Patient-Ventilator Asynchrony Detection via Similarity Search Methods

Chenyang Wang, Uwe Aickelin, Ling Luo, and Goce Ristanoski
University of Melbourne, Melbourne, Australia
Email: chenyangw3@student.unimelb.edu.au, {uwe.aickelin, ling.luo}@unimelb.edu.au

Mark E. Howard and David Berlowitz
Austin Health, Melbourne, Australia

***Abstract*—Patient-Ventilator Asynchrony (PVA) is a common cause of ventilation-related medical complications and are traditionally only able to be reliably identified by trained clinicians. The need for constant monitoring and limited access to trained experts are major challenges in managing PVA, both of which can potentially be solved by automating the detection process. In this research, we propose a new data-driven approach to PVA detection using several similarity and randomness measures, including how unusual a time window is in the series and randomness of the time window. We found that all these similarity or randomness measures can be estimated with variants of the highly efficient Matrix Profile (MP) algorithm, and that one base routine can be repeated to generate all the features used in classification. We show that MP-based features, when used in combination with basic statistical and spectral features, can achieve an F-2 score of over 0.9 for two classes of PVA events in a sample of participants with moderate to high rate of PVA occurrence.**

***Index Terms*—patient-ventilator asynchrony, matrix profile, anomaly detection**

I. Introduction

Patient-Ventilator Asynchrony (PVA) are anomalous events during mechanical ventilation where the ventilator cycle is desynchronised from the patient's breathing cycles. It has been shown that frequent PVA events can lead to many adverse consequences for the patient, such as reduced sleep quality, lung injury, and even increased ICU and hospital mortality rate [1]. Detecting and managing PVA often rely on bedside waveform monitoring by trained clinicians, and the availability of such experts also poses a challenge. Therefore, research on automating PVA detection would greatly benefit the patients, as well as relieving the stress on the medical staff.

In this work, our research aims to explore the most consistent distinguishing properties of PVA events versus non-PVA monitoring data by examining similarity and randomness measures. We find that the local inter-class similarity, the local randomness measure, and the multi-channel frequency difference over time are three properties that capture the fundamental differences between asynchrony and synchronised breaths. We then propose a new PVA detection algorithm based on exploiting these three properties and show that it is possible to build a highly interpretable anomaly detector for PVA events that is also competitive against existing machine learning detection algorithms using domain-specific features.

II. Related Work

The majority of existing PVA automatic detection methods use either rule-based threshold detectors or machine learning classifiers using hand-crafted features. Most of previous research works focus on defining a set of rules based on domain knowledge to detect particular patterns in the ventilator readings and trigger an event detection when a preset threshold is reached, such as in [2]-[4]. These methods are often limited in the types of events they can detect, or the quality of input data required.

Alternatively, machine-learning algorithms based on hand-crafted features [5], [6] or raw data inputs [7] have been proposed to tackle this problem. They often achieve over 90% precision and recall in most cases, proving that effectively detecting PVA automatically is indeed feasible. However, the predictions made by these algorithms can be hard to interpret for domain experts, and the complexity of these models can make them unsuitable to be adapted to different patient types or sensor setups.

In recent years, there have been several efforts on PVA detection by estimating and comparing the data-generation process of PVA and non-PVA breathing cycles. Marchuk *et al.* [8] proposed using a Hidden Markov Chain model to identify hidden state changes from low PVA-risk state to high PVA-risk state and vice versa. Although this model is only able to model the count of PVA events over a longer time window, rather than performing fine-grained individual event detection, this nevertheless shows that it is possible to trace the observed distinctions between PVA and normal data back to their different data-generation processes. Sarlabous *et al.* [9] proposed that instead of estimating the actual data-generation process, we can estimate the randomness or

Manuscript received July 30, 2021; revised December 3, 2021.

doi: 10.12720/jait.13.1.15-20

predictability changes of the underlying process via Sample Entropy and detect potential PVA events by exploiting the high variation / low predictability of PVA events versus regular breaths.

Overall, existing research traditionally rely more on identifying features of given PVA event types, but there are also promising results exploiting the high variation and inter-class dissimilarity of PVA events, taking an approach more similar to outlier detection problems. We believe that the latter approach has the potential to be more robust, adaptive and interpretable, but is not yet adequately explored. In this work, we aim to quantify such similarity and variability, and integrating them into the feature set for PVA detection.

III. Data and Methods

A. Data Collection and Annotation

The data for our analysis was collected from 59 participants undergoing non-invasive ventilation. Multiple channels of ventilator readings such as mask pressure and airflow were collected. Additional sensors on the patient also recorded more data channels such as abdominal and thoracic movement, leg movement, EMG, etc. Fig. 1 shows a sample data with four channels.

A group of respiratory experts subsequently examined the data and added annotations (consisting of timestamps and event type) over breathing cycles in the monitoring data they deemed to be PVA events. By definition, PVA events are events that introduce mismatches between the ventilation cycles and the patient's breathing efforts.

B. Change of Frequency Difference Detection

Given a proxy measure of ventilation cycles and patient's breathing cycles (for example "mask pressure" and "thoracic movement"), the most fundamental distinction between a PVA event and a normal ventilation cycle should be that the PVA event has a non-zero frequency difference between the two data channels. Many of the previous researches, such as [2], [6], [7] did not have access to proxies of breathing efforts, and did not take advantage of this PVA detection method. We have access to two measures of breathing efforts, thoracic and abdominal movement in our dataset, therefore we include this direct detection approach in our analysis. Fig. 1 gives an example of direct detection approach.

To determine the frequency of each data channels over time, we opt to use Filtered Hilbert Transform to extract phase sequences from detrended data channels, and then use peak detection to find end points of each cycle. We then apply a windowed moving average to count the number of completed cycles around each time point, as use the number of neighbouring cycles as a measure of instantaneous frequency. We found this approach to be more robust than directly calculating frequency from phase angles. We then calculate the phase difference between two channels by taking the difference of their respective windowed moving average cycle counts. A positive difference typically corresponds to Autocycle (AT) or Double Trigger (DT) events, whereas a negative difference often corresponds to Ineffective Effort (IE) events.

C. Self-Join Similarity Search via Matrix Profile

The direct approach is straightforward, but often fails when the signal is highly irregular or there are alternating behaviours over time. To detect PVA events more reliably, we desire to identify the irregularities in the data and determine which ones correspond to PVA events. Hence, we explore performing similarity search via matrix profile.

As suggested by the results of [9], PVA event intervals are less predictable and more irregular than normal ventilation cycles, and therefore will likely also have a lower inter-class similarity compared to normal cycles. PVA events are also usually the minority class across the duration of the patient monitoring time series, therefore it is more likely for time windows of regular cycles to be similar to each other, and for time windows containing PVA events to be more distinct from other time windows.

The Matrix Profile algorithm [10] is well-suited for anomaly detection tasks such as this. It compares each fixed-length time windows to all other time windows in the same time series (or with a different time series), and either finds the distance of each time window to their nearest neighbour (we denote this as MP_{1NN}), or the accumulated correlation to all other windows (denoted as MP_{sum}). The MP_{1NN} and MP_{sum} of matrix profiles are defined as below:

$$MP_{1NN} = \langle \min_{j,|i-j|>d} \left\| x_{i:i+m} - x_{j:j+m} \right\| \rangle, i \in [0, l-m]$$

$$MP_{sum} = \langle \sum_{j,|i-j|>d} \left\| x_{i:i+m} - x_{j:j+m} \right\| \cdot I \left\| x_{i:i+m} - x_{j:j+m} \right\| > \alpha \rangle$$

where i, j are indices of time series elements, $\boldsymbol{x}$ are length-m subsequences of the time series, α is a threshold value for summation to exclude low and/or negative correlations, and I is the indicator function.

Since PVA events are typically rarer and more dissimilar to each other, they will more likely show up as having a higher MP_{1NN} and very low MP_{sum} compared to normal cycle time windows.

As classification features, the self-join matrix profiles capture how unusual a given time window is compared to the rest of the time series. From the matrix profile distribution histograms in Fig. 2 and Fig. 3, we can see that certain event classes like Autocycle (AC) and Double Trigger (DT) has noticeably higher MP_{1NN} and $MPsum$ compared to the unlabelled class (OTHER). However, we also noticed that for the Ineffective Effort (IE) event class, neither matrix profile measures demonstrate a distinct distribution from the unlabelled class, therefore these features may not have much discriminative power for IE event.

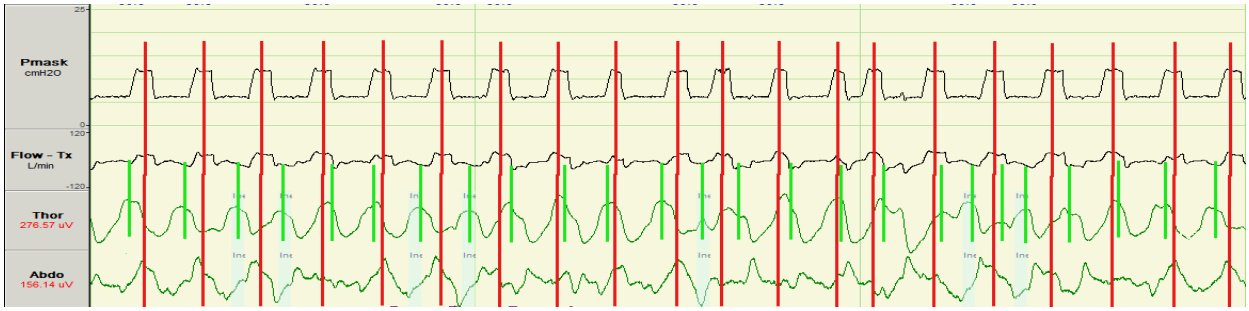

Figure 1. PVA detection of Ineffective Effort events via frequency difference detection. The red lines in the figure are lined up with mask pressure cycles, whereas the green lines are lined up with thoracic signal cycles. As we see, there are20 Pmask cycles in this snippet, but there are 25 corresponding Thor cycles, indicating there are missed ventilation cycles.

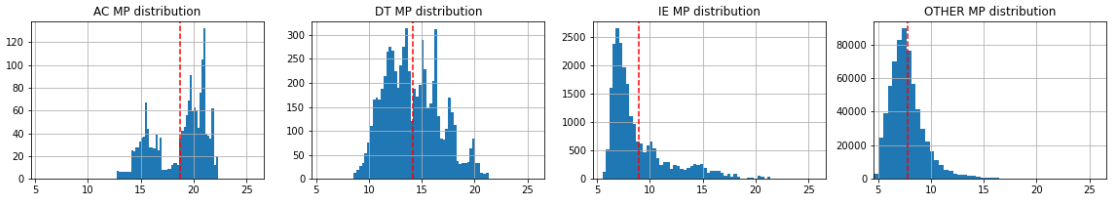

Figure 2. MP_{1NN} Distribution histogram for each event class (AC - Autocycle, DT - Double Trigger, IE - Ineffective Effort, OTHER - not labelled as events) from observation s2g2p55.

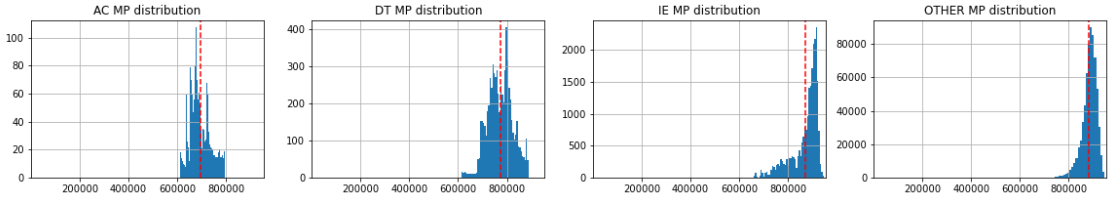

Figure 3. MP_{sum} Distribution histogram for each event class, from observation s2g2p55.

D. *Efficient Modified Approximate Entropy via Matrix Profile*

Matrix Profile provides us with a highly efficient tool for comparing self-join similarities between rolling time windows. However, it does not directly measure the randomness or regularity of the time series, only how it manifests in the distribution of time series snippets. MP_{1NN} and MP_{sum} are still dependent on the particular distribution of time series "shapes". For example, a non-PVA pattern might be highly regular, but because it only appears a few times, it may not have a high MP_{sum} score. Conversely, a PVA event might be quite irregular, but it might happen to match up well with one single pattern by chance, and end up with a much lower MP_{1NN} score. Is there a way to directly measure randomness with matrix profile, similar to the Sample Entropy used in [9]. We believe this is possible.

Approximate Entropy [11] and Sample Entropy [12] are regularity statistics for time series. They measure how likely patterns that are similar to each other will evolve over time in a similar way. According to [13], approximate entropy and sample entropy can be defined as follows:

$$SampEm(\mathrm{X}, m, r) = -log \frac{\sum_{i=0}^{l_X-m-1} \sum_{j=0, j\neq i}^{l_X-m-1} I\{\|\mathrm{x}_{i:i+m+1} - \mathrm{x}_{j:j+m+1}\|_\infty < r\}}{\sum_{i=0}^{l_X-m-1} \sum_{j=0, j\neq i}^{l_X-m-1} I\{\|\mathrm{x}_{i:i+m} - \mathrm{x}_{j:j+m}\|_\infty < r\}}$$

$$ApproxEm(\mathrm{X}, m, r) = -\frac{1}{N-m} \sum_{i=0}^{N-m-1} log \frac{\sum_{j=0}^{l_X-m-1} I\{\|\mathrm{x}_{i:i+m+1} - \mathrm{x}_{j:j+m+1}\|_\infty < r\}}{\sum_{j=0}^{l_X-m-1} I\{\|\mathrm{x}_{i:i+m} - \mathrm{x}_{j:j+m}\|_\infty < r\}}$$

where m is the time window size and r is a predefined distance threshold that determines what subsequences we consider similar. Essentially, approximate entropy sums up the log probability of length-m time windows which are close to each other staying close to each other when extended to length $m+1$. Sample entropy moves the summation inside the log function, with a few other minor changes to the calculation. Notice that in both equations above, the calculation involves comparing each time window to all other time windows in the same series, which is exactly the task that can be efficiently computed with matrix profile. Also, instead of summarising the entire time series by adding up all the log probabilities over the full time length, we can perform the summation / averaging over a rolling time window or a given set of episodes (like breathing cycles), and obtain a new time series of evolving approximate entropy.

We define our modified approximate entropy pre-summation as follows:

$$Modified.ApproxEm = \left\langle -log \frac{\sum_{j=0}^{l_X-m-1} Corr(\mathrm{x}_{i:i+m+1}, \mathrm{x}_{j:j+m+1}) \cdot I\{Corr(\mathrm{x}_{i:i+m+1}, \mathrm{x}_{j:j+m+1}) > r\}}{\sum_{j=0}^{l_X-m-1} Corr(\mathrm{x}_{i:i+m}, \mathrm{x}_{j:j+m}) \cdot I\{Corr(\mathrm{x}_{i:i+m}, \mathrm{x}_{j:j+m}) > r\}} \right\rangle$$

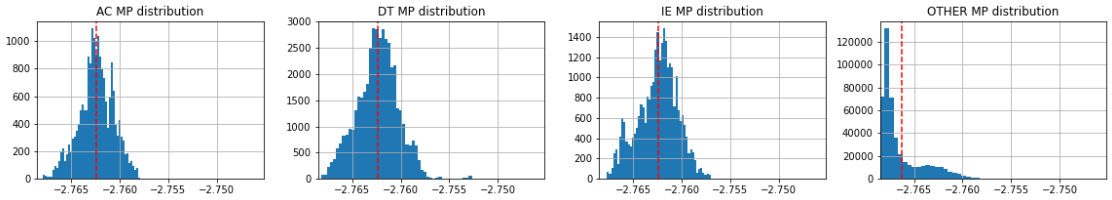

Figure 4. Modified approximate entropy distribution, from observation s2g2p55.

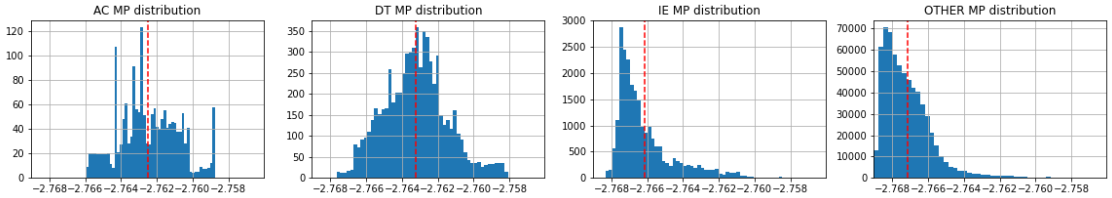

Figure 5. Modified approximate entropy distribution, from observation s2g1p19.

Notice that we also changed the subsequence a counting with distance threshold of $\|\cdot\|_\infty$ to summation of correlations larger than a threshold.

Fig. 4 and Fig. 5 show that when applied to the PVA dataset, modified approximate entropy generally has distinct distributions for AT and DT events versus non-event time windows, but the distribution for IE events might differ less from the corresponding non-event distribution for certain participants.

E. *Feature Generation and Classification*

We aggregate the features generated from previous sections per ventilation cycle, including signal phase, amplitude, phase difference and peak count, as well as the two MP-based features, MP_{1NN} and modified approximate entropy as features for final classification. We first select valid sections of observation data and segment the data into individual ventilation cycles using a mean squared peak detector over the mask pressure channel. We then calculate the mean, max, min and standard deviation of simple features listed above, for each data channel. The classification methods can be generic machine learning models that are easy to interpret. For example, we use a LightGBM classifier [14] as the final detector. We label each ventilation cycle based on whether it contains a PVA event.

IV. Experiment

A. *Experimental Setup*

We conduct experiment using the data from the data collection section. We select 44 participants with sufficient valid ventilation cycles in their observation data to be used in the experiment.

We sequentially perform leave-one-out testing on each of the participants, where the classifier is trained and five-fold validated on all but one participant, then tested on the remaining participants. We use F-2 score ($(1+2^2)\cdot(precision\cdot recall)/(2^2\cdot precision+recall)$) as the primary performance metric due to the higher importance of recall for our applications, then calculate the best precision and recall given the best F-2.

To evaluate the effectiveness of our model, we apply the same classifier and the same validation and testing settings, using the feature sets offered in the ventMAP model of [15]. The model extracts 16 shape-based features per ventilation cycle from pressure and flow values. It does not make use of the abdominal or thoracic sensor readings in our dataset, however, thus a direct comparison of performance would not be fair. Therefore, we also combine our feature sets with that of ventMAP to see whether and by how much the new features can improve upon the performance of ventMAP features alone.

B. *Results*

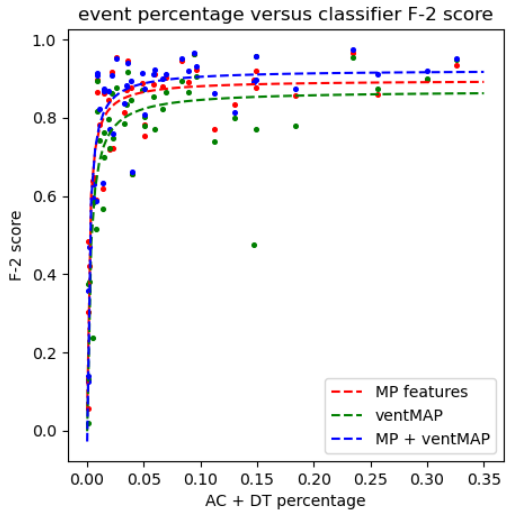

Figure 6. Event (AC+DT) percentage of all ventilation cycles versus the performance of the classifier. As we see, the classifier performs better on participants with higher event prevalence.

The overall classification performance and comparisons are shown in Fig. 6. For a significant portion of the participants tested, our model is able to detect combined AC and DT events at over 90% recall and over 80% F-2 score. As we can see in Fig. 6, generally, event detection is significantly better in participants with a higher prevalence of PVA (i.e., AC+DT percentage > 5%). Looking at the testing precision and recall, we see

that the classifier can achieve over 90% recall for most participants, but the detection precision drops as the ratio of PVA events as a percentage of all ventilation cycles decreases to less than 3-5%.

We can also observe from Fig. 7 that the MP-based features perform slightly better than ventMAP features. If we combine the MP-based and ventMAP features, we see a noticeable F-2 score increase across all event prevalence levels. The precision of detection is still low for low-event-rate participants, but with the combined features, even there it is much improved.

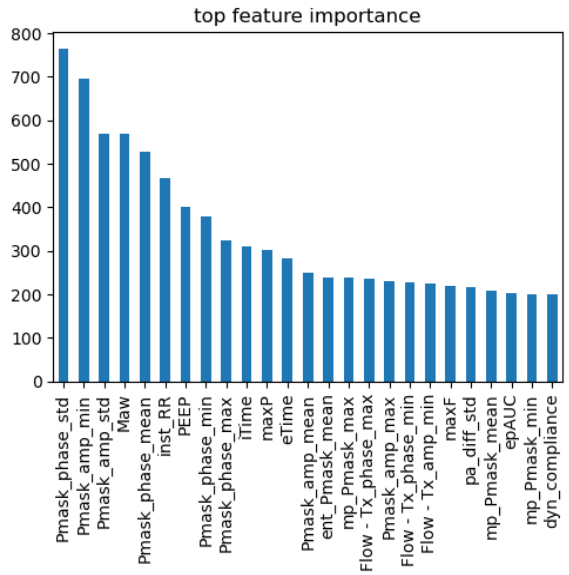

Figure 7. Feature importance for combined feature set.

If we look at the feature importance of the classifier for the combined feature set (Fig. 7), we see that the classifier is making use of features from all sources, including basic shape, ventMAP, phase difference, self-join matrix profile and modified approximate entropy.

V. Discussion

We have shown that our MP-based features, self-join matrix profile and modified approximate entropy, when coupled with basic statistics and signal features, can be effective for Autocycle and Double Trigger event detection for participants with high event rates. The performance is at least comparable to the results with the same classifier using ventMAP features. Generally, our method is able to achieve very high (90%+) test recall for most participants. From Fig. 7 we can also find the MP-based features being actively used by the classifier as some of the most relevant features when we mix both MP and ventMAP features in the same feature set. The classifier trained on combined feature set performs better than with separate feature sets, and make use of features from both sets, indicating that the new MP-based features are indeed providing some information not captured by the ventMAP features.

Unfortunately, the precision suffers when the participants' PVA event rate is very low. The reason for poor precision in low-PVA participants has to be investigated further, as normally it should be easier to detect anomalies when most of the data is regular. Currently we believe there could be two non-exclusive explanations. The first is that the low-PVA patients often have isolated PVA events without a sustained period of irregular breathing, which could be easier to miss. The second explanation is that as in the previous explanation, PVA events in low-event-rate participants might behave quite differently, and there might not be enough samples in the data for the classifier to learn well enough, because they happen so rarely. We hope that with more data, data augmentation techniques and stratified training, we will be able to achieve higher performance even on participants with low PVA rates.

We have not addressed event identification problem with the current classifier. Different PVA events like Autocycle or Double Trigger can appear very similar based on some basic features like pressure standard deviation or cycle difference, but the waveforms will have different shape characteristics. The self-join MP-based features are great for anomaly detection, but alone, they cannot effectively determine the type of the anomaly,. Therefore our current MP-based feature set is unfit for high-precision PVA event identification.

In future work, we are going to explore using matrix profile as an AB comparison algorithm that measures the similarity between two different time series and their subseries. Using this approach, we will be able to use annotated data as templates and find the closest match of a suspected PVA event in a collection of labelled time series, making event identification easier. We are actively investigating ways we can incorporate both similarity self-match and cross-match in our methods.

Another task for future work is to include Ineffective Effort detection in our method. The main challenge is that Ineffective Effort events can be much subtler than Autocycle or Double Trigger events, and often on a shorter time scale. Our current approach only focuses on anomalies on the scale of whole ventilation cycles, but for sub-cycle anomalies, we will have to adjust how we calculate our features, for example, using matrix profile with lower time window sizes. This will be challenging, as shorter time window distances are more affected by random noise, which we will have to mitigate.

VI. Conclusion

In this work, we proposed quantifying the variability of the ventilator output time series with two similarity-based features, self-join matrix profile and modified approximate entropy, and showed that they serve as great addition to existing shape-based features in PVA detection through machine learning. Even though these features do not achieve high detection precision on their own, they are powerful for detecting suspected PVA events, and can potentially boost other methods that are better at classifying events. In future works, we aim to extend similarity-based methods to supervised nearest-neighbour search for better event confirmation and identification.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

Chenyang Wang conducted the research and drafted most of the paper. Uwe Aickelin, Ling Luo and Goce Ristanoski participated in algorithm selection, data analysis and also edited the final paper. Mark E Howard and David Berlowitz provided the clinical data and offered professional advice on data analysis.

REFERENCES

[1] L. Zhang, *et al.*, "Detection of patient-ventilator asynchrony from mechanical ventilation waveforms using a two-layer long short-term memory neural network," *Computers in Biology and Medicine*, p. 103721, 2020.

[2] C. M. Yeh, *et al.*, "Matrix profile I: All pairs similarity joins for time series: A unifying view that includes motifs, discords and shapelets," in *Proc. IEEE 16th International Conference on Data Mining*, 2016.

[3] C. Sinderby, S. Liu, D. Colombo, G. Camarotta, A. S. Slutsky, P. Navalesi, and J. Beck, "An automated and standardized neural index to quantify patient-ventilator interaction," *Critical Care*, vol. 17, pp. 1-9, 2013.

[4] L. Sarlabous, *et al.*, "Development and validation of a sample entropy-based method to identify complex patient-ventilator interactions during mechanical ventilation," *Scientific Reports*, vol. 10, pp. 1-12, 2020.

[5] J. S. Richman and J. R. Moorman, "Physiological time-series analysis using approximate entropy and sample entropy," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 278, pp. H2039-H2049, 2000.

[6] S. M. Pincus, "Approximate entropy as a measure of system complexity," *Proceedings of the National Academy of Sciences*, vol. 88, pp. 2297-2301, 1991.

[7] Q. Mulqueeny, P. Ceriana, A. Carlucci, F. Fanfulla, M. Delmastro, and S. Nava, "Automatic detection of ineffective triggering and double triggering during mechanical ventilation," *Intensive Care Medicine*, vol. 33, pp. 2014-2018, 2007.

[8] Y. Marchuk, *et al.*, "Predicting patient-ventilator asynchronies with hidden Markov models," *Scientific Reports*, vol. 8, pp. 1-7, 2018.

[9] G. Ke, *et al.*, "LightGBM: A highly efficient gradient boosting decision tree," in *Advances in Neural Information Processing Systems*, vol. 30, 2017.

[10] S. Imani, F. Madrid, W. Ding, S. Crouter, and E. Keogh, "Matrix profile XIII: Time series snippets: A new primitive for time series data mining," in *Proc. IEEE International Conference on Big Knowledge*, 2018.

[11] B. Gholami, T. S. Phan, W. M. Haddad, A. Cason, J. Mullis, L. Price, and J. M. Bailey, "Replicating human expertise of mechanical ventilation waveform analysis in detecting patient-ventilator cycling asynchrony using machine learning," *Computers in Biology and Medicine*, vol. 97, pp. 137-144, 2018.

[12] M. Dres, N. Rittayamai, and L. Brochard, "Monitoring patient-ventilator asynchrony," *Current Opinion in Critical Care*, vol. 22, pp. 246-253, 2016.

[13] A. Delgado-Bonal and A. Marshak, "Approximate entropy and sample entropy: A comprehensive tutorial," *Entropy*, vol. 21, p. 541, 2019.

[14] A. Cuvelier, L. Achour, H. Rabarimanantsoa, C. Letellier, J. F. Muir, and B. Fauroux, "A noninvasive method to identify ineffective triggering in patients with noninvasive pressure support ventilation," *Respiration*, vol. 80, pp. 198-206, 2010.

[15] J. Y. Adams, *et al.*, "Development and validation of a multi-algorithm analytic platform to detect off-target mechanical ventilation," *Scientific Reports*, vol. 7, p. 14980, 2017.

Chenyang Wang is a PhD student at University of Melbourne. His current research focus is on data mining with machine learning approaches, specifically combining similarity search and machine learning for time series anomaly detection.

Prof. Uwe Aickelin is the Head of School of Computing and Information Systems at the University of Melbourne. He has worked for more than twenty years in the fields of Artificial Intelligence, Optimisation and Data Mining, and has authored more than 280 publications. His specific expertise is in the modelling stages of problems with a focus on robust methods to overcome uncertainty.

Dr. Ling Luo is a lecturer at the School of Computing and Information Systems, University of Melbourne, Australia. Before joining UoM, Ling was an associate lecturer at UTS, and a postdoctoral research fellow at Analytics, Data61 (formerly NICTA), CSIRO. Her research focuses on data mining, machine learning, temporal modelling, stochastic processes and user behaviour pattern analysis.

Dr. Goce Ristanoski received his PhD at The University of Melbourne, Australia in 2014, specializing in Machine Learning Methods for time series forecasting. He has since working in both industry and academic roles. His work experience covers both the research components of using Machine Learning and AI decision support tools, as well as development and production of ML based products that deliver prediction and optimization functions. He has been working in the area of applied AI for medical tasks for the past few years.

Associate Professor Mark Howard is the Director of the Victorian Respiratory Support Service at Austin Health. He is a specialist physician in respiratory and sleep medicine with interests in long-term ventilation of patients with respiratory failure and the impact of sleep disorders on driving and occupational health and safety.

Prof. David Berlowitz is a Physiotherapist with the Victorian Respiratory Support Service who holds the University of Melbourne Chair in Physiotherapy at Austin Health in Melbourne Australia. His research encompasses respiratory physiology, sleep, health systems research, and clinical trials of therapies and care models, especially in neuromuscular diseases such as Spinal Cord Injury and Motor Neuron Disease.