An AI-enabled predictive analytics dashboard for acute neurosurgical referrals

Supplementary Document

Software Demonstration

A trial dashboard using synthetic data can be accessed on: [https://referralsdash.herokuapp.com/](https://referralsdash.herokuapp.com/forecast) via a desktop web browser. Please note it can take up to a minute for the dashboard to load on some internet browsers.

A video demonstrating the functionality of the dashboard presenting the data outlined in this manuscript is available on<https://youtu.be/Th2vsCpLHbI>

Supplementary Methods

Python libraries and dependencies

The following dependencies were used in the creation of this dashboard (see code snippet below)

Data pre-processing

Following anonymisation, referral data was uploaded as a *pandas* data-frame. Redundant columns, duplicates and erroneous entries were removed, and all dates and times were transformed to python date-time data-types for further manipulation. Specialist working diagnoses are designated by the on-call neurosurgical registrar when receiving the referral and include a total of 138 different options. The diagnosis is based on the information received at the point of the referral and may be modified as further information is shared or after senior review. Specialist diagnoses were aggregated into 13 primary diagnostic categories: brain tumour, cauda equina syndrome, congenital, subdural haematoma, cranial trauma, degenerative spine, hydrocephalus, infection, spinal trauma, stroke, neurovascular and 'not neurosurgical' (Supplementary Appendix).

<CODE>

```
#Upload anonymised file - either saved as .csv or .pkl
```

```
df_all = pd.read_pickle(filename)
```
#Drop duplicates

df all.drop duplicates(inplace=True)

```
#Drop redundant columns
df all.drop(columns = ['Referring Doctor Name','Bleep or Telephone No','MobileNo','Subsequent Doctor
Grade Name','Subsequent Bleep Number','Subsequent Mobile No','Subsequent Dr Email Address','Subsequent 
Consultant Email Address'], inplace = True)
#Transform date-time entries to datetime datatype
df all = transform to datetime(df all, 'Referral Time')
#Convert specialist working diagnosis into primary diagnostic classification based on diagnosis table -
see Appendix table
diagnosis table = pd.read csv('diagnoses table.csv', low memory=False)
df all = add classification level(df all, diagnosis table,
                                      'Primary Classification')
## RELEVANT PROCESSING FUNCTIONS
def match classification(diagnosis table, classification level,
                              diagnosis):
       diagnosis level = diagnosis table[
       diagnosis table['Specialist working diagnosis'] ==
        diagnosis][classification_level]
       if (len(diagnosis level.values) > 0):
       return diagnosis level.values[0]
       return 'no_match'
def add classification(input df, diagnosis table, classification level):
       df_copy = copy.deepcopy(input_df)
       partial_func = partial(match_classification, diagnosis_table,
                              classification_level)
       df copy[classification\ level] = df\ copy[ 'Specialist Working Diagnosis'].apply(partial_func)
       return df_copy
def transform_to_datetime(df, time_col):
       copy = df.copy()copy[time_col] = pd.to datetime(copy[time_col], dayfirst=True)
       return copy
```
Geographical information

Using the name of the referring site, an application programming interface (API) request is made to *openstreetmap.org* to derive the latitude and longitude of referral site locations. This location data is then cached and parsed to a geographical plotting function.

<CODE>


```
geocount.reset_index(inplace = True)
       #Generate empty columns to fill location data in
       geocount['add'] = 0geocount['lon'] = 0geocount['lat'] = 0geocount = geocount.sort_values(by = 'total', ascending = False)
       geocount.reset_index(drop=True, inplace=True)
       #Generate list of unique hospitals from dataframe
       hosplist = geocount['Referring Hospital'].unique()
       hosplist = hosplist.tolist()
       #For each unique hospital, perform an API request
       for i,v in enumerate(hosplist):
        address = v
        url = 'https://nominatim.openstreetmap.org/search/' + urllib.parse.quote(address)+'?format=json'
        response = requests.get(url).json()
        geocount.loc[i,['add', 'lon', 'lat']] = [address,response[0]["lon"],response[0]["lat"]]
       #Create seperate dataframe to save location data to cache
       locmatch = pd.DataFrame()
       locmatch['Referring Hospital'] = geocount['add']
       locmatch['lon'] = geocount.lon
       locmatch['lat'] = geocount.lat
       locmatch.to_csv('locmatch2.csv')
       return geocount, hosplist, locmatch
##GENERATE GEOGRAPHICAL FIGURE
```

```
def geospatial(df, date1, date2,classification):
       #select data by time
       geocount = single_period(df, date1, date2)
       #filter df by primary classification and sort
       if classification != "all":
           geocount = geocount[geocount['Primary Classification'] == classification]
       geocount = geocount.groupby(by=['Primary Classification','Referring 
Hospital'])[['Age']].count().unstack(level=0)
       geocount.columns = geocount.columns.droplevel()
       geocount.fillna(value=0,inplace=True)
       geocount['total'] = geocount.sum(axis=1)
       geocount.reset_index(inplace = True)
       geocount = geocount.sort_values(by='total', ascending=False)
       geocount.reset_index(drop=True, inplace=True)
       geocount = geocount.merge(locmatch, on='Referring Hospital')
       #create figure, can be scaled by color or size. Center is the receiving hospital
       fig5 = px.scatter_mapbox(geocount,
                              lat="lat",
                              lon="lon",
                              hover_name="Referring Hospital",
                              hover_data=["total"],
```
 zoom=9, height=300,

size=geocount.total,

```
size max=40,
                               color="total",
                               center={
                                'lat': 51.6,
                               'lon': -0.26
, where \{ \} , we have the contract of \} , \{ \}opacity=0.7)
       #update layouts
       fig5.update layout(mapbox style='carto-positron')
       fig5['data'][0]['showlegend'] = False
       fig5['data'][0]['name'] = 'Referring Site'
       fig5.update layout(margin={"r": 0, "t": 0, "l": 0, "b": 0})
       fig5.update layout(autosize=True, width=800, height=800)
       return fig5
## RELEVANT PROCESSING FUNCTIONS
def single_period(df, date1, date2):
       return df[(df['Referral Time'] >= date1) & (df['Referral Time'] < date2)]
```
Implementation of time-series forecasting models

Three forecasting algorithms were trialled in this work: an automated pipeline which combined Seasonal and Trend decomposition using Loess (STL) with an automatic Autoregressive Integrated Moving Average (Auto-ARIMA) model, a Convolutional Neural Network - Long Short-Term Memory (CNN-LSTM) network and Prophet. In this section we describe how each model was implemented.

Supplementary Table 1. Median weekday and weekend volumes.

All referrals and the four highest referring categories are shown. p values shown are Bonferroni multiple comparison corrected following univariate Mann-Whitney U tests. (NS = not significant)

STL + ARIMA

We performed an exploratory analysis of the time-series using auto-correlation and partial auto-correlation plots in combination with augmented Dickey-Fuller testing to determine the degree of stationarity in the data and assist in defining initial parameters for seasonal decomposition and upper and lower parameter limits for the auto-ARIMA grid search.

```
### STL/Auto-ARIMA model
#Run EDA on weekly time-series first to manually check seasonality
#Set variables
res = []#STL period corresponds to expected seasonality. 4 chosen to reflect monthly seasonal changes.
##Also can use 52 for yearly or 26 for 6-monthly seasonality
period = 4#How long into future/out-of-sample to make forecast
future = \theta#95% Confidence interval
confidence = 0.05#STL decomposition with default parameters and period - can be further tuned using grid search
res = STL(df, period = period, robust = False).fit()#Seasonal auto-ARIMA, stepwise can be changed to True for more thorough grid search. Upper and lower 
limits regarding p, q, d determined by initial exploratory analysis of data set
smodel = pm.auto arima(res.seasonal,
                       start_p=0, max_p=5,
                       start_q=0, max_q=5,
                       seasonal=False,
                       stepwise = False,
                      start d=0, max d=5,
                       trace=False, error_action='ignore');
#Trend auto-ARIMA
tmodel = pm.auto_arima(res.trend,
                       start_p=0, max_p=5,
                      start q=0, max q=5,
                       seasonal=False,
                       stepwise = False,
                       start_d=0, max_d=5,
                       trace=False, error_action='ignore');
#Residual auto-ARIMA
rmodel = pm.auto_arima(res.resid,
                      start p=0, max p=5,
                      start q=0, max q=5,
                       seasonal=False,
                       stepwise = False,
                       start_d=0, max_d=5,
                       trace=False, error_action='ignore');
#Modelling seasonality
modelsea = SARIMAX(res.seasonal, order = smodel.order, seasonal_order= smodel.seasonal_order).fit()
#If Auto-ARIMA fails then use simple differenced d=1 model for trend
try:
       modeltrend = ARIMA(res.trend, order = tmodel.order, freq=interval).fit()
except:
       modeltrend = ARIMA(res.trend, order = (0,1,0), freq=interval).fit()#Modelling residual
modelres = ARIMA(res.resid, order = rmodel.order, freq=interval).fit()
#Forecasting and recomposition
forecast_season = modelsea.forecast(future, alpha=confidence)
```
forecast trend, std err trend, confidence int trend = modeltrend.forecast(future, alpha=confidence) forecast_resid, std_err_resid, confidence_int_resid = modelres.forecast(future, alpha=confidence) forecast final = forecast season + forecast trend + forecast resid conf = confidence int trend + confidence int resid

CNN - LSTM

```
###CNN-LSTM implementation
#Relevant imports
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import LSTM, Dense, Flatten, TimeDistributed, Conv1D, MaxPooling1D
# define input sequence from dataframe
sequence = df['all'].to_list()
# Set number of steps, keep even
n_steps = 52
# split into an array of subsequences, X = input
X, y = sequence split(sequence, n steps)features = 1n_seq = 2
# divided subsequence into 2 subsamples
n steps2 = n steps/2
# reshape input data for CNN layer
X = X.read. reshape((X.shape[0], n\_seq, n\_steps2, features))
# set up sequential stack model
model = Sequential()
#CNN layer with 64 output filters, kernel size corresponds to length of convolutional window. Input 
shape must match shape from reshape step
model.add(TimeDistributed(Conv1D(filters=64, kernel_size=1, activation='relu'), input_shape=(None, 
n_steps2, n_features)))
# Down samples by pool size
model.add(TimeDistributed(MaxPooling1D(pool_size=2)))
#Flatten to single 1D vector
model.add(TimeDistributed(Flatten()))
#Single LSTM layer with 64 neurons
model.add(LSTM(64, activation='relu'))
#NN dense layer
model.add(Dense(1))
#ADAM optimisation using mse as a cost function
model.compile(optimizer='adam', loss='mse')
model.fit(X, y, epochs=500, verbose=0)
## RELEVANT PROCESSING FUNCTIONS
def sequence_split(timeseries, n_steps):
```

```
#Prepare list variables
X, y = list(), list()for i in range(len(timeseries)):
 # find index at sequence end
end_index = i + n_steps # stop code if has gone past total length of sequence
 if end_index > len(timeseries)-1:
     break
 # divide sequence into subsamples
sub x, sub y = timeseries[i:end index], timeseries[end index]
 X.append(sub_x)
 y.append(sub_y)
return np.array(X), np.array(y)
```
Prophet

<CODE>

```
### Prophet implementation
#Specify dataframe and convert to prophet input
propletely the total index()
prophetdf.columns = ['ds', 'y']
#Specify weeks to predict
prediction = 1
#Specify lockdown period
lockdown = pd.DataFrame({
        'holiday': 'lockdown',
        'ds': pd.to_datetime(['2020-03-23']),
       'lower_window': 0,
       'upper window': 84,
       })
#Set model parameters. Note data is already in weekly format.
model = Property (yearly seasonality=True,weekly seasonality=False,
                daily_seasonality = True,
                seasonality_mode='additive',
                interval_width=0.95,
               changepoint_prior_scale= 0.05,
               seasonality prior scale= 0.1,
               holidays = lockdown)
#Fit model
model.fit(prophetdf)
future = model.make_future_dataframe(periods=prediction,freq='W')
#Make predictions
forecast = model.predict(future)
```
Usability, acceptability and feasibility

This study employed a mixed-method design to assess dashboard usability, acceptability and feasibility. Participants were recruited from the local neurosurgical centre through mailing lists and were included if they had an adequate experience of using the electronic referral system (> 6 months). Participants were excluded if they were aware of the development of the dashboard.

In each testing session, a demonstration of the dashboard's capabilities were shown $($ \sim 10-minutes). As an example which would simulate a typical service evaluation, participants were shown how to use features to audit a particular diagnostic category or time-period. Using a think-aloud protocol, participants were invited to explore the functions of the dashboard independently, after which they completed an electronic questionnaire that incorporated three validated instruments: the System Usability Scale (SUS), Acceptability of Intervention Measure (AIM) and Feasibility of Intervention Measure (FIM) adapted for use. The SUS asks participants to respond to a set of 10 statements using a 5 point Likert scale, with a composite score above 70 defined as "good" usability.

In each of the AIM and FIM scales, participants were presented with 4 statements in reference to the 'intervention' (dashboard) and asked to rate these according to a 5-point Likert Scale. These statements have been previously assessed for substantive and discriminant content validity³[.](https://app.readcube.com/library/829d0e09-047f-41e4-a790-84f30ccd2829/all?uuid=7558850617045383&item_ids=829d0e09-047f-41e4-a790-84f30ccd2829:d0bfdcb7-4042-4ad3-a487-b7f7e519a9af) Two white-box questions were also incorporated into the questionnaire: "Which aspects or features of the dashboard did you find useful?" and "Do you have any suggestions for improving the dashboard?". The questionnaire has been outlined in full in the Supplementary Appendix.

Web application and synthetic data set

A trial version of the dashboard was hosted using Heroku [\(www.heroku.com\)](http://www.heroku.com/), an online service allowing developers to deploy, manage and scale applications. A synthetic data set was created by taking the original anonymised data set and scrambling demographic and clinical variables, while keeping frequency of aggregate diagnostic classes and outcomes the same. Referral locations were shuffled and replaced with names and locations of English Premier League football stadiums to preserve referral site anonymity.

Supplementary Results

Supplementary Figure 1.

Out-of-sample one-year referral projections using all three forecasting algorithms trained on all available data.

User experience and implementation

Analysis of coded user feedback explores possible reasons why the dashboard scored well (Supplementary Table 2). Many users highlighted the 'clarity', 'usefulness' and 'variety' of graphs and figures [M1, C3, C4, R1, R6, R7, R10, R12]. Others commented on dashboard interactivity [R6, R8], in particular the use of drill-down features as being particularly positive. Some users found that the dashboard would help with auditing and research. In particular they found that it gave important insights 'into previously inaccessible big-data' [R1], that it highlighted 'areas of improvement for staff allocation' [C3], suggested 'directions for more focused audit and research' [C5] and that it demonstrated 'why we need to liaise with local referring sites' [R3]. A few users commented on the AI implementation and time-series forecasting functions stating that it would be 'useful in anticipating demand' [R1], and that it 'could be implemented easily' [R8] but 'unsure how it would be applied day-to-day' [R12]. Some users did express concerns about dashboard access in the department [C1, C2], whereas others thought there should be additional functionality to 'export data' or review it in more detail [R2, R11].

Supplementary Table 2. User feedback and interview responses

Supplementary References

[1. Box, G. E. P., Jenkins, G. M., Reinsel, G. C. & Ljung, G. M.](https://app.readcube.com/library/?style=Nature%20Communications) *[Time Series Analysis Forecasting and Control](https://app.readcube.com/library/?style=Nature%20Communications)*[. \(John Wiley &](https://app.readcube.com/library/?style=Nature%20Communications) [Sons, Inc., Hoboken, New Jersey, 2016\).](https://app.readcube.com/library/?style=Nature%20Communications)

[2. Taylor, S. J. & Letham, B. Forecasting at Scale.](https://app.readcube.com/library/?style=Nature%20Communications) *[Am Statistician](https://app.readcube.com/library/?style=Nature%20Communications)* **[72](https://app.readcube.com/library/?style=Nature%20Communications)**[, 37–45 \(2018\).](https://app.readcube.com/library/?style=Nature%20Communications)

[3. Weiner, B. J.](https://app.readcube.com/library/?style=Nature%20Communications) *[et al.](https://app.readcube.com/library/?style=Nature%20Communications)* [Psychometric assessment of three newly developed implementation outcome measures.](https://app.readcube.com/library/?style=Nature%20Communications) *[Implement Sci](https://app.readcube.com/library/?style=Nature%20Communications)* **[12](https://app.readcube.com/library/?style=Nature%20Communications)**[,](https://app.readcube.com/library/?style=Nature%20Communications) [108 \(2017\).](https://app.readcube.com/library/?style=Nature%20Communications)

Supplementary Appendix

Appendix 1. Diagnostic classification.

Specialist working diagnoses typically made by on-call neurosurgical registrar, aggregated into diagnostic classes for further analysis. Where a working diagnosis was fitting more than one class, the most likely class was used.

