



LJMU Research Online

Olier, I, Ortega-Martorell, S, Margereson, G, Bellfield, RAA, Welters, ID and Lip, GYH

The integrated multiple event representation framework (IMERF): a case study on critically-ill patients

<https://researchonline.ljmu.ac.uk/id/eprint/27373/>

Article

Citation (please note it is advisable to refer to the publisher's version if you intend to cite from this work)

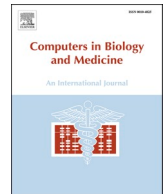
Olier, I ORCID logo**ORCID: <https://orcid.org/0000-0002-5679-7501>, Ortega-Martorell, S ORCID logo****ORCID: <https://orcid.org/0000-0001-9927-3209>, Margereson, G ORCID logo****ORCID: <https://orcid.org/0009-0004-5117-4032>, Bellfield. RAA ORCID logo****ORCID: <https://orcid.org/0000-0002-9945-914X>.**

LJMU has developed **LJMU Research Online** for users to access the research output of the University more effectively. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LJMU Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain.

The version presented here may differ from the published version or from the version of the record. Please see the repository URL above for details on accessing the published version and note that access may require a subscription.

For more information please contact researchonline@ljmu.ac.uk

<http://researchonline.ljmu.ac.uk/>



The integrated multiple event representation framework (IMERF): a case study on critically-ill patients

Ivan Olier^{a,b,*}, Sandra Ortega-Martorell^{a,b}, George Margereson^{a,b}, Ryan A. Bellfield^{a,b}, Ingeborg D. Welters^b, Gregory Y.H. Lip^{b,c,d}

^a Data Science Research Centre, Liverpool John Moores University, Liverpool, UK

^b Liverpool Centre for Cardiovascular Science at University of Liverpool, Liverpool John Moores University and Liverpool Heart & Chest Hospital, Liverpool, UK

^c Danish Centre for Health Services Research, Department of Clinical Medicine, Aalborg University, Denmark

^d Department of Cardiology, Lipidology and Internal Medicine, Medical University of Białystok, Białystok, Poland

ABSTRACT

This study introduces the Integrated Multiple Event Representation Framework (IMERF), a novel methodological approach for developing risk prediction models for multiple clinical events. Using a two-stage process involving multi-task learning and dimensionality reduction, IMERF creates a visual representation of predicted event risks and identifies clusters based on overlapping risks. The proposed framework is showcased through a case study modelling nine adverse events in critically ill patients admitted to intensive care units (ICUs). Stage 1 was implemented using convolutional neural networks, which displayed superior performance to logistic regression and random forest algorithms. The generative topographic mapping (GTM) algorithm was implemented in stage 2 for data visualisation and clustering. It revealed clear patterns of adverse event risk clusters. GTM in combination with class activation maps was also employed to trace input factors influencing cluster membership, highlighting distinct risk profiles among patients. Macro-clusters representing distinctive combinations of adverse event risk levels were also identified by performing a hierarchical clustering on the GTM results. In conclusion, IMERF could represent a significant advancement in multiple event risk modelling by enabling simultaneous prediction and characterisation of overlapping events and providing an interpretable framework for understanding their complex patterns. Its application in ICUs underscores its potential for broader clinical use, including modelling clusters of conditions or multiple instances of events.

1. Introduction

A common question in medical research is how to predict the risk of patient events within a particular study. Events, which can be either beneficial or adverse, refers to discrete, observable occurrences or outcomes that are relevant to the study being conducted. For instance, if a patient undergoes a medical procedure, it is crucial to assess their risk of experiencing complications, such as bleeding, infection or even death, which are considered adverse events. One approach to addressing such questions is by implementing a risk prediction model that utilises statistical or machine learning (ML) algorithms on data collected from observational studies [1]. Having a model that can anticipate the risk of a given event is fundamental to serving as a medical decision support tool [2].

Most algorithms model the risk of an event as a single-output function of a set of input variables or factors. In cases where a patient is at risk of multiple events, it is generally assumed that a model will be built for each event separately [3]. However, events may not necessarily be independent. For example, individuals at risk of developing multiple

co-morbidities tend to have a less favourable prognosis than those at risk of a single condition. Moreover, modelling individual co-morbidities independently fails to identify specific factors associated with an increased risk of several co-morbidities simultaneously. A similar challenge arises in intensive care, where patients are at a higher risk of developing multiple adverse events simultaneously, for instance, a patient requiring mechanical ventilation and developing atrial fibrillation (AF) within a short time frame [4]. Detecting and characterising patients at risk of multiple adverse events introduces additional complexity to modelling, which single-output risk prediction models are not equipped to handle efficiently.

The use of multi-task learning (MTL) in medical and healthcare research is not new. Early works by Wiens et al. [5], Wang et al. [6] and Zhou et al. [7], to mention a few, paved the way for its application. The advent of deep learning (DL) and its widespread adoption in recent years have significantly increased the use of MTL in risk prediction for clinical outcomes [8–10]. MTL has also expanded to address closely related topics, such as time-to-event modelling [11], trajectory analysis [12], and risk profiling [13]. Most publications focus on using MTL to enhance

* Corresponding author. Data Science Research Centre Liverpool John Moores University James Parsons Building, 3 Byrom Street, Liverpool, L3 3AF I, UK.
E-mail address: I.Olier@ljmu.ac.uk (I. Olier).

model performance, yet limited attention has been given to integrating and interpreting the model predictions. However, to the best of our knowledge, no existing framework specifically addresses risk prediction modelling for multiple events that is also capable of identifying clusters of risks and uncovering factors associated with specific clusters.

In this research paper, we propose the *Integrated Multiple Event Representation Framework* (IMERF): a novel methodological approach for developing risk prediction models for multiple events. IMERF is a novel two-stage methodological approach involving: 1) a multi-task learning (MTL) model to predict the risks of multiple events simultaneously, followed by 2) a dimensionality reduction (DR) model to project the predicted risks onto a visualisation map and to identify event clusters. IMERF can be complemented with interpretable machine learning techniques to assist in identifying and characterising clusters and to explain possible associations between factors and event clusters.

IMERF is capable of realising and characterising potential associations not only between input factors and multiple events simultaneously—associations that cannot be identified when modelled separately. Moreover, modelling the risks of multiple events together has the potential to enhance prediction quality, as learning can be shared across the various tasks. Traditionally, multiple events are modelled by creating a composite variable, often defined by aggregating predicted risks using hand-crafted weights. These composite events are then used as the output variable in a statistical or ML risk prediction model [14].

Although the proposed approach can be applied to various clinical scenarios, this research focuses on modelling multiple adverse events in patients admitted to intensive care units (ICUs) as a case study. It serves as a testbed for evaluating the added value of our proposed framework. ICU patients are critically ill, suffering from conditions that threaten their lives. Data from ICU stays are highly heterogeneous—patients may be admitted for a wide range of reasons, such as cardiac surgery, sepsis, or emergency department referral, among others. Moreover, ICU patients are at higher risk of experiencing multiple negative outcomes due to adverse events, which often leads to poor performance in traditional risk prediction models [15]. This further strengthens the case for using MTL to model ICU data more effectively.

2. Methods

2.1. Overall methodological approach of IMERF

We propose IMERF as a methodological approach for modelling multiple event risks, identifying event clusters, and determining factors associated with their risk increase. As mentioned previously, IMERF is a two-stage approach combining MTL and DR for visualisation and clustering. The MTL algorithm implements the risk prediction model for multiple events. In the second stage of the approach, a new dataset is generated using the predicted event probabilities, where the number of

columns corresponds to the number of events. A data visualisation model based on DR techniques is then applied to map this dataset onto a two-dimensional latent space (map). The objective is to identify regions within the map where specific events occur or where two or more events are likely to occur simultaneously in a group of patients. In principle, stages 1 and 2 of IMERF can be implemented using most MTL algorithms and data visualisation and clustering techniques, respectively. Fig. 1 illustrates the overall methodological approach underpinning IMERF.

2.2. Single-task vs multi-task learning modelling

Learning a task is the process of fitting a statistical or ML model. In this context, a task encapsulates a dataset along with details about input variables and a target (or outcome). Single-task learning (STL) involves learning one task at a time, and most ML algorithms are traditionally designed for STL. In contrast, MTL models several tasks, allowing them to share information during the learning process. By sharing information and parameters across tasks, MTL is generally more efficient than modelling tasks separately (i.e. STL). In other words, MTL tasks can achieve equivalent performance to those implemented individually via STL but with significantly fewer parameters. Consequently, an MTL multi-outcome model would require less data and be less prone to overfitting compared to equivalent multiple STL models [16].

Popular ML algorithms like Random Forest, XGBoost, and Support Vector Machines (SVM) are typically designed and used for STL problems, although MTL variants have also been proposed [17]. In recent years, MTL models have predominantly been developed using deep learning algorithms [16,18].

2.3. Visualisation modelling for the predicted events

DR algorithms can be used for data visualisation by mapping high-dimensional data onto a 2- or 3-dimensional space. In our proposed framework, the purpose of using a DR algorithm is to map the MTL risk predictions onto a 2-dimensional space, to identify regions of interest associated with different risk levels for events. Commonly used DR algorithms for visualisation include Principal Component Analysis (PCA), t-distributed Stochastic Neighbour Embedding (t-SNE [19]), and Uniform Manifold Approximation and Projection (UMAP [20]) are commonly used for data visualisation [21].

Here, we propose employing the Generative Topographic Mapping (GTM [22]) algorithm. Although other DR algorithms could, in principle, be utilised, GTM inherently integrates clustering and visualisation into a single model. Unlike traditional clustering algorithms such as k-means and hierarchical clustering, GTM is grounded in a probabilistic framework allowing for soft assignments of data observations to clusters with associated probabilities. Furthermore, in contrast to recent techniques such as t-SNE and UMAP, the visualisation maps produced by GTM are highly interpretable and globally consistent.

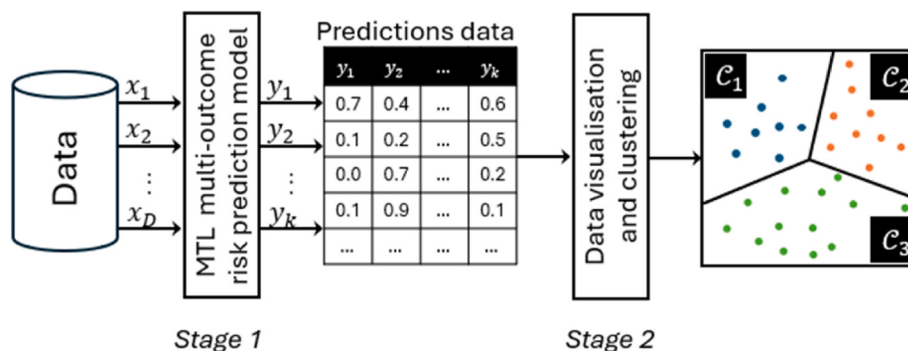


Fig. 1. Overall methodological approach of IMERF, comprising an MTL risk prediction model (Stage 1), followed by a DR model for data visualisation and clustering (Stage 2).

2.4. Case study: modelling multiple adverse event risks in critically-ill patients

2.4.1. Data source

In this study, our proposed framework was utilised to model the risk of multiple adverse events in patients admitted to the ICU. A total of nine commonly occurring adverse events in the ICU were considered: AF, acute kidney injury (AKI), acute respiratory distress syndrome (ARDS), the need for vasoactive medications (VAM), ICU length of stay greater than 3 days (LOS-3d) and 7 days (LOS-7d), in-hospital death, as well as 30-day (30d) and 1-year (1yr) mortality.

AF refers to an irregular and often rapid heart rhythm that can lead to blood clots, stroke, or other complications. AKI is a sudden decline in kidney function, resulting in the accumulation of waste products and fluid imbalances. ARDS is a severe lung condition characterised by inflammation-mediated disruptions in alveolar membrane permeability, leading to impaired oxygen exchange. VAMs are required to stabilise blood pressure and maintain organ perfusion in critically ill patients experiencing shock or haemodynamic instability. ICU length of stay is particularly challenging to model due to its highly skewed distribution. To address this, we defined ICU length of stay as two binary events with cutoffs at 3 and 7 days. We applied our model to a cohort extracted from the Medical Information Mart for Intensive Care IV (MIMIC-IV [23]), a freely available database of de-identified electronic health records from patients admitted to the Beth Israel Deaconess Medical Centre in Boston, Massachusetts. We used version 2.2, released in January 2023, which includes 299,712 patients, 431,231 hospital admissions, and 73,181 ICU stays.

2.4.2. Formation of the data cohort

For this study, we extracted sequences of vital signs, laboratory test results, and other clinical characteristics of patients admitted to the hospital's ICU. The data used for modelling was formatted as a three-dimensional array, with dimensions representing patient admissions, time points (hours), and variables. Patients under 18 years of age, those with an ICU stay of less than 24 h, and patients with multiple admissions (only their first admission was included) were excluded. Consequently, for the purpose of this analysis, each patient was considered to have only one hospital and ICU admission.

From the electronic health records and bedside monitoring data, a total of 28 features were selected. Relevant static data, such as age and sex, were included alongside dynamic vital signs and laboratory features commonly recorded in ICU settings. Dynamic features were organised into 1 h non-overlapping time series bins during data extraction from the MIMIC-IV database. This binning was performed to account for varying sampling frequencies in the available data and to balance the trade-off between missing data points and bin size [24]. Missing time points were backfilled, while variables with a level of missingness exceeding 65 % were excluded from the analysis. Values deemed implausible were assumed to be outliers, likely due to measurement and/or recording errors, and were replaced with plausible extreme values. Additionally, a second version of the dataset was created by converting all time-varying variables into tabular representations by extracting their means and standard deviations. This aggregated version was used to develop baseline models, enabling performance comparisons with our proposed pipeline.

A pairwise correlation followed by a variance inflation factor (VIF [25]) analysis was performed on the aggregated dataset to exclude highly correlated variables. Using a threshold value of 5, variables with VIF values above this threshold were sequentially excluded, and VIF was recalculated repeatedly until all remaining variables had VIF values below the threshold.

For modelling purposes, the dynamic and aggregated data cohorts were randomly split into training (80 %) and test (20 %) subsets. The training subset was used for model development, while the test subset was reserved for independent model evaluation. To ensure a fair

comparison, all models utilised the same data splits. Variables with no available time points for a given admission, and therefore unable to be backfilled, appeared as missing values when aggregated. To address this, a multivariate imputation by chained equations (MICE [26]) model was applied to the training subset using a ridge regressor algorithm as the estimator and was subsequently used to impute missing values in both the training and test subsets. Additionally, the training and test data subsets were standardised using the column means and standard deviations from the training subset.

2.4.3. Formation of the adverse event data

Patient adverse events were recorded in two two-dimensional arrays: one for model development (training subset) and another for model evaluation (test subset), with columns representing each event. The timing of event onset was an additional criterion for determining patient admission inclusion or exclusion. Starting at ICU admission, we defined a 24 h variable extraction (input) window during which patient input data were gathered. This was followed by a 2 h gap window, after which adverse events could be considered for analysis. The gap window was implemented to account for imprecisions in data recording times and to minimise the risk of data leakage (i.e., attempting to predict an already known event).

Admissions where all adverse event onsets occurred within the input and gap windows were excluded from the analysis. We define the outcome window as the period following the gap window. Any adverse event occurring within the outcome window was recorded as “1” in the outcome data arrays, while a “0” indicated that the patient did not experience an adverse event at any time during the study period (i.e. input, gap, and outcome windows). As expected in an ICU setting, many patients had already developed at least one adverse event before the end of the gap window. These patients were excluded only when the relevant event was being modelled. Such cases were recorded as “1” in the outcome data arrays. Fig. 2 provides a graphical representation of this process. Additionally, Table S2 (Supplementary Material) shows how the adverse events were defined using the MIMIC-IV database.

2.4.4. Multiple event risk prediction modelling

The MTL model was implemented using convolutional neural networks (CNNs [27]) following a hard parameter-sharing MTL architecture. In this approach, all tasks share a common model backbone, with task-specific heads added for generating the final outputs. For performance comparison, STL CNN models were also implemented, one for each adverse event. Additionally, baseline STL models were developed using the aggregated data and trained with logistic regression (LR) and random forest (RF) algorithms. All models were optimised via hyperparameter tuning. A randomised search with 5-fold cross-validation was used on the training subset to determine the optimal hyperparameter values for the baseline models, while STL and MTL CNN architectures were optimised using the hyperband search strategy. All models were evaluated using the areas under the receiver operating characteristic (AUC) and the precision-recall (PR-AUC) curves. Performance results on the independent test subset are reported. Details of the hyperparameter sets and values considered in this analysis are provided in Section S3 of the Supplementary Material.

2.4.5. Data visualisation model using generative topographic mapping

A visualisation model was built using the predicted outputs of the MTL model (i.e., event risks) from the training subset, and trained with GTM. The GTM assumes that the observed data is generated through a nonlinear and topology-preserving mapping from a low-dimensional latent space in $\mathcal{M}^L$ onto a manifold embedded in the high-dimensional space, $\mathcal{M}^D$, where the observed data reside. The GTM latent space is constrained to form a uniform discrete grid of M centres. Each of these centres is responsible for generating a spherical Gaussian density function in the D -dimensional data space. Collectively, these centres form M

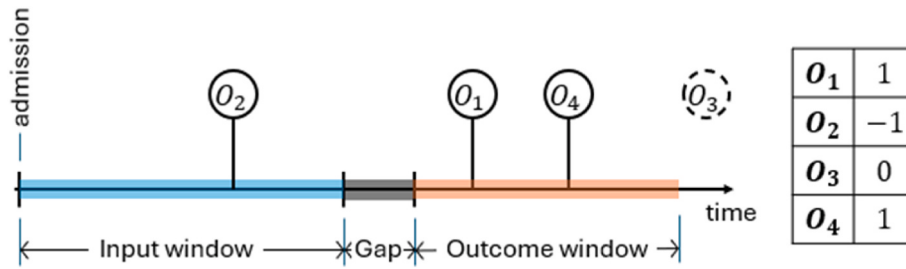


Fig. 2. Example of an ICU stay with four outcome events. Event O_2 is excluded for this admission, as its onset occurred during the input window ($O_2 = -1$). Events O_1 and O_4 are coded as 1, as their onsets occurred during the outcome window, whereas O_3 is coded as 0, as it did not occur during the outcome window.

clusters in the data space, with each cluster corresponding to a Gaussian component in the model. In this sense, the GTM can be understood as a special case of a Gaussian mixture model in which each component in the mixture defines the probability of an observable data point given a latent centre.

The GTM not only can assign data points to clusters but also visualises them in a cluster membership map by projecting the latent centres. The GTM latent space can be used for visualisation purposes when its dimensionality is two ($L = 2$). In this case, the mode probability (i.e., the highest cluster probability) is used to determine the cluster membership of each data point. A GTM membership map is constructed as a two-dimensional histogram, with node size representing the number of data points allocated to each cluster. In addition, cluster centres in the data space, henceforth referred to as reference vectors, act as data prototypes. Reference maps associated with each variable can be generated based on the components of the reference vectors. These reference maps are typically visualised as heatmaps, with variable values represented through colour coding. Reference maps are particularly useful for interpreting the relationship between each variable and the corresponding GTM clusters.

To train the GTM model, the number of GTM clusters was set to 400, arranged in a 20×20 square grid. Although other grid sizes could be considered, it is important to emphasise that GTM is relatively insensitive to the number of clusters selected. Due to GTM's topographic relationships between clusters, altering the number of clusters effectively changes the resolution of the visualisation maps. Additionally, the

number of radial basis functions (RBFs) of the Gaussian mixture was set to 100, ensuring at least four latent points per RBF. The RBF widths and regularisation hyperparameters were optimised using a grid search strategy. Hierarchical clustering was applied to the GTM prototypes (reference vectors) to identify macro-clusters representing combined risks of adverse events. As proposed by Bellfield et al. [28], hierarchical clustering can be applied to group GTM clusters into macro-clusters, further enhancing the interpretability of the GTM data visualisation model.

2.4.6. Model interpretation

GTM reference maps were produced to aid in identifying associations between the predicted event risks and the clustering structure established by the GTM. These maps facilitate the identification of regions where patients are at risk of not only a specific adverse event but also multiple adverse events concurrently. To trace back how variations in the input variables may influence the GTM clusters, we extracted class activation maps (CAMs). Specifically, we used HiResCAM [29] to highlight specific segments of the input time series that are most strongly associated with the model's predictions and clustering patterns.

2.4.7. Overall IMERF implementation with ICU data

Fig. 3 illustrates the IMERF approach for modelling multiple patient adverse event risks in the ICU. Python 3.11 was used to implement all models. Specifically, traditional ML models were implemented using *scikit-learn* 2.5, CNN models were implemented using *TensorFlow* 2.9,

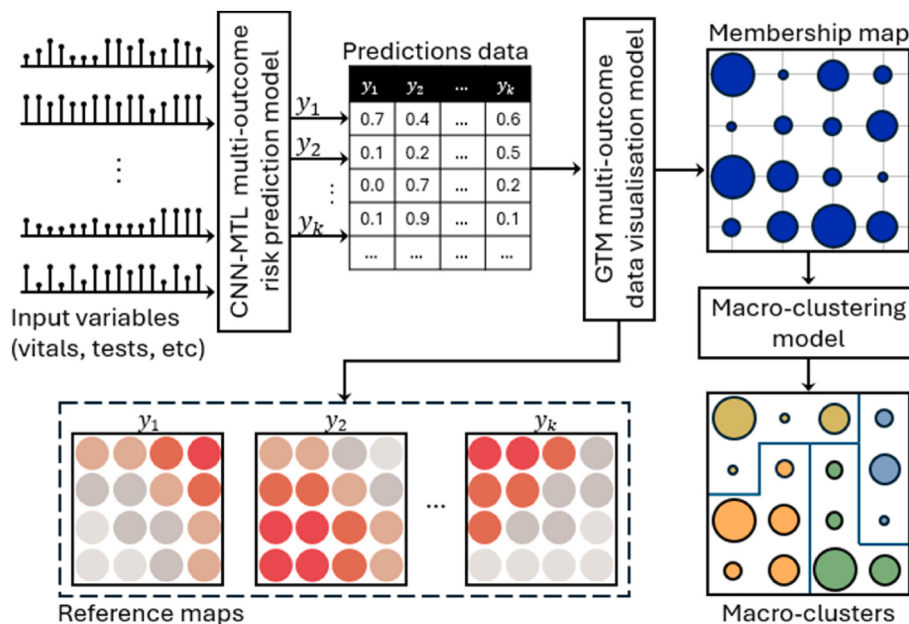


Fig. 3. IMERF implementation using CNN-MTL for stage 1 and GTM for stage 2. GTM reference maps and macro-clusters are extracted to aid in the interpretation of IMERF's results.

and the GTM model was implemented using the *ugtm* package, which was extended to support macro-clusters as described in Ref. [30].

3. Results

3.1. Descriptive analysis

The VIF analysis led to the exclusion of the mean blood pressure (mean BP) variable. The final list of variables included in the analysis is presented in Table 1, which also provides their medians, interquartile ranges (IQRs), and minimum and maximum values. The results of the pairwise correlation and VIF analyses are detailed in Section S1 of the Supplementary Material.

The final cohort used for the analysis comprised 37,253 admissions, predominantly consisting of older male adults (67 years old, 56.92 %). The summary statistics in Table 1 indicate that this is a highly heterogeneous patient cohort, with some variables exhibiting extreme values that deviate significantly from their medians, particularly among laboratory test results.

Table 2 presents the prevalence of the adverse events considered in this study, along with the number of admissions per event included in the analysis. It can be observed that most AF and AKI events occurred after the data collection and gap windows (i.e. after 26 h). In contrast, a

Table 1
Variables used in the analysis, including their descriptive statistics. The total number of patient admissions included in the analysis was N = 37,253.

Variable	Statistics		
Demographics			
	Frequency	Proportion (%)	
Sex: [Female]	16,050	(43.08)	
[Male]	21,203	(56.92)	
	Median	IQR	Extreme values
Age	67.00	(55.00,78.00)	(18.00,89.00)
Weight [kg]	79.23	(66.35,94.20)	(30.10,296.80)
Height [cm]	170.09	(162.78,177.90)	(120.00,231.07)
Vitals			
Capillary refill rate (CRT)	0.00	(0.00,0.00)	(0.00,1.00)
Diastolic blood pressure [mmHg]	61.00	(53.00,71.00)	(30.00,120.00)
Heart rate [bpm]	83.00	(72.00,96.00)	(20.00,200.00)
Respiratory rate (RR) [breaths/min]	18.00	(15.00,22.00)	(4.00,60.00)
Oxygen saturation (SO2) [%]	97.00	(95.00,99.00)	(40.00,100.00)
Systolic blood pressure [mmHg]	116.00	(103.00,131.00)	(50.00,250.00)
Temperature [Celsius]	36.83	(36.56,37.17)	(26.50,42.30)
Glasgow Coma Scale (GCS) (total)	14.00	(10.00,15.00)	(3.00,15.00)
Oxygen therapy settings			
Fraction of inspired oxygen (FiO2)	50.00	(40.00,60.00)	(21.00,100.00)
Positive end-expiratory pressure (PEEP) [cm H2O]	5.00	(5.00,5.15)	(0.00,20.00)
Laboratory tests			
Alanine aminotransferase (ALT) [U/L]	29.00	(17.00,67.00)	(7.00,5000.00)
Anion gap [mEq/L]	14.00	(11.00,16.00)	(5.00,40.00)
Calcium ion (CA2+) [mmol/l]	1.12	(1.07,1.18)	(0.80,2.50)
Glucose [mmol/l]	126.00	(105.00,156.50)	(40.00,1000.00)
Haemoglobin [g/dL]	10.60	(9.20,12.20)	(4.00,20.00)
Lactate [mmol/L]	1.70	(1.20,2.50)	(0.30,15.00)
Magnesium [mmol/L]	2.00	(1.80,2.30)	(0.40,5.00)
Partial pressure of oxygen (PO2) [mmHg]	107.00	(71.00,160.00)	(30.00,600.00)
pH	7.36	(7.29,7.41)	(6.80,7.80)
Phosphate [mmol/L]	3.40	(2.80,4.10)	(0.80,8.00)
Platelets [count/L]	187.00	(136.00,249.00)	(10.00,1000.00)
Potassium [mmol/L]	4.10	(3.80,4.50)	(2.00,6.50)
Prothrombin time (PT) [s]	13.70	(12.40,15.70)	(8.00,40.00)
Serum creatinine (Scr) [mg/dL]	0.90	(0.70,1.30)	(0.30,15.00)

Table 2
Event frequencies and prevalences (percentages in brackets).

Event	Included	No event	Event
AF	31,731 (85.18)	29,002 (91.40)	2729 (8.60)
AKI	36,560 (98.14)	33,920 (92.78)	2640 (7.22)
ARDS	21,940 (58.89)	20,649 (94.12)	1291 (5.88)
VAM	23,950 (64.29)	21,630 (90.31)	2320 (9.69)
LOS-3d	37,253 (100.00)	23,148 (62.14)	14,105 (37.86)
LOS-7d	37,253 (100.00)	32,903 (88.32)	4350 (11.68)
In-hospital death	37,253 (100.00)	33,042 (88.70)	4211 (11.30)
30d mortality	37,253 (100.00)	31,055 (83.36)	6198 (16.64)
1yr mortality	37,253 (100.00)	27,190 (72.99)	10,063 (27.01)

substantial proportion of patients developed earlier ARDS (~41 %) or VAM (~35 %) and were consequently excluded from the analysis. Overall, the reported prevalences indicate a significant degree of class imbalance among the tasks.

3.2. Model performances

Tables 3 and 4 report model performance results as measured by the AUC and PR-AUC, respectively. All scores are accompanied by their corresponding confidence intervals. Overall, CNN-based models demonstrated better performance than LR and RF models, although differences in the PR-AUC of AF models are negligible. Additionally, STL and MTL models performed very similarly when implemented using CNNs. The performance results also indicate that ARDS was the most challenging event to predict, as evidenced by the low average values of AUC and PR-AUC, along with wide confidence intervals. In contrast, death-related event models exhibited the best performance, regardless of the algorithm used. Furthermore, the LOS-7d models outperformed LOS-3d models based on the AUC.

The performance of LOS-7d models was much lower when evaluated using the PR-AUC, which is expected given that the prevalence of the LOS-7d event is approximately one-third of that of LOS-3d (Table 2).

3.3. Multiple event visualisation model

Results of the GTM model trained on the predicted event risks are shown in Fig. 4. The resulting GTM membership map is scattered, although the number of admissions seems to be slightly higher on the left-hand side of the map. Additionally, a relatively significant number of clusters allocated less than 100 patients, which could be a symptom of high heterogeneity in the data. The resulting GTM reference maps (Fig. 4b) show a clear pattern of how GTM arranged the risk data on the

Table 3
Model performance report, measured using AUC with corresponding confidence intervals in brackets.

Event	LR	RF	CNN-STL	CNN-MTL
AF	0.74 (0.72, 0.76)	0.75 (0.73, 0.77)	0.77 (0.75, 0.78)	0.77 (0.75, 0.78)
AKI	0.74 (0.72, 0.76)	0.76 (0.74, 0.78)	0.78 (0.76, 0.80)	0.77 (0.75, 0.79)
ARDS	0.63 (0.58, 0.67)	0.68 (0.64, 0.72)	0.72 (0.68, 0.75)	0.71 (0.68, 0.75)
VAM	0.74 (0.71, 0.76)	0.73 (0.71, 0.76)	0.76 (0.74, 0.78)	0.77 (0.75, 0.79)
LOS-3d	0.72 (0.71, 0.73)	0.72 (0.71, 0.74)	0.77 (0.75, 0.78)	0.76 (0.75, 0.77)
LOS-7d	0.77 (0.75, 0.79)	0.77 (0.76, 0.79)	0.81 (0.80, 0.83)	0.82 (0.80, 0.83)
In-hospital death	0.86 (0.85, 0.87)	0.85 (0.83, 0.86)	0.89 (0.88, 0.90)	0.89 (0.88, 0.90)
30d mortality	0.85 (0.84, 0.86)	0.84 (0.83, 0.85)	0.88 (0.87, 0.89)	0.88 (0.87, 0.89)
1yr mortality	0.82 (0.81, 0.83)	0.82 (0.81, 0.83)	0.85 (0.84, 0.86)	0.85 (0.85, 0.86)

Table 4

Model performance report, measured using PR-AUC with corresponding confidence intervals in brackets.

Event	LR	RF	CNN-STL	CNN-MTL
AF	0.22 (0.18, 0.25)	0.22 (0.18, 0.26)	0.21 (0.17, 0.25)	0.22 (0.19, 0.26)
AKI	0.21 (0.17, 0.25)	0.20 (0.16, 0.24)	0.24 (0.20, 0.28)	0.23 (0.19, 0.26)
ARDS	0.12 (0.07, 0.18)	0.13 (0.08, 0.18)	0.16 (0.11, 0.20)	0.15 (0.10, 0.20)
VAM	0.29 (0.25, 0.32)	0.21 (0.17, 0.25)	0.32 (0.28, 0.35)	0.32 (0.29, 0.36)
LOS-3d	0.62 (0.61, 0.63)	0.64 (0.62, 0.65)	0.70 (0.68, 0.71)	0.69 (0.68, 0.70)
LOS-7d	0.33 (0.30, 0.36)	0.34 (0.31, 0.37)	0.40 (0.37, 0.43)	0.40 (0.37, 0.43)
In-hospital death	0.49 (0.46, 0.52)	0.49 (0.46, 0.51)	0.57 (0.55, 0.59)	0.58 (0.55, 0.60)
30d mortality	0.56 (0.53, 0.58)	0.55 (0.53, 0.57)	0.62 (0.60, 0.63)	0.64 (0.62, 0.66)
1yr mortality	0.64 (0.62, 0.65)	0.63 (0.61, 0.64)	0.69 (0.67, 0.70)	0.70 (0.69, 0.71)

membership map. By their inspection, it is observed that the membership map's bottom left represents patients at very low risk of any of the considered events. In addition, ARDS risk almost splits the membership map into two well-defined regions, with the top and top-right ones allocated to higher risks. A similar scenario is presented with AKI, although with higher risks concentrated in the top-right corner. GTM reference maps also show some level of correlation between LOS-3d and LOS-7d, which is anticipated as patients with LOS longer than 7 days would have stayed longer than 3 days. A similar situation is observed between In-hospital death, 30d mortality and 1yr mortality. It is also interesting to see the representation of AF risk by GTM, which seems to be split into several areas that overlap with other event risks. Overall, Fig. 4b indicates a high degree of overlap between two or more events, suggesting that the number of patients at risk of several events is significant.

Fig. 5 presents the resulting class activation maps generated using HiResCAM for five randomly selected patient admissions. According to the figure, the implemented CNN-MTL model predicted an overall low risk of adverse events for Patient I (bottom-left in the membership map),

with the model focusing primarily on lower diastolic BP, highest haemoglobin and GCS Total variable values. In contrast, high risks of ARDS, VAM, and in-hospital death were predicted for Patient II (bottom-right in the membership map), with predominant attention on high lactate and low magnesium values. GTM placed Patient III in the bottom-centre of the membership map, which is associated with a predicted high risk of developing AF. The CNN-MTL model focused on glucose changes, early heart rate increase, low magnesium and PO₄ levels, as well as patient age and sex. According to the membership map, Patient IV is at high risk of developing AKI only. The HiResCAM results indicate that glucose changes, SCr levels, initial systolic BP, and low temperature values were particularly relevant to the risk predictions. In contrast, Patient V is predicted to be at risk of multiple adverse events: ARDS (high risk), VAM (medium), LOS-3d (high), LOS-7d (medium-high), in-hospital death (medium), and 30-day and 1-year mortalities (medium-high). The CNN-MTL model focused particularly on early low ALT, changes in anion gap, diastolic BP and systolic BP, decreasing FIO₂, heart rate variability, and changes in magnesium, platelets, and PO₂ values. Additionally, some attention was directed towards late temperature increases and changes in GCS Total.

The results of the macro-clustering are displayed in Figs. 6 and 7. The hierarchical clustering algorithm partitioned the GTM prototype data into 10 macro-clusters (Fig. 6). Fig. 7 shows the predicted event risks averaged across the estimated macro-clusters. According to these results, Cluster 1 represents patients simultaneously at medium risk of AF and AKI, high risk of VAM, LOS-3d/-7d, In-hospital death, 30d/1yr mortalities, and very high risk of ARDS. In contrast, Cluster 8 groups patients with the lowest predicted risk of any adverse events. Patients in Cluster 3 are at very high risk of ARDS and VAM, those in Cluster 4 are at very high risk of AF, while Cluster 7 is characterised by very high risk of VAM and Cluster 9 by very high risk of AKI. Moreover, Cluster 6 represents patients at an increasing risk of death (medium risk for In-hospital death, high risk for 30d mortality, and higher risk for 1yr mortality). Patients in this cluster are also predicted to be at significant risk of AF, VAM, and LOS-3d. In contrast, Cluster 2 groups patients at low risk of death but very high risk of ARDS and VAM, and high risk of LOS-3d. Finally, Cluster 10 represents patients at medium-low to medium risk of developing multiple adverse events.

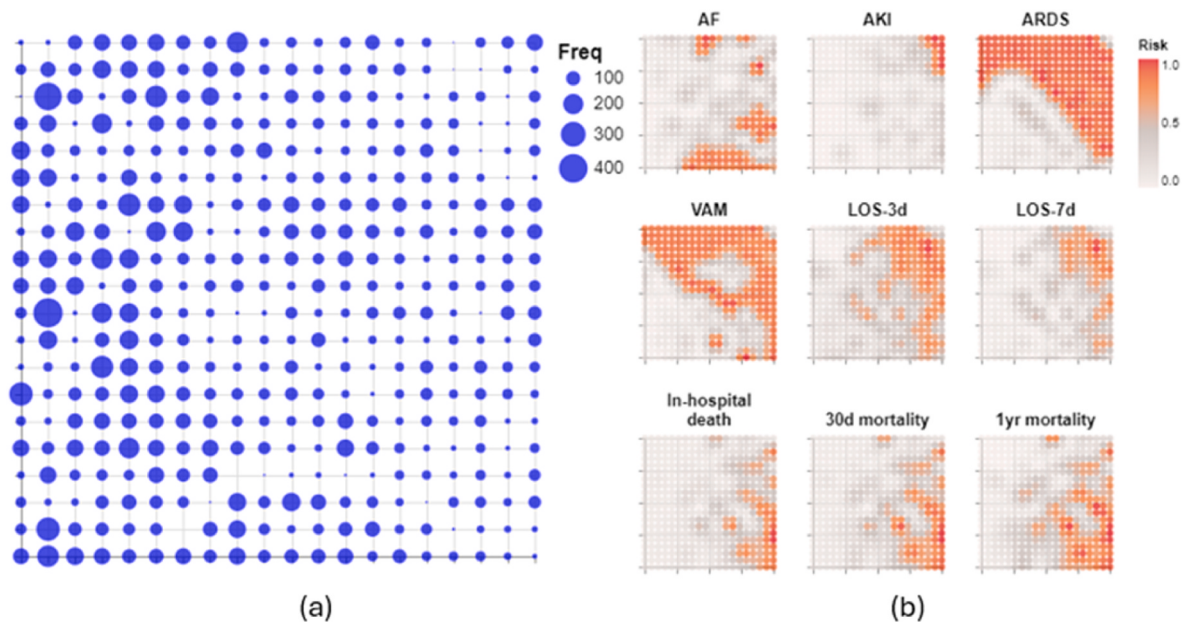


Fig. 4. Resulting GTM membership (a) and reference maps (b). In (a), node size represents the number of admissions allocated to each cluster. Event risks (probabilities) are colour-coded from light grey to red, indicating low to high risk, respectively.

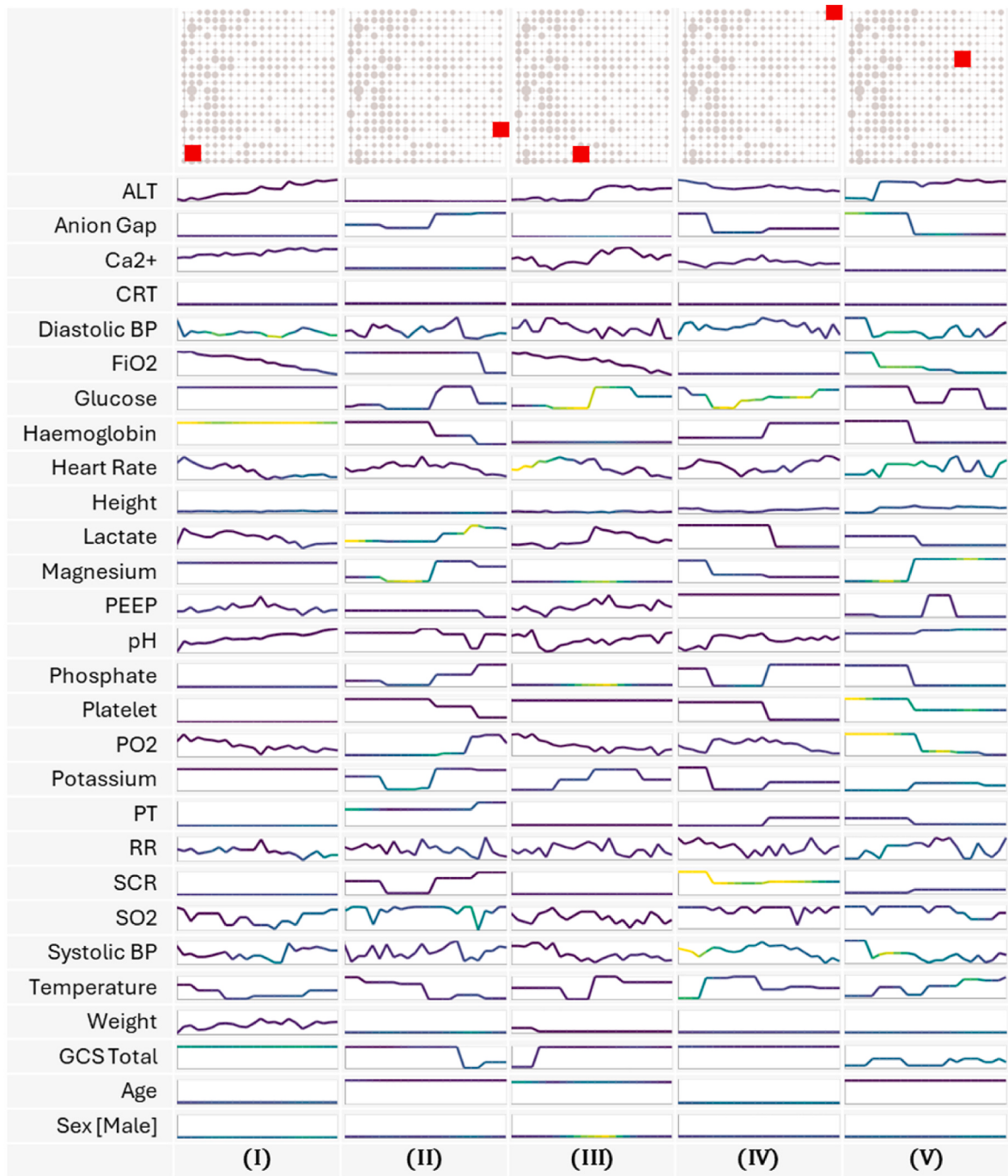


Fig. 5. Random selection of five patient admissions, their cluster allocations in the GTM membership map (top row), and the resulting HiResCAM plots, one per variable and admission. Admission locations in the membership map are represented with red squares overlaid on the membership map, shown in light grey. In each line plot, the horizontal and vertical axes represent time (24 h) and normalised variable values (ranging from 0, lowest, to 1, highest), respectively. HiResCAM continuous values are colour-coded from purple to light yellow, indicating low to high attention values. Abbreviations as in Table 1.

4. Discussion

In this paper, we proposed IMERF, a novel integrated machine learning framework for modelling and representing multiple events. This innovative methodological approach advances the development of risk prediction models by enabling the simultaneous prediction and characterisation of complex associations between input factors and multiple events. It uncovers relationships that remain invisible when events are modelled independently, offering unprecedented insights into risk patterns. By leveraging shared learning across tasks, it pushes the boundaries of prediction accuracy, outperforming traditional methods.

Unlike outdated approaches that rely on manually constructed composite variables with arbitrary weights, this innovative framework directly models multiple events, delivering superior interpretability, precision, and transformative potential for risk modelling. Although IMERF was used for modelling multiple adverse events in patients admitted to the ICU to showcase its capabilities, our proposed framework has significant potential for addressing many other clinical questions. IMERF is an ideal framework for modelling clusters of medical conditions and discovering new ones. We firmly believe that IMERF is a much-needed approach for integrated multiple event risk modelling. For example, an IMERF model could be implemented for integrated risk

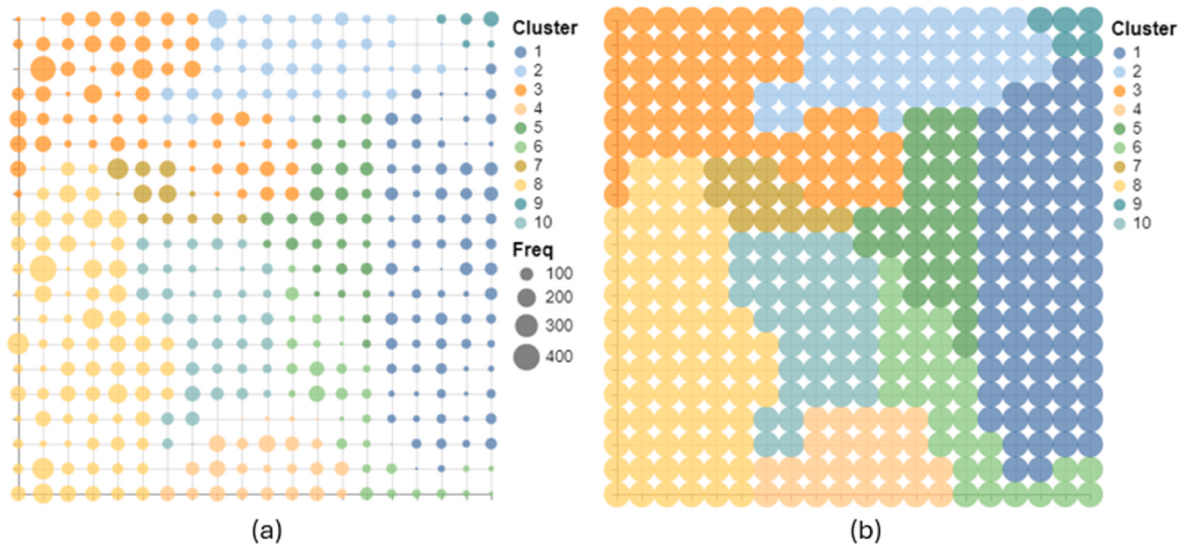


Fig. 6. Results of the macro-clustering. Macro-clusters are colour-coded and overlaid onto the GTM membership map (a). To enhance clarity, the same map is presented with all GTM nodes represented at an equal size (b).

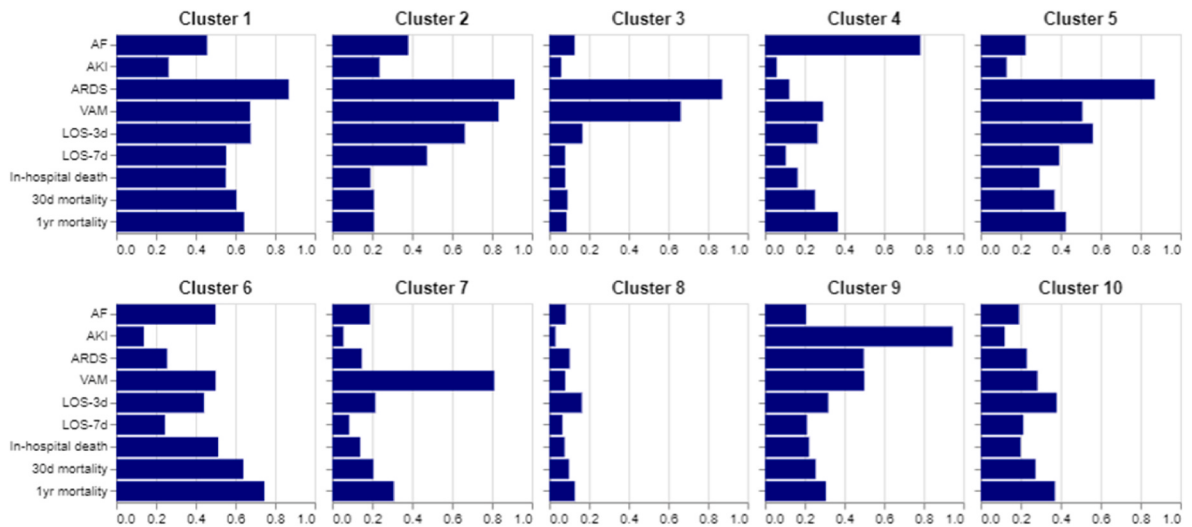


Fig. 7. Average predicted event risks per macro-cluster. In each bar plot, the horizontal axis represents the average risk for each event.

modelling of individuals at risk of developing multiple co-morbidities. Additionally, IMERF can be considered for modelling several instances of a single event, such as helping to characterise patients at risk of developing paroxysmal, persistent, or permanent AF; or conditions with a concurrent nature, such as elevated blood pressure (BP) and stroke.

The proposed IMERF is a two-stage approach comprising a multiple-event risk prediction model and a data visualisation and clustering model. Although the aim of this study was not to develop a predictive model that maximises performance, we took the usual measures towards that goal. We can speculate that by testing a broader range of hyperparameters and considering more ML algorithms, higher performances could still be achieved. However, we believe that the reported results are close to their performance ceilings, given the data and variables used. When comparing the performance of models that treat the input variables as sequences (CNN-STL and CNN-MTL), rather than as aggregated tabular data (LR and RF), it is not surprising that the former produced higher scores, which aligns with results previously reported in Refs. [31–33] when using similar data.

Additionally, the implementation of the multiple event risk prediction model using an MTL architecture offers clear advantages over a

multiple STL approach, even when both architectures achieve equivalent performance scores: MTL models tend to be more compact, requiring significantly fewer parameters to be learned than their multiple STL counterparts. Consequently, the implemented CNN-MTL model would be considerably less prone to overfitting than the CNN-STLs. It is also known that MTL can enhance the performance of less well-defined, smaller tasks by leveraging the larger, better-defined ones [16,34,35]. However, we did not observe this effect in our results, which may indicate little room for performance improvement in the data.

With regard to Stage 2 of the IMERF, we previously mentioned that any algorithm for data visualisation and clustering could, in principle, be used. However, we believe that the use of GTM as our preferred choice is justified. Although non-linear, GTM is inherently interpretable, making it a valuable tool for exploring complex relationships within the data. In this case study, reference maps were generated to assist in identifying associations between the predicted event risks and the clustering structure established by the GTM. These maps enabled the identification of regions where patients are at risk not only of a specific adverse event but also of multiple adverse events occurring simultaneously. This interpretability aids in understanding how clusters

represent combined risks, which is crucial for clinical decision-making in ICU settings.

Furthermore, depending on the algorithm employed for the multi-event risk prediction model, it is possible to trace how input variables influence the GTM clusters. This traceability adds another layer of interpretability, allowing researchers and clinicians to better understand the relationships between patient features and adverse event risks. In our case study, we leveraged the CNN-MTL model to extract class activation maps via HiResCAM. We used this approach to highlight specific segments of the input time series data that were most strongly associated with the model's predictions. By combining the interpretability of GTM reference maps with the explanatory insights provided by HiResCAM, we obtained a more comprehensive understanding of the underlying relationships between input features, event risks, and clustering patterns.

The results of applying our proposed IMERF to model critically-ill patients highlight the utility and need for a novel methodological approach that addresses the analysis and interpretation of multiple events simultaneously. A key aspect of IMERF lies in its ability to identify clusters based on overlapping risks of multiple events. When IMERF was applied in our case study, the results revealed clearly differentiated clusters based on the considered events. For instance, patients II and III (Fig. 5) were both at risk of AF, yet they presented significantly different risks of death, with patient II showing a much higher risk. This distinction was confirmed by the event reference maps shown in Fig. 4.

According to the class activation map results from the CNN-MTL model, increasing levels of lactate appeared to be more relevant for patient II than for patient III. Elevated lactate is a common marker of tissue hypoxia, impaired oxygen utilisation, or metabolic stress and can result from conditions such as sepsis, hypovolaemia, cardiogenic shock, or liver dysfunction, all strongly linked to high mortality rates [36]. However, both patients exhibited low levels of magnesium, which was relevant to the CNN-MTL model. This could indicate hypomagnesaemia, which is associated with an increased risk of arrhythmias, including AF [37]. Additionally, the CNN-MTL model highlighted patient III's changes in heart rate and glucose levels, both of which are known to contribute to AF by affecting metabolic, autonomic, and inflammatory pathways [38].

A similar analysis can be performed with patient IV, who was predicted to have a very high risk of AKI, as indicated in the top-right corner of the GTM membership map. The HiResCAM results for this patient showed an elevated serum creatinine (SCR) level. An increase in SCR by 0.3 mg/dL or more within 48 h, or to 1.5 times the baseline or more, indicates the patient has AKI [39]. Changes in glucose levels (glycaemic variability) also emerged as another relevant factor for the CNN-MTL model. Large fluctuations in glucose levels are known to be more strongly associated with higher rates of AKI than conditions such as hyperglycaemia [40].

Finally, patient V presented a very high risk of ARDS and a high risk of requiring VAMs. The CNN-MTL model identified variables in this patient that serve as markers of haemodynamic instability, such as low magnesium, platelet, and systolic BP levels, which are associated with the need for VAMs. Additionally, fluctuations in heart rate, which could indicate shock or inadequate perfusion, and an elevated anion gap, which may suggest metabolic acidosis, could also signal the need for VAMs [41]. Furthermore, patient V presented clear indications of ARDS risk, with a dramatic decrease in FiO₂ and PO₂ levels during the input time window, which could indicate symptoms of hypoxia, a hallmark of ARDS [42]. The relationship between ARDS and the need for VAMs is well-documented, as haemodynamic instability can both exacerbate and be exacerbated by ARDS. This is also confirmed by the GTM reference maps (Fig. 4b), which indicate a high correlation between the risks of ARDS and the need for VAMs [43].

An important consideration is to identify how much of the patterns revealed by the GTM visualisation model are attributable to the type of

intensive care unit in which the patient stayed and, or their mechanical ventilation status. We therefore performed a stratified analysis to examine how reference maps and macro-clusters varied across ICUs, and a similar analysis to assess the influence of mechanical ventilation status on these visualisations and macro-cluster structures. The results of these analyses are presented in Section S4 of the Supplementary Material. Apart from the CVICU, no clear correlation was observed between ICU type and the GTM visualisation maps (Figs. 4 and 6 in the main manuscript and Fig. S2 in the Supplementary Material). A significant correlation was found between mechanical ventilation and elevated ARDS risk, likely because the clinical criteria for ARDS diagnosis incorporate mechanical ventilation parameters and oxygenation measures. Correlations between CVICU admissions, mechanical ventilation, and predicted macro-clusters 2 and 3 (Fig. 6 in the main manuscript and Figs. S4 and S12 in the Supplementary Material) may be explained by the higher prevalence of post-surgical and advanced cardiovascular cases in these groups, which are associated with both invasive respiratory support and increased risk of multiple adverse events.

A potential limitation of the current IMERF architecture is that Stage 1 and Stage 2 are optimised independently. It could be argued that better overall performance might be achieved if the two stages were optimised concurrently. However, the downside of this approach is that it would necessitate the development of new formulations to enable such integrated optimisation, thereby significantly restricting the range of algorithms available for IMERF implementation.

Additionally, the IMERF approach implemented in the presented case study does not explicitly model the chronological order of overlapping events. This is partly due to the algorithm used in Stage 2 of IMERF. In future work, the GTM Through Time [44], a version of GTM that explicitly models sequences, could be considered as an alternative to the traditional GTM implementation used in this case study. Moreover, restricting the analysis to only the first admission for each patient likely reduces the observed rate of multiple comorbidities and adverse events, which represents another limitation of the case study design.

Finally, further research could explore how to fully leverage GTM's potential in modelling probabilities. For instance, GTM could be used to determine whether a patient is likely to transition to another cluster with a higher (or lower) risk, thus providing deeper insights within the IMERF methodological approach. The proposed framework will be instrumental in addressing the modelling of multiple adverse events within extensive research programmes, such as the EU projects ARISTOTELES [45] and TARGET [46].

5. Conclusion

We proposed IMERF, a novel ML methodological approach for modelling risk predictions for multiple events. This approach employs a two-stage process comprising an MTL model followed by a DR model for data visualisation and clustering. We successfully demonstrated the utility of IMERF in clustering events based on their predicted risk levels, as observed in the presented case study involving critically ill patients. Furthermore, the integration of GTM with the CNN-MTL's HiResCAM proved to be a highly interpretable framework that facilitates understanding of factors associated with multiple event risks occurring simultaneously. Finally, the extracted macro-clusters enhance IMERF's interpretability by segmenting the GTM visualisation map into regions based on the average levels of multiple event risks. Future work includes the use of algorithms such as the GTM Through Time to model sequentially multiple event risks.

CRedit authorship contribution statement

Ivan Olier: Writing – review & editing, Writing – original draft, Visualization, Validation, Methodology, Formal analysis, Data curation, Conceptualization. **Sandra Ortega-Martorell:** Writing – review & editing, Visualization, Validation, Methodology. **George Margereson:**

Writing – review & editing, Software. **Ryan A. Bellfield:** Writing – review & editing, Visualization, Software. **Ingeborg D. Welters:** Writing – review & editing. **Gregory Y.H. Lip:** Writing – review & editing.

Ethics statement

This paper uses patient data from the MIMIC-IV database, available on the PhysioNet portal (<https://physionet.org/content/mimiciv/2.2/>) for credentialed users. It does not involve primary data collection.

The authors confirm that the work complies with the ethical standards for scholarly publishing, including:

- **Originality and Plagiarism:** The manuscript is original, has not been published elsewhere, and does not contain plagiarised material. All sources are appropriately cited and referenced.
- **Conflict of Interest:** There are no conflicts of interest that could influence the outcomes or interpretations presented in this paper.
- **Acknowledgment of Sources:** All studies, data, and other materials referenced in this paper are duly acknowledged to ensure transparency and credit to the original authors.
- **Compliance with Journal Guidelines:** The manuscript adheres to the ethical requirements and policies of the journal.

Disclaimer

Funded by the European Union. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the Health and Digital Executive Agency. Neither the European Union nor the granting authority can be held responsible for them.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT to review the grammar. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

Declaration of competing interest

I.O. - Methodological lead and co-PI of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244) and partner lead in the ARISTOTELES project on artificial intelligence for the management of chronic long-term conditions (grant agreement no. 101080189), both funded by the EU's Horizon Europe Research & Innovation programme. No fees are received personally.

S.O.M. - Principal Investigator of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244) and senior investigator in the ARISTOTELES project on artificial intelligence for the management of chronic long-term conditions (grant agreement no. 101080189), both funded by the EU's Horizon Europe Research & Innovation programme. She is also a member of the board of the ART (Ageing Research Translation) of Healthy Ageing Network funded by the Biotechnology and Biological Sciences Research Council (BBSRC). No fees are received personally.

G.M. - Team member of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244).

R.A.B. - Team member of the ARISTOTELES project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101080189).

I.D.W. - Team member of the TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant

agreement no. 101136244).

G.Y.H.L. - Consultant and speaker for BMS/Pfizer, Boehringer Ingelheim, Daiichi-Sankyo, and Anthos. No fees are received personally. He is a National Institute for Health and Care Research (NIHR) Senior Investigator and co-PI of the AFFIRMO project on multimorbidity in AF (grant agreement no. 899871), TARGET project on health virtual twins for personalised management of atrial fibrillation and stroke (grant agreement no. 101136244) and ARISTOTELES project on artificial intelligence for the management of chronic long-term conditions (grant agreement no. 101080189), which are all funded by the EU's Horizon Europe Research & Innovation programme.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.combiomed.2025.111196>.

References

- [1] B.A. Goldstein, A.M. Navar, M.J. Pencina, J.P.A. Ioannidis, Opportunities and challenges in developing risk prediction models with electronic health records data: a systematic review, *J. Am. Med. Inf. Assoc.* 24 (2017) 198, <https://doi.org/10.1093/JAMIA/OCW042>.
- [2] R. Giddings, A. Joseph, T. Callender, S.M. Janes, M. van der Schaar, J. Sheringham, N. Navani, Factors influencing clinician and patient interaction with machine learning-based risk prediction models: a systematic review, *Lancet Digit. Health* 6 (2024) e131–e144, [https://doi.org/10.1016/S2589-7500\(23\)00241-8](https://doi.org/10.1016/S2589-7500(23)00241-8).
- [3] A.-K. Ozga, M. Kieser, G. Rauch, A systematic comparison of recurrent event models for application to composite endpoints, *BMC Med. Res. Methodol.* 18 (2018) 2, <https://doi.org/10.1186/s12874-017-0462-x>.
- [4] J.C. Jentzer, S. van Diepen, G.W. Barsness, J.N. Katz, B.M. Wiley, C.E. Bennett, S. V. Mankad, L.J. Sinak, P.J. Best, J. Herrmann, A.S. Jaffe, J.G. Murphy, D. A. Morrow, R.S. Wright, M.R. Bell, N.S. Anavekar, Changes in comorbidities, diagnoses, therapies and outcomes in a contemporary cardiac intensive care unit population, *Am. Heart J.* 215 (2019) 12–19, <https://doi.org/10.1016/j.ahj.2019.05.012>.
- [5] J. Wiens, J. Guttat, E. Horvitz, Patient risk stratification with time-varying parameters: a multitask learning approach, *J. Mach. Learn. Res.* 17 (2016).
- [6] X. Wang, F. Wang, J. Hu, A multi-task learning framework for joint disease risk prediction and comorbidity discovery, in: *Proceedings - International Conference on Pattern Recognition*, 2014, <https://doi.org/10.1109/ICPR.2014.47>.
- [7] J. Zhou, J. Liu, V.A. Narayan, J. Ye, Modeling disease progression via multi-task learning, *Neuroimage* 78 (2013), <https://doi.org/10.1016/j.neuroimage.2013.03.073>.
- [8] R. Miotto, F. Wang, S. Wang, X. Jiang, J.T. Dudley, Deep learning for healthcare: review, opportunities and challenges, *Briefings Bioinf.* 19 (2017), <https://doi.org/10.1093/bib/bbx044>.
- [9] A. Triantafyllidis, H. Kondylakis, D. Katehakis, A. Kouroubali, L. Koumakis, K. Marias, A. Alexiadis, K. Votis, D. Tzovaras, Deep learning in mHealth for cardiovascular disease, diabetes, and cancer: systematic review, *JMIR Mhealth Uhealth* 10 (2022), <https://doi.org/10.2196/32344>.
- [10] Y. Zhao, X. Wang, T. Che, G. Bao, S. Li, Multi-task deep learning for medical image computing and analysis: a review, *Comput. Biol. Med.* 153 (2023), <https://doi.org/10.1016/j.combiomed.2022.106496>.
- [11] A.B. Teshale, H.L. Htun, M. Vered, A.J. Owen, R. Freak-Poli, A systematic review of artificial intelligence models for time-to-event outcome applied in cardiovascular disease risk prediction, *J. Med. Syst.* 48 (2024) 68, <https://doi.org/10.1007/s10916-024-02087-7>.
- [12] A. Amirahmadi, M. Ohlsson, K. Etmnani, Deep learning prediction models based on EHR trajectories: a systematic review, *J. Biomed. Inf.* 144 (2023), <https://doi.org/10.1016/j.jbi.2023.104430>.
- [13] T.H. Pham, C. Yin, L. Mehta, X. Zhang, P. Zhang, A fair and interpretable network for clinical risk prediction: a regularized multi-view multi-task learning approach, *Knowl. Inf. Syst.* 65 (2023), <https://doi.org/10.1007/s10115-022-01813-2>.
- [14] H.G.M. Walker, A.J. Brown, I.P. Vaz, R. Reed, M.A. Schofield, J. Shao, V. B. Nanjaya, A.A. Udy, T. Jeffcote, Composite outcome measures in high-impact critical care randomised controlled trials: a systematic review, *Crit. Care* 28 (2024), <https://doi.org/10.1186/s13054-024-04967-3>.
- [15] H. Harutyunyan, H. Khachatrian, D.C. Kale, G. Ver Steeg, A. Galstyan, Multitask learning and benchmarking with clinical time series data, *Sci. Data* 6 (2019), <https://doi.org/10.1038/s41597-019-0103-9>.
- [16] Y. Zhang, Q. Yang, A survey on multi-task learning, *IEEE Trans. Knowl. Data Eng.* 34 (2022) 5586–5609, <https://doi.org/10.1109/TKDE.2021.3070203>.
- [17] D. Xu, Y. Shi, I.W. Tsang, Y.-S. Ong, C. Gong, X. Shen, Survey on multi-output learning, *IEEE Transact. Neural Networks Learn. Syst.* (2019) 1–21, <https://doi.org/10.1109/TNNLS.2019.2945133>.
- [18] M. Crawshaw, Multi-Task Learning with Deep Neural Networks: a Survey, 2020.
- [19] L. Van Der Maaten, G. Hinton, Visualizing Data Using t-SNE, 2008.
- [20] J. Healy, L. McInnes, Uniform manifold approximation and projection, *Nature Rev. Method. Primers* 4 (2024), <https://doi.org/10.1038/s43586-024-00363-x>.

- [21] F. Anowar, S. Sadaoui, B. Selim, Conceptual and empirical comparison of dimensionality reduction algorithms (PCA, KPCA, LDA, MDS, SVD, LLE, ISOMAP, LE, ICA, t-SNE), *Comput Sci. Rev.* 40 (2021), <https://doi.org/10.1016/j.cosrev.2021.100378>.
- [22] C.M. Bishop, M. Svensén, C.K.I. Williams, GTM: the generative topographic mapping, *Neural Comput.* 10 (1998) 215–234, <https://doi.org/10.1162/089976698300017953>.
- [23] A.E.W. Johnson, L. Bulgarelli, L. Shen, A. Gayles, A. Shammout, S. Horng, T. J. Pollard, B. Moody, B. Gow, L. wei H. Lehman, L.A. Celi, R.G. Mark, MIMIC-IV, a freely accessible electronic health record dataset, *Sci. Data* 10 (2023), <https://doi.org/10.1038/s41597-022-01899-x>.
- [24] S. Ortega-Martorell, M. Pieroni, B.W. Johnston, I. Olier, I.D. Welters, Development of a risk prediction model for new episodes of atrial fibrillation in medical-surgical critically ill patients using the AmsterdamUMCdb, *Front. Cardiovasc. Med.* 9 (2022), <https://doi.org/10.3389/fcvm.2022.897709>.
- [25] M.O. Akinwande, H.G. Dikko, A. Samson, Variance inflation factor: as a condition for the inclusion of suppressor variable(s) in regression analysis, *Open J. Stat.* 5 (2015) 754–767, <https://doi.org/10.4236/ojs.2015.57075>.
- [26] S. van Buuren, K. Groothuis-Oudshoorn, Mice: multivariate imputation by chained equations in R, *J. Stat. Software* 45 (2011) 1–67, <https://doi.org/10.18637/jss.v045.i03>.
- [27] A. Derry, M. Krzywinski, N. Altman, Convolutional neural networks, *Nat. Methods* 20 (2023) 1269–1270, <https://doi.org/10.1038/s41592-023-01973-1>.
- [28] R.A.A. Bellfield, P.R. Hormiga, I. Olier, R. Lotto, I. Jones, G.Y.H. Lip, S. Ortega-Martorell, AI-driven clustering and visualisation of ECG signals to enhance screening for atrial fibrillation: the supermarket/hypermarket opportunistic screening for atrial fibrillation (SHOPS-AF) study, *Heart Rhythm O2* (2025), <https://doi.org/10.1016/j.hroo.2025.07.003>, 0.
- [29] R.L. Draelos, L. Carin, *Use Hirescam Instead of Grad-CAM for Faithful Explanations of Convolutional Neural Networks*, 2020.
- [30] R.A.A. Bellfield, I. Olier, R. Lotto, I. Jones, E.A. Dawson, G. Li, A.M. Tuladhar, G.Y. H. Lip, S. Ortega-Martorell, AI-based derivation of atrial fibrillation phenotypes in the general and critical care populations, *EBioMedicine* 107 (2024), <https://doi.org/10.1016/J.EBIO.2024.105280>.
- [31] H. Harutyunyan, H. Khachatrian, D.C. Kale, G. Ver Steeg, A. Galstyan, Multitask learning and benchmarking with clinical time series data, *Sci. Data* 6 (2019) 96, <https://doi.org/10.1038/s41597-019-0103-9>.
- [32] S. Ortega-Martorell, I. Olier, B.W. Johnston, I.D. Welters, Sepsis-induced coagulopathy is associated with new episodes of atrial fibrillation in patients admitted to critical care in sinus rhythm, *Front Med (Lausanne)* 10 (2023), <https://doi.org/10.3389/fmed.2023.1230854>.
- [33] N. Aikodon, S. Ortega-Martorell, I. Olier, Predicting decompensation risk in intensive care unit patients using machine learning, *Algorithms* 17 (2023) 6, <https://doi.org/10.3390/a17010006>.
- [34] N. Sadawi, I. Olier, J. Vanschooren, J.N. Van Rijn, J. Besnard, R. Bickerton, C. Grosan, L. Soldatova, R.D. King, Multi-task learning with a natural metric for quantitative structure activity relationship learning, *J. Cheminf.* 11 (2019), <https://doi.org/10.1186/s13321-019-0392-1>.
- [35] T.H. Pham, C. Yin, L. Mehta, X. Zhang, P. Zhang, A fair and interpretable network for clinical risk prediction: a regularized multi-view multi-task learning approach, *Knowl. Inf. Syst.* 65 (2023), <https://doi.org/10.1007/s10115-022-01813-2>.
- [36] S. Kushimoto, S. Akaishi, T. Sato, R. Nomura, M. Fujita, D. Kudo, Y. Kawazoe, Y. Yoshida, N. Miyagawa, Lactate, a useful marker for disease mortality and severity but an unreliable marker of tissue hypoxia/hypoperfusion in critically ill patients, *Acute Med.Surg.* 3 (2016) 293–297, <https://doi.org/10.1002/ams2.207>.
- [37] A.G. Negru, A. Pastorici, S. Crisan, G. Cismaru, F.G. Popescu, C.T. Luca, The role of hypomagnesemia in cardiac arrhythmias: a clinical perspective, *Biomedicine* 10 (2022) 2356, <https://doi.org/10.3390/biomedicine10102356>.
- [38] J.-Y. Huang, A.-P. Cai, C.T.W. Tsang, M.-Z. Wu, W.-L. Gu, R. Guo, J.-N. Zhang, C.-Y. Zhu, Y.-M. Hung, G.Y.H. Lip, K.-H. Yiu, The association of haemoglobin A1c variability with adverse outcomes in patients with atrial fibrillation prescribed anticoagulants, *Eur. J. Prev. Cardiol.* (2024), <https://doi.org/10.1093/eurjpc/zwae249>.
- [39] S. Verma, J.A. Kellum, Defining acute kidney injury, *Crit. Care Clin.* 37 (2021) 251–266, <https://doi.org/10.1016/j.ccc.2020.11.001>.
- [40] L. Wen, Y. Li, S. Li, X. Hu, Q. Wei, Z. Dong, Glucose metabolism in acute kidney injury and kidney repair, *Front Med (Lausanne)* 8 (2021) 744122, <https://doi.org/10.3389/fmed.2021.744122>.
- [41] K.M.L. Leunissen, J.P. Kooman, W. Van Kuijk, F. Van Der Sande, A.J. Luik, J.P. Van Hooff, Preventing haemodynamic instability in patients at risk for intra-dialytic hypotension, in: *Nephrology Dialysis Transplantation*, Oxford University Press, 1996, pp. 11–15, <https://doi.org/10.1093/ndt/11.suppl2.11>.
- [42] J.F.F.M. Ferraz, M.T. Siuba, S. Krishnan, R.L. Chatburn, E. Mireles-Cabodevila, A. Duggal, Physiologic markers of disease severity in ARDS, *Respir. Care* 68 (2023) 1708–1718, <https://doi.org/10.4187/respcare.11100>.
- [43] G.A. Cortes-Puentes, R.A. Oeckler, J.J. Marini, Physiology-guided management of hemodynamics in acute respiratory distress syndrome, *Ann. Transl. Med.* 6 (2018), <https://doi.org/10.21037/atm.2018.04.40>, 353–353.
- [44] I. Olier, A. Vellido, Advances in clustering and visualization of time series using GTM through time, *Neural Netw.* 21 (2008) 904–913, <https://doi.org/10.1016/j.neunet.2008.05.013>.
- [45] G. Boriani, D.A. Mei, G.Y.H. Lip, ARISTOTELES consortium, artificial intelligence in patients with atrial fibrillation to manage clinical complexity and comorbidities: the ARISTOTELES project, *Eur. Heart J.* (2024), <https://doi.org/10.1093/eurheartj/ehae792>.
- [46] S. Ortega-Martorell, I. Olier, M. Ohlsson, G.Y.H. Lip, TARGET consortium, TARGET: a major European project aiming to advance the personalised management of atrial fibrillation-related stroke via the development of health virtual twins technology and artificial intelligence, *Thromb. Haemost.* 125 (2025) 7–11, <https://doi.org/10.1055/a-2438-5671>.