Using Associative Classification and Odds Ratios for In-Hospital Mortality Risk Estimation

Oliver Haas [1 2] **Andreas Maier** [2] **Eva Rothgang** [1]

Abstract

We propose a novel method based on associative classification in combination with odds ratios, a well-understood epidemiological metric, as an interpretable method for in-hospital mortality estimation, which is influenced by thousands of clinical variables. We tested and validated the method for cases in intensive and emergency care. The resulting model achieves an area under the receiver operating characteristic curve of 0.98. The model is easy to interpret in the form of one-to-one rules and the corresponding odds ratios. This study shows that associative classification combined with epidemiological metrics can be used as effective and interpretable machine learning models in the presence of outcomes that are influenced by thousands of variables.

1. Introduction

Intensive care units (ICUs) and emergency departments (EDs) provide care for the critically ill. In-hospital mortality, which is the term for when a patient dies during their stay in the hospital, is part of this care. In a complex clinical environment, different factors play a vital role for in-hospital mortality. Socio-demographic factors describe the patients as well as their background. Diagnoses, medication, and procedures, describe the clinical care the patients experience. Both the patients as well as their care influence in-hospital mortality. When viewed as data recorded in electronic health records or other forms of clinical data, this plethora of factors results in thousands of different variables that may or may not be associated with in-hospital mortality.

Scoring systems and machine learning methods allow providers to use clinical data to arrive at an objective estimation of the in-hospital mortality risk based on these variables. Various scoring systems have been developed to estimate the mortality risk or a case's severity in different scenarios (Salluh & Soares, 2014). Such a risk score can objectively assess the urgency, mortality risk, and patients' expected need of care.

Next to these standardized analog scoring systems, various machine learning methods have been used in the last years (Xie et al., 2017; Johnson et al., 2017; Keuning et al., 2020). These models provide scoring systems tailored to the respective clinic because they are learned directly from the data recorded in this clinic. Interpretability of these methods has been discussed both in in-hospital mortality prediction (Xie et al., 2017) and early warning score systems (Fu et al., 2020), mentioning that there is a compromise between interpretability and predictive performance. Decision trees show the best interpretability among the commonly used models, but often lack predictive performance (Xie et al., 2017).

Apart from this compromise, there is another challenge. ICUs and EDs are, by design, caring for patients with very different conditions and illnesses. Mortality risk estimation methods need to incorporate many different variables of different types, as many variables are associated with mortality. This hinders interpretability, as more variables that are associated with the outcome of interest make models even more complex.

One class of machine learning models showing high interpretability are associative classifiers (Thabtah, 2007). They are built of association rules, i.e., rules of the form "$A \Rightarrow B$" where A is a set of variables and B the outcome(s) of interest. These rules are first mined from the dataset to build a set of rules that fulfill configurable quality constraints. Given a new observation, this set of rules is then evaluated on the observation and ranked or aggregated to form a prediction.

The advantages of associative classifiers are their analytic nature and their high interpretability. These advantages stem from the fact that the algorithm's decision can be understood in the form of human-readable if-then-rules. Additionally, associative classifiers can incorporate large numbers of heterogeneous variables.

[1]Institute of Medical Engineering, Department of Industrial Engineering and Health, Technical University Amberg-Weiden, Germany [2]Pattern Recognition Lab, Department of Computer Science, Technical Faculty, Friedrich-Alexander University, Erlangen, Germany. Correspondence to: Oliver Haas <o.haas@oth-aw.de>.

Workshop on Interpretable ML in Healthcare at International Conference on Machine Learning (ICML).

Associative classification methods have previously been used in the field of healthcare. Notable fields of application include disease and wellness prediction (Meena et al., 2019; Dua et al., 2009; Rea & Huff, 2010; Ujager & Mahmood, 2019), outcome prediction and adverse drug reactions (Kadkhoda et al., 2020; El Houby, 2014; Uriarte-Arcia et al., 2014; Lin et al., 2012), and biochemistry and genetics (Yu & Wild, 2012a;b; He & Hui, 2009; ElHefnawi & Sherif, 2014; Kianmehr & Alhajj, 2008). In the field of in-hospital mortality prediction, Cheng et. al. used association rules based on twelve lab test results (2015). While this study shows the feasibility of associative classification, the present study expands the approach by incorporating heterogeneous variables and much more variables in general.

We want to assess the association that a variable has with in-hospital mortality. In epidemiology, odds ratios (ORs) are often used to measure the strength and direction of an association between variables. We use this metric to measure the quality of the association rules in this study. ORs have previously been used with associative classification (Lin et al., 2012). They are widely used in epidemiology and can be used to measured both the direction and the strength of an association. We give a short introduction to ORs as they are less often used in machine learning.

Let $\mathcal{D}$ denote the clinical dataset under analysis. $\mathcal{D}$ is a collection of clinical cases, with each case being a set comprised of the variables that occurred during the case. Let $y = 1$ denote that the patient died during the stay, and $y = 0$ that they did not. In this way, we can assign to each case A in $\mathcal{D}$ the number 1 or 0, depending on the outcome of the case A. This leads to the following definition.

Given a variable x and y either 0 or 1, we denote by $\text{supp}(x, y)$ the number of cases that include x and have outcome y, i.e.

$$\text{supp}(x, y) = |\{A \in \mathcal{D} \mid x \in A, A \text{ has outcome } y\}|.$$

The term $\text{supp}(x, y)$ is called the *support of x with outcome y*. We additionally define

$$\text{supp}(y) = |\{A \in \mathcal{D} \mid A \text{ has outcome } y\}|,$$

which is the number of all cases with outcome y without imposing any criterion on the items included in the case.

Similarly, we want to look at cases that do not contain the variable x, indicated by $\neg x$. This leads to the definition

$$\text{supp}(\neg x, y) = \text{supp}(y) - \text{supp}(x, y).$$

Then, the OR can be expressed as

$$\text{OR}(x) = \frac{ad}{bc} = \frac{\text{supp}(x, 1) \cdot \text{supp}(\neg x, 0)}{\text{supp}(\neg x, 1) \cdot \text{supp}(x, 0)},$$

using the outcome-based support function given above.

In a fictional example with 1000 patients and some drug A, imagine that 100 patients got drug A and 200 patients did not get drug A, after which they died during the stay. Additionally, 400 patients got drug A and 300 patients did not get drug A and did not die during the stay. Then, we have

$$\begin{aligned} a &= \text{supp}(\text{drug A}, 1) = 100, \\ b &= \text{supp}(\text{drug A}, 0) = 400, \\ c &= \text{supp}(\neg \text{ drug A}, 1) = 200, \\ d &= \text{supp}(\neg \text{ drug A}, 0) = 300, \end{aligned}$$

and

$$\text{OR}(\text{drug A}) = \frac{100 \cdot 300}{400 \cdot 200} = 0.375,$$

which indicates that drug A is negatively associated with mortality during the stay because the OR is lower than 1.0. This calculation allows us to assess the association between drug A and in-hospital mortality.

We use the same principle in this study to assess the association between various clinical variables and in-hospital mortality. Next to the prediction and the interpretability of the method, the goal of the present study is to analyze how such a rule-based model can be used to study which variables influence in-hospital mortality in a clinical dataset.

2. Materials and Methods

2.1. Data

This study uses data from the MIMIC-IV project, version 0.4 (Johnson et al., 2020). MIMIC-IV is a publicly available clinical dataset comprising around 525,000 ICU and ED cases collected between 2008 and 2019. Recorded variables include vital signs, textual notes, diagnoses, procedures, and various organizational information. We included all available cases in the study. The age of a patient is not used to filter the cases. This allows us to assess the in-hospital mortality risk for the most general and heterogenous population possible that can be expected in ICUs and EDs. This yields 524,520 cases that were used in this study. Out of these, 9369 (1.79%) were reported to have died during the hospital stay. While this is a considerable amount of cases, general clinical data also contains a lot of variables.

The types of variables used in the present study are listed in Table 1. This set of variables types includes variables that can be expected to exists in all ICUs and EDs. It does not include quickly changing information like vital signs or unstructured data like textual notes. Like diagnoses and procedures, some information might change throughout the stay but can be expected to do so slowly.

Table 1. The types of variables used in the present study as well as a description and the number of variables of this type in MIMIC-IV. ICD: International Classification of Diseases.

VARIABLE TYPE	DESCRIPTION	NUMBER OF VARIABLES
DIAGNOSIS	DIAGNOSES (ICD)	86,751
ETHNICITY	PATIENT'S ETHNICITY	8
GENDER	RECORDED AS BINARY: MALE/FEMALE	2
INSURANCE TYPE	RECORDED AS MEDICARE, MEDICAID, OR OTHER	3
LANGUAGE	RECORDED AS BINARY: ENGLISH/OTHER	2
MARITAL STATUS	RECORDED AS SINGLE, MARRIED, DIVORCED, WIDOWED, MISSING	5
PRESCRIPTION	PRESCRIBED DRUGS	10,259
PROCEDURE	PROCEDURES (ICD)	82,763
SERVICE	SERVICES: MEDICAL, PSYCHOLOGICAL ETC.	21
WARD	WARDS: EMERGENCY DEPARTMENT, SURGERY ETC.	43
OVERALL		179,857

Overall, 179,857 variables were included in the method. Each variable is treated the same, irrespective of its type. Additionally, in-hospital mortality is defined for each clinical case in MIMIC-IV. This variable is used as the outcome of interest. In comparison to other datasets, this is a vast number of variables, resulting in only around three cases per variable. Many commonly used software packages struggle with such sparse data, rendering the usage of popular algorithms like logistic regression and decision trees difficult. We circumvent this by applying ARM to the dataset, which considers all variables independently and based on commonly used metrics.

2.2. Method

The method is composed of three steps, which are shown in Figure 1. In the first step, the OR between each variable and in-hospital mortality is calculated. ORs which take the value 0 or $+\infty$ are discarded. This corresponds to at least one of the values included in the OR being 0. In the second step, these odds ratios are filter based on a statistical significance test, resulting in a prediction model composed of rules with a statistically significant odds ratio. This prediction model is used in the third step by aggregating the odds ratios of all rules that apply to the given case.

In the first step, the OR of each variable with in-hospital mortality is calculated as explained above. After the dataset has been fully searched, the second step consists of applying a statistical test to each rule to determine if it is statistically significant. The normal approximation of the log odds ratio (Morris & Gardner, 1988) is used to test for statistical significance, analyzing the calculated OR against the null hypothesis "$OR = 1.0$". The returned two-sided p-value is compared to a configurable value p_{max} and only rules with $p < p_{\text{max}}$ are kept. A method to correct for multiple hypothesis testing may be used when configuring p_{max} to ensure valid results despite the high number of statistical tests. The user can thus choose whether to use Bonferroni correction

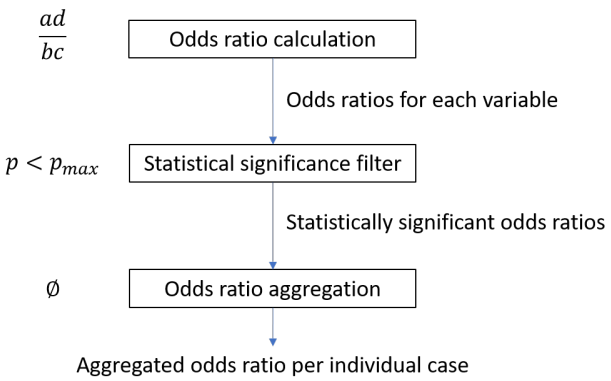

Figure 1. A visual representation of the proposed method.It contains three steps, namely the calculation of the odds ratio between each variable and in-hospital mortality, a filter for statistical significance of the calculated odds ratio, and an aggregation of all odds ratio that apply to a given case.

(Shaffer, 1995). If it is enabled, the user-chosen p-value is changed before the statistical tests are executed. If p is the original p-value, the new p-value is chosen to be p divided by the number of tests to be executed. This statistical significance test (with or without Bonferroni correction) can filter out rules for which the test deemed that there is not enough evidence available to determine that "$OR \neq 1.0$" and that the OR being different from 1.0 could therefore have been observed due to random chance.

2.3. Risk estimation

These calculations result in a set of rules of the form "variable $\Rightarrow$ in-hospital mortality," each with an OR that measures the strength as well as the direction of the association between the variable and in-hospital mortality. The third and last step of the method consists of the actual predic-

tion of in-hospital mortality for an individual case. Given a new clinical case, the algorithm estimates the risk as follows. First, the method checks which rules apply to the case, i.e., for which rules the corresponding variable occurred in the case. Second, the average OR of all these applying rules is taken. This average OR is then used as a mortality risk score. A cutoff value is needed to transform this score into a binary prediction (low risk/high risk). The cutoff value depends on the users' needs and has to be determined empirically; see below.

2.4. Experiments

Ten-fold cross-validation was used. The hyperparameters were configured as follows. The use of Bonferroni correction could be turned on ("yes" in the plots below) and off ("no"). The maximal allowed p-value threshold was set to 10^{-n} for $n \in \{0, \ldots, 10\}$ as well as the commonly used threshold 0.05. The metric of interest was the area under the receiver operating characteristic curve (AUC). The experiments resulted in $12 * 2 * 10 = 240$ AUC values.

3. Results

Figure 2 shows the resulting AUCs. Overall, the predictive performance was very good, with AUCs around 0.98. Even though there is some variation in the average AUCs, the different configurations do not impact the predictive qualities of the model.

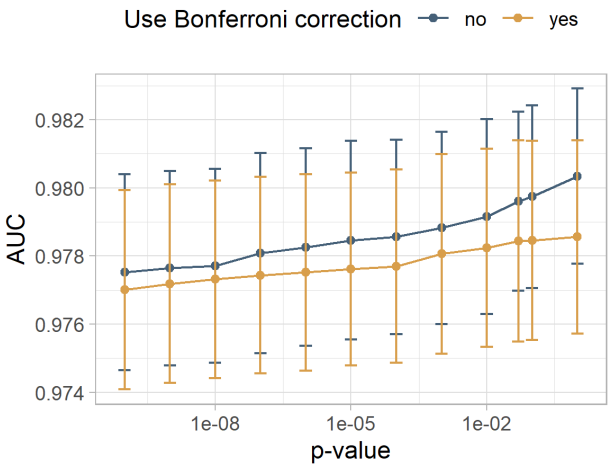

Figure 2. The AUC values returned by the experiments are shown on the y-axis. The x-axis is the p-value threshold used to filter the rules, the color indicates whether Bonferroni correction was used in the experiments. The error bars denote plus/minus one standard deviation. Higher p-values lead to slightly increased AUC values.

The configured p-value threshold and the use of Bonferroni correction have a major impact on the number of rules in the model, as shown in Figure 3. The number of rules increases approximately exponentially with increasing p-value threshold. The use of Bonferroni correction slows this increase down. The standard deviation of the number of rules is small, which indicates that the model size is stable in different folds of the dataset. In summary, statistical significance tests can simplify the model considerably without compromising the predictive performance.

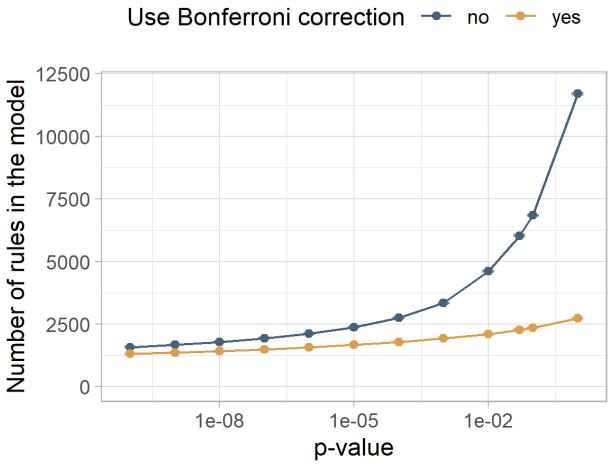

Figure 3. The y-axis shows the number of rules in the model. The x-axis is the p-value threshold used to filter the rules, the color indicates whether Bonferroni correction was used in the experiments. The error bars (which are barely visible) denote plus/minus one standard deviation. The number of rules increases exponentially with the p-value.

The high number of rules results from the complex and highly heterogeneous scenario with many variables for which this model is built. This can be seen in Figure 4. Even though there are thousands of rules in the model, only around 23 to 36 rules apply to each case, dependent on the hyperparameters. This makes the model easily understandable and interpretable for each case.

We further analyze one model that was trained with Bonferroni correction and a maximal p-value of 10^{-10}. As the AUC is only weakly affected by the hyperparameters while the number of rules decreases quickly for smaller p-value thresholds, this configuration was chosen to achieve a model with a good predictive performance and as few rules as possible. This model achieved an AUC of 0.98 with 1302 rules. The corresponding receiver operating characteristic curve (ROC) can be seen in Figure 5. It shows the sensitivity and specificity for different risk score cutoff values.

The choice of a cutoff value depends on the usage context. One common method to choose a cutoff values is by Youden

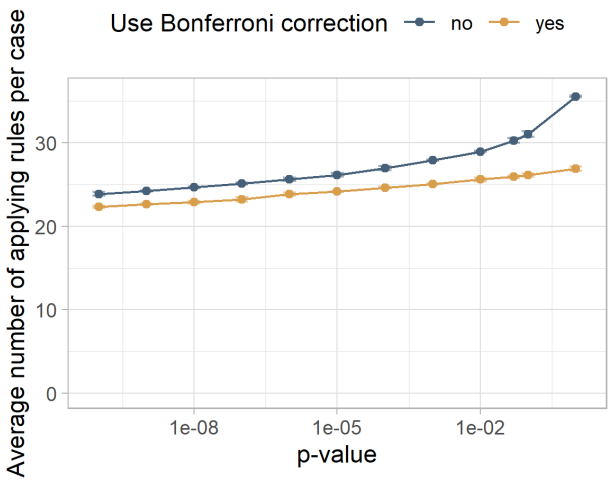

Figure 4. The average number of rules that apply to one clinical case. The x-axis is the p-value threshold used to filter the rules, the color indicates whether Bonferroni correction was used in the experiments. The error bars (which are barely visible) denote plus/minus one standard deviation. The average number of rules increases only slightly with the p-value. The standard deviations are very low, indicating that most cases have a similar number of rules that apply.

Table 2. The four-fold table between in-hospital mortality and the predicted mortality risk with the cutoff value 5.31, which minimizes the Youden index.

	DIED	SURVIVED
HIGH RISK	929	3,345
LOW RISK	54	48,036

index, which is calculated as sensitivity + specificity − 1 (Youden, 1950). The maximal Youden index was 0.87 for the cutoff value 5.31. This cutoff value resulted in a sensitivity of 95% and specificity of 93%. The corresponding four-fold table can be seen in Table 2.

Table 3 shows the types of variables in the model. Out of the 179,857 variables in MIMIC-IV, 1302 were included in the model due to their association with in-hospital mortality. Out of these, 1260 denoted positive associations (i.e., OR >1, the variable's occurrence often co-occurs with in-hospital mortality), and 42 denoted negative associations (i.e., OR <1, the variable's occurrence rarely co-occurs with in-hospital mortality).

The rules are heterogeneous. Every type of variable included in the study occurred in the final rule set. This shows the heterogeneous factors that are associated with mortality in ICUs and EDs.

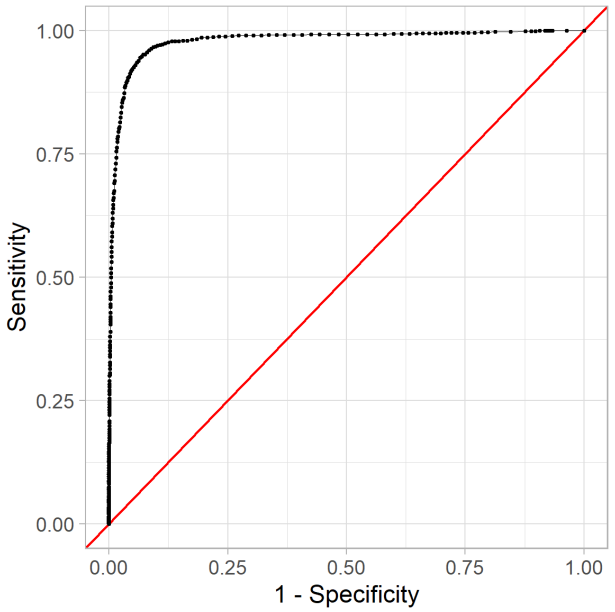

Figure 5. The receiver operating characteristic curve of one of the resulting models. The model contains 1302 rules. The x-axis shows one minus the specificity, the y-axis shows the sensitivity for different threshold values, which are shown as dots. The red line indicates the curve of a random coin toss. The ROC is very close to the top left corner, indicating a high predictive performance.

Two of the top five rules, namely the ones with brain deaths, highlight two curiosities of the dataset. First, the ICD (International Classification of Diseases) coding scheme was updated from version 9 to version 10 during the data collection. Second, both variables for brain death (ICD-9 and ICD-10) appear in the top five positive rules. This is surprising, as one would assume that brain death always co-occurs with in-hospital mortality, which would lead to an OR of $+\infty$. As this value is not meaningful for our purposes, it would have been discarded. While the ORs are very high, they are finite, indicating that there were cases in which brain death was diagnosed, but the patient survived the case. These cases were organ donors. After their death, they were recorded as being re-admitted to the hospital for the donation, and, according to the data, the patient survived one of the cases. Hence, these two rules are the result of a rare documentation error. This shows how the interpretable nature of the association rules used for prediction also allow us to identify patterns in the dataset.

The rules describe differences in in-hospital mortality between different patient groups in the dataset. We give some examples of rules of different types. Males had higher mortality odds than females (OR 1.25 vs. 0.80). The effect of marital status (single: 0.50; widowed: 2.05) could

Table 3. The types of variables contained in the model as well as the number of overall, positive (OR >1) and negative (OR <1) rules.

VARIABLE TYPE	VARIABLES IN MIMIC-IV	OVERALL RULES	POSITIVE RULES	NEGATIVE RULES
DIAGNOSIS	86,751	656	650	6
ETHNICITY	8	4	2	2
GENDER	2	2	1	1
INSURANCE TYPE	3	3	1	3
LANGUAGE	2	2	1	1
MARITAL STATUS	5	2	1	1
PRESCRIPTION	10,259	461	442	19
PROCEDURE	82,763	144	144	0
SERVICE	21	9	6	3
WARD	43	19	12	7
OVERALL	179,857	1302	1260	42

Table 4. The rules with the five highest and lowest odds ratios (ORs) in the model. Diagnosis codes are given as International Classification of Diseases code, in the format version:code.

VARIABLE	OR
PRESCRIPTION: MORPHINE INFUSION	856.26
DIAGNOSIS: BRAIN DEATH, 10:G93.82	668.41
DIAGNOSIS: SUBDURAL HEMATOMA, 9:852.25	582.14
DIAGNOSIS: BRAIN DEATH, 9:348.82	413.38
PRESCRIPTION: ANGIOTENSIN II	332.77
...	...
PRESCRIPTION: TRIPLE DYE	0.02
WARD: OBSTETRICS	0.02
SERVICE: OBSTETRICS	0.01
PRESCRIPTION: HEPATITIS B VACCINE	0.01
PRESCRIPTION: PHYTONADIONE	0.00

have been confounded by age, as could the effect of insurance type (Medicare: 2.40; Medicaid: 0.57; other: 0.50). English-speaking patients were more likely to survive (0.73 vs. 1.36). The ethnicity categories of hispanic/latino (0.47) and black/african american (0.58) were negatively associated with in-hospital mortality. The other retained ethnicity rules indicate missing values (unknown: 4.59; unable to obtain: 2.25). These rules might hint at critically ill patients who were unconscious at the time of admission. The rules of the types service and ward are similar to each other. One group of rules describes childbirth with a very low mortality risk with the services "Newborn" (0.17) and "Obstetrics" (0.01) and the wards Obstetrics (0.02), Nursery (0.02), and Labor & Delivery (0.04). As is to be executed, various ICU wards show higher ORs (cardiac vascular 2.80; trauma-surgical 7.55; surgical 8.68; medical/surgical 10.42; neuro-surgical 10.80; medical 13.25). In comparison, EDs are much less associated with in-hospital mortality (ED 1.90; ED observation 0.17). All 144 rules of type procedure were positive. Of these, 81 were coded with ICD-9 and 63 with ICD-10. In the ICD-9 rules, common patterns include catheterizations (eight rules, OR 2.59-21.11), biopsies (four rules, OR 4.07-11.46), and infusions (seven rules, OR 1.91-25.68), while common patterns in the ICD-10 rules are drainages (18 rules, OR 5.04-22.06) and insertions of feeding, infusion, monitoring, and other devices (14 rules, OR 3.69-73.80). Diagnoses are heterogeneous and describe different conditions, with some doubling occurring due to two ICD versions being used in the dataset. Out of the 656 rules, there are only six negative associations, with three being about single lifeborns (OR 0.08-0.16). The other rules are encounters for immunizations (ICD-10 Z23, OR 0.06), suicidal ideation (ICD-9 V62.84, OR 0.07), and unspecified chest pain (ICD-9 Z86.50, OR 0.11). Prescription rules are also varied. Notable groups include Heparin (14 rules, OR 2.11-68.58), sodium chloride (19 rules, OR 1.98-94.94), and lidocaine (eleven rules, OR 2.10-15.55). Out of the 461 rules, 19 were negative associations.

The set of all rules can be found in the supplementary materials.

4. Discussion

Associative classification in combination with ORs has proven useful for in-hospital mortality prediction with very high AUCs. Statistical significance tests have greatly reduced the number of rules in the model while keeping the predictive performance high. This high performance shows that the proposed method can cope with a high number of variables like the almost 180,000 included in the study.

To highlight the model's interpretability, we give a fictional example of how it assesses the mortality risk and how this reasoning can be presented to users.

Assume we have a patient who has been in the ICU for two hours. The rules in our model have been compared to the patient's electronic health records. All nine rules that apply to the case are shown to providers as in Table 5. These rules allow us to easily understand what this case is about: it is a male patient with acidosis and an-/oliguria treated with

Table 5. All the rules that apply to our fictional example case. The codes for diagnoses are from the International Classification of Diseases, version 10.

VARIABLE TYPE	VARIABLE	OR
DIAGNOSIS	ACIDOSIS, E87.2	10.96
DIAGNOSIS	AN-/OLIGURIA, R34	17.27
GENDER	MALE	1.25
INSURANCE TYPE	OTHER	0.50
LANGUAGE	OTHER	1.36
MARITAL STATUS	SINGLE	0.50
PRESCRIPTION	FUROSEMIDE	4.96
PRESCRIPTION	SODIUM BICARBONATE	11.54
WARD	EMERGENCY DEPARTMENT	1.90

Furosemide and Sodium Bicarbonate, respectively. The average OR of these nine rules is 5.58. This is larger than the cutoff value of 5.31 mentioned above, which indicates that this patient (currently) has a high in-hospital mortality risk. Note that additional variables that are added at a later point in time might change the risk score.

The objective in-hospital mortality risk assessment can be used in different clinical processes. Examples include staff planning (as more high-risk patients mean more care is needed), benchmarking and statistics, resource allocation (to provide every patient as much care as they currently need), and decision-making in individual cases, e.g. if palliative care is to be started.

Apart from the simple algorithm and the widely known metric used in the rules, the interpretability is further improved by associating every variable by itself with in-hospital mortality. Unlike decision trees, which form more complex rules, this allows providers to assess each variable's effect on the prediction in isolation. This can be seen in the Results section. The simple form of the risk estimation model allows us to use it to analyze the underlying data set and reveal rules that are relevant to in-hospital mortality.

While the method was created for and tested with ICU and ED data, it is also applicable to general clinical contexts. The variables in Table 1 are very common and should be available in most clinical contexts. The rules and the performance of the model are dependent on the scenario and the quality of the dataset, but the method itself is usable in general clinical contexts. Similarly, it is also applicable to variables other than in-hospital mortality after learning appropriate rules. In-hospital mortality is an example of a clinical outcome that is influenced by many different variables. This property also holds for other clinical variables, including various diagnoses.

Despite the high predictive performance and interpretability of the model, the present study has some major limitations. First, we assume that all the information about the case is readily available. As the case could span over a significant amount of time, this is not the case in realistic scenarios. Instead, one would assume that some basic information (including gender, ethnicity, marital status, language, insurance type, service, ward, and first diagnosis) is available, with other information being added as soon as the corresponding variables occur in the case. Future research is needed to analyze how a case's risk score evolves as more and more information is added to the case.

Second, we only incorporate categorical data in our variable selection process. This is in part due to the nature of the underlying association rules. Quantitative values could be used after binning them, i.e., forming bins of similar values and using those bins as categorical values. For the use case at hand, this was not needed, as the model showed very high performance using only categorical data. Future research should focus on how quantitative values can be included in the model and how they affect the prediction of in-hospital mortality.

5. Conclusion

We developed and evaluated a novel risk estimation method for in-hospital mortality in the presence of thousands of variables. Associative classification in combination with ORs is a feasible method for in-hospital mortality risk estimation. The learned model provides a very high predictive performance. Due to it being built of easy-to-interpret rules, the model itself can easily be interpreted by providers and used to analyze the underlying dataset.

Acknowledgements

The authors wish to thank the reviewers for their valuable comments.

This project is funded by the Bavarian State Ministry of Science and the Arts and coordinated by the Bavarian Research Institute for Digital Transformation (bidt) and supported by the Bavarian Academic Forum (BayWISS) – Doctoral Consortium "Health Research".

References

Cheng, C. W. and Wang, M. D. Improving Personalized Clinical Risk Prediction Based on Causality-Based Association Rules. *ACM BCB*, 2015:386–392, Sep 2015. doi: 10.1145/2808719.2808759.

Dua, S., Singh, H., and Thompson, H. Associative classification of mammograms using weighted rules. *Expert Systems with Applications*, 36(5):9250–9259, 2009. doi: 10.1016/j.eswa.2008.12.050.

El Houby, E. M. A framework for prediction of response

to HCV therapy using different data mining techniques. *Adv Bioinformatics*, 2014:e181056, 2014. doi: 10.1155/2014/181056.

ElHefnawi, M. and Sherif, F. F. Accurate classification and hemagglutinin amino acid signatures for influenza A virus host-origin association and subtyping. *Virology*, 449:328–338, 2014. doi: 10.1016/j.virol.2013.11.010.

Fu, L.-H., Schwartz, J., Moy, A., Knaplund, C., Kang, M.-J., Schnock, K. O., Garcia, J. P., Jia, H., Dykes, P. C., Cato, K., Albers, D., and Rossetti, S. C. Development and validation of early warning score system: A systematic literature review. *Journal of Biomedical Informatics*, 105: e103410, 2020. doi: 10.1016/j.jbi.2020.103410.

He, Y. and Hui, S. C. Exploring ant-based algorithms for gene expression data analysis. *Artif Intell Med*, 47(2): 105–119, Oct 2009. doi: 10.1016/j.artmed.2009.03.004.

Johnson, A., Bulgarelli, L., Pollard, T., Horng, S., Celi, L. A., and Mark, R. MIMIC-IV (version 0.4). *PhysioNet*, 2020. doi: 10.13026/A3WN-HQ05.

Johnson, A. E. W., Pollard, T. J., and Mark, R. G. Reproducibility in critical care: a mortality prediction case study. In Doshi-Velez, F., Fackler, J., Kale, D., Ranganath, R., Wallace, B., and Wiens, J. (eds.), *Proceedings of the 2nd Machine Learning for Healthcare Conference*, volume 68 of *Proceedings of Machine Learning Research*, pp. 361–376, Boston, Massachusetts, 18–19 Aug 2017. PMLR.

Kadkhoda, M., Akbarzadeh-T., M. R., and Sabahi, F. FLeAC: A Human-Centered Associative Classifier Using the Validity Concept. *IEEE Transactions on Cybernetics*, pp. 1–12, 2020. doi: 10.1109/TCYB.2020.3025479.

Keuning, B. E., Kaufmann, T., Wiersema, R., Granholm, A., Pettilä, V., Møller, M. H., Christiansen, C. F., Castela Forte, J., Snieder, H., Keus, F., Pleijhuis, R. G., van der Horst, I. C. C., and consortium, H. Mortality prediction models in the adult critically ill: A scoping review. *Acta Anaesthesiologica Scandinavica*, 64(4):424–442, 2020. doi: 10.1111/aas.13527.

Kianmehr, K. and Alhajj, R. CARSVM: A class association rule-based classification framework and its application to gene expression data. *Artificial Intelligence in Medicine*, 44(1):7–25, 2008. doi: 10.1016/j.artmed.2008.05.002.

Lin, W. Y., Li, H. Y., Du, J. W., Feng, W. Y., Lo, C. F., and Soo, V. W. iADRs: towards online adverse drug reaction analysis. *Springerplus*, 1(1):e72, Dec 2012. doi: 10.1186/2193-1801-1-72.

Meena, K., Tayal, D. K., Gupta, V., and Fatima, A. Using classification techniques for statistical analysis of Anemia. *Artificial Intelligence in Medicine*, 94:138–152, 2019. doi: /10.1016/j.artmed.2019.02.005.

Morris, J. A. and Gardner, M. J. Calculating confidence intervals for relative risks (odds ratios) and standardised ratios and rates. *Br Med J (Clin Res Ed)*, 296(6632):1313–1316, May 1988. doi: 10.1136/bmj.296.6632.1313.

Rea, S. and Huff, S. Cohort Amplification: An Associative Classification Framework for Identification of Disease Cohorts in the Electronic Health Record. *AMIA Annual Symposium proceedings*, 2010:862–6, 11 2010.

Salluh, J. I. and Soares, M. ICU severity of illness scores: APACHE, SAPS and MPM. *Curr Opin Crit Care*, 20(5):557–565, Oct 2014. doi: 10.1097/MCC.0000000000000135.

Shaffer, J. P. Multiple Hypothesis Testing. *Annual Review of Psychology*, 46(1):561–584, 1995. ISSN 0066-4308. doi: 10.1146/annurev.ps.46.020195.003021.

Thabtah, F. A review of associative classification mining. *Knowledge Eng. Review*, 22:37–65, 03 2007. doi: 10.1017/S0269888907001026.

Ujager, F. S. and Mahmood, A. A Context-Aware Accurate Wellness Determination (CAAWD) Model for Elderly People Using Lazy Associative Classification. *Sensors (Basel)*, 19(7):e1613, Apr 2019. doi: 10.3390/s19071613.

Uriarte-Arcia, A. V., López-Yáñez, I., and Yáñez-Márquez, C. One-Hot Vector Hybrid Associative Classifier for Medical Data Classification. *PLOS ONE*, 9(4):1–13, 04 2014. doi: 10.1371/journal.pone.0095715.

Xie, J., Su, B., Li, C., Lin, K., Li, H., Hu, Y., and Kong, G. A review of modeling methods for predicting in-hospital mortality of patients in intensive care unit. *Journal of Emergency and Critical Care Medicine*, 1(8):e18, 2017. doi: 10.21037/jeccm.2017.08.03.

Youden, W. J. Index for rating diagnostic tests. *Cancer*, 3 (1):32–35, 1950. doi: 10.1002/1097-0142(1950)3:1⟨32::AID-CNCR2820030106⟩3.0.CO;2-3.

Yu, P. and Wild, D. J. Fast rule-based bioactivity prediction using associative classification mining. *J Cheminform*, 4 (1):e29, Nov 2012a. doi: 10.1186/1758-2946-4-29.

Yu, P. and Wild, D. J. Discovering Associations in Biomedical Datasets by Link-based Associative Classifier (LAC). *PLOS ONE*, 7(12):1–11, 12 2012b. doi: 10.1371/journal.pone.0051018.