Students:

- Mahi Chowdhury
- Hala Awad
- George Nzaji
- Joshlin Ronisha Devadhasan Merwin Britto
- Catherine Ojo

Group Name:

- Health Data Science

Project Name:

- NHS Covid Predictions

Company Involved:

- NHS Lancashire Teaching

Source Code

- github.com/premodelling/infections: The size of the group's data set and source code is quite large, hence the link.

INTRODUCTION

Since late January 2020, there have been over 17,515,199 cases of SARS-CoV-2 infections in the United Kingdom (UK) (Flynn et al. 2020). The virus causes high fever, coughing, shortness of breath, pneumonia, as well as serious respiratory infections. The periods of high UK hospitalisation numbers were a serious threat to an already overburdened healthcare system. Indeed, the National Health Service (NHS) cannot meet the needs of many patients with urgent medical conditions due to overcrowded hospitals.(Anderson et al. 2021)

It is essential to monitor and forecast new inpatient admissions, during an infectious disease outbreak or pandemic, in order to manage hospital resources efficiently, reduce overcrowding, and improve the quality of care received. To this end, this project focuses on the development of an inpatients' prediction model that may help a hospital's contingency planning during an infectious disease outbreak/pandemic.

The project's aim is the development of a prediction model that forecasts what the expected number of patient admissions will/might be - per day, N days ahead, and per NHS Trust - during an infectious disease pandemic. The aim underlies the research question - how many future admissions should a NHS trust expect during an infectious disease pandemic? The infectious disease in focus is coronavirus 19 disease; and by virtue of the characteristics of SAR-CoV-2 infections & coronavirus 19 disease the project's objectives are to

- Investigate the question, how many days of infection history lead to the most accurate forecasts?
- Forecast the expected daily new admissions for the next 15 days

The next section outlines the project's methods. In relation to model development, we focused on algorithms that use historical data to predict future events. The best results are due to the developed Long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997); the testing phase mean absolute errors are below 0.75, whilst the validation phase mean absolute errors are below 0.5. For the projects code please visit github.com/premodelling/infections.

METHODS

The schematic illustration of *fig. 1* outlines the project's data engineering, modelling, and evaluation steps, which underlie the project's research strategy. This section briefly discusses the research strategy, and the steps.
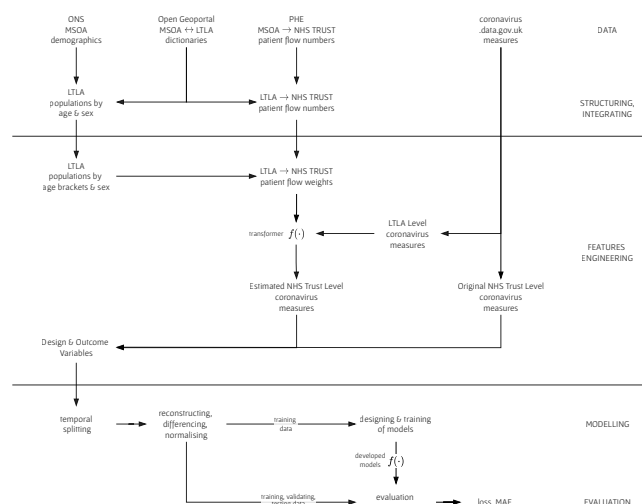

Figure 1: The project's processing, analysis, modelling, and evaluation steps. Please refer to the methodologies section for a brief description of (a) the patient flow weights, and (b) the estimation of NHS trust level measures via flow weights and LTLA level measures. MSOA: middle layer super output area, LTLA: lower tier local authority, ONS: office for national statistics, NHS: national health service, PHE: Public Health England.

RESEARCH STRATEGY

Foremost, understand the problem; guided by domain expertise. Next, collect the required data. Finally, develop a set of appropriate models, and evaluate their results. (Oates 2006)

DATA ENGINEERING

Data Collection.

The data sources are: (a) the coronavirus.data.gov.uk application programming interface (API) for England's SARS-CoV-2 infections measures, (b) the office for national statistics (ONS) for population estimates, (c) Public Health England (PHE) for the annual intake of patients from one or more middle layer super output areas to an NHS Trust, and (d) the Open Geography Portal (geoportal) for the middle layer super output area (MSOA) ↔ lower tier local authority (LTLA) geographic codes mappings. The SARS-CoV-2 infections measures

of this project span the period 1 March 2020 until 20 January 2022.

Structuring & Integrating.

The structuring and integrating segment of $fig.1$ ensures that all the data sets

- have a structured data file set up, and
- are appropriately mapped

as illustrated.

Features Engineering.

The aim of *fig. 1's* feature engineering segment is the construction of the design matrix & outcome vector variables. The design matrix variables are the set of predictors, i.e., independent variables. The variables are

- covidOccupiedBeds: The no. of beds occupied by coronavirus disease patients.

- covidOccupiedMVBeds: The no. of mechanical ventilation beds occupied by coronavirus disease patients.

- estimatedNewAdmissions: **The outcome variable**. Estimated by NHS England.

- EDC0-4, EDC5-9, ..., EDC90+: The estimated daily cases (EDC) by age group.

- newDeaths28DaysByDeathDate: The no. of estimated daily deaths, such that each death occurred *within 28 days of a first positive laboratory-confirmed test.*

- EDV12-15, EDV16-17, ..., EDV90+: The estimated no. of daily vaccinations (EDV) by age group; second vaccinations.

The first three are original NHS Trust level measures available via the coronavirus.data.gov.uk application programming interface, whereas the remaining variables are project estimated NHS Trust level measures. An estimated NHS Trust value is

$$e_t = \sum_l \lambda_{l, t} m_l \qquad (1)$$

whereby

$$\lambda_{l, t} = \frac{\beta_{l, t}}{\rho_l} \qquad (2)$$

and

variable	description
l	A LTLA
t	A NHS Trust
m_l	A LTLA measure from coronavirus.data.gov.uk, e.g., the no. of daily cases.
e_t	An estimate NHS Trust level measure w.r.t. LTLA regions from whence it has received patients.
$\lambda_{l, t}$	The flow weight from a LTLA to a NHS Trust.
ρ_l	The patient population of a LTLA.
$\beta_{l, t}$	The number of LTLA patients $\rightarrow$ NHS Trust w.r.t. a specified year

MODELLING

The Algorithms

The SARS-CoV-2 infections measures have both spatial and temporal features, i.e., the spatially spread set of NHS Trusts, and the infection dynamics, respectively. A number of algorithms have been developed for spatio-temporal prediction problems. The project focuses on **(a)** Long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997), **(b)** Gated Recurrent Unit (GRU) (Cho et al. 2014), and **(c)** [Temporal] Convolutional Neural Networks (CNN) (Bai, Kolter, and Koltun 2018), because they were developed for

- Spatio-temporal problems.
- Learning from the past.

Forecasting & History

The outlined algorithms forecast 15 days into the future w.r.t. "varying days of history" (*fig. 2*).
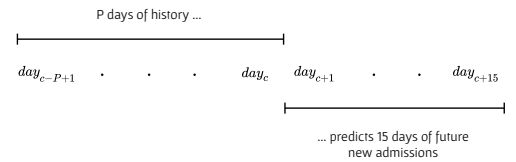
Figure 2: Prediction windows logic

Pre-modelling Procedures

The pre-modelling procedures are

- Temporal Splitting
- Differencing
- Reconstruction
- Normalisation

A temporal splitting step splits a data set into training, validation, and testing data sets along the time dimension. A training, validation, and testing data set, per NHS Trust, is created. In relation to this project, after differencing each variable point is

$$x_{t+1} - x_t \tag{3}$$

Reconstruction refers to the merging of the data splits per trust into 3 data sets; a training, validation, and testing data set. Merging is not by concatenation, it's a positional merge that is appropriate for developing an overarching model instead of a model per entity. For example, instead of a model for each of the 140 NHS Trusts, a single overarching model.

Finally, the normalisation step ensures a standard scale across all variables.

EVALUATION

The results section summarises the modelling results. Model evaluation is via the error measures

$$\text{loss} = \frac{1}{N} \sum_{n=1}^{N} (y_t(n) - y_p(n))^2 \tag{4}$$

and

$$\text{MAE} = \frac{1}{N} \sum_{n=1}^{N} |y_t(n) - y_p(n)| \tag{5}$$

wherein

variable	description
y_t	a true outcome value
y_p	a predicted outcome value
loss	the mean squared error
MAE	the mean absolute error
N	the length of the outcome vector

RESULTS

MODEL EVALUATION

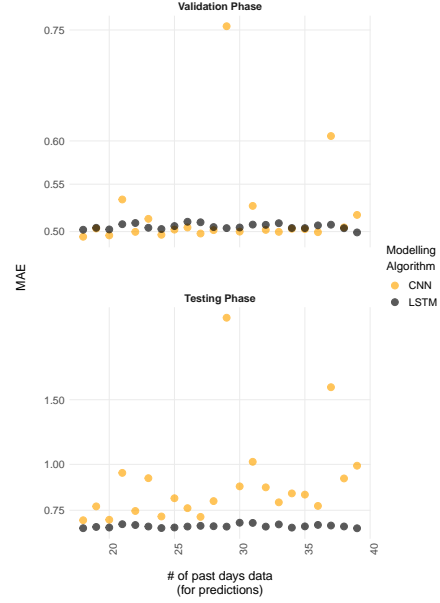Let the *history window* be the number of past days of data used for predictions.

Figure 3: The mean absolute errors of the validation and testing phase.

The graphs of *fig. 3* summarise the validation & testing phase MAE. Per algorithm type, and regardless of the history window, the testing phase MAE values are higher than the validation phase values. The LSTM error values are consistently low. Whereas, the CNN error values are rather haphazard, especially the testing phase values. The pattern of the CNN errors might be due to over-fitting, or perhaps a different CNN architecture might be appropriate. The lowest LSTM MAE value is due to a 39-day history window.

The pattern of the loss & MAE values are quite similar (Cf. *fig. 3* & *fig. 4*). Once again, the LSTM error values are consistently low (*fig. 4*). The CNN testing phase errors are quite haphazard, but its validation phase errors are almost consistent.

Per model, the graphs of *fig. 5* summarise each history window's MAE per epoch. The CNN MAE values are spread-out, whereas the LSTM MAE values are compact.
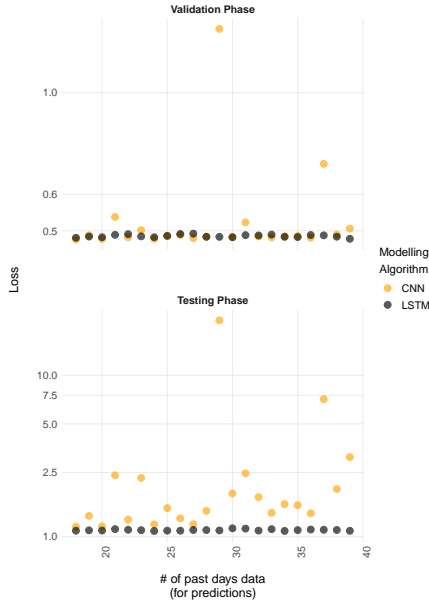
4

The LSTM model converges faster than the CNN model, hence the fewer epochs.
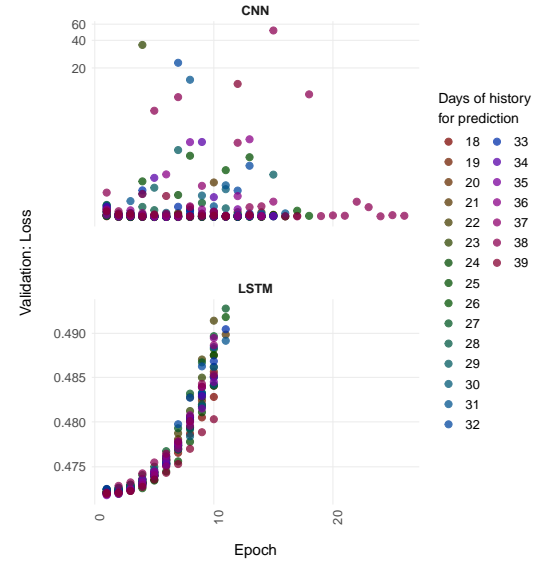

Figure 6: The loss errors

Figure 4: Loss w.r.t. the ...

The distribution patterns of the loss values (*fig. 6*) are similar to those of the MAE values. Altogether, it seems

- It is possible to predict future patient admissions via the LSTM model, in conjunction with the predictors listed within the feature engineering section.
- A slightly more complex CNN architecture might lead to lower, and more stable, CNN model error measures.
- The estimated NHS Trust level values might be good estimates of the real values; further tests required.

Due to time and insufficient computing resources, the GRU model runs had to be stopped. However, the GitHub repository includes the GRU class.
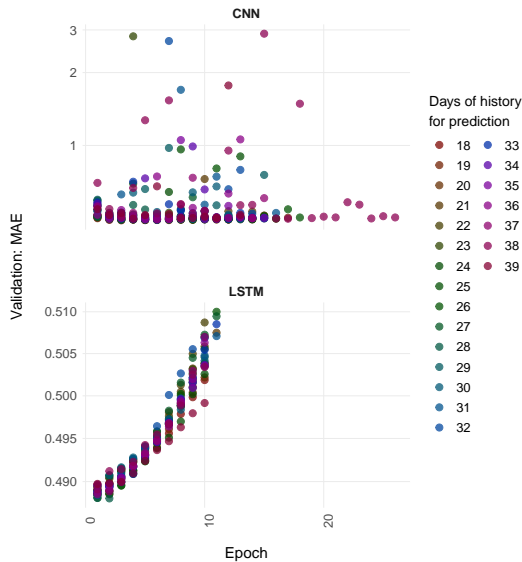
BIASES & VALIDITY

In order to combat any biases that may hinder the validity of our study we followed the Transparent Reporting of multivariate prediction model for Individual Prognosis or Diagnosis (TRIPOD) (Collins et al. 2015). This is important because the systematic review by Wynants et al. (Wynants et al. 2020) noted that almost all COVID-19 prediction models they reviewed exhibited high bias. In terms of validity

- Internal Validity: The degree to which a cause and

Figure 5: Pending

effect relationship can be discerned between the predictors and the outcome variable, i.e., expected new admissions, is limited. It is not the question in focus.

- Construct Validity: The operational setting does reflect the construct. Each model predicts - the number of new inpatient admissions per day, per trust, for the next 15 days. This is in line with the project's aim.

- External Validity: The study's conclusions can be generalised to England, the data covers the entire region. The study did include narrow external validation - the time points of the testing data do not overlap with the earlier time points of the validation or training data. However, evaluation is in relation to England only, to generalise further the models should be tested in different geographic regions.

- Conclusion Validity: The loss and mean absolute errors associated with the developed prediction models, especially the LSTM model, are *small*. Hence, it is probably reasonable to conclude that the predictors listed within the feature engineering section are plausible predictors.

CONCLUSIONS

In general, it is nigh impossible to state whether one of the machine learning algorithms is better than the other. Partly because each algorithm's architecture can have numerous designs.

An option that would have led to a better model is boosting, i.e., a model built via an ensemble of models via methods such as AdaBoost (Freund and Schapire 1996).

REFERENCES

Anderson, M, E Pitchforth, M Asaria, C Brayne, B Casadei, and A Charlesworth. 2021. "LSE–Lancet Commission on the Future of the NHS: Re-Laying the Foundations for an Equitable and Efficient Health and Care Service After COVID-19." *The Lancet Commissions* 397. https://doi.org/10.1016/S0140-6736(21)00232-4.

Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun. 2018. "An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling." *CoRR* abs/1803.01271. http://arxiv.org/abs/1803.01271.

Cho, Kyunghyun, Bart van Merrienboer, Çaglar Gülçehre, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. 2014. "Learning Phrase Representations Using RNN Encoder-Decoder for Statistical Machine Translation." *CoRR* abs/1406.1078. http://arxiv.org/abs/1406.1078.

Collins, Gary S., Johannes B. Reitsma, Douglas G. Altman, and Karel G. M. Moons. 2015. "Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD): The TRIPOD Statement." *Annals of Internal Medicine* 162 (1): 55–63. https://doi.org/10.7326/M14-0697.

Flynn, Darren, Eoin Moloney, Nawaraj Bhattarai, Jason Scott, Matthew Breckons, Leah Avery, and Naomi Moyd. 2020. "COVID-19 Pandemic in the United Kingdom." *Health Policy and Technology* 9 (4): 673–91. https://doi.org/10.1016/j.hlpt.2020.08.003.

Freund, Yoav, and Robert E. Schapire. 1996. "Game Theory, on-Line Prediction and Boosting." In *Proceedings of the Ninth Annual Conference on Computational Learning Theory*, 325–32. COLT '96. New York, NY, USA: Association for Computing Machinery. https://doi.org/10.1145/238061.238163.

Hochreiter, Sepp, and Jürgen Schmidhuber. 1997. "Long Short-Term Memory." *Neural Computation* 9 (8): 1735–80. https://doi.org/10.1162/neco.1997.9.8.1735.

Oates, Briony J. 2006. *Researching Information Systems and Computing.* SAGE.

Wynants, Laure, Ben Van Calster, Gary S Collins, and et al. 2020. "Prediction Models for Diagnosis and Prognosis of Covid-19: Systematic Review and Critical Appraisal." *BMJ* 369. https://doi.org/doi.org/10.1136/bmj.m1328.