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). This virus can cause high fever, coughing, shortness of breath, pneumonia, as well as serious respiratory infections MISSING REF: Original reference required. The growing number of daily infections and hospitalisations in the UK constitute 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 MISSING REF: Original reference required.

It is essential to monitor and forecast new inpatient admissions in order to manage hospital resources efficiently, reduce overcrowding, and improve the quality of care received MISSING REF .... Therefore, this study focuses on the development of a coronavirus 19 disease new inpatients' prediction model. Addressing this research question may help improve NHS performance and patient outcomes by providing more efficient and higher-quality patient care and optimising the allocation of limited resources to meet the growing demand for hospital places MISSING REF: Huang2019.

The succeeding section outlines the project's research strategy, and the data engineering, modelling, and evaluation steps. By virtue Thereafter, ...

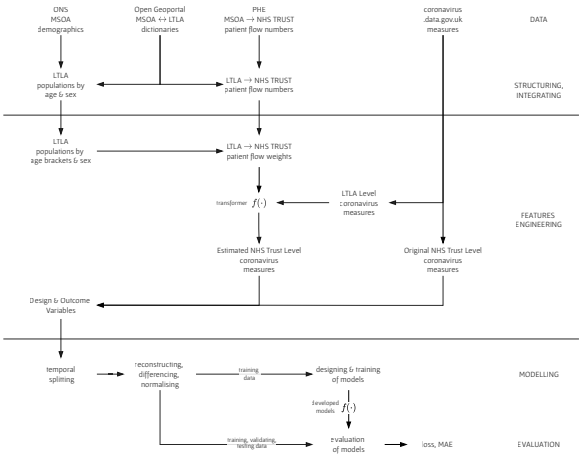

Figure 1: The project's processing, analysis, modelling, and evaluation steps. Please refer to the methodologies section for a brief description of (a) the patient flow weights, and (b) the estimation of NHS trust level measures via flow weights and LTLA level measures. MSOA: middle layer super output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

METHODS

The schematic illustration of *figure ...* 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.

RESEARCH STRATEGY

In progress ... includes research strategy (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.

Structuring & Integrating.

The structuring and integrating segment of $fig.4$ ensures that all the data sets

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

as illustrated.

Features Engineering.

The aim of the 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. ... *fig. 1* outlines the variables-construction steps, and 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 API, whereas the remaining variables are project estimated NHS Trust level measures. The project estimates rely on (a) the LTLA Level measures of the API, and (b) ...

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 such prediction challenges ... Long short-term memory (LSTM) (Hochreiter and Schmidhuber 1997), Gated Recurrent Unit (GRU) (Cho et al. 2014), Convolutional Neural Networks (CNN) (Bai, Kolter, and Koltun 2018).

LSTM: In RNN, each layer takes the input data and the hidden state of past hidden state and outputs the hidden state, then this hidden state is given along with the input to the next layer. This hidden state holds the short-term memory. For the solution that requires long term memory, RNN would not be feasible option. To overcome this shortcoming we introduce LSTM. In LSTM, each layer outputs cell state which is responsible for holding the long-time memory in addition to the hidden state. A LSTM cell contains simple RNN cells, cell states, forget gates, input gates and output gates. Input data and hidden state is given as input to all the components in the LSTM cell. Forget gate determines what to dismiss from the input and hidden state and hand that information to the cell state. Input gate in addition with the simple RNN cell determines what new information should be added to the cell state. This is outputted as cell state of the layer. Cell state value is also given as the input to the output gate to create a hidden state for the current layer. Hidden state and the cell state is given as the input to the next layer and the process continuous.

GRU: In contrast with LSTM, GRU has only hidden state. But it holds both the long term and short-term memory unlike the traditional RNN. GRU has two gates, the update gate and reset gate. The update gate determines how much to retain of the past memory. Reset gate determines how much past memory to forget. Reset gate output is used with the past hidden state and the input data to find the current memory content. This current memory content is used along with the update gate output and past hidden state to find the current hidden state output. This hidden state is then passed as the input to the next layer and the process continuous.

CNN: Though the CNN is famous for computer vision, the one-dimensional convolution is useful for time-series ... CNN has filters that slides over the data to capture the features of the data. These filters are the learnt by training. The filters are followed by pooling to reduce the size of the parameters that are learnt by the algorithm.

Forecasting & History

The outlined algorithms ... addressed via modelling w.r.t. varying historical data windows. Altogether ... forecasting 15 days into the future w.r.t. "varying days of history" ... via window logic

Pre-modelling Procedures

- temporal splitting ... training, validation, and testing data sets
- reconstructing
- differencing
- normalisation

EVALUATION

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

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

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

wherein

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

Table 1: The error formulae terms.

RESULTS

MODEL EVALUATION

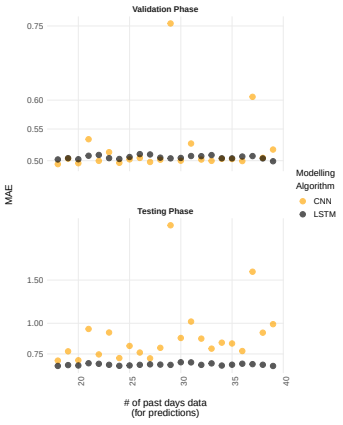

Figure 2: MAE w.r.t. the ...

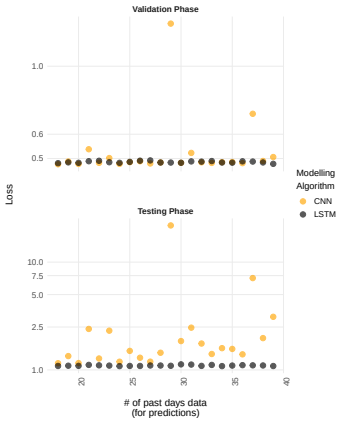

Figure 3: Loss w.r.t. the ...

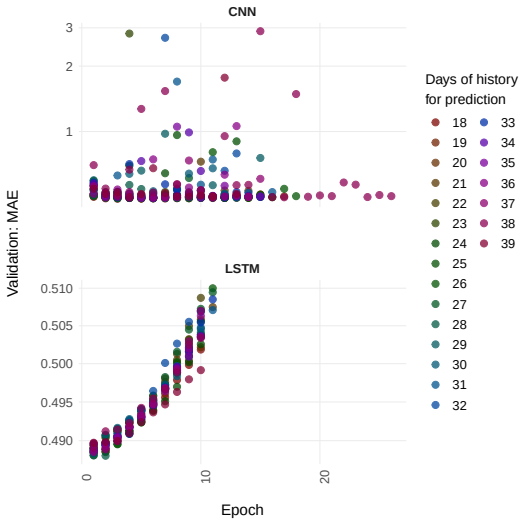

Figure 4: Pending

Project guide:

- Explain your results and what your analysis revealed ()
- What implications would your analysis' results have? How do your findings relate to the original question?

BIASES & VALIDITY

- Were there potential biases in your work?
- Validity (remember to discuss what would have been done differently to address identified limitations)

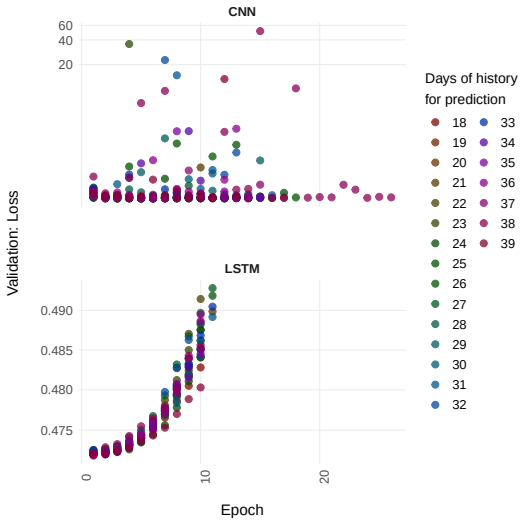

Figure 5: The loss errors

CONCLUSIONS

Project Guide:

- Reflection on the approach taken. (Appropriate?)
- How would you have improved the approach in future? (Alternative methodologies, models, etc)

REFERENCES

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.

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.

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.