Investigation of Machine Learning Techniques in Forecasting of Blood Pressure Time Series Data

Shamsul Masum, John P. Chiverton, Ying Liu and Branislav Vuksanovic

School of Energy and Electronic Engineering, University of Portsmouth
Anglesea Building, Anglesea Road,
Portsmouth (UK) PO1 3DJ
shamsul.masum@myport.ac.uk,john.chiverton@port.ac.uk

Abstract. The aim of this paper is to investigate different machine learning based forecasting techniques for forecasting of blood pressure and heart rate. Forecasting of blood pressure could potentially help a clinician to take preventative steps even before dangerous medical situations occur. This paper examines forecasting blood pressure 30 minutes in advance. Univariate and multivariate forecast models are considered. Different forecast strategies are also considered. To compare different forecast strategies, LSTM and BI-LSTM machine learning algorithms were included. Then univariate and multivariate LSTM, BI-LSTM and CNN machine learning algorithms were compared using the two best forecasting strategies. Comparative analysis between forecasting strategies suggest that MIMO and DIRMO forecast strategies provide the best accuracy in forecasting physiological time series data. Results also appear to show that multivariate forecast models for blood pressure and heart rate are more reliable compared to blood pressure alone. Comparative analysis between MIMO and DIRMO forecasting strategies appear to show that DIRMO is more reliable for both univariate and multivariate cases. Results also appear to show that the forecast model that uses BI-LSTM with the DIRMO strategy is the best overall.

Keywords: time series forecasting; univariate data; multivariate data; forecast strategies; LSTM; BI-LSTM; CNN; blood pressure; heart rate.

1 Introduction

Long-term time series forecasting of physiological data could potentially help health care professionals to predict and perhaps even prevent needing to treat patients based on their diagnosis. Forecasting physiological data 30 minutes in advance could potentially help a health care professional in this way. For instance, for general decision making or possibly intervening dangerous clinical events such as hypotensive events [22]. However, physiological times series analysis has been mostly conducted on event prediction thus limited to short-term and single-step prediction [36, 5, 17, 39]. Most of these works directly map input physiological signals to output values. Moreover, they are unable to model the underlying temporal dependencies in time series, such as those present in physiological data dynamics. These works have difficulty in modelling contextual information and sequential measurements simultaneously. This results in a decay in accuracy over time and requires frequent re-calibration. There is a very limited number of works that actually perform forecasting on continuous values of physiological data. In contrast to this, continuous monitoring is often a crucial part of clinical decision making. Examples could potentially include glucose monitoring [32] and EEG monitoring in the ICU [43].

Forecasting across multi-step and for a long-term horizon is very challenging [7, 3, 29]. Also, the literature that forecasts physiological time series data does not usually consider different forecast strategies [23, 8, 24]. This can be an important consideration [30, 29]. Taieb et al. described different forecast strategies and showed that forecast strategies play a vital role in long-term forecasting scenario [42]. Comparison of different forecast strategies can help to show which strategy is best for the forecast model. This is therefore a consideration here, particularly for the case of physiological time series data.

Forecasting with univariate and multivariate time series data has been a great consideration for researchers in forecasting for many years. Preez and Witt compared univariate and multivariate time series data in forecasting international tourism demand and found forecasting based on univariate time series data outperforms multivariate data [10]. Aboagye-Sarfo et al. found multivariate time series data outperformed univariate time series data in forecasting emergency department demand [2].

However, such comparisons are lacking when considering physiological data. Billis and Bamidis forecast artificial univariate blood pressure time series data [8] thus missing the multivariate comparison. Li et al. forecast blood pressure (BP) with multivariate data [24] but did not compare the results with the univariate case. Li et al. considered a number of different machine learning techniques, combined with a Contextual Layer. This enabled relative constants, such as BMI to be included at a different point in learning process. Lee and Mark performed forecasting of 30 minutes continuous mean arterial blood pressure (MAP) values. They used multivariate physiological time series [23]. Their work included age and medication information as well. However they did not compare against univariate physiological time series performance or forecast strategy. In more recent work, Su et al. [41] predicted Systolic BP and Diastolic BP sequences using a multi-layer BI-LSTM network. They used a univariate BP dataset extracted from 84 and 12 healthy people. It would be of interest to compare univariate and multivariate approaches in time series forecasting of blood pressure. Forecast strategy is also another important consideration. It is therefore of interest to explore whether an additional vital sign such as Heart Rate (HR) could improve the forecast accuracy of the response variable or perhaps only the past data of the response variable is good enough in forecasting.

There are quite a wide range of machine learning algorithms. This includes Naive Bayes [25], Support Vector Machines [34], Support Vector Regression [40], Gradient Boosted Regression Tree [11], Factorisation Machine [35] and Multilayer Perceptrons. They can be applied to time series forecasting of blood pressure but they are not specifically designed to deal with temporal data. On the other hand, it was also discovered that machine learning based sequential techniques are well suited for such problems [13, 4, 20]. Examples include Gaussian Processes (GP)s, Hidden Markov Models (HMM)s, Conditional Random Fields (CRF)s. Unfortunately they are unable to handle long-term dependencies. A bit more recent development has been Long-Short-Term-Memory (LSTM) and Bidirectional- LSTMs (BI-LSTM). These are based on Recurrent Neural Networks (RNN)s. They have gained attention for time series forecasting in different fields. This has included traffic speed prediction [27], solar power forecasting [12], electric load forecasting [30] and natural language processing [44]. LSTMs and BI-LSTMs have also been applied to medical time series data. Lipton et al. used LSTM networks to assign diagnostic information learned using multivariate time series of clinical measurements [26]. Nguyen et al. used both LSTMs and BI-LSTMs to predict mortality outcomes of patients in Intensive Care Units (ICU)s by modelling physiological time-series data [33]. Zhu et al. considered supervised BI-LSTM RNNs to predict ICU mortality [45].

Convolutional Neural Networks (CNN)s have been used for time series forecasting. Applications have included ECG classification [19], structural health monitoring [1] and motor-fault detection [15]. These techniques were found to be effective in capturing long term dependencies and the nonlinear dynamics especially in comparison to classical machine learning algorithms. However, typically, they are used to perform classification rather than actual forecasting of physiological data. Time series can also be used in combination with e.g regression to forecast future values of the physiological state of a patient. However, to the best of the authors' knowledge, relatively few works actually consider this.

In contrast, time series forecasting of physiological data for both BP and HR is considered 30 minutes in advance. This paper compares different forecast strategies in order to identify the best strategy. Following this, to compare univariate and multivariate approaches, forecast models with LSTM, BI-LSTM and CNN algorithms are compared. The two best performing forecast strategies, Multiple Input Multiple Output (MIMO) and DIRMO are also included. These forecast models are used to forecast blood pressure 30 minutes in advance for both univariate (i.e. BP) and multivariate (i.e. BP and HR) cases.

2 Forecasting Strategies

Taieb et al. [42] compared five forecast strategies: recursive strategy, Direct strategy, DirRec strategy, MIMO strategy and DIRMO strategy. They applied their work to neural network based time series modelling consisting of cash machine withdrawals. The same set of strategies are also considered here but with application to physiological time series forecasting. The details of each strategy and their differences can be seen in e.g. [42]. Among these strategies, MIMO strategy produces multiple outputs from a single-step forecast. All the other strategies need to be performed in multiple steps to forecast multiple outputs.

3 Machine Learning Algorithms

Three machine learning algorithms are considered here.

Long-Short-Term-Memory (LSTM) was introduced by Hochreiter & Schmidhuber [14] in 1997. LSTMs are able to learn long-term dependencies better than the simpler Recurrent Neural Network (RNN) architecture. In theory, an RNN appears to pass some potentially useful properties for long-term forecasting. Perhaps it is even capable of handling long-term dependencies [16]. In practice, however, these characteristics do not hold as shown by Bengio, et al. [6]. The motivation behind developing the LSTM was to remove the vanishing gradient issues that occur with RNNs when processing long-term dependencies. The standard RNN consists of a chain of repeating modules of the neural network, where each module consists of a single hyperbolic tangent layer structure. This can be compared with the LSTM module structure. It is relatively more complex where each module consists of four layers rather than a single layer as for an RNN module. LSTM modules or memory blocks consist of an input gate, a forget gate, an output gate and the cell state. All these layers interact in a particular way, see e. g. [30] for more details. Information that will be added or removed to the cell state is controlled by three gates. Different combinations of these gates can be used for the memory cells to deal with data with a longer horizon.

Bidirectional Long-Short-Term-Memory (BI-LSTM) shares some similarities in terms of the mechanisms as a bidirectional RNN [38] where the data sequence is fed in both forward and backward directions using two separate hidden layers. These are then connected to an output layer. The typical unrolled architecture of a bidirectional LSTM consists of a forward LSTM layer and a backward LSTM layer. The forward layer output sequence is calculated using inputs in a forward sequence from time $t-n$ to time $t-1$. The backward layer output sequence is calculated using the reversed sequence from time $t-1$ to $t-n$. Outputs of both layers are calculated by using the standard LSTM equations [30]. An output vector $\boldsymbol{Y_n}$ is generated from forward and backward LSTM layers using $\boldsymbol{Y_n} = \sigma(\boldsymbol{h_n}, \boldsymbol{h_n})$ where, n is the time step, σ is a function used to combine the outputs of forward and backward LSTM layer. It can be a concatenating function, a summation function, an average function or a multiplication function. Concatenate is considered here for the model.

Convolutional Neural Networks are feed-forward artificial neural networks consisting of alternating convolutional and sub-sampling layers, comparable with simple and complex cells in the human visual cortex. CNNs can be considered to mimic the human visual system and have achieved state-of-the-art performance with recognising patterns, structures and other functions such as tracking [21]. CNNs were primarily developed for 2D signals but recently 1D CNNs have been used for applications such as ECG classification [19], structural health monitoring [1] and motor-fault detection [15]. 1D CNNs have a very simple structure and can be trained with a limited amount of data compared to 2D CNNs. CNNs are a very good candidate in time series forecasting because of the filter feature extraction and composition ability. CNNs are also easier to train in comparison to RNNs because CNNs use convolution operations as opposed to recursion. The sliding window approach used with RNNs could also be used to train CNNs for time series forecasting.

4 Methodology

Time series forecast analyses on physiological data sets (BP and HR) are performed with 10 forecast models by combining two machine learning algorithms each with five forecast strategies. This has been considered to find the best forecast strategy in physiological time series forecasting. Following this, time series forecasting on physiological data sets are performed to find the best approach between univariate and multivariate data. Here, 12 forecast models are considered by combining three machine learning algorithms, two forecast strategies each implemented for both the univariate and multivariate cases. Forecast performance of all models in forecasting blood pressure are compared.

4.1 Data Set

The MIMIC II database is a freely accessible critical care database which consists of various vital sign information. Patients' data were collected from a variety of ICUs in Beth Israel Deaconess Medical Center in Boston, Massachusetts [37]. Advantages of using the MIMIC II database is that all data are anonymized and open to researchers. For this paper, minute by minute MAP and heart rate time series data of 30 patients have been extracted. Experimental data sets were selected from the hypotension group of an ICD-9 code. The MAP is a measure of blood pressure [28] which is calculated from systolic and diastolic pressure following the equations 1.

$$MAP = [2(DP) + SP]/3 \tag{1}$$

During the data selection procedure here, it was ensured that there were no missing values in the selected time series. Moreover, length of the time series data of each data set are variable. The time series were scaled to values between -1 and 1, this is because LSTM and BI-LSTM algorithms require data to be within the scale of the activation function of the network [31].

Supervised learning is very common in practical machine learning. In supervised learning, a machine learning algorithm is used to learn the mapping function between the input variables (X) and output variables (Y). The aim is to teach the model well during the training process so that for a new input data (X), the model can predict the output variables (Y) for that data. To undertake supervised learning, a time series data set needs to be processed to a form that can be used in a supervised training process. Extracted time series data has been used to create samples for the prediction models. Each sample consisted of two-time intervals; observation window (X) and target window (Y) achieved through the sliding window method, see e. g. [9] for more details. The observation window is also known as the input and its size in the sample depends on the user-defined sequence (for this work 30 minutes observation window is considered). The target window is known as the output and its size in the sample depends on the forecast strategies. For example, a forecast model with a MIMO forecast strategy to predict a 30 minute window will require a 30 minute target window (Y). This process was applied to all 30 patients' time series data.

Both blood pressure and heart rate are considered for the univariate case. The univariate time series data sets are converted to supervised data sets consisting of samples. The samples include an observation window (X) of 30 minutes of BP or HR for all strategies. However, the target window (Y) varies following different strategies. For example, the MIMO strategy requires 30 minutes of BP or HR as a target window whereas recursive strategy only requires 1 minute of BP or HR as the target window.

The source of the multivariate data are blood pressure and heart rate time series data which are used to forecast future blood pressure. The multivariate time series data set is converted to supervised data set consist of samples. The samples consist of an observation window of 60 minutes which includes 30 minutes of BP and HR each. The 60 minutes of observations (X) is consistent for all strategies. However, the target window (Y) varies depending on the strategy and the response variable. For example, if BP is the response variable and MIMO is the forecast strategy then the target window requires 30 minutes of BP values.

4.2 Forecast Model Formulation

First, the forecast models were built to investigate the scope of forecasting strategies in forecasting physiological time series data. LSTM and BI-LSTM algorithm are used here in conjunction with the aforementioned time series forecasting strategies to build the forecast models. This gives a total of 10 forecast models to compare the forecast strategies performance on univariate physiological time series data (HR and BP). Then following the outcome of different forecast strategies, 12 forecast models were built to investigate the scope of univariate and multivariate approaches in forecasting physiological time series data. LSTMs, BI-LSTMs and CNNs are used here in conjunction with the best two forecast strategies (MIMO and DIRMO).

The forecast models with LSTM and BI-LSTM RNNs are designed here with a network structure consisting of 1 hidden layer with 10 LSTM units, then an output layer with a hyperbolic tangent activation and target window (Y) as output values which varies following the forecasting strategy. LSTM is stateful in the designed network and the network was fitted with 5 epochs. The forecast model with the CNN algorithm was designed with a network structure consisting of one convolutional hidden layer followed by a max pooling layer. The filter maps are then flattened before being interpreted by a dense layer and outputting a prediction. The output layer consisting of a tanh activation and output values equal to the target window (Y). The batch size of the networks was set to 1.

The target window was varied according to the number of time steps over which a forecast was required depending on the forecasting strategy. The number of neurons in the output layer also differs depending on the forecasting strategy.

Iterative, Direct and DirRec strategies predict one step ahead at a time so one neuron is required at the output layer of the model. Whereas, MIMO and DIRMO models predict multiple points so more than one neuron is required at the output layer of the model. More specifically in MIMO, the number of neurons in the output layer is equal to the number of predictions needed in each regression. For DIRMO, the number of neurons in the output layer are calculated by dividing the number of prediction points by the number of models. The network also uses the Root Mean Squared Error (RMSE) as a loss function and the ADAM algorithm [18] as an optimiser. The parameters of the developed forecast models were not tuned and random parameters were set. This is because the experiments in this paper were not performed for a specific medical problem. Rather the main aim of the experiments was to compare the univariate and multivariate approaches along with forecast strategies and machine learning algorithms. However, parameter tuning is essential when physiological time series forecasting is applied to specific medical applications.

All models were developed using the Python ecosystem [31]. To perform forecasting using the models, each data set were split into train and test. The aim is to forecast 30 minutes in advance and rest of the samples were used to train the models. Samples consisting of the last 30 minutes of data of the data set were used for test. The test data were used for performance characterisation. In Direct, DirRec and DIRMO strategies the training data are used multiple times as these strategies require multiple models to forecast the required target window (Y). In the testing phase, the predict function of the model is called to make predictions on given input values (X). The forecasting process varies following different strategies. The Recursive and DirRec strategies use past predicted data to feed back into the model during the multi-step forecast. MIMO, DIRMO and DirectH strategies are not recursive strategies so the predicted data do not feedback to the model. The predicted value is then re-scaled to get the actual predicted output. To measure the performance of the models, the RMSE is calculated here. using $\bar{e} = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(y_i - \widehat{y}_i)^2}$ where, y_i are actual values, $\widehat{y}_i$ are forecast values and m is the number of target output data. The standard deviation of the RMSE was also calculated.

4.3 Results

The aim of all the developed models is to forecast physiological data (HR and BP) 30 minutes in advance. In comparing the forecast strategies, average RMSE and the RMSE standard deviation of patients following all models are shown in Fig. 1. Average RMSE and the RMSE standard deviation of patients following all models are shown in Fig. 2 in comparing the univariate and multivariate approaches. To compare DIRMO and MIMO based strategies in detail, average

RMSE and the RMSE standard deviation of patients of univariate and multivariate approaches are also plotted in Fig. 3.

Comparison of Forecast Strategies To compare forecast strategies, the performance of the 10 different forecast models was assessed by forecasting HR and BP. Average RMSE and the RMSE standard deviation of patients following forecast models combining LSTM and BI-LSTM algorithm along with all aforementioned strategies are shown in Fig. 1 in comparing the forecast strategies. It can be observed that the MIMO and DIRMO forecast strategies appear to exhibit lower RMSE forecast performance. Traditional strategies like Recursive, Direct and DirRec forecast performance were poor compared to MIMO and DIRMO strategy in forecasting both Blood Pressure and Heart Rate. Following the outcome, MIMO and DIRMO forecast strategies were considered in this paper to develop forecast models for different other comparison scenarios.

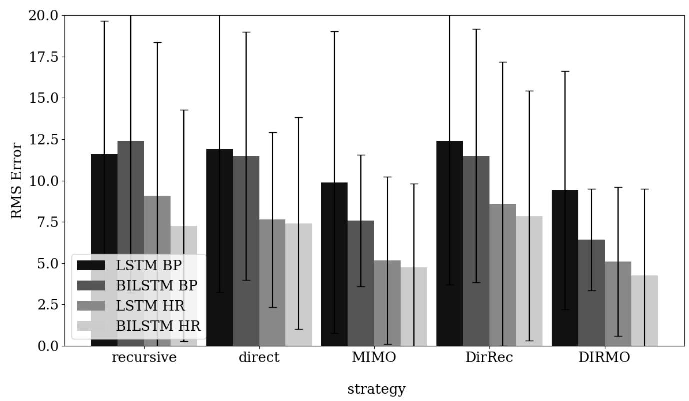

Fig. 1: Performance of forecast models with different forecast strategies in forecasting Blood Pressure (BP) and Heart Rate (HR) 30 minutes in advance.

Comparison of Forecasting with Univariate and Multivariate Data Twelve forecast models have been tested combining 3 machine learning algorithms and 2 forecast strategies, covering both univariate and multivariate techniques. Average RMSE and the RMSE standard deviation from all patients are taken into consideration and are shown in Fig. 2. Careful observation of the average RMSE and the RMSE standard deviation of all patients appears to show that the multivariate approach outperforms the univariate approach in all forecast scenario. So overall it appears that multivariate techniques can provide better performance in forecasting blood pressure 30 minutes in advance compared to univariate techniques.

The comparison shown in Fig. 2 includes the results for the DIRMO and MIMO forecasting strategies. The results are shown differently in Fig. 3 to en-

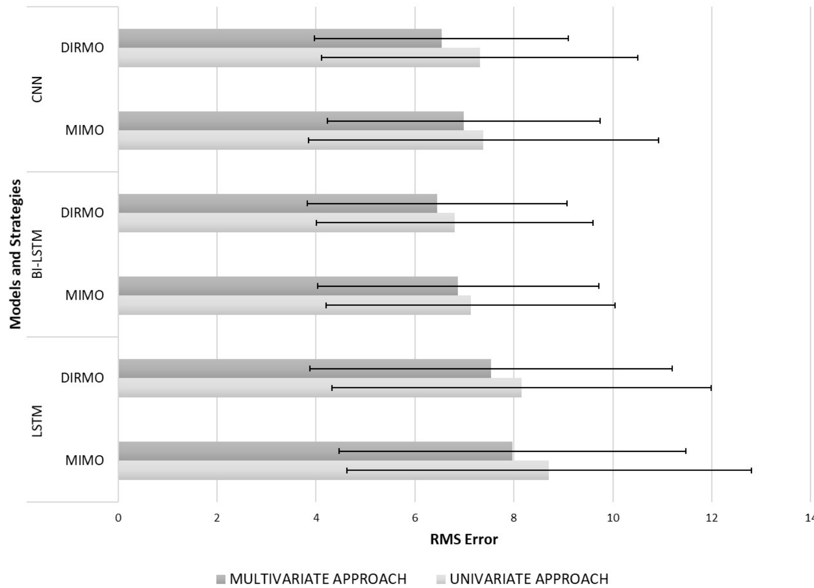

Fig. 2: Comparison of univariate and multivariate cases considering the average RMSE and standard deviation of all patients

able easier comparison between these two strategies. These show that DIRMO forecasting strategy provides lower RMSE and less standard deviation in comparison to MIMO forecasting in all scenarios. Thus, it can be tentatively concluded that DIRMO forecasting performs better than MIMO based forecasting in forecasting blood pressure 30 minutes in advance.

Best Forecasting Model can be considered given these empirical results seen here. It is observable in Fig. 2 that the forecast model with the BI-LSTM algorithm outperforms the CNN and LSTM algorithms. From Fig. 2 it is also observable that the forecasting model with the BI-LSTM algorithm along with a DIRMO strategy and multivariate configurations performs best. Moreover, in Fig. 2 and Fig. 3 it appears that the forecast model with the BI-LSTM algorithm outperforms the models with the CNN and LSTM algorithms. This is true for both the univariate and multivariate approaches. Overall, it can therefore be tentatively concluded that BI-LSTM forecasting using DIRMO strategy is the best model. Furthermore the multivariate case seems to enhance the results even further.

So far only a single forecast horizon has been considered for all forecast models. This has involved forecasting of physiological data which was 30 minutes. However, a further range of forecast horizons were also considered. Forecast horizon up to 2 hours with intervals of 10 minutes were also considered for forecasting BP with the multivariate data. The forecasting error across this range is shown in Fig. 4 for the best forecast model (BI-LSTM algorithm, DIRMO strategy, multivariate data). It is observable that, as the forecast horizon increases,

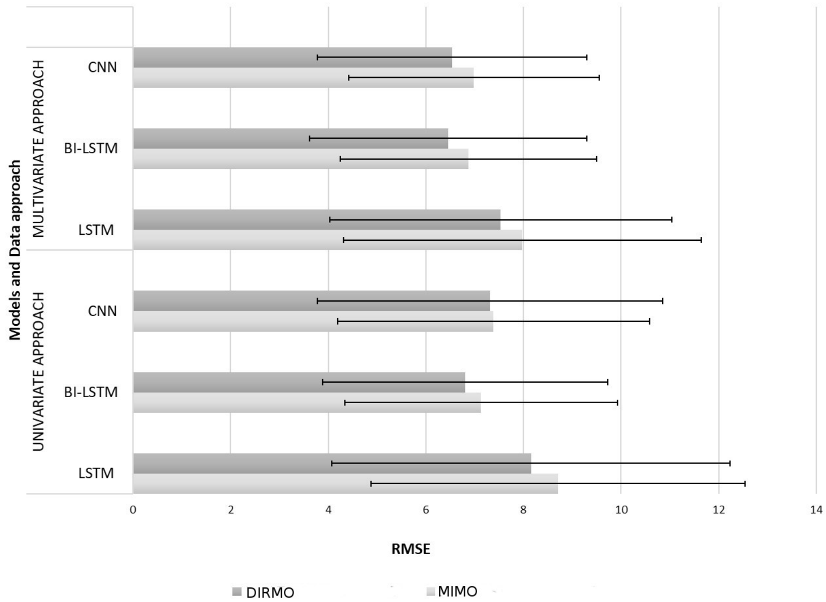

Fig. 3: Comparison of DIRMO and MIMO forecasting strategies considering using RMSE and standard deviation of all patients.

the RMSE increases. Furthermore, from 60 to 120 minutes the RMSE increases linearly.

5 Discussion

Analysis of different forecast strategies can help in the selection of the best forecast strategy. This is important when building a forecast model for a particular application such as for forecasting of physiological data. The results appear to suggest the best performing strategy is DIRMO. This appears to be in agreement with the work of Taieb et al. [42]. There, various different types of MIMO and DIRMO were compared along with DIR, REC and DIRREC forecasting strategies. There DIRMO was also found to outperform MIMO but not in all cases: only when no input selection had taken place which is also the case here. Multivariate forecasting with machine learning algorithms and forecast strategies in forecasting of blood pressure might also have another potential advantage. The inter-relationship between blood pressure and heart rate can also be considered. The dependency between the time series on each other is implicitly modelled here for the multivariate approaches. This has resulted in improved forecast accuracy in forecasting blood pressure. Lee and Mark in [23] also made use of HR and BP. They also included a number of different manually derived measures derived from either BP, HR or a combination of both. However they did not

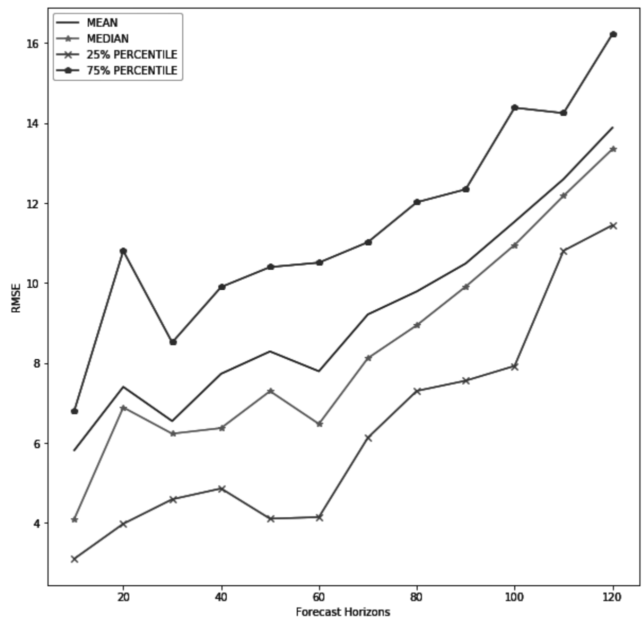

Fig. 4: RMSE of forecasting of blood pressure for different forecast horizons using BI-LSTM with DIRMO strategy with multivariate data.

provide any comparison with forecasting for the univariate case. Their investigation also only considered a single forecast strategy and machine learning model. The multivariate approach presented here, which uses the BI-LSTM algorithm and DIRMO forecast strategy has proved to be the best model in forecasting blood pressure 30 minutes in advance. Such modelling might be a useful technique which could potentially help health care professionals in planning, decision making and predicting the event.

The multivariate approach provides insight into the dynamic relationships of the used variables but in such cases, more variables, data points and data sets are required. Data points of all variables need to be measured at the same time period and this is the subject of ongoing work.

6 Conclusions

The results shown here appear to demonstrate that multivariate time series modelling is more reliable in forecasting blood pressure 30 minutes in advance. The multivariate approach along with a BI-LSTM algorithm and DIRMO strategy provide more accurate forecasting performance. This is in comparison to the univariate approaches and the other machine learning algorithms and forecast strategies. It is also observed that the BI-LSTM algorithm with a DIRMO strategy provides the smallest standard deviation which makes this combination favourable for forecasting blood pressure. Overall, the comparison of forecast strategies and data approaches contributes to improving the current techniques in the applied forecasting and machine learning literature.

References

1. Osama Abdeljaber, Onur Avci, Serkan Kiranyaz, Moncef Gabbouj, and Daniel J Inman. Real-time vibration-based structural damage detection using one-dimensional convolutional neural networks. *Journal of Sound and Vibration*, 388:154–170, 2017.
2. Patrick Aboagye-Sarfo, Qun Mai, Frank M Sanfilippo, David B Preen, Louise M Stewart, and Daniel M Fatovich. A comparison of multivariate and univariate time series approaches to modelling and forecasting emergency department demand in western australia. *Journal of biomedical informatics*, 57:62–73, 2015.
3. Nesreen K Ahmed, Amir F Atiya, Neamat El Gayar, and Hisham El-Shishiny. An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6):594–621, 2010.
4. Moez Baccouche, Franck Mamalet, Christian Wolf, Christophe Garcia, and Atilla Baskurt. Sequential deep learning for human action recognition. In *International workshop on human behavior understanding*, pages 29–39. Springer, 2011.
5. Jules Bassale. Hypotension prediction arterial blood pressure variability. *Tecnhical report*, 2001.
6. Yoshua Bengio, Patrice Simard, Paolo Frasconi, et al. Learning long-term dependencies with gradient descent is difficult. *IEEE transactions on neural networks*, 5(2):157–166, 1994.
7. Victor L Berardi and Guoqiang Peter Zhang. An empirical investigation of bias and variance in time series forecasting: modeling considerations and error evaluation. *IEEE Transactions on Neural Networks*, 14(3):668–679, 2003.
8. Antonis Billis and Panagiotis D Bamidis. Employing time-series forecasting to historical medical data: an application towards early prognosis within elderly health monitoring environments. In *AI-AM/NetMed@ ECAI*, pages 31–35. Citeseer, 2014.
9. Thomas G Dietterich. Machine learning for sequential data: A review. In *Joint IAPR international workshops on statistical techniques in pattern recognition (SPR) and structural and syntactic pattern recognition (SSPR)*, pages 15–30. Springer, 2002.
10. Johann Du Preez and Stephen F Witt. Univariate versus multivariate time series forecasting: an application to international tourism demand. *International Journal of Forecasting*, 19(3):435–451, 2003.
11. Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
12. André Gensler, Janosch Henze, Bernhard Sick, and Nils Raabe. Deep learning for solar power forecastingan approach using autoencoder and lstm neural networks. In *2016 IEEE international conference on systems, man, and cybernetics (SMC)*, pages 002858–002865. IEEE, 2016.
13. Tim Hill, Marcus O'Connor, and William Remus. Neural network models for time series forecasts. *Management science*, 42(7):1082–1092, 1996.
14. Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.
15. Turker Ince, Serkan Kiranyaz, Levent Eren, Murat Askar, and Moncef Gabbouj. Real-time motor fault detection by 1-d convolutional neural networks. *IEEE Transactions on Industrial Electronics*, 63(11):7067–7075, 2016.

16. Lakhmi C Jain and Larry R Medsker. *Recurrent neural networks: design and applications*. CRC Press, 2000.
17. Amin Janghorbani, Abdollah Arasteh, and Mohammad Hassan Moradi. Prediction of acute hypotension episodes using logistic regression model and support vector machine: A comparative study. In *2011 19th Iranian Conference on Electrical Engineering*, pages 1–4. IEEE, 2011.
18. Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
19. Serkan Kiranyaz, Turker Ince, and Moncef Gabbouj. Real-time patient-specific ecg classification by 1-d convolutional neural networks. *IEEE Transactions on Biomedical Engineering*, 63(3):664–675, 2016.
20. John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.
21. Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.
22. Joon Lee and RG Mark. A hypotensive episode predictor for intensive care based on heart rate and blood pressure time series. In *2010 Computing in Cardiology*, pages 81–84. IEEE, 2010.
23. Joon Lee and Roger G Mark. An investigation of patterns in hemodynamic data indicative of impending hypotension in intensive care. *Biomedical engineering online*, 9(1):62, 2010.
24. Xiaohan Li, Shu Wu, and Liang Wang. Blood pressure prediction via recurrent models with contextual layer. In *Proceedings of the 26th International Conference on World Wide Web*, pages 685–693. International World Wide Web Conferences Steering Committee, 2017.
25. Zachary C Lipton, John Berkowitz, and Charles Elkan. A critical review of recurrent neural networks for sequence learning. *arXiv preprint arXiv:1506.00019*, 2015.
26. Zachary C Lipton, David C Kale, Charles Elkan, and Randall Wetzel. Learning to diagnose with lstm recurrent neural networks. *arXiv preprint arXiv:1511.03677*, 2015.
27. Xiaolei Ma, Zhimin Tao, Yinhai Wang, Haiyang Yu, and Yunpeng Wang. Long short-term memory neural network for traffic speed prediction using remote microwave sensor data. *Transportation Research Part C: Emerging Technologies*, 54:187–197, 2015.
28. Douglas L Mann, Douglas P Zipes, Peter Libby, and Robert O Bonow. *Braunwald's Heart Disease E-Book: A Textbook of Cardiovascular Medicine*. Elsevier Health Sciences, 2014.
29. Shamsul Masum, Ying Liu, and John Chiverton. Comparative analysis of the outcomes of differing time series forecasting strategies. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*, pages 1964–1968. IEEE, 2017.
30. Shamsul Masum, Ying Liu, and John Chiverton. Multi-step time series forecasting of electric load using machine learning models. In *International Conference on Artificial Intelligence and Soft Computing*, pages 148–159. Springer, 2018.
31. Wes McKinney. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. " O'Reilly Media, Inc.", 2012.
32. Kylie McLACHLAN, Alicia Jenkins, and David ONEAL. The role of continuous glucose monitoring in clinical decision-making in diabetes in pregnancy. *Australian and New Zealand journal of obstetrics and gynaecology*, 47(3):186–190, 2007.

33. Phuoc Nguyen, Truyen Tran, and Svetha Venkatesh. Deep learning to attend to risk in icu. *arXiv preprint arXiv:1707.05010*, 2017.
34. Femke Ongenae, Stijn Van Looy, David Verstraeten, Thierry Verplancke, Dominique Benoit, Filip De Turck, Tom Dhaene, Benjamin Schrauwen, and Johan Decruyenaere. Time series classification for the prediction of dialysis in critically ill patients using echo statenetworks. *Engineering Applications of Artificial Intelligence*, 26(3):984–996, 2013.
35. Steffen Rendle. Factorization machines. In *2010 IEEE International Conference on Data Mining*, pages 995–1000. IEEE, 2010.
36. Teresa Rocha, Simao Paredes, Paulo De Carvalho, and Jorge Henriques. Prediction of acute hypotensive episodes by means of neural network multi-models. *Computers in biology and medicine*, 41(10):881–890, 2011.
37. Mohammed Saeed, Christine Lieu, Greg Raber, and Roger G Mark. Mimic ii: a massive temporal icu patient database to support research in intelligent patient monitoring. In *Computers in cardiology*, pages 641–644. IEEE, 2002.
38. Mike Schuster and Kuldip K Paliwal. Bidirectional recurrent neural networks. *IEEE Transactions on Signal Processing*, 45(11):2673–2681, 1997.
39. Costas Sideris, Haik Kalantarian, Ebrahim Nemati, and Majid Sarrafzadeh. Building continuous arterial blood pressure prediction models using recurrent networks. In *2016 IEEE International Conference on Smart Computing (SMARTCOMP)*, pages 1–5. IEEE, 2016.
40. Alex J Smola and Bernhard Schölkopf. A tutorial on support vector regression. *Statistics and computing*, 14(3):199–222, 2004.
41. Peng Su, Xiao-Rong Ding, Yuan-Ting Zhang, Jing Liu, Fen Miao, and Ni Zhao. Long-term blood pressure prediction with deep recurrent neural networks. In *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 323–328. IEEE, 2018.
42. Souhaib Ben Taieb, Gianluca Bontempi, Amir F Atiya, and Antti Sorjamaa. A review and comparison of strategies for multi-step ahead time series forecasting based on the nn5 forecasting competition. *Expert systems with applications*, 39(8):7067–7083, 2012.
43. Paul M Vespa, Val Nenov, and Marc R Nuwer. Continuous eeg monitoring in the intensive care unit: early findings and clinical efficacy. *Journal of Clinical Neurophysiology*, 16(1):1–13, 1999.
44. Peng Zhou, Wei Shi, Jun Tian, Zhenyu Qi, Bingchen Li, Hongwei Hao, and Bo Xu. Attention-based bidirectional long short-term memory networks for relation classification. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, volume 2, pages 207–212, 2016.
45. Yao Zhu, Xiaoliang Fan, Jinzhun Wu, Xiao Liu, Jia Shi, and Cheng Wang. Predicting icu mortality by supervised bidirectional lstm networks. In *AIH@IJCAI*, 2018.