

On the Harmonisation of Time Series Data for the Optimisation of Machine Learning Using the Example of Rejection Prediction After Kidney Transplantation

Darian Liehr, Matthias Katzensteiner, Oliver J. Bott

Suggested citation:

Liehr, Darian, Matthias Katzensteiner, and Oliver J. Bott. 2025. "On the Harmonisation of Time Series Data for the Optimisation of Machine Learning Using the Example of Rejection Prediction After Kidney Transplantation." In *Intelligent Health Systems – From Technology to Data and Knowledge (Studies in Health Technology and Informatics ; 327)*, 68–72. IOS Press. <https://doi.org/10.25968/opus-3628>.

Abstract

A significant risk following a kidney transplantation is graft loss. The Screen Reject Project has developed a Clinical Data Warehouse (CDWH) as a foundation for a clinical decision support system designed to improve the diagnosis of graft rejections. The CDWH integrates patient data and event records of $n = 141$ kidney transplant patients. These data are not directly comparable within the cohort as they consist of irregular time series, particularly of laboratory values. Therefore, a pre-processing routine was developed which divides a relative time window before the last biopsy (the relevant end event of the reference period for subsequent machine learning procedures) into equal time intervals for each patient. For each of these intervals a representative value is calculated from the contained laboratory values. These representative values are used to train models for predicting kidney rejection. The comparison with an existing study from the project, in which a classification model was developed without considering the temporal dependencies, shows an improved sensitivity and specificity in predicting kidney rejection for the harmonised data using the same random forest model.

Terms of use

CC BY-ND 4.0

On the Harmonisation of Time Series Data for the Optimisation of Machine Learning Using the Example of Rejection Prediction After Kidney Transplantation

Darian LIEHR^{a,1}, Matthias KATZENSTEINER^a, and Oliver J. BOTT^a

^aData|H Institute for Applied Data Science, Faculty III – Media, Information and Design, University of Applied Sciences and Arts Hanover, Germany

ORCID ID: Darian Liehr <https://orcid.org/0009-0003-0350-436X>,

Matthias Katzensteiner <https://orcid.org/0000-0002-0665-6110>,

Oliver J. Bott <https://orcid.org/0000-0002-8823-5247>

Abstract. A significant risk following a kidney transplantation is graft loss. The Screen Reject Project has developed a Clinical Data Warehouse (CDWH) as a foundation for a clinical decision support system designed to improve the diagnosis of graft rejections. The CDWH integrates patient data and event records of $n = 141$ kidney transplant patients. These data are not directly comparable within the cohort as they consist of irregular time series, particularly of laboratory values. Therefore, a pre-processing routine was developed which divides a relative time window before the last biopsy (the relevant end event of the reference period for subsequent machine learning procedures) into equal time intervals for each patient. For each of these intervals a representative value is calculated from the contained laboratory values. These representative values are used to train models for predicting kidney rejection. The comparison with an existing study from the project, in which a classification model was developed without considering the temporal dependencies, shows an improved sensitivity and specificity in predicting kidney rejection for the harmonised data using the same random forest model.

Keywords. time series data, secondary use, machine learning, data harmonisation, data management, kidney transplant, rejection diagnostics

1. Introduction

Kidney transplantation is the most common of all organ transplants [1]. Due to demographic change with an ageing population, an increase in diseases such as end-stage renal disease (ESRD) and therefore also an increase in the number of kidney transplants is to be expected [2]. Graft loss is a significant risk following transplantation, often leading to a shortened lifespan of the transplant. For this reason, early diagnosis is crucial. A biopsy remains the gold standard for diagnosis and is still superior to other methods, such as clinical diagnostics [3].

¹ Corresponding Author: Darian Liehr, University of Applied Sciences and Arts Hanover, Expo Plaza 12, 30539 Hanover, Germany; E-mail: darian.liehr@hs-hannover.de.

As part of the Screen Reject Project, Hanover University of Applied Sciences and Arts has developed a Clinical Data Warehouse (CDWH) to serve as the foundation for a clinical decision support system designed to improve the diagnosis of graft rejections [4]. The CDWH integrates patient data and especially event recordings, including biopsies, transplantations, or rejection episodes of kidney transplant patients. The data sources include clinical primary systems as well as the Enterprise Clinical Research Data Warehouse (ECRDW) [5] of the Hannover Medical School (MHH), which serves as a pre-aggregated data environment. The research data are collected during routine clinical practice and are not directly comparable within the cohort, as they are gathered at varying intervals and frequencies depending on individual clinical indication. This applies in particular to numerical observation data such as laboratory data.

This work aims to ensure temporal comparability by developing a data pre-processing routine in the form of a time series harmonisation as a proof of concept to improve machine learning based predictions of the presence of graft rejection processes.

2. Methods

2.1. Time Series Harmonisation

The mentioned CDWH contains about 900,000 records of $n = 141$ patients. The data for the pre-processing routine is available in the form of CSV files. The implementation is done in *Python* using the libraries *pandas*, *NumPy*, *scikit-learn*, and *SciPy*.

The starting point for harmonising a time series, such as a series of laboratory values, is the patient-specific definition of an observation period or a time window with a selected start and end event or start and end time. The start and end point can either be a patient specific date or an event, such as a transplantation, biopsy, or rejection. Additionally, a relative time window, such as six months before the last biopsy (the end event), can be defined. This time window is then divided into equally sized intervals, whereby the number of intervals or the length of the intervals can also be freely selected. A representative value is calculated for each artificial interval from the laboratory values it contains. In the simplest case, this is the mean value. In addition, a linear regression is performed to obtain a value for the trend of the course within an interval using the slope. A polynomial regression of a selectable degree is also implemented. The mean value of the function on the respective interval is calculated and used as the representative value for that interval. All calculated representative values are used as input data for a model to predict kidney rejection. Figure 1 shows an outline of the time series harmonisation methodology.

2.2. Evaluation of the Time Series Harmonisation

The time series harmonisation was evaluated by comparison with an existing study from the project, in which a classification model for predicting kidney rejection was developed without considering the temporal dependencies of the laboratory values [6]. Instead, only the last laboratory value before a biopsy was used for each patient. The same random forest model as the one developed there was trained in the evaluation with a data set with harmonised time series, i.e. the representative values of the artificial intervals introduced above. The accuracy, recall, specificity, and precision are compared to explore which data set provides a better prediction in relation to the target variable *rejection*.

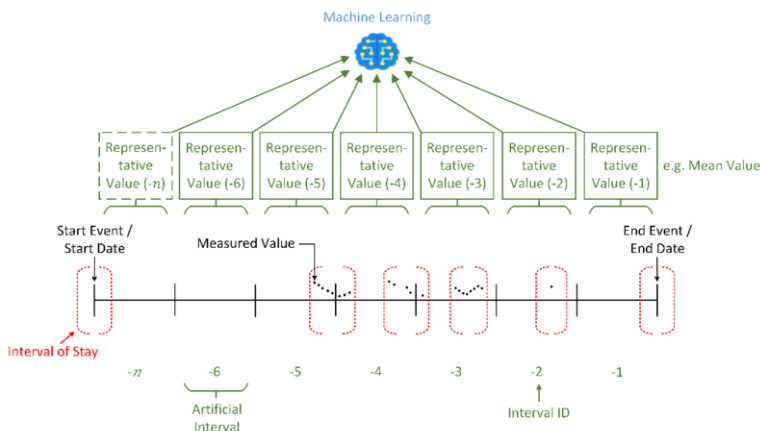


Figure 1. Methodology of the time series harmonisation.

3. Results

The time series harmonisation was implemented for different laboratory parameters, which are usually used in the context of kidney transplantation: creatinine, urea, potassium, sodium, albumin, and the estimated glomerular filtration rate. Freely definable parameters of the time series harmonisation are the start and end point of the considered time window, as well as the number or the length of the artificial intervals, the degree of the polynomial regression, and the imputation method. Mean imputation, k -nearest neighbours imputation and Multivariate Imputation by Chained Equations are available for use and application in data processing. The harmonisation can be applied to one or more patients, and one or more laboratory parameters, providing maximum flexibility in its application. Table 1 shows an example of the results of a time series harmonisation for a specific patient and the laboratory parameter serum creatinine. The time window considered is twelve months before the last biopsy performed. The time window was divided into four intervals. The number of values per interval is displayed in addition to the representative values, i.e. mean, slope of the linear regression, and mean of the third-degree polynomial function.

Table 1. Example result for the interval-wise calculation of representative values of a patient's serum creatinine value. The time window covers 12 months before the last biopsy.

Interval ID	Start Date	End Date	No. of Values	Mean [$\mu\text{mol/l}$]	Slope (Linear Regression)	Mean of the Third-Degree Polynomial Function
-1	05/10/2016	04/01/2017	26	337.65	-9.38	336.92
-2	05/07/2016	04/10/2016	5	398.40	61.00	409.78
-3	04/04/2016	04/07/2016	80	171.29	-0.67	170.98
-4	03/01/2016	04/04/2016	26	203.86	-0.69	204.30

For the evaluation, a harmonised data set was created with data from all available 141 patients. A relative time window was selected for each patient, starting six months before and ending at the time of the last biopsy. The time window was divided into two intervals. Missing values were filled by mean imputation. The representative values for

each implemented laboratory parameter were included in the data set. In addition to laboratory values, the data set contains information on diagnoses such as diabetes and polycystic kidney disease, the patient's age, number of previous rejections, and information on HLA findings. Instead of the harmonised laboratory values, the comparison data set only contains the last serum creatinine value. Both data sets were used to train a random forest model, which was developed in a previous project, when the comparison data set was created [6].

The harmonised data set reaches a recall of 95.8 % and a specificity of 72.3 %. The comparison data set has a recall of 95.2 % and a specificity of 64.3 % (see table 2). Accuracy and precision are also higher for the harmonised data (88.6 % vs. 82.9 % and 88.4 % vs. 80.0 %).

Table 2. Results for the prediction of a graft rejection with a Random Forest. Comparison of the performances of a data set with a harmonised laboratory value course and a comparison data set, which does not take the aspect of time into account.

Data Set	Accuracy	Recall/Sensitivity	Specificity	Precision
Data set with harmonised time series	88.6 %	95.8 %	72.3 %	88.4 %
Comparison data set	82.9 %	95.2 %	64.3 %	80.0 %

4. Discussion

This work shows that harmonisation of laboratory value series has the potential to improve recall and specificity in the prediction of kidney rejection. A limiting factor is the data quantity and quality. The number of patients included in this study is low at $n = 141$. This could lead to an overfitting of the used random forest model. In addition, there are many missing values for some laboratory parameters and patients, which were filled by mean imputation. Also, the data are unbalanced, as $n = 89$ patients had a rejection reaction and $n = 52$ did not. Overall, the quality of the data used for the evaluation limits the generalizability of the method. Therefore, the developed methodology should be tested with additional, larger and preferably more balanced samples.

As the objective of this work was not to develop specific prediction models, it should be noticed that no feature selection was performed for the harmonised data set, so that all implemented laboratory parameters and representative values were included. Rather, the method for harmonising the time series is to be evaluated as a proof of concept by analysing its suitability for improving machine learning methods using a classification model and a data set that does not take the temporal aspect into account. As the model was also not adapted to the data, further work should optimise the pre-processing of the data and select and train models specifically suitable for time series data with the aim of further improving the prediction results in this specific clinical scenario. This primarily includes exploring and testing additional algorithms besides the random forest algorithm, which was only used in the present study because it performed best in the comparative work in which the comparative data set was developed [6].

Besides further tests with larger samples of better data quality and additional machine learning algorithms, future work should also include a comprehensive literature review on state-of-the-art time series modelling and data pre-processing practices, such as the use of gated recurrent neural networks [7]. This could provide the basis for a structured comparison with the method developed here.

5. Conclusions

This first proof of concept shows promising results regarding an improved prediction of kidney transplant rejection by using harmonised time series generated from irregular routine data instead of just one data point. This can potentially improve rejection diagnostics prior to the decision to perform a biopsy.

The time series harmonisation approach presented here is not limited to renal rejection diagnostics. As the use of routine clinical data is becoming increasingly important for research the approach could also be adapted and tested in other medical fields in which time series and the temporal progression of measured values such as laboratory values play a role. The Health Data Utilisation Act (GDNG) [8] passed by the German Federal Ministry of Health in March 2024 promotes the broad use of routine clinical data for research purposes. In order to be able to use this real world data effectively, appropriate data pre-processing is essential. The proposed time series harmonisation is one step towards making routine data more comparable with data from clinical studies.

However, beyond the pre-processing of time series harmonisation, other critical aspects of the datasets – such as data quality, completeness and reliability – must also be considered to ensure robust and reliable research results.

References

- [1] Deutsche Stiftung Organtransplantation. Jahresbericht Organspende und Transplantation in Deutschland 2022 [Internet]. Frankfurt/Main, Germany; 2023 May [cited 2024 Oct 1]. Available from: <https://www.dso.de/SiteCollectionDocuments/DSO-Jahresbericht%202022.pdf>
- [2] Peters E, Pritzkuleit R, Beske F, Katalinic A. Demografischer Wandel und Krankheitshäufigkeiten: Eine Projektion bis 2050. Bundesgesundheitsbl. [Internet]. 2010 May [cited 2024 Oct 1];53(5):417-26, doi: 10.1007/s00103-010-1050-y.
- [3] Al-Awwa IA, Hariharan S, First MR. Importance of allograft biopsy in renal transplant recipients: Correlation between clinical and histological diagnosis. Am J Kidney Dis [Internet]. 1998 Jun [cited 2024 Oct 1];31(6):S15-8, doi: 10.1053/ajkd.1998.v31.pm9631859.
- [4] Katzensteiner M, Schewe N, Zubke M, Nörtemann K, Frenzel B, Bott OJ. Screen Reject Data Warehouse – Klinisches Data Warehouse zur Abstoßungsdiagnostik nach NTx: Ein Werkstattbericht. In: Wartena C, Bruns R, editors. Smart Data Analytics: Schriften des Forschungsclusters Smart Data Analytics 2020 [Internet]. Hannover, Germany: Hochschule Hannover; 2020 [cited 2024 Oct 1]. p. 102-24, doi: 10.25968/opus-1789.
- [5] Gerbel S, Laser H, Schönfeld N, Rassmann T. The hannover medical school enterprise clinical research data warehouse: 5 years of experience. In: Auer S, Vidal ME, editors. Data Integration in the Life Sciences: 13th International Conference, DILS 2018, Hannover, Germany, November 20-21, 2018, Proceedings. In: Lecture Notes in Bioinformatics, vol. 11371, Cham: Springer International Publishing, 2019, S. 182–194, doi: 10.1007/978-3-030-06016-9.
- [6] Avci F. Die Anwendung von Machine Learning Verfahren für die Vorhersage einer Nierentransplantatabstoßungsreaktion. Bachelor's thesis, Hochschule Hannover - University of Applied Sciences and Arts, Hannover, Germany, 2021.
- [7] Weerakody PB, Wong KW, Wang G, Ela W. A review of irregular time series data handling with gated recurrent neural networks. Neurocomputing [Internet]. 2021 Mar [cited 2025 Jan 3];441:161-78, doi: 10.1016/j.neucom.2021.02.046.
- [8] Federal Ministry of Health (Germany). Gesetz zur verbesserten Nutzung von Gesundheitsdaten. BGBl. 2024 I Nr. 102 [Internet]. Bonn, Germany; 2024 Mar 15 [cited 2024 Oct 1]. Available from: <https://www.recht.bund.de/bgbl/1/2024/102/VO.html>