A Machine Learning Approach for the Classification of Disease Risks in Time Series

Lejla Begic Fazlic*, Ahmed Hallawa†, Matthias Dziubany*, Marlies Morgen*, Jens Schneider*, Marvin Schacht*, Anke Schmeink†, Lukas Martin‡, Arne Peine‡, Thomas Vollmer§ Stefan Winter§ and Guido Dartmann*
*ISS, Trier University of Applied Sciences, Trier, Germany
†ISEK Research and Teaching Area, RWTH Aachen University, Aachen, Germany
§Philips GmbH Innovative Technologies, Aachen, Germany
‡Department of Intensive and Intermediate Care, University Hospital Aachen, Aachen, Germany

***Abstract*—In this work, a new hybrid algorithm for disease risk classification is proposed. The proposed methodology is based on Dynamic Time Warping (DTW). This methodology can be applied to time series from various domains such as vital sign time series available in medical big data. To validate our methodology, we applied it to risk classification for sepsis, which is one of the most challenging problems within the area of medical data analysis. In the first step the algorithm uses different statistical properties of time series features. Furthermore, using differently labeled training data sets, we created a DTW Barycenter Averaging (DBA) on each feature. In the second step, validation data sets and DTW are used to validate the precision of classification and the final results are compared. The performance of our methodology is validated with real medical data and on six different criteria definitions for the sepsis diseases. Results show that our algorithm performed, in the best case, with precision and recall of 96,38% and 90,90%, respectively.**

***Index Terms*—Machine Learning, Time Series , Dynamic Time Warping, Data Mining, Sepsis**

I. Introduction

Time series data extracted from electronic health records play an important role for improving medical care. Using different statistical methods, these retrospective data have been used to understand the relationship between inputs and outcomes or to find similar patterns for specific patient groups. If an accurate diagnosis is provided in the right time, the appropriate treatment can be provided and the patient has the best chance for a positive health outcome. Since early treatment of sepsis increases the chance of positive outcomes, a rapid diagnosis is crucial.

As extracting and labeling sepsis data is not a trivial task, clinical risk prediction is very complex and depends on experience, how one chooses criteria, the time of prediction and the prediction horizon. In this research, we developed a novel methodology for disease risk classification using retrospective data [1], [2]. The algorithm is based on the principles of DBA, DTW and additional statistical methods. In the training phase, we merged all the patients' features data by creating DBA in a positive and negative sense. In the validation phase, we used validation data to validate the precision of the classification. If the sample "is positive and it is classified as positive, it is counted as a true positive (TP); if it is classified as negative, it is considered as a false negative (FN)" [3]. If the sample "is negative and it is classified as negative it is considered as true negative (TN); if it is classified as positive, it is counted as false positive (FP)" [3]. Recall of a classifier "represents the positive correctly classified samples to the total number of positive samples" [3]. Precision "represents the proportion of positive samples that were correctly classified to the total number of positive predicted samples" [3]. At the final step, all results by precision and recall for different labeled data are compared. We also show the high impact of the different formulation of disease criteria (differently labeled data) on performance.

The paper is organized as follows: In Section I, we describe related problems and the clinical challenge of sepsis identification. Furthermore, we review existing research studies on different approaches of combination of statistical and Machine Learning (ML) approaches. In Section III, we explain the algorithm design and methods used in our research, where we describe the proposed methodology including data acquisition and pre-processing. Section IV proposes our methodology, where we describe the algorithm design architecture. Section V illustrates and compares the numerical results for different training data according to the different criteria for the sepsis disease. Finally, a conclusion is given in Section VI.

II. Related Works

The mortality rates due to some diseases like, for example heart diseases or sepsis [4], are very high worldwide, so risk prediction plays a very important role. Their diagnosis requires a lot of experience, time and knowledge. For example, authors in research work [5] evaluated the relative validity of sepsis identification criteria in a large database with intensive care unit patients. The monitoring and clinical challenge of sepsis identification is also presented by [6]. The authors in [7] developed a novel traumatic sepsis score (TSS) whose validation results allow a reliable prediction of the sepsis risk. Furthermore they constructed a model using logistic regression based on a LASSO analysis. The authors in [8] used a statistical approach where they tried to derive and internally validate the sepsis risk score to predict future sepsis events. Using recorded vital signs and results of lab values from blood tests, C-statistic models and software-aided risk scores for

the prediction of sepsis were developed in [9]. Developing and validating another risk prediction model using a statistical approach is proposed by [10], [11]. The concepts of Hidden Markov models for disease risk classification and prediction are proposed by the authors in [12], [13]. A deep learning approach for identifying risk factors in Electronic Medical Records is suggested by [14].

A Support Vector Machine (SVM) was implemented by the authors in [15] to find classes of cancer subtypes. The authors in [16] developed an online incremental learning for the prediction of health parameters using a regression approach. To reduce the error they used a feedback method. The useful technique of finding an optimal alignment between two given time-dependent sequences under certain conditions is called DTW. The authors in [17] used DTW to calculate the similarity of dynamic gait data and to predict the risk of falling for older people. DTW is also used to train kNN classifiers in dialysis treatment as well as to predict patient's risks [18]–[20]. To achieve fast and more precise results in classification tasks, the authors in [21], [22] developed a global averaging method for DTW.

III. Methodology

A. Data preparation

Our data set is based on the large MIMIC-III clinical database [1]. The MIMIC III database (MDB3) contains information of patients admitted to different care units in hospitals. The MDB3 contains multiple measurements such as vital signs, patient demographics, laboratory test results, fluid balance as well as reports, diagnostics and notes. The MIMIC -III Waveform Database [2] (MDB3W) contains thousands of physiological waveforms recorded at bedside patient monitors and corresponding vital signs. As clinical data is collected separately from the patient's electronic health records, in our research, we used parts of a 'matched subset' consisting of the data from 370 patients. This subset allows a connection between the waveform records and the related clinical records. We used three vital signs: heart rate, mean arterial pressure and respiratory rate as inputs for our algorithm. As extracting and labeling of sepsis data is not a trivial task, we used different identification criteria for the sepsis disease to define for each patient the presence or absence of illness as described in [5], where different methods for identifying sepsis are presented. We used cross-validation as a method to estimate the performance of our model and its parameters which is also a good measure of how robust a model with its parameters is.

B. Dynamic Time Warping

DTW is an algorithm based on the Levensthein distance and was originally used in the speech recognition domain [21]. The DTW algorithm can find the optimal alignment between two time series even when they did not appear at the same instance. The applicability of DTW for data analysis is limited, since DTW does not include averaging techniques. In [21] a methodology for averaging a set of time series was derived to avoid iterative pairwise averaging. The authors present a global averaging methodology (DBA) for a length reduction of the averaged time series. A very good introduction to DTW and DBA is presented in [21], [22]. In our research, we used a combination of DTW, which allows to define a similarity measure for time series on the one side, and DBA to extend the methodology with an averaging method on the other side.

IV. Algorithm design

A. Step 1: Data Preparation

The first step of our proposed model (no.1 in Fig. 1) includes the pre-processing of the time series data given by $\hat{\mathbf{s}}_{i,p} \in \mathbb{R}^{N_{i,p}\times 1}$ for a patient p and a vital sign i and defines different training data sets according to the "different sepsis identification criteria" provided by [5]. Here we labeled patients as "$-$" if the patient was healthy, and "$+$" if the patient was ill. For simplicity, we considered only one data set in our diagram. As the waveform data has different lengths, the best approach is to aggregate the data and take one segment of *M* observations.

B. Step 2: Statistical transformation

The lengths $N_{i,p}$ of the patients' p time series data $\hat{\mathbf{s}}_{i,p}$ are different, so the idea is to aggregate each feature for patients, and then use the last *M* observations. As it is shown in (no.2 in Fig. 1), we describe the relationship among variables using descriptive statistics. We applied standard deviation, mean, median, trend, kurtosis, range and skew statistic to sets of ten measurements of each feature from patients. Afterwards, we applied multiple combinations of statistical approaches. Based on this step, we obtained the top $M = 200$ observations $\mathbf{s}_{i,p} \in \mathbb{R}^{M\times 1}$. Finally, we transformed each aggregated feature to keep it in the interval $[0, 1]$ of the training data. As we have information which subjects were positive or negative for the disease, we merged all data of the patient by features in positive and negative "feature pools". We use these transformed data samples which we obtained by the above described methodology for the next DBA step.

The matrix $\mathbf{E}_i^+ \in \mathbb{R}^{M\times N_+}$ (positive feature pool i) contains the vector of best statistical transformations $\mathbf{t}(i) \in \mathbb{R}^{M\times 1}$ of the corresponding vital values $i = 1, 2, 3$ in the columns. The same is done for the matrix of the negative patients $\mathbf{E}_i^- \in \mathbb{R}^{M\times N_-}$.

C. Step 3: DBA algorithm

First the data sets are divided in a training data (70%) and validation data set (30%) using cross validation. That means we take 70% of the columns (patients) of $\mathbf{E}_i^+$. The same is done for the negative patients $\mathbf{E}_i^-$. The remaining columns are used as validation data. For the learning step, we use the training data from $\mathbf{E}_i^+$ and $\mathbf{E}_i^-$ to create the DBA $\mathbf{d}_i^+ = f^{\mathrm{DBA}}(\mathbf{E}_i^+) \in \mathbb{R}^{M\times 1}$ (where $f^{\mathrm{DBA}}()$ denotes the DBA algorithm). In the same way we create $\mathbf{d}_i^-$. Here, DTW is applied between the individual time series and the average time series in order to find correlations between the coordinates of the averaged time series and the coordinates of the set of time series [21]. Then each coordinate of the averaged time

series ($\mathbf{d}_i^+, \mathbf{d}_i^-$) is updated as the Barycenter of the coordinates assigned to it in the previous step [21]. As a result of using DBA, we got two time series: one as model example for "+"-labeled patients and one as model example for "−"-labeled patients. As we use 3 vital signs as features, we got 6 "feature example" agents and representatives in total (3 in positive) - marked as $\mathbf{d}_i^+$ and 3 in negative direction - denoted as $\mathbf{d}_i^-, i = 1, 2, 3$.

D. Step 4: Classification and Validation

We used DTW to calculate the distance between the features of the patients from the validation data and $\mathbf{d}_i^+$ and $\mathbf{d}_i^-$ with $i = 1, 2, 3$. If the comparison result for a "+" patient is closer to $\mathbf{d}_i^+$, we marked it as 1 otherwise 0. Similarly, we calculated the distances between "−" patients and $\mathbf{d}_i^+$ and $\mathbf{d}_i^-$. Here $f^{\mathrm{DTW}}(\mathbf{s}_{i,p}, \mathbf{d}_i^+) \in \mathbb{R}^+$ in Fig. 1 (4a) denotes the DTW algorithm [21]. The final result for each patient p is represented in the form $y_p^* = (l, j, k) \in \{0, 1\}^3$, $* \in \{+, -\}$ (where l, j, k are the results for vital sign (feature) 1, 2, 3). The patient was classified as positive if (l, j, k) contains minimum two values of 1 in 3-tuple. Similarly, the patient was classified as negative, if (l, j, k) contains minimum two values of 1. Using the conditional probability, for "+" patients we calculated the risk of a disease as the ratio between the sum of all "+" patients from the validation data that have the same combination (l, j, k) and the total sum of all patients from the validation data that have the same combination. After Step 4 we calculated the risks:

$$r((l,j,k)) = \frac{\sum_{p \in P^+} \mathbb{1}_{\{y_p=(l,j,k)\}}}{\sum_{p \in P^+} \mathbb{1}_{\{y_p=(l,j,k)\}} + \sum_{p \in P^-} \mathbb{1}_{\{y_p=(l,j,k)\}}}$$

where the sets P^+ and P^- contain the positive and negative patients of the validation data respectively and $\mathbb{1}_{\{x=y\}}$ is the indicator variable defined by:

$$\mathbb{1}_{\{x=y\}} = \begin{cases} 1 & \text{if } x = y \\ 0 & \text{else} \end{cases}$$

An exemplary risk classification for "+" patient is represented in Fig. 1(4b).

V. Discussion and Results

As the identification of the sepsis disease is not a trivial task, our algorithm is trained on differently labeled data using different criteria of the disease [5]. For each criteria, using the proposed model, we created DBA centers in a positive and negative direction, creating thereby classifiers.The results showing the ability of the algorithm to recognize patients with disease are presented in Table I.The best recall result of 90,90% is obtained for labeled data in the SOFA sense [5], where 8 patients were not recognized as ill patients but were actually ill. The worst results were obtained for the Explicit criteria [5]. We also analysed specificity, recall, as well as total accuracy to estimate the risk of too many false positives. The best precision results 96,38% in the SOFA sense [5] are presented in Table II. The example of risk pattern for positive patients over the features is presented in Table III. The graphical representation of the best and the worst total accuracy results are given in Fig 2. The results and algorithm code generated during the current study are available from the first author. We also showed that different formulations of the disease criteria have a high impact on the performance of the algorithm.

TABLE I: The results of precision: Range-Std-Range

Attribute [5]	Negative [5]	Positive [5]	TP	FP	Precision
Angus	82	28	17	46	26,98%
Martin	96	14	10	67	12,98%
CDC	60	51	31	25	55,35%
Explicit	102	8	3	94	7,84%
Sepsis3	33	77	67	9	72,72 %
SOFA	22	88	80	3	96,38 %

TABLE II: The results of recall: Range-Std-Range

Attribute [5]	Negative [5]	Positive [5]	TP	FN	Recall
Angus	82	28	17	11	60,71%
Martin	96	14	10	4	71,42%
CDC	60	51	31	20	60,78 %
Explicit	102	8	3	5	37,5%
Sepsis3	33	77	67	10	87,01 %
SOFA	22	88	80	8	90,90 %

TABLE III: Example of risk pattern for criterion SOFA [5]

Risk pattern of the positive patients				
combination	(1,1,0)	(1,0,1)	(0,1,1)	(1,1,1)
risk	85,71%	90%	62,50%	81,75%

VI. Conclusion

A novel classification technique based on combined statistical and DBA approach for disease risk classification is reported. The proposed methodology exploits the advantages of DBA and a statistical approach, where the model performance is investigated by precision, recall and total accuracy functions. Results show that our algorithm preformed, in the best case, with precision and recall of 96.38% and 90.90%, respectively. The next phase of our research will be focused on applying the proposed algorithm to other domains and datasets, for example, in IoT environments.

Acknowledgment

This project was funded by federal ministry of education and science (BMBF) grant 13GW0280. We gratefully acknowledge the students Mr. Stefan Martens and Mr. Marcel Simon for providing necessary resources to accomplish this research work.

References

[1] A. L. Goldberger, L. A. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mietus et al., *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals*, Circulation, pp. 215- 220, http://circ.ahajournals.org/cgi/content/full/101/23/e215, 2000 .

[2] A. E. W. Johnson, T. J. Pollard, L. Shen, Li-wei. H. Lehman, M. Feng, M. Ghassemi et al. *MIMIC-III, a freely accessible critical care database*, Scientific Data, doi: 10.1038/sdata.2016.35, 2016.

[3] A. Tharwat, Classification assessment methods, Applied Computing and Informatics, Elsevier, doi:10.1016/j.aci.2018.08.003, 2018.

[4] C. V. Seymour, V. X. Liu, T. J. Iwashyna, et al. *The Third International Consensus Definitions for Sepsis and Septic Shock (Sepsis-3)*, JAMA, pp. 762- 774, doi:10.1001/jama.2016.0287, 2016.

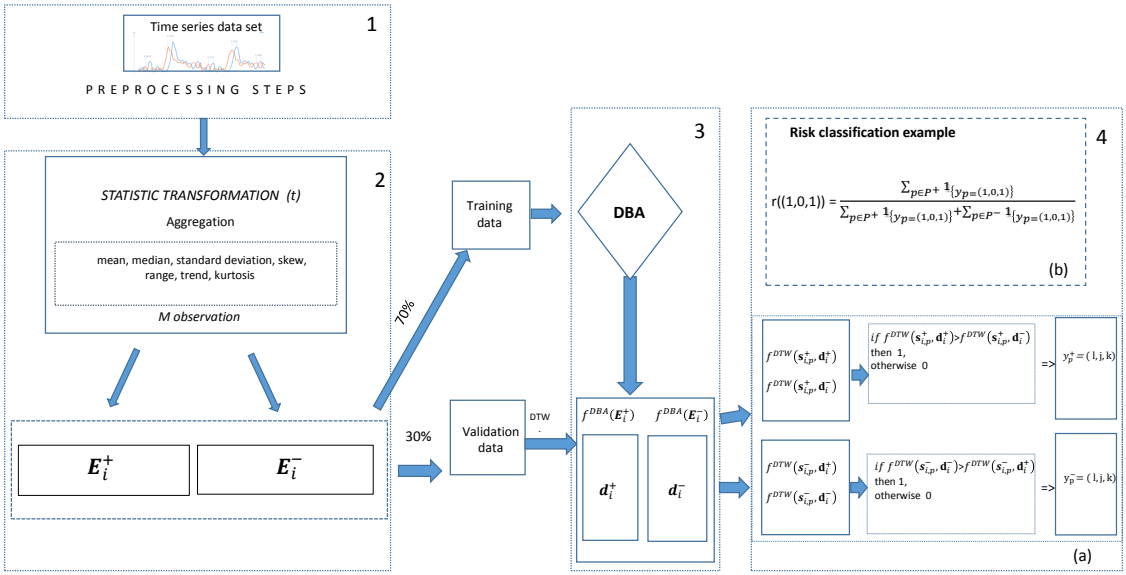

Fig. 1: Proposed risk classification system design

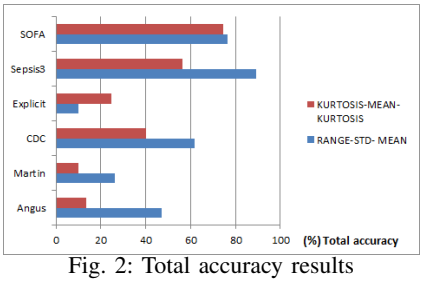

Fig. 2: Total accuracy results

[5] Alistair E. W. Johnson, J. Aboab,J. D. Raffa, Tom J. Pollard, Rodrigo O. Deliberato, Leo A. Cheli, David J. Stone *A comparative analysis of sepsis identification methods in an electronic database*, Crit Care Med. 2018 Apr; 46(4), pp. 494–499, doi: 10.1097/CCM.0000000000002965, 2018.

[6] J. L Vincent,*The Clinical Challenge of Sepsis Identification and Monitoring*, PLoS Med. 2016; 13(5): e1002022, doi:10.1371/journal.pmed.1002022, 2016.

[7] H. Lu, D. Wen, J. Sun, M.Chen, *Development and validation of a novel predictive score for sepsis risk among trauma patients*, World Journal of Emergency Surgery, vol. 14, doi: 10.1186/s13017-019-0231-8, 2019.

[8] H. E. Wang, J.P. Donnelly, R. Griffin, E. B. Levitan, N. I. Shapiro, G. Howard, M. M. Safford, *Derivation of novel risk prediction scores for community-acquired sepsis and severe sepsis*, Crit Care Med. 2016; 44(7), pp. 1285-1294, 2016.

[9] M. Faisal, A. Scally, D. Richardson, K. Beatson, R. Howes, K. Speed M.A. Mohammed, *Development and external validation of an automated computer-aided risk score for predicting sepsis in emergency medical admissions using the patient's first electronically recorded vital signs and blood test results*, Crit Care Med. 2018; 46(4), pp. 612-618, 2018.

[10] Stuart W. Grant, Gary. S. Collins, Samer A.M. Nashef, *Statistical Primer: developing and validating a risk prediction model*, European Journal of Cardio-Thoracic Surgery, Vol. 54, Issue 2, doi: 10.1093/ejcts/ezy180, pp. 203-208, 2018.

[11] X. S. Zhang, F. Tang, H. Dodge, J. Zhou, F. Wang, *MetaPred: Meta-Learning for Clinical Risk Prediction with Limited Patient Electronic Health Records*, Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data

[12] A. Singh, T. Tamminedi, G. Yosiphon, A. Ganguli, J. Yadegar, *Hidden Markov Models for modeling blood pressure data to predict acute hypotension*,IEEE International Conference Acoustics Speech and Signal Processing (ICASSP) 2010 doi:10.1109/ICASSP.2010.5495603, Dallas, USA, 2010.

[13] A. Sultana, H.L. Abdelwahab, M. Couture, *An improved Hidden Markov Model for anomaly detection using frequent common patterns*, IEEE International Conference on Communications (ICC) 2012, doi:10.1109/ICC.2012.6364527, Otawa, Canada, 2012 .

[14] T. Chokwijitkul, Anthony. N . Nguyen, H. Hassanzadeh, S. Perez, *Identifying Risk Factors For Heart Disease in Electronic Medical Records: A Deep Learning Approach*, Proceedings of the BioNLP 2018 workshop, pp. 18-27,Melbourne, Australia, 2018 .

[15] S. Huang, N. Cai, P.P. Pacheco, S. Narandes, Y. Wang, W. Xu, *Applications of Support Vector Machine (SVM) Learning in Cancer Genomics*, Cancer Genomics Proteomics. 2018 Jan-Feb; 15(1), pp. 41–51, doi: 10.21873/cgp.20063, 2018.

[16] K. Raghuraman, M. Senthurpandian, M. Shanmugasundaram, Bhargavi V. Vaidehi *Online incremental learning algorithm for anomaly detection and prediction in health care*, 2014 International Conference on Recent Trends in Information Technology, pp. 1-6,Chennai, India, doi:10.1109/ICRTIT.2014.6996092,2014.

[17] S. Jiang, B. Zhang, D. Wei, *The Elderly Fall Risk Assessment and Prediction Based on Gait Analysis*,IEEE 11th International Conference on Computer and Information Technology 2011, pp. 176 - 180, Pafos, Cyprus, doi:10.1109/CIT.2011.82, 2011.

[18] J. Wiens, E. Horvitz,, J. V. Guttag, *Patient Risk Stratification for Hospital-Associated C. diff as a Time-Series Classification Task*, Advances in Neural Information Processing Systems 25 (NIPS 2012), pp. 467-475, 2012.

[19] T. Knorr, L. Schmidt-Thieme, C. Johner, *Identifying Patients at Risk*, Cooperation in Classification and Data Analysis, pp. 131-140, Germany, 2009.

[20] A. Giannoula, A. Gutierrez-Sacristán, A. Bravo, F. Sanz, Laura I. Furlong, *Identifying temporal patterns in patient disease trajectories using dynamic time warping: A population-based study*, Sci Rep. 2018; 8: 4216, doi: 10.1038/s41598-018-22578-1, 2018.

[21] F. Petitjean, A. Katterlin, P. Gancarski *A global averaging method for dynamic time warping, with applications to clustering*, Pattern recognition, pp. 678-693, 2011.

[22] F. Petitjean, G. Forestier, Geoffrey I. Webb, Ann E. Nicholson, Y. Chen, E. Keogh, *Dynamic Time Warping Averaging of Time Series Allows Faster and More Accurate Classification.*, IEEE International Conference on Data Mining, ICDM, pp. 470-479, doi: 10.1109/ICDM.2014.27, 2015.