Data Augmentation for Pathology Prioritisation: An Improved LSTM-Based Approach

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

Abstract. Public hospitals receive large volumes of pathology results everyday. It is therefore challenging for doctors to comprehensively analyse all this data. Pathology data prioritisation would seem to provide at least a partial solution. It has been suggested that deep learning techniques can be used to construct pathology data prioritisation models. However, due to the resource required to obtain sufficient prioritisation training and test data, the usage of deep learning, which requires large labelled training data sets, was found not to be viable. The idea presented in this paper is to use a small seed set of labelled data and then to augment this data. The motivation here was that data augmentation had been previously employed successfully to address data scarcity problems. Four data augmentation methods are considered in this paper and used to train deep learning pathology data prioritisation models. Evaluation was conducted using Urea and Electrolytes pathology data.

Keywords: Data prioritisation · Data augmentation · LSTM · Pathology Data.

1 Introduction

In any hospital, many of the decisions on patients care are based on the pathology test results. Doctors will interpret the results and look for abnormalities that may cause disease, such as cancer and other chronic illnesses, or health risks, such as pre-diabetes. As our ability to collect pathology data, driven by scientific and technological advances, becomes increasing more sophisticated, the quantity of pathology data that clinicians are expected to reference presents an increasingly significant problem .

In order to solve the problem, some form of automated pathology result prioritisation is suggested. The hypothesis is that there exist certain pattern in the patients pathology tests records, and that these pattern can be identified and utilised using the tools and techniques of machine learning. However, the challenge of achieving such a prioritisation system by adopting machine learning techniques is frequently 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. This is largely due to the resource required but also the challenge of any rigorous definition of what a priority pathology result looks like. The phrase "I can't define what a priority pathology result is but I know one when I see one" is encountered. The challenge is compounded by the fact that typical pathology data prioritisation scenarios comprise a number of pathology results and that it is also important to take into account the individual patient's pathology history. It is not simply a matter of considering a single result, instead a number of time series data sequences need to be considered together.

Some attempts have been made aimed at address the "no ground truth problem". [10] it was assumed that a high priority pathology result equated to an anomalous result and an anomaly detection approach using supervised learning, directed at static and time series data, was advocated. However, a criticism that can be directed at this approach is that given a large number of priority pathology results these would no longer be considered to be anomalous but as "standard" results, and therefore not be prioritised. In [9] the use of a proxy ground truth was proposed to which established deep learning techniques were applied to generate a three class (high, medium, low) pathology data prioritisation model. The proxy ground truth was based on the known outcomes of previous patients; whether they became emergency patients (high priority), in-patients (medium priority) or an out-patents (low priority). However, the criticism that can always be directed at system that use ground truth proxy data is that it is difficult to know whether the proxy data accurately reflect the "on the ground" situation unless ground truth data is available for comparison (which then obviate the need for a proxy ground truth!).

The view taken in this paper is that, although it is acknowledged that collecting a comprehensive pathology data prioritisation data set is time-consuming and costly, it is possible to collect a small number of examples. A set of examples that, on its own, is insufficient to to generate and validate machine/deep learning model, but which can be used as a seed set from which a usable ground truth can be generated (grown), a process known as *data augmentation* [2][8]. Data augmentation is the process of generating new data points from existing data, and is seen as an effective solution for dealing with the data scarcity problems. This idea is explored in this paper. A seed set was obtained founded on the Urea and Electrolytes (U&E) pathology application domain. Four different data augmentation techniques were applied to this seed set: (i) Jittering, (ii) the Synthetic Minority Oversampling Technique (SMOTE), (iii) the Deviation From Mean (DFM) mechanism and (iv) Guide Warping. Evaluation was conducted by generating and comparing pathology data prioritisation models generated using the Long Short Term Memory (LSTM) recurrent neural network framework to the augmented data.

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 U&E pathology application domain used as a focus for the work presented here. Details of the considered data augmentation techniques and the proposed pathology data prioritisation process are the provided in Section 4. The

conducted evaluation is then reported on 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 general topic of the work presented in this paper broadly falls into the area of AI-based medical data classification [15]. AI-driven techniques for aiding disease detection and diagnosis has become a popular area of research, where the main objective is to help professionals make more informed decisions concerning medical diagnoses, treatments and triage. It is this last which is most relevant with respect to the work presented in this paper. Traditional machine learning techniques, such as ensemble random forest, Support Vector Machines (SVM) and Logistic regression have all been extensively used for medical triage research[15]. For example, in [1] it was proposed to used ensemble random forests to triage patients in emergency departments so as to reduce waiting time [1]. In [14] a triage approach founded on predicting anomalous patients, using a SVM model and Principal Component Analysis (PCA), was proposed. In [11] a machine learning system was presented for prioritising patients with serious (unstable) conditions from patients with stable conditions. In [**?**] a deep learning model was proposed in the context ophthalmology referral (triage). Some reported studies have adopted Natural Language Processing (NLP) techniques, directed at patient records, to prioritise patients [15, 7].

With respect to most existing studies related to medical data prioritisation, one of the most important factors for achieving a triage system utilising machine learning techniques is to have appropriate "ground truth" training data. In the case of the existing work on triage systems for emergency departments, triage levels were usually established in advance by experienced clinicians according to the criticality of patients [11]. Where NLP techniques have been adopted the data used for training were comprised labelled doctors' reviews. However, as already noted, for some areas within the triage domain, such as pathology results prioritisation, ground truth data is usually not available because of the resource required (especially given the current COVID-19 pandemic which has placed extra strain on health services). This is turn means that established techniques, such as those listed above, are difficult to be used directly.

Several approaches have been proposed to address the absence of a "ground truth" problem min the context of pathology data prioritisation. As noted in the introduction to this paper, these include anomaly detection approaches [10] and ground truth proxy approaches [9]. In [10] unsupervised learning techniques were used to generate a cluster configuration for pathology results. Then, if a new patient's pathology result could not be fitted into an existing cluster it was assumed to be anomalous (an outlier) and therefore a priority result. One of the criticisms of anomaly detection-based approaches s that when there are a large number of anomalous pathology results these would be grouped into cluster configuration and therefore any new patient record would not be considered

anomalous. In [9] a data set of the final destination of patients was used to categorise patients as Emergency Department, In Patient or Out Patient patients. A training set was then derived of pathology history time series, each labelled with these final destination. A classifier was then built using this training data which was then used of the patient which was then equated to a prioritisation level, high, medium or low. However, a criticism that can be directed at proxy ground truth-based approaches is that it is challenging to determine how representative of the "real situation" the proxy ground truth data actually is.

Another direction for dealing with the unavailability of a appropriate training data is to generate synthetic data. However, this can only be done given a comprehensive understanding of the domain under consideration which requires significant input from domain experts. This in turn entails resource which, as already noted, in the case of the pathology data prioritisation domain is not available. The approach advocated in this paper is to augment a small seed set; the generation of which does not require significant resource. Data augmentation is the process of supplementing a dataset with similar data that is created from the information held in the seed set [5]. There are two application areas for data augmentation. One is for addressing the data imbalanced problem [4]. For example to use under-sampling techniques to increase the number of samples in minority class (classes). The second is for addressing the small sample size problem [3]. The latter is the focus for the work presented in this paper.

It was noted in the introduction to this paper that when clinicians are required to prioritise current pathology data they need to take into account previous pathology data results. In other words we are talking about time series data. Thus, only data augmentation techniques which are suitable for time series data are considered in this paper. Popular data augmentation methods fall into four categories: Random Transformation, Pattern Mixing, Generative Models and Decomposition [4]. In [4] it was noted that Pattern Mixing methods are appropriate for short time series, which means that such methods are well suited to pathology data. Generative models require large amounts of (training) data, data not available in the case of the pathology data prioritisation application considered here. Decomposition models are intended for time series forecasting applications, whereas classification is under consideration here. Therefore, three alternative pattern mixing methods were considered with respect to the work presented in this paper: (i) Synthetic Minority Oversampling Technique (SMOTE), (ii) Deviation From Mean (DFM) and (iii) Guide Warping. The Jittering random transformation model was also considered. This was selected because of its simplicity and its ability for generalisation in the context of deep learning models [4].

3 Application Domain

The pathology data prioritisation application focus for the work presented in this paper Urea and Electrolytes pathology testing (U&E testing); a commonly used test 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 result, the 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 pathology test; each U&E test result comprises five pathology values. Abnormal levels in any of the tasks may indicate kidney disease.

In more detail, 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:

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

Where T_{so} to T_{bi} are five multi-variate time series and c is the class label taken from a set of classes C. For the work presented in this paper a three-level prioritisation is assumed $\{high, medium, low\}$. The dimensions in each multi-variate time series were: (i) test value, (ii) upper bound and (iii) lower bound. The upper bound and lower bound indicate a "band" in which pathology results are expected to fall. These values are less volatile than the pathology result values, but can change over time.

4 Pathology Data Augmentation and Prioritisation

In this section the proposed approach to pathology data prioritisation, using data augmentation, to expand a given seed set is presented. For the evaluation presented in the following section, Section 5, four data augmentation methods were considered: (i) Jittering, (ii) the Synthetic Minority Oversampling Technique (SMOTE), (iii) the Deviation From Mean (DFM) mechanism and (iv) Guide Warping. Each is therefore described individually in the following four sub-sections, Sub-sections 4.1 to 4.4. In each case the process for applying the method to U&E pathology is included.

This section is concluded, Sub-section 4.5, with a review of the the adopted process for generating pathology data prioritisation models using the LSTM framework.

4.1 Jittering

Jittering is one of the simplest methods from the random transformation category of time series augmentation [13]. The fundamental idea of Jittering is to add random noise to a *query time series*, a time series is selected from the seed set (the source data set). The query time series is then adapted in some way to form an additional time series referred to as the *new time series*. by "jittering" the values in the time series. The idea is that the generated *new time series* (the additional time series generated) will only vary from the query time series by a

factor equivalent to the noise parameters used. The method can be defined as follows:

$$x' = x_1 + \epsilon_1, x_2 + \epsilon_2, ..., x_T + \epsilon_T \tag{2}$$

where ϵ is typically Gaussian noise added to each time stamp value x_T in a query time series. The adopted process of applying Jittering to U&E pathology data was as follows:

1. For each patient time series P_n add Gaussian noise $\epsilon \in \mathcal{N}(0, \sigma^2)$ to each time stamp value within each pathology task. The value 0.03 was used for the value of σ, as suggested in [5]: $\hat{P}_{nJT} = \langle \hat{T}_{So}, \hat{T}_{Po}, \hat{T}_{Ur}, \hat{T}_{Cr}, \hat{T}_{Bi} \rangle$, where $\hat{T}_n = V_{i_1} + \epsilon_1, V_{i_n} + \epsilon_2, ..., V_{i_n} + \epsilon_n$.

4.2 SMOTE

The Synthetic Minority Oversampling Technique (SMOTE) [6] is a pattern mixing data augmentation technique. Unlike Random Transformation techniques, pattern mixing techniques preserve the distribution of the original time series when generating new time series. Though SMOTE was originally designed to deal with the imbalanced classes problem by interpolating patterns for under-represented classes, it can equally be used in the context of the small sample size problem. SMOTE can be defined as follows:

$$x' = x + \lambda \left| x - x_{NN} \right| \tag{3}$$

where sample x is a random sample selected from each class and x_{NN} is a random sample selected from the kth nearest neighbours of x. λ is a random value taken from the range $[0, 1]$. The process of applying SMOTE to the U&E pathology data considered in this paper was as follows:

1. For each query time series P_n, use $k^{st}NN$ to find the reference neighbours T_{NN} for the query time series. Here k is chosen according using the identity $k = \sqrt{n}$, where n represents the number of data items in the seed data set.
2. Randomly select a "reference neighbour" from the k nearest neighbours identified in step 1.
3. Calculate the difference d_n between the current (query) time series and its reference neighbour.
4. Apply d_n to the query time series to give: $\hat{P}_{n_{SMOTE}} = \langle \hat{T}_{So}, \hat{T}_{Po}, \hat{T}_{Ur}, \hat{T}_{Cr}, \hat{T}_{Bi} \rangle$, where $\hat{T}_n = V_{i_1} + \lambda \left| d_1 \right|, V_{i_n} + \lambda \left| d_2 \right|, ..., V_{i_n} + \lambda \left| d_n \right|$.

4.3 DFM

Deviation From Mean (DFM) [5], as the name implies, generates new time series from "mean curves" identified in the seed data. To be more specific, first a Savitzky-Golay filter [12] is used to smooth all of the seed samples of the class of interest. Then the bounding curves with max values and min values of the

smoothed time series data are calculated. A mean curve is then computed from the bounding curves. The DFM of each time series is derived from computing the difference between each time series and the mean curve of the corresponding class. Random segments of DFMs are then combined to create a surrogate DFM curve. Finally, new time series are generated from the multiplication of the surrogate DFM and the class mean curve. The process of applying DFM to U&E pathology data was then:

1. For each time series P_n, use oversampling to ensure that all the time series are of the same length (max length). Use the Savitzky-Golay filter to smooth the time series.
2. Compute the maximum (max) and minimum (min) of all of the smoothed time series, for the class under consideration, so as to generate bounding curves. Then compute the mean curve; the mean of the bounding curves for all time series.
3. Calculate the DFM curve for all of the smoothed time series.
4. Generate a surrogate DFM curve by combining randomly selected sections from the DFM curves from Step 3, and apply a linear gradient to the selected section to scale the start and end points from the different sections.
5. Compute a new time series from the surrogate DFM and the mean curve.

4.4 Guided Warping

Guided Warping [5] generates new time series by "time warping" a query time series by a "teacher pattern" using Dynamic Time Warping (DTW)[16]. DTW is a popular time series similarity measurement technique which takes into account that the time series to be compared may be offset from one another by using a dynamic programming technique. Guided warping warps the elements of a query time series to the elements of the teacher pattern selected by a DTW alignment function. In this manner the aligned elements of the query time series can be replaced by the corresponding teacher time series, thus a new reference time series can be generated. The process of applying Guided Warping to the U&E data was as follows:

1. Given a query time series P_n, and randomly selected teacher patters P_i for the same class, calculate the minimum DTW warping distance $D(I, J)$, where I and J are the lengths of the two time series respectively.
2. Generate a new reference time series by warping the query time series time series with the aligned elements from $D(I, J)$ so that the new reference time series will features values from the query time series but at different time stamps.

4.5 LSTM for Prioritisation

The architecture for the proposed pathology data prioritisation model comprised five LSTMs one for each task. The results from the LSTMs are combined in a final

"decision layer". A similar structure was proposed in [9]. The whole structure is shown in Figure **??**. Each LSTM comprised four layers: (i) an input layer, (ii) two hidden layers and (iii) a Softmax layer. The input was five set of time series, one for each component task, $T_{j_{So}}$, $T_{j_{Po}}$, $T_{j_{Ur}}$, $T_{j_{Cr}}$ and $T_{j_{Bi}}$, made up of the augmented training data. Thus for each task, the input was a multi-variate time series $T_i = \{V_1, V_2, ..., V_m\}$ as described in Section 3. Each time series T_i was padded to the maximum length in T_n. The last layer of the architecture is the decision layer where the final labels for patients are predicted. After obtaining all of the five outputs and predicted labels from each of the five LSTM models, a decision logic module was used to decide the final prioritisation level of the patient. The logic rule is: "*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*". In order to maximise the effectiveness of a LSTM, hyper-parameter tuning is an important element. There are 4 parameters to be considered: (i) learning rate, (ii) batch size, (iii) number of hidden units (nodes per hidden layer) and (iv) number of epochs. Generally, the parameters can be tuned by observing the loss plots of the training and validation data, where cross-entropy can be used as the loss function. Section 5 will provide more details. For the optimization of LSTM, Adam optimization was chosen due to its efficiency and the nature of the adaptive learning rate.

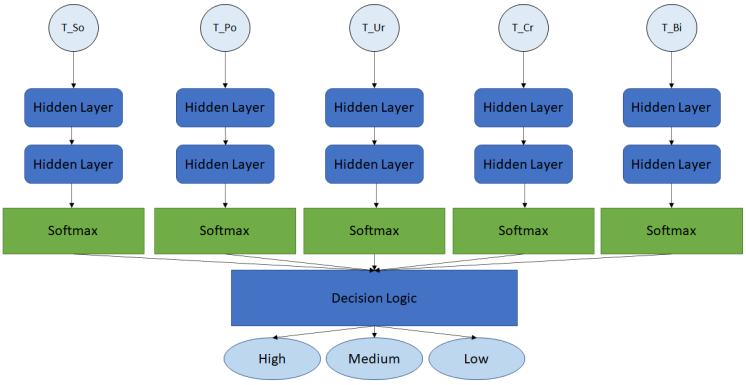


Fig. 1. LSTM architecture

5 Evaluation

The performance evaluation of the proposed approach to pathology data prioritisation is presented in this section. The evaluation data set, as noted earlier, was provided by Arrowe Park Hospital in Merseyside in the UK. 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. Five cross validation was used through-out, and 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 to identify the most appropriate data augmentation strategy for addressing the problem of data scarcity in the context of the pathology data prioritisation problem.

5.1 Evaluation Data Set

A general format of the data was presented in Section 3. The evaluation data set D comprised 3,734 patient records, each with five U&E task results (time series) per patient (no labels available). The seed data set, to which the data augmentation models were to be applied, were extracted from the original data set. Doctors supporting the work were then asked to prioritise the data according to a three-level prioritisation: high, medium, low. The resulting labelled seed set, D_{source}, comprised 30 patients, 6 labelled as high priority, 7 as medium priority and 17 as low priority. Augmentation was then applied to this data. However, it was found that in some cases only a limited number of additional (new) time series could be generated; insufficient to support the generating of a data prioritisation model. This was the found to be the case with respect to SMOTE and Guided Warping because of the way in which these augmentation techniques operate. To address this, proxy data of the form used in [9] was added. Recall that for the proxy ground truth labelling, the final destinations of the patients within the U&E data set was 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 mentioned earlier.

5.2 Best LSTM Settings

The performance of LSTM models depends highly on parameter setting. Analysing the learning curve of the training and validation data is one the most popular methods to find the best parameters for a given training data set. One of the metrics for diagnosing models is the training and validation loss over time. Loss indicates prediction error. A low loss value thus indicates that more learning is required. A value of 0 indicates a model that is perfectly matched the training data. Figure 2 shows an example set of a loss plots derived from training five LSTMs, one per task, on data augmented using SMOTE. For each graph, the x-axis gives the number of times the training samples were "viewed" during LSTM generation, and the y-axis the loss value. From the figures, it can be seen that:

1. Oscillations appear in all of the loss plots.
2. For Loss plots (c), (d) and (e), loss of training and validation decreases to a point of stability and the gap between them are relatively small, which means a good performance of the model.
3. In the case of Loss plot (d) where the validation data set does not contain sufficient information for evaluating the ability of the model. The possible

cause of this problem is that the patterns generated for the *Sodium* results are for the LSTM models to generalise.
4. In the case of Loss plot (a) overfitting has occurred.

Thus, from the foregoing it can be concluded that the effectiveness of SMOTE data augmentation on different task is not consistent. Similar observations were identified with respect to loss plots (not shown here) generated using the other three augmentation techniques considered. Whatever the case the best parameter settings for each data augmentation techniques is given in Table 1.

Table 1. Parameter setting for LSTMs

Parameter Name	**Data Augmentation methods**			
	JT	***SMOTE***	***DFM***	***GW***
Batch size	64	128	256	128
Learning Rate	0.001	0.001	0.001	0.001
Epochs	500	500	500	500
Hidden State	64	64	64	64

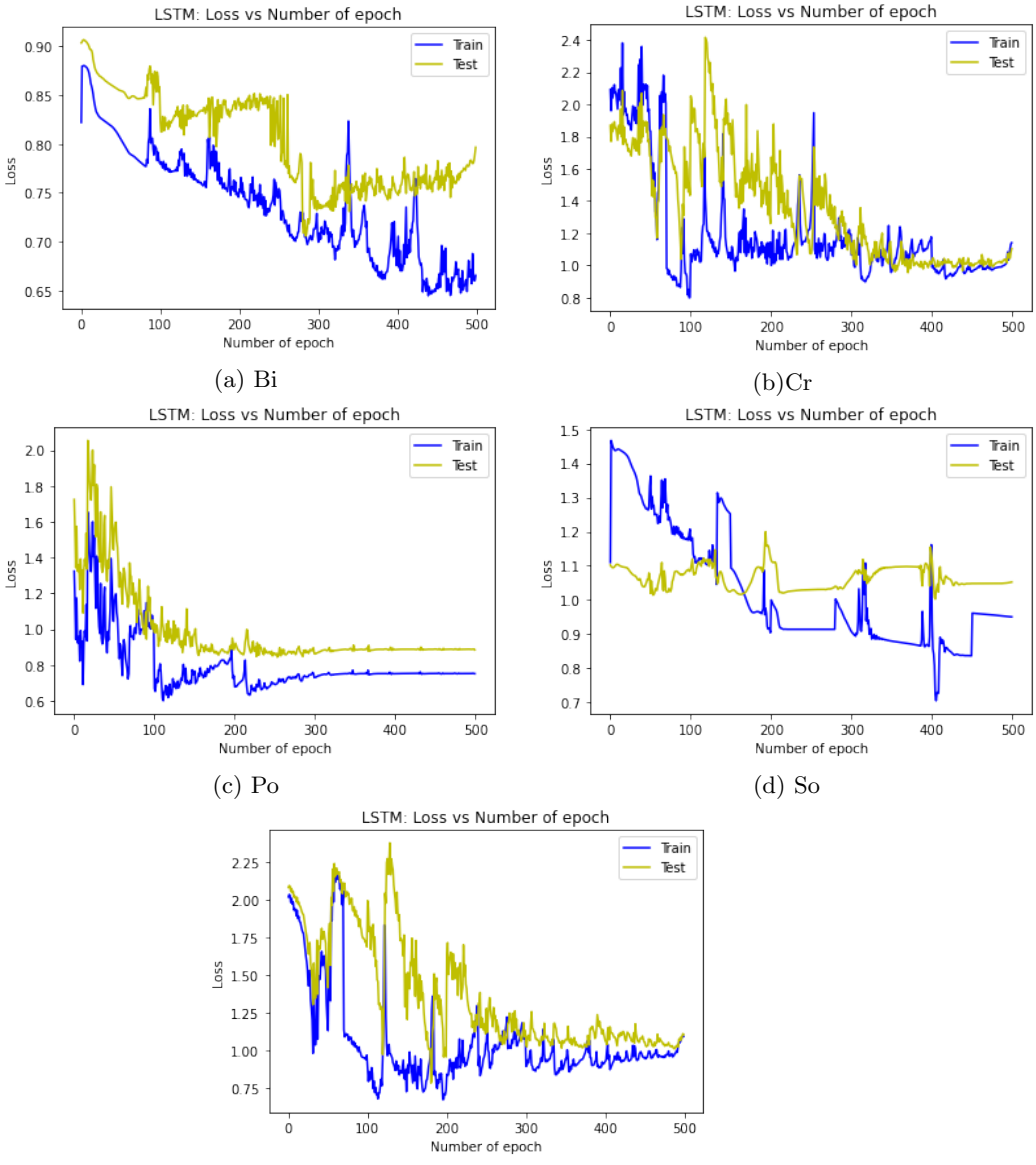
5.3 Comparison of the Overall Performance of Prioritisation

As noted earlier, four data augmentation techniques were used to generate additional training data. Consequently four training data sets were generated: D_{JT}, D_{SMOTE}, D_{DFM} and D_{GW}. Table 2 show the LSTM performance using the different augmented data set. From the table, it can been that:

1. The performance of the pattern mixing techniques (SMOTE, DFM and Guide Warping) was overall better than the Random transformation technique (Jittering) considered; although both SMOTE and Guided Warping required the inclusion of further proxy data. This, it is argued, is because of the high level of randomness in the random transformation technique.
2. The recall and precision of the medium level is lower than the other two classes. This is probably caused by the ambiguous features between the medium class and the other two classes.

6 Conclusions

The work presented in this paper has sought to address the pathology data prioritisation in the absence of a ground truth problem. The reason for the absence of a ground truth was the resource required, on the behalf of clinicians, to generate such data. The central idea promoted is to use a seed set a small seeds set of labelled examples, through which resource was available, and then



(a) Bi

(b)Cr

(c) Po

(d) So

(e) Ur

Fig. 2. Loss curves for five LSTM models based on SMOTE

Table 2. Comparison Precision and Recall of four data augmentation methods

Methods.	Acc.	Pre. High	Pre. Medium	Pre. Low	Rec. High	Rec. Medium	Rec. Low
JT	0.48	0.40	0.37	0.32	0.57	0.42	0.46
SMOTE	**0.66**	**0.55**	0.45	**0.71**	**0.72**	0.46	**0.73**
DFM	0.64	0.51	**0.53**	0.69	0.56	**0.48**	0.63
GW	0.58	**0.55**	0.41	0.51	0.64	0.42	0.60
Ave	0.59	0.51	0.45	0.56	0.62	0.45	0.61

to grow this using data augmentation techniques. Four different augmentation techniques were considered: (i) Jittering, (ii) SMOTE, (iii) DFM and (iv) Guide Warping. The proposed approaches was evaluated using U&E pathology test data which comprised five tasks. The final comparative results demonstrated the best performance was obtained using SMOTE, which provided a best recall and precision of 0.73 and 0.71 respectively. The work has demonstrated that the idea of generating artificial pathology data through data augmentation is a feasible option. The proposed approach is dependent on the quality of the seed set. For future work consideration will be given to the nature of the required seed set. Future work will also entail a comprehensive collaborate with clinicians to obtain feedback regarding the prioritisation produced and to testing of the utility of the best performing mechanism in a real setting will be considered. 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. Dhifaf Azeez, KB Gan, MA Ali, and MS Ismail. Secondary triage classification using an ensemble random forest technique. *Technology and Health Care*, 23(4):419–428, 2015.
2. Markus Bayer, Marc-André Kaufhold, and Christian Reuter. A survey on data augmentation for text classification. *ACM Computing Surveys*, 2021.
3. Edmund Do, Jack Boynton, Byung Suk Lee, and Daniel Lustgarten. Data augmentation for 12-lead ecg beat classification. *SN Computer Science*, 3(1):1–17, 2022.
4. Brian Kenji Iwana and Seiichi Uchida. An empirical survey of data augmentation for time series classification with neural networks. *Plos one*, 16(7):e0254841, 2021.
5. Brian Kenji Iwana and Seiichi Uchida. Time series data augmentation for neural networks by time warping with a discriminative teacher. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 3558–3565. IEEE, 2021.
6. Xubin Jiao and Jinguo Li. An effective intrusion detection model for class-imbalanced learning based on smote and attention mechanism. In *2021 18th International Conference on Privacy, Security and Trust (PST)*, pages 1–6. IEEE, 2021.
7. Norawit Kijpaisalratana, Daecha Sanglertsinlapachai, Siwapol Techaratsami, Khrongwong Musikatavorn, and Jutamas Saoraya. Machine learning algorithms

for early sepsis detection in the emergency department: A retrospective study. *International Journal of Medical Informatics*, 160:104689, 2022.

8. Renkun Ni, Micah Goldblum, Amr Sharaf, Kezhi Kong, and Tom Goldstein. Data augmentation for meta-learning. In *International Conference on Machine Learning*, pages 8152–8161. PMLR, 2021.
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. Jing 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. Zulfany Erlisa Rasjid, Reina Setiawan, and Andy Effendi. A comparison: prediction of death and infected covid-19 cases in indonesia using time series smoothing and lstm neural network. *Procedia computer science*, 179:982–988, 2021.
13. Edgar Talavera, Guillermo Iglesias, Ángel González-Prieto, Alberto Mozo, and Sandra Gómez-Canaval. Data augmentation techniques in time series domain: A survey and taxonomy. *arXiv preprint arXiv:2206.13508*, 2022.
14. Shen-Tsu Wang. Construct an optimal triage prediction model: A case study of the emergency department of a teaching hospital in taiwan. *Journal of medical systems*, 37(5):1–11, 2013.
15. Chee Keong Wee, Xujuan Zhou, Ruiliang Sun, Raj Gururajan, Xiaohui Tao, Yuefeng Li, and Nathan Wee. Triaging medical referrals based on clinical prioritisation criteria using machine learning techniques. *International Journal of Environmental Research and Public Health*, 19(12):7384, 2022.
16. Xinyu Yang, Zhenguo Zhang, Xu Cui, and Rongyi Cui. A time series data augmentation method based on dynamic time warping. In *2021 International Conference on Computer Communication and Artificial Intelligence (CCAI)*, pages 116–120. IEEE, 2021.