Proceedings Paper:
Wójcik, Z., Dimitrova, V. orcid.org/0000-0002-7001-0891, Warrington, L. orcid.org/0000-0002-8389-6134 et al. (3 more authors) (2025) Time-Series Patient-Reported Data and LSTM Predicting Hospital Utilisation During Chemotherapy. In: 2025 IEEE 13th International Conference on Healthcare Informatics (ICHI). 2025 IEEE 13th International Conference on Healthcare Informatics (ICHI), 18-21 Jun 2025, Rende, Italy. Institute of Electrical and Electronics Engineers (IEEE), pp. 228-234. ISSN: 2575-2634 EISSN: UNSPECIFIED

https://doi.org/10.1109/ichi64645.2025.00034

Time-Series Patient-Reported Data and LSTM Predicting Hospital Utilisation During Chemotherapy

Zuzanna Wójcik
School of Computer Science
University of Leeds
Leeds, United Kingdom
ORCID: 0009-0007-6214-7736
sczw@leeds.ac.uk

Vania Dimitrova
School of Computer Science
University of Leeds
Leeds, United Kingdom
ORCID: 0000-0002-7001-0891
v.g.dimitrova@leeds.ac.uk

Lorraine Warrington
Leeds Institute of Medical Research
University of Leeds
Leeds, United Kingdom
ORCID: 0000-0002-8389-6134
l.warrington@leeds.ac.uk

Galina Velikova
Leeds Institute of Medical Research
University of Leeds
Leeds, United Kingdom
ORCID: 0000-0003-1899-5942
g.velikova@leeds.ac.uk

Kate Absolom
Leeds Institute of Medical Research
University of Leeds
Leeds, United Kingdomy
ORCID: 0000-0002-5477-6643
k.l.absolom@leeds.ac.uk

Samuel D Relton
Leeds Institute of Health Sciences
University of Leeds
Leeds, United Kingdom
ORCID: 0000-0003-0634-4587
s.d.relton@leeds.ac.uk

***Abstract*—Adverse effects of chemotherapy often require acute hospital admissions, which can negatively impact patients' well-being and increase healthcare burden. Identifying the risk of hospital utilisation could support prevention of patients' deterioration and alert medical teams about potential admissions. This study uses patients' clinical and demographic data, patient-reported outcome measures, and time-series symptom severity reports to predict the risk of hospital admissions and triage events within 14 days from completed symptom severity report. Hospital utilisation at any time during chemotherapy was also predicted. The performance of long-short term memory (LSTM) and extreme gradient boosting (XGBoost) models was compared. Nested cross-validation enabled robust hyperparameter tuning and model evaluation using unseen data. Patient representatives and a clinical oncologist were consulted during the study design to support its clinical relevance. LSTM outperformed XGBoost at short-term predictions of hospital admission (balanced accuracy=0.780, AUC=0.845) and triage (balanced accuracy=0.706, AUC=0.779). However, XGBoost performed better at long-term predictions. The results suggest that LSTM processed complex data with sudden fluctuations better than XGBoost. However, classical ML might be sufficient for longer-term outcome predictions. If further explored, these models could prompt hospital contact prior to patient's deterioration and prevent admission or alert medical team of potential hospitalisation.**

***Index Terms*—patient-reported data, time-series data, LSTM, hospital utilisation predictions**

I. Introduction

Chemotherapy toxicity is associated with physical, psychological, and psychosocial symptoms negatively impacting patients' quality of life (QoL). Adverse effects of chemotherapy can lead to initial or prolonged hospitalisation, permanent disability, life-threatening situations, or death [1]. Unplanned healthcare utilisation during chemotherapy can be detrimental to healthcare quality, through increased burden on hospitals. Identifying symptoms that contribute to acute presentations to hospitals of cancer patients can support planning for emergency admissions, increase the quality of care, and reduce healthcare costs [2]. Predictions of acute admissions may help to prompt hospital contact, activate emergency alerts to the medical team, and prevent patient deterioration [3].

Understanding symptom experiences and severity is possible due to the use of patient-reported data collected throughout chemotherapy. Questionnaires, such as symptom severity reports or patient-reported outcome measures (PROMs) can capture patients' perspective on their own health status [3]. These have been useful in clinical practice through increasing patient involvement in shared treatment decision making. There is also existing evidence that patient-reported data can add predictive value to machine learning (ML) models [4], [5]. However, there is limited research on deep learning (DL) approaches applied on such data [6].

To explore the potential of DL models applied to time-series patient-reported data, this study investigates the performance of long-short term memory (LSTM) model, a type of recurrent neural network (RNN), in the short-term predictions of acute hospital admissions and triage events occurring within 14 days from completing a symptom severity report. Predictions of admissions and clinical triage events following a report at any time during chemotherapy are also explored. All models use time-series symptom severity reports and PROMs, as well as static data, such as clinical and demographic information. LSTM performances are compared with extreme gradient boosting (XGBoost) [7], which acts as a baseline in this study.

A. Related work

Hospital utilisation is a common outcome in medical prediction models [8], suggesting that AI has a potential to support hospitals in improving healthcare quality. Hospital utilisation predictions can involve detecting changes in data happening over a long period of time, or sudden events. According to a literature review conducted by Teo et al., [8] classical statistical and ML models can perform well in predicting outcomes not limited in time (e.g., cardiovascular disease readmission). However, the LSTM model was superior in predicting outcomes that required capturing sudden changes in patterns of data (e.g., intensive care unit readmissions).

Studies predicting hospital utilisation often use statistical models, or classical ML methods [9], [10] with limited capability to capture complex dependencies and patterns in multidimentional or time-series data [11]. There are existing studies exploring DL models for predictions of oncology outcomes, but the main focus has been on toxicity symptoms [12]–[14], rather than hospital utilisation. These studies have used LSTM model, which is the most common approach to process time-series data in healthcare [11]. Of these studies, only Wang et al., in a paper predicting head and neck cancer symptoms [12], and its follow-up study [13] used patient-reported data. Both studies described a common challenge of using longitudinally collected patient-reported data, which is the irregularity in patients' reports. Therefore, they explored missing data imputation techniques. Nevertheless, these techniques might introduce bias and prevent the model from learning from the patterns of missingness, which could hold information on patients' health status.

Furthermore, Wang et al., [13] have only used patient-reported data without the inclusion of clinical or demographic information. Combining time-series data with static data (e.g., demographic) adds a challenge in DL model development. A simple approach is a repetition of the same static information for each time-point generated by the same patient [11]. However, there are possibilities to use the DL network architecture to handle static data inclusion more efficiently. Time-series and static data can be processed by different network layers as separate inputs. This is the most common approach in medical outcome predictions using static and temporal data [11]. For example, Li et al., [15] used a multi-modal fusion approach that uses the representation of static data as the initial hidden state of RNN. Another study put static input separately through a fully connected layer, and time-series input though LSTM and convolutional neural network layers, combining both outputs at the end [16].

Cancer prediction and diagnosis models rarely include patient-reported data [6], missing the information on patients' perspective on their health. Even though such studies predict clinical outcomes, there is also no evidence of patient or clinician engagement during the study design. Since the development of AI models is not based on patients' perspective or clinical knowledge, it might not actually serve its purpose or address needs of diverse group of people in an equal and fair way. Involving patients and public input is essential not only for the feasibility of novel models, but also for increasing public trust in AI research [17].

B. Contributions

To our knowledge, there is no existing research predicting chemotherapy toxicity related hospital utilisation using DL models and time-series patient-reported data combined with static clinical and demographic information. The proposed approach uses LSTM models to predict admissions and triage events within 14 days from the completion of a symptom severity report. A successful system that can alert to potential complications during chemotherapy can be extremely useful to prevent serious complications or hospitalisation [3]. Therefore, the main contribution of this study is the short-term prediction of hospital utilisation based on longitudinal patient-reported symptom severity reports using LSTM models. Admissions and triage events happening at any time of the chemotherapy were also predicted. In this study the information from missing reports was used, and not imputed, providing the model with an opportunity to learn from the irregularities in reporting. This approach was decided following the consultations with patient representatives and a clinical oncologist during the study design. They suggested that the irregularity of reporting might provide the information of patients' health status (e.g., feeling too fatigued to report).

II. Methods

A. Overall methodology

This study explores the performance of LSTM and XGBoost using clinical, demographic, and longitudinally collected patient-reported data predicting admissions and triage happening within 14 days from a completed symptom severity report (short-term predictions) and at any time during chemotherapy (long-term predictions). Figure 1 illustrates the overall methodology of this research. LSTM is the most common model to process time-series data in healthcare [11], so was selected as a DL model in this study. XGBoost is a baseline model, as it is often compared to DL models when predicting outcomes from numerical, tabular data [7]. XGBoost also performed very well in previous studies predicting hospital utilisation, outperforming other ML models [18]. Patient representatives were consulted during the study design to provide the patients' perspective on the subject. The decisions regarding outcomes and feature engineering were influenced by these consultations. Furthermore, a clinical oncologist was involved, which improved the robustness of the design and its relevance to clinical practice.

B. Dataset

The data were collected from 256 patients initiating systemic treatment for colorectal, breast, or gynecologic cancers in an eRAPID clinical trial [3] at Leeds Cancer Centre (United Kingdom) between January 22, 2015, and June 11, 2018.

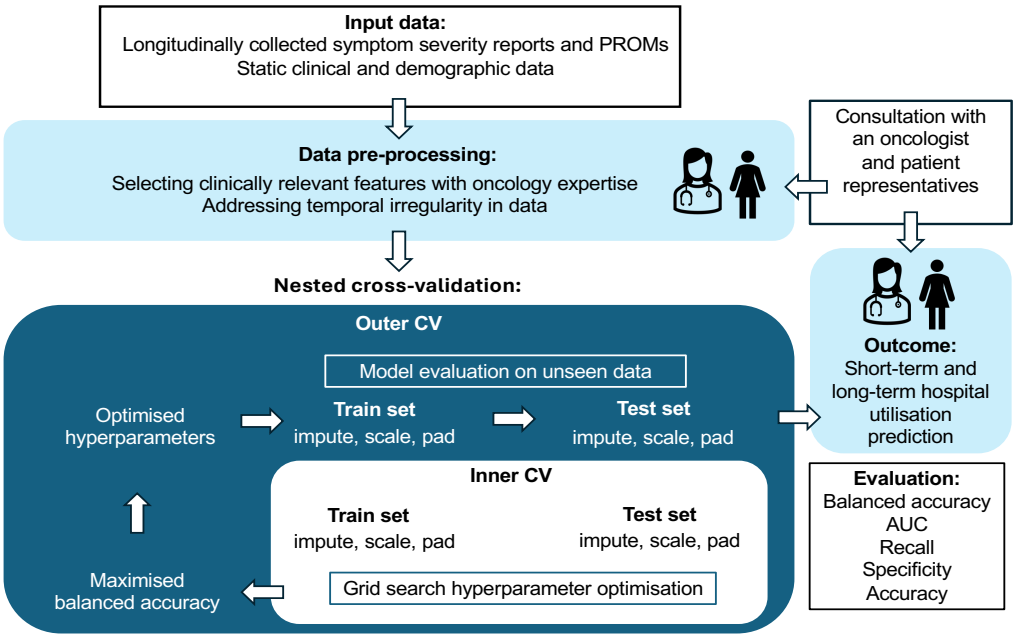

Fig. 1. Flow diagram illustrating data flow and the nested cross-validation (CV) pipeline for implementing LSTM and XGBoost models in this study.

The static data are clinical and demographic information derived from electronic healthcare records. Time-series features include symptom severity reports and PROMs. The patients were asked to complete symptom severity reports weekly for 18 weeks. There were 3260 reports in total. Each included multiple symptoms (pain, nausea, vomiting, diarrhoea, temperature, chills, physical ability, appetite, fatigue, sore mouth, indigestion, shortness of breath) and a scale for assessing severity from 0 (no symptom) to 3 (severe). PROMs were completed by participants at 4 time-points (baseline at the start of the trial, and at 6-, 12-, and 18-week follow-up). For each time-point, 6 variables were from Five-dimensional Visual Analogue Scale (EQ-5D-VAS), including mobility, self-care, usual activities, pain/discomfort, anxiety/depression, and self-rated health status; 4 variables from Functional Assessment of Cancer Therapy - General 28 items (FACT-G), including aggregated scores of physical, social, emotional and functional well-being; and 2 remaining variables were social and role scale from EORTC Core QoL Questionnaire (QLQ-C30). See eRAPID clinical trial report [3] for details on questionnaires.

C. Outcome variables

The outcomes of this study are the short-term and long-term predictions of hospital utilisation:

Short-term predictions of hospital utilisation:

- Admission within 14 days (whether a patient was admitted to hospital within 14 days from completing the report)
- Clinical triage within 14 days (whether a patient contacted hospital within 14 days from completing the report)

Long-term predictions of hospital utilisation:

- Admission (whether a patient was admitted to hospital at any time during chemotherapy, following the report)
- Clinical triage (whether a patient contacted hospital at any time during chemotherapy, following the report)

The 14-day window of admission was chosen by a clinical oncologist as a useful time frame for detecting symptom deterioration in clinical practice, which would enable preparing for or avoiding potential admission. These outcomes were computed from the dates of admissions and triage episodes during the clinical trial. A binary column was created to indicate if the event happened (1) or not (0), following the completion of the symptom report. The long-term predictions don't limit the time frame to any specific amount of days, so the patterns in data can be less complex to capture [8].

D. Data pre-processing

1) Feature engineering: Since symptom severity reports and PROMs were collected at different time-steps, a PROMs value was allocated to each report depending on the time-step of completion (for example, for each report completed between baseline and 6 weeks, the PROMs collected at baseline were

allocated; for reports between 6 and 12 weeks, the PROMs collected at 6 weeks were allocated). The static data were kept separately during the data pre-processing stage for LSTM. However, for XGBoost, to each report, repeated values of patient's static data were added. From all available features, the selection of clinically relevant ones was performed by a clinical oncologist.

2) Addressing irregularity in data: LSTM model requires all data sequences to have the same length for each subject. However, the number of reports submitted by each patient varied due to missed or frequent reporting. The challenge of irregularity can be addressed in 2 different ways: 1) by aggregating the reports that were completed more than once a week (e.g., taking average or the lowest score out of all reports completed during that week) and imputing the missing reports [12], or 2) by leaving the irregularity in data, padding to longest sequence, and including a variable which indicates the time interval between the current and previous report. Due to a significant variability in frequency of reporting, option 1) would create too much synthetic data and would prevent the model from learning on real data. Therefore, a variable was created to indicate the number of days since the previous report. It provides the opportunity for the model to learn from the temporal pattern of reporting. Missing reports were not imputed, but k-nearest neighbors (KNN, k=5) was used to impute any sporadic missing data. There are very few cases of missingness in this dataset. Three patients were outliers, and therefore their data were removed. The longest sequence (43 reports) was then selected and all sequences were padded (filled with a particular value) to match its length within the nested cross-validation loops. This is a common approach to adjust the size of the data in DL model development [19].

E. Model implementation pipeline

1) LSTM model architecture: The model accepts 2 separate inputs: static data (demographic and clinical information) and time-series data (symptom severity reports and PROMs). The LSTM layer processes time-series data, and outputs the hidden representation of time-series sequences. The fully connected layer receives the concatenated input of static data and the LSTM output (processed temporal dependencies). The output is converted by sigmoid activation to a probability for binary classification. Therefore, final output is the binary prediction of event not happening (0) or happening (1). The model architecture is presented in Figure 2. The hidden size, number of layers, learning rate were established through hyperparameter tuning. Due to large class imbalance in data (out of 3260 reports, there were: 640 admissions, 175 admission within 14 days, 1212 triage events, 454 triage events within 14 days), class weight factor was set to 1.

2) Nested cross-validation: To ensure robust hyperparameter optimisation, prevent data leakage and over-estimation of results, the models were implemented through the loops of nested cross-validation [20] (Figure 1). The hyperparameter

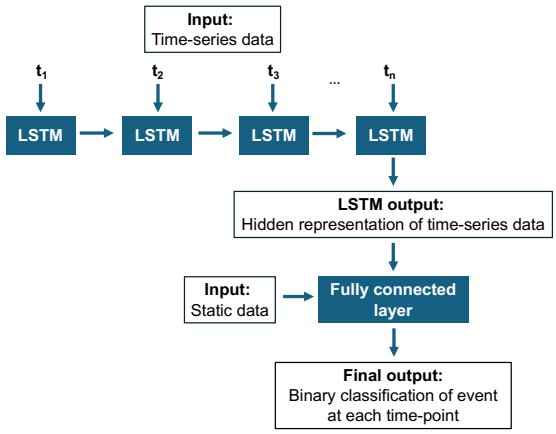

Fig. 2. Flow diagram illustrating the LSTM model. Time-series input is processed through LSTM layer, which output is then passed through fully connected layer, receiving separate static input. The model hyperparameters were established during nested cross-validation process.

tuning was performed in the inner cross-validation loop and in outer cross-validation loop model was evaluated on unseen test set. In each loop a stratified split of patients to train and test sets was completed. The continuous variables were scaled to unit variance, and missing values were imputed with KNN(k=5). The sequences were padded with -999 in order to avoid confusion with zeros existing in the dataset. To further avoid model learning from padded data, masking was applied on padded values. All these steps were performed within loops to avoid data leakage to test sets.

The hyperparameter tuning in the inner loop involved the following hyperparameter sets: hidden size: [64, 128], number of layers: [1, 2], and learning rate: [0.001, 0.0001]. The number of inner folds was 2. Each iteration involved 20 epochs, but early stopping was introduced if the validation loss was not improved in 5 epochs. In the outer loop, the model with hyperparameters selected in the inner loop was trained during 50 epochs and evaluated based on the average of the performance metrics over 3 iterations. The model was trained with "BCEWithLogitsLoss" loss function combining sigmoid activation and binary cross-entropy loss, due to the required binary output (event happening or not happening).

The pipeline for XGBoost was the same, with some adjustments. In each loop, the training set was balanced using Synthetic Minority Oversampling Technique (SMOTE), to help the model learn from the originally imbalanced data. This step was unnecessary in LSTM, as the class weight factor was adjusted. Furthermore, XGBoost does not require as much running time as LSTM, so the nested cross-validation included 5 iterations of both inner and outer loops. In LSTM, to support code efficiency, the number of iterations was limited [7]. Another difference between the 2 models was combining

static and temporal data. XGBoost cannot take 2 separate inputs, so the static data was repeated for each time-step. The hyperparameters grid for XGBoost involved: number of estimators: [50, 100], learning rate: [0.01, 0.05], maximum depth: [3, 4], subsample: [0.7, 0.8], and percentage of features used for building each tree: [0.7, 0.8].

F. Evaluation metrics

The models were evaluated with 5 different performance metrics. Balanced accuracy was selected as the main one, as it is useful at assessing performance of models serving as a screening system, being able to capture model's bias towards one class. Area under the ROC curve (AUC) was also used, as it is a very common metric in ML studies. Even though it is not recommended for imbalanced data, it can be used for between-studies comparison. Recall and specificity are considered as important metrics which explain the model's abilities to discriminate between classes. Plotted together, they can help visualise if the model is struggling with a positive or negative class. Accuracy was also reported due to its common use in research, but is not discussed, as it is not relevant for models tested on imbalanced data.

III. Results

A. Short-term predictions

Table I and Figure 3 present all results for short-term predictions of admission and triage. Overall, LSTM performed better than XGBoost. Balanced accuracy indicated advantage of using LSTM (0.780 for admission within 14 days and 0.706 for triage within 14 days) over XGBoost (0.545 for admission within 14 days and 0.532 for triage within 14 days) for short-term predictions. AUC of 0.845 suggests a very good performance of LSTM for short-term admissions prediction, which is also higher than the AUC of XGBoost (0.724). LSTM for short-term predictions of triage also resulted with higher AUC (0.779) than XGBoost (0.687). Specificity of LSTM predicting admission within 14 days is high (0.846), with lower recall (0.715), suggesting that the model is slightly better at detecting negative class. However, the overall performance is balanced, as all metrics are considerably high. The specificity of XGBoost (0.991), which means that the model is very good at detecting negative class. However, it has an extremely low recall of 0.098, suggesting that the model tends to predict no admissions within 14 days from most reports, generating a lot of false negative results. For triage, XGBoost has a similar bias towards positive class, with recall of 0.085 and specificity of 0.979. LSTM predicted triage within 14 days with much higher recall (0.699), and lower specificity (0.712). Overall, the results of LSTM are much more balanced than of XGBoost, without bias towards any class.

Overall, short-term predictions of admissions had a higher performance than of triage. However, balanced accuracy was only slightly higher for admission within 14 days than triage within 14 days for both LSTM and XGBoost, with the difference lower than 0.1. All other metrics consistently suggest better performance of both models for short-term predictions of admissions than triage. It suggest that predicting admission within 14 days led to fewer false negative and false positive results than predicting triage within 14 days (Figure 3).

TABLE I
LSTM and XGBoost performance of short-term and long-term predictions of hospital utilisation.

Evaluation metric	Model	Admission within 14 days	Triage within 14 days	Admission	Triage
Balanced accuracy	LSTM	0.780	0.706	0.717	0.776
	XGBoost	0.545	0.532	0.813	0.836
AUC	LSTM	0.845	0.779	0.775	0.839
	XGBoost	0.724	0.687	0.939	0.917
Recall	LSTM	0.715	0.699	0.618	0.823
	XGBoost	0.098	0.085	0.654	0.795
Specificity	LSTM	0.846	0.712	0.816	0.729
	XGBoost	0.991	0.979	0.973	0.877
Accuracy	LSTM	0.836	0.709	0.773	0.765
	XGBoost	0.943	0.853	0.914	0.912

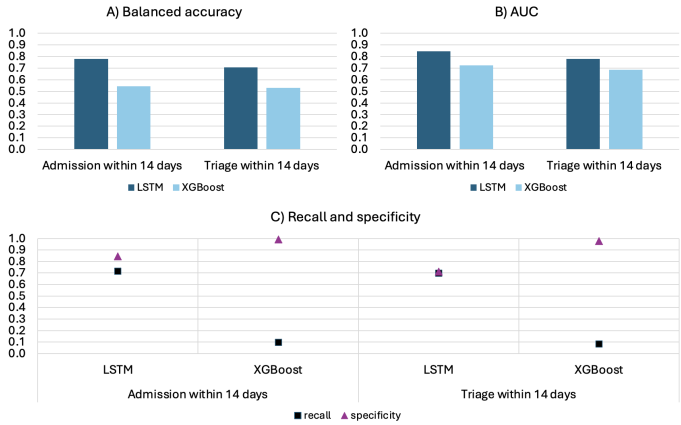

Fig. 3. Plotted balanced accuracy, AUC, recall and specificity of LSTM and XGBoost predicting admission and triage within 14 days.

B. Long-term predictions

Figure 4 and Table I present the results of long-term hospital utilisation predictions. Overall, the long-term predictions were good for both models. However, unlike in short-term predictions, XGBoost performed better than LSTM, achieving higher balanced accuracy, AUC and specificity for both admission and triage predictions. Recall was very similar for both models. XGBoost resulted with higher recall for admissions prediction, and LSTM for triage prediction. The balanced accuracy of 0.8.13 and AUC of 0.939 suggest that XGBoost can successfully predict admission during chemotherapy with a good performance. Similarly, balanced accuracy of 0.836 and AUC of 0.917 suggest a good performance of XGBoost predicting clinical triage. The balanced accuracy for LSTM was 0.717 (admission prediction), and 0.776 (triage prediction) which is only slightly lower than for XGBoost. The recall in predicting admission was moderate for both models (between 0.6 and 0.7), suggesting a bias towards a negative class. Models predicting triage had a higher performance than predicting admissions, based on balanced accuracy. However, the difference was lower than 0.1 for both models.

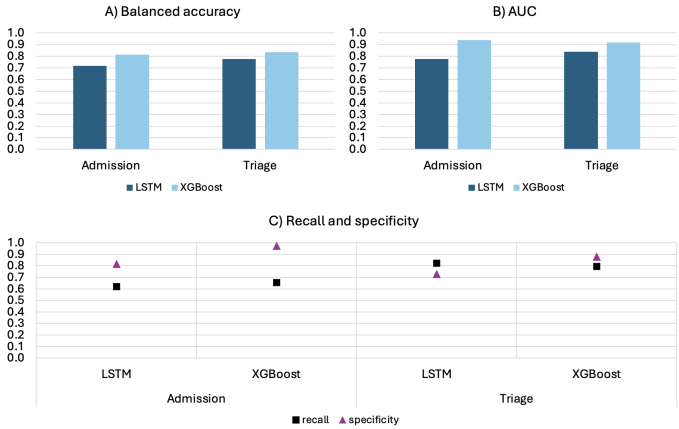

Fig. 4. Plotted balanced accuracy, AUC, recall and specificity of LSTM and XGBoost predicting admission and triage.

IV. Discussion

The results of this study indicate that LSTM model is successful at short-term predictions of hospital utilisation and its performance is superior to XGBoost. However, LSTM performed slightly worse than XGBoost at predicting hospital utilisation happening at any time during chemotherapy.

LSTM performing better than XGBoost in short-term hospital predictions is consistent with literature suggesting that DL models are able to process more complex and sudden fluctuations in the dataset [8]. Therefore, in time-series scenarios where the outcomes happen shortly after the input time-step, deep network architectures, such as LSTM, have more potential to succeed. On the other hand, the performance of long-term predictions of admission and triage was better for XGBoost than LSTM. However, the difference in these performances was not as high as for short-term predictions. There is previous evidence that XGBoost often outperforms DL models when predicting outcomes from numerical, tabular data [7]. The complexity of sudden outcome following report was removed in these scenario, which might have strengthen the performance of XGBoost. Another explanation for the differences in the model performances can be derived from significance testing of the differences between variables for each class of the outcomes. Table II presents the number of variables which are statistically significantly different for classes in each outcome. It indicates that more features differed significantly for classes in admission and triage than for classes in admission and triage within 14 days. These differences might have provided a more structured discriminatory advantage for XGBoost in predicting admission and triage. For short-term predictions more complex dependencies in data were necessary to process, which LSTM succeeded in.

The overall performance of all models is good, which can be compared with other studies. Peterson et al., [10], predicted emergency department visits and hospital admissions during chemotherapy from clinical, demographic and patient-reported data using ML models. The reported AUC for the best model was 0.783, which is lower than the highest AUC of LSTM for short-term admission predictions (0.845), and XGBoost for admission during chemotherapy predictions (0.939). In another study predicting emergency visits and hospital admissions during cancer treatment [9], the AUC of gradient tree boosting was 0.798, which is also lower, compared to the results in this study. Good performance of these models might be the result of robust design with nested cross-validation, as this approach can reduce bias in model selection [21].

TABLE II
Number of features within different significance levels of the difference between classes for each outcome.

	Admission within 14 days	Triage within 14 days	Admission	Triage
$p<0.001$	14	16	20	21
$p<0.01$	6	6	8	8
$p<0.05$	5	6	1	4
$p>0.05$	18	15	14	10

The strengths of this study include the prediction of an outcome within the short time frame of patient-reported symptom severity. If applied in clinical practice, such system could help guide the identification of patient deterioration and support to prevent hospital admission. The involvement of patient representatives and a clinician in a study design provided perspectives of stakeholders which is necessary for any AI system to be included in clinical practice. Furthermore, the model implementation with the nested cross-validation pipeline enabled robust hyperparameter tuning and minimised the risk of data leakage [20]. Additionally, multiple evaluation metrics reported enable in-depth explanation of model performance and between-study comparison.

The interpretation of the study should also consider its limitations. The inherent bias of participant samples recruited to clinical trials could affect the representability of the sample. Therefore, future work should validate the models on an external dataset. The sample size in this study is also relatively small, which could have negatively affected the performance of LSTM, as DL models require more data than classical ML models [6]. Furthermore, even though LSTM and XGBoost were implemented through the same pipeline, the differences included the combination of static and time-series features, and handling class imbalance. The discrepancies could have affected the performance of the models, but were unavoidable due to different model architectures. The number of iterations was also inconsistent between the 2 models, which could have prevented LSTM from learning effectively. In the future, the trade-off between model efficiency and the robustness of the design should be explored. Future work should also involve more structured focus groups with patients and clinicians, including the discussion around results interpretation.

V. Conclusion

This study used time-series patient reported data, as well as static demographic and clinical information to predict hospital utilisation during chemotherapy. LSTM model, by detecting sudden changes in data, successfully predicted admissions and

triage events happening within 14 days from symptom severity reports, outperforming XGBoost. However, admissions and triage predictions at any time during chemotherapy resulted with higher performance of XGBoost. It suggests that simple ML models might be better for less complex predictions. If further explored, this study could support reducing acute hospital utilisation during chemotherapy by detecting risk of deterioration in a timely manner. Furthermore, it encourages patient and clinician involvement in AI research and supports informed shared treatment decision-making.

VI. Acknowledgements

This research is supported in part by UK Research and Innovation [EP/S024336/1]. Ethics approval was granted by School of Medicine Research Ethics Committee [MREC 22-098], University of Leeds. The authors thank patient representatives from Patient Centred Outcomes Research (PCOR) in the Leeds Institute of Medical Research for consultation, and the patients and clinicians involved in the eRAPID clinical trial for providing the data. The authors also thank Allan Pang and Alexander Coles for the help with study design.

References

[1] I. Ingrand, G. Defossez, C. Lafay-Chebassier, F. Chavant, A. Ferru, P. Ingrand, and M.-C. Pérault-Pochat, "Serious adverse effects occurring after chemotherapy: A general cancer registry-based incidence survey," *British Journal of Clinical Pharmacology*, vol. 86, no. 4, pp. 711–722, 2020, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/bcp.14159. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/bcp.14159

[2] P. H. Dufton, M. F. Gerdtz, R. Jarden, and M. Krishnasamy, "Methodological approaches to measuring the incidence of unplanned emergency department presentations by cancer patients receiving systemic anti-cancer therapy: a systematic review," *BMC Medical Research Methodology*, vol. 22, no. 1, p. 75, Dec. 2022. [Online]. Available: https://bmcmedresmethodol.biomedcentral.com/articles/10.1186/s12874-022-01555-3

[3] K. Absolom, L. Warrington, E. Hudson, J. Hewison, C. Morris, P. Holch, R. Carter, A. Gibson, M. Holmes, B. Clayton, Z. Rogers, L. McParland, M. Conner, L. Glidewell, B. Woroncow, B. Dawkins, S. Dickinson, C. Hulme, J. Brown, and G. Velikova, "Phase III Randomized Controlled Trial of eRAPID: eHealth Intervention During Chemotherapy," *Journal of Clinical Oncology*, vol. 39, no. 7, pp. 734–747, Mar. 2021, publisher: Wolters Kluwer.

[4] J.-a. Sim, Y. A. Kim, J. H. Kim, J. M. Lee, M. S. Kim, Y. M. Shim, J. I. Zo, and Y. H. Yun, "The major effects of health-related quality of life on 5-year survival prediction among lung cancer survivors: applications of machine learning," *Scientific Reports*, vol. 10, p. 10693, Jul. 2020. [Online]. Available: https://www.nature.com/articles/s41598-020-67604-3

[5] Z. Wójcik, V. Dimitrova, L. Warrington, G. Velikova, and K. Absolom, "Using Machine Learning to Predict Unplanned Hospital Utilization and Chemotherapy Management From Patient-Reported Outcome Measures," *JCO Clinical Cancer Informatics*, no. 8, p. e2300264, Apr. 2024, publisher: Wolters Kluwer. [Online]. Available: https://ascopubs.org/doi/10.1200/CCI.23.00264

[6] D. Krepper, M. Cesari, N. J. Hubel, P. Zelger, and M. J. Sztankay, "Machine learning models including patient-reported outcome data in oncology: a systematic literature review and analysis of their reporting quality," *Journal of Patient-Reported Outcomes*, vol. 8, no. 1, p. 126, Nov. 2024. [Online]. Available: https://doi.org/10.1186/s41687-024-00808-7

[7] R. Shwartz-Ziv and A. Armon, "Tabular data: Deep learning is not all you need," *Information Fusion*, vol. 81, pp. 84–90, May 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1566253521002360

[8] K. Teo, C. W. Yong, J. H. Chuah, Y. C. Hum, Y. K. Tee, K. Xia, and K. W. Lai, "Current Trends in Readmission Prediction: An Overview of Approaches," *Arabian Journal for Science and Engineering*, vol. 48, no. 8, pp. 11 117–11 134, Aug. 2023. [Online]. Available: https://doi.org/10.1007/s13369-021-06040-5

[9] J. C. Hong, D. Niedzwiecki, M. Palta, and J. D. Tenenbaum, "Predicting Emergency Visits and Hospital Admissions During Radiation and Chemoradiation: An Internally Validated Pretreatment Machine Learning Algorithm," *JCO Clinical Cancer Informatics*, no. 2, pp. 1–11, Dec. 2018.

[10] D. J. Peterson, N. P. Ostberg, D. W. Blayney, J. D. Brooks, and T. Hernandez-Boussard, "Machine Learning Applied to Electronic Health Records: Identification of Chemotherapy Patients at High Risk for Preventable Emergency Department Visits and Hospital Admissions," *JCO Clinical Cancer Informatics*, no. 5, pp. 1106–1126, Dec. 2021. [Online]. Available: https://ascopubs.org/doi/10.1200/CCI.21.00116

[11] M. A. Morid, O. R. L. Sheng, and J. Dunbar, "Time Series Prediction Using Deep Learning Methods in Healthcare," *ACM Transactions on Management Information Systems*, vol. 14, no. 1, pp. 1–29, Mar. 2023. [Online]. Available: https://dl.acm.org/doi/10.1145/3531326

[12] Y. Wang, L. Van Dijk, A. S. R. Mohamed, C. D. Fuller, X. Zhang, G. E. Marai, and G. Canahuate, "Predicting late symptoms of head and neck cancer treatment using LSTM and patient reported outcomes," *Proceedings. International Database Engineering and Applications Symposium*, vol. 2021, pp. 273–279, Jul. 2021.

[13] Y. Wang, L. V. Dijk, A. S. R. Mohamed, M. Naser, C. D. Fuller, X. Zhang, G. E. Marai, and G. Canahuate, "Improving Prediction of Late Symptoms using LSTM and Patient-reported Outcomes for Head and Neck Cancer Patients," in *2023 IEEE 11th International Conference on Healthcare Informatics (ICHI)*, Jun. 2023, pp. 292–300, iSSN: 2575-2634. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/10337216

[14] S. Chae, W. N. Street, N. Ramaraju, and S. Gilbertson-White, "Prediction of Cancer Symptom Trajectory Using Longitudinal Electronic Health Record Data and Long Short-Term Memory Neural Network," *JCO Clinical Cancer Informatics*, no. 8, p. e2300039, Mar. 2024, publisher: Wolters Kluwer. [Online]. Available: https://ascopubs.org/doi/abs/10.1200/CCI.23.00039

[15] D. Li, P. Lyons, J. Klaus, B. Gage, M. Kollef, and C. Lu, "Integrating Static and Time-Series Data in Deep Recurrent Models for Oncology Early Warning Systems," in *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, ser. CIKM '21. New York, NY, USA: Association for Computing Machinery, Oct. 2021, pp. 913–936. [Online]. Available: https://dl.acm.org/doi/10.1145/3459637.3482441

[16] C. Lin, Y. Zhang, J. Ivy, M. Capan, R. Arnold, J. M. Huddleston, and M. Chi, "Early Diagnosis and Prediction of Sepsis Shock by Combining Static and Dynamic Information Using Convolutional-LSTM," in *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, Jun. 2018, pp. 219–228, iSSN: 2575-2634. [Online]. Available: https://ieeexplore.ieee.org/document/8419365

[17] S. Banerjee, P. Alsop, L. Jones, and R. N. Cardinal, "Patient and public involvement to build trust in artificial intelligence: A framework, tools, and case studies," *Patterns*, vol. 3, no. 6, p. 100506, Jun. 2022. [Online]. Available: https://linkinghub.elsevier.com/retrieve/pii/S2666389922000988

[18] S. Han, T. J. Sohn, B. P. Ng, and C. Park, "Predicting unplanned readmission due to cardiovascular disease in hospitalized patients with cancer: a machine learning approach," *Scientific Reports*, vol. 13, no. 1, p. 13491, Aug. 2023, publisher: Nature Publishing Group. [Online]. Available: https://www.nature.com/articles/s41598-023-40552-4

[19] M. A. Islam, M. Kowal, S. Jia, K. G. Derpanis, and N. D. B. Bruce, "Position, Padding and Predictions: A Deeper Look at Position Information in CNNs," *International Journal of Computer Vision*, vol. 132, no. 9, pp. 3889–3910, Sep. 2024. [Online]. Available: https://doi.org/10.1007/s11263-024-02069-9

[20] S. Kapoor and A. Narayanan, "Leakage and the Reproducibility Crisis in ML-based Science," Jul. 2022, arXiv:2207.07048 [cs]. [Online]. Available: http://arxiv.org/abs/2207.07048

[21] S. Varma and R. Simon, "Bias in error estimation when using cross-validation for model selection," *BMC Bioinformatics*, vol. 7, no. 1, p. 91, Feb. 2006. [Online]. Available: https://doi.org/10.1186/1471-2105-7-91