Pathology Data Prioritisation: A Study Using Multi-Variate Time Series

Jing Qi[1], Girvan Burnside[2], and Frans Coenen[1]

[1] Department of Computer Science,
The University of Liverpool, Liverpool L69 3BX, UK
[2] Department of Health Data Science, Institute of Population Health,
The University of Liverpool, Liverpool L69 3BX, UK

Abstract. A mechanism to support the prioritisation of multi-variate pathology data, in the absence of a ground truth prioritisation, is presented. The motivation is the ever increasing quantity of pathology data that clinicians are expected to consider. The fundamental idea, given a previously unseen pathology result and the associated pathology history, is to use a deep learning model to predict future pathology results and then use the prediction to classify the new pathology result according to a pre-defined set of prioritisation levels. A further challenge is that patient pathology history, expressed as a multi-variate time series, tends to be irregularly time stamped and of variable length. The proposed approach used a Recurrent Neural Network to make predictions and a bounding box technique for the classification. The approach was evaluated using Urea and Electrolytes pathology data. The operation of the proposed approach was also compared with previously reported approaches, and was found to outperform these previous approaches.

Keywords: Data Ranking · Multivariate Time Series · Deep Learning · Pathology Data.

1 Introduction

Pathology results play an important role for decision making in any clinical environment. Clinicians use pathology results to diagnose patient conditions and decide on best next measures. Large amounts of pathology results are generated on a daily basis. For many conditions, such as kidney disease, pathology results are generated at regular intervals, sometimes over many years. Many pathology results comprise a set of values, not just one; in other words they are multivariate. The amount of pathology data that clinicians are expected to look at on a daily bases presents a significant information overload problem. A problem that is compounded by our ever increasing technical ability to collect pathology data; not helped by the recent COVID-19 pandemic which has put further strain on resources. In order to solve the problem, it is suggested that some form of automated pathology result prioritisation is required, and that this can best be achieved using the tools and techniques of machine learning whereby results can

be classified using a prioritisation scale of some kind. However, the challenge of the application of deep learning to pathology data is the absence of a "ground truth", a set of examples illustrating what a priority pathology result looks like, and what it does not look like. The reason for this is the resource required to generate such a ground truth.

There has been some previous work directed a pathology result prioritisation in the absence of a ground truth [7, 8]. In [8] it was assumed that high priority pathology results equated to anomalous priority results and hence an anomaly detection mechanism was adopted. However, given a large number of priority pathology results these would no longer be considered to be anomalous and therefore not be prioritised. In [7] a proxy ground truth was used based on the known outcomes of previous patients; whether they became emergency patients, in-patients, out-patents or remained with their General Practitioner (GP). The proxy ground truth was used to train a deep learner. Some improvement was reported over the work presented in [8]. However, the way the proxy ground truth was calculated meant that possible correlations between different pathology values were not considered.

An alternative pathology result prioritisation mechanism, to that given in [7] and [8], is presented in this paper directed at patients that have conditions where pathology results, each comprised of a set of values, are generated as an ongoing part of a care programme. In other words we have time series of previous pathology results. The fundamental idea is to predict whether the next pathology result in the sequence will be out of the anticipated normal range with respect to the pathology test under consideration. To be more specific, the use of an RNN-based pathology result forecasting model is advocated to predict follow-on results which can then be compared to the expected range. To distinguish prioritisation levels, a novel "Bounding Box" mechanism, which can distinguish between levels of prioritisation, is proposed. For the evaluation presented later in this paper three prioritisation levels were considered: high, medium and low. However, the bounding box technique will also work with any numbers of levels of two or more.

From the foregoing, the hypothesis that this paper seeks to establish is that there are patterns (trends) in a patients' historical pathology data which can be considered to be markers that are indicative of future pathology values. To act as a focus for the work the application domain of Urea and Electrolytes (U&E) pathology testing was considered. The proposed approach was evaluated using U&E data provided by Arrowe Park Hospital in Merseyside in the UK. This application domain was selected because it is the most commonly undertaken biochemistry test used to provide essential information on renal function.

The remainder of this paper is organised as follows. A review of relevant previous work is presented in Section 2. This is followed, in Section 3, by a review of the Urea and Electrolytes pathology application domain, used as a focus for the work. The proposed approach is considered in Section 4, and the evaluation of the proposed approach in Section 5. The paper is concluded in

Section 6 with a summary of the main findings and some suggested directions for future work.

2 Previous Work

The broad area of research into which the work presented in this paper falls, is that of big data prioritisation [12], where the aim is to determine which data items take priority over other data items. Data prioritisation can be applied in many areas [1, 6, 10], while in the medical area, the concept is more similar to patient triage [3] or prognosis [2], which supports decision-making through predicting the severity or risk of a given patient's condition. There has been some previous work directed at using machine learning for patient triage [4, 5, 14]. In [4] various forms of multinomial Logistic Regression (LR) were used: multinomial LR, eXtreme gradient boosting (XGBoost), random forests (RFs) and Gradient-Boosted Decision Trees (GBDTs) were explored to identify high-risk emergency department patients with suspected cardiovascular disease. In [5], natural language prediction methods were adopted to predict admission to a Neurosciences intensive Care Unit. In [14] a machine Learning based AutoScore–Derived triage tool was developed for predicting mortality risk after patients admitted to emergency departments. Most of the work aimed at predicting triage adopted supervised learning using a predefined training data labeled by domain experts, a resource intensive process which does not scale up to give general applicability. One of the challenges for the prioritisation of pathology data is the absence of training data.

Another challenge of time series pathology data is that it is usually irregular [15]; the spacing of observations is not constant. Most time series prediction methods, the technology adopted with respect to the approach presented in this paper, assume unit-spaced (regular) time series data [11]. Pathology time series data also tends to be multivariate in nature; the time series have more than one time-dependent variable. Each value depends not only on its past values but also the values for the associated variables. There are very few reported studies where irregular time series have been used directly. The majority of studies adopt some form of imputation so that spacing is of a unit length to ensure that time series are all of the same length. Or alternatively some form of padding and masking is used [13]. A number of alternatives are considered later in this paper.

3 U&E Testing Application Domain

The work presented in this paper is focused on Urea and Electrolytes pathology test data, U&E testing. U&E testing is usually performed to confirm normal kidney function or to exclude a serious imbalance of biochemical salts in the bloodstream. The U&E test data considered in this paper comprised five values per record: (i) Bicarbonate (bi), (ii) Creatinine (cr), (iii) Potassium (po), (iv) Sodium (so) and (v) Urea (ur). The measurement of each is referred to as a "task", thus we have five tasks per test. Abnormal levels in any of these tasks

may indicate that the kidneys are not working properly. However, a one time abnormal result does not necessarily need to be prioritised. A new task value that is out of range for a patient who has a previous recent history of out of range task values, but the latest result indicates a trend back into the normal range, may not be a priority result. Conversely, a new task value that is within the normal range for a patient who has a history of normal range task values, but the latest result indicates a trend heading out of the normal range, may need to be prioritised. Given a new set of pathology values for a U&E test we wish to determine the priority to be associated with this set of values.

The U&E data used for evaluation purposes with respect to the work presented in this paper comprised a set of clinical patient records, $\mathbf{D} = \{P_1, P_2, \dots\}$, where each record $P_i \in \mathbf{D}$ was of the form:

$$P_i = \langle PatientID, History, TestResult, ReferencedRange \rangle \tag{1}$$

Where: (i) $PatientID$ is the ID for the patient in question; (ii) $History$ is a set previously obtained pathology results expressed as a set a multivariatre time series $T = [t_1, t_2, \dots]$, where each $t_i \in T$ comprised a 5-tuple of the form $\langle v_{bi}, v_{cr}, v_{po}, v_{so}, v_{ur} \rangle$, (iii) $TestResult$ is a current previously unseen pathology result R also comprised of an n tuple of the form $\langle v_{bi}, v_{cr}, v_{po}, v_{so}, v_{ur} \rangle$, and (iv) $ReferencedRange$ is a set of bounds defining the normal range for the patient in questionn for the values associated with each task represented as two sets, $L = [l_1, l_2, \dots]$ and $U\,[u_1, u_2, \dots]$, where L holds the minimum (normal low) values and U holds the maximum (normal upper) values. There is a one-to-one correspondence with T. The normal low and high dimensions indicate a "band" in which pathology results are expected to fall given a healthy patient. These bands vary from task to task, will not be the same for each patient and may change for a given patient over the course of time. A training record $P_i \in \mathbf{D}$ will also include a task label c taken from a set of classes C.

4 Prioritisation for U&E Pathology Patients Results

The fundamental idea, for prioritising pathology results, presented in this paper is that prioritisation can be achieved by predicting whether future patient pathology results will be out of the normal range. An overview of the proposed process is given in Figure 1. The process commences with a new multi-variate pathology result $R = \langle v_1, v_2, \dots, v_n \rangle$. This is combined with the pathology history of the patient in question to give a time series $T = [t_1, t_2, \dots, R]$. This is then passed to a prediction model where the next set of results $P = \langle v_1, v_2, \dots, v_n \rangle$ is predicted. A bounding-box technique is then used to classify R according to P. Thus, Figure 1, there are two main stages within the overall process:

1. **Future Result Prediction**: The process of predicting future pathology results given a new, previously unseen, multi-variate pathology result and the pathology result history for a given patient.

2. **Bounding Box Classification**: The process of assigning a priority level to a previously unseen pathology result using the predicted future pathology results.

Each of the above is discussed in further detail in the following two sub-sections.

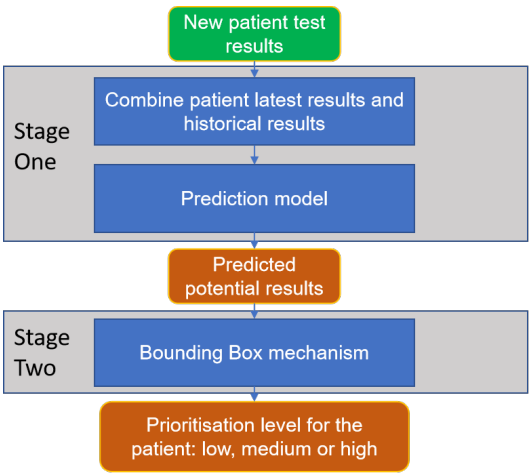


Fig. 1. Schematic outlining the proposed pathology data prioritisation process

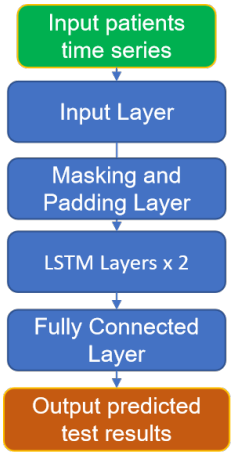


Fig. 2. LSTM architecture for future pathology results prediction

4.1 Future Results Prediction

There are a range of prediction mechanism that could have been adopted with respect to Stage 1 of the proposed process, a Long Short Term Memory (LSTM) mechanism was adopted. The architecture of the proposed LSTM is showed in Figure 2. From the figure it can be seen that the proposed LSTM comprised four layers: (i) the input layer which receives a time series T which includes the new pathology result R of interest, (ii) the padding layer where the time series were transformed from irregular time series to regular time series, (iii) the hidden layer comprised of LSTM cells arranged in two sub-layers, and (iv) the fully connected layer where the predicted future test results are generated. The hidden layer is where the training takes place. The training set is divided into a set of overlapping input/output *samples*. During training, each sample forms a *prediction step* in the overall LSTM model generation process. The hidden layers compute the intermediate results and pass them on to the next iteration, which makes it possible for the network to maintain memory of the state of earlier historical records, so that the effect of the early results can be considered for the prediction.

The adopted LSTM architecture comprised 16 cells arranged in two sub-layers in the hidden layer, and n neurons in the output layer for predicting a set of n pathology task values ($n = 5$ with respect to the U&E data set used

to evaluate the proposed approach). The "input shape" of the data is a time series $T = [t_1, t_2, \dots]$, where each point t_1 comprises a n-tuple. An important aspect of training a LSTM network is the iterative updating of weights using the training data. To achieve this, the Adam stochastic gradient descent was used. The adopted loss function was the Mean Squared Error (MSE) loss function. The structure and parameters used were selected a as consequence of a number of preliminary experiments (not reported here), and because these had been adopted in related work [9].

4.2 Bounding Box Classification

Once a set $P = \langle v_1, v_2, v_3, ..., v_n \rangle$ of predicted test results have been obtained, derived from a new pathology result to be prioritised, the next step is to assign a priority class to the new pathology result. As noted earlier, the prioritisation idea considered in this paper is to use the normal range associated with a particular task and patient. The normal range will vary from task to task and from patient to patient, but in each case will be defined by a minimum and maximum value. Thus, the normal range "zone" is defined by a n dimensional bounding box, where n is the number of tasks and each side will equate to the normal range associated with a task defined by the minimum and maximum value for the task in question. If a predicted pathology result falls entirely within this bounding box the pathology result will be deemed to have a "low" priority. Anything outside can then be labelled as "high" priority. However, a binary classification (high-low) is considered too coarse a classification; we require more than one class label for results that fall outside of the "low priority bounding box". In this paper we will consider a three class prioritisation, $C = \{\text{high}, \text{medium}, \text{low}\}$. Thus if a pathology result falls outside of the "low priority bounding box" it will be either medium or high priority. The question is how this can best be calculated. One idea is to simply calculate the Euclidean distance from the geometric centre of the low priority bounding box and use a threshold of some kind to distinguish high priority pathology results from low priority results. However, this will mean that the distance from a pathology result close to a corner of the bounding box to the geometric center, will be treated the same as a result some way away from a side of the bounding box. Thus to distinguish between medium priority and high priority pathology result a second bounding box, the "medium priority bounding box" was defined by expanding the low priority box by a factor χ.

The pseudo code for the Bounding Box Comparison approach is given in Algorithm 1. Note that the algorithm assumes $n = 5$. The inputs to the algorithm are: (i) a predicted pathology result $P = \{v_1, v_2, \dots v_n\}$, (ii) the set $L = \{l_i, l_2, \dots l_n\}$ of low normal range values, (iii) the set $U = \{u_i, u_2, \dots u_n\}$ of upper normal range values and (iv) the expansion factor χ to be applied. The algorithm commences, line 2, by defining a default class of "high". Then, line 3, the algorithm determines whether P falls inside the low priority bounding box using a function *inNormalZone*. This returns a result tuple comprised of a set of binary values, 0 = inside and 1 = outside. In the U&E case the tuple is of size $n = 5$, but it can be any other values according to the type of pathology

Algorithm 1 Bounding Box Comparison

```
1: input P, L, U χ
2: class = high                                          ▷ Default class
3: result = inNormalZone(P, U, L)                        ▷ Determine if P in normal range
4: if result == ⟨0, 0, 0, 0, 0⟩ then
5:     class = low                                       ▷ Predicted point entirely within normal zone
6: else
7:     result = ⟨1, 1, 1, 1, 1⟩
8:     for ∀v_i ∈ P do
9:         offset = (u_i − l_i) × χ
10:        if (u_i + offset) ≥ t_i ≥ (l_i − offset) then
11:            result_i = 0
12:        end if
13:    end for
14:    if result == ⟨0, 0, 0, 0, 0⟩ then
15:        class = medium                                ▷ Predicted point entirely within medium zone
16:    end if
17: end if
18: return class
```

under consideration. If the resulting tuple equates to $\langle 0, 0, 0, 0, 0 \rangle$ then P is entirely within the low priority bounding box and allocated the class "low" (line 5). Otherwise the result tuple is set to $\langle 1, 1, 1, 1, 1 \rangle$ (line 7) and we proceed to expand the low priority bounding box in each direction in an iterative manner (lines 8 to 13). On each iteration the expansion is conducted using an offset applied to each normal low and upper value. The offset, with respect to each task is calculated as shown in Equation 2, where $l_i \in L$, $u_i \in U$, and χ is a predefined multiplier (factor). On each iteration (line 10), $t_i \in R$ is compared with the expanded range and the outcome added to the result tuple. On completion, if the tuple equates to $\langle 0, 0, 0, 0, 0 \rangle$ the new pathology record P is allocated the class "medium" (line 15). Otherwise the default class, "high", is used. The class is then returned (line 18). It is easy to see how the process can be repeated and further classes added if desired.

$$offset = (u_i - l - i) \times \chi \tag{2}$$

A value for *chi* can be established empirically. However, for the evaluation presented here a proxy ground truth was used (more on this in Sub-section 5.1). The proxy training data set was of the form presented in Sub-section 3. A value for χ was then "learnt", for each class, by clustering all potential χ values and then determining the mid point between the two cluster centroids. For the evaluation presented in the following section a binary classification scenario was considered $C = \{c_1, c_2\}$. The adopted process is illustrated in Algorithm 2. The input to the algorithm is the training set $\mathbf{D}$ and the set of classes C (see Section 3). A value for χ for each value v_k in each time series T_j in each set of time series T for each patient record D_i in $\mathbf{D}$ was calculated using the Equation 3 where

u_k and l_k are the upper and lower range limits associated with v_k, and $dist$ is the distance of v_k from the mid-point between u_k and l_k calculated as shown in Equation 4. The average value χ for each class is calculated, lines 12 and 13 in Algorithm 2. The mean of the two averages is then the final value for χ to be used. For future work the intention is to investigate the potential of using different values for chi for different tasks. In the present study the same value of χ was used throughout.

Algorithm 2 Factor χ Generation

1: **Input D**, C
2: **Chi** $= \{Chi_1, Chi_2\}$
3: **for** $D_i \in$ **D do**
4: **for** $T_j \in$ **T**, **T** $\in D_i$ **do**
5: **for** $v_k \in T_j$ **do**
6: $dist = abs\left(v_k - \frac{u_k - l_K}{2}\right)$
7: $\chi = \frac{2 \times dist}{u_k - l_k}$
8: $Chi_i = Chi_i \cup \chi$, i = class ID for class $c \in D_i$
9: **end for**
10: **end for**
11: **end for**
12: $ave_{c_1} = average(Chi_1)$ ▷ cluster centre for $c_1 \in C$
13: $ave_{c_2} = average(Chi_2)$ ▷ cluster centre for $c_2 \in C$
14: $\chi = \frac{ave_{c_1} + ave_{c_2}}{2}$
15: **return** χ

$$\chi = \frac{2 \times dist}{u_k - l_k} \qquad (3) \qquad\qquad dist_i = abs\left(v_k - \frac{u_k - l_K}{2}\right) \qquad (4)$$

5 Evaluation

The evaluation of the proposed approach is reported on in this section. For the evaluation, as noted earlier, a U&E data set provided by Arrowe Park Hospital in Merseyside in the UK was used. The data set was entirely anonymised and ethical approval for its usage in anonymised form obtained by Arrowe Park Hospital. Details concerning this data set are given in Sub-section 5.1 below. The Objectives of the evaluation were

1. To investigate the most appropriate imputation strategy for addressing the unequal length of the pathology time series to be considered.
2. To determine the overall performance of the proposed approach using a proxy ground truth, as proposed in [7], and comparing with previously proposed approaches.

The first is considered in Sub-section 5.2 and the second in Sub-section 5.3. The value for χ was determined using the classifications in the training data. Five-cross validation was used through out. The average value for χ was found to be

0.57. All the experiments were run using a windows 10 desktop machine with a 3.2 GHz Quad-Core IntelCore i5 processor and 24 GB of RAM. For the LSTM, a GPU was used fitted with a NVIDA GeForceRTX 2060 unit.

5.1 Evaluation Data Set

The evaluation data set used was provided by Arrowe Park Hospital in Merseyside in the UK. A general format of the data was presented in Section 3. The data set comprised 3,734 patient records with five U&E task results (time series) per patient. The operation of prediction models is typically conducted using a test set that features known values for the variable to be predicted which can be compared with the predicted values produced by the model. However, as noted in the introduction to this paper such test data is typically not available because of the resource required. Indeed this was the motivation for the work presented in this paper. As also noted earlier, in [7] a proxy ground truth was used. The same approach was therefore adopted with respect to the evaluation of the proposed approach. The final destinations of the patients within the U&E data set were used to create a proxy ground truth; whether they ended-up as emergency, in or out patients; equating to high, medium and low priority respectively ($C = \{\text{high}, \text{medium}, \text{low}\}$). The proxy ground truth data set comprised 255 patients with high priority, 123 with medium priority and 3,356 with low priority, covering all five tasks.

5.2 Data Imputation

As noted earlier, the interval between pathology results (points in the multivariate time series), and the overall length of the pathology multi-variate time series, was variable. This is illustrated, using the U&E data set, in Figures 3 and 4. Figure 3 shows the number of days between the individual patients considered. For the figure the patients were ordered according to the maximal interval in their pathology history and each given a sequential ID number. In Figure 3, sequential ID numbers are listed on the x-axis, and maximal intervals on the y-axis. From the figure it can be seen that the majority of patients have a maximum pathology interval of less than 100 days. Figure 4 shows the range of time series lengths. The figure was generated by grouping the time series lengths, in the U&E evaluation data, into 10 bins of 6 starting with length 1, this covering time series from 1 to 60. From, the figure it can be seen that the majority of time series fell into the first bin, lengths 3 to 6.

LSTM model generation requires all time series to be of the same length. We would also like our time series to reflect the correct spacing between pathology results (we could have simply assumed a unit spacing). To engineer this we experimented with five alternative strategies:

1. **Mean imputation**: Using the mean value in the time series to impute additional values.

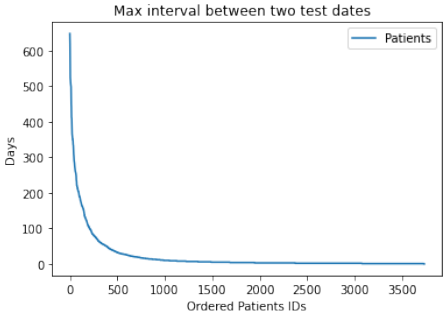


Fig. 3. Maximal interval in days per patient

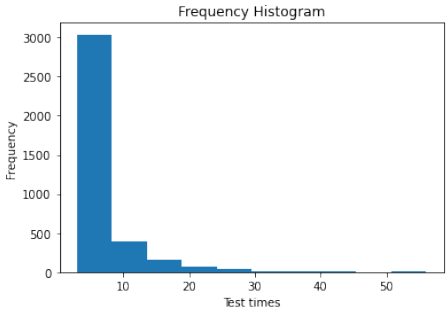


Fig. 4. Frequency of occurrence of different time series lengths

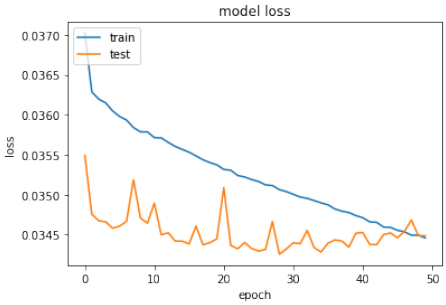
2. **Median imputation**: Using the medium value in the time series to impute additional values.
3. **Mode imputation**: Using the mode value in the time series to impute additional values.
4. **Zero imputation**: Using the value 0 to impute additional values.
5. **Padding and masking**: Skip missing time series values.

To evaluate these different strategies 'loss plots" were generated for each strategy as shown in Figure 5. The number of epochs is given on the x-axis, and the loss in terms of Mean Squared Error (MSE) on the y-axis (note the y-axis scale is different for each graph). The same number of epochs was used in each case. The plots show the "training history" of the LSTMs. Each plot shows the loss between the training and test (validation) data as the model generation progressed. We want the loss to be minimal once model generation is complete. From the figure it can be seen that in all cases the training and test curves converge. Closer inspection indicates that the "padding and masking" method achieved the minimum loss.

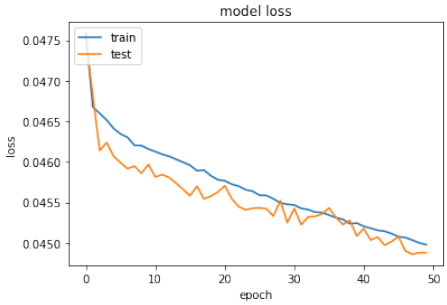
Table 1 gives the classification accuracy, precision and recall values obtained, using the five imputation strategies when the proposed bounding box classification was applied (best results in bold font). The reported results are averages obtained using five-cross validation. From the table it can be seen that best performance was obtained using padding and masking (confirming the results from Figure 5). As noted in Section 4, Padding and Masking is considered to offer the advantage that it preserves the original length and irregular spacing of the time series; this seems to be the most appropriate method for dealing with the differing lengths of time series as it does not change the original information provided by the data itself. Zero imputation produced the worst performance.

5.3 Overall Performance

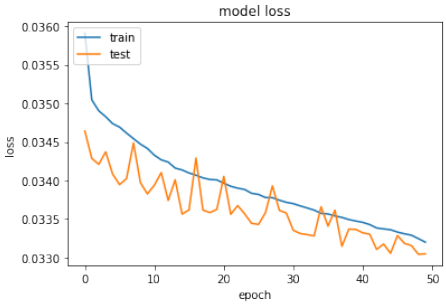
Table 2 gives the best results from Table 1 compared to the results reported in [7] and [8] where a very similar proxy ground truth was used; best results are again highlighted in bold font. In [7] two classification models were considered a



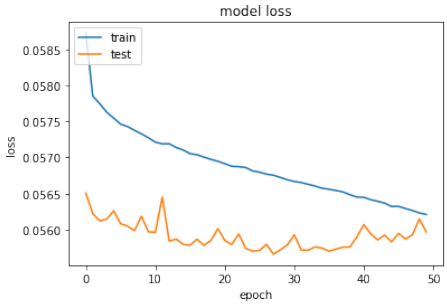
(a) Loss of using mean imputer



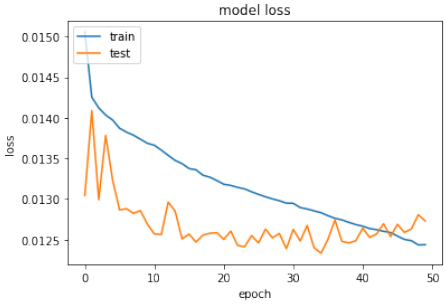
(b) Loss of using median imputer



(c) Loss of using mode imputer



(d) Loss of using zero imputer



(c) Loss of using padding and masking

Fig. 5. Performance comparison of the five alternative imputation strategies in terms of LSTM model generation

Imputation Strategy	Acc.	Precision			Recall		
		High	Med.	Low	High	Med.	Low
Mean	0.56	0.60	0.51	0.53	**0.59**	0.59	**0.75**
Median	0.33	0.57	0.29	0.16	0.31	0.38	0.29
Mode	0.60	0.60	0.51	0.41	0.55	0.57	0.50
Zero	0.21	0.28	0.21	0.17	0.27	0.24	0.31
Pad. & Mask.	**0.73**	**0.69**	**0.62**	**0.61**	0.58	**0.69**	0.47
Ave	0.48	0.55	0.43	0.38	0.43	0.49	0.46

Table 1. Performance comparison of the five alternative imputation strategies in terms of accuracy, precision and recall

k Nearest Neighbour model and an LSTM model, the best results for both are included in Table 2. In a [8] an Anomaly Detection (AD) approach was proposed. Two versions were considered, a point-based AD approach and a time series AD approach. The best reported results for both are also included in Table 2. Note that in [8] only average precision and recall were reported.

Method	Acc.	Precision			Recall		
		High	Med.	Low	High	Med.	Low
LSTM and Bounding Box	**0.73**	**0.69**	**0.62**	0.61	0.58	**0.69**	0.47
LSTM [7]	0.61	0.58	0.55	**0.69**	**0.79**	0.59	0.63
kNN [7]	0.60	0.42	0.51	0.85	0.70	0.55	**0.75**
Point-based AD [8]	0.34	0.35			0.43		
Time Series AD [8]	0.45	0.45			0.43		

Table 2. Average Accuracy, Precision and Recall results compared with the results reported in [7] and [8]

From Table 2 it can be seen that the proposed LSTM prediction coupled with bounding box classification produced the best overall accuracy. Closer inspection of the table indicates that good precision was obtained, using the proposed method, with respect to high and medium priority classes. The poor performance of the anomaly detection approaches (Point-based AD and Time Series AD) is probably because anomalous pathology results do not necessarily equate to priority pathology results.

6 Conclusions

The work presented in this paper was directed at multi-variate pathology data prioritisation in the absence of a ground truth. This is typically the case because of the resource required. A proposed approach has been presented that used LSTM prediction couple with a novel bounding box classification mechanism. The main contributions of this paper are: (i) the idea of using LSTM predicted test results as a marker for prioritisation and (ii) the derived "bounding box" technique for prioritisation classification. A further challenge was the irregular sampling interval of pathology data, and the variable length of the pathology history associated with a given pathology result. A number of alternative imputation techniques were therefore considered. The operation of the proposed approach was compared with alternative techniques from the literature using a proxy ground truth. The results indicated a better overall performance than that achieved by earlier work. A best accuracy of 73% was obtained. Padding and masking was found to be the most appropriate method for ensuring all time series were of the same size. For future work the authors intend to investigate: (i) the generation of artificial evaluation data sets to provide for a more comprehensive evaluation, and (ii) a comprehensive collaborate with clinicians to obtain feed back regarding the prioritisation produced and to test the utility of the best performing mechanism in a real setting. The authors are currently liaising with domain experts on the practical impact of the proposed pathology data prioritisation mechanism presented this paper.

References

1. Reem Ahmed, Fuzhan Nasiri, and Tarek Zayed. A novel neutrosophic-based machine learning approach for maintenance prioritization in healthcare facilities. *Journal of Building Engineering*, 42:102480, 2021.
2. Mrs. Chandralekha and N. Shenbagavadivu. Data analytics for risk of hospitalization of cardiac patients. *Iete Journal of Research*, pages 1–10, 2021.
3. Riccardo Doyle et al. Machine learning–based prediction of covid-19 mortality with limited attributes to expedite patient prognosis and triage: retrospective observational study. *JMIRx med*, 2(4):e29392, 2021.
4. Huilin Jiang, Haifeng Mao, Huimin Lu, Peiyi Lin, Wei Garry, Huijing Lu, Guangqian Yang, Timothy H Rainer, and Xiaohui Chen. Machine learning-based models to support decision-making in emergency department triage for patients with suspected cardiovascular disease. *International Journal of Medical Informatics*, 145:104326, 2021.
5. Eyal Klang, Benjamin R Kummer, Neha S Dangayach, Amy Zhong, M Arash Kia, Prem Timsina, Ian Cossentino, Anthony B Costa, Matthew A Levin, and Eric K Oermann. Predicting adult neuroscience intensive care unit admission from emergency department triage using a retrospective, tabular-free text machine learning approach. *Scientific reports*, 11(1):1–9, 2021.
6. Sunday Ochella, Mahmood Shafiee, and Chris Sansom. Adopting machine learning and condition monitoring pf curves in determining and prioritizing high-value assets for life extension. *Expert Systems with Applications*, 176:114897, 2021.
7. Jing Qi., Girvan Burnside., Paul Charnley., and Frans Coenen. Event-based pathology data prioritisation: A study using multi-variate time series classification. In *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KDIR,*, pages 121–128. INSTICC, SciTePress, 2021.
8. Jing Qi, Girvan Burnside, and Frans Coenen. Ranking pathology data in the absence of a ground truth. In *International Conference on Innovative Techniques and Applications of Artificial Intelligence*, pages 209–223. Springer, 2021.
9. Haşim Sak, Andrew Senior, and Françoise Beaufays. Long short-term memory based recurrent neural network architectures for large vocabulary speech recognition. *arXiv preprint arXiv:1402.1128*, 2014.
10. Á Sobrinho, L Dias da Silva, E Candeia, A Perkusich, et al. Machine learning classification models for covid-19 test prioritization in brazil. *Journal of Medical Internet Research*, 2021.
11. Philip B Weerakody, Kok Wai Wong, Guanjin Wang, and Wendell Ela. A review of irregular time series data handling with gated recurrent neural networks. *Neurocomputing*, 441:161–178, 2021.
12. Carla Wilkin, Aldónio Ferreira, Kristian Rotaru, and Luigi Red Gaerlan. Big data prioritization in scm decision-making: Its role and performance implications. *International Journal of Accounting Information Systems*, 38:100470, 2020.
13. Jie Wu, Na Li, and Yan Zhao. Missing data filling based on the spectral analysis and the long short-term memory network. In *2021 International Symposium on Computer Technology and Information Science (ISCTIS)*, pages 198–202. IEEE, 2021.
14. Feng Xie, Marcus Eng Hock Ong, Johannes Nathaniel Min Hui Liew, Kenneth Boon Kiat Tan, Andrew Fu Wah Ho, Gayathri Devi Nadarajan, Lian Leng Low, Yu Heng Kwan, Benjamin Alan Goldstein, David Bruce Matchar, et al. Score

for emergency risk prediction (serp): An interpretable machine learning autoscore-derived triage tool for predicting mortality after emergency admissions. *medRxiv*, 2021.
15. Muhammad Adib Uz Zaman and Dongping Du. A stochastic multivariate irregularly sampled time series imputation method for electronic health records. *BioMedInformatics*, 1(3):166–181, 2021.