BRAIN. Broad Research in Artificial Intelligence and Neuroscience
e-ISSN: 2067-3957 | p-ISSN: 2068-0473
Covered in: Web of Science (ESCI); EBSCO; JERIH PLUS (hkdir.no); IndexCopernicus; Google Scholar; SHERPA/RoMEO; ArticleReach Direct; WorldCat; CrossRef; Peeref; Bridge of Knowledge (mostwiedzy.pl); abcdindex.com; Editage; Ingenta Connect Publication; OALib; scite.ai; Scholar9; Scientific and Technical Information Portal; FID Move; ADVANCED SCIENCES INDEX (European Science Evaluation Centre, neredataltics.org); ivySCI; exaly.com; Journal Selector Tool (letpub.com); Citefactor.org; fatcat!; ZDB catalogue; Catalogue SUDOC (abes.fr); OpenAlex; Wikidata; The ISSN Portal; Socolar; KVK-Volltitel (kit.edu)
2026, Volume 17, Issue 3, pages: 215-235
Submitted: May 24th, 2026 | *Accepted for publication:* July 3rd, 2026

MITOF: An Integrated AI Framework for Forecasting and Structured Reporting in Clinical Immunosuppressive Therapy Management

Rareș Arvinte*
Alexandru Ioan Cuza University of Iași, Romania.
raresarv1@gmail.com,
https://orcid.org/0000-0002-8494-4441

Diana Trandabăț
Alexandru Ioan Cuza University of Iași, Romania.
diana.trandabat@gmail.com,
https://orcid.org/0000-0002-9874-0642

Dan Cristea
Alexandru Ioan Cuza University of Iași, Romania;
Romanian Academy, Institute of Computer Science Iași, Romania.
https://orcid.org/0009-0009-3362-3304

* *Corresponding author*

Abstract: *This retrospective framework study presents Medical Integrated Therapy Optimisation and Forecasting (MITOF), an integrated clinical artificial-intelligence framework for immunosuppressive therapy in renal transplantation, in which patient risk stratification, dosage optimisation, dosage forecasting and structured clinical reporting are combined into a single workflow with role-based views for doctors, patients and hospital management. The framework was validated on Romanian retrospective clinical datasets, with most components grounded in a renal-transplant cohort of 22,885 transplant follow-up hospitalisation records between 2007 and 2022. The predictive components were validated retrospectively against recorded outcomes: the survivability classifier reached an accuracy of 92.9% and an area under the curve of 0.843, the LSTM forecaster outperformed an ARIMA baseline on irregular tacrolimus series, and the dosage model was most stable with the Huber loss and the Adam optimiser. We are careful about what each result establishes: the dosage model reproduces clinical practice rather than proven optimality, and the quality and usefulness of the generated report, together with the clinical value of the platform, will be evaluated through a planned clinician reader study and a prospective pilot. The main contribution is the integration itself, with role-based reporting and value-level traceability, validated module by module on Romanian transplant data. At the platform level, MITOF is organised around three clinical-AI requirements: clinical reviewability, role-specific abstraction and value-level provenance, so that model outputs remain inspectable rather than acting as autonomous decisions.*

Keywords: *immunosuppressive therapy; clinical report generation; dosage forecasting; optimisation; hospital data; clinical decision support.*

How to cite: Arvinte, R., Trandabăț, D., & Cristea, D. (2026). MITOF: An integrated AI framework for forecasting and structured reporting in immunosuppressive therapy management. *BRAIN. Broad Research in Artificial Intelligence and Neuroscience*, *17*(3), 215-235. https://doi.org/10.70594/brain/17.3/14

1. Introduction

Renal-transplant follow-up provides a clinical setting in which decision support must remain longitudinal, individualised and reviewable. Immunosuppressive therapy, especially with tacrolimus, cyclosporine and mycophenolate, must be adjusted in relation to laboratory findings, prescription history, hospitalisation context and patient trajectory. A useful clinical platform therefore must assist the physician in reviewing these elements faster, without replacing the medical decision-making process.

The corresponding computer science problem is that the relevant evidence is not stored as a single, analysis-ready object. In the Romanian hospital environment considered here, the patient history is distributed across hospital files, laboratory analyses, medication records, prescriptions and short epicrisis texts. The manual assembly of this information is slow, and the need to repeatedly complete or interpret medical records reduces the time available for direct patient care. Obtaining relevant patient information quickly, and generating structured documentation from verified data, could improve the efficiency of the clinical workflow.

Existing clinical decision-support and report-generation solutions usually address only one aspect of this problem. Dosage estimation, risk stratification, time-series forecasting and clinical text generation are often studied as separate tasks, with limited connection to the patient record and to the type of report that a doctor, a patient or a hospital manager can actually use. For immunosuppressive therapy, this separation is especially problematic because dosage decisions are longitudinal, the time series are often irregular, and the final output must remain traceable to the source data.

This retrospective framework study on Medical Integrated Therapy Optimisation and Forecasting (MITOF) is presented as an integrated framework that connects these directions into a single workflow. The platform is structured as a three-layer model. The data layer accesses hospital files, anonymises them, links the required parameters, and transforms them into structures that can be used by several modules. The analysis layer applies risk stratification, dosage estimation, dosage forecasting, and epicrisis processing methods. The reporting layer organises the outputs into role-based views for clinical review, patient communication and management-level resource planning.

To make the framework suitable for clinical AI use, MITOF follows three design requirements. Clinical reviewability means that the system prepares evidence for medical inspection rather than replacing clinical judgement. Role-specific abstraction means that the same source data are transformed into different views for doctors, patients, and management. Value-level provenance means that numerical values displayed in the report remain traceable to their source rows and are clearly separated from model predictions and generated explanations.

Through this organisation, the same source data can support several user needs. Physicians require concise clinical summaries that highlight laboratory values, medication history, trends, and warnings. Patients need simplified visual explanations of their treatment progress without unnecessary technical terminology. Hospital management requires aggregated information about medication consumption, possible future resource needs, and logistical planning. The platform therefore does not alter the clinical process but offers a structured interface through which the existing data and model outputs can be reviewed.

The reporting layer is designed to preserve data provenance. Each displayed value is bound to its source from the hospital dataset and, when possible, to the corresponding hospitalisation period and the national medical registry. Predictive values are clearly distinguished from measured values, and generated explanations are separated from clinical observations. This distinction is essential because the report must remain useful while making visible the difference between data, model output and medical judgement.

This article is positioned as a retrospective framework study describing an integrated platform. It separates the technical contributions of data preprocessing, dosage-estimation experiments, forecasting experiments and structured report generation from the proposed evaluation

and implementation pathway. The predictive components are validated retrospectively against recorded outcomes, while the quality and usefulness of the generated report, together with the clinical value of the integrated platform, will be evaluated through a planned clinician reader study and a prospective pilot.

The contributions of this work are threefold. First, it integrates modules that are usually studied in isolation, namely survival risk, dosage estimation, dosage forecasting and structured clinical reporting, around a single role-based report for doctors, patients and management, with every displayed value traceable to its source. Second, it validates the predictive modules retrospectively on multiple Romanian clinical datasets, validation being reported separately because different modules rely on different sources instead of a single dataset. Third, it shows that the dosage model reproduces clinical practice and the quality of the report and the clinical value of the platform need to be further clinically validated.

2. State of the Art

2.1. Immunosuppressive Therapy

Renal-transplant follow-up provides the clinical background for the framework, as immunosuppressive treatment must be interpreted in the context of longitudinal laboratory results and medication history. This background explains why optimisation and forecasting are relevant to the platform and why their outputs must be presented through a medically reviewable report.

Immunosuppressants are drugs administered in the context of organ transplantation with the aim of preventing allograft rejection (Halloran, 2004). Following transplantation, the immune system of the patient may recognise the donated organ as foreign tissue. Donor cells express alloantigens that differ from those of the patient which will prompt the action of cytotoxic T lymphocytes (CTL) to specifically target the foreign tissue. Consequently,specific immunosuppressive agents are required to manage allograft rejection, as complete immunological tolerance to the transplanted organ is not yet achievable in clinical practice (Starzl & Zinkernagel , 2001).

Immunosuppressants fall into various groups, namely those with induction agents and those with maintenance agents. Induction medications aim to eradicate alloreactive lymphocytes from the host immune system when it is first exposed to alloantigens following transplantation. Maintenance medications are designed to prevent lymphocyte-mediated graft rejection.

These drugs are beneficial for immunosuppression; however, they may eventually require dosage escalation or a transition to alternative medications. In this context, these medications possess adverse effects that necessitate the administration of additional pharmaceuticals to mitigate their cardiovascular and metabolic impact, as they compromise patient immunity over time and render it vulnerable to external agents (Miller, 2002).

For this research, however, we focused on maintenance immunosuppressants, specifically cyclosporine and tacrolimus. Cyclosporine (CsA) was identified in 1972 in Switzerland and revolutionised the domain of organ transplantation. However, cyclosporine exhibits nephrotoxic effects, which can be managed by adjusting the administered dosage or discontinuing treatment. Consequently, throughout the years, the primary factor contributing to the inadequate long-term survival of grafts has been chronic calcineurin (CNI) nephrotoxicity (Tedesco & Haragsim,, 2012). Besides its impact on immunological function, CsA exhibits various other toxic effects. Acute and chronic nephrotoxicity are the primary concerns, although additional issues may encompass gingival hyperplasia, hyperuricemia, hyperlipidaemia, neurotoxicity, thrombotic microangiopathy, hypertension, hypomagnesemia, and hyperkalaemia (Kahan,1989). Consequently, there is a desire to determine the best well balanced dosage to maintain a patient's health effectively.

Tacrolimus is an immunosuppressive agent akin to cyclosporine, yet far more powerful, that also suppresses the immune response. Due to its higher potency, there are challenges in dosing it as effectively as possible. It binds the cytoplasmic immunophilin FKBP12 forming a complex that

inhibits calcineurin signalling which reduces calcium-dependent T-cells from activating. Therefore, tacrolimus is regarded as an option for persons undergoing retransplantation who encounter liver or kidney graft rejection or drug-induced toxicity from therapy. This immunosuppressant, similar to cyclosporine, induces nephrotoxicity, with a comparable frequency of adverse effects observed in both instances of drug administration. Tacrolimus and cyclosporine should not be administered in the same treatment scheme because of the risk of nephrotoxicity (Peters et al., 1993).

This brief presentation indicates that the medications assisting transplant recipients have various adverse effects on the immune system, which can be observed in the long term. Dosing still depends mostly on the physician's own experience and on repeated manual adjustment for each patient. An automatic system that gives faster suggestions would clearly help, but the decision-support tools reported so far tend to stop at a single decision and rarely follow the dosage over time, which is exactly what transplant follow-up needs.

2.2. Dosage Prediction and Forecasting

A second direction of work directly covers predicting drug levels and dosages.Yoon et al. (2024) developed a machine-learning model that predicts tacrolimus concentration and suggests a dose based on the values from the previous three days for hepatic transplantation, and Choshi et al. (2025) adopted a similar approach using an LSTM for personalised tacrolimus dosing after lung transplantation. These are the closest works to what the present framework addresses, but both assume that the input data arrive at fairly regular intervals, and both treat dosing as an isolated task, without linking the prediction to the rest of the patient record or to a report that a doctor would actually use. Regarding loss functions, Meyer (2021) provides a probabilistic interpretation of the Huber loss, which supports our preference for Huber over standard MAE or MSE when modelling noisy dosage values.

For the time-series part, ARIMA and its relatives are still the usual baseline. Riaz et al. (2023) compare ARIMA, SARIMA, and Holt-Winters for epidemiological forecasting and report satisfactory results when the series are smooth, which is consistent with the observed behaviour in the present experiments, as ARIMA performs well only when dosage trends exhibit limited variability. A separate line of work forecasts hospital resources rather than single patients. Koç and Türkoğlu (2022) forecast medical-equipment demand with a deep LSTM, while Murray et al. (2023) and Johnson et al. (2023) forecast ICU census and ward-level bed needs for planning. These confirm that demand forecasting is useful for management, but they work at the aggregate level from the start; none of them builds the management view by summing up per-patient predictions, which is the link we try to make between the clinical and the administrative side of the platform. Together with the report-generation functionality, the recurring gap is integration: dosing, forecasting and the resource view are each solved separately, rarely on irregular transplant data, and almost never inside a single workflow.

2.3. Clinical Report Generation

The reporting component depends on natural language processing, since part of the patient record is textual and the final clinical document must be readable by doctors, patients, and researchers. This requirement connects structured prediction outputs with the textual information already present in clinical records.

Regarding language technologies and their impact on medical research, the focus is on using different NLP tools to extract and generate features,especially when the medical dataset contains only unstructured data (Yu et al., 2014). In practice, standard methods are applied to search for specific words or words with an increased frequency (Friedman et al., 1994). This raises the issue that medical images must be accompanied at least by the expertise of a doctor in order to be able to build a set of features using computational linguistics.

Recently, we have seen a shift from simple text generators to integrated reports generated in platform systems. Rodrigues and Lopes (2025) studied summary generation approaches with large

language models that process clinical documentation rather than just a generic summarisation task. Palm et al. (2025) evaluate AI-generated clinical notes with a high focus on quality that can be improved by humans. Both works, however, consider discharge text in English and judge mostly the language quality, not whether the numbers are right. For transplant follow-up this is the weak point: a fluent note that states the wrong dose is worse than no note, so a quality score on its own does not tell us if the report is safe to use.

A second direction focuses on the integration of multiple tools in the same systems. RadAlign introduces concept alignment for radiology report generation, while RadVLM explores a conversational vision-language model for multiple radiology functions (Gu et al., 2025; Deperrois et al., 2025). The design extends beyond radiology; a clinical reporting platform should preserve notes between source, concepts, and the final generated report. Both are still tied to radiology images and to one modality, while our input is free-text epicrises together with laboratory and medication tables. Their alignment between image and text cannot be carried over directly to this case.

Salme et al. (2025) explore a report generator adapted to using low-resource language models, arguing that clinical languages, available datasets, and local documentation patterns can influence report quality. This observation is relevant in the context of Romanian hospital data, where clinical records and prescribing practices are reported using conventions that vary from institution to institution. Their evaluation, though, stays on radiology corpora. Romanian free-text epicrises, which are short, written in a compact and somewhat irregular style, have to our knowledge not been used as a basis for report generation, and this is one of the gaps the present work tries to fill.

A certain concern remains grounded in report generation. As Bose et al. (2025) focus on visual alignment in medical vision-language models, the importance of connecting the generated report to the available visual evidence is highlighted. It proposes that each report component should be explained through the patient dataset, analysis dataset, medication dataset, or explicit model output. In the present case, the difference is that the grounding is not to an image but to a row in the patient, analysis or medication data, so every value in the report can be traced back to where it came from. This kind of provenance is rarely treated explicitly in the works above.

Pulling these observations together, a few gaps stay open. Most clinical report generation is built for radiology and for high-resource languages, so Romanian free-text epicrises are left aside, and the generative models read well but give no guarantee that a dose or a date is correct, which is the part that matters in transplant care. Dosage optimisation and forecasting are usually studied on their own and on regular time series, while transplant data are irregular and the two tasks are rarely brought together in one workflow a doctor can actually review. Predictive outputs are seldom connected to a reporting layer that serves the doctor, the patient and the administration from the same source, with values that can be traced back. And negative results, such as unsupervised clustering that does not separate infected patients on imbalanced data, are rarely reported, so the limits of these methods stay unclear. The present work addresses these gaps by bringing the prediction, forecasting and reporting modules together around a role-based report, and by validating each module on retrospective data, including the cases where it did not work.

3. Data and Ethics

The study design is retrospective and uses a single-centre cohort from the Transplant Department, covering admissions recorded between January 2007 and July 2022. The data were provided as four linked CSV datasets: patients (25,644 admission records), medication (60,442 rows), analyses (1,197,540 rows), and prescriptions (8,914 rows), all keyed on the combination of the FO (personal health record) and PreCode identifiers. Inclusion was limited to records of patients hospitalised for kidney transplantation; after excluding non-day admissions, 22,885 admission records remained for modelling. The patients had a mean age of approximately 40 years (range: 3–93 years); the most frequently performed laboratory analysis was creatinine measurement, and

prednisone was the most frequently administered drug, accounting for approximately 3% of all administrations. The retrospective use of anonymised clinical records was approved under the institutional research agreement with the hospital administration and was conducted exclusively for non-commercial scientific purposes. The records were anonymised before release, with all direct identifiers such as names removed in line with the General Data Protection Regulation (EU Regulation 2016/679); access was restricted to authorised research personnel, and the data were held in an access-controlled environment. The need for individual consent was waived because the data were anonymised before release and no intervention or patient contact was involved.

The principal problem addressed at the institutional level is one of optimisation, namely how immunosuppressant dosages can be optimised using neural networks. Because of the large volume of records, data access was specified as a minimal, explicit list of the fields required for the study. The requested data involved patients who were hospitalised for kidney transplant. The original data structure was preserved, with the records stored as CSV files organised into three datasets: patients, analyses, and medication.

The **patient dataset** includes personal information, such as age, gender, diagnoses and hospitalisation period, the analysis dataset contains all the details about the analysis performed on patients during their hospitalisation period, and the medication dataset comprises the administered medications along with their corresponding dosages. Every entry in the patient dataset is associated with multiple entries from the analyses and medication datasets, which are linked to patient-specific medical records. The data layer is deliberately described in detail because the value of the platform depends on how well hospital records can be identified, linked, cleaned, and made usable for the prediction and reporting modules.

Medical department: This field contains two text values: "kidney transplant" and "transplant day admission". The "kidney transplant" type medical files have a higher consistency of data because these files are monitored for several days, more precisely for the entire period that the patient was hospitalised after the transplant. The "transplant day hospitalisation" type medical files represent the file of a post-transplant patient who came for periodic control and these files contain repetitive entries viewed from the administered treatment perspective, the differences being in the possible dosages, depending on the analysis performed. A combination of these two columns and the year of patient registration provides a unique identifier across all CSV files for reliable longitudinal patient tracking.

Age: This column records the patient's age and is important because patients are monitored over several years; consequently, age changes over time and may influence automated dosage recommendations.

Epicrises: A noteworthy detail is presented in the column referring to the epicrises, which is a succinct text edited in professional terminology, detailing a patient trajectory during hospitalisation. Doctors may subsequently use this document during hospital transfers or when reviewing the patient's medical history. This column contains complete textual data from 2018 onwards, when recording such information became mandatory.These entries usually follow a specific pattern in how they are written, which makes them suitable for developing textual data extraction systems.This column may support future analyses, particularly from the perspective of computational linguistics and clinical text generation. It can also be used to create datasets for future analyses, or even for generating voice-based summaries that can be transmitted to doctors. Such tools can improve the efficiency of clinical workflows. Figure 1 displays the word-count distribution for each epicrises entry. The x-axis represents the number of words per epicrises entry, and the y-axis represents the frequency of records. The distribution shows that epicrisis texts are generally short and formulaic, which supports template-based extraction and summary generation.

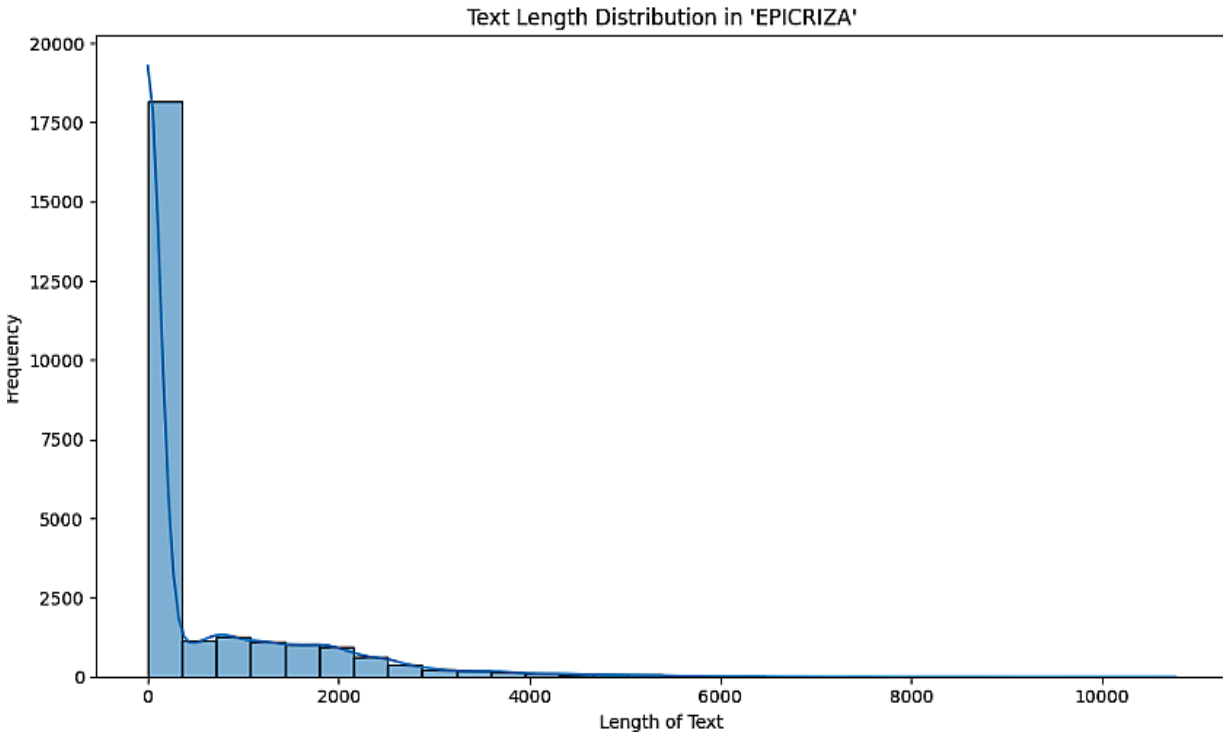


Figure 1. Epicrises text length distribution.

The patient dataset contains several columns, including year, hospital department, FO (personal health record), admission type, admission date, discharge type, discharge date, age, and primary and secondary diagnoses, with almost all fields being complete. The only columns with incomplete data are the epicrisis field (which became mandatory only after 2018) and the ATI field, which is completed only when a patient requires Intensive Care Unit (ICU) admission.

The **analysis dataset** contains all the entries related to the analysis performed by the patients during hospitalisation, including the collection date and their results. The main columns in this dataset are FO, age, ANANUME, and the list of laboratory analyses with their corresponding values. The **ANANUME** column contains the full name of each laboratory analysis. When a patient enters the hospital, a standard package of blood and urine tests is done. Following the results, other analysis dependent on the previous results will be made later.For example, if the white blood cell count exceeds the reference range, additional laboratory analyses may be required, highlighting the need to address such cases in future studies. Creatinine measurement is the most frequently performed laboratory analysis and, together with other clinical parameters, contributes to the assessment of graft function and guides dosage adjustment.

The **medication dataset** contains all the necessary information about the medicines administered to a patient (these include medicines prescribed to be taken outside the hospital), the name of the medicine according to its supplier, followed by the amount of active substance in its component or, in the case of intravenous solutions , the actual amount, the method of administration and the category of the drug. The date of administration, together with the administered dose and medication name, is essential for subsequent automated decision-making.

Since medication records contained missing values, a fourth dataset of prescriptions was obtained from the same provider to complete them. After some additional cleaning, this **prescription dataset** of 8,914 entries was used, with the following identification columns:

ActiveSubstance: as the name of the column implies, this column contains information about the active substance of the drug, not the name of the drug itself. This column can be used to improve the linkage between medication records and prescription records.

Concentration: the column contains information to identify the concentration of active substance in the tablets, which can be used to further convert the data into units of measurement. For the medications analysed in this study, concentrations are consistently recorded in milligrams.

Figure 2 presents the details of the prescription dataset; in this case, there are no missing data, as prescriptions were recorded individually. In this case, the data were not standardised, unlike those in the medication dataset.

drugName	643 non-null	object
activeSubstance	8914 non-null	object
prescriptionPackage	8914 non-null	int64
pharmaceuticalForm	8914 non-null	object
concentration	8914 non-null	object
icdName	8914 non-null	object
drug_code	8914 non-null	object
prescribedQuantity	8914 non-null	int64
dosage	8914 non-null	object

Figure 2. Columns of the prescription dataset

No personal data, such as names, contact details, or any other personal identifiers, are present in the datasets. Before being released for research purposes, this data underwent an extensive anonymisation process that removed all information capable of directly or indirectly identifying patients. Furthermore, this information is used strictly for scientific analytical purposes, without affecting individuals in any way. Thus, the data-processing activities are aligned with the GDPR principles of data minimisation and restricted access. According to GDPR requirements, access to the data is restricted to authorised research personnel, and all data is stored in an access-controlled environment.

4. MITOF Platform Architecture and Reporting Workflow

The MITOF architecture is organised around the path from anonymised hospital records to role-based reports. Data is collected from hospital files, preprocessed into usable structures, analysed by the predictive and optimisation modules, and finally presented through a reporting interface that can be reviewed by doctors, patients, and hospital management.

The platform does not introduce a separate clinical decision process. It uses the patient's medical history as input, links it with laboratory, medication, prescription and epicrises information, and exposes model outputs in a form that can be reviewed by a clinician. Figure 3 presents the platform as a layered workflow, from anonymised hospital data to analysis layer and role-specific reporting outputs.

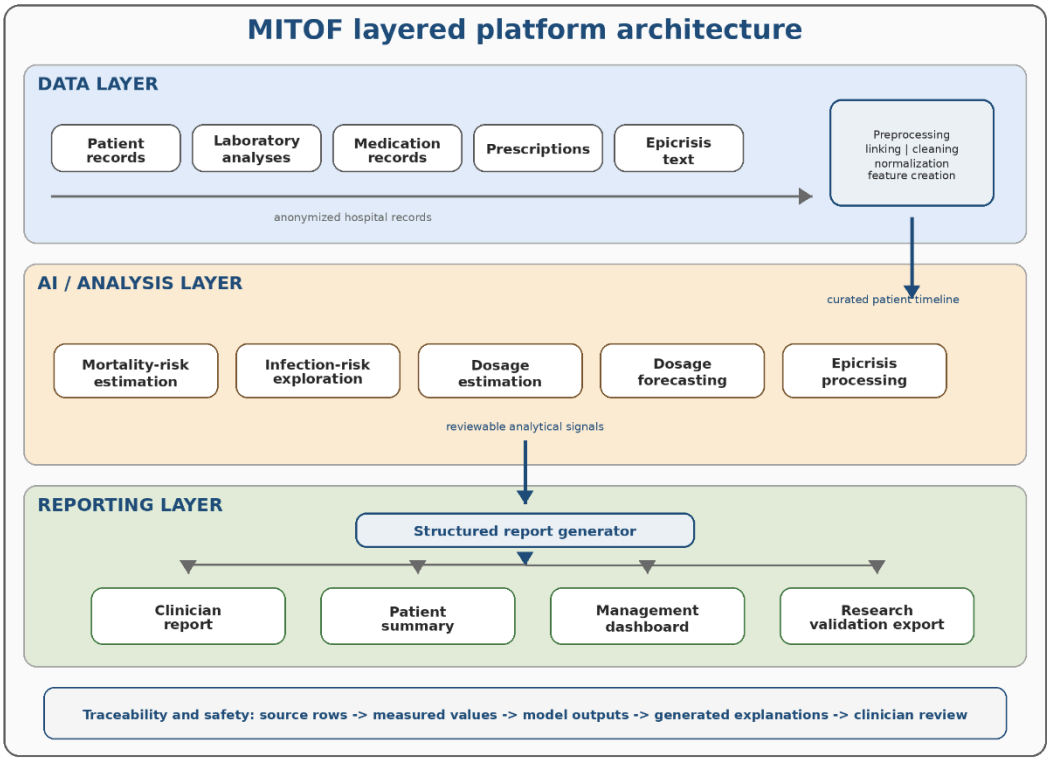


Figure 3. Layered MITOF platform architecture and reporting workflow. The bottom traceability path indicates that source rows, measured values, model outputs and generated explanations remain linked and visible as distinct categories for clinical review.

4.1. Platform Organisation

As shown in Figure 3, the platform can be described through three connected layers: a data layer, an analysis layer, and a reporting layer. The data layer prepares and links hospital records; the analysis layer transforms the linked data into risk, dosage, and forecasting signals; and the reporting layer converts these outputs into user-specific views. This separation is important because it keeps medical evidence, model outputs, and generated explanations clearly distinguishable.

These layers are governed by the same three clinical AI requirements. Clinical reviewability is addressed by keeping the doctor as the final interpreter of the report. Role-specific abstraction is addressed by producing different levels of detail from the same anonymised data source. Value-level provenance is addressed by preserving the link between displayed values, source records and model outputs.

Design rationale and user roles. The same patient history must serve different users. Physicians need a concise clinical view with diagnoses, recent laboratory analyses, medication history, warnings, and model outputs. Patients need a simplified explanation of treatment progress without unnecessary medical terminology. Hospital management needs aggregated information about medication use, resource consumption, and possible future demand. The role-based structure therefore adapts the same evidence base to different levels of detail.

Data layer. The data layer accesses the anonymised patient, analysis, medication and prescription files described in Section 3. It links records through the available hospital identifiers, normalises relevant analyses and medication names, handles missing or irregular values, and prepares the features required by the downstream modules. The epicrises column is treated as a textual source that can be summarised or attached to the report when available.

Figure 4 illustrates the type of longitudinal structured information that the data layer prepares for later interpretation. Together with Figures 1 and 2, it provides a broader overview of the clinical parameters and their distribution.

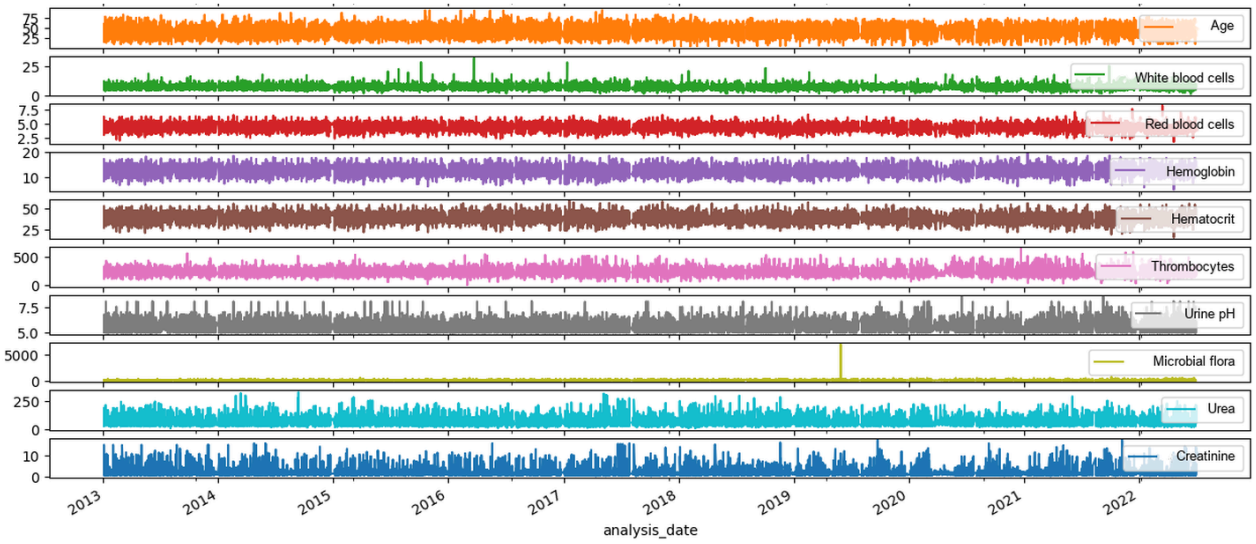


Figure 4. Longitudinal distribution of dataset features over time.

Analysis layer. The analysis layer contains the reusable AI modules of the platform. Their role inside MITOF is not to issue an automatic treatment decision, but to provide reviewable signals that are later checked in the validation setup and results sections. In this section, the modules are therefore described only by their place in the workflow: mortality-risk estimation, exploratory infection-risk grouping, dosage estimation and dosage forecasting.

Reporting layer. The reporting layer assembles the structured values, textual information and model outputs into a report. In the doctor's view, the report prioritises clinical review: patient status, diagnoses, laboratory findings, current medication and warnings. The patient view is intended for simpler explanation, while the management view exposes only aggregated or logistical information and should not include sensitive patient-level data.

Traceability and safety constraints. Each displayed value should remain connected to its source row, hospitalisation period or registry entry when this information is available. Measured values, predicted values and generated explanations are visually and conceptually separated. This design reduces the risk that the report appears as an autonomous medical decision rather than as a structured decision-support artifact.

Platform workflow example. When a doctor searches for a transplant follow-up patient, the system retrieves previous admissions, recent analyses and medication history, processes the available epicrisis text, runs the relevant modules, and generates a structured report. The doctor can then review the values, compare them with the previous visit, evaluate the model outputs, and add or correct clinical observations before the information is used further.

4.2. Validation Setup

For reproducibility and to clarify the statistical outputs, the model-level details that are not part of the platform architecture are consolidated in this subsection. The architecture section defines how the modules are connected in the workflow, while the validation setup records the datasets, preprocessing choices, splits, hyperparameters and baselines, used to evaluate each component retrospectively. For the classification modules, accuracy, precision, recall, F1-score and AUC were used to interpret the results. Regression and forecasting modules are interpreted with error-based measures against recorded clinical data. The clustering module uses silhouette score for internal structures and AUC after cluster labels are mapped, to detect the presence of the toxic outcome label.

Modules included in MITOF were not all developed on the same dataset. Dosage estimation, forecasting and structured reporting use a renal transplant dataset. Mortality prediction uses a renal

comorbidities dataset and infection-risk analysis was developed on a separate retrospective infectious patients dataset, all being integrated as reusable framework components.

For the dosage estimation, forecasting and structured reporting modules, the models were trained on the curated cohort of 22,885 day-admission records; numeric features were standardised, missing analysis values were imputed with the patient mean and set to NaN when an analysis had never been performed, and missing medication values were imputed with the patient average or with zero when no administration was recorded. The data were split into 80% for training and validation and 20% for testing.

Survival prediction used a feed-forward neural network with the all-cause mortality column as the binary target, trained with binary cross-entropy. The final feature set was a reduced selection of laboratory and demographic variables, for example white-cell count, age, gender and environment, rather than the full 240-column table, after duplicate columns were removed and zero-filled placeholders were replaced by column means. The baseline is the same network trained on the raw, imbalanced data, which reached only 43% accuracy before cleaning and re-balancing.

Dosage optimisation used a feed-forward regressor with an input layer and two hidden layers, predicting the dose for tacrolimus (Prograf), mycophenolate and cyclosporine, with mycophenolate added as an input when modelling Prograf. Training used the Adam optimiser with a learning rate of 1e-3 and the Huber loss over 100 epochs; longer runs of 500, 1000 and 1500 epochs were also tested. The comparators were the same model trained with mean-squared-error and mean-absolute-error losses, and with stochastic gradient descent instead of Adam, while the historically administered doses serve as the clinical reference.

Table 1. Model configuration and retrospective validation setup

Module	Task / model	Data and features	Preprocessing / split	Hyperparameters / baseline
Mortality	Feed-forward NN classifier (binary, all-cause death)	Dialysis survivability dataset containing 4,114 patient records with demographic, clinical, and laboratory variables. Lab and demographic features (white cells, age, gender, environment) extracted from 240 columns	Duplicate columns removed; zero placeholders to column mean; standardised; 80/20 split	Binary cross-entropy; baseline = raw imbalanced model (43% accuracy)
Dosage optimisation	Feed-forward regressor (input + 2 hidden layers)	Renal-transplant cohort from main dataset - 22,885 transplant day-admission records Tacrolimus (Prograf), mycophenolate, cyclosporine; mycophenolate added as input for Prograf	Per-patient mean / zero imputation; standardised; 80/20 split	Adam, lr 1e-3, Huber loss, 100 epochs (also 500/1000/1500); baselines = MSE, MAE, SGD; reference = administered doses
Forecasting	LSTM vs ARIMA baseline	Same renal-transplant cohort - 22,885 records Tacrolimus and mycophenolate concentration series (irregular)	Per-series; 80/20 split	LSTM: Adam, Huber, batch 32, ~40 epochs; baseline = ARIMA (3 previous values)
Infection risk	K-Means + DBSCAN (unsupervised)	Infectious-diseases clinical corpus, 6,835 instances after TF-IDF representation; PCA-reduced features; binary toxin-presence label	PCA; cluster-to-label mapping by majority voting	DBSCAN eps 0.5, min samples 8; 19 clusters, metrics = silhouette, AUC =0.465
Report generation	Template / rule-based generator	Renal-transplant cohort - 22,885 records Structured fields from patient, analysis and medication data	Values bound to their source row (traceability)	Comparator = general LLM (ChatGPT-GPT-4); evaluation = reader study (planned)

Forecasting compared an LSTM with an ARIMA baseline on the tacrolimus and mycophenolate concentration series. The LSTM used the Adam optimiser, the Huber loss, batches of 32 and about 40 epochs, while ARIMA used the three previous observations as input. The series are irregular, without a fixed sampling interval, which is the main difficulty of this task. The

nosocomial-infection module used K-Means and DBSCAN (eps 0.5, minimum samples 8) on PCA-reduced features, and cluster labels were mapped to the binary toxin-presence outcome by majority voting so that supervised metrics could be computed.

Because the modules were developed and validated on partly distinct datasets, Table 1 reports the validation source separately for each component. The renal-transplant cohort of 22,885 day-admission records supports the dosage-estimation, forecasting and structured-reporting components, whereas the survivability and nosocomial-infection modules originate from separate retrospective datasets and are included as reusable framework components rather than as results obtained from the transplant cohort itself.

5. Clinical Report Generation for Medical Use

5.1. Report Generation Principles

The reporting layer is the point at which the predictive modules become clinically interpretable. Its role is not to replace medical judgement, but to assemble patient history, laboratory trends, medication history, model outputs, and editable observations into a structured document adapted to each user role.

Raw outputs from the previous workflows are converted into a clinical and user-friendly form. The report generator structures survival information, infection exploration, tacrolimus and mycophenolate forecasts, and dosage-estimation outputs into a readable medical document. It improves interpretability by displaying not only received data, but also brief explanations, contributions and alerts.

The design is intended to reduce the risk of unsupported generated text. Instead of asking a language model to generate a medical narrative, the system assembles a report from patient data, templates and model outputs. Measured data, predictive values and generated explanations remain separable and visible to the doctor through the interface and the labels used in the report.

This design is also supported by the comparison between the structured report generator and text generation models such as ChatGPT (version: GPT-4), as discussed in relation to Tanno et al. (2025). Automatically generated reports without a stable structure do not have a clear basis for justifying their correctness, while the structured generator preserves the expected fields and the corresponding source values. Ganzinger et al. (2025) report a related approach based on modifiable drafts, whereas the present system operates on several predictive outputs, including risk, dosing and forecast values, not only narrative discharge text. As a future improvement, the generated report could be evaluated using BERTScore against clinician-corrected reference report.

5.2. Doctor Report

The doctor-facing report should be concise and evidence-based. It focuses on laboratory values, trends, warnings and a brief epicrisis, while leaving the final interpretation to the clinician. The platform-generated report is split into six blocks. The first one identifies the patient record through anonymised technical identifiers. The second summarises the most relevant historical data, particularly previous hospitalisations. The third presents laboratory values and highlights changes from one visit to the next. The fourth presents medication and prescription information, including current medication and quantities that may be needed in the future. The fifth displays model outputs for optimisation and forecasting. The final block is reserved for new clinical observations, so that generated content can be validated and updated before official recording.

To make the reporting component easier to interpret, Figures 5 to 7 present an example of how the generated report can be displayed to a hospital doctor. The example uses a transplant follow-up report translated into English and adapted to the clinical platform interface. It is included as a doctor-facing example for clinical review and research validation.

Figure 5 shows a mock-up interface and does not contain real patient data. The datasets used for model development and retrospective validation were anonymised before research. In the future

in a deployed hospital environment, information such as the patient name would be visible only to authorised clinical users under GDPR compliant access policies. The doctor can enter a patient identifier bound to the hospital system, and the system generates a quick overview of the patient status, including a short epicrises generated from previous visits.

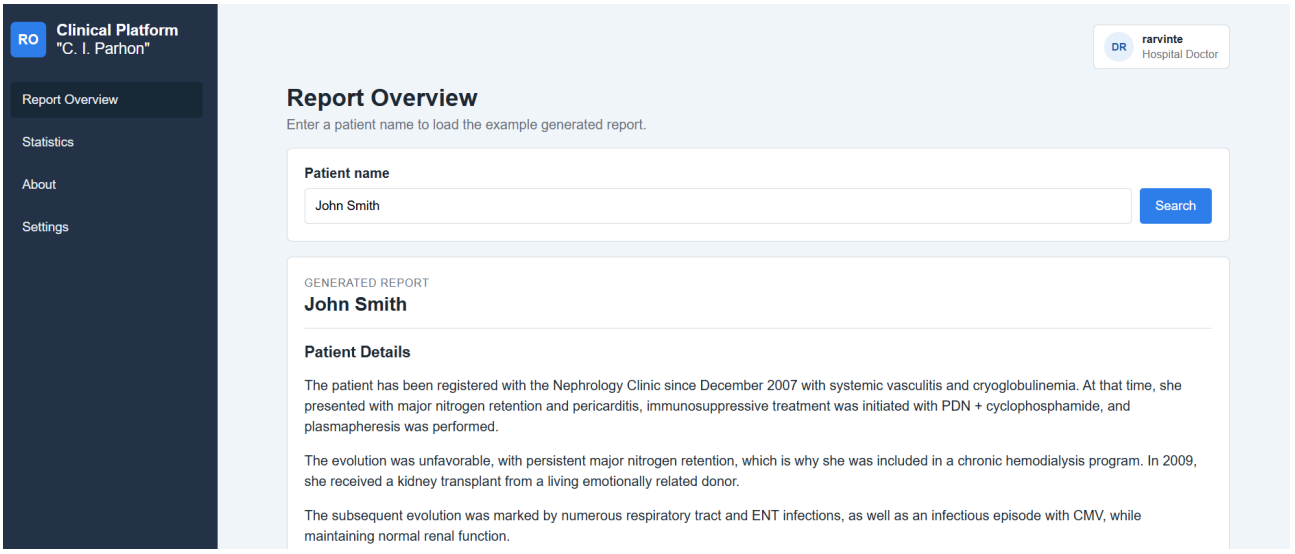


Figure 5. Doctor-facing report overview inside the clinical platform.

Figure 6 continues the preview of the generated report. The main and secondary diagnoses are separated into two blocks, allowing the doctor to rapidly verify the clinical context before proceeding to the laboratory analysis and medication sections.

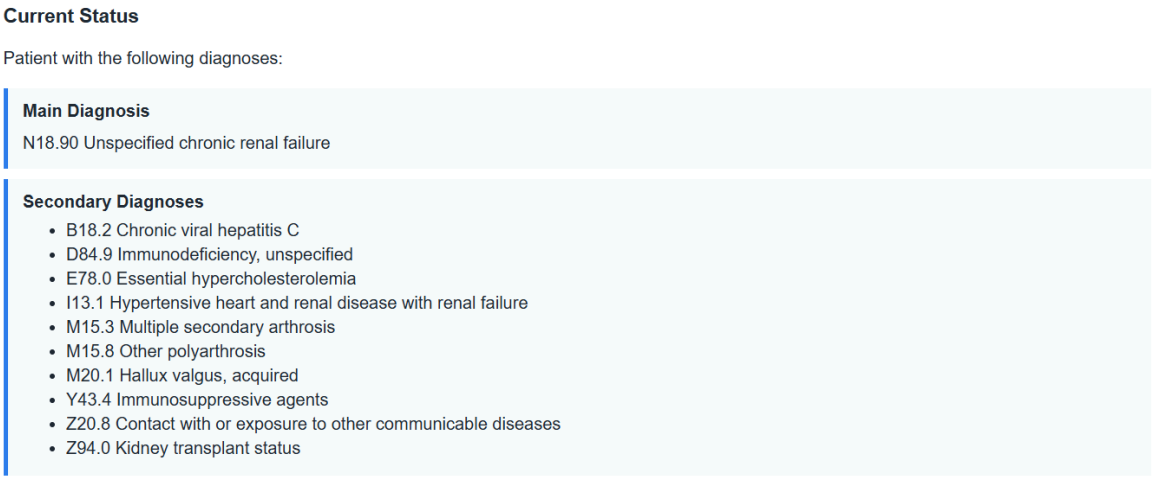
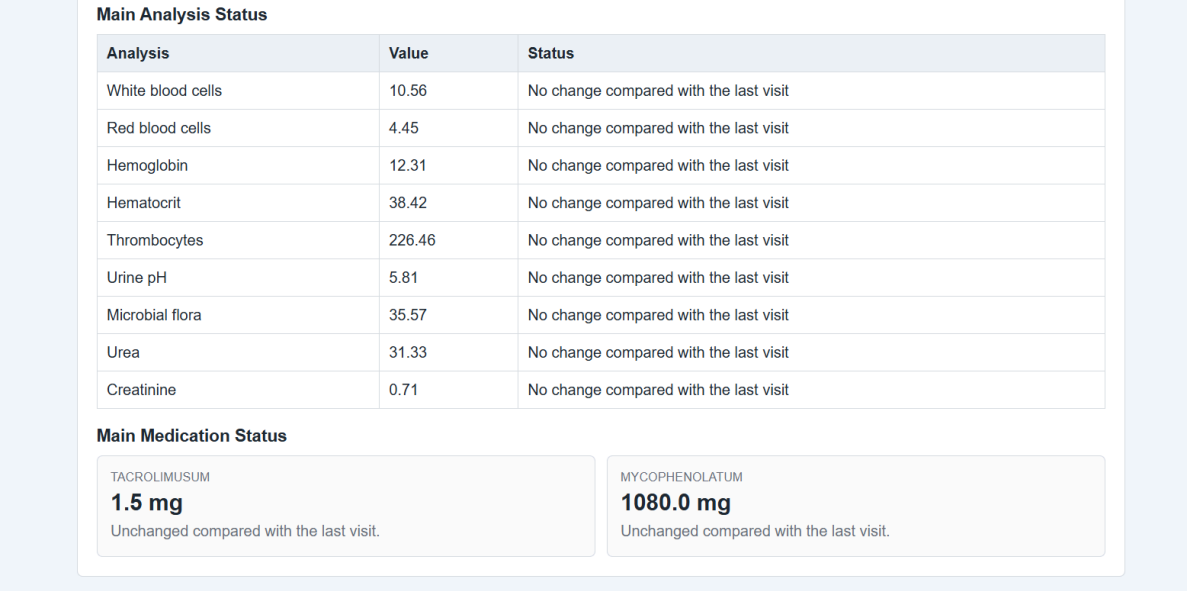


Figure 6. Current status and diagnosis section displayed in the generated report.

Figure 7 presents the main analysis and medication sections. Laboratory values are displayed in table format and compared with those from the previous visit. Immunosuppressant-related values are shown separately because the template is customised for the renal-transplant environment and for the needs of those patients.



Main Analysis Status

Analysis	Value	Status
White blood cells	10.56	No change compared with the last visit
Red blood cells	4.45	No change compared with the last visit
Hemoglobin	12.31	No change compared with the last visit
Hematocrit	38.42	No change compared with the last visit
Thrombocytes	226.46	No change compared with the last visit
Urine pH	5.81	No change compared with the last visit
Microbial flora	35.57	No change compared with the last visit
Urea	31.33	No change compared with the last visit
Creatinine	0.71	No change compared with the last visit

Main Medication Status

TACROLIMUSUM
1.5 mg
Unchanged compared with the last visit.

MYCOPHENOLATUM
1080.0 mg
Unchanged compared with the last visit.

Figure 7. Main analysis and medication status sections displayed in the generated report.

5.3. Patient Report

The patient report uses the same structured source, but with a simplified terminology, and does not include the technical epicrises. Its role is to explain the progress of treatment, the main laboratory trends, and possible recommendations in a visual and comprehensible manner. The language remains close to the clinical facts, but without exposing unnecessary technical details or creating unsupported medical advice.

5.4. Management Report

The management report does not expose sensitive patient-level information. It focuses on aggregated medication use, forecasted dosage needs, and resource-oriented indicators, enabling hospital administrators to anticipate future costs, restocking needs, and logistical constraints. In this manner, the same source data produce different abstractions for doctors, patients, and management while preserving the separation between clinical care and resource planning.

6. Results

This section reports the retrospective results obtained for each module integrated into the platform. All values were obtained on anonymised hospital data, and full experimental details are available in (Arvinte & Trandabăț, 2023) for the optimisation module or (Arvinte & Trandabăț, 2025) for a first version of the forecasting module. The results are organised per direction and are reported as obtained. Retrospective backtesting is the right tool for the predictive and forecasting modules, where each prediction can be checked against a recorded outcome. The report-generation module and the wider claim of clinical usefulness are not of this kind; they are treated separately, as a clinician reader study and a prospective evaluation, in the following section.

Patient health evolution module. A neural network classifier was trained to predict all-cause mortality during dialysis follow-up. Class balancing and feature cleaning were decisive: test accuracy increased from 43% on the raw, imbalanced data to approximately 94% after duplicate-column removal, mean-value imputation, and binary re-labelling. On the held-out test set (20% of the data) the classifier reached an accuracy of 92.9%, an AUC of 0.843, a precision of 0.87, a recall of 0.64, and an F1-score of 0.738, with a mean absolute error of 0.074 and a mean squared error of 0.067. These values indicate that the module separates surviving and deceased patients well above chance and can support the risk-stratification view offered to clinicians and administrators.

Nosocomial-infection module. An unsupervised clustering approach (K-Means and DBSCAN) was evaluated for infection-risk grouping on a Romanian medical corpus. DBSCAN

produced nine clusters with a silhouette score of 0.91, indicating internally coherent groups. However, when these clusters were mapped to the binary toxin-presence label, the resulting AUC was 0.465, that is, no better than chance. This negative result is reported deliberately: with the current feature representation the unsupervised clusters do not align with the clinical target, so the module is retained as an exploratory signal rather than a decision input.

Dosage-optimisation module. A neural network regressor estimated immunosuppressant dosages for tacrolimus (Prograf), mycophenolate, and cyclosporine. Across 100 epochs, the Huber loss combined with the Adam optimiser produced the most stable convergence, outperforming mean-squared-error and mean-absolute-error objectives; stochastic gradient descent converged poorly and was discarded as unsuitable for the dosage distribution. Cyclosporine data, owing to their more uniform distribution, yielded a faster and more stable learning rate than the Prograf data. These findings show that stable optimisation and disciplined feature selection matter more than network size for clinically usable dosage estimation.

Figure 8 reports one mycophenolate dosage-estimation example after 100 epochs, showing the type of model output used when comparing optimization configurations.

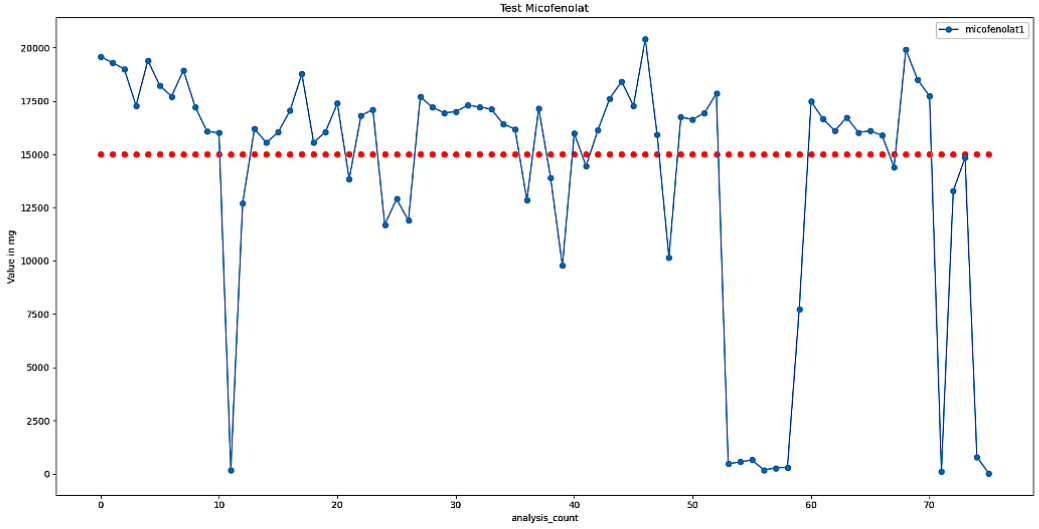


Figure 8. Mycophenolate dosage-estimation results after 100 training epochs

Forecasting module. Time-series models were compared for forecasting tacrolimus and mycophenolate concentrations. The LSTM model outperformed ARIMA in capturing the non-linear fluctuations of tacrolimus levels, with lower error across most tested configurations, whereas ARIMA produced acceptable results only when the dosage trend was approximately linear and non-fluctuating. A hybrid LSTM-ARIMA model is identified as a promising future direction.

Figure 9 reports the ARIMA forecasted values, which are used as the statistical baseline for the forecasting module.

Report-generation module. The structured, template-based report generator was compared with a general-purpose generative model (ChatGPT, version: GPT-4). The structured generator preserved data fidelity, in that every displayed value is bound to its source and predictive outputs are kept exactly as computed, whereas the unstructured generative output lacked a verifiable basis for its content. This supports the design decision to keep the structured report as the authoritative artefact and to use generative models only for optional rewriting or simplification.

Because the forecasting module estimates per-patient dosage trajectories, these trajectories can be aggregated across a department to anticipate medication demand and to inform restocking and budgeting. This management-facing view is a direct extension of the per-patient forecasts and constitutes the integrating contribution of the platform.

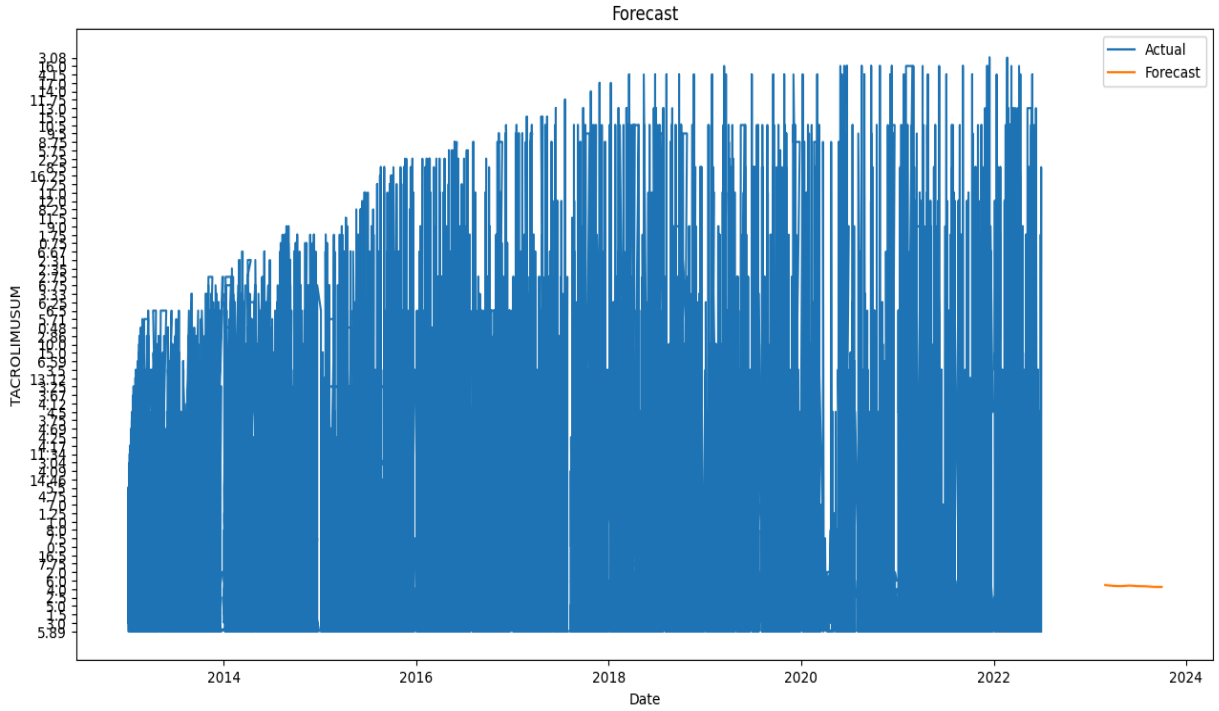


Figure 9. ARIMA forecasting results used as the statistical baseline

7. Proposed Platform Evaluation

Because the integration is presented as a modular framework, its evaluation should not use only one prediction metric (see Table 2). A complete evaluation can use multiple perspectives, like model plausibility, report quality, clinical readability, or time required by a doctor to review the medical report.

The predictive and forecasting modules were already validated retrospectively in Section 6. What remains, and is described below, is a clinician reader study for the generated report and a prospective pilot for the platform as a whole; these are planned rather than carried out here.

A comprehensive evaluation framework can be organised around four validation layers. The data layer can be evaluated by measuring completeness of the patient, analysis, medication, and prescription datasets after preprocessing. The model layer can be evaluated retrospectively by comparing forecasted values or recommended dosage trends with historical clinical evolution. The report layer can be evaluated by clinicians using completeness, correctness, readability, and usefulness scores. The clinical-use layer can be evaluated using a structured review test in which doctors, patients, and research reviewers compare platform reports with manually assembled medical information.

An additional evaluation dimension should address the physician's cognitive load during report interpretation. Because the platform combines tables, trends, forecasts, and generated explanations, clinical usefulness should not be assessed only by correctness or readability. A report that is factually accurate can still be unsafe or inefficient if it overloads the physician during time-constrained review.

The proposed clinician-reader study should therefore include a cognitive-load assessment, for example using the NASA Task Load Index (NASA-TLX), with subscales for mental demand, temporal demand, effort, frustration, and perceived performance. The same cases can be presented in different interface conditions, such as mainly textual reports versus reports with visual trend summaries, and compared by task-completion time, interpretation errors, self-reported mental effort, and confidence in the decision.

Table 2. Proposed evaluation layers for the integrated platform

Layer	Evaluation target	Possible metric	Data source	Interpretation
Data	Completeness and usability after preprocessing	Missing-value rate; valid date fields; linked patient-analysis-medication records	Anonymised hospital datasets	Shows whether the reporting pipeline has enough structured input to operate reliably.
Model	Forecasting and dosage optimisation	Retrospective deviation from historical dosage trends; error by medication and patient period	Historical prescriptions and analyses	Checks whether model outputs are clinically reviewable before report use.
Report	Clinical documentation quality	Completeness; correctness; readability; traceability; review time	Generated platform reports reviewed by clinicians	Determines whether the report helps doctors interpret patient status faster.
Cognitive load	Mental effort and frustration during report interpretation	NASA-TLX subscales; task-completion time; interpretation-error rate; comparison of visual and textual report conditions	Clinician reader study using generated platform reports	Evaluates whether the interface reduces cognitive burden while preserving clinically relevant information.
Patient view	Accessible explanation	Readability score; comprehension questions; user feedback	Patient-oriented report export	Evaluates communication without excessive medical terminology.
Research validation	Traceability and reporting readiness	Time to assemble information; report-generation coverage; data-quality indicators	Research review of anonymised clinical records	Measures whether the platform supports reproducible medical analysis and report validation.

The expected outcome is not to demonstrate that the platform replaces clinical judgement. We aim to determine whether the platform can shorten the process of reviewing multiple patients' historical data by providing only the necessary information and optimised suggestions for the clinician to consult.

Future evaluations could report both numerical and qualitative outcomes. The numerical side can include data completion and value comparison. The qualitative evaluation can focus on clinical review of the reports by doctors and patients, which can come back with feedback. These results should be presented separately so that model performance is not confused with documentation usefulness.

The first empirical evaluation should use retrospective records because they allow the predicted or summarised output to be compared with the patient's known history. For a sample of selected transplant follow-up cases, the platform can generate specialised reports, and ask clinicians to score the reports for correctness, completeness, usefulness, and time saved. A second reviewer can inspect whether the values displayed in the report can be traced back to the original anonymised datasets. This creates a quality-control process that is closer to medical practice than a single automatic score.

A second evaluation could be conducted by an independent researcher, who reviews the reports and assigns a score to each. The reviewer can also document missing or incorrect information identified during the evaluation. These evaluations should be conducted independently and would improve the medication-optimisation process.

A final evaluation should follow failed reports. Failed reports should be evaluated separately, focusing on the missing text values in the epicrises, which could provide insights into how the epicrises would have been written in the past, and the module to learn to improve them. Recording these cases is useful because it shows where the platform needs human correction or

additional data. A realistic results chapter should therefore include both successful outputs and limitations, since a hospital platform is valuable only when its uncertainties are visible.

7.1. Reader Study Design

The report-generation module is the component that cannot be adequately evaluated using retrospective metrics, so we describe here a reader study that can be conducted on the same anonymised records without collecting new patient data. The aim is to ask transplant specialists to evaluate the generated reports directly, as they would evaluate a colleague's summary, while separately verifying that every value in the report can be traced back to the source data.

The primary outcomes are the correctness and traceability of the structured report; the secondary outcomes are its usefulness and review time compared with the two alternative reports, doctors manually prepared and the LLM generated version. Agreement between raters will be assessed using, for example, Cohen's kappa for categorical variables and the intraclass correlation coefficient (ICC) for the one-to-five rating scales; disagreements will be resolved by a third specialist. Because the sample is expected to be small, differences between the three report versions will be analysed using non-parametric tests, such as the Friedman or Wilcoxon signed-rank test, rather than assuming a normal distribution. Reports that fail or are incomplete, usually because the epicrisis text is missing, will be analysed separately, since they indicate where the module requires human correction or additional data. The whole study runs on anonymised records under the existing institutional agreement, and no identifiable information leaves the controlled environment.

7.2. Implementation and Validation Path

A practical implementation would begin with a limited pilot deployment in a renal transplant unit rather than hospital-wide implementation. The deployed version could use anonymised records from patients who agree to participate in the study, generate reports for a fixed number of transplant follow-up cases, and compare the platform output with reports manually assembled by clinicians. Its purpose is to collect relevant information, present the model outputs in a controlled clinical environment, and allow the clinician to determine whether the generated content is useful.

The system administrator can also define roles for each clinical and research user category. The clinician reviews the patient-level report and corrects any values or explanations that are not appropriate. The patient-facing version is reviewed for clarity and to ensure it does not contain unnecessary technical language. The research reviewer inspects indicators such as missing analyses, incomplete medication records, and the number of cases for which a report can be generated. This role separation makes the platform easier to evaluate because each user group is asked to judge only the part of the platform that corresponds to its medical or research activity.

Another implementation requirement is version control to track the generated report, record when it was edited, and maintain multiple versions for backup and recovery. This is relevant for future research and accountability. This can also provide evidence regarding data quality and indicate whether the platform is reliable or requires further improvement. The platform should remain adaptable to future improvements, allowing the models to evolve while maintaining compatibility with new clinical requirements. A modular architecture is adopted in this implementation. This also prevents the clinical workflow from requiring complete redesign whenever individual modules are updated.

This evaluation serves a practical purpose. Instead of claiming that LLM solutions can automatically manage immunosuppressive therapy, the study can show how different resources can be assembled into a reporting workflow.

The most important outcome would be a platform that makes relevant information easier to access, clearly communicates uncertainty, and supports clinicians in remaining the final decision-makers during patient evaluation.

8. Conclusion and Further Discussion

The main advantage of the proposed framework is that it connects previously separate research directions through a structured clinical report while making the current stage of the project explicit. The predictive, optimisation-oriented, and forecasting modules are supported by retrospective technical experiments, whereas the integrated clinical workflow, clinician-facing report evaluation, cognitive-load assessment, and hospital implementation remain prospective validation tasks. Optimisation and forecasting modules remain useful only if their outputs can be interpreted, verified, and acted upon. The reporting layer provides a common interface for reviewing model output, patient history, and source data references together.

Empirically, the predictive modules were validated retrospectively on the transplant cohort and performed above their baselines, while one module, the unsupervised infection clustering, is reported as a negative result that marks its current limit. The usefulness of the generated report and the clinical value of the integrated platform were deliberately not claimed from these retrospective results; they are left to the clinician reader study and the prospective pilot described in Section 7.

One of the main limitations is institutions' cooperation and access to anonymised data. MITOF currently uses retrospective hospital records and does not integrate real-time monitoring data. This remains a future clinical-integration task, requiring update intervals, warning thresholds, audit trails, and clinician confirmation. Report personalisation is also limited to the described doctor, patient, and management views; future work can add configurable templates, specialty-specific thresholds, and patient-specific explanation levels while preserving the separation between generated text and extracted predicted data.

From an implementation and statistical perspective, the reporting approach is suitable for unifying multiple machine-learning modules in a hospital-oriented workflow. The evidence should be interpreted module by module: retrospective metrics support components with recorded targets, while platform-level usefulness still requires the clinician-reader study and prospective pilot. The present article does not claim completed clinical deployment or proven reduction of workload and costs. Instead, it reports retrospective technical evidence and proposes a structured prospective evaluation pathway for clinical utility, safety, cognitive effort, and resource-management impact.

In this research, we explored several directions that converge towards immunosuppressive care, reflecting the growing role of technology in modern hospital care. These range from medication forecasting and dosage optimisation to the prevention of hospital-acquired infections. Finally, natural language processing supports automated clinical documentation and structured report generation for modern healthcare.

In conclusion, this paper represents research towards building intelligent healthcare systems that can generate a report based on multiple modules, which produces predictions and optimisations for the medical professional to review for a final decision. By combining the developed optimisation and forecasting modules oriented through immunosuppressive therapy, we aim to contribute to a new generation of medical tools that can assist clinicians and hospitals in improving the treatment speed and efficiency without replacing clinical decision-making, but supporting it.

This study therefore presents a platform-focused framework for integrated immunosuppressive therapy support. The demonstrated contribution lies in connecting retrospective AI modules with a structured reporting layer, while the proposed contribution lies in the future validation model required before routine clinical adoption. This distinction should guide both interpretation of the current results and the design of subsequent pilot studies.

Therefore, the platform's contribution should also be read as a clinical AI design contribution: it combines clinical reviewability, role-specific abstraction and value-level provenance in a single reporting workflow for immunosuppressive therapy.

Speech recognition is retained as a future-work direction rather than as part of the current platform implementation. Incomplete or inconsistently structured hospital records could eventually be complemented through a controlled voice-input module that allows clinicians to dictate missing observations and then review the generated text before it is stored.

Future development should first evaluate Romanian medical speech-recognition accuracy background-noise resistance, clinician-specific adaptation, patient acceptability, and the mapping of dictated content to structured report fields. Because these requirements are not yet validated in the present study, speech recognition is not treated as a demonstrated component of the current framework.

References

Arvinte, R., & Trandabăț, D. (2023). Using neural networks to personalize immunosuppressive dosing in renal transplanted patients. *Procedia Computer Science, 225*, 3967–3976. https://doi.org/10.1016/j.procs.2023.10.392

Arvinte, R., & Trandabăț, D. (2025). Forecasting tacrolimus levels: A time series analysis using deep learning and statistical methods. *Procedia Computer Science, 270*, 2346–2355. https://doi.org/10.1016/j.procs.2025.09.356

Bose, S., Rajendran, R. K., Debnath, B., Karydis, K., Roy-Chowdhury, A. K., & Chakradhar, S. (2025). *Visual alignment of medical vision-language models for grounded radiology report generation*. arXiv. https://doi.org/10.48550/arXiv.2512.16201

Choshi, H., Miyoshi, K., Tanioka, M., Arai, H., Tanaka, S., Shien, K., Suzawa, K., Okazaki, M., Sugimoto, S., & Toyooka, S. (2025). Long short-term memory algorithm for personalized tacrolimus dosing: A simple and effective time series forecasting approach post-lung transplantation. *The Journal of Heart and Lung Transplantation, 44*(3), 351–361. https://doi.org/10.1016/j.healun.2024.10.026

Deperrois, N., Matsuo, H., Ruipérez-Campillo, S., Vandenhirtz, M., Laguna, S., Ryser, A., Fujimoto, K., Nishio, M., Sutter, T. M., Vogt, J. E., Kluckert, J., Frauenfelder, T., Blüthgen, C., Nooralahzadeh, F., & Krauthammer, M. (2025). RadVLM: A multitask conversational vision-language model for radiology. *arXiv*. https://doi.org/10.48550/arXiv.2502.03333

European Parliament and Council of the European Union. (2016). Regulation (EU) 2016/679, General Data Protection Regulation.

Friedman, C., Alderson, P. O., Austin, J. H. M., Cimino, J. J., & Johnson, S. B. (1994). A general natural-language text processor for clinical radiology. *Journal of the American Medical Informatics Association, 1*(2), 161–174. https://doi.org/10.1136/jamia.1994.95236146

Ganzinger, M., Kunz, N., Fuchs, P., Lyu, C. K., Loos, M., Dugas, M., & Pausch, T. M. (2025). Automated generation of discharge summaries: Leveraging large language models with clinical data. *Scientific Reports, 15*(1), Article 16466.

Gu, D., Gao, Y., Zhou, Y., Zhou, M., & Metaxas, D. (2025). RadAlign: Advancing radiology report generation with vision-language concept alignment. *arXiv*. https://doi.org/10.48550/arXiv.2501.07525

Halloran, P. F. (2004). Immunosuppressive drugs for kidney transplantation. *The New England Journal of Medicine, 351*(26), 2715–2729. https://doi.org/10.1056/NEJMra033540

Johnson, M. R., Naik, H., Chan, W. S., Greiner, J., Michaleski, M., Liu, D., Silvestre, B., & McCarthy, I. P. (2023). Forecasting ward-level bed requirements to aid pandemic resource planning: Lessons learned and future directions. *Health Care Management Science, 26*(3), 477–500. https://doi.org/10.1007/s10729-023-09639-2

Kahan, B. D. (1989). Cyclosporine. *New England Journal of Medicine, 321(25),* 1725-1738.

Koç, E., & Türkoğlu, M. (2022). Forecasting of medical equipment demand and outbreak spreading based on deep long short-term memory network: The COVID-19 pandemic in Turkey. *Signal, Image and Video Processing, 16*(3), 613–621. https://doi.org/10.1007/s11760-020-01847-5

Meyer, G. P. (2021). An alternative probabilistic interpretation of the Huber loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 5261–5269). https://doi.org/10.1109/CVPR46437.2021.00522

Miller, L. W. (2002). Cardiovascular toxicities of immunosuppressive agents. *American Journal of Transplantation, 2*(9), 807–818. https://doi.org/10.1034/j.1600-6143.2002.20902.x

Murray, L. L., Wilson, J. G., Rodrigues, F. F., & Zaric, G. S. (2023). Forecasting ICU census by combining time series and survival models. *Critical Care Explorations, 5*(5), e0912. https://doi.org/10.1097/CCE.0000000000000912

Palm, E., Manikantan, A., Pepin, M. E., Mahal, H., & Belwadi, S. S. (2025). Assessing the quality of AI-generated clinical notes: A validated evaluation of a large language model scribe. *arXiv*. https://doi.org/10.48550/arXiv.2505.17047

Peters, D. H., Fitton, A., Plosker, G. L., & Faulds, D. (1993). Tacrolimus: A review of its pharmacology and therapeutic potential in hepatic and renal transplantation. *Drugs, 46*(4), 746–794. https://doi.org/10.2165/00003495-199346040-00009

Riaz, M., Hussain Sial, M., Sharif, S., & Mehmood, Q. (2023). Epidemiological forecasting models using ARIMA, SARIMA, and Holt–Winter multiplicative approach for Pakistan. *Journal of Environmental and Public Health, 2023*, Article 8907610. https://doi.org/10.1155/2023/8907610

Rodrigues, T., & Teixeira Lopes, C. (2025). Large language model-based generation of discharge summaries. *arXiv*. https://doi.org/10.48550/arXiv.2512.06812

Salmè, M., Sicilia, R., Soda, P., & Guarrasi, V. (2025). Evaluating vision language model adaptations for radiology report generation in low-resource languages. *arXiv*. https://doi.org/10.48550/arXiv.2505.01096

Starzl, T. E., & Zinkernagel, R. M. (2001). Transplantation tolerance from a historical perspective. *Nature Reviews Immunology, 1*(3), 233–239. https://doi.org/10.1038/35105088

Tanno, R., Barrett, D. G. T., Sellergren, A., Ghaisas, S., Dathathri, S., See, A., Welbl, J., Lau, C., Tu, T., Azizi, S., Singhal, K., Schaekermann, M., May, R., Lee, R., Man, S. W., Mahdavi, S., Ahmed, Z., Matias, Y., Barral, J., ... Ktena, I. (2025). Collaboration between clinicians and vision-language models in radiology report generation. *Nature Medicine, 31*(2), 599–608. https://doi.org/10.1038/s41591-024-03302-1

Tedesco, D., & Haragsim, L. (2012). Cyclosporine: A review. *Journal of Transplantation, 2012*, Article 230386. https://doi.org/10.1155/2012/230386

Yoon, S. B., Lee, J. M., Jung, C. W., Suh, K. S., Lee, K. W., Yi, N. J., Hong, S. K., Choi, Y. R., Hong, S. Y., & Lee, H. C. (2024). Machine-learning model to predict the tacrolimus concentration and suggest optimal dose in liver transplantation recipients: A multicenter retrospective cohort study. *Scientific Reports, 14*(1), Article 19996. https://doi.org/10.1038/s41598-024-71032-y

Yu, S., Kumamaru, K. K., George, E., Dunne, R. M., Bedayat, A., Neykov, M., Hunsaker, A. R., Dill, K. E., Cai, T., & Rybicki, F. J. (2014). Classification of CT pulmonary angiography reports by presence, chronicity, and location of pulmonary embolism with natural language processing. *Journal of Biomedical Informatics, 52*, 386–393. https://doi.org/10.1016/j.jbi.2014.08.001