Enhancing early warning systems: predicting next vital signs using recurrent neural networks and attention models

Enhancing Early Warning Systems: Predicting Next Vital Signs Using Recurrent Neural Networks and Attention Models

Abstract

Vital signs have proven to be highly precise indicators of patient deterioration. To accurately identify high-risk patients in hospital wards, early warning scores have been introduced. However, these scores often rely solely on current vital sign readings and seldom incorporate trends in vital signs over time. In this work, the prediction of patient vital signs which include Diastolic Blood Pressure, Systolic Blood Pressure, Heart Rate, Respiratory Rate, Oxygen Saturation and Temperature over the MIMIC-III dataset involves utilizing past vital signs, patient demographic data, and admission details for the next early warning score prediction. Various deep learning models were trained to perform multi task learning for this purpose, leveraging clinical data available prior to the admission diagnosis. The resultant model demonstrates strong predictive performance, showcasing its robust capabilities in forecasting forthcoming vital signs. Among the deep learning models utilized Long Short Term Memory (LSTM), Bidirectional Long Short Term Memory (BiLSTM), Multi-head Attention, and Long Short Term Memory-Attention (LSTM-ATTN), among these the LSTM-ATTN model yielded the most promising outcomes. It achieved a total mean squared error of 0.0022, surpassing the performance of the other three models. This underscores its potential for deployment in deployment not only in hospital settings but also in the context of virtual ward management, where real-time prediction of patients' next vital signs and early detection of deterioration can be invaluable.

Keywords: Vital Signs, Bidirectional Long Short Term Memory, Long Short Term Memory, Multi-head Attention, Attention, Early Warning Score, Virtual Wards.

1. Introduction

Vital signs play a crucial role in monitoring a patient's health status and can provide valuable insights into their recovery or deterioration. It is essential for identifying early warning signs of deterioration, assessing treatment response, detecting complications, guiding clinical decision-making, and monitoring patient recovery. Patient decline and negative outcomes are often heralded by unusual vital signs. These indicators typically manifest hours to days before the event such as infection, death or a cardiac arrest (Barfod et al., 2012), allowing for timely intervention. A patient's vital signs are typically monitored continuously using a bedside monitoring system, supplemented by intermittent manual checks by clinical staff. The bedside monitor employs straightforward thresholds for each vital sign, like a heart rate (HR) below 40 bpm, to identify patient abnormalities. These thresholds are often established based on clinical expertise or derived from data on stable patients worldwide. The irregularities in vital signs have consistently been noted in patients preceding significant clinical occurrences like infection, cardiac arrest, and mortality. Consequently, early warning systems (EWS) (Petersen, 2018), have emerged. These systems monitor vital signs at regular intervals and use predefined criteria to alert clinicians of patient deterioration.

EWS commonly monitors Heart Rate (HR), Respiratory Rate (RR), Blood Pressure (BP), Peripheral Oxygen Saturation (SpO2), Temperature, and sometimes Consciousness level. Aggregate-weighted EWS combines multiple vital signs and patient characteristics, each assigned a weight based on predefined thresholds. An overall risk score is then calculated by summing the weighted scores.

There are several early warning score system such as National early warning score (NEWS) (Smith et al., 2019), Centile-based early warning score (CEWS) (Tarassenko et al., 2011), Modified Early Warning Score (MEWS) (Gardner-Thorpe et al., 2006), Manual centile-based early warning scores (MCEWS) (Watkinson et al., 2018), Age-based early warning score (AEWS), and Hamilton Early Warning Score (HEWS) (Tam et al., 2017). Various EWS have been utilized in the United Kingdom (UK) to date, among which the National Early Warning Score (NEWS) has demonstrated superiority over 33 other EWS (Kolic et al., 2015). NEWS is designed as a scoring mechanism aimed at preventing and promptly identifying patients at risk of developing or presenting with acute illness. In this study we are focusing on forecasting the next vital sign so that we can perform the next NEWS score calculation an hour ahead of time.

However, aggregate-weighted EWS have limitations. They provide a snapshot of the patient's current risk but fail to capture trends or predict future risk trajectories. Moreover, these scores do not consider correlations between parameters, as each parameter's score is independently calculated, disregarding potential interactions. Furthermore, most of the works that aim to forecast these vital signs or NEWS scores do not consider other factors such as patient demographics, admission details, and medical history.

In this study, we aim to overcome the noted limitation in EWS score calculation, alterations therein in response to abrupt pattern shifts, allowing medical professionals to devise patient treatment plans accordingly. For this purpose, we employed various deep learning models to conduct multitask learning, predicting upcoming vital signs for NEWS score computation. The dataset used incorporates static patient demographics, admission details, and dynamic patient vital signs predicting the anticipated trajectory of a patient's forthcoming vital signs which could enable medical personnel to monitor deviations from the anticipated pattern. Such observations could assess the likelihood of physiological decline, enabling clinicians to tailor an optimal treatment strategy. Figure 1 provides a comprehensive depiction of a patient's journey from hospital admission to the prediction of vital signs.

2. Literature Review

Using machine learning algorithms on vital signs or Early Warning Scores (EWS) to predict medical outcomes like mortality (Alghatani et al., 2021), chronic respiratory disease (Youssef Ali Amer et al., 2021), cardiac arrest (Ahmed et al., 2023), sepsis (Barton et al., 2019) and other emergencies can be considered crucial, especially in emergency department settings where longer waits can cause fatal for a patient. In situations such as disasters or mass casualty incidents, where demand exceeds available resources, this technology could become essential.

The NEWS has been widely validated in acute medical and pre-hospital settings, highlighting its effectiveness, yet there is an opportunity to further explore its application in the postoperative surgical population. To address this, researchers (Chiu et al., 2020) employed simple logistic regression to model the relationship between the NEWS physiological variables and the likelihood of a serious patient event within the following 24 hours for cardiac surgical population and also assessed the discriminatory power of each model for predicting events in the next 6 or 12 hours. The study found that a logistic version of the NEWS, as opposed to the current additive model, more effectively discriminates patients after cardiac surgery who may die, experience cardiac arrest, or require unplanned readmission to intensive care.

Vital signs are crucial for assessing patient risk, as specific patterns of abnormalities emerge due to strong correlations between multiple vital signs. By considering short-term summary statistics and correlations of all vital signs simultaneously, Forkan and Khalil, 2017 generated a feature vector suitable for multi-label classifiers. These features are then used to build machine learning models that predict short-term vital sign threshold values reducing false alerts in monitoring stations and aiding in the early detection of clinical dangers.

Intelligent monitoring solutions are necessary to efficiently manage hospital resources, particularly within crowded intensive care units (ICUs), a challenge exacerbated during the global COVID-19 pandemic. One such work by Youssef et al., 2021 used machine learning models to predict future vital sign values in COVID-19 ICU patients, focusing on heart rate, respiration rate, and oxygen saturation. Different approaches were tested, including models considering multiple vital signs and those focusing solely on the three mentioned. Results showed acceptable prediction performance. These models offer potential for integration into monitoring systems to provide real-time health condition predictions for COVID-19 ICU patients with a limited set of vital signs.

In our study, we didn't just rely on patient vital signs; we also incorporated patient demographics and admission particulars like admission type (Emergency, Newborn, Elective, or Urgent), length of stay, and Insurance. These factors are crucial for predicting future vital signs using multitask learning, allowing

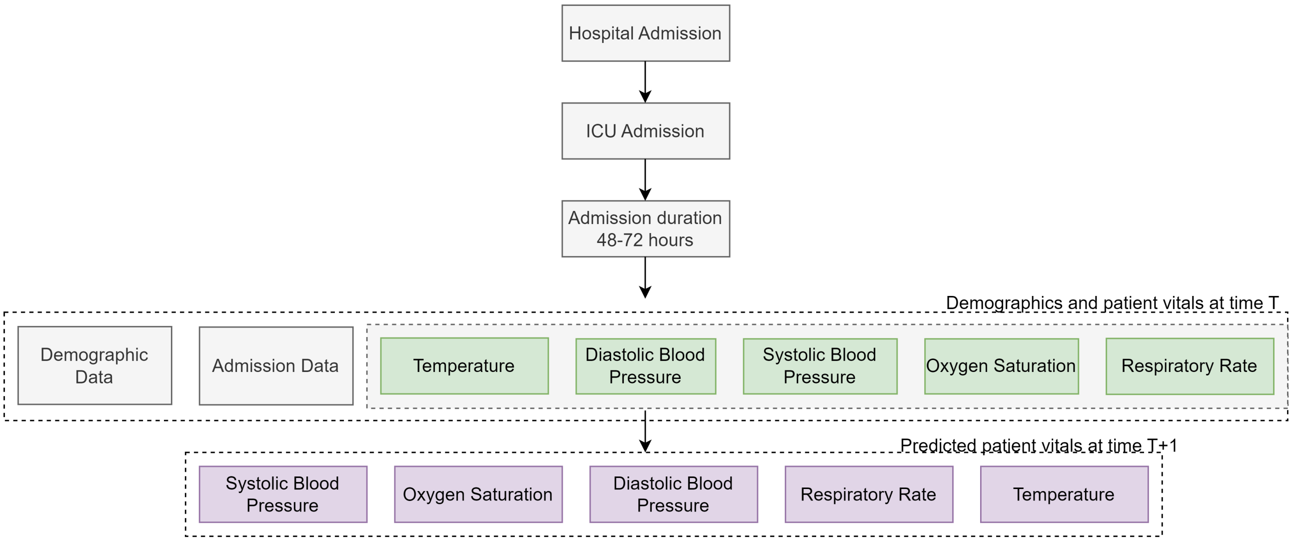

Figure 1. Patient Journey from admission to vital signs prediction.

for a patient-focused prediction. This approach will inturn enable us to forecast NEWS score one hour in advance, ensuring timely treatment administration in not just hospital setting but can be used in virtual ward also.

3. Methodology

In this section, we delineate our methodology encompassing the entire process from pre-processing to model architecture and training. We provide a comprehensive overview of our approach, detailing the steps involved in preparing the data, designing the model architecture, and training the models. This section serves as a foundational framework for understanding our research methodology and the subsequent analysis of results.

3.1. Dataset

The MIMIC-III dataset (Johnson et al., 2016) comprises extensive medical records freely available for research purposes. It encompasses detailed hospital stay data of patients admitted to the intensive care units of the Beth Israel Deaconess Medical Center, a tertiary care hospital, spanning from 2001 to 2012. The dataset contains a wide range of information, including deidentified patient demographics, hourly bedside vital signs, and clinical notes in free-text format. The data was extracted from the MIMIC tables using BigQuery, followed by extensive pre-processing and then forecasting of the vital signs of patient data.

Inclusion Criteria: We hold the view that acute care events should manifest during the ICU stay, and the sooner they can be anticipated, the more beneficial they are. Hence, our emphasis is on extracting the initial 48-72 hours of data for forecasting future vital signs, rather than utilizing the entire admission dataset. Out of all the patients we chose the patients we chose the patients whose available vital signs were in the range of 48-72 hours so that we get an optimum range for predicting the future vital signs for the NEWS score calculation.

3.2. Preprocessing

This section encompass data extraction, cleansing, missing value imputation, and structuring the data according to the problem statement. Typically, data processing is performed using Python libraries like Pandas and NumPy. The initial phase involves extracting data from the source system, this was executed using SQL due to the MIMIC-III dataset's hosting on Google BigQuery. We used the data from dynamic and static patient data by using the information from Admissions, chartevents and d_items table.

A vital aspect of the forecasting model is obtaining the patient's vital signs, such as systolic blood pressure, diastolic blood pressure, temperature, and heart rate. Subsequently, we perform data cleansing, which entails standardizing vital signs to a single unit of measurement. For instance, temperature readings may be recorded in Fahrenheit or Celsius, requiring conversion using appropriate formulas.

After standardization, outlier handling becomes important to ensure the accuracy of forecasting models. Outlier ranges are defined for each vital sign a separate table as mentioned in the work Wang et al., 2020, enabling the replacement of outlier values with null or valid high/low values. The missing values are also tackled by repeating the previous values for a specific patient. Furthermore, as vital readings are recorded at varying hours and frequencies for each ICU patient, this makes aggregation necessary. Aggregating data involves computing mean vital readings for each hour per patient. Following aggregation, data pivoting is conducted to transform each vital sign into a column, with rows representing readings at specific hours during an ICU stay. This restructuring prepares the data for training machine learning algorithms. Following pre-processing, the sequence length for the time series data was kept 48.

3.3. Model Architecture

Upon further study different types of data were gathered for this model including demographics such as Hospital admission ID (HADM_ID) which is our primary key, 'hour from intime (which keeps count of instances of vital signs recording), icustay id, Age, Length of total stay, Admission type (Emergency, elective, Newborn, or urgent) Gender of the patient (Male or Female), Religion (Catholic, Not specified, Unobtainable, and Others), Insurance (Medicare, Private, Medicaid, or others), Marital Status (married, single, unknown, or others), Language and ethnicity. The vital signs or output include diastolic blood pressure, heart rate, oxygen saturation, systolic blood pressure, respiratory rate, and temperature. Figure 2 illustrates the model architecture utilized in our study, featuring a two-layered design with distinct dense layers dedicated to each vital sign forecasting task.

3.4. Model Training

Several models, including LSTM, BiLSTM, multi-head attention, and LSTM-attention, were utilized in this study for implementation. Each of these models was trained at 15 epochs, and the hyperparameters, such as learning rate, sequence length, and evaluation metrics, were kept the same for a fair comparison.

Hardware: The training process utilized a computing infrastructure consisting of a Google Collaboratory Pro version, equipped with a V100 GPU and 50GB of RAM. This configuration allowed for efficient training of the model, leveraging the computational power necessary for complex deep learning tasks. Detailed description and structure of these models are discussed below

- **LSTM**

  An LSTM, which stands for Long Short-Term Memory (Elsworth and Güttel, 2020), is a form of recurrent neural network (RNN) proficient at analysing sequences of information such as text, speech, or time series data. Its name, "Long Short-Term Memory", reflects its capability to manage both short-term and long-term dependencies within the data.

  Memory Cells: The core element of an LSTM lies in its memory cells, which are capable of retaining information over prolonged duration. These cells aid the model in learning from sequences containing intervals between significant occurrences.

  Gates: LSTMs are equipped with three types of gates regulating the inflow and outflow of information within the memory cells:

  - *Forget Gate:* Determines the data to discard from the memory cells.

  - *Input Gate:* Determines the new information to incorporate into the memory cells.

  - *Output Gate:* Determines the information to output from the memory cells.

  From the current input word representation x_i, the previous hidden state h_{i-1}, and the preceding memory cell n_{i-1}, the current hidden state h_i and memory cell n_i at time step i are generated. The feature vector e_i, forget gate f_i, output gate o_i, and input gate i_i are defined in equations (1) – (6).

  $$i_i = Sigmoid(W_j x_i + U_j h_{(i-1)} + b_i) \quad (1)$$

  $$f_i = Sigmoid(W_f x_i + U_f h_{(i-1)} + b_f) \quad (2)$$

  $$O_i = Sigmoid(W_o x_i + U_o h_{(i-1)} + b_o) \quad (3)$$

  $$e_i = \tanh(W_g x_i + U_g h_i + b_g) \quad (4)$$

  Where W and U are weights and b is the bias. The hidden state hi and current state ni are calculated as

  $$n_i = f_i \bigodot n_i + i_i \bigodot g_i \quad (5)$$

  $$h_i = O_i \bigodot \tanh(C_{(i-1)}) \quad (6)$$

  Training: Throughout the training process, the LSTM adjusts the parameters of its gates to effectively retain crucial information while disregarding irrelevant details within the input sequences.

- **Multi-Head Attention**

  Multi-head attention (Zeng et al., 2022) is a technique used in deep learning, notably in

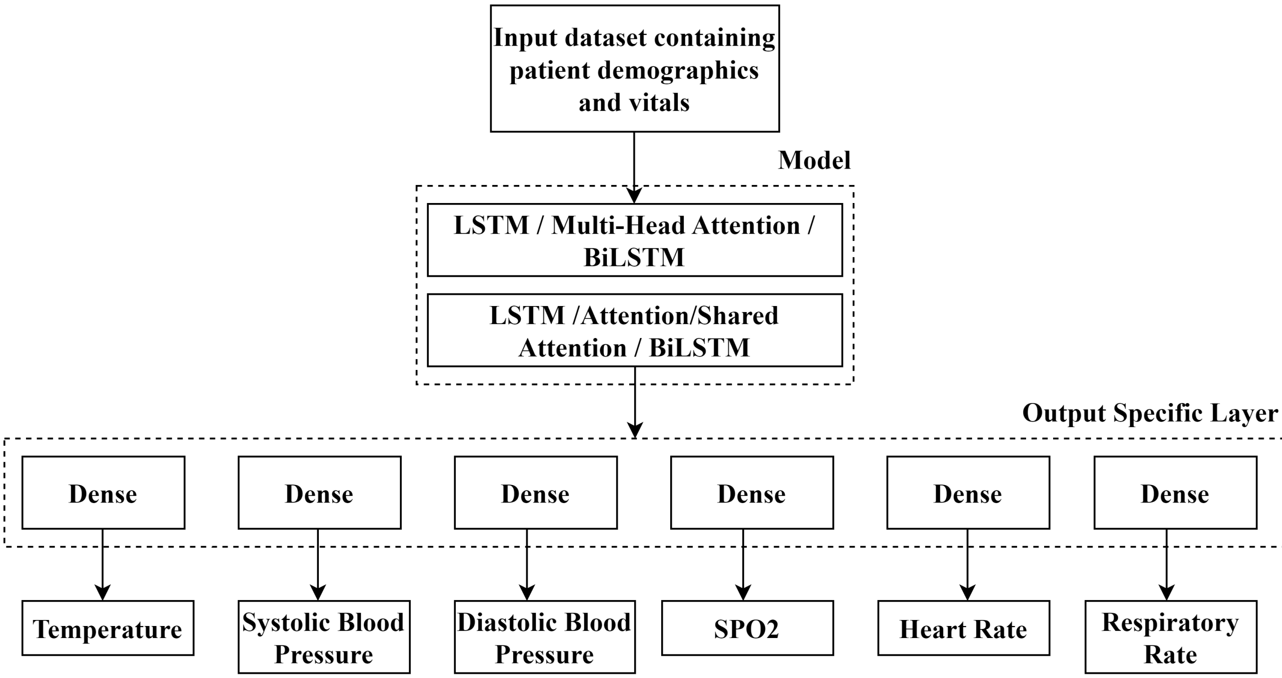

Figure 2. Model Architecture.

transformer models, to simultaneously capture various aspects of the input data. It improves the model's capacity to focus on multiple sections of the input sequence, enabling it to grasp intricate patterns and relationships.

Multi-head attention significantly boosts the capabilities of neural network models, particularly in tasks involving sequential data like natural language processing, machine translation, and sequence generation. It enables the model to comprehend diverse patterns and dependencies within the input sequence, resulting in enhanced performance across a wide array of tasks.

The process includes Splitting Heads of Key (K), Query (Q) and Value (V), then scaling Dot-Product Attention for each head and finally, Concatenation and Linear Transformation.

Mathematically, the multi-head attention operation is defined as follows in the equations (7) to (11):

$$split(Q) = [h_1, h_2, ..., h_{num\_h}] \tag{7}$$

$$split(K) = [h_1, h_2, ..., h_{num\_h}] \tag{8}$$

$$split(V) = [h_1, h_2, ..., h_{num\_h}] \tag{9}$$

$$head_i = softmax\left(\frac{Q \cdot K^T}{\sqrt{d_h}}\right) \cdot V \tag{10}$$

$$MultiHead(Q, K, V) =$$

$$\text{Concat}([\text{h}_1, h_2, ..., h_{num\_h}]) \cdot W^O (11)$$

where W^O is the output weight matrix with dimensions $d_{model} \times d_{model}$.

- **LSTM-Attention** LSTM-Attention (Zhang et al., 2019) is a fusion model that merges the capabilities of LSTM networks and attention mechanisms.
  LSTM: This component adeptly manages sequential data by grasping temporal dependencies, excelling at preserving information for extended periods, thus ideal for tasks dealing with time-series or sequential data.

  Attention Mechanism: By selectively concentrating on distinct segments of the input sequence during predictions, the attention mechanism empowers the model to attribute varying levels of significance to different segments, thereby augmenting its aptitude to apprehend pertinent details.

- **Bidirectional LSTM** Bidirectional LSTM or BiLSTM (Siami-Namini et al., 2019) enhances the LSTM framework by concurrently analyzing input sequences in both forward and backward directions. It comprises two LSTM layers: one handles the input sequence from start to finish (referred to as the forward LSTM), while the other processes it in reverse (known as the backward LSTM). The outcomes of both LSTM layers are fused or combined to generate the ultimate output sequence. Through this bidirectional processing, BiLSTM can grasp insights from both past and future contexts, enabling a more comprehensive modeling of temporal relationships in sequential data. BiLSTM proves particularly effective when accurate predictions or classifications rely on considering the context from both preceding and succeeding time steps.

4. Results and Analysis

The models were evaluated using Mean Squared Error (mse) in equation (12) and the results were noted for each of the models.

$$MSE = \sqrt{\frac{\sum(s-\hat{s})^2}{m}} \tag{12}$$

Where m is the total number of data points, s is the actual output value, $\hat{s}$ is the predicted output value.

The table 1 displays the mse values for each of the models and vitals across the test dataset, while Figures 3-6 illustrate the evolution of training and validation errors over increasing epochs. It is apparent from the graphs that the error values decrease with each epoch until reaching a stable state.

As observed from the table below, the individual mean squared errors over unseen test data for each model are 0.0022, 0.0024, 0.0026, and 0.0031 for LSTM-ATTN, LSTM, BiLSTM, and Multi-head Attention models respectively. LSTM Attention demonstrates the most promising results, followed by LSTM, BiLSTM, and finally Multi-head Attention. The individual mean squared error values for vital signs for LSTM-ATTN are 0.00024, 0.00054, 0.00058, 0.00072, 0.00013, and 0.00023 for Diastolic Blood Pressure, Heart Rate, Oxygen Saturation, Systolic Blood Pressure, Respiratory Rate, and Temperature respectively. Upon analyzing the individual mean squared errors of vital signs, it is evident that LSTM and LSTM-ATTN perform best for Diastolic Blood Pressure, while BiLSTM outperforms the other models for Heart Rate, and Oxygen Saturation, while LSTM-attention outperforms others for Systolic Blood Pressure, and Respiratory Rate, and LSTM for temperature. The graph of actual value VS predicted values for LSTM Attention model can also be visualized in Figures 7-12 wherein blue line depicts the actual value and the orange line depicts the predicted value.

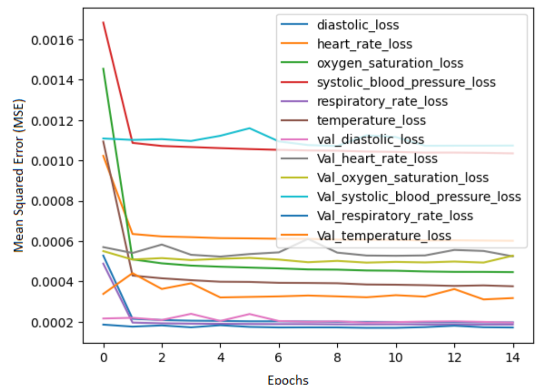

Figure 3. Variation of mean squared error with each epoch for each of the variables for LSTM model.

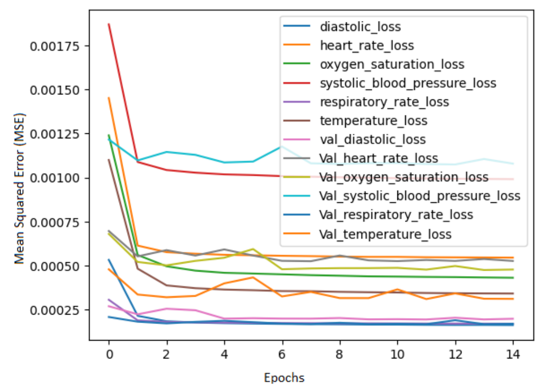

Figure 4. Variation of mean squared error with each epoch for each of the variables for LSTM attention model.

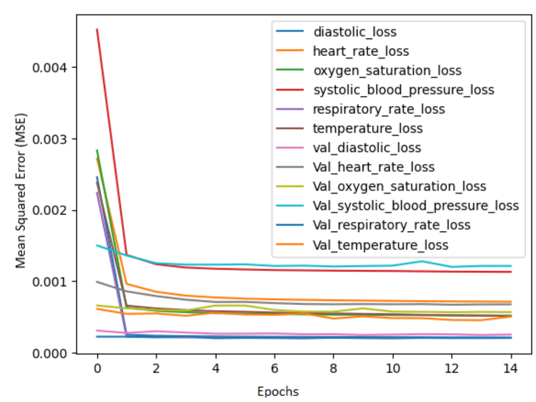

Figure 5. Variation of mean squared error with each epoch for each of the variables for Multihead attention model.

Table 1. Mean Squared Error Values for each of the predictions.
Wherein DBP = Diastolic Blood Pressure, HR = Heart Rate, SPO2 = Oxygen Saturation, SBP = Systolic Blood Pressure, RR = Respiratory Rate, Temp = Temperature

Model	Test MSE	DBP MSE	HR MSE	SpO2 MSE	SBP MSE	RR MSE	Temp MSE
LSTM Model	0.0024	0.00024	0.00054	0.00058	0.00072	**0.00013**	**0.00023**
BiLSTM Model	0.0026	**0.00016**	**0.00052**	**0.00042**	0.00099	0.00016	0.00032
Multi-head Attention	0.0031	0.00019	0.00082	0.00068	0.00072	0.00016	0.00050
LSTM -Attention	**0.0022**	**0.00016**	0.00060	0.00037	**0.00068**	**0.00013**	0.00026

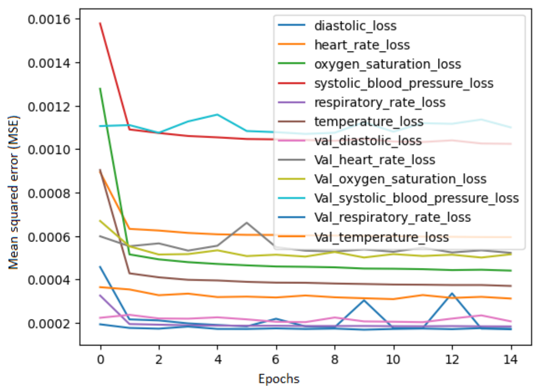

Figure 6. Variation of mean squared error with each epoch for each of the variables for BiLSTM model.

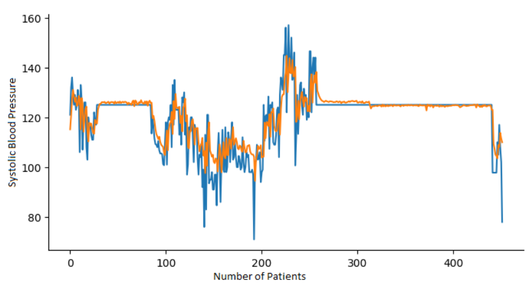

Figure 7. Actual value and predicted value of Systolic Blood Pressure.

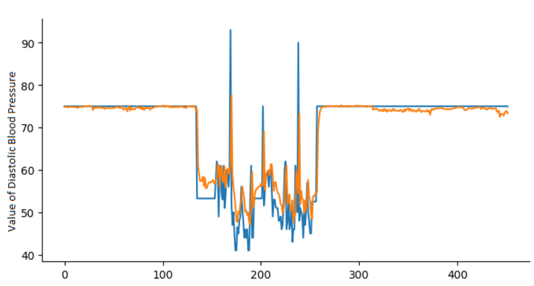

Figure 8. Actual value and predicted value of diastolic Blood Pressure.

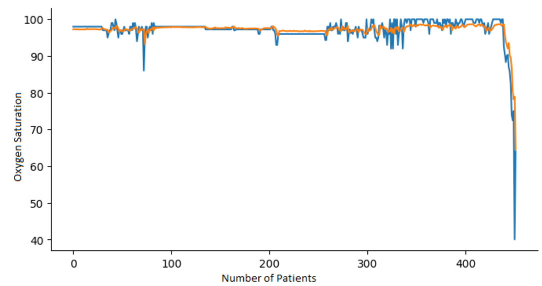

Figure 9. Actual value and predicted value of Oxygen Saturation.

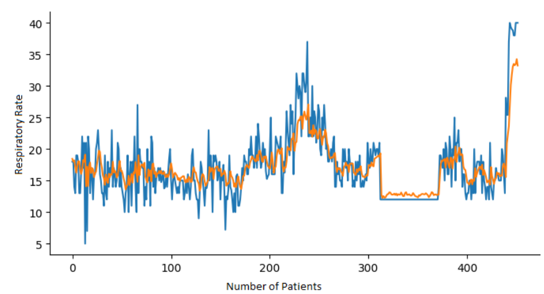

Figure 10. Actual value and predicted value of Respiratory Rate.

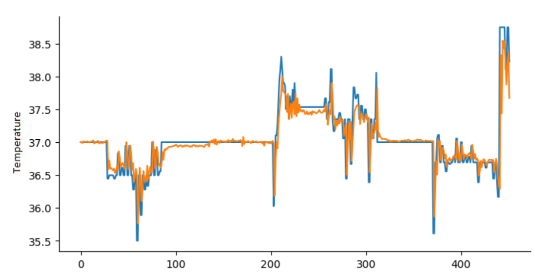

Figure 11. Actual value and predicted value of Temperature.

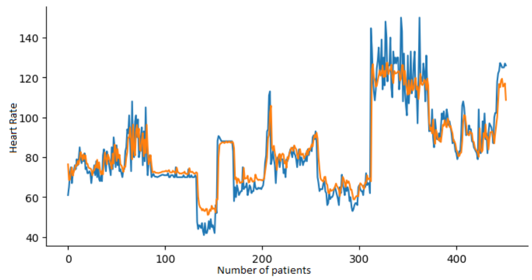

Figure 12. Actual value and predicted value of Heart Rate.

Below graphs (Figures 7-12) illustrate the comparison between predicted and actual values of vital signs. On the graph, the X-axis represents individual data points, while the Y-axis represents the corresponding vital sign values. In this work standardized units are used to measure each vital sign across all patients, ensuring consistency (e.g., temperature measured in Celsius).

Upon examination, a marginal variance between the actual and predicted values of each vital sign is observable. This discrepancy suggests that while the prediction model provides an overall estimation, individual patient conditions may influence the precise values recorded. Given that the test data encompasses multiple patients, it includes a spectrum of vital sign readings. Some values fall within typical or ”normal” ranges, reflecting stable patient conditions, while others may extend to extremes, indicating more acute or varied health statuses. This range underscores the diverse health scenarios captured within the dataset, offering valuable insights into the breadth of patient conditions and the challenges of accurate prediction across varied medical contexts.

5. Conclusion

In conclusion, we performed a forecasting of patient vital signs data over the MIMIC -III data using past vital signs, patient demographic data, and admission details. Multiple deep learning models including LSTM, LSTM -attention and Multi head Attention were trained to forecast the forthcoming vital signs using clinical data accessible prior to the admission diagnosis. The results exhibit robust predictive capabilities. Subsequent research endeavours could bolster the model's reliability by enhancing the interpretability of features extracted from clinical notes. Due to the lack of a few variables such as Air or Oxygen, SPO2 Scale 2, and consciousness we did not move further with the total NEWS score calculation. However, the performance highlights the models capability for implementation in hospital environments, where the ability to predict patients' next vital signs in real-time can be extremely valuable.

6. Future Work

Despite yielding favourable outcomes there is a lot that can be done in future to address certain limitations such as the current model being constrained to predicting only one hour ahead and attempts to extend this prediction window in an iterative manner yielded unsatisfactory results as the predictions suffered from error propagation wherein small error in the present position piled up and became larger in subsequent predictions. To address this limitation, implementing a transformer-based model (Phetrittikun et al., 2021) could allow us to perform multi horizon time series forecasting, enabling visualization of future trajectories.

To ensure trust in model predictions, interpretability is essential. Incorporating interpretability in the model can be achieved through methods such as the Temporal Fusion Transformer (Lim et al., 2021).

Given the inherent complexities within medical data, particularly in time series datasets characterized by irregular intervals, using a robust framework (Zhang et al., 2021) is imperative to address such challenges effectively. Additionally, it is crucial to confront issues related to missing data. In our study, we employed a straightforward imputation method to handle missing values (Zhang et al., 2023). However, exploring and integrating alternative frameworks into our methodology is essential to enhance the performance of the final model and yield improved results.

Considering the correlation between different vital signs (Forkan and Khalil, 2017) based on patient abnormality can also help us to impute missing values if other vital signs are present.

While the current study focuses solely on structured data, there is a need to broaden its applications to encompass data from diverse sources, including time varying International Classification of Diseases (ICD) codes (Fritz, 2000), clinical notes, radiology data (such as X-rays, MRIs, CT-scans, etc), and laboratory tests. Utilizing multiple diagnostic models for images, semi-structured data, and an embedding model for clinical or document-format data could facilitate this expansion.

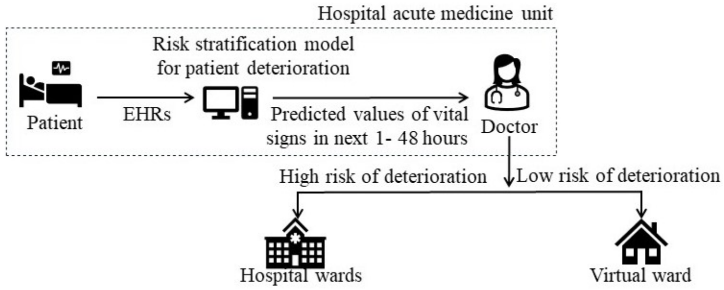

Figure 13. Application of our model in hospital wards for virtual ward selection

7. Application to Virtual Wards

Virtual wards also known as hospital at home (NHS, 2022) which are an innovative approach to health care delivery, have the ability to harness technology to extend the reach and effectiveness of medical care beyond traditional hospital settings. This healthcare service model provides hospital-level care to patients in their own homes rather than traditional hospital settings. Though not a new concept, utilizing virtual wards (Lewis et al., 2013) during the COVID-19 pandemic has proven to be a successful initiative for the National Health Service in the United Kingdom. Extending this practice to other medical conditions has also shown promise. The virtual ward selection process involves medical and nursing teams assessing patients' suitability, confirming self-sufficiency, and arranging a technical setup visit at the patient's residence. The assessment includes several baseline characteristics such as patient demographics, medical history, time varying vital signs, and referrals.

Our predictive model can also be used in the virtual ward context. As shown in Figure 13, predicting patients' vital signs ahead of time, healthcare providers can assess whether they require hospitalization or can be safely discharged to a virtual ward based on the prediction of the deterioration risk. The purpose of this model is to provide hospital staff with the necessary information to make decisions regarding patient management. This approach will not only optimizes resource allocation within healthcare facilities but also ensure timely and appropriate care delivery tailored to individual patient needs.

References

Ahmed, U., Lin, J. C.-W., & Srivastava, G. (2023). Multivariate time-series sensor vital sign forecasting of cardiovascular and chronic respiratory diseases. *Sustainable Computing: Informatics and Systems*, *38*, 100868.

Alghatani, K., Ammar, N., Rezgui, A., Shaban-Nejad, A., et al. (2021). Predicting intensive care unit length of stay and mortality using patient vital signs: Machine learning model development and validation. *JMIR medical informatics*, *9*(5), e21347.

Barfod, C., Lauritzen, M. M. P., Danker, J. K., Sölétormos, G., Forberg, J. L., Berlac, P. A., Lippert, F., Lundstrøm, L. H., Antonsen, K., & Lange, K. H. W. (2012). Abnormal vital signs are strong predictors for intensive care unit admission and in-hospital mortality in adults triaged in the emergency department-a prospective cohort study. *Scandinavian journal of trauma, resuscitation and emergency medicine*, *20*, 1–10.

Barton, C., Chettipally, U., Zhou, Y., Jiang, Z., Lynn-Palevsky, A., Le, S., Calvert, J., & Das, R. (2019). Evaluation of a machine learning algorithm for up to 48-hour advance prediction of sepsis using six vital signs. *Computers in biology and medicine*, *109*, 79–84.

Chiu, Y.-D., Villar, S., Brand, J., Patteril, M., Morrice, D., Clayton, J., & Mackay, J. (2020). Logistic early warning scores to predict death, cardiac arrest or unplanned intensive care unit re-admission after cardiac surgery. *Anaesthesia*, *75*(2), 162–170.

Elsworth, S., & Güttel, S. (2020). Time series forecasting using lstm networks: A symbolic approach. *arXiv preprint arXiv:2003.05672*.

Forkan, A. R. M., & Khalil, I. (2017). A clinical decision-making mechanism for context-aware and patient-specific remote monitoring systems using the correlations of multiple vital signs. *Computer methods and programs in biomedicine*, *139*, 1–16.

Fritz, A. G. (2000). *International classification of diseases for oncology: Icd-o*. World Health Organization.

Gardner-Thorpe, J., Love, N., Wrightson, J., Walsh, S., & Keeling, N. (2006). The value of modified early warning score (mews) in surgical in-patients: A prospective observational study. *The Annals of The Royal College of Surgeons of England*, *88*(6), 571–575.

Johnson, A. E., Pollard, T. J., Shen, L., Lehman, L.-w. H., Feng, M., Ghassemi, M., Moody, B., Szolovits, P., Celi, L. A., & Mark, R. G. (2016). Data descriptor: Mimic-iii, a freely accessible critical care database. *Scientific data*, *3*(160035), 1–9.

Kolic, I., Crane, S., McCartney, S., Perkins, Z., & Taylor, A. (2015). Factors affecting response to national early warning score (news). *Resuscitation*, *90*, 85–90.

Lewis, G., Vaithianathan, R., Wright, L., Brice, M. R., Lovell, P., Rankin, S., & Bardsley, M. (2013). Integrating care for high-risk patients in england using the virtual ward model: Lessons in the process of care integration from three case sites. *International journal of integrated care*, *13*.

Lim, B., Arık, S. Ö., Loeff, N., & Pfister, T. (2021). Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting*, *37*(4), 1748–1764.

NHS. (2022). Virtual ward including hospital at home [[Online; accessed 14-June-2024]]. https://www.england.nhs.uk/wp-content/uploads/2021/12/B1478-supporting-guidance-virtual-ward-including-hospital-at-home-march-2022-update.pdf/.

Petersen, J. A. (2018). Early warning score. *Dan Med J*, *65*(2).

Phetrittikun, R., Suvirat, K., Pattalung, T. N., Kongkamol, C., Ingviya, T., & Chaichulee, S. (2021). Temporal fusion transformer for forecasting vital sign trajectories in intensive care patients. *2021 13th Biomedical Engineering International Conference (BMEiCON)*, 1–5.

Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019). The performance of lstm and bilstm in forecasting time series. *2019 IEEE International conference on big data (Big Data)*, 3285–3292.

Smith, G., Redfern, O., Pimentel, M., Gerry, S., Collins, G., Malycha, J., Prytherch, D., Schmidt, P., & Watkinson, P. (2019). The national early warning score 2 (news2). *Clinical Medicine*, *19*(3).

Tam, B., Xu, M., Kwong, M., Wardell, C., Kwong, A., & Fox-Robichaud, A. (2017). The admission hamilton early warning score (hews) predicts the risk of critical event during hospitalization. *Canadian Journal of General Internal Medicine*, *11*(4), 24–27.

Tarassenko, L., Clifton, D. A., Pinsky, M. R., Hravnak, M. T., Woods, J. R., & Watkinson, P. J. (2011). Centile-based early warning scores derived from statistical distributions of vital signs. *Resuscitation*, *82*(8), 1013–1018.

Wang, S., McDermott, M. B., Chauhan, G., Ghassemi, M., Hughes, M. C., & Naumann, T. (2020). Mimic-extract: A data extraction, preprocessing, and representation pipeline for mimic-iii. *Proceedings of the ACM conference on health, inference, and learning*, 222–235.

Watkinson, P. J., Pimentel, M. A., Clifton, D. A., & Tarassenko, L. (2018). Manual centile-based early warning scores derived from statistical distributions of observational vital-sign data. *Resuscitation*, *129*, 55–60.

Youssef, A., Kouchaki, S., Shamout, F., Armstrong, J., El-Bouri, R., Taylor, T., Birrenkott, D., Vasey, B., Soltan, A., Zhu, T., et al. (2021). Development and validation of early warning score systems for covid-19 patients. *Healthcare Technology Letters*, *8*(5), 105–117.

Youssef Ali Amer, A., Wouters, F., Vranken, J., Dreesen, P., de Korte-de Boer, D., van Rosmalen, F., van Bussel, B. C., Smit-Fun, V., Duflot, P., Guiot, J., et al. (2021). Vital signs prediction for covid-19 patients in icu. *Sensors*, *21*(23), 8131.

Zeng, P., Hu, G., Zhou, X., Li, S., Liu, P., & Liu, S. (2022). Muformer: A long sequence time-series forecasting model based on modified multi-head attention. *Knowledge-Based Systems*, *254*, 109584.

Zhang, X., Zeman, M., Tsiligkaridis, T., & Zitnik, M. (2021). Graph-guided network for irregularly sampled multivariate time series. *arXiv preprint arXiv:2110.05357*.

Zhang, X., Li, S., Chen, Z., Yan, X., & Petzold, L. R. (2023). Improving medical predictions by irregular multimodal electronic health records modeling. *International Conference on Machine Learning*, 41300–41313.

Zhang, X., Liang, X., Zhiyuli, A., Zhang, S., Xu, R., & Wu, B. (2019). At-lstm: An attention-based lstm model for financial time series prediction. *IOP Conference Series: Materials Science and Engineering*, *569*(5), 052037.