RetainEXT: Enhancing Rare Event Detection and Improving Interpretability of Health Records using Temporal Neural Networks

Suraj Ramchand*, Gavin Tsang*, Duncan Cole†, Xianghua Xie*
*Computer Science, Swansea University, Swansea, United Kingdom
†School of Medicine, Cardiff University, Cardiff, United Kingdom
866648@swansea.ac.uk, Gavin.Tsang@swansea.ac.uk, ColeDS1@cardiff.ac.uk, X.Xie@swansea.ac.uk

***Abstract*—A recurring theme during the pandemic was the shortage of hospital beds. Despite all efforts, the healthcare system still faces 25% of resource strain felt during the first peak of coronavirus. Digitisation of Electronic Healthcare Records (EHRs) and the pandemic have brought about many successful applications of Recurrent Neural Networks (RNNs) to predict patients' current and future states. Despite their strong performance, it remains a challenge for users to delve into the black box which has heavily influenced researchers to utilise more interpretable techniques such as 1D-Convolutional neural networks. Others focus on using more interpretable machine learning techniques but only achieve high performance on a select subset of patients. By collaborating with medical experts and artificial intelligence scientists, our study improves on the REverse Time AttentIoN EX model, a feature and visit level attention network, for increased interpretability and usability of RNNs in predicting COVID-19-related hospitalisations. We achieved 82.40% area under the receiver operating characteristic curve and showcased effective use of the REverse Time AttentIoN EXTension model and EHRs in understanding how individual medical codes contribute to hospitalisation risk prediction. This study provides a guideline for researchers aiming to design interpretable temporal neural networks using the power of RNNs and data mining techniques.**

***Index Terms*—Artificial intelligence, Data mining, Electronic health records, COVID-19, Attention networks**

I. Introduction

Despite Coronavirus Disease 2019 (COVID-19) prevention and risk mitigation measures, hospital resource utilisation remains a significant concern. Although the hospitalisation rate has reduced from 36.68 to 8.20 per 100,000, United Kingdom (UK)'s healthcare system still faces 25% of the strain felt during the first wave. With the resurgence of COVID-19 cases, patients at high risk of hospitalisation due to the virus must be rapidly identified for early intervention and risk mitigation.

Numerous studies developed Artificial Intelligence (AI) predictive tools to identify patients at risk of severe outcomes due to COVID-19. These tools utilise highly interpretable Machine Learning (ML) algorithms, such as decision trees [1]–[3], gradient boosting decision trees [4]–[6] and logistic regression [7], [8]. These models yielded Area Under the Curve (AUC) scores between 0.74 and 0.92 and discovered critical prognostic markers essential for predicting a patient's COVID-19 outcome, including white cell differential count, creatinine phosphate and lymphocyte proportion.

However, identifying risk factors leading to COVID-19 hospitalisation proved to be difficult for ML models primarily when large populations and multiple comorbidities are involved. Willete *et al.* [9] employed a permutation-based linear discriminant analysis to predict COVID-19 and hospitalisation risk. When trained on a subset of participants with an antibody titter, they achieved an AUC of 0.969 (95% CI 0.934–1.000), but when trained on a more significant portion of the population, the AUC dropped to 0.803 (95% CI 0.663–0.943). Similarly, Wollenstein *et al.* [7] only achieved 61% accuracy predicting COVID-19 hospitalisations.

Deep Learning (DL) techniques, specifically Recurrent Neural Networks (RNNs), are known for their ability to model long temporal sequences in heterogeneous patient data [10]–[12]. Besides being termed as 'black box models' due to the lack of interpretability, there are limitations around how DL models handle dimensionally varying or missing data. These constraints have led to the use of 1D Convolutional Neural Network (1D-CNN)s, which generate a feature importance similar to logistic regression models. 1D-CNNs, however, are unable to model temporal patterns well and often underperform RNNs.

This study aims to employ the capabilities of RNNs in predicting COVID-19 hospitalisations whilst providing interpretability around the model's decisions by unboxing the black box and improving on REverse Time AttentIoN model (Retain), an interpretable RNN [11] and its successor REverse Time AttentIoN EX model (RetainEX) [12]. Retain and RetainEX adopt a temporal attention generation mechanism to learn the importance of each General Practice (GP) visit and each medical code. However, they lack architectural depth, visit-level risk scores and a means of learning from imbalanced datasets. Hence, this study aims to improve the predictability and interpretability of the Retain models and exhibit the use of an enhanced version of the model, REverse Time AttentIoN EXTension model (RetainEXT).

The findings of this study will aid in identifying individuals with a high risk of hospitalisation from the virus, allowing for

This work was supported by Amicus Therapeutics UK Operations and the Engineering and Physical Sciences Research Council (EPSRC) centre for doctoral training in enhancing human interactions and collaborations with data and intelligence-driven systems, grant number EP/S021892/1.

early interventions to mitigate risks of post-infection complications.

Section II includes the study design and population and a description of RetainEXT. Section III evaluates and compares the results and feature importance extrapolated from the best-performing RetainEXT model. Finally, Section IV concludes and elaborates on future work.

II. Methodology

A. Study Design

The study design was a retrospective longitudinal Self-Controlled Case Series (SCCS). Data were obtained from the Secure Anonymised Information Linkage databank [13], which contains Electronic Health Records (EHR) of 80% of the Welsh primary care data. We included demographic, primary and secondary care data between 2009-2020 to follow the patient's from their early interactions with the National Health Service (NHS) up to and including their first COVID-19-related hospitalisation.

We used the following inclusion criteria: 1) A minimum of 2 GP interactions in the data collection period; 2) Aged 18 and above at the start of the data collection period; 3) A hospital admission within 14 days before a positive COVID-19 test, undertaken during the hospitalisation.

To limit confounding factors, we excluded medical codes including COVID-19 infection or hospitalisation collected within the 14 days before a positive test result. The resulting cohort comprised 2,277 female and 2,071 male patients with a mean age of 69.7 years.

B. Data Structure

Similar to the structures in Choi *et al.* [11] we modelled each patient's EHR as Encounter Sequence Modelling (ESM) [10], where the sequence of patient visits is represented by a set of a varying number of medical codes $d^1, ..., d^l$, where l is the number of diagnosis codes per GP visit,x_t and d^j is the j^{th} code from the dictionary of all codes, D. Therefore, the total number of possible codes is $r = |D|$.

Traditionally, ESM models each visit $x_t \in \{0,1\}^{|D|}$ as a binary vector, where the value 1 in the j^{th} coordinate indicates that d^j was documented in the t^{th} visit. Given a sequence of visits $x_1, ..., x_T$, with T as the total number of visits, the goal of ESM is to predict the codes occurring at the following visit $x_2, ..., x_{T+1}$, with the number of labels $y = |D|$.

As $|D|$ contains 27,317 unique read codes, it would be resource heavy and impractical to train a model on large binary vectors. Instead, we decided that a patient's visit, x_t, should only include the medical codes recorded during that visit.

Raw medical codes contain a mixture of alphanumeric characters, therefore, we decided to encode $|D|$ into sequential numerical values with arbitrary meaning. Having defined each patient's visit as $x_t = d_1, ..., d_s$, we set the model to predict the patient risk of hospitalisation, $\hat{y_t}$, at each visit, x_t.

Following each visit, x_t, was passed through an embedding layer to learn the representation and later visualise any clusters of medical codes. This generates v_t for each x_t, which is then concatenated to the patient's age and gender. Both variables were included at every time step due to the patient's varying age and possibly gender.

Furthermore, the model's comprehension of the time between visits is vital to determining the risk of hospitalisation at each visit. For instance, a series of visits to the GP over a short period may indicate comorbidity or severe illness. Long hibernation may suggest good health and influence the model to predict lower risk scores. To harness temporal information, we incorporate visit dates as an additional feature.

Given a sequence of T events $t_1, t_2, ..., t_T$, we obtain $\Delta t_i = t_i - t_{i-1}$ for each successive visit. We assume that the first visit is unaffected by time constraints by fixing Δt_1 to 1. We explored the benefit of additional representations of time, which are (1) Δt_i (time interval between visits) [11], (2) $1/\Delta t_i$ (its reciprocal value) [14], and (3) $1/\log e + \Delta t_i$ (an exponentially decaying value) [14]. These values are concatenated to the embedding,v_t for each x_t, to enrich the information for our model.

While handling the data, we found a critical improvement necessary to improve the epidemiological study design of Retain and RetainEX.

1) Per-event risk assessment: Retain and RetainEX both utilise an Learn to Diagnose (L2D) [15] approach and are limited to predicting a risk score for each patient. This suggests that both case and control groups were used to learn high-risk markers, hindering clinicians from evaluating outcome severity at each GP visit. Thus, our study decided to modify the algorithm's output to model a risk score at each GP visit using an SCCS study design.

The SCCS is a case-only method in which confounders are automatically controlled for [16]. This allows us to investigate the association between a transient exposure and an outcome event which aids clinicians in making better-informed decisions.

C. RetainEXT

Fig. 1 (A) shows our model takes in a patient visit sequence as C dimensional vectors $x^1, x^2, ..., x^T$. An embedding matrix $W_{emb} \in R^{m \times C}$ is used to convert all 27,317 unique medical codes linearly into a matrix of size m × C, where m is the number of units in the embedding layer, resulting in $v_t = W_{emb} \cdot x_t$. The patient's age and gender, n_t, are also appended to v_t at each visit and passed through a dropout layer to improve the model's generalisability. Additionally, each successive visit generates a set of representations that characterise the time between visits, Δt^t, to offer additional insight into the patient's state; this is concatenated to v_t and n_t.

Following the structure of Retain and RetainEX, we computed two attention types, α and β. Fig. 1 (B) and (C) represent the stacked Bi-LSTM network that takes in the age, gender and time-attached visit representations and returns attention values (e.g. contribution scores).

α_t is a single value representing the importance of each GP visit. β_t is an m-dimensional vector that quantifies the

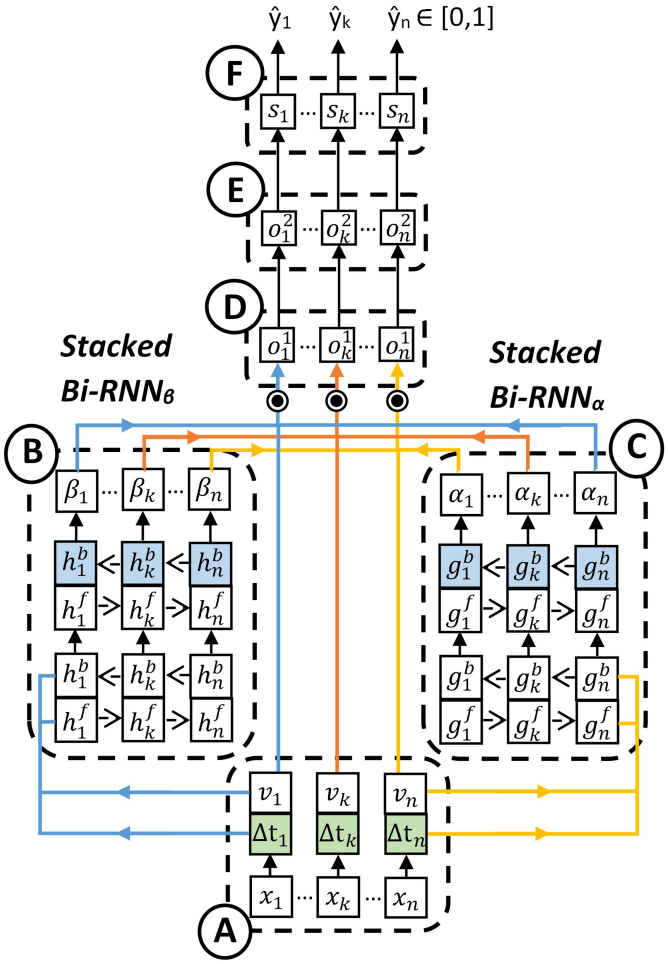

Fig. 1. Overview of RetainEXT. (A) Using a single embedding layer, a binary vector x_t is represented as embedding vectors v_t, with time interval information appended to the former. (B, C) v_t is input into two Bi-LSTM layers to obtain scalar α and vector β attention weights. (D) α, β and v_t are multiplied over all time-steps, and then each time step is passed through a dense layer. (E) Output from the first dense layer is dimensionally reduced into a single output per time step. (F) Each time step output is non-linearly transformed to a risk score $\hat{y}$.

significance of each medical code within a specific visit. To benefit from both visit and feature level importance requires separate stacked Bi-LSTM to compute each class of attention.

For each $[v_t; n_t; \Delta t^t]$, the stacked α-Bi-LSTM computes the forward and backward hidden states of the first α—Bi-LSTM, then passes it on to the second α-Bi-LSTM. The final hidden state vectors, g_t^{f2} and g_t^{b2} are concatenated into a single 2m-dimensional vector, which is passed on to a dense layer.

The parameter $w_\alpha \in R^{2m}$ was used to compute a scalar value for each time step as $e_t = w_\alpha\ [g_t^{f2}; g_t^{b2}]$. Next, the softmax function is applied to all scalar values $\{e^1, ..., e^T\}$ to obtain $\{\alpha_1, \alpha_2, ..., \alpha_T\}$, a distribution of attention values that sum to one. Similarly, the concatenated hidden state vectors generated by the stacked β-Bi-LSTM are multiplied by $w_\beta \in R^{mx2m}$ and return an m-dimensional vector β_t for the t^{th} visit as $\beta_t = w_\beta\ [g_t^{f2}; g_t^{b2}]$.

After obtaining both α_t and β_t values, we performed element-wise multiplication with the concatenated array $[v_t; n_t; \Delta t^t]$, and pass it through a dense layer, with weights w_o^1 (Fig. 1D). The output is then passed to the final dense layer, with weights w_o^2. The additional dense layer increased the complexity of providing interpretability; thus, we defined w_{out}^t as each event's aggregated and combined weight. In mathematical terms, $w_{out}^t = \Sigma_z^U \Sigma_g^T w_{o,g}^1 \cdot w_{o,z}^2$, where U is the number of units in the 2^{nd} dense layer.

Lastly, the contribution score for each visit was computed, $s_t = w_{out}^t \cdot o_t^1$. The t^{th} visit is passed through a dense layer with sigmoid activation to dimensionally reduce the vector at each time step, allowing us to compute a normalised prediction value, $\hat{y}_t$, ranging between 0 and 1 where $w_{out}^t \in R^m$. The predicted value indicates the patient's risk of hospitalisation in that particular visit, with a value closer to 1 indicating a higher risk. We trained our model to minimise binary cross-entropy loss and conducted hyperparameter tuning for a fraction of dropout, regularisation, embedding units and stacked LSTM units. Here, we found improvements needed to limit the input data to only relevant medical codes and improve performance, especially when modelling rare events.

1) Stacked and Deeper Architecture: The original implementations of Retain and RetainEX lacked depth in their architectures. Stacked Bi-LSTM have increased the capacity to identify complex nested patterns in patients that single-layer networks may overlook. Using an additional dense layer before the output layer also facilitates in understanding non-linear patterns in the dataset. However, the added layers increase training time. In Section III we compare the models' performance gain and training times before and after including the additional layers.

2) Ragged Tensors: Another novelty of this study is the adoption of ragged tensors to compute patients with a variable number of visits and medical codes in each visit. This eliminates the need for padding and masking or binary encoding of medical codes, reduces training time, and improves model performance.

3) Rare event weighting: RetainEXT implements the use of sample weighting, which not only attributes a single weight to a patient but also at each visit. Hospitalisation is considered a rare event as only 1.59% GP visits of the entire dataset of medical codes are related to hospitalisation due to COVID-19. Thus, rare event weighting helps the model learn significantly more from a few samples. In contrast, the preceding Retain and RetainEX models struggle to model an imbalanced scenario.

D. Interpretability

The backbone of Retain and RetainEX is their long-standing ability to provide feature and visit level importance scores. With this foundation, we focused on understanding the global feature importance, which is an information-rich measure of how clinically aligned our model is.

1) Local attention: RetainEXT and its predecessors achieve its transparency by multiplying the final layer of the stacked Bi-LSTM generated attention weights α_t and β_t to the visit vectors v_t to obtain the context vector o_t^1, which are used

instead of the Bi-LSTM hidden state vectors to make predictions. Each input vector x_t has a linear relationship with the final contribution score, S. Thus, we derive an equation that measures the contribution score of the code d at time step t to S by reformulating the aforementioned equations as $S_t^d = \alpha_t w_{out}(W_{emb}[d,:] \cdot \beta_t)$, where $W_{emb}[d,:]$ is the d^{th} row of W_{emb}.

Additionally, we generate a visit-level contribution score S_t by aggregating contribution scores of codes for each visit as $S_t = \Sigma_{d \in x_t} S_t^d$.

2) Global feature importance: Retain and RetainEX perform dimensionality reduction on the embedding weight of all clinical markers to visualise possible clusters in the data. However, Kwon *et al.* [12] mentioned that limitations exist in visualising clusters if numerous patients or codes are fed in, limiting our ability to understand critical high-risk medical codes.

ML researchers have used global feature importance [17] to assess the model's ability to mimic a clinician's diagnostic pathways. Leveraging on the linear relationship between the embedding space and the context vector, the global feature importance is defined as $S^d = (\Sigma_t^T \alpha_t) w_{out}(W_{emb}[d,:] \cdot (\Sigma_t^T \beta_t))$.

III. Results and Evaluation

This study demonstrated a high-performing risk prediction model to identify patients susceptible to hospitalisation due to COVID-19 before being hospitalised and suffering from life-long morbidities. While using ten years of historical GP interactions and a large sample of heterogeneous EHR, our interpretable Temporal Neural Network (TNN), RetainEXT, shows statistically significant improvements in traditional model evaluative metrics (AUROC, F1 Score, Sensitivity, Specificity) compared to RetainEX and other state-of-the-art TNN. Of note, without sample weighting, the compared models severely over-fit, hence we trained all models with sample weighting.

A. Model Performance

Insights from previous Retain iterations helped set a hyperparameter space to allow optimisation using a hyperband tuner.

We found that excluding dropout and l2 regularisation acutely improved model performance. Iterations of RetainEXT with 0 and 0.4 dropout resulted in a 0.17 reduction in f1 score, indicating the data may be very heterogeneous and thus weighting of features is sparse, and the model requires multiple features to assess risk. Similar to the original implementation, convergence on local minima occurs quickly, and dropout requires a longer training time to work best, indicating the possibility that the model is stuck at a local minimum and requires longer to diverge out.

In table I, the cascading LSTM [18] has shown to perform well in imbalanced scenarios, albeit, less interpretable. Its sensitivity is leveled with RetainEX and the RetainEXT models, where the only models surpassing it have deeper architectures. The cascading LSTM was trained for 200 epochs compared to 20 epochs for the RetainEXT model, which further emphasises the significance of the attention pathways in the convergence of optimal features.

The 1D-convolutional LSTM benefits from both spatial and sequential learning and can be easily interpreted by extracting the kernel weights. Although, even after training for over 200 epochs, the F1 score is comparatively low, suggesting that successively aggregating medical codes, utilising a fixed kernel size, assume that the patient's condition is limited to that visit and does not take into account the future impact of that medical code.

Further, adding two extra time representations to the time input layer results in a 5.5% increase in sensitivity that is attributed to the model's increased awareness of the interval between GP visits. Including time representations that are normalised between 0 and 1 will generate smaller weight updates on back-propagation, which drives the model closer to the global/local minima.

We also note that RetainEXT outperforms RetainEX in both stacked bi-LSTM configurations, achieving a 0.09 higher F1 score in the optimal configurations. The added LSTM and dense layers allow the RetainEXT model to understand complex non-linear patterns in the patient's EHR. Both models present with similar sensitivity, specificity and AUROC. To determine if the improved performance of RetainEXT is significantly different from other models, we conducted a 5×2 Cross-Fold F-test and calculated an F-score for each evaluation metric. DL models tend to have higher degrees of freedom; hence a critical value of 0.05 was chosen, and the F-score threshold was set at 1.00.

Sensitivity, AUROC and F1-Score all produced an F-value below the threshold of 1.00, which suggests our implementation of RetainEXT is significantly different from RetainEX. Though both Positive Predictive Value and specificity yield an F-score that surpasses the threshold, hence both models are within a margin of error in these metrics. This suggests both models have a similar capability to learn low-risk features that are present before hospitalisation, whereas, RetainEXT is more sensitive to high-risk visits or medical codes which would allow for early intervention and possibly a reduction in the risk of hospitalisation.

The increase in performance between RetainEX and our model is mainly attributed to the stacked LSTM layers and the time-interval representations. RetainEXT provides the added ability to understand the patient's risk after each GP visit, which is critical for the frequency analysis. The model discovered that patients with multiple secondary care referrals are at high risk of being hospitalised due to COVID-19. Additionally, the sample weights have assisted in learning from instances that make up 1.59% of the dataset.

B. Global Feature Importance

Previous studies have reported that age and underlying comorbidities, such as hypertension, diabetes and cardiovascular diseases, are risk factors for patients admitted due to COVID-19 [19]. Congruent with their findings, the average

TABLE I
RETAINEXT AND BASELINE PERFORMANCES ON PREDICTING RISK OF HOSPITALISATION DUE TO COVID-19

	Cascading LSTM [18]	1D-CNN + LSTM	RetainEXT (1-Time Diff.)	RetainEXT (3-Time Diff.)				RetainEX [12]	RetainEXT (Extra Emb)	F-Test Best RetainEX vs EXT
Emb. Units			200	200	200	128	128	128, 128	128,128	
LSTM Units			64,64	64,64	64, 64	64,64	128,128	128	128,128	
Dense Units			1	1	50, 1	50,1	50,1	1	50,1	
Sensitivity	55.97%	46.66%	44.83%	58.79%	62.67%	66.03%	56.19%	62.10%	54.10%	0.63
Specificity	88.92%	73.25%	99.23%	98.92%	99.24%	98.77%	99.17%	98.65%	99.32%	2.43
Positive Predictive Value	40.60%	19.11%	52.71%	51.02	61.44%	50.71%	56.44%	46.74%	60.25%	1.39
AUROC	72.00%	59.96%	72.03%	78.86%	80.96%	82.40%	77.68%	80.37%	76.71%	0.65
F1 Score	0.47	0.27	0.48	0.55	0.62	0.57	0.56	0.53	0.57	0.96
Run-time (Min/epoch)	12	3	10	10	12	12	25	13	32	

Table I illustrates the performance metrics for RetainEXT, its predecessor, RetainEX and other state-of-the-art temporal classification models. The smaller embedding space with 128 units significantly improves model sensitivity. Additionally, F-test scores for Area Under the Receiver Operating Characteristic curve (AUROC) are below F-score at 0.05 critical value, which suggests the null hypothesis can be rejected and assume a significant difference between RetainEX and our model, RetainEXT.

patient age in our dataset is 69.8 years and care home administrative codes were observed in the global feature importance. Elderly patients have poorer immune responses and experience an increased number of age-related comorbidities and therefore are at higher risk of hospitalisation from COVID-19 or other infections.

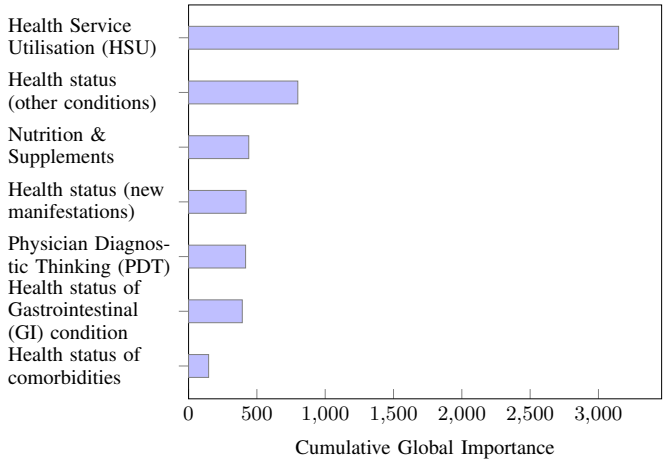

Fig. 2. Global grouped feature importance for predicting the risk of hospitalisation due to COVID-19. HSU group features are significantly more important as the patient will have multiple encounters with the NHS before being hospitalised. Of note, the high importance of Nutrition & Supplements may relate to the patient's diet playing a vital role in mitigating the risk of adverse outcomes.

Fig. 2 shows a set of grouped markers and their overall importance in assessing the risk of hospitalisation due to COVID-19. These groups were created under clinical supervision to understand aggregated markers and medications instead of singular events. For example, a patient receiving GI treatment for a particular condition and a new medication such as an over-the-counter anti-acid suggests that the patient is experiencing a slight deterioration of the condition but that the new drug is a determinant of the GI condition.

The most important and prevalent category of markers is the hospital service utilisation group, which includes anything from receiving a letter from a specialist to seeing a respiratory physician. These administrative codes are vital in understanding the patient's journey through the healthcare system, and the quantity of referrals may suggest the severity of the condition.

The next vital group of markers is the other conditions, which include long-term conditions or medication, such as sleep apnoea medication. While these may not be serious illnesses, the need for constant medication can deteriorate immune response over time.

Of interest is the involvement of a nutritionist or supplements in predicting hospitalisation risk. Correct nutrition can have a significant influence on immune response. The supplement ferrous fumarate indicates anaemia, which is strongly associated with hospitalisation. Anaemia leads to severe outcomes due to COVID-19 [20].

Features in the Physician Diagnostic Thinking (PDT) group are clinical tests to assess the patient's state resulting from a clinician's suspicion. Tests such as urea microscopy are generally conducted when a physician suspects an infection without knowing the cause. White cell differential count, creatinine phosphate and lymphocyte proportion are also among the top 100 important markers, which are also found in the literature [1], [2]. Nonetheless, a patient usually undergoes a Full Blood Count (FBC) just before admission or during the hospital stay, which may make these tests confounders of hospitalisation.

Finally, the frequency analysis showed that most patients have had telephone interactions with NHS staff or have received a specialist's letter before being hospitalised. This underscores the significance of the Health Service utilisation group markers.

C. Limitations

- The available data is not well documented; hence confounding codes may be present in the prediction model and overestimate feature importance.

- The models compared with RetainEX were trained using an encoded binary vector for each visit and may differ in performance and time to convergence.
- The use of global feature importance requires all dimensions of the embedding to be added together. This merely resembles an estimate of the significance of the feature, and further analysis of the latent space may provide deeper insights.

IV. Conclusion and Future Work

In the initial phases of the pandemic, COVID-Related hospitalisations have heavily overwhelmed the healthcare systems, the impact of which is still felt today. With the resurgence of COVID-19 infections and reinfections globally, it is crucial that healthcare resources are used effectively. Presently, the best performing model only achieves a 61% accuracy in predicting hospitalisations due to COVID-19.

We conducted fundamental improvements on an interpretable TNN, RetainEX, and provided additional tools to learn from imbalanced EHR when predicting the risk of adverse outcomes. We leveraged on the predictive power of RNNs and combined it with a sophisticated attention generation process. RetainEXT obtained 82.40% AUROC on assessing risks of hospitalisation due to COVID-19. The model identified key features including the importance of telephone calls with patients to assess their severity, poor GI conditions and low levels of ferrous fumarate indicating anaemia which are consistent with existing literature. Additionally, PDT pathways provide valuable insights into possible infections.

In sum, our model was able to outperform existing models in predicting COVID-19 hospitalisations while providing better interpretability.

Future include developing visualisation tools to better interpret the findings and applying it to more diverse sets of medical records. This increases the model's reliability and helps in understanding other rare diseases. Additional improvements include using a cascading architecture and custom loss functions that reward early diagnosis. We believe the lessons from this study can guide future researchers in building interpretable recurrent neural network models.

References

[1] B. Mahboub, M. T. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of covid-19 hospital length of stay and risk of death using artificial intelligence-based modeling," *Frontiers in Medicine*, vol. 8, p. 389, 5 2021.

[2] G. Wu, P. Yang, Y. Xie, H. C. Woodruff, X. Rao, J. Guiot, A.-N. Frix, R. Louis, M. Moutschen, J. Li, J. Li, C. Yan, D. Du, S. Zhao, Y. Ding, B. Liu, W. Sun, F. Albarello, A. D'Abramo, V. Schininà, E. Nicastri, M. Occhipinti, G. Barisione, E. Barisione, I. Halilaj, P. Lovinfosse, X. Wang, J. Wu, and P. Lambin, "Development of a clinical decision support system for severity risk prediction and triage of covid-19 patients at hospital admission: an international multicenter study," *European Respiratory Journal*, p. 2001104, 7 2020.

[3] D. Assaf, Y. Gutman, Y. Neuman, G. Segal, S. Amit, S. Gefen-Halevi, N. Shilo, A. Epstein, R. Mor-Cohen, A. Biber, G. Rahav, I. Levy, and A. Tirosh, "Utilization of machine-learning models to accurately predict the risk for critical covid-19," *Internal and Emergency Medicine*, vol. 15, pp. 1435–1443, 11 2020.

[4] A. Stachel, K. Daniel, D. Ding, F. Francois, M. Phillips, and J. Lighter, "Development and validation of a machine learning model to predict mortality risk in patients with covid-19," *BMJ Health & Care Informatics*, vol. 28, p. e100235, 5 2021.

[5] F. Rahimian, G. Salimi-Khorshidi, A. H. Payberah, J. Tran, R. A. Solares, F. Raimondi, M. Nazarzadeh, D. Canoy, and K. Rahimi, "Predicting the risk of emergency admission with machine learning: Development and validation using linked electronic health records," *PLOS Medicine*, vol. 15, p. e1002695, 11 2018.

[6] O. Noy, D. Coster, M. Metzger, I. Atar, S. Shenhar-Tsarfaty, S. Berliner, G. Rahav, O. Rogowski, and R. Shamir, "A machine learning model for predicting deterioration of covid-19 inpatients," *Scientific Reports*, vol. 12, p. 2630, 12 2022.

[7] S. Wollenstein-Betech, C. G. Cassandras, and I. C. Paschalidis, "Personalized predictive models for symptomatic covid-19 patients using basic preconditions: Hospitalizations, mortality, and the need for an icu or ventilator." *medRxiv : the preprint server for health sciences*, 5 2020.

[8] Y. Fu, W. Zhong, T. Liu, J. Li, K. Xiao, X. Ma, L. Xie, J. Jiang, H. Zhou, R. Liu, and W. Zhang, "Early prediction model for critical illness of hospitalized covid-19 patients based on machine learning techniques," *Frontiers in Public Health*, vol. 10, 5 2022.

[9] A. A. Willette, S. A. Willette, Q. Wang, C. Pappas, B. S. Klinedinst, S. Le, B. Larsen, A. Pollpeter, T. Li, J. P. Mochel, K. Allenspach, N. Brenner, and T. Waterboer, "Using machine learning to predict covid-19 infection and severity risk among 4510 aged adults: a uk biobank cohort study," *Scientific Reports*, vol. 12, p. 7736, 12 2022.

[10] E. Choi, M. T. Bahadori, A. Schuetz, W. F. Stewart, and J. Sun, "Doctor ai: Predicting clinical events via recurrent neural networks," *CoRR*, vol. 56, 11 2015.

[11] E. Choi, M. T. Bahadori, J. A. Kulas, A. Schuetz, W. F. Stewart, and J. Sun, "Retain: An interpretable predictive model for healthcare using reverse time attention mechanism," in *Proceedings of the 30th International Conference on Neural Information Processing Systems*, ser. NIPS'16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 3512–3520.

[12] B. C. Kwon, M.-J. Choi, J. T. Kim, E. Choi, Y. B. Kim, S. Kwon, J. Sun, and J. Choo, "Retainvis: Visual analytics with interpretable and interactive recurrent neural networks on electronic medical records," *IEEE Transactions on Visualization and Computer Graphics*, vol. 25, pp. 299–309, 1 2019.

[13] R. A. Lyons, K. H. Jones, G. John, C. J. Brooks, J.-P. Verplancke, D. V. Ford, G. Brown, and K. Leake, "The sail databank: linking multiple health and social care datasets," *BMC Medical Informatics and Decision Making*, vol. 9, p. 3, 12 2009.

[14] I. M. Baytas, C. Xiao, X. Zhang, F. Wang, A. K. Jain, and J. Zhou, "Patient subtyping via time-aware lstm networks," in *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '17. New York, NY, USA: Association for Computing Machinery, 2017, p. 65–74.

[15] Z. C. Lipton, D. C. Kale, C. P. Elkan, and R. C. Wetzel, "Learning to diagnose with lstm recurrent neural networks," *CoRR*, vol. abs/1511.03677, 2016.

[16] I. Petersen, I. Douglas, and H. Whitaker, "Self controlled case series methods: an alternative to standard epidemiological study designs," *BMJ*, p. i4515, 9 2016.

[17] W. L. Cava, C. Bauer, J. H. Moore, and S. A. Pendergrass, "Interpretation of machine learning predictions for patient outcomes in electronic health records." *AMIA ... Annual Symposium proceedings. AMIA Symposium*, vol. 2019, pp. 572–581, 2019.

[18] G. Tsang and X. Xie, "Deep learning based sepsis intervention: The modelling and prediction of severe sepsis onset," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 8671–8678.

[19] D. Wang, B. Hu, C. Hu, F. Zhu, X. Liu, J. Zhang, B. Wang, H. Xiang, Z. Cheng, Y. Xiong, Y. Zhao, Y. Li, X. Wang, and Z. Peng, "Clinical characteristics of 138 hospitalized patients with 2019 novel coronavirus-infected pneumonia in wuhan, china." *JAMA*, vol. 323, pp. 1061–1069, 2020.

[20] M. F. Dinevari, M. H. Somi, E. S. Majd, M. A. Farhangi, and Z. Nikniaz, "Anemia predicts poor outcomes of covid-19 in hospitalized patients: a prospective study in iran," *BMC Infectious Diseases*, vol. 21, p. 170, 12 2021.