Pathology Data Prioritisation: A Study of Using Multi-Variate Time Series without a Ground Truth

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 Biostatistics, Institute of Translational Medicine,
The University of Liverpool, Liverpool L69 3BX, UK

Abstract. In any hospital, pathology results play an important role for decision making. However, it is not unusual for clinicians to have hundreds of pathology results to review on a single shift; this "information overload" presents a particular challenge. Some form of pathology result prioritisation is therefore a necessity. One idea to deal with this problem is to adopt the tools and techniques of machine learning to identify prioritisation patterns within pathology results and use these patterns to label new pathology data according to a prioritisation classification protocol. However, in most clinical situations there is an absence of any pathology prioritisation ground truth. The usage of supervised learning therefore becomes a challenge. Unsupervised learning methods are available, but are not considered to be as effective as supervised learning methods. This paper considers two mechanisms for pathology data prioritisation in the absence of a ground truth: (i) Proxy Ground Truth Pathology Data Prioritisation (PGR-PDP), and (ii) Future Result Forecast Pathology Data Prioritisation (FRF-PDP). The first uses the outcome event, what happened to a patient, as a proxy for a ground truth, and the second forecasted future pathology results compared with the known normal clinical reference range. Two variation of each are considered: kNN-based and LSTM-based PGR-PDP, and LSTM-based and Facebook Profit-based FRF-PDP. The reported evaluation indicated that the PGR-PDP mechanism produced the best results with little distinction between the two variations.

Keywords: kNN, LSTM-RNN, Facebook Prophet, Time Series, Pathology Data.

1 Introduction

The problem of information overload has become a global phenomena, fueled by the large quantities of data that are produced currently, on a continuous basis. One specific example is the increasing number of pathology results generated in hospitals. Clinicians usually use the results to provide support for decision

making when treating patients, such as whether some urgent intervention is required, or some particular medication needs to be prescribed or further tests need to be carried out. However, even for experienced doctors, this procedure is complex and time consuming. One solution is to prioritise pathology results using the tools and techniques of machine learning so as to provide a "look at these first" style of help to accelerate the process of analysing pathology results. Highlighting significant results to be considered first will prevent the condition of patients from worsening due to the delays in treatment.

However, the biggest challenge for the prioritisation of pathology data using machine learning is, in many cases, the absence of ground truth data. In practice, experienced clinicians can recognise a priority result when they see one, but do not have the resource to use their experience to generate bespoke training data sets. Thus traditional, well established, supervised learning techniques can not be used directly. One proposed solution, reported in [10], is to use an unsupervised anomaly detection approach whereby anomalous records are considered to be priority records. In [10] a set of clusters was generated using historical data and any new pathology result which could not be readily fitted into a cluster labeled as an outlier and therefore assumed to be a priority record. The flaw in this approach was that, just because a record was unusual, it did not necessarily mean it was a priority record (and vice-versa). In [9] the idea of using a proxy ground truth was proposed; an idea that addresses the criticism directed at the anomaly detection approach described in [10]. This paper builds on the ideas presented in [9] by reconsidering the proxy ground truth concept. This paper also presents an alternative to this previous work founded on the idea of forecasting future pathology results and comparing these with the expected normal clinical range.

In more detail, two mechanisms are presented in this paper: (i) Proxy Ground Truth Pathology Data Prioritisation (PGT-PDP) and (ii) Future Result Forecast Pathology Data Prioritisation (FRF-PDP). The PGT-PDP mechanism, founded on ideas first proposed in [9], uses a proxy training data set to build a pathology data classification model. The proxy ground truth in this case was obtained from meta-knowledge about patients concerning the "final destination" of patients, what is known as the *outcome event* for each patient. Three outcome events were considered: Emergency Patient (EP), an In-Patient (IP) or an Out Patient (OP). Then, given a new pathology result and the patient's pathology history, it would be possible to predict the outcome event and then use this to prioritise the new pathology result. For example if we predict the outcome event for a patient to be EP, then the new pathology result would be assigned a *high* priority; however, if we predict that the outcome event will be IP the new pathology result would be assigned *medium* priority, and otherwise *low* priority. The hypothesis that we seek to establish here is that there are patterns in patients' historical lab test results which are markers as to where the patient "ended up", and which can hence be used for prioritisation. Two variations of the PGT-PDP mechanism are considered for classification model generation, Long Short-Term Memory (LSTM) and k Nearest Neighbour (kNN). The first is a well established deep

learning approach, and kNN is a classic way of classifying time series data that has previously been used in the medical domain [19]. A value of $k = 1$ was adopted for the kNN, as suggested in [1], and Dynamic Time Warping (DTW) was used as the similarity measure.

The FRF-PDP mechanism uses the normal range for a test result. This will vary from test to test, and patient to patient, and may even change with time. A "quick and easy" solution would be to prioritise a new pathology results if it fell outside of the normal range. However, waiting till this happens may be too late. Instead it is argued here that a better approach would be to use the patient's history and, given a new pathology result, predict what the next pathology result value in the sequence will be. If this next value is out of range then we have a priority pathology result. In other words, the idea promoted by this mechanism is to use time series forecasting as a tool for prioritising pathology data. More specifically, the mechanism utilises the idea of time series regression. Two regression approaches are considered: (i) LSTM regression and (ii) Facebook Prophet (FP). LSTM regression is a well established regression approach. FP forecasting is founded on a novel Bayesian forecasting model whereby the influencing factors that affect the trend associated with a set of time series are decomposed so as to identify patterns. Unlike LSTM regression, FP can operate using irregular time intervals (pathology data is collected in an irregular manner). In the case of the LSTM regression a unit time interval needed to be assumed.

The main contribution of this paper are thus: (i) a more detailed consideration of the event-based pathology prioritisation than that presented in [9], (ii) a new mechanism to prioritise pathology data using future result forecasting, and (iii) an in depth comparison of the operation of the two mechanisms. To act as a focus for the comparison the domain of Urea and Electrolytes (U&E) pathology testing was considered; a domain that features five types of pathology result. More specifically U&E data provided by Arrowe Park hospital in Merseyside in the UK was used.

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 presented in this paper. A formalism for the pathology prioritisation problem is also presented.The proposed mechanisms, PGT-PDP and FRF-PDP, are then presented in Sections 4 and 5 respectively. The comparative evaluation of the mechanisms is presented and discussed in Section 6. The paper is concluded, in Section 7, with a summary of the main findings and some suggested avenues for future work.

2   Previous Work

This previous work section commences by considering the general concept of data prioritisation, followed by some discussion of existing work directed at pathology data prioritisation. The reminder of this previous work section considers the

classification model generation and forecasting approaches utilised in this paper with respect to the two proposed mechanism, PCT-PDP and FRF-PDP.

Data prioritisation is concerned with the applications of techniques to a data source to determine which examples should be "ranked" higher than others. The techniques used range from the application of simple heuristics to the application of sophisticated models generated using the tools and techniques of machine learning. In the medical domain, the term "triage" is frequently used instead of prioritisation; especially in the context of hospital emergency departments. The Emergency Severity Index (ESI), a five-level emergency department triage algorithm, that arranges patients into five tiers from 1 (most urgent) to 5 (least urgent), is frequently used. In [11] several machine learning models, Random Forest and Lasso regression, were adopted to categorise patients using the five-level ESI stratification. Favourable comparisons were made with the ESI algorithm. In [4] a support vector machine, coupled with free text analysis, was used to build an emergency department prioritisation prediction model. However, the work in [11] and [4] was directed at supervised machine learning approaches that assumed a suitable training set was available, not the case with respect to the focus of the work presented in this paper. The work reported in [11] and [4] was also confined to emergency department triage only.

To the best knowledge of the authors there is little work directed at the use of machine learning to prioritise pathology data in the absence of a "ground truth". However, one notable approach, already referenced in the introduction to this paper, is the Anomaly Detection (AD) mechanism proposed in [10]. In [10] the assumption was that an anomalous record was a priority record. This was identified by generating a cluster configuration of pathology results and attempting to fit a new pathology result into this cluster configuration; if the new result could not be fitted to an existing cluster it was assumed to be anomalous and therefore the pathology result should be prioritised. Thus the pathology result prioritisation problem was defined as a two class problem, priority versus non-priority. Two variations were considered: (i) Point-based AD and (ii) Time Series-based AD. The distinction being that the second considered patient history while the first did not. However, given a large number of priority pathology results these would no longer be considered to be anomalous and therefore not be prioritised. It can also be argued that just because a record is unusual does not necessarily mean it is a priority record. The opposite can also be argued, just because a record is common it does not necessarily mean it is not a priority record. The results reported in [10] are referred to later in this paper for evaluation purposes to compare with the operation of the PGT-PDP and FRF-PDP mechanisms described. In response to the above criticism of the AD mechanism, in [9], an event-based proxy ground truth prioritisation mechanism was proposed[9], the PGT-PDP mechanism also considered in this paper and discussed further in Section 4.

The pathology data considered in this paper is in the form of multi-variate time series. There are two popular time series formats: (i) instance-based and (ii) feature-based. Using the instance-based format the original time series format is

maintained. While using the feature-based representation, properties of the time series are used [17]. In this paper, the instance based format was adopted.

In the case of the PGT-PD mechanisms discussed in this paper, two approaches for generating the desired pathology data classification models were considered, kNN and LSTM. In the case of the proposed FRF-PDP mechanism, two approaches for forecasting pathology results were considered: LSTM regression and Facebook Profit (FP). Each of these approaches is considered in some further derail in the remainder of this literature review section.

The fundamental idea of kNN classification is to compare a previously unseen record, which we wish to label, with a "bank" of records whose labels are known. The class labels from the identified k most similar records are then used to label the previously unseen record. Usually $k = 1$ is adopted because it avoids the need for any conflict resolution. A significant issue when using kNN with respect to time series classification is the nature of the similarity (distance) measure to be used [18]. There are a number of similarity measure options, such as Euclidean, Manhattan and Minkowski distance measurement, but Dynamic Time Warping (DTW) is considered to be the most effective with respect to the instance-based format time series, and offers the additional advantage that the time series considered do not have to be of the same length [18]. DTW was therefore used for the work presented in this paper.

Among the various deep learning models that are available, the Long Short Term Memory (LSTM) Recurrent Neural Networks (RNN) model is one of most effective with respect to sequence (time series) data. LSTM can be applied in both the context of classification and forecasting. There are many examples in the literature where RNNs have been used with respect to time series analysis [14, 13, 20]. The ability to "memorise" and select which data in a sequence is important for predicting future values makes the LSTM model an obvious candidate for both the PGT-PDP and the FRF-PDP mechanism.

In recent years Facebook Prophet (FP) has attracted increasing attention due to its demonstrated better performance with respect to certain application domains [7, 8]. FP models can deal with "regime shift" where the characteristics of a time series (mean and variance) changes over time, in some cases suddenly and in others gradually. Intuitively, pathology data time series features such regime shifts. FP was therefore experimented with in the context of the proposed FRF-PDP mechanism. An alternative might have been Autoregressive Integrated Moving Average (ARIMA) model [2, 6].

3   Application Domain

This section presents the U&E application domain used as a focus for the work presented in this paper. The section is divided into two sub-sections. Sub-section 3.1 considers the practice of U&E testing, whist Sub-section 3.2 presents a formalism for the data.

3.1   U&E Testing

Urea and Electrolytes pathology testing (U&E testing) is commonly used to detect abnormalities of blood chemistry, primarily kidney (renal) function and dehydration. 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, for each pathology test, measurement of levels of: (i) Bicarbonate (bi), (ii) Creatinine (cr), (iii) Sodium (so) (iv) Potassium (po) and (v) Urea (ur). The measurement of each is referred to as a "task", thus we have five tasks per test. Therefore each U&E test result comprises five pathology values. Pathology results comprised of a number of tasks are not unusual. Normally, for each of the task results, there is a clinical reference range indicating whether the corresponding result value is abnormal or not. In the case of U&E data, abnormal levels in any of the tasks may indicate that the kidneys are not working properly. However, a one time abnormal result for a single task does not necessarily indicate priority. A new task result that is out of range for a patient who has a previous recent history of out of range task results, but the latest result indicates a trend back into the normal range, may not be a priority result either. Conversely, a new task result that is within the normal range for a patient who has a history of normal range task results, but the latest result indicates a trend heading out of the normal range, maybe a priority result. It is suggested that U&E pathology results can be prioritised more precisely if the trend of the historical records is taken into consideration. Thus for the mechanisms presented in this paper, historical results are considered.

3.2   Formalism

In the context of the U&E Testing domain, the data used for evaluation purposes with respect to the work presented in this paper comprised a set of clinical patient records, $\mathbf{D} = \{P_1, P_2, \dots\}$. Each patient record $P_j \in \mathbf{D}$ was of the form:

$$P_j = \langle Id, Date, Gender, T_{So}, T_{Po}, T_{Ur}, T_{Cr}, T_{Bi}, c \rangle \tag{1}$$

Where T_{so} to T_{bi} are five multi-variate time series representing, in sequence, pathology results for the five tasks typically found in a U&E test: Sodium (So), Potassium (Po), Urea (Ur), Creatinine (Cr) and Bicarbonate (Bi) and c is the class label(proxy) taken from a set of classes C. Each time series T_i has three dimensions: (i) pathology result, (ii) normal low and (iii) normal high. The normal low and high dimensions indicate a "band" in which pathology results are expected to fall. These values are less volatile than the pathology result values themselves, but do change for each patient over time. Thus each times series T_i comprises a sequence of tuples, of the form $\langle v, nl, nh \rangle$ (pathology result, normal low and normal high respectively).

   In the case of the PGT-PDP mechanism the class label for each record $P_j \in \mathbf{D}$ for the proxy ground truth was derived from the outcome event(s) associated with each patient. Three outcome events were considered: (i) Emergency Patient (EP), an In-Patient (IP) or an Out Patient (OP) which were correlated with

the priority descriptors "high", "medium" and "low" respectively. Hence $C = \{high, medium, low\}$. The proxy data, formulated as described above was also used for the evaluation pf FRF-PDP mechanism.

4   Proxy Ground Truth Pathology Data Prioritisation

The fundamental idea underpinning the PGT-PDP mechanism, as already noted, was that although no ground truth training data was available, the final destinations of patients were known, and hence these could act as a proxy for a ground truth. In order to validate this idea, two classification model generation methods were considered, the kNN-DTW approach and the LSTM-RNN approach. Thus two variations of the PGT-PDP mechanism were considered: (i) kNN-Based and (ii) LSTM based. Each is discussed in further detail in the following two sub-sections.

4.1 kNN-Based Proxy Ground Truth Pathology Data Prioritisation

The kNN-based variation of the PGT-PDP mechanism used the well known kNN classification algorithm. As already noted, kNN used a parameter k, the number of best matches we are looking for. For the work presented here $k = 1$ was used because it is most frequently used in the context of time series analysis [1]. DTW was used for the similarity measure because of its ability to operate with time series of different length and because it has been shown to be more effective than alternatives such as Euclidean distance measurement [18]. The disadvantage of DTW, compared to the Euclidean distance measurement, is its high computational time complexity of $O(x \times y)$ where x and y are the lengths of the two time series under consideration. Two strategies for addressing this computational overhead were adopted with respect to the work presented here: (i) early-abandonment [12] and (ii) LB-Keogh lower bounding [16]. The first is a strategy whereby the accumulative distance between two time series is repeatedly checked as the calculation progresses and if the distance exceeds the best distance so far the calculation is "abandoned" [12]. The second involves preprocessing the time series to be considered by comparing the time series using an alternative "cheaper" technique and pruning those that are unlikely to be close matches and applying DTW to the remainder. One example of this, and that adopted with respect to the work presented in this paper, is the lower bounding technique proposed in [16], the so called LB-Keogh technique. This operates by superimposing a band, defined by a predefined offset value referred to as the lower bound, over each time series in the bank and calculating the complement of the overlap with the new time series. Where the calculated value exceeds a given threshold ϵ the associated time series is pruned.

The traditional manner in which kNN is applied in the context of time series analysis is to compare a query time series with the time series in the kNN bank. In the case of the U&E test data prioritisation scenario considered here the process involved five comparisons, once for each time series in the query record

P_j, $T_{q_{so}}$, $T_{q_{po}}$, $T_{q_{ur}}$, $T_{q_{cr}}$ and $T_{q_{bi}}$. In addition, traditional kNN is applied to univariate time series, in the U&E pathology case each task time series was a three-dimensional multi-variate time series: (i) pathology value, (ii) normal low and (iii) normal high. Thus, from the foregoing, for each comparison five distance measures were obtained and combined to obtain a final prioritisation. These five distance measures therefore need to be combined to give a final prioritisation.

The final prioritisation was decided using a rule called the "High priority first and voting second" rule (Rule 1), which will be explained in more details in Sub-section 4.3 below. The application of kNN for prioritising pathology results P_j was as follows:

1. Calculate the average LB-keogh overlap for the five component time series separately and prune all records in D where the overlap for any one time series was greater than a threshold ϵ, to leave D'.
2. Apply DTW, with early-abandonment to each pair $\langle T_{q_i}, T_j \in D' \rangle$ where i indicates the U&E task.
3. Assign the class label c to the time series T_{q_i}, of a patient record P_j, associated with the most similar time series $T_i \in D'$.
4. Use the final prioritisation rule, Rule 1, to decide the final priority level for P_j (see Sub-section 4.3 below for further details).

With respect to the above the choice of the value for ϵ is of great importance as it affects the efficiency and the accuracy of the similarity search. According to [5], there is a threshold value for ϵ whereby the time complexity for the lower bounding is greater than simply using DTW distance without lower bounding. The experiments presented in [5] demonstrated that this threshold occurs when the value for ϵ prunes 90% of the time series in D. For the work presented in this paper $\epsilon = 0.159$ was used because, on average, this resulted in 10% of the time series in D being retained.

4.2   LSTM-Based Proxy Ground Truth Pathology Data Prioritisation

Using the LSTM classification approach, and given the U&E pathology prioritisation scenario, the PGT-PDP mechanism required the training of five LSTM models, one for each task: $LSTM_{so}$, $LSTM_{po}$, $LSTM_{ur}$, $LSTM_{cr}$ and $LSTM_{bi}$. Figure 1 illustrates the adopted LSTM architecture.

From Figure 1 it can be seen that the overall structure is expressed in four "layers" (i) the input layer, (ii) the model layer, (iii) the Softmax layer and (iv) the decision layer. The input is the five component task parts $T_{q_{so}}$, $T_{q_{po}}$, $T_{q_{ur}}$, $T_{q_{cr}}$ and $T_{q_{bi}}$ of data set D. Thus for each task, a multi-variate time series $T_i = \{V_1, V_2, ..., V_m\}$ was established, where $V - J$ is a tuple of the form presented earlier, and $m \in [l_{min}, l_{max}]$. Where necessary each time series T_i is padded to the maximum length, l_{max} using the mean values for the pathology test values, normal low and normal high values in T_i. Each time series T_i is then passed to the model layer and the LSTM constructed. And for each LSTM two hidden
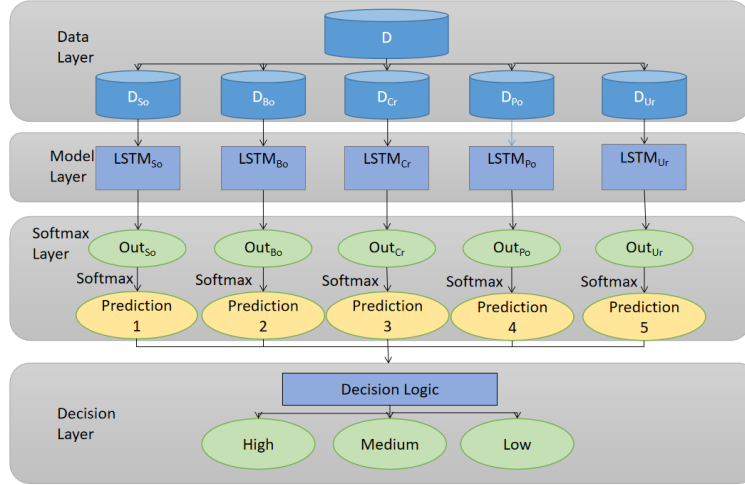
Fig. 1. LSTM architecture for the PGT-PDP mechanism [10]

layers of cells were adopted. The output from the LSTM layer is then passed to a Softmax layer, where the first predictions will be made with respect to the task level. The Softmax function for normalising the output of each single task LSTM model was as follows:

$$y_i = \frac{e^{a_i}}{\Sigma_{k=1}^{|C|} e^{a_k}} \quad \forall i \in 1...C \tag{2}$$

Where: (i) $|C|$ is the number of classes (three in this case) and (ii) a_i is the output of the LSTM layer.

The last layer in Figure 1 is the decision layer where the final label is derived. A logits component rule (Rule 2) in this layer is used to determine the final pathology result prioritisation label. Sub-section 4.3 presents further detail regrading Rule 2.

4.3   PGT-PDP Prioritisation Rules

Two rules were utilised in the foregoing to assign an overall pathology result prioritisation: (i) the final prioritisation rule, Rule 1, for kNN-based PGT-PDP, and (ii) the logits component rule, Rule 2, for LSTM-based PGT-PDP. Both are discussed in further detail in this sub-section. Rule 1 for prioritisation using kNN was as follows:

Rule 1 If one of the class labels is "high" the overall class label is high, otherwise voting will be adopted to derive the overall class label.

Rule 2 For prioritisation using $LSTM$ was a folows:

Rule 2 If there exists a prediction that equates to 'High' for one of the tasks then the overall prediction is high, otherwise average the five outputs produced by the Softmax function and choose the class with the maximum probability.

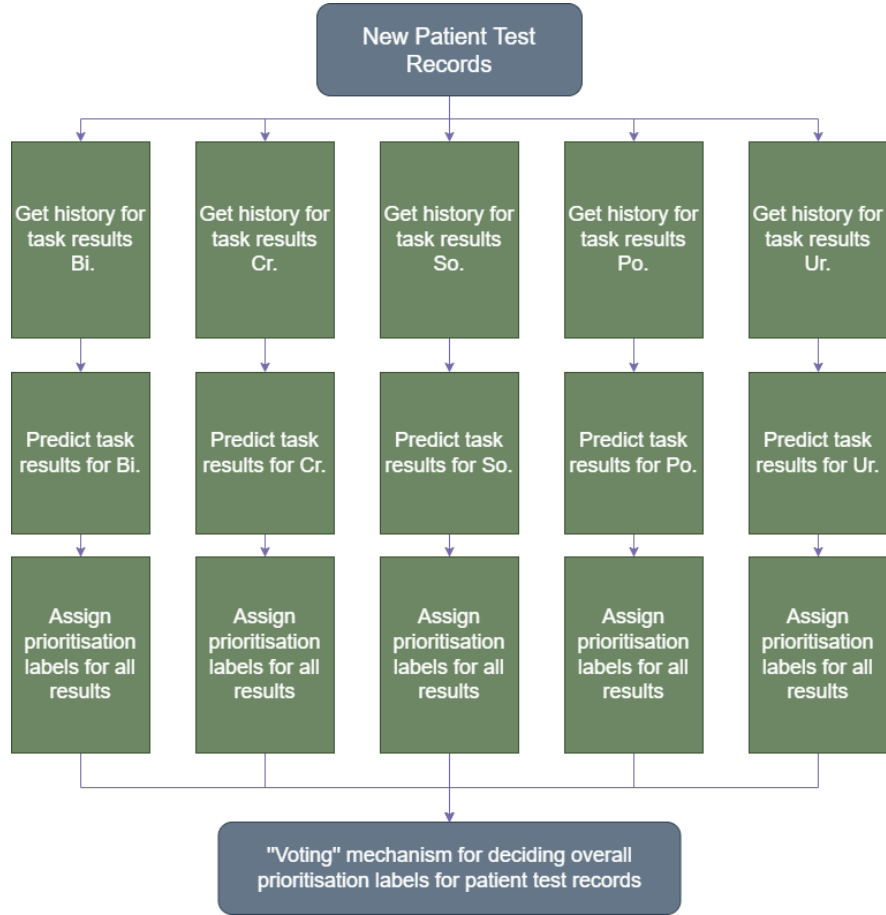## 5   Future Result Forecast Pathology Data Prioritisation

The fundamental idea promoted using the FRF-PDP mechanism is that, given a current pathology result P_j, comprised of a number of task value sequences, the sequences can be appended to using forecasted next values, the $n+1$ values. If a predicted $n+1$ value is out of the normal clinical reference range the associated current task value should be labelled as a "priority" value. Thus given a new patient pathology result, P_j, the historical records for the patient were used to predict the $n+1$ values which were then used to label P_j. The prioritisation problem thus becomes a "one-step" time series forecasting problem. The entire FRF-PDP process is illustrated in Figure 2. The forecasting can be done using any appropriate forecasting algorithm, however two were considered with respect to the work described here, LSTM regression and FP. Two variation of the FRF-PDP mechanisms were thus considered: (i) LSTM-based and (ii) FP-based. Further detail concerning these two variations is presented in the following two sub-sections, Subsections 5.1 and 5.2.

5.1   LSTM-Based Forecasting Future Results Pathology Data Prioritisation

LSTM classification was considered with respect to the PGT-PDP mechanisms described previously in Section 4. However, due to its versatility, LSTM can also be adopted for regression/prediction purposes. Hence it was adopted with respect to the FRF-PDP mechanism. LSTM prediction was applied to single tasks individually because earlier work, not shown here, had demonstrated that considering all tasks simultaneously produced a poor model. This was thought to be because of potential correlations between tasks that caused the model to become over-complex and lead to unstable solutions, consequently compromising the generalisability of the prioritisation model.

Thus individual LSTM models were generated for each task (as indicated in Figure 2). For the evaluation data used in this paper this resulted in five LSTM models: $LSTM_{so}$, $LSTM_{po}$, $LSTM_{ur}$, $LSTM_{cr}$ and $LSTM_{bi}$. The overall prioritisation for each pathology result was then obtained by applying a predefined rule (discussed at the end of this sub-section). Further challenges, as in the case of LSTM-based PGT-PDP, were that:

1. Pathology (task) data features irregular time spacing, the time period between points was not uniform, both with respect to individual patients and between patients. LSTM prediction assumed regular spaced time series data. For the evaluation set used with respect to this paper the spacing varied from 3 to 52 days.
2. The overall length of the time series for each patient was not the same. LSTM requires all time series to be of the same length. Thus a padding techniques was applied to the time series so that they were all of the same length and therefore could be used to generate a LSTM model.

Fig. 2. Future Result Forecast Pathology Data Prioritisation process

The overall architecture of the generated LSTM models was similar to the architecture of the LSTM used with respect to the PGT-PDP mechanism as shown in Figure 1. The LSTM comprises three layers: (i) the input layer, (ii) the model layer and (iii) the decision layer. Each model took as input a task time series $T_i = [t_{i_1}, t_{i_2}, \ldots, t_{i_n}, t_{i_{n+1}}$, where t_{i_n} is the current test result value ,and $t_{i_{n+1}}]$ is the task value to be predicted. Where necessary, each time series T_i was padded to the maximum length, l_{max}, using the mean values for the pathology test values in T_i. For each LSTM model, 2 layers of LSTM blocks were used for the model layer, linked by the default Sigmoid activation function.

After obtaining all of the predicted future result values, one from each LSTM model, a "label computation function" was applied to predict the prioritisation labels for each task result. The equation for deriving a prioritisation label c taken from a set of labels $C = \{priority, non\text{-}priority\}$ is given by Equation

3. Where: (i) $t_{i_{n+1}}$ is the predicted task next value, and (ii) x_l and x_h are the referenced clinical normal low and normal high values for the corresponding task and patient.

$$c = \begin{cases} \text{non-priority,} & \text{if } x_l < t_{i_{n+1}} \leq x_h \\ \text{priority,} & \text{otherwise} \end{cases} \tag{3}$$

Given the above, the process for generating a pathology result prioritisation label using a set of LSTM models (one per task) was as follows:

1. Given a new pathology test result for a patient j comprised of m tasks, thus m values, $\langle v_1, v_2, \ldots, v_m \rangle$, where v_k is a pathology value, add these values to the historical data for the patient j to generate m time series of the form $T_i = \{t_{i_1}, t_{i_2} \ldots t_{i_n}\}$ where $t_{i_n} = v_k$.
2. Input each time series into the corresponding LSTM model to produce predicted values $t_{i_{n+1}}$.
3. Use Equation 3 to assign a label to each predicted values $t_{i_{n+1}}$.
4. Apply "Voting" to the task prioritisation labels to acquire the final prioritisation for the pathology result (see Sub-section 5.3 below for further detail).

5.2   FP-Based Future Result Forecast Pathology Data Prioritisation

For the FP-based FRF-PDP variation Facebook Profit (FP) was adopted. FP was created by Facebook's Core Data Science team for univariate time series forecasting. FP is an open source library for forecasting with time series that feature significant seasonal variation and that span a number of seasons. Unlike LSTMs, FP operates well with missing values. Models generated using FP operate in an "additive regressive" manner. The fundamental FP equation is as follows:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t \tag{4}$$

Where: $g(t)$ is the trend factor, and $s(t)$ and $h(t)$ are the seasonal and holiday components respectively, and ϵ_t is an error term [15, 3].

Although specifically designed for use with time series that feature seasonality, the components $s(t)$ and $h(t)$ can be ignored. For the work presented in this paper the focus was on $g(t)$. There are two trend models for $g(t)$: (i) a non-linear saturating growth model and (ii) a piece-wise linear model. For the experiments reported on later in this paper the non-linear growth model was used, because of the non-linear nature of the task time series under consideration. Overall, the adopted procedure for pathology forecasting using FP was similar to that used with respect to LSTM-based FRF-PDP. Equation 3 was again used. The differences were: (i) the date of each point in each patient time series was added to the input training data, and (ii) no padding was required as FP was able to deal with time series data with different lengths.

5.3   Final FRF-PDP Classification

Once we have forecast the next value in a task time series and used Equation 3 to label each task value in a new pathology result using a two class classification, priortity versus non-priority, the next stage was to determine the overall three-class (high, medium, low) prioritisation of the given pathology result. To this end a simple conditional rule was defined as follows:

$$c = \begin{cases} high, & \text{if } m \geq 4 \\ medium, & \text{if } 3 \leq m < 4 \\ low, & \text{otherwise} \end{cases} \tag{5}$$

where m is the number of tasks classified as "priority" tasks (see above). In the case of the evaluation presented in the following section, the U&E evaluation data used featured $m = 5$.

6   Evaluation

This section presents the evaluation of the two proposed mechanism and their two variation in each case: (i) kNN-based PCT-PDP, (ii) LSTM-based PCT-PDP, (iii) LSTM-based FRF-PDP and (iv) FP-based FRF-PDP. As established in the foregoing, the pathology prioritisation problem was conceptualised as a three class problem; hence solutions to the problems could be evaluated using the "standard" accuracy, precision, recall and F1 metrics. 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. The objectives of the evaluation were as follows:

1. To identify the optimum parameter settings for the proposed approaches.
2. To compare the operation of the kNN and LSTM-based PGT-PDP approaches.
3. To compare the operation of the FP and LSTM-based FRF-PDP approaches.
4. To compare the overall performance of the proposed mechanisms.

For the evaluation U&E pathology data provided by the Wirral Teaching Hospital in Merseyside in the UK was used. From this raw data an evaluation data D was created.

The remainder of this section is organised as follows. An overview of the U&E evaluation data sets is given in Sub-section 6.1. The LSTM parameter setting are considered in Sub-section 6.2. The results with respect to remainder of the above objectives are then discussed in Sub-sections 6.3, 6.4 and 6.5 respectively.

6.1   Evaluation Dataset

The Wirral Teaching Hospital U&E pathology test data comprised four data tables: (i) Emergency Data (ED), (ii) In-Patient (IP), (iii) Out-Patient (OP) and

Laboratory Results (LR). The data tables comprised of 180,865, 226,634, 955,318 and 532,801 records respectively. The primary table used for the evaluation of the proposed data prioritisation mechanisms considered in this paper was the LR data table. A single pathology task record in this table was of the form:

$$\langle ID, Task, Date, Value, Max, Min, Gender \rangle \tag{6}$$

This was used to generate a database $\mathbf{D}$ where, as noted in Section 3.2, each patent record $P_j \in \mathbf{D}$ was of the form:

$$P_j = \langle Id, Date, Gender, T_{So}, T_{Po}, T_{Ur}, T_{Cr}, T_{Bi}, c \rangle \tag{7}$$

Some data cleaning, such as removing patients with missing or non-numeric task values and feature scaling was undertaken with respect to $\mathbf{D}$.

For the FRF-PDP mechanism, as noted in Section 5, an initial two class prioritisation, priority versus non-priority, was undertaken. The two-class distribution, per task, is given in Table 1. This two-class prioritisation was then converted into the desired three-class prioritisation, high, medium and low. In the case of the PGT-PDP mechanism the classification models were generated directly by using this three-class prioritisation. Table 2 shows the three-class distribution per patient.

Table 1. Two-class (priority, non-priority) distribution per task in the U&E LAB Dataset

Task Name	Task Labels		Total
	Prior.	*Non-prior.*	
Bicarbonate (bi)	1,099	2,635	3734
Creatinine (cr)	1,639	2,095	3734
Sodium (so)	861	2873	3734
Potassium (po)	397	3,337	3734
Urea (ur)	1,373	2,361	3734
Total	5,369	13,301	18,670

From the Table 2, it can be seen that there was a significant imbalance between the number of patients within each class. This is not an issue for the forecasting mechanism as the data was not required to train a classification model. But for the PGT-PDP mechanism, especially when using LSTMs, it was an issue, as highly imbalanced data may pose a bias towards the majority class. Thus we used an oversampling techniques to address this issue with respect to the LSTM-based PGT-PDP variation.

6.2   Parameter Settings

Two classification model generator approaches, kNN and LSTM; and two prediction mechanisms, LSTM and FP, were considered with respect to the two

Table 2. Three-class (high, medium, low) distribution per patient in the U&E LAB Dataset

Event Proxy Truth	No. Patients Test Records
High	255
Medium	123
Low	3356
Total	3734

approaches presented in this paper, PGT-PDP and FRF-PDP. All these approaches require parameters. As already discussed above kNN required a value for k and a value for the associate $\epsilon = 0.159$ threshold. The values for these parameters are fairly easy to specify. As discussed above, $k = 1$ and $\epsilon = 0.159$ were used.
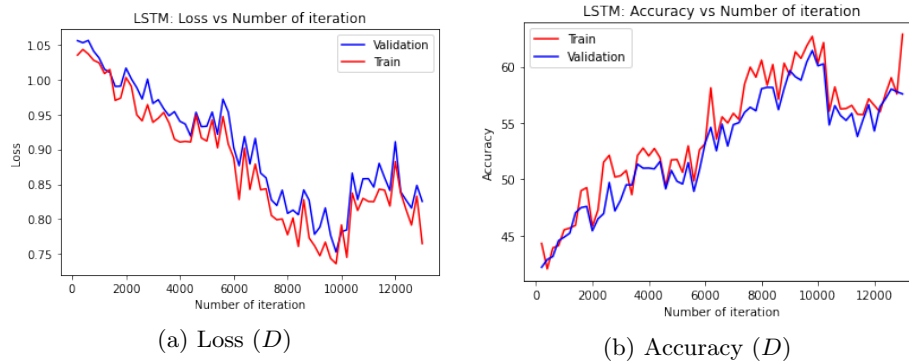
Fig. 3. Loss and Accuracy curves for LSTM-based PGT-PDP as reported in [9]

LSTMs require three parameters: (i) the batch size, (ii) the learning rate and (iii) the number of epochs. The general way of finding the best LSTM parameters is to analyse the learning curves and accuracy plots of the training and validation data. The most popular learning curve used for this purpose is the "loss over time" curve. Loss measures the model error, which represents the performance of the model. As loss is usually a function indicating the difference between the derived values and the actual values, the lower the loss the better the model performance. A grid search was used to find the best combination of parameters according to the generated loss over time curves. Although the drawback of the grid search method is that it is computationally expensive, it is a simple and effective technique when there is a prior knowledge of the parameters range. Grid search, coupled with loss over time curves, were therefore used with respect to the work presented here. This was also the approach taken with respect to the

Table 3. Parameter setting for LSTM-based PGT-PDP

Parameter Name	Task				
	Bi	*Cr*	*So*	*Po*	*Ur*
Batch size	512	128	256	512	512
Learning Rate	0.01	0.01	0.01	0.01	0.01
Epochs	1,000	1,000	1,000	1,000	1,000

Table 4. Parameter setting for LSTM-based FRF-PDP

Parameter Name	Task				
	Bi	*Cr*	*So*	*Po*	*Ur*
Batch size	128	64	128	128	128
Learning Rate	0.01	0.01	0.01	0.01	0.01
Epochs	45	45	45	45	45

work described in [9]. In this manner the most appropriate parameters with respect to each pathology task were identified. The average loss and accuracy plots obtained, as reported in [9], are presented in Figure 3. In the figure the x-axis gives the number of times the weights in the network were updated, and the y-axis the loss value. From the figures, it can be seen that convergence is not obvious in all cases. Possible reasons could be: (i) the oversampling techniques used for addressing the class imbalanced problem was not suitable, and the information insufficient problem still exists, so there is not enough information for the LSTM to learn from; or (ii) that the proxy ground truth data set may not be entirely representative. Whatever the case, the derived LSTM parameters for the PGT-PDP mechanism are given in Table 3, and those for the FRF-PDP mechanism in Table 4. From the table it can be observed that: (i) the most appropriate batch size was different for each task, (ii) the learning rate was the same in all cases, and (iii) the number of epochs was required to be higher for PGT-PDP than for FRF-PDP.

The parameters for FP prediction approach were fairly straight forward to set. The original FP algorithm required three categories of parameters, which controlled the trend, seasonality and holiday effects. However, As noted above, in the case of the FP-based FRF-PDP mechanism only the trend part was utilised. Therefore the two parameters of concern were: (i) the number of trend changes within the data, and (ii) the change points scale which determines the flexibility of the trend, and in particular how much the trend changes at the trend change points. To identify the most appropriate FP parameter settings ranges of values wee considered. For the number of trend changes the selected range was from 3 to 52, incrementing in steps of 1. For the change points scale the selected range was from 0.01 to 0.5, incrementing in steps of 0.01. The identified parameters are given in Table 5.

Table 5. Parameter settings for FP-based FRF-PDP

Parameter Name	Task				
	Bi	*Cr*	*So*	*Po*	*Ur*
N_changepoints	3	3	3	3	3
ChangePoints Scale	0.01	0.03	0.05	0.06	0.05

6.3 Comparison of Proxy Ground Truth Pathology Data Prioritisation Approaches

The comparison results for the kNN and LSTM-based PGT-PDP approaches, as reported in [9], are given in Tables 6 and 7 (best results for each fold highlighted in bold font). Five fold cross validation was used to generate these results. The overall average (Ave) and standard deviation (SD) are also presented in the last two rows. Recall that a low SD values indicates that little variation existed across the folds. It can be observed from the table that LSTM-based PGT-PDP consistently out performed the kNN-based PGT-PDP. It might be argued that the recall and precision values are relatively low, this could be because the irregular distribution of the time stamps within the training data may have had an adverse effect.

Table 6. Average Precision and Recall of kNN-based PGT-PDP [9]

Fold Num.	Acc.	Pre. High	Pre. Medium	Pre. Low	Rec. High	Rec. Medium	Rec. Low
1	0.585	**0.414**	0.400	0.545	**0.637**	0.577	0.666
2	0.632	**0.534**	0.688	0.578	**0.678**	0.467	0.714
3	0.576	**0.412**	0.541	0.674	**0.588**	0.535	0.647
4	0.523	**0.598**	0.541	0.634	**0.712**	0.4688	0.505
5	0.566	**0.444**	0.384	0.598	**0.541**	0.487	0.785
Ave	0.576	**0.480**	0.510	0.605	**0.631**	0.507	0.663
SD	0.039	**0.082**	0.124	0.050	**0.068**	0.047	0.103

Table 7. Average Precision and Recall of LSTM-based PGT-PDP [9]

Fold Num.	Acc.	Pre. High	Pre. Medium	Pre. Low	Rec. High	Rec. Medium	Rec. Low
1	0.671	**0.578**	0.374	0.711	**0.811**	0.641	0.412
2	0.642	**0.475**	0.552	0.735	**0.758**	0.468	0.577
3	0.622	**0.553**	0.577	0.708	**0.669**	0.547	0.703
4	0.608	**0.615**	0.714	0.699	**0.712**	0.563	0.697
5	0.645	**0.466**	0.766	0.596	**0.699**	0.476	0.778
Ave	0.638	**0.538**	0.597	0.690	**0.730**	0.539	0.633
SD	0.024	**0.065**	0.120	0.054	**0.056**	0.071	0.143

6.4   Comparison of Future Results Forecast Pathology Data Prioritisation Approaches

Using the FRF-PDP mechanism forecast models were built for each task and then the results combined to produce a final priority classification. Hence, the evaluation was conducted on two levels, the Local Task level and the Global Test level. Accuracy, Recall, Precision, F1 score and Mean Square Error(MSE) were again used as the evaluation metrics, together with Five fold cross validation.

The results obtained at the local task level, using LSTM-based and FP-based FRF-PDP, are given in Tables 8 and 9 respectively. The tables are divided into two parts. The right hand side gives the average MSE values recorded with respect to the prediction for each task. The left had side gives the the average accuracy (Acc), Precision (Pre), Recall (Rec) and F1 values recorded with respect to the prioritisation classification.

From the Tables 8 and 9 it can be seen that the overall recall using FP-based FRF-PDP is much higher than when using LSTM-based FRF-PDP. In the context of pathology data prioritisation high recall is generally considered desirable over precision, given that it would be acceptable to classify a non-priority record to be a priority record mistakenly, rather than the contrary. But it can also be noticed that the average MSE using FP-based FRF-PDP, for the majority of the tasks is higher than when using LSTM-based FRF-PDP, especially for the $MSE_{(so)}$, where the difference is considerable. Perhaps this is because FP takes account of the individual date stamps of values and because it does not require padding. Another possible reason for the better performance of FP is that the predicted values derived by FP for different tasks are further away from the true mean values ("outside" of the true values), than when LSTM is used. This would mean that the predicted values are likely to fall outside of the clinical reference range and hence be classified as "priority" values. This in turn would explain the relatively higher recall values when using FP compared to LSTM, as there is a higher chance of a priority record being labelled correctly; using FP the high priority classification is therefore favoured. However, this conjecture may also cause concerns about the stability and generalisability of the model. Whatever the case, in summary it can be seen that the overall performance of FP-based FRF-PDP, at least at the Local Task level, is better than LSTM-based FRF-PDP.

Recall that at the global level three proxy ground truth class labels were used (high, medium and low). The evaluation results obtained are given in Tables 10 and 11; best results for each fold highlighted in bold font. From the Tables it can be seen that the average recall for LSTM and FP-based FRF-PDP is higher than the precision for almost all of the three prioritisation classes. The average recall for the classification of the high prioritisation class was 0.739 using LSTM-based FRF-PDP, and 0.712 using FP-based FRF-PDP; both higher than for the other classes; which, as noted earlier, is desirable as it entails a lower chance for missing any high priority pathology results. However, the average precision for the high prioritisation, using both approaches, is far lower than their recall, which may cause low prioritisation pathology results to be classified as higher level results.

Table 8. Local Task Level Accuracy, Recall, Precision, F1-score and MSE using LSTM-based FRF-PDF

Fold Num.	Acc.	Pre.	Rec.	F1	MSE(bi)	MSE(cr)	MSE(so)	MSE(po)	MSE(ur)
1	0.611	0.524	0.371	0.412	16.29	421.56	4.92	0.37	4.71
2	0.657	0.544	0.333	0.502	21.01	224.68	6.47	0.25	5.68
3	0.555	0.639	0.278	0.530	26.36	602.24	24.31	0.61	7.98
4	0.586	0.410	0.362	0.408	21.43	478.12	33.75	0.78	13.21
5	0.512	0.682	0.403	0.577	9.78	541.45	9.74	1.28	12.56
Ave	0.584	0.560	0.345	0.486	18.974	453.61	15.838	0.658	8.828
SD	0.049	0.095	0.042	0.066	5.594	129.494	11.285	0.362	3.485

Table 9. Local Task Level Accuracy, Recall, Precision, F1-score and MSE using FP-based FRF-PDF

Fold Num.	Acc.	Pre.	Rec.	F1	MSE(bi)	MSE(cr)	MSE(so)	MSE(po)	MSE(ur)
1	0.674	0.523	0.701	0.511	15.31	602.31	81.32	3.41	9.45
2	0.712	0.612	0.682	0.576	17.45	487.56	77.45	1.74	8.73
3	0.667	0.662	0.674	0.611	21.68	512.37	98.56	2.01	6.55
4	0.721	0.549	0.660	0.554	19.21	345.87	74.31	0.93	10.32
5	0.732	0.522	0.736	0.568	23.85	324.63	88.63	1.21	3.96
Ave	0.701	0.574	0.691	0.564	19.500	454.548	84.054	1.86	7.802
SD	0.026	0.055	0.026	0.0324	3.017	104.839	8.688	0.863	2.291

Thus the benefit of the prioritisation might not be significant if all results were classified as high priority. In the extreme situation, the prioritisation might thus be meaningless!

6.5 Comparison of the Overall Performance of The Proposed Mechanisms

Table 12 shows a comparison of the performance between the four approaches proposed in this paper. The table includes the results, reported in [10], obtained when using the two Anomaly Detection (AD) pathology prioritisation approaches: (i) Point-based AD and (ii) Time Series-based AD. Note that in [10] only the overall precision and recall values were reported; and hence these have been reported in Table 12. From the table, it can be seen that the recall for the high prioritisation level using the both the PCT-PDP and FRF-PDP mechanisms is good, over 70%, although the overall precision is lower than when using the FRF-PDP mechanism. The AD mechanisms produced the worst performance. The reason behind the poor performance of the Point-based AD and Time Series AD is probably because, as noted earlier, anomalous pathology results do not necessarily equate to priority pathology results. From the table, the best performing mechanism was PGT-PDP, with little distinction between the LSTM abd the kNN variations. The results given in the Table 12 of course need

Table 10. Global Test Level Precision and Recall for the overall prioritisation using LSTM-based FRF-PDP

Fold Num.	Acc.	Pre. High	Pre. Medium	Pre. Low	Rec. High	Rec. Medium	Rec. Low
1	0.512	**0.310**	0.325	0.338	**0.694**	0.546	0.516
2	0.572	**0.317**	0.412	0.454	**0.711**	0.597	0.537
3	0.582	**0.447**	0.334	0.497	**0.824**	0.611	0.498
4	0.611	**0.301**	0.378	0.523	**0.765**	0.547	0.564
5	0.504	**0.477**	0.364	0.437	**0.702**	0.511	0.536
Ave	0.556	0.370	0.363	0.450	0.739	0.562	0.530
SD	0.041	0.076	0.031	0.064	0.049	0.037	0.022

Table 11. Global Test Level Precision and Recall for the overall prioritisation using FP-based FRF-PDP

Fold Num.	Acc.	Pre. High	Pre. Medium	Pre. Low	Rec. High	Rec. Medium	Rec. Low
1	0.587	**0.456**	0.441	0.663	**0.714**	0.402	0.537
2	0.572	**0.437**	0.467	0.538	**0.820**	0.331	0.584
3	0.513	**0.525**	0.493	0.507	**0.551**	0.227	0.623
4	0.538	**0.533**	0.541	0.463	**0.698**	0.380	0.609
5	0.607	**0.488**	0.552	0.497	**0.775**	0.534	0.497
Ave	0.563	0.488	0.499	0.534	0.712	0.375	0.570
SD	0.034	0.037	0.042	0.069	0.091	0.100	0.047

Table 12. Comparison of PGT-PDP, FRF-PDP and Anomaly Detection-based PDP

Method	Acc.	Precision			Recall		
		High	Med.	Low	High	Med.	Low
LSTM-based FRF-PDP	0.56	**0.37**	0.36	0.45	**0.74**	0.56	0.53
FP-based FRF-PDP	0.56	0.49	0.50	**0.53**	**0.71**	0.38	0.57
LSTM-based PGT-PDP [9]	0.61	0.58	0.55	**0.69**	**0.79**	0.59	0.63
kNN-based PGT-PDP [9]	0.60	0.42	0.51	**0.85**	0.70	0.55	**0.75**
Point-based AD [10]	0.34	0.35			0.43		
Time Series AD [10]	0.45	0.45			0.43		

to be tempered with the observation that the ground truth used was a proxy for the real ground truth.

7   Conclusions

This paper has discussed two mechanisms for prioritising pathology results in the context of the absence of a ground truth; pathology results made up of a number of task values. The first mechanism, the Proxy Ground Truth Pathology

Data Prioritisation (PGR-PDP) mechanism was underpinned by the fundamental idea that new pathology test results could be classified (using a three level priority classification) according to the anticipated outcome event associated with the result. Two variations were considered, kNN-based and LSTM-based PGR-PDP. The second mechanism, the Future Result Forecast Pathology Data Prioritisation (FRF-PDP) mechanism was underpinned by the fundamental idea that new pathology results could be classified according to whether predicted future test values were inside or outside the expected normal range. Two variations were again considered, LSTM-based and FP-based FRF-PDP. The proposed approaches were comparatively evaluated using U&E pathology test data which comprised five tasks. The final comparative results demonstrated that the PGR-PDP mechanism produced the best recall and precision of 0.79 and 0.58 respectively for the high priority class. For future work the authors intend to: (i) investigate the generation of artificial evaluation data sets to provide for a more comprehensive evaluation, and (ii) undertake a comprehensive collaborate with clinicians to obtain feedback regarding the prioritisations 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 mechanisms presented in this paper.

References

1. Anthony Bagnall, Jason Lines, Aaron Bostrom, James Large, and Eamonn Keogh. The great time series classification bake off: a review and experimental evaluation of recent algorithmic advances. *Data Mining and Knowledge Discovery*, 31(3):606–660, 2017.
2. Domenico Benvenuto, Marta Giovanetti, Lazzaro Vassallo, Silvia Angeletti, and Massimo Ciccozzi. Application of the arima model on the covid-2019 epidemic dataset. *Data in brief*, 29:105340, 2020.
3. Mustafa Daraghmeh, Anjali Agarwal, Ricardo Manzano, and Marzia Zaman. Time series forecasting using facebook prophet for cloud resource management. In *2021 IEEE International Conference on Communications Workshops (ICC Workshops)*, pages 1–6. IEEE, 2021.
4. Steven Horng, David A Sontag, Yoni Halpern, Yacine Jernite, Nathan I Shapiro, and Larry A Nathanson. Creating an automated trigger for sepsis clinical decision support at emergency department triage using machine learning. *PloS one*, 12(4):e0174708, 2017.
5. Zheng-xin Li, Shi-hui Wu, Yu Zhou, and Chao Li. A combined filtering search for dtw. In *2017 2nd International Conference on Image, Vision and Computing (ICIVC)*, pages 884–888. IEEE, 2017.
6. Zhenwei Li, Jing Han, and Yuping Song. On the forecasting of high-frequency financial time series based on arima model improved by deep learning. *Journal of Forecasting*, 39(7):1081–1097, 2020.
7. Sakib Mahmud. Bangladesh covid-19 daily cases time series analysis using facebook prophet model. *Available at SSRN 3660368*, 2020.
8. JC Park, BP Chang, and N Mok. 144 time series analysis and forecasting daily emergency department visits utilizing facebook's prophet method. *Annals of Emergency Medicine*, 74(4):S57, 2019.

9. 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.

10. JKing Qi, Girvan Burnside, Paul Charnley, and Frans Coenen. Ranking pathology data in the absence of a ground truth. In *Artificial Intelligence XXXVIII*, pages 209–223. Springer International Publishing, 2021.

11. Yoshihiko Raita, Tadahiro Goto, Mohammad Kamal Faridi, David FM Brown, Carlos A Camargo, and Kohei Hasegawa. Emergency department triage prediction of clinical outcomes using machine learning models. *Critical care*, 23(1):1–13, 2019.

12. Thanawin Rakthanmanon, Bilson Campana, Abdullah Mueen, Gustavo Batista, Brandon Westover, Qiang Zhu, Jesin Zakaria, and Eamonn Keogh. Searching and mining trillions of time series subsequences under dynamic time warping. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 262–270, 2012.

13. Bhargava K Reddy and Dursun Delen. Predicting hospital readmission for lupus patients: An rnn-lstm-based deep-learning methodology. *Computers in biology and medicine*, 101:199–209, 2018.

14. Murtaza Roondiwala, Harshal Patel, and Shraddha Varma. Predicting stock prices using lstm. *International Journal of Science and Research (IJSR)*, 6(4):1754–1756, 2017.

15. Toni Toharudin, Resa Septiani Pontoh, Rezzy Eko Caraka, Solichatus Zahroh, Youngjo Lee, and Rung Ching Chen. Employing long short-term memory and facebook prophet model in air temperature forecasting. *Communications in Statistics-Simulation and Computation*, pages 1–24, 2020.

16. Sharad Vikram, Lei Li, and Stuart Russell. Handwriting and gestures in the air, recognizing on the fly. In *Proceedings of the CHI*, volume 13, pages 1179–1184, 2013.

17. Liang Wang, Xiaozhe Wang, Christopher Leckie, and Kotagiri Ramamohanarao. Characteristic-based descriptors for motion sequence recognition. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 369–380. Springer, 2008.

18. Xiaoyue Wang, Abdullah Mueen, Hui Ding, Goce Trajcevski, Peter Scheuermann, and Eamonn Keogh. Experimental comparison of representation methods and distance measures for time series data. *Data Mining and Knowledge Discovery*, 26(2):275–309, 2013.

19. Wenchao Xing and Yilin Bei. Medical health big data classification based on knn classification algorithm. *IEEE Access*, 8:28808–28819, 2019.

20. Hanzhong Zheng and Dejia Shi. Using a lstm-rnn based deep learning framework for icu mortality prediction. In *International Conference on Web Information Systems and Applications*, pages 60–67. Springer, 2018.