

Chapter 22

Data Fusion Techniques for Early Warning of Clinical Deterioration

Peter H. Charlton, Marco Pimentel and Sharukh Lokhandwala

Learning Objectives

Design and evaluate early warning score (EWS) algorithms which fuse vital signs with additional physiological parameters commonly available in hospital electronic health records (EHRs).

1. Extract physiological, demographic and biochemical variables from the MIMIC II database.
2. Extract patient outcomes from the MIMIC II database.
3. Prepare EHR data for analysis in Matlab[®].
4. Design data fusion algorithms in Matlab[®].
5. Compare the performances of data fusion algorithms.

22.1 Introduction

Acutely-ill hospitalized patients are at risk of clinical deteriorations such as infection, congestive heart failure and cardiac arrest [1]. The early detection and management of such deteriorations can improve patient outcomes, and reduce healthcare resource utilization [2, 3]. Currently, early warning scores (EWSs) are used to assist in the identification of deteriorating patients. EWSs were designed for use at the bedside: they can be calculated by hand, and the required inputs (vital signs) can be easily measured at the bedside. Now that EHRs are becoming more widespread in acute hospital care there is scope to develop improved EWSs by using more complex algorithms calculated by computer, and by incorporating additional physiological data from the EHR.

Most methods for detection of deteriorations are based on the assumption that changes in physiology are manifested during the early stages of deteriorations. This assumption is well documented. Schein et al. published landmark results in 1990

that 84 % of patients “had documented observations of clinical deterioration or new complaints” in the eight hours preceding cardiac arrest [4]. This was further supported by a study by Franklin et al. [5]. Physiological abnormalities have also been observed prior to other deteriorations such as unplanned Intensive Care Unit (ICU) admissions [6] and preventable deaths [7]. Evidence of deterioration can be observed 8–12 h before major events [8, 9].

It was proposed that the incidence of deteriorations could be reduced by recognising and responding to early changes in physiology [10–12]. Subsequently, EWSs were developed to allow timely recognition of patients at risk of deterioration. EWSs are aggregate scores calculated from a set of routinely and frequently measured physiological parameters, known as vital signs. The higher the score, the more abnormal the patient’s physiology, and the higher the risk of future deterioration. EWSs are now in widespread use in acute hospital wards [13].

Current EWSs correlate with important patient-centered endpoints such as levels of intervention [14], hospital mortality [14, 15], and length of stay [15], and have been shown to be a better predictor of cardiac arrest than individual parameters [16]. However, there is scope for improving their performance since most EWSs use simple formulae which can be calculated by hand at the bedside, and use only a limited set of vital signs as inputs [17]. Now that electronic health records (EHRs) are becoming widely used in acute hospital care, there is opportunity to use more complex, automated algorithms and a broader range of inputs. Consequently, algorithms have been proposed in the literature which improve performance by using data fusion techniques to combine vital signs with other parameters such as biochemistry and demographic data [18, 19].

The remainder of this chapter is designed to equip the reader with the necessary tools to develop and evaluate data fusion algorithms for prediction of clinical deteriorations.

22.2 Study Dataset

Data was extracted from the MIMIC II database (v. 2.26) [21], which is publicly available on PhysioNet [22]. This database was chosen because it contains routinely recorded EHR data for thousands of patients who, being critically-ill, are at high risk of deterioration. Data extraction was performed using the three SQL queries `cohort_labs.sql`, `cohort_vitals.sql`, and `cohort_selection.sql`. For ease of analysis data were extracted from only 500 patients. Only adult data were extracted since paediatrics have different normal physiological ranges to those of adults. The parameters extracted from the database, listed in Table 22.1, were chosen in line with those used previously in the literature [18, 19].

Traditionally the performance of EWSs has been assessed using three outcome measures with which rapid response systems have been assessed: mortality, cardiopulmonary arrest and ICU admission rates [20]. However, cardiopulmonary arrests are difficult to reliably identify in the MIMIC II dataset, and the dataset only

Table 22.1 EHR Parameters extracted from the MIMIC II database records for input into data fusion algorithms

Biochemistry	Vital signs
Albumin	Respiratory rate
Anion gap	Heart rate
Arterial pCO ₂	Blood pressure—systolic and diastolic
Arterial pH	Temperature
Aspartate aminotransferase (AST)	Oxygen saturation
Bicarbonate	Level of consciousness
Blood urea nitrogen (BUN)	Demographics
Calcium	Age
Creatinine	Gender
Glucose	
Hemoglobin	
Platelets	
Potassium	
Sodium	
Total bilirubin	
White blood cell count (WBC)	

contains data from patients already staying on the ICU. Therefore, mortality, which can be reliably and easily extracted from the dataset, was chosen as the outcome measure for this case study.

22.3 Pre-processing

Data analysis was conducted in Matlab[®]. The first pre-processing step was to import the CSV files generated by the SQL query into Matlab[®] (using LoadData.m). The purpose of this step was to create:

1. A design matrix of predictor variables (the parameters listed in Table 22.1): This MxN matrix contained values for each of the N parameters at each of M time points. This was performed using the methodology in [19]: the time-points were calculated as the end times of successive four-hour periods spanning each patient’s ICU stay; parameter values at the time-points were set to the last measured value during that time period.
2. An Mx3 response matrix of the three easily acquired dependent variables, namely, binary variables of death in ICU and death in ICU within the next 24 h, and a continuous variable of time to ICU death.

The remaining pre-processing steps and analyses were conducted using only data from within these matrices.

Further pre-processing was required to prepare the data for analysis (PreProcessing.m). Firstly, it was observed that the temperature values exhibited a bimodal distribution centred on 37.1 and 98.8 °C, indicating that some had been measured in Celsius, and others in Fahrenheit. Those measured in

Fahrenheit were converted to Celcius. Secondly, the dataset contained blood pressures (BPs) acquired invasively and non-invasively. Invasive measurements were retained since they had been acquired more frequently. Non-invasive measurements were replaced with surrogate invasive values by correcting for the observed biases between the two measurement techniques when both had been used in the same four-hour periods (the median differences between invasive and non-invasive measurements were 2, 7 and 6 mmHg for systolic, diastolic and mean BPs respectively). Finally, the dataset contained missing values where parameters had not been measured within particular four-hour periods. These missing data had to be imputed since the analysis technique to be used, logistic regression, requires a complete data set. To do so, we followed the approach proposed previously of imputing the last measured value, unless no value had yet been measured in which case the population median value was imputed [19]. Note that this approach could be applied to a dataset in real-time.

22.4 Methods

Novel data fusion algorithms were created using `CreateDataFusionAlgs.m`. Generalized linear models were used to fuse both continuous and binary variables to provide an output indicative of the patient’s risk of deterioration. A training dataset, containing 50 % of the data, was used to create the algorithms.

Logistic regression was used to estimate the probability of each of the binary response variables of “death in ICU”, and “death in ICU within 24 h” being true. Logistic regression differs from ordinary linear regression in that it bounds the output to be between 0 and 1, thus making it suitable for estimation of the probability of a response variable being true. Logistic regression provides an estimate for

$$y = \ln \left[\frac{p(x)}{1 - p(x)} \right]$$

where $p(x)$ is the probability of the response variable being true and x is a vector of predictor variables. Notice that $p(x)$ is constrained to be between 0 and 1 for all real values of y .

When using logistic regression one must decide how to model the relationships between the n predictor variables contained within x , and the output, y . The simplest method is to assume that y is linearly related to the predictor variables as $y = \alpha + \sum_{i=1}^n \beta_i x_i$, where α is the intercept term, and β is a vector of coefficients. For variables such as diastolic blood pressure the assumption of a linear relationship is reasonable because they consistently change in one particular direction during a deterioration. However, other variables such as sodium level could change in either

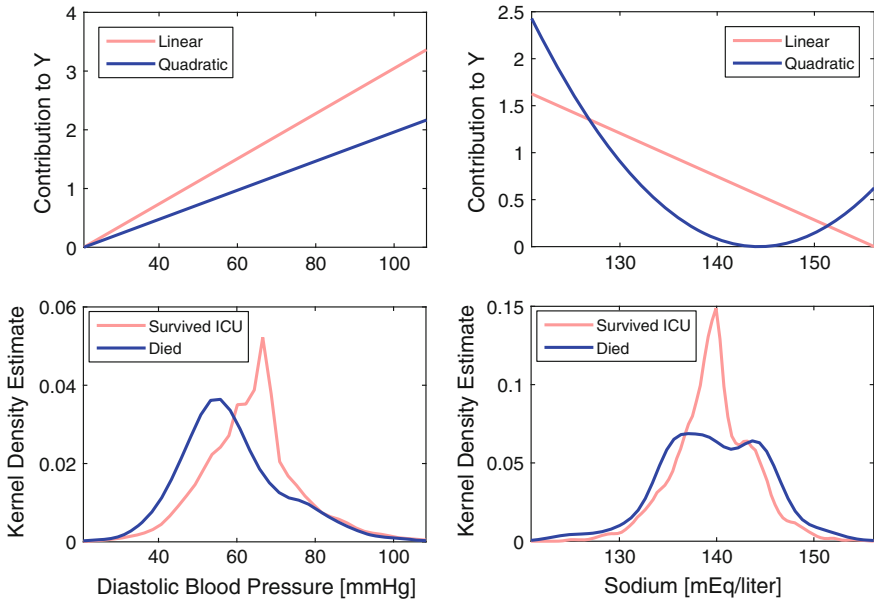


Fig. 22.1 A comparison of the contributions of input variables to the algorithm output, Y , under the assumptions of either a linear or a non-linear relationship between the input variables and Y . The choice of relationship had little impact on the contribution of Diastolic Blood Pressure (*above left*), since it tended to be reduced in those patients who died (*below left*). However, a quadratic relationship provided a very different contribution for Sodium Level (*above right*), since the Sodium Levels of those patients who died exhibited a bimodal distribution indicating either an increase or a decrease away from the normal range (*below right*)

direction away from normality. For these variables a non-linear relationship is more appropriate, such as the quadratic

$$y = \alpha + \sum_{i=1}^n \beta_i x_i + \sum_{i=1}^n \gamma_i x_i^2,$$

where γ is a vector of coefficients for the squares of the predictor variables. Note that this ‘purely quadratic’ relationship does not contain interaction terms such as $x_i x_j$. The importance of the choice of relationship between the predictor variables and the estimate is demonstrated in Fig. 22.1.

In this case study separate algorithms were created using linear and quadratic relationships. Firstly, only the parameters which are used in EWSs (vital signs) were included. Secondly, all the extracted EHR parameters were included. Thirdly, step-wise regression was used to avoid including terms which do not increase the performance of the model. This consisted of building a model by including terms until no further terms would increase the performance of the model, and then removing terms whose removal would not significantly decrease the performance of the model.

22.5 Analysis

EWS algorithms must trigger an effective clinical response in order to impact patient outcomes. Typically, a particular response is mandated when the algorithm's output is elevated above a threshold value. The response may include clinical review by ward staff or a centralised rapid response team. The following analysis is based on the assumption that the algorithms would be used to mandate responses such as this.

The performance of each algorithm was analysed using the latter 50 % of the data—the validation dataset. At all 4 h time points the model was used to estimate the probability of a patient dying during their ICU stay. Figure 22.2 shows exemplary plots of the output for four patients throughout their ICU stays. Throughout the analysis, each time point was classified as either positive or negative, indicating that the model predicted that the patient either subsequently died on ICU, or survived to ICU discharge. Hence, a true positive is identified at a particular time point when the model correctly predicts the death of a patient who died on ICU, whereas a false positive is identified when the model incorrectly predicts the death of a patient who survived to ICU discharge. True and false negatives were similarly identified.

Table 22.2 shows the performances of each algorithm assessed using the area under the receiver operating characteristic (ROC) curve (AUROC). The algorithm with the highest AUROC of 0.810 used stepwise inclusion of parameters and the quadratic relationship. The ROC curves for this algorithm and the corresponding algorithm using vital signs alone are shown in Fig. 22.3. Algorithms using all available parameters as inputs had higher AUROCs than those using vital signs

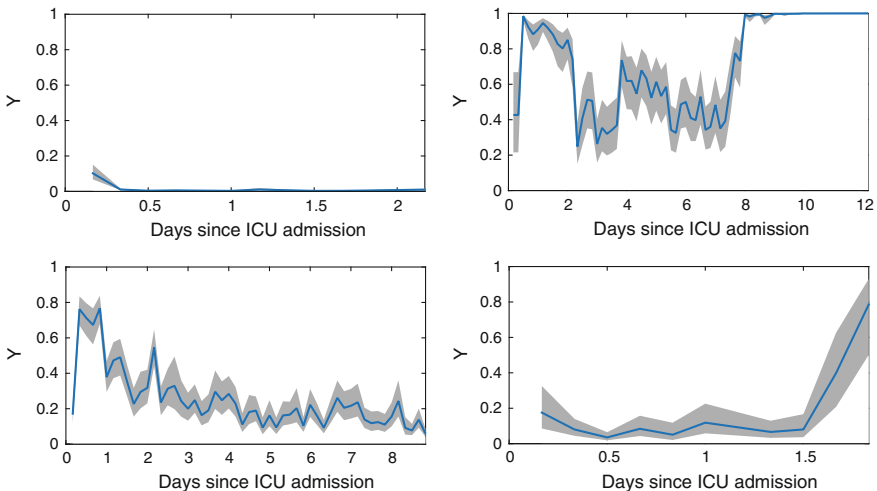
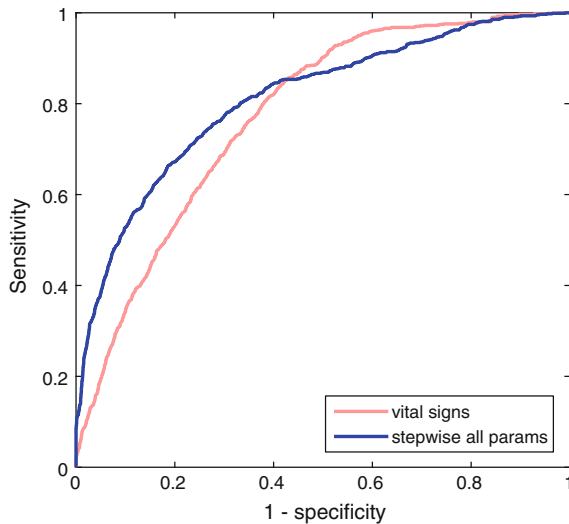


Fig. 22.2 Exemplary plots of the output of algorithm outputs (Y) over the duration of patients' ICU stays. The *left hand* plots show patients who survived their ICU stays, whereas the *right hand* plots show patients who died. The *upper plots* show examples in which the algorithm performed well, whereas the *lower plots* show examples in which the algorithm did not perform well

Table 22.2 The performances of data fusion algorithms for prediction of death in ICU, given as the area under the receiver-operator curve (AUROC), and the maximum sensitivities when the algorithms were constrained to satisfy the clinical requirements of a PPV ≥ 0.33 , and an alert rate of $\leq 17\%$

Relationship between predictor variables and output	Candidate predictor variables	Number of predictor variables included	AUROC	Maximum Sensitivities [%]	
				PPV ≥ 0.33	Alert rate $\leq 17\%$
Linear	Vital signs only	6	0.757	14.4	42.5
Linear	All	25	0.800	46.6	49.7
Linear	Stepwise inclusion of all	23	0.800	45.8	48.9
Purely quadratic	Vital signs only	6	0.774	13.2	41.4
Purely quadratic	All	25	0.799	55.5	53.9
Purely quadratic	Stepwise inclusion of all	21	0.810	59.3	56.3

Fig. 22.3 Receiver operating characteristic curves showing the performances of the best algorithms using stepwise inclusion of all parameters, and vital signs alone. These algorithms assumed a quadratic relationship between the predictor variables and the output



alone, demonstrating the benefit of fusing vital signs with additional parameters. In most instances the use of a quadratic relationship resulted in a higher AUROC. Furthermore, stepwise selection of parameters did reduce the number of parameters required, whilst maintaining or improving the AUROC.

Other metrics for comparison of algorithms have been suggested including sensitivity, positive predictive value (PPV) and alert rate [23]. However, these are more difficult to use since each metric varies according to the threshold value. A useful method for comparing algorithms using these metrics is to compare their sensitivities when a threshold is used which provides algorithmic performance in line with clinical requirements. In the case of EWS algorithms, key clinical requirements are that the PPV is at or above a minimum acceptable level, and the alert rate is at or below a maximum acceptable level. In the absence of evidence-based values, for demonstration purposes we used a minimally acceptable PPV of 0.33, indicating that one in three alerts is a true positive, and a maximally acceptable alert rate of 17 %, indicating that one in six observation sets results in an alert. Table 22.2 shows the sensitivities provided by each algorithm when constrained to satisfy these clinical requirements. The PPVs and alert rates at all thresholds are shown in Fig. 22.4 for the best performing algorithms using vital

Fig. 22.4 A comparison of the PPVs and alert rates for algorithms using vital signs alone and using all parameters. Exemplary clinical requirements of a $PPV \geq 0.33$ and an alert rate $\geq 17\%$ are shown by the dashed lines. The quadratic algorithm using vital signs alone has a much lower sensitivity of 13.2 % than the equivalent algorithm using stepwise inclusion of all parameters, at 59.3 % when the PPV criterion is met. Similarly, when the alert rate criterion is used, the sensitivity of the vital signs algorithm is 41.4 %, also lower than that of the algorithm using stepwise inclusion of all parameters, at 56.3 %

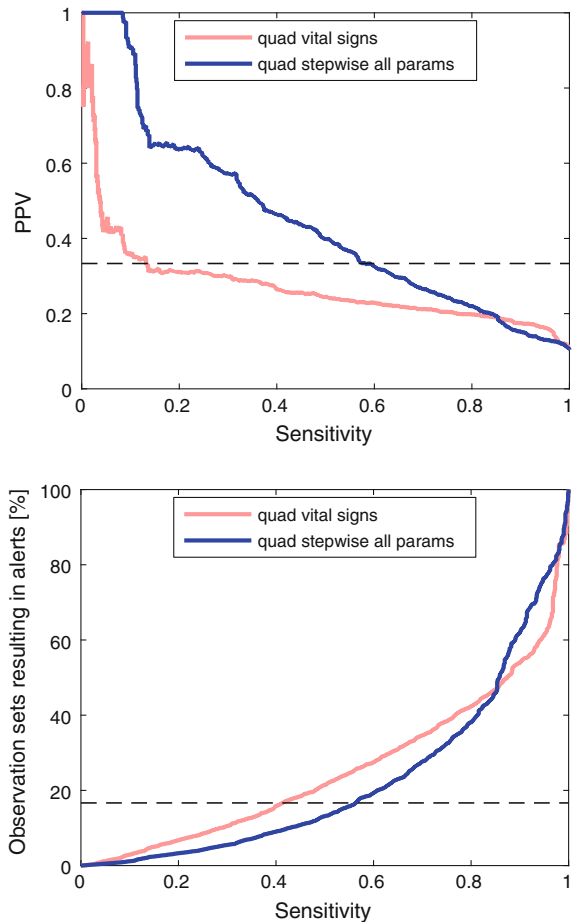


Fig. 22.5 Mean algorithm outputs during the 48 h prior to death on ICU (after exponential smoothing). A lower choice of threshold for alerting results in more advanced warning of deterioration

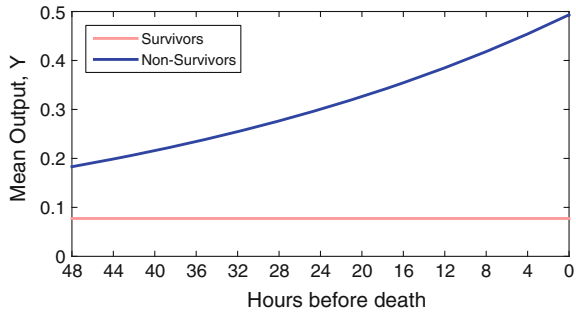
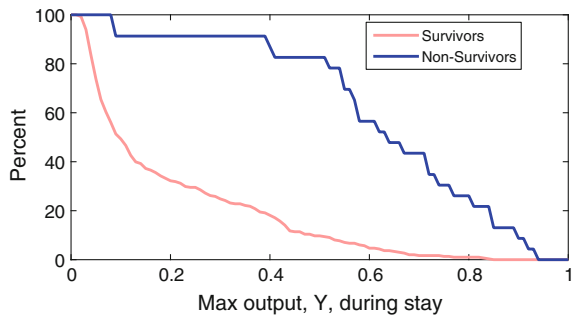


Fig. 22.6 The proportion of survivors and non-survivors who reached each algorithm output value during their ICU stay. A lower choice of threshold for alerting results in more false alerts, and fewer true alerts



signs alone and using stepwise inclusion of all parameters. The highest sensitivities were achieved when using stepwise inclusion of all parameters, with a purely quadratic relationship. The benefit of using additional parameters beyond vital signs is clearly shown by the algorithms’ sensitivities at the minimum acceptable PPV, which were 13.2 % when using vital signs alone, and 59.3 % when using stepwise inclusion of all parameters.

In [19] additional visualisations were used to demonstrate the effect of choosing different thresholds. Firstly, the dependent variable of time before death on ICU was used to examine how the output changed with time before death, as shown in Fig. 22.5. This shows that a lower threshold results in more advanced warning of deterioration. Secondly, the proportion of patients who reached each output during their stay was presented, as shown in Fig. 22.6. This suggests that a lower threshold results in more false alerts and fewer true alerts.

22.6 Discussion

The introduction of EHRs has provided opportunity to improve the clinical algorithms used to identify deteriorations. The data fusion algorithms described in this chapter estimate the probability of a patient dying during their ICU stay every 4 h.

The inclusion of additional physiological parameters beyond vital signs alone resulted in improvements in algorithm performance in this study when assessed using the AUROC, as also observed previously [18, 19], and when assessed using the minimum sensitivities corresponding to clinical requirements.

This case study has demonstrated the fundamental steps required to design and evaluate data fusion algorithms for prediction of deteriorations. During pre-processing the required data were extracted from the raw data files, and processed into matrices ready for analysis. It was important to perform this step separately to the analysis to reduce the time required for algorithm design. During this step we identified deficiencies in the dataset. Unfortunately, there is no systematic way to ensure that all deficiencies have been identified. We recommend that firstly the distributions of each variable are inspected to identify obvious discrepancies such as the different units used for temperature in this dataset. Secondly, it is helpful to plot the raw data over time to identify any changes in practice that may have occurred during data acquisition. Thirdly, it is often valuable to seek the guidance of a clinician or database curator at the host institution, or a researcher who has worked with the dataset before.

The results presented here cannot be generalised to a hospital-wide patient population for two reasons. Firstly, the dataset consists of data from critically-ill patients, whereas EWSs are primarily designed to identify deteriorations in acutely-ill patients. Since the disease processes of critically-ill patients are more advanced and they have additional clinical interventions such as mechanical ventilation and organ support, both the baseline physiology and the physiological changes accompanying deteriorations may differ in this population compared to acutely-ill patients. Secondly, death in ICU was used as the dependent variable in this study. Death is the latest possible stage of deterioration, and therefore an algorithm which predicts death may not predict the onset of deteriorations early enough to be of clinical utility in acutely-ill patients.

The choice of statistical methods to assess the performance of EWSs is the subject of debate [23]. The AUROC has often been used to quantify the performance of EWS algorithms, such as in [17]. This statistic is calculated from an algorithm's sensitivities and specificities at a range of threshold values. However, it has been recently suggested that the AUROC is misleading due to the low prevalence of deteriorations [23]. In [23] alternative statistical measures were proposed to account for the clinical requirements of EWS algorithms. Statistical measures should firstly assess the benefits and costs of using EWSs. The benefit is that EWSs can act as a safety net to catch deteriorating patients who have been missed in routine clinical assessments. This requires a high sensitivity (the proportion of EWS assessments of deteriorating patients which do alert). The cost of EWSs is the time taken to respond to false alerts. This cost is relatively small, since the additional clinical assessment triggered by an alert takes only a short amount of time. This means that a high specificity (the proportion of negative tests which are true negatives) is not of great importance. Secondly, it is important to ensure that the positive predictive value (the proportion of alerts which are true) is high enough to prevent caregivers suffering from desensitisation to alerts, which may result in less

effective responses to patients who are correctly identified as deteriorating [24]. Thirdly, the alert rate must be manageable to avoid excessive resource utilization. In this case study we presented the AUROC and the maximum sensitivities when algorithms were constrained to a minimally acceptable PPV and a maximally acceptable alert rate [23].

22.7 Conclusions

This case study has demonstrated the potential utility of data fusion techniques to predict clinical deteriorations. Currently identification of deteriorations is achieved using EWSs which take vital signs as inputs. The performance of the data fusion algorithms assessed in this study was improved by increasing the set of inputs to include physiological parameters which are routinely available in EHRs, but are not measured at the bedside.

The fundamental techniques for design and evaluation of data fusion algorithms have been demonstrated. Logistic regression algorithms were used to predict a binary response variable, death in ICU. The use of both linear and quadratic relationships between the predictor and response variables were demonstrated as well as the use of stepwise inclusion of variables. A range of statistical measures were presented for evaluation of algorithms, illustrating the benefits of using alternative statistical measures to the commonly used AUROC.

The results should not be interpreted as representative of the results that could be expected when EWSs are used in acute settings since the study dataset consists of critically-ill patients, and death in ICU was used as the dependent variable. However, the techniques used to design and evaluate algorithms can be easily applied to a wide range of patient settings, providing a basis for further work.

22.8 Further Work

Two particular areas have been identified for further research. Firstly, the work could be repeated using a dataset acquired from acutely-ill, rather than critically-ill patients, and by using a dependent variable other than death. This would facilitate design of algorithms that are generalisable to the target hospital population. Secondly, a range of additional functions could be explored to model the relationship between the predictor variables and the output. More complex functions than the linear or purely quadratic functions such as higher order polynomials or

logistic functions may improve performance. In addition it would be prudent to investigate the effect of the inclusion of interaction terms to account for the relationships between predictor variables.

22.9 Personalised Prediction of Deteriorations

The algorithms presented here are limited in scope by the input parameters. Currently they obtain a detailed description of a patient's physiological state from the vital signs and biochemistry values, which make up 23 out of the 25 inputs. However, these parameters provide very little differentiation between individual patients according to their state on admission to hospital. In contrast, additional information present upon hospital admission is used by clinicians during a patient's hospital stay to contextualise physiological assessments.

To illustrate this, consider the response of the algorithms to two fictional 65-year old males, patients A and B. Patient A has a history of hypertension, and a high systolic blood pressure (SBP) prior to hospital admission of 147 mmHg. Patient B has led an active life, has a healthy diet, and has a relatively low SBP prior to admission of 114 mmHg. During their hospital stay, the SBP of both patients is measured to be 114 mmHg. The algorithms cannot distinguish whether this is representative of patient A during a significant deterioration, such as the early stages of hypotension preceding septic shock, or whether it is representative of patient B's usual state in the absence of any deterioration. If the algorithms used a wider range of inputs indicative of patient state prior to admission, such as the presence or absence of co-morbidities (existing medical conditions) including hypertension, they might be able to differentiate between patients A and B in this situation.

This illustrates the potential benefit of incorporating additional inputs indicating co-morbidities. Even greater benefit may be derived by also personalising EWS algorithms according to physiological state prior to admission. Personalised EWS algorithms would not only stratify patients using additional inputs to contextualise physiology, but would also personalise the regression coefficients according to a patient's physiological state measured previously at a time of relative health.

Open Access This chapter is distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, duplication, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, a link is provided to the Creative Commons license and any changes made are indicated.

The images or other third party material in this chapter are included in the work's Creative Commons license, unless indicated otherwise in the credit line; if such material is not included in the work's Creative Commons license and the respective action is not permitted by statutory regulation, users will need to obtain permission from the license holder to duplicate, adapt or reproduce the material.

Code Appendix

The code used in this case study is available from the GitHub repository accompanying this book: <https://github.com/MIT-LCP/critical-data-book>. Further information on the code is available from this website. The following key scripts were used to extract data from the MIMIC II database:

- `cohort_selection.sql`: used to identify a cohort of patients for whom data would be extracted.
- `cohort_labs.sql`: used to extract laboratory test results.
- `cohort_vitals.sql`: used to extract vital signs.

Data was extracted in CSV format. Subsequent analysis was performed in Matlab[®] using `RunFusionAnalysis.m`. It contains the following script:

- `SetupUniversalParams`: used to set universal parameters (in this case, file paths), which are used to load and save files throughout the analysis). These parameters should be adapted when using the code.

It then called the following scripts:

- `LoadData.m`: used to load CSV data into Matlab[®] for analysis.
- `PreProcessing.m`: performs pre-processing to prepare data for analysis.
- `CreateDataFusionAlgs.m`: creates data fusion algorithms using training data.
- `AnalysePerformances.m`: analyses the performances of data fusion algorithms using validation data.

References

1. Silber JH et al (1995) Evaluation of the complication rate as a measure of quality of care in coronary artery bypass graft surgery. *JAMA* 274(4):317–323
2. Khan NA et al (2006) Association of postoperative complications with hospital costs and length of stay in a tertiary care center. *J Gen Intern Med* 21(2):177–180
3. Lagoe RJ et al (2011) Inpatient hospital complications and lengths of stay: a short report. *BMC Res Notes* 4(1):135
4. Schein RM et al (1990) Clinical antecedents to in-hospital cardiopulmonary arrest. *Chest* 98(6):1388–1392
5. Franklin C et al (1994) Developing strategies to prevent in-hospital cardiac arrest: analyzing responses of physicians and nurses in the hours before the event. *Crit Care Med* 22(2):244–247
6. Buist MD et al (1999) Recognising clinical instability in hospital patients before cardiac arrest or unplanned admission to intensive care. A pilot study in a tertiary-care hospital. *Med J Aust* 171(1):22–25
7. Hillman KM et al (2001) Antecedents to hospital deaths. *Intern Med J* 31(6):343–348
8. Hillman KM et al (2002) Duration of life-threatening antecedents prior to intensive care admission. *Intensive Care Med* 28(11):1629–1634

9. Whittington J et al (2007) Using an automated risk assessment report to identify patients at risk for clinical deterioration. *Jt Comm J Qual Patient Saf* 33(9):569–574
10. Smith AF et al (1998) Can some in-hospital cardio-respiratory arrests be prevented? A prospective survey. *Resuscitation* 37(3):133–137
11. Patient Safety Observatory (2007) Safer care for the acutely ill patient: learning from serious incidents. National Patient Safety Agency, London
12. Whittington J et al (2007) Using an automated risk assessment report to identify patients at risk for clinical deterioration. *Jt Comm J Qual Patient Saf* 33(9):569–574
13. Royal College of Physicians (2012) National early warning score (NEWS): standardising the assessment of acute-illness severity in the NHS”, Report of a working party. RCP, London
14. Goldhill DR et al (2005) A physiologically-based early warning score for ward patients: the association between score and outcome. *Anaesthesia* 60(6):547–553
15. Paterson R et al (2006) Prediction of in-hospital mortality and length of stay using an early warning scoring system: clinical audit. *Clin. Med.* 6(3):281–284
16. Churpek MM et al (2012) Predicting cardiac arrest on the wards: a nested case-control study. *Chest* 141(5):1170–1176
17. Smith GB et al (2013) The ability of the national early warning score (NEWS) to discriminate patients at risk of early cardiac arrest, unanticipated intensive care unit admission, and death. *Resuscitation* 84(4):465–470
18. Alvarez CA et al. (2013) Predicting out of intensive care unit cardiopulmonary arrest or death using electronic medical record data. *BMC Med Inform Decis Mak* 13(28)
19. Churpek MM et al (2014) Multicenter development and validation of a risk stratification tool for ward patients. *Am J Respir Crit Care Med* 190:649–655
20. Maharaj R et al (2015) Rapid response systems: a systematic review and meta-analysis. *Crit Care* 19(1):254
21. Saeed M et al (2011) Multiparameter intelligent monitoring in intensive care II: a public-access intensive care unit database. *Crit Care Med* 39(5):952–960
22. Goldberger AL et al (2000) PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *Circulation* 101(23):E215–E220
23. Romero-Brufau S, Huddleston JM, Escobar GJ, Liebow M (2015) Why the C-statistic is not informative to evaluate early warning scores and what metrics to use. *Crit Care* 19(1):285
24. Cvach M (2012) Monitor alarm fatigue: an integrative review. *Biomed Instrum Technol* 46(4):268–277