in the literature toward operationalising AHP for practical, computationally efficient implementation within larger decision-support pipelines, particularly when the weighting process must be integrated with downstream predictive models rather than functioning as a standalone decision tool. In parallel with developments in Random Forests and SHAP, Chen and Guestrin [7] introduced XGBoost. This scalable gradient-boosting framework has become one of the most performant algorithms for structured, tabular classification and regression tasks. Unlike the parallel tree construction employed by Random Forest, XGBoost builds trees sequentially, with each successive tree trained to correct the residual errors of the preceding ensemble, guided by a regularised objective function that balances predictive accuracy against model complexity [1]. This design, combined with computational-efficiency optimisations, has made XGBoost a preferred choice in numerous healthcare prediction systems, where it frequently outperforms alternative classifiers on structured clinical datasets. Its inclusion alongside Random Forest and SVM in the present study's model comparison reflects established practice in the literature of benchmarking multiple high-performing classifiers rather than relying on a single algorithm.

Beyond specific algorithmic contributions, several studies have examined the broader application of data-driven methods within clinical medicine. Bellazzi and Zupan [8] explored predictive data mining in clinical settings, demonstrating how machine learning techniques could be systematically applied to tasks spanning diagnosis, prognosis, and treatment planning. Their work was notable for framing predictive modelling not as a replacement for clinical expertise but as a complementary analytical layer capable of uncovering patterns within clinical data that might otherwise remain undetected through conventional statistical or manual review. This framing anticipated a broader shift in the field, later articulated more fully by Topol, who introduced the concept of high-performance medicine, describing an emerging paradigm in which artificial intelligence and clinical expertise operate synergistically rather than in competition. Breiman [1] emphasised that the most effective and sustainable use of AI in healthcare lies not in autonomous decision-making but in augmenting clinicians' judgment, providing them with richer, more comprehensive information to support, rather than supplant, their decisions. This principle directly informs the design philosophy of the present study, which positions its machine learning and decision-support components explicitly as tools to assist, rather than replace, clinical decision-makers in the organ allocation process. The application of more advanced computational architectures to healthcare data has also been extensively documented.

Esteva et al. [10] provided a comprehensive review of deep learning applications in healthcare, illustrating how neural network-based models have been successfully applied to complex medical data types, including imaging, genomics, and unstructured clinical records, often achieving performance comparable to or exceeding that of human experts in specific diagnostic tasks. Their review underscored both the substantial potential and the practical challenges of deploying such models at scale, including issues of data quality, generalizability, and interpretability, the latter of which reinforces the continued relevance of explainability frameworks such as SHAP even as model complexity increases. Rajkomar et al. [11] extended this discussion by examining the practical application of machine learning across a range of clinical environments, emphasising its capacity to process large-scale, heterogeneous healthcare data and its potential to meaningfully improve clinical outcomes when properly integrated into existing healthcare workflows. Their subsequent work, Rajkomar et al. [12], demonstrated this potential concretely by applying deep learning techniques directly to electronic health record (EHR) data, showing that large-scale, longitudinal patient records could be leveraged to build highly accurate predictive models for a range of clinical outcomes. This body of work establishes the empirical foundation for using structured patient data, as captured in the feature set used in the present study, to build clinically meaningful predictive systems. The broader informatics infrastructure underlying such systems is addressed by Shortliffe and Cimino [13], whose foundational work on biomedical informatics examined the design and application of computer-based systems within healthcare more generally.

Their contributions established many of the conceptual and architectural principles that continue to underpin modern clinical decision-support systems, including considerations of system usability, integration with clinical workflows, and the appropriate scope of automated decision assistance. This work provides an important methodological grounding for the present study's system-level design, which similarly seeks to embed predictive and explanatory components within a coherent decision-support architecture rather than treating them as isolated analytical tools. Finally, the ethical and policy dimensions of organ allocation are addressed directly by the World Health Organization. The WHO's examination of organ transplantation systems [14] highlighted the persistent global challenges associated with equitable organ allocation, noting that disparities in access, inconsistent prioritisation criteria, and limited transparency remain significant barriers to fair transplant practice across health systems. This was further formalised in the WHO Guiding Principles on Human Cell, Tissue and Organ Transplantation [15], which explicitly articulate the ethical standards expected to govern organ allocation, emphasising fairness, accountability, non-discrimination, and transparent prioritisation criteria as foundational requirements for any legitimate transplant system. These principles provide the normative framework against which any computational decision-support system for organ allocation must be evaluated, and they directly motivate the emphasis placed in this study on transparency and explainability, not merely as technical features but as ethical necessities.

Taken together, the reviewed literature demonstrates that substantial progress has been made independently across three relevant domains: machine learning methods such as Random Forest and XGBoost offer strong predictive capability for clinical

classification tasks; SHAP and related explainability frameworks provide rigorous, computationally feasible means of interpreting model predictions; and AHP/SAHP offer a validated approach to structuring expert judgment into consistent, quantifiable criterion weights [2]. However, the literature also reveals a clear gap: existing studies tend to address these components in isolation, applying machine learning for prediction, AHP-type methods for weighting, or SHAP for explanation, but rarely integrating all three within a single, unified decision-support framework specifically tailored to organ transplant prioritisation. Studies emphasising the ethical imperatives of transparent and fair allocation [4, 14, 15] further underscore the need for such integration, as prediction alone, however accurate, is insufficient without both principled weighting and interpretable justification. This paper directly addresses this gap by combining Random Forest-based machine learning, SAHP-based criteria weighting, and SHAP-based explainability into a unified framework, thereby providing a system that is simultaneously predictive, structured by expert clinical judgment, and transparent in its reasoning, to support fair and accountable organ transplant prioritisation decisions [1].

3. Methodology

Figure 1 represents the complete workflow of the proposed Smart Organ Transplant Priority Analyser. The process begins with patient data collection, including demographic and clinical details such as urgency level, disease severity, compatibility score, and survival probability. The collected data is then passed through preprocessing and feature preparation stages, followed by SAHP-based weighting of important medical criteria. After weighting, machine learning models analyse the data to predict the transplant priority category. SHAP is then applied to explain the model's decision, and the system finally generates the patient's priority score, urgency category, recommendation, and ranked waiting-list position. Figure 1 provides an overview of how the system operates from input to final decision-support output. The proposed system architecture follows a modular pipeline in which patient data is processed through parallel expert-driven and data-driven pathways before being merged into a single machine learning prediction stage, which is subsequently rendered interpretable through post-hoc explainability methods. This hybrid design was adopted specifically because organ allocation is an ethically sensitive task for which prediction accuracy alone is insufficient; the system must also remain auditable and consistent with established clinical allocation principles. The pipeline begins with the acquisition of patient-level data, comprising demographic attributes (age, gender, blood group), organ-specific clinical indicators (organ needed, medical urgency level, disease severity, organ failure stage), a temporal factor (cumulative waiting time), compatibility metrics (compatibility score, survival probability), situational flags (ICU requirement, donor availability), and physiological/laboratory values (BMI, hemoglobin, creatinine, blood pressure, and SpO_2).

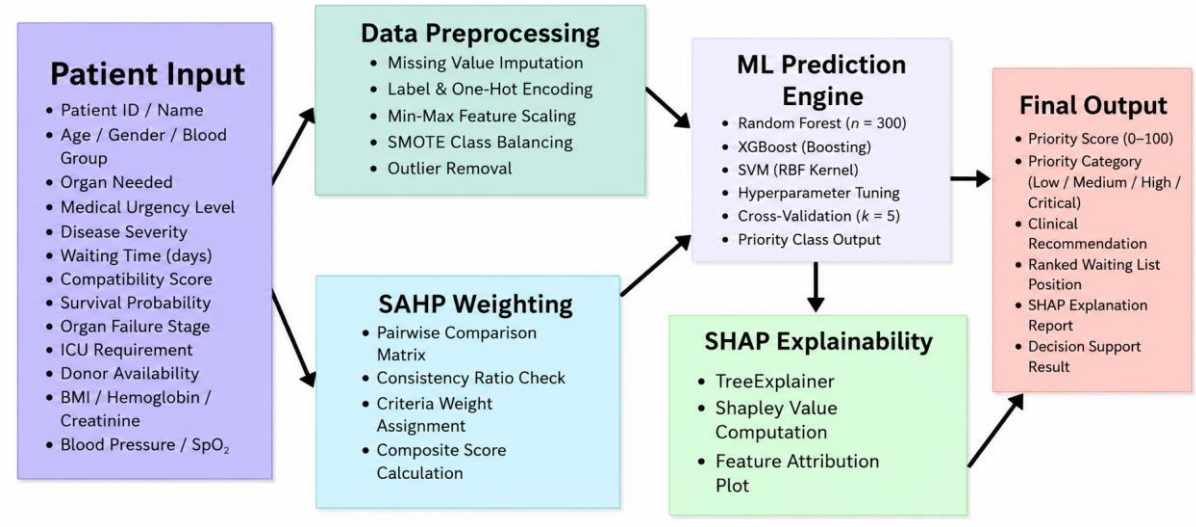

Figure 1: Block diagram – smart organ transplant priority analyser

Identifier fields, such as patient ID and name, are retained solely for record traceability and excluded from the predictive feature set to avoid leakage and unintended bias. This heterogeneous mixture of categorical, ordinal, and continuous variables is inherently unsuitable for direct model ingestion, necessitating a dedicated preprocessing stage. During preprocessing, missing values are imputed to prevent the systematic exclusion of incomplete records, which would otherwise disproportionately remove more severe or complex cases from the dataset. Categorical variables, including blood group and gender, are transformed via label and one-hot encoding to avoid the imposition of false ordinal relationships. Continuous features are normalized using min-max scaling, a step of particular importance for distance-based classifiers, such as the support vector machine employed later in the pipeline. Given the natural class imbalance typical of priority-labelled clinical datasets, where critical cases are underrepresented relative to low- or medium-priority cases, the Synthetic Minority Oversampling Technique

(SMOTE) is applied to synthetically balance the training distribution rather than relying on naive duplication. Finally, outliers are removed to eliminate physiologically implausible or erroneous entries that would otherwise distort feature scaling and downstream model training. In parallel with data preprocessing, an independent expert-knowledge pathway is constructed using a Saaty-based Analytic Hierarchy Process (SAHP), which encodes structured clinical judgment into a quantitative weighting scheme.

Criteria relevant to transplant prioritization, such as disease severity, urgency, compatibility, and waiting time, are compared pairwise to construct a comparison matrix that reflects their relative importance. The internal consistency of these pairwise judgments is verified using a consistency ratio derived from the matrix's principal eigenvector, ensuring that the resulting weights are not the product of contradictory or illogical comparisons. Once validated, criterion-specific weights are extracted and used to compute a composite priority score for each patient, providing a transparent, policy-anchored scoring mechanism that operates independently of any patterns the machine learning model may extract from historical data. The preprocessed feature set and the SAHP-derived composite scores converge at the machine learning prediction stage, where three classifier types are trained and evaluated: a random forest ensemble comprising 300 decision trees, an XGBoost gradient-boosting model, and a support vector machine with a radial basis function kernel. Each model undergoes hyperparameter tuning and is evaluated using five-fold cross-validation to obtain robust performance estimates and mitigate the risk of overfitting inherent to limited clinical datasets. The output of this stage is a predicted priority class for each patient, reflecting the integrated influence of both the learned statistical patterns and the expert-derived weighting scheme. To address the inherent opacity of ensemble and kernel-based models, the system incorporates a SHAP (SHapley Additive exPlanations) explainability module. A Tree Explainer, optimised for tree-based architectures such as random forest and XGBoost, is used to compute exact Shapley values efficiently. These values, grounded in cooperative game theory, quantify the marginal contribution of each input feature to a given patient's predicted priority score relative to the model's average output (Figure 2).

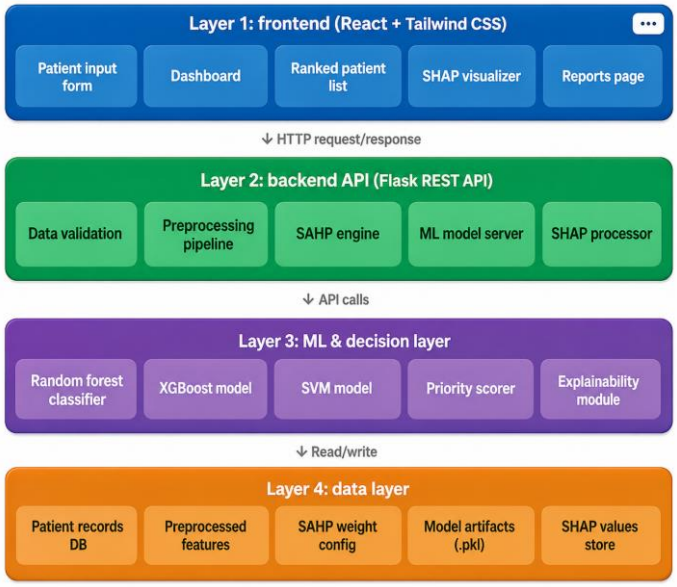

Figure 2: Architecture diagram – system layer overview

The resulting feature attribution plots allow clinicians to interpret which specific factors, such as organ failure stage or waiting time, most strongly influenced an individual prediction, thereby supporting clinical trust and accountability in the model's recommendations. The final output stage consolidates these components into a comprehensive decision-support package, comprising a numerical priority score on a 0–100 scale, a corresponding categorical designation (low, medium, high, or critical), a ranked position within the transplant waiting list, an associated clinical recommendation, and a SHAP-based explanation report. Rather than functioning as an autonomous decision-making system, the architecture is explicitly designed to support, rather than replace, clinical judgment, providing both a quantitative recommendation and the interpretive evidence necessary for clinicians to critically evaluate that recommendation before acting upon it. This architecture diagram illustrates the layered structure of the proposed system and shows how the application is organised into multiple functional modules. The frontend layer includes user interface pages such as patient input, the dashboard, result visualization, and the ranked patient list. The backend layer manages preprocessing, SAHP analysis, machine learning prediction, and SHAP-based explainability. The machine learning layer contains trained classification models such as Random Forest, XGBoost, and SVM, while the data layer stores patient records, processed datasets, and model outputs. This architecture ensures modularity, scalability, and easy integration of the system into a practical healthcare decision-support environment. The methodology begins with patient data

collection, preprocessing, SAHP-based weighting, machine learning prediction, and SHAP-based explainability. Patient attributes, including medical urgency, disease severity, compatibility score, waiting time, survival probability, organ failure stage, ICU requirement, and recovery chance, are used to classify patients into Low, Medium, High, and Critical Priority categories.

3.1. Experimental Setup

3.1.1. Training

The training phase begins by partitioning the preprocessed patient dataset into three distinct subsets: training, validation, and testing. This three-way split, rather than a simple train-test split, is essential for developing models that generalize reliably to unseen patients rather than merely memorizing patterns specific to the available data. The training subset is used to fit the parameters of each candidate model, the validation subset is used iteratively during hyperparameter tuning to select optimal model configurations without contaminating the final performance assessment, and the testing subset, held out entirely until final evaluation, provides an unbiased estimate of how each model would perform on genuinely new patients encountered in practice. Typically, such splits are stratified by priority class to ensure that the relative proportions of low, medium, high, and critical cases are preserved across all three subsets, a consideration that is particularly important given the class imbalance inherent in transplant priority data and addressed earlier through SMOTE-based oversampling during preprocessing. Three classification algorithms are trained on the structured transplant patient data: Random Forest, XGBoost and Support Vector Machine (SVM). Each of these algorithms was selected for its distinct strengths and its established track record in structured clinical data classification. Random Forest, an ensemble of decision trees trained on bootstrapped samples with randomised feature subsets, offers robustness against overfitting and provides an intuitive mechanism for later feature importance analysis. XGBoost, a gradient boosting framework that builds trees sequentially to correct the residual errors of prior iterations, typically achieves superior predictive accuracy on tabular datasets of this nature, though at some cost to training simplicity. SVM, employing a radial basis function kernel, is included for its ability to model complex, nonlinear decision boundaries between priority classes, particularly when class separability is not easily captured by tree-based partitioning alone.

Training each of these algorithms on the same preprocessed dataset allows for a controlled comparison of their relative predictive performance under identical data conditions. Following training, each model's performance is rigorously evaluated using a suite of standard classification metrics: accuracy, precision, recall, F1-score, and AUC-ROC. Accuracy provides a general measure of overall correct classification. Still, it is deliberately not relied upon in isolation, given its known tendency to be misleading in the presence of class imbalance, a concern directly relevant to this dataset despite prior SMOTE correction. Precision quantifies the proportion of patients predicted to belong to a given priority class who actually do, which is particularly critical for the "Critical" category, where false positives could result in the inappropriate allocation of scarce organs. Recall, conversely, measures the proportion of genuinely critical patients correctly identified as such, a metric of paramount importance in this clinical context, since failing to identify a truly critical patient (a false negative) carries far graver consequences than a false positive. The F1-score, as the harmonic mean of precision and recall, offers a single balanced measure that is particularly informative when the costs of false positives and false negatives are both clinically high, as they are here. Finally, the area under the receiver operating characteristic curve (AUC-ROC) evaluates each model's ability to discriminate between classes across all possible classification thresholds, providing a threshold-independent measure of overall discriminative power that is especially valuable when comparing models intended for deployment across varying clinical risk tolerances. Collectively, these metrics enable a multidimensional comparison of the three candidate models, informing the selection of the best-performing algorithm or combination of algorithms for subsequent deployment.

3.1.2. Implementation

Once trained and validated, the best-performing model, or an ensemble combination of the trained models, is integrated into a web-based decision support system designed for practical clinical use. This implementation phase translates the offline, research-oriented model development process into an operational tool capable of processing real-time patient data. Within this system, new patient records, comprising the same structured feature set used during training (demographic, clinical, compatibility, and laboratory variables), are submitted through a web interface and passed through the identical preprocessing pipeline established during model development, ensuring consistency between the conditions under which the model was trained and those under which it is subsequently applied. This includes the same encoding schemes, scaling transformations, and, where applicable, the SAHP-derived composite weighting, applied consistently to incoming data to preserve the integrity of the model's learned decision boundaries. Once preprocessed, incoming patient data is passed through the trained classification model, which outputs a predicted urgency category for the patient, functioning as the operational core of the decision-support tool. This prediction is not presented to clinical users in isolation; consistent with the explainability framework established earlier in the system architecture, each classification output is accompanied by a SHAP-based feature attribution report, allowing the clinician interacting with the system to understand the specific factors driving a given patient's predicted

urgency level. The web-based implementation is designed with clinical usability as a central consideration, providing an accessible interface through which hospital staff or transplant coordinators can input patient information, review model outputs, and access ranked waiting list positions without requiring any direct interaction with the underlying computational models. This design choice reflects the broader principle, established in the literature and reiterated throughout this study, that such systems are intended to augment, rather than replace, clinical decision-making, and their implementation must therefore prioritise transparency, accessibility, and ease of interpretation for end users who are clinical professionals rather than data scientists.

3.1.3. Evaluation

The final phase of the methodology involves a comparative evaluation of the proposed hybrid framework against traditional, established prioritization approaches, namely MELD-based scoring and manual, expert-driven prioritization methods. This comparative evaluation is essential for establishing the practical value of the proposed system beyond its performance in isolation; a model that performs well on standard classification metrics is of limited clinical significance unless it can be shown to offer tangible improvements over the methods currently in use. The Model for End-Stage Liver Disease (MELD) score and its analogous organ-specific counterparts represent the current clinical standard for prioritization in several transplant contexts, calculated from a small set of objective laboratory values to produce a numerical severity score that correlates with short-term mortality risk. While MELD-based scoring offers the considerable advantages of simplicity, objectivity, and clinical familiarity, it has been widely noted in the literature to rely on a comparatively narrow set of input variables, potentially overlooking relevant clinical, temporal, or compatibility-related factors that a more comprehensive, multi-criteria system, such as the one proposed here, is explicitly designed to incorporate. Manual prioritisation methods, by contrast, rely on the clinical judgment of transplant committees or individual physicians, offering the benefit of contextual, case-by-case nuance but introducing corresponding risks of inconsistency, subjectivity, and potential bias between different decision-makers or institutions, a concern directly raised in the ethical guidelines established by the World Health Organisation (Table 1).

Table 1: Patient inputs

Attribute	Patient 1	Patient 2	Patient 3
Age	56	49	63
Gender	Male	Female	Male
Organ Needed	Kidney	Liver	Heart
Medical Urgency	High	Critical	High
Disease Severity	Severe	Critical	Moderate
Compatibility Score	82	76	88
Waiting time (days)	210	145	320
Survival Probability	74%	61%	80%
Organ Failure Stage	Stage 4	Stage 5	Stage 3
ICU Requirement	Yes	Yes	No

The proposed framework is evaluated against these two baseline approaches along multiple dimensions, including predictive accuracy relative to observed patient outcomes, consistency of prioritisation decisions across comparable patient profiles, and the degree of transparency afforded to the decision-making process. Where retrospective patient outcome data is available, the priority classifications generated by each method (the proposed hybrid system, MELD-based scoring, and manual prioritisation) can be compared against actual clinical outcomes to assess which approach most accurately identifies patients at greatest need. Additionally, given that the proposed system's central contribution lies not only in predictive performance but in its integration of structured expert weighting (via SAHP) and post-hoc explainability (via SHAP), the evaluation also considers qualitative dimensions, such as the degree to which the system's recommendations can be traced to specific, clinically meaningful factors, a property that neither MELD-based scoring nor manual prioritization inherently guarantees in a systematic, reproducible form. Through this multidimensional comparative evaluation, the study aims to demonstrate that the proposed framework offers measurable improvements in both predictive performance and procedural transparency relative to existing prioritisation practices, thereby substantiating its potential value as a complementary decision-support tool within real-world transplant allocation workflows.

4. Results and Discussions

The results indicate that the proposed AI-based framework performs effectively in transplant priority classification. Random Forest achieved the strongest performance among the tested models, followed by XGBoost and SVM. The system demonstrates high predictive accuracy and strong generalisation ability.

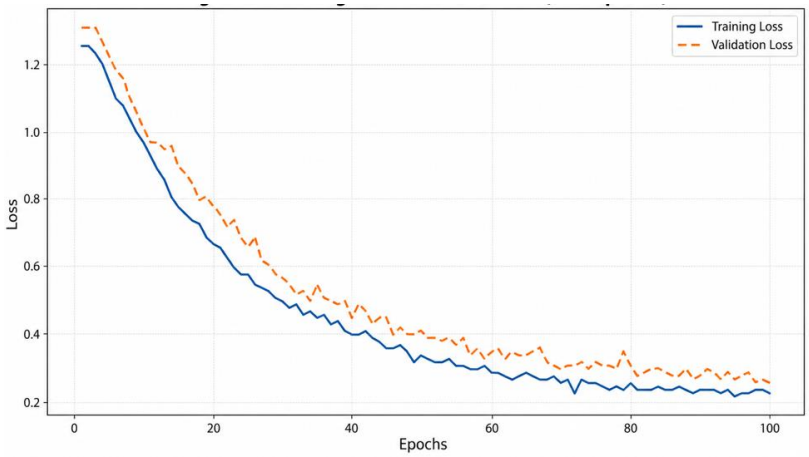

Figure 3: Training and validation loss (100 epochs)

Figure 3 shows the model's learning behaviour over 100 epochs. The curve demonstrates a steady reduction in loss during the early epochs, indicating that the model quickly learns the important patterns in the transplant patient dataset (Figure 4).

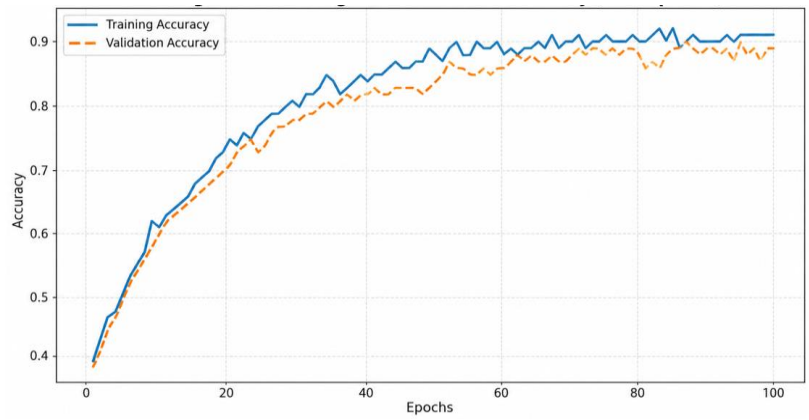

Figure 4: Training and validation accuracy (100 epochs)

As training continues, the loss gradually stabilises, showing that the model is converging effectively without major instability. The validation loss follows a similar trend to the training loss, indicating that the model generalizes well to unseen patient data and does not suffer from severe overfitting.

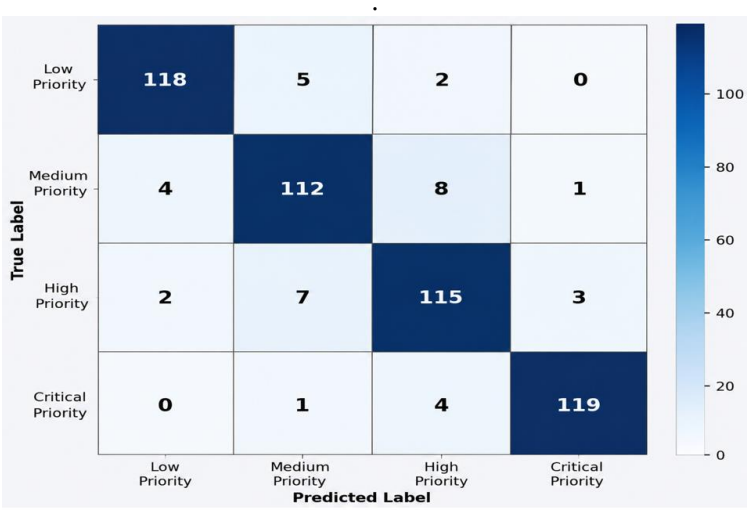

Figure 5: Confusion matrix – transplant priority classification

This result confirms the stability and reliability of the training process. Figure 5 shows the training and validation accuracy of the proposed model across 100 epochs. The accuracy increases significantly during the initial training phase, indicating that the model is learning useful relationships among the patient features. As the epochs progress, the accuracy continues to improve and stabilises at a high level, reflecting strong predictive capability. The validation accuracy remains close to the training accuracy, suggesting that the model performs consistently on unseen data and has good generalization ability. This result demonstrates that the machine learning model is effective at classifying transplant priorities. The confusion matrix illustrates the classification performance of the proposed transplant prioritisation model across the defined urgency categories. Most values are concentrated along the diagonal, indicating that the majority of patient cases are correctly classified into their actual priority levels. Only a small number of cases are misclassified into neighboring classes, as expected in borderline clinical situations where patient conditions may overlap. The matrix confirms that the model performs well in distinguishing between Low, Medium, High, and Critical transplant priority groups, thereby supporting its reliability for clinical decision support.

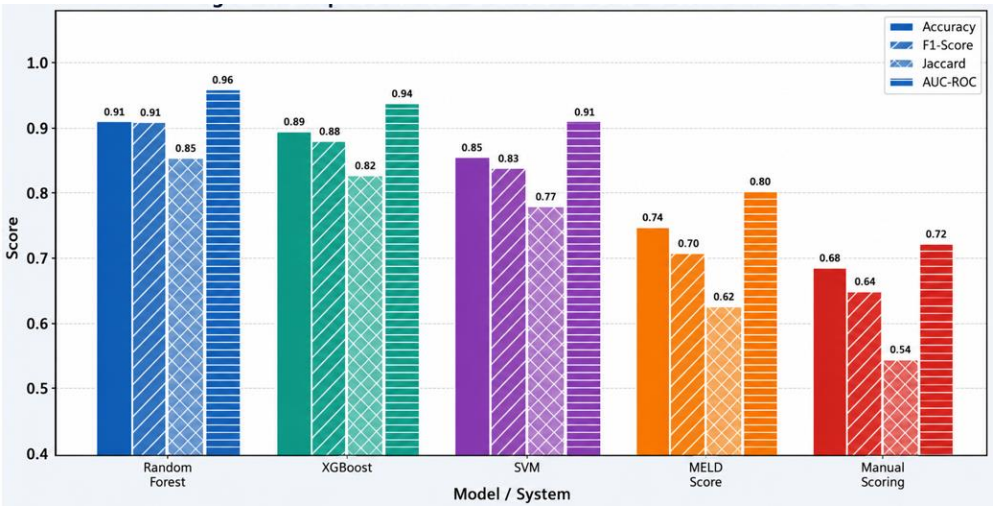

Figure 6: Comparative performance of models across evaluation metrics

Figure 6 compares the performance of different prioritisation approaches across multiple evaluation metrics including Accuracy, F1-Score, Jaccard Score, and AUC-ROC. The Random Forest model achieves the strongest overall performance, followed by XGBoost and SVM, while conventional approaches such as MELD score and manual scoring perform significantly lower. The result demonstrates that machine learning-based prioritization offers superior predictive reliability compared to traditional systems (Figure 7).

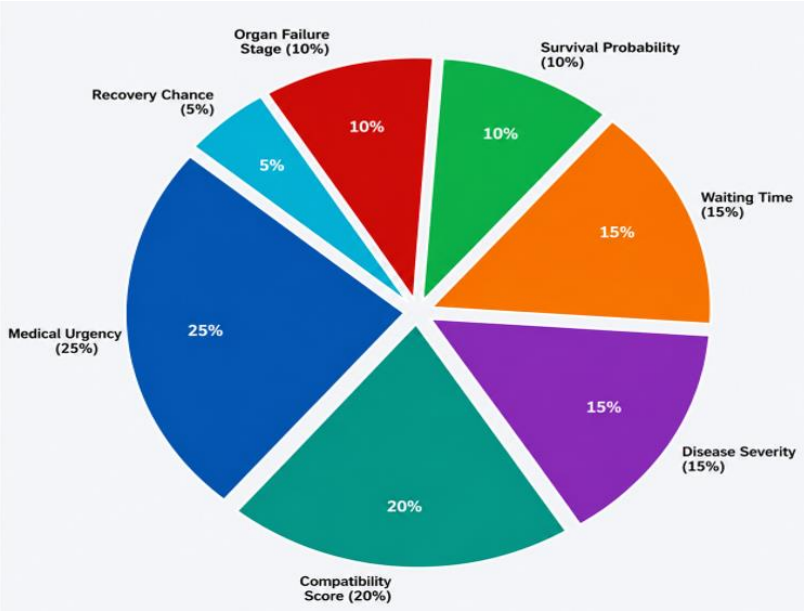

Figure 7: SAHP criteria importance weights

This pie chart illustrates the SAHP-assigned importance weights for the major transplant decision criteria. Medical urgency receives the highest weight, followed by Compatibility Score, Waiting Time, and Disease Severity. Survival Probability, Organ Failure Stage, and Recovery Chance are also included as supporting decision factors. These weights ensure that the final patient prioritization aligns with clinically meaningful transplant decision-making logic.

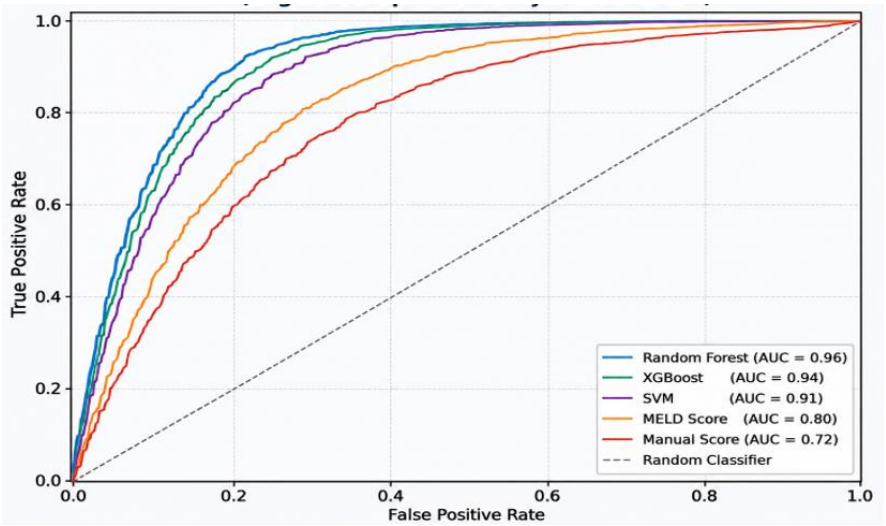

Figure 8: ROC curves – multi-model comparison

The ROC curve comparison evaluates the ability of the different models to distinguish between transplant priority classes. A higher area under the curve (AUC) indicates better classification performance. The Random Forest model achieves the highest AUC among the compared methods, demonstrating its superior ability to correctly classify transplant urgency categories. XGBoost and SVM also perform well, while conventional scoring methods exhibit lower discrimination capability. Figure 8 confirms that the proposed machine learning framework is highly effective for transplant-priority prediction and supports selecting Random Forest as the best-performing model.

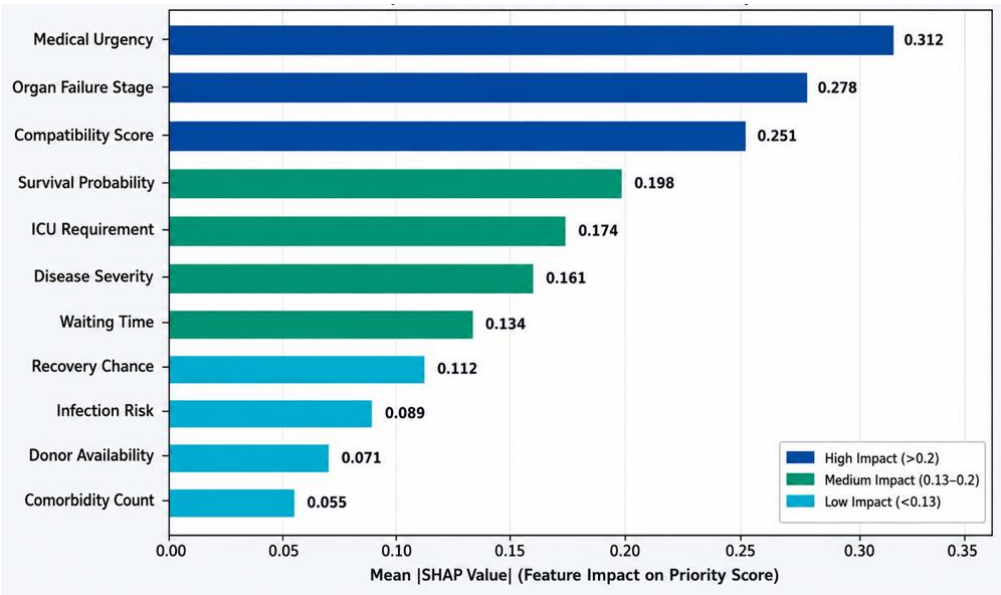

Figure 9: SHAP feature importance – mean absolute Shapley values

This SHAP feature importance plot shows how each input feature influences the model's final prediction. Features such as Medical Urgency, Organ Failure Stage, Compatibility Score, and Survival Probability are shown to have the greatest impact on transplant priority outcomes. Lower-ranked features still contribute to the prediction but with relatively smaller influence. Figure 9 is important because it enhances transparency and helps clinicians understand why the model assigned a particular patient to a specific urgency category. Thus, SHAP enhances the interpretability and trustworthiness of the proposed system.

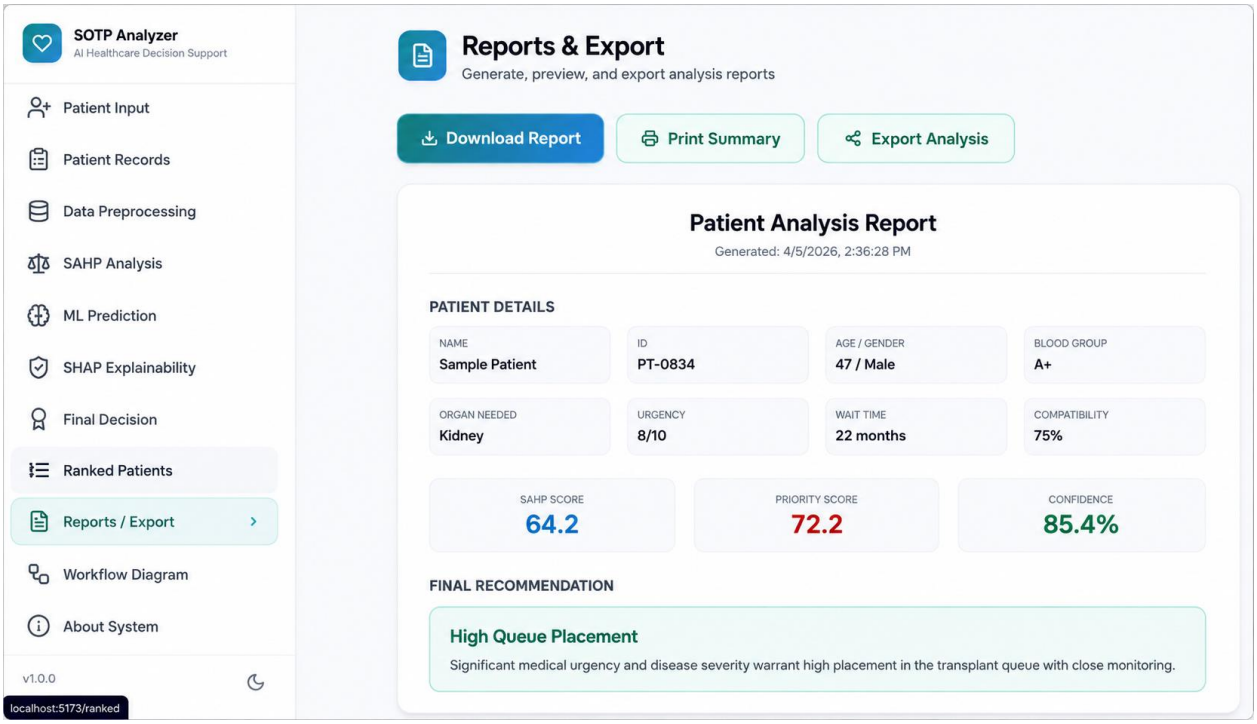

Figure 10: Output of SOTP analyser

Figure 10 shows the final output interface of the Smart Organ Transplant Priority Analyser, which displays the patient's transplant priority results after analysis (Table 2).

Table 2: Comparison of results of organ transplant priority prediction

Metric	Our Paper	Existing Models
Algorithm Used	Random Forest, XGBoost, SVM	MELD, Manual Scoring
Accuracy	0.89 – 0.91	0.68 – 0.80
F1-Score	0.88 – 0.91	0.64 – 0.70
AUC-ROC	0.91 – 0.96	0.72 – 0.80

The interface presents key outputs, including the patient's priority score, urgency category, confidence level, recommendation, and other supporting details, in a user-friendly format. It helps users and clinicians quickly understand the system's final decision. This output screen demonstrates the practical usability of the application and shows how the system's analytical components are translated into a clear, interactive healthcare decision-support tool.

5. Conclusion

This paper presents a Smart Organ Transplant Priority Analyser with Machine Learning, SAHP, and SHAP developed to mitigate the issue of fair and efficient organ allocation. Given the scarcity of donor organs and the growing number of patients on transplant waiting lists, it is necessary to prioritise patients correctly. The proposed system considers a set of medical parameters, including urgency, disease severity, waiting time, compatibility score, survival probability, and recovery chances. The system thus aims to provide a structured, more consistent approach to transplant decision-making by integrating these factors into a single analytical framework. Using SHAP enables the system to assign appropriate weights to each medical criterion, so that the most critical factors have greater influence on the final decision. Random Forest is the classification model used for transplant priority prediction. Besides, the introduction of SHAP enhanced transparency by providing explanations of each feature's contribution to the final prediction. Thus, prediction plus explanation makes the system more trustworthy and easier to understand, considering its sensitivity in the medical field. Overall, the developed web-based application is an exemplary case of how Artificial Intelligence and decision analysis techniques can be leveraged in medical decision support systems. Besides increasing accuracy in patient prioritisation, it also provides transparency in the use of the suggested solutions. The paper highlights the potential of combining Machine Learning with its explainable and structured methods to support fair, transparent, and data-driven healthcare solutions.

Acknowledgment: The authors sincerely acknowledge the academic support and institutional contributions provided by Purdue University, SRM Institute of Science and Technology at Ramapuram, S.A. Engineering College, S. Khan Lab, Khyber Medical University, Beanibazar Cancer and General Hospital, and Zewail City of Science and Technology.

Data Availability Statement: The data for this study can be made available upon request to the corresponding author.

Funding Statement: This manuscript and research paper were prepared without any financial support or funding.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare. This work represents a new contribution by the authors, and all citations and references are appropriately included based on the information utilized.

Ethics and Consent Statement: This research adheres to ethical guidelines, obtaining informed consent from all participants. Confidentiality measures were implemented to safeguard participant privacy.

References

1. L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 10, pp. 5–32, 2001.
2. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," *arXiv preprint arXiv:1705.07874*, 2017. [Accessed by 05/05/2025]
3. S. M. Lundberg, G. G. Erion, H. Chen, A. J. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "Explainable AI for trees: From local explanations to global understanding," *arXiv preprint arXiv:1905.04610,* 2019. [Accessed by 06/05/2025]
4. A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable AI systems for the medical domain?", *arXiv preprint arXiv:1712.09923*, 2017. [Accessed by 08/05/2025]
5. T. L. Saaty, "Decision making with the analytic hierarchy process," *International Journal of Services Sciences*, vol. 1, no. 1, pp. 83–98, 2008.
6. T. L. Saaty, "The analytic hierarchy process: planning, priority setting, resource allocation", *McGraw-Hill International Book Co,* New York, United States of America, 1980.
7. T. Chen and C. Guestrin, "XGBoost: A scalable tree boosting system," in *Proc. 22nd ACM SIGKDD Int. Conf. Knowledge Discovery and Data Mining*, California, United States of America, 2016.
8. R. Bellazzi and B. Zupan, "Predictive data mining in clinical medicine: Current issues and guidelines," *International Journal of Medical Informatics*, vol. 77, no. 2, pp. 81–97, 2008.
9. E. J. Topol, "High-performance medicine: The convergence of human and artificial intelligence," *Nature Medicine*, vol. 25, no. 1, pp. 44–56, 2019.
10. A. Esteva, A. Robicquet, B. Ramsundar, V. Kuleshov, M. DePristo, K. Chou, C. Cui, G. Corrado, S. Thrun, and J. Dean, "A guide to deep learning in healthcare," *Nature Medicine*, vol. 25, no. 1, pp. 24–29, 2019.
11. A. Rajkomar, J. Dean, and I. Kohane, "Machine learning in medicine," *New England Journal of Medicine*, vol. 380, no. 14, pp. 1347–1358, 2019.
12. A. Rajkomar, E. Oren, K. Chen, A. M. Dai, N. Hajaj, M. Hardt, P. J. Liu, X. Liu, J. Marcus, M. Sun, P. Sundberg, H. Yee, K. Zhang, Y. Zhang, G. Flores, G. E. Duggan, J. Irvine, Q. V. Le, K. Litsch, A. Mossin, J. Tansuwan, D. Wang, J. Wexler, J. Wilson, D. Ludwig, S. L. Volchenboum, K. Chou, M. Pearson, S. Madabushi, N. H. Shah, A. J. Butte, M. D. Howell, C. Cui, G. S. Corrado, and J. Dean, "Scalable and accurate deep learning with electronic health records," *NPJ Digital Medicine*, vol. 1, no. 18, pp. 1-10, 2018.
13. E. H. Shortliffe and J. J. Cimino, "Biomedical Informatics: Computer Applications in Health Care and Biomedicine," *Springer*, London, United Kingdom, 2021.
14. World Health Organization, "Transplantation," *Health Topics,* 2026. [Accessed by 15/06/2026]
15. World Health Organization, "WHO guiding principles on human cell, tissue and organ transplantation," *WHO Publication*, 2010. [Accessed by 20/05/2025]