Predicting Health Risk in Patients with Coronavirus or Influenza using Artificial Intelligence

Prof Dr Harald Braun*, Prof David Patterson[ω], Dr Aoife Molloy[†], Keith Davies[ε]

i5 Analytics[ε], The Whittington Hospital[ω], Imperial College[†], University of Chester[*]

Overview: About 1.03m people are infected with SARS-CoV-2 virus and have developed Covid-19 symptoms at the time of writing (2nd April March 2020) [1]. This study researched the health risk after admission to hospital for patients with medium to severe symptoms based on length of stay in hospital [2]. The AI that was trained in this paper can be used for **health systems** to support prioritisation, to inform hospital treatment and to optimise bed management or for **people** to estimate a level of quarantine based on their medical history. The AI is available online for free on https:\\www.coronavirusrisk.org for individuals accessible with an easy user interface. For processing large population datasets a Webservice endpoint is available from https://www.i5analytics.com/free-healthcare-ai.

Objective: The use of this tool is to provide estimations of peoples' health risks in the event of infection from SARS-CoV-2 and development of Covid-19 and admission to hospital or people without symptoms to predict their vulnerability and estimate their quarantine levels. While quarantine has been used to reduce spread of infection, stringent blanket quarantines are difficult to implement and have sparked violence and tension between states and citizens in the past [3]. This information could be used to support the decision to increase the level of quarantine for people at high risk of infection by identifying risk factors.

Results: The best performing Neural Network had 255 inputs (features), with 244 diagnosis and 11 age band codes. This network was able to re-categorise 98.9% of patients in the training data set into three bands of outcomes: medium, severe and died. The accuracy of re-categorisation of new patients into the three bands was 80.1% with an error rate of 19.9%. Levels of prediction will be improved once new data on SARS-CoV-2 is available [4].

Keywords

Artificial Intelligence (AI); Healthcare; Coronavirus; SARS-CoV-2; Covid-19; Health Risk; Neural Networks (NN); National Health Service (NHS); England; UK; Epidemic; Pandemic; Health Policy; Disease Control; Population Health Management (PHM)

1. Introduction

The aim of this work was to identify clinical characteristics from medical history to predict health risk among patients with acquired flu or Coronavirus [5]. In mid December 2019 a novel Coronavirus (SARS-CoV-2) emerged in Wuhan, China, that has the ability to spread more easily between people than similar strains in the past [6]. By providing a health risk estimate for Covid-19, people at high risk can be advised to take additional measures to protect themselves from infection [7][8].

Two key elements are required for efficient disease control: immunisation and containment [9]. Since there is no immunisation (or even cure) available in March 2020, containment seems to be the only realistic option. Containment can be achieved by individuals protecting themselves when around people carrying SARS-CoV-2 and / or isolating people carrying SARS-CoV-2 [10]. This can significantly reduce the need for healthcare resources to keep healthservices functioning and cut mortality.

Not much is know how SARS-CoV-2 is spreading, it is thought to spread from person to person in close contact (within 2 metres) through respiratory droplets. It may also spread through contact with contaminated surfaces or objects. It is assumed the the virus is able to survive longer once outside the body of a carrier. Droplets carrying cold and flu viruses remain infectious for several hours if they land on non-absorbent materials like stainless steel or plastic whereas Coronavirus survives for much longer [11][12].

Viruses do not replicate outside living cells but infectious virus may persist on contaminated environmental surfaces. The human Coronavirus associated with the common cold (229E or OC43) was reported to remain viable only for 3 hours on environmental surfaces after drying, compared to SARS CoV which survived for up to 28 days [13].

* harald.braun@i5analytics.com; harald.braun@nhs.net
ω d.patterson@ucl.ac.uk
† aoifemolloy@nhs.net
ε keith.davies@i5analytics.com

Through containing the virus at a population level to maintain healthcare capacity, people at risk such as older people, diabetics, and those with cardiovascular diseases require efficient protection from infection. [14]

High temperature combined with high relative humidity promotes inactivation of SARS CoV while lower temperatures and low humidity support prolonged survival of the virus on contaminated surfaces. Studies show that SARS CoV is more stable and can survive at least two weeks after drying at favourable temperature and humidity [15][17].

Studies on air temperature (AT) and relative humidity (RH), showed that SARS CoV persisted for up to 28 days at AT of 22–25◦C and RH of 40–50%. Conversely, the virus viability dropped rapidly at higher AT and higher RH of 38 ◦C and >95% RH. Other studies showed that it can also be inactivated by ultraviolet light, alkaline (pH > 12), or acidic (pH < 3) conditions [17].

Through understanding the risk of infection from the environemnt and the risk of health outcomes based on past medical history, people at high risk from both should increase protection against infection [17]. This study uses past Influenza and human Coronavirus as surrogates whilst awaiting detailed medical records from Covid-19 patients.

This study is based on Hospital Episode Statistics (HES) in England from 2016/17-2018/19. It includes hospitalised patients with medium to severe respiritory symptoms e.g. pneumonia. As no data is available for patients with no or mild symptoms that were not hospitalised [19][20].

2. Background

Since there is limited data in the UK (12th March 2020) on patients with SARS-CoV-2, surrogate viral infections relating to the Influenza virus and human Coronavirus (229E, NL63, OC43, HKU1) have been used to build predictive models.

The use of data from several surrogate viruses has the possibility of developing models to inform people about their potential health risk when contracting SARS-CoV-2.

The following paragraphs describe the data on Influenza and human Coronavirus and compares the two infections. This data was used to build the predictive models for the Covid-19 health risk predictor.

As shown in Table 1, in the last three years,a total of 216,812 people were hospitalised in England due to Influenza, of which 7,852 died of the effects of Influenza (3.63%). At the same time, 4,831 patients were hospitalised due to Human Coronavirus (229E, NL63, OC43, HKU1) of which 138 died (2.86%).

Year	Influenza			Coronavirus		
	People	Died	Mort Rate	People	Died	Mort Rate
2016/17	36,110	1,434	3.97%	1,283	39	3.04%
2017/18	96,612	3,797	3.93%	1,720	49	2.85%
2018/19	84,090	2,621	3.12%	1,828	50	2.74%
Total	**216,812**	**7,852**	**3.62%**	**4,831**	**138**	**2.86%**

Table 1. Hospitalised patients due to Influenza or Coronavirus in England over a period of 3 years**.**

Table 1 shows that the -all ages- mortality rate of Influenza is higher than that of Human Coronavirus (3.62% vs 2.86%) and the number of deaths due to Influenza is 57 times higher. Early research of SARS-CoV-2 show that transmission rates are higher than Influenza and, although mortality rates are relatively low, many more people can be affected. Age band specific mortality rates of SARS-CoV-2 or case-fatality rate (CFR) varies significantly from 8%-15% [21] [22].

Figure 1 below shows the number of patients admitted to hospital for Influenza and human Coronavirus in England between April 2016 and March 2019. It can be noted that both charts correlate strongly (93.8%) and both peak during the winter months. The number of patients admitted for Influenza is, on average, 44.9 times higher than for human Coronavirus which demonstrates that Influenza is much more contagious.

The new forms of Coronavirus SARS CoV, MERS-CoV and SARS-CoV-2 spread much faster than human Coronavirus with an infection rate of R0 of 3.2 people infected per carrier, which is higher than the infection rate of Influenza R0 of 1.80 people. With the assumption that SARS-CoV-2 can be as contagious as Influenza, affected patient numbers can significantly increase in 2020.

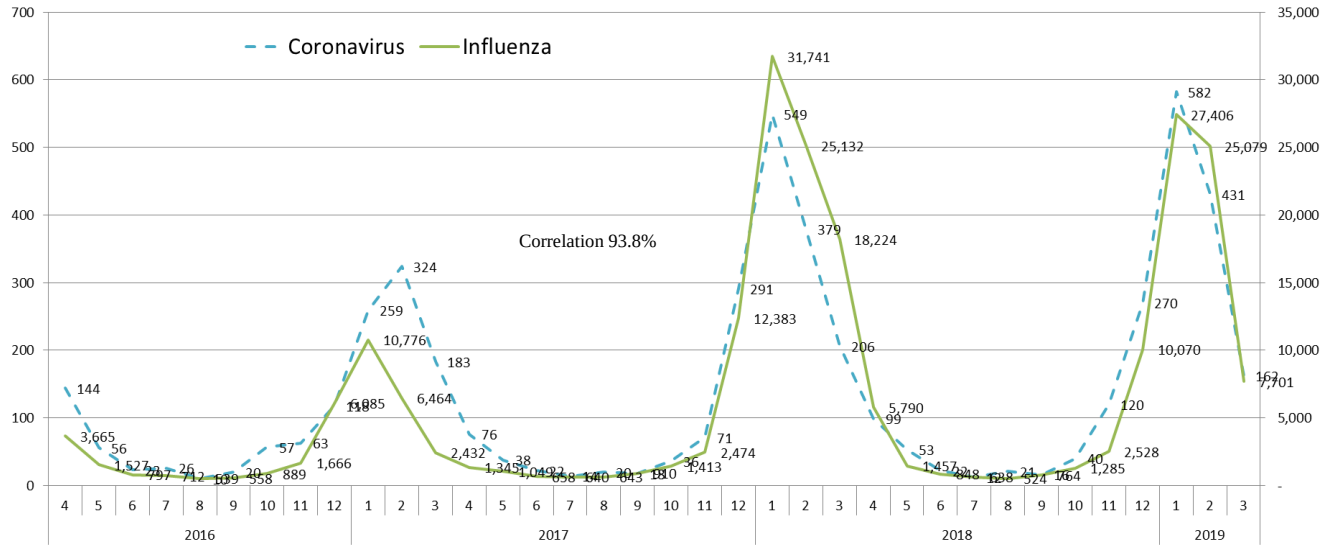

Figure 1. Number of people hospitalised due to Influenza or Coronavirus from April 2016 till March 2019 in England.

Figure 2 shows the number of admitted patients due to Influenza and Coronavirus between Apr 2016 and March 2019 in England. Differences in admission rates can be seen for very young patients 0-9 years of age. The profiles of patients with Influenza and Coronavirus of 10 years and above can be seen to correlate strongly at a rate of 98.3% (excluding children).

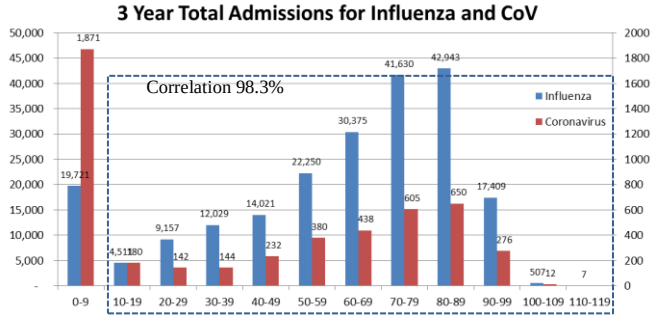

Figure 2. Age bands of people admitted due to Influenza and Coronavirus between 2016/17 and 2018/19 in England**.**

Figure 3 shows patient that died in hospital from Influenza and human Coronavirus between Apr 2016 and Mar 2019 in England. In line with the number of admissions shown in figure 2, young children up to 1 year are disproportionately high (58.3% of the 0-9y bracket). The profiles of patients 10 years and above with Influenza and Coronavirus correlate at a rate of 94.6% (excl. children under 10y).

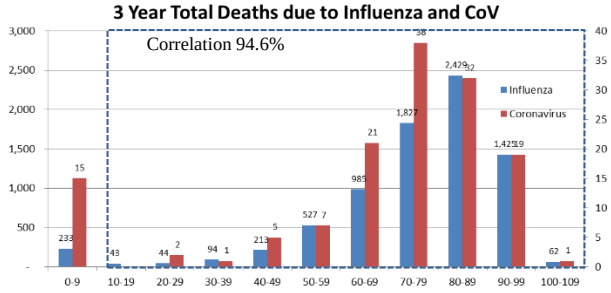

Figure 3. Age bands of people that have died due to Influenza and Coronavirus between 2016/17 and 2018/19 in England.

Figure 4 shows the case fatality rate (CTF) of Influenza and Coronavirus by age band in England between Apr 2016 and Mar 2019 in England.

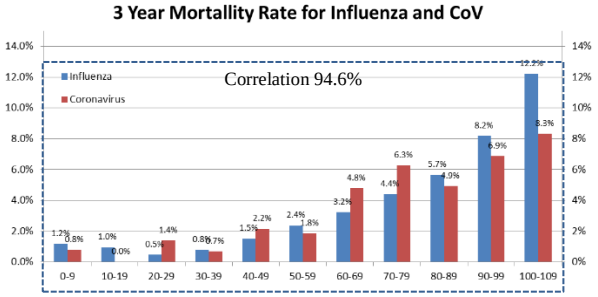

Figure 4. Case fatality rate by age band for patients died due to Influenza and Coronavirus between 2016/17 and 2018/19 in England.

It can be seen that the case fatality rate for human Coronavirus is higher compared to Influenza in the lower age bands 20-29, 40-49, 40-69 and 70-79 with rates of 1.4%, 2.2%, 4.8%, 6.3% respectively. The reverse is true for higher age bands. The Influenza case fatality rate is higher for the ages of 80 years and above (5.7%, 8.2%, 12.2%) when compared to human Coronavirus (4.9%, 6.9%, 8.3%).

On aggregate, the combined case fatality rate in the age bands between 0 and 59 years of age is 1.4%, which is low compared to the mortality rate in the age bands between 60-109 years, as shown in Table 2.

Aggregated Age Bands	Age Band	
	0-59	60-109
Admitted	84,638	134,852
Died	1,184	6,839
Mortality Rate	1.40%	5.07%

Table 2. Case fatality rates between Apr 2016 and Mar 2019 due to Influenza or Coronavirus in England.

Figure 5 shows the percentage of patients that were diagnosed with either Influenza or Coronavirus that died in hospital. It is noticeable that three times more children died of Coronavirus compared to Influenza. It can be seen that patients between 70-79 years of age had the highest number of mortalities, as reflected in figure 5 which shows that 27% of all patients in that age bracket died of Coronavirus. It can also be seen that the case fatality rate of children is particularly high. 15 children between 0-9 years of age died of Coronavirus, which represents 10.6% of all patients that died of Coronavirus.

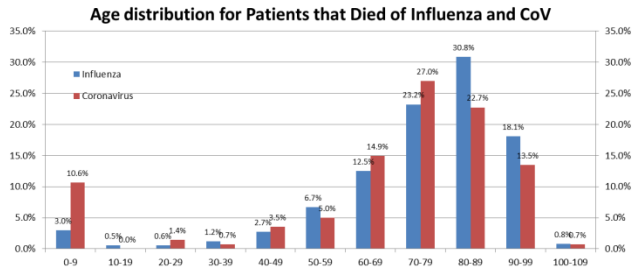

Figure 5. Age distribution of patients that died in hospital due to Influenza and Coronavirus by age band between 2016/17 and 2018/19 in England.

Excluding the outlier for children, the age band distribution for patients that died of Influenza and Coronavirus correlates strongly at 94.6%.

Since SARS-CoV-2 infections follow a clustering pattern that affects older people with comorbidities, characteristics of patients who died were in line with the MuLBSTA score, an early warning model for predicting mortality in viral pneumonia. This re-iterates that pre-existing conditions provide reliable markers for prediction [23].

Figure 6, below, shows the number of co-morbidities diagnosed in patients with Influenza and Coronavirus for all age bands in England between Apr 2016 and Mar 2019. These co-morbidities were present up to three years prior to admission and at the day of admission for their disease. All numbers represent the count of a disease chapter as defined for ICD-10 clinical codes by the WHO [24].

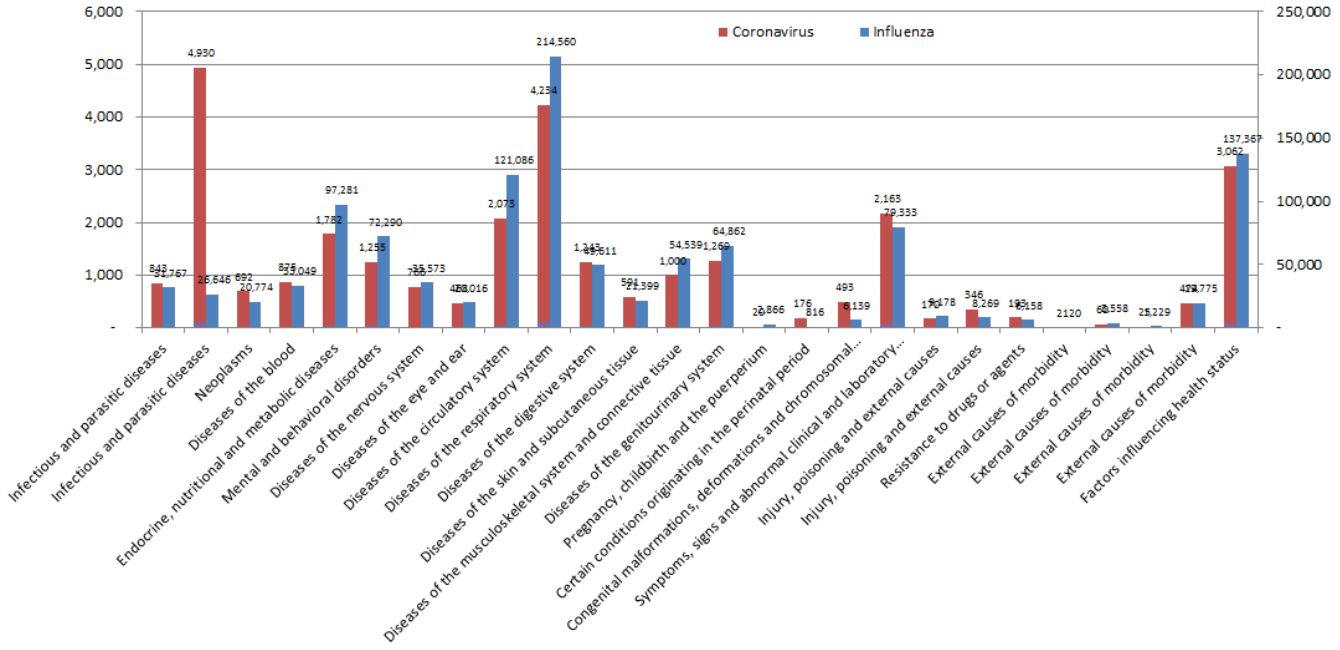

Figure 6. Co-morbidities in all patients with Influenza or Coronavirus from April 2016 till March 2019 in England.

The 219,490 patients in this study have 1,948,307 diagnosed co-morbidities, with an average of 8.9 conditions per patient. To compare the clinical characteristics between patients with Influenza and Coronavirus, the disease profiles grouped by disease chapter are shown in figure 6.

The side by side comparison of co-morbidities identifies two outliers. The first one relates to Infections in the Coronavirus group (Chapter B) since the clinical codes for Coronavirus lie in this group. The second outlier is in Diseases of the Respiratory System (Chapter J) in the Influenza group since the clinical codes for Influenza lie in this group.

For purpose of illustration, Figure 7 shows a sample of medical records of patients that had severe complications or died after hospitalisation due to Influenza or Coronavirus. The presence of conditions identified as contributing factors to such complications are shown as red boxes. The sample shows complex patients with various permutations of conditions which are used to train AI to develop a means of predicting outcomes.

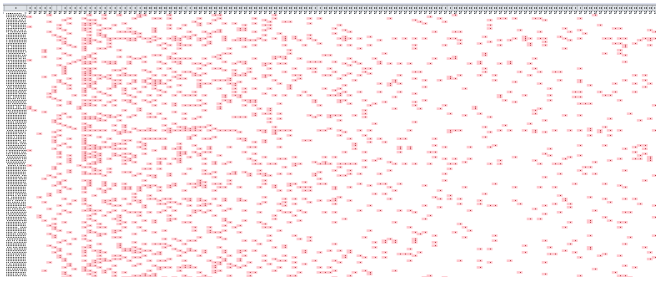

Figure 7. Sample of 150 patients' co-morbidities, shown as red squares for presence of underlying disease.

Figure 8 shows the co-morbidity distribution of patients between the ages of 50-59, which correlates closely with all ages shown in Figure 5. Other age bands have similar distributions as in the all ages diagram in Figure 5. If the chapter relating to Infections is excluded, the patient cohorts with Influenza and Coronavirus correlate closely at 98.2% across all age groups.

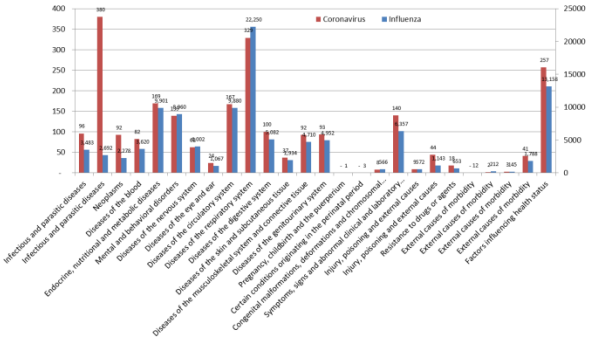

Figure 8. Co-morbidities in patients between 50 and 59 years of age with Influenza or Coronavirus 3 year aggregate in England.

Due to the strong correlation in seasonality (Figure1), incidence numbers (Figure 2 and 3), mortality rate (Figure 4 and 5) and medical history (Figure 6 to 8) between the families of viruses, the medical history of patients with Influenza and Coronavirus have been used to identify pre-existing conditions that contribute to the health risk. Those pre-existing conditions were subsequently used to train an Artificial Intelligence model to predict the health risk for admitted patients taking into account the complex relationships and permutations amongst any relevant pre-existing conditions.

3. AI Development Methodology

To estimate the health risk for patients admitted to hospital due to medium to severe symptoms of Influenza and Coronavirus, past medical history is used. Due to the strong correlation of admission profile, outcomes and medical history, both population profiles were combined to obtain sufficient training and test data.

The methodology used for this paper is based on an existing implementation of Neural Networks for preliminary diagnosis to support targeted screening in the NHS in North West London, UK [27]. It is based on existing information whilst maintaining clinical governance (IG, CG) standards in the UK and is split broadly into five stages.

1) Obtain data that is de-identified and consists of a representative cross-section of patients.
2) Identify the number of pre-existing conditions that are relevant for prediction of health risk.
3) Split the data for AI training, monitoring of overfitting and performance testing.
4) Create a balanced set of datasets to avoid bias.
5) Train various AI models and topologies to evaluate which combination works best for predicting the health risk.
6) For implantation, facilitate re-identification of patients, and inform them about Support Programmes (e.g. self-isolation, reduced social interaction, etc.) in high risk areas.

The steps described above cover all stages and produce a patient cohort for health risk prediction, which from a Public Health or Disease Control perspective, would benefit from behaviour change for the at risk population. Figure 9 shows the process flow from obtaining medical records to support programmes for self-isolation.

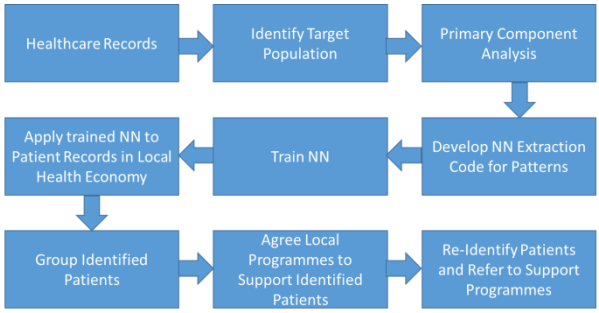

Figure 9. The methodology suggested for implantation of AI based risk management of high risk populations.

4. Data Preparation

For an AI system to be usable within a range of healthcare settings with variations in data quality, completeness and coding standards, the training data should be covering as many relevant pre-existing conditions as possible [26].

To perform prediction of health risk, the Neural Networks (NN) used in the AI system require training on patients with various outcomes. In this study, patients were categorised into three cohorts: medium severity, high severity and patient died. The datasets were balanced across all three categories to obtain optimal separation between Sensitivity and Specificity.

The data used for training of the NNs was based on medical records from hospitals in England which consists of over 200 fields including 21 diagnosis and 21 procedures [25].

Longitudinal healthcare records were used covering three years of patient history and included any diagnosis that was present prior to the diagnosis of Influenza or Coronavirus, as shown in Figure 10 below.

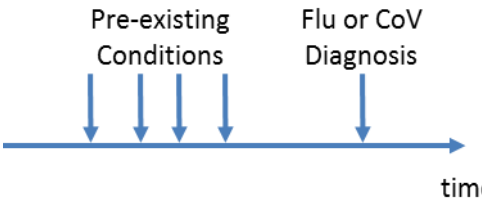

Figure 10. Diagnosis codes prior and at the diagnosis of Influenza or Coronavirus were used.

The patient history diagnosis fields used were coded in ICD-10 standard, which covers over 14,400 conditions. [28].

The dataset for training of the NNs should contain only relevant pre-existing conditions. This has been achieved by performing a Principal Component Analysis (PCA) on the 1,948,307 diagnosis codes contained in the sample population. The objective is to reduce the size of the dataset to only those conditions that are pertinent for the prediction. The design of the PCA is vital to ensure that the conditions chosen are both accurate and relevant [29].

This study included predictive systems based on 255 inputs consisting of previous diagnoses and age bands of patients. The 255 inputs into the first layer of the NN consists of 244 diagnosis and 11 age band nodes. Each input is binary (0 or 1) to indicate if a specific diagnosis and age band is true.

Table 3 below shows the datasets used for training, testing and validation. The training dataset was used to update the weights of the NN during supervised training; testing was used during training to measure the generalisation error and to define the stopping criteria; and validation was used to measure the health risk prediction ability of the NN against a set of patients with known outcomes.

Purpose	Number of Records	Medium Severity	High Severity	Patient Died
Training	50,310	22,005	27,053	3,216
Testing	5,630	2,510	2,983	327
Validation	8,423	3,755	4,463	489
Total	**64,363**	**28,270**	**34,499**	**4,032**

Table 3. Summary of training, testing and validation data used.

Before data can be used for training, it should be balanced to contain admitted patients with various outcomes to ensure that the NN output will not be biased to predict outcomes of a population that is over-represented.

Balancing the input dataset ensures that the network does not unduly favour one outcome over another. In the training data set for this study, 43.7% of patients had medium severity, 53.8% of patients had high severity and 6.4% died. This does not add up to 100% since some patients fell into more than one category, e.g. severe outcomes and patient died.

Figure 11 below shows the three datasets by outcome categorised into three groups. It can be that categories Medium and Severe have similar numbers of cases with less cases available where a patient died.

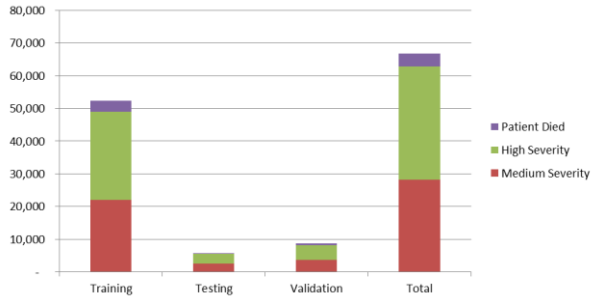

Figure 11. Distribution of patients with and without HF in the datasets used for training, testing and validation.

The design of the input dataset also requires the size of the data file to be defined. A larger dataset will provide improved learning but will also require more time to learn. A balance between learning capability and processing times has been found whereby processing time was limited to 2 hours on a GeForce RTX GPU.

All stages of data preparation, training and validation are shown in Figure 12 and consist of:

1) Training data preparation, including pre-existing conditions prior to Influenza or Coranavirus infection.
2) A database view to create flexibility in database schema naming.
3) Mapping of clinical codes to identify if the patient has a specific condition e.g. hypertension. This mapping enables inclusion of ICD-10, SNOWMED, Read, OPCS or local codes.
4) Training of the NNs with various topologies.
5) Obtaining results of the training set and validation set.

Shown in Figure 13 are the top 20 diagnoses of the 244 that were identified as main contributors to the three categories of outcomes (medium, severe, death). Those 244 diagnoses define the number of input neurons for the Neural Network.

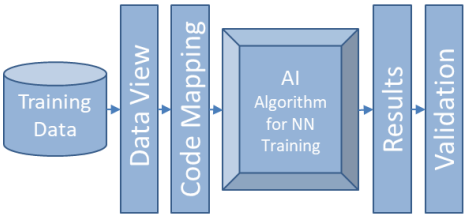

Figure 12. Data flow of source data for NN training and validation.

It can be seen that hypertension, psychoactive substance abuse, past renal failure, diabetes and Asthma are amongst the key drivers for predicting the health risk.

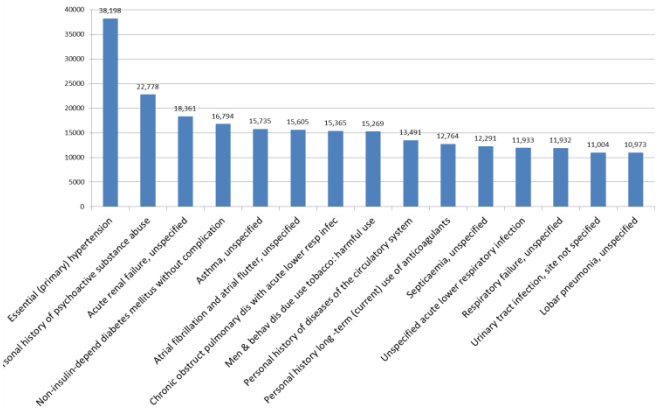

Figure 13. Distribution of top 20 pre-existing conditions, out of 244 identified by PCA, that have an effect on outcomes.

The inputs of the NN correlate to the presence of a diagnosis e.g. if hypertension is present either a 0 or 1 is presented to the input neuron. This binary input to the Neural Network allows mapping from various other coding schemes to be added later. Such schemes include SNOMED, Read or local codes used by a hospital.

Supervised artificial Neural Networks require a fixed length vector representation of the input and the output. Hence the input dataset is created containing a record for each patient with a 1/0 (true/false) value for each input field that was identified as significant during the PCA and three binary output fields for medium severity <= 4 days length of stay, high severity >4 days length of stay, and patient died.

Because quarantine was not applied to those patients in the dataset, the length of stay can be used as a proxy measure for outcomes. This is no longer the case if patients are quarantined for 2 weeks or longer and other means for measuring outcomes are required for future datasets.

The benefits of using binary representation are the simplification of the data preparation, such as avoidance of normalisation, conversion of diagnostic or treatment codes into numerical values and the use of an AI classifier [30].

5. Neural Network Configuration

Neural Networks (or NNs) provide a powerful technique for implementation of Artificial Intelligence (AI) for modelling nonlinear relationships. Neural Networks are, therefore, commonly applied in non-linear prediction to analyse the complex relationships that exist between patient medical history and health risk.

There are many different AI systems that can be used for the creation of an AI classifier to predict the health risk. The chosen topology was a supervised, multilayer network with binary input with 244 neurons based on pre-existing conditions and 11 age bands with sigmoid activation functions for all neurons and standard backpropagation learning algorithm. Figure 14 is an illustration of a multi-layer Neural Network [31].

The training data base of patient medical records has been a set of patients with various outcomes. They were chosen from a large set of patients across England and span across age groups, areas of deprivations, ethnic groups, case-mix of pre-existing conditions, etc. in order to avoid bias.

The number of permutations across patients, medical conditions and age bands drives the number of neurons required in the hidden layer. The more complex the permutations, the more hidden neurons are needed.

If the number of hidden neurons is too large, the network will memorise each permutation and will not be able to make a prediction for patients it has not seen before. If the number of hidden neurons is too small, it will not be able to re-identify the outcomes patients had that were part of the training set - resulting in poor recognition performance [32].

Several Neural Network topologies were tested with various numbers of hidden neuron. The chosen topologies included in this paper were based on representing sufficiently distinctive results and were all based on 255 inputs.

Each network topology was trained with three variations in hidden neurons resulting in 255-50-3, 255-80-3 and 255-100-3 topologies - see Table 4 below.

	Topology 1	Topology 2	Topology 3
Input	255	255	255
Hidden	50	80	100
Output	3	3	3

Table 4. Neural Network topologies.

After initialisation of the Neural Network weights with small numbers and Gaussian distribution, the Neural Networks were trained. During training, the generalisation error was measured and training was stopped once minimum was found.

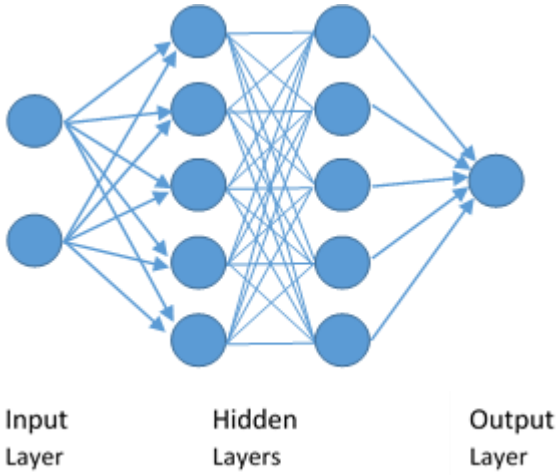

Figure 14. An illustration of a multi-layer Neural Network used for predicting health risk in patients with Influenza or Coronavirus.

6. Network Training

During training, the learning rate, momentum and other Neural Network parameters were set to very low numbers initially which were self-adjusting in order to yield the best training results in the shortest time. Table 5 below shows the initialisation and learning parameter settings at the beginning of training.

The initialisation parameters were set between 0.007 and 0.003 with a Gaussian distribution.

	Neural Network Topology		
	255-50-3	255-80-3	255-100-3
Initialisation	0.007	0.005	0.003
Learning Rate	0.02	0.015	0.012
Momentum	0.01	0.008	0.006

Table 5. Initial initialisation and training parameters.

During training, an adjustment algorithm was used to prevent weights becoming too large and developing over-sensitive inputs. This avoidance of large weight counters the effects of overtraining and over-weighted groups of input patterns [33].

After starting training, a stopping criteria has been defined based on crossover validation and minimum generalisation error from a set of 5,630 unseen patterns (testing data) which are assessed during training. Figure 15 below shows the Sum Square Error (error on training data) and Generalisation Error (error on the testing data). A snapshot of the weight matrix was stored when the generalisation error reached its minimum e.g. 9,843 iterations for the 255-80-3 NN topology.

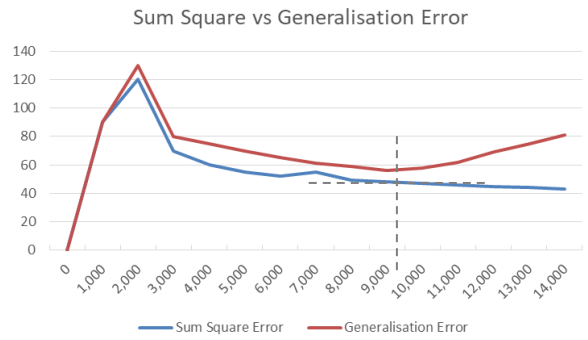

Figure 14. Sum Square and Generalisation error during training of the 255-80-3 Neural Network stopped after 9,843 iterations.

After training, Sensitivity, Specificity and total mis-classification errors were measured for various classification thresholds. The classification thresholds selected were the minimum number of errors observed. Table 6 below shows a summary of the various topologies and training results that were included in this study.

To obtain the classification results for each of the three outputs (medium, severe, died), each Neural Network output had a separate threshold that is set for optimum results. For example, the thresholds used for the 255-80-3 NN with 9,843 iterations were 0.50 for 1/0 of medium, 0.38 for severe and 0.25 for died. This Neural Network has been the most accurate of the set.

		Training Data			Testing Data		
Topology	Iterations	Sensitivity	Specificity	Thresholds	Sensitivity	Specificity	Thresholds
255-50-1	6,943	95.6%	91.16%	0.50, 0.37, 0.25	72.67%	70.99%	0.50, 0.37, 0.25
	7,839	95.7%	90.84%	0.50, 0.41, 0.25	71.04%	71.73%	0.50, 0.41, 0.25
	8,937	96.2%	92.34%	0.50, 0.39, 0.25	73.89%	77.20%	0.50, 0.39, 0.25
255-80-3	9,843	98.9%	95.80%	0.50, 0.38, 0.25	80.09%	78.20%	0.50, 0.38, 0.25
	9,345	97.4%	92.61%	0.50, 0.36, 0.31	76.31%	78.56%	0.50, 0.36, 0.31
	8,782	97.4%	95.14%	0.50, 0.39, 0.27	70.36%	78.15%	0.50, 0.39, 0.27
255-100-3	8,923	98.1%	91.09%	0.50, 0.33, 0.28	74.52%	71.61%	0.50, 0.33, 0.28
	9,530	97.8%	91.16%	0.50, 0.38, 0.24	76.11%	78.65%	0.50, 0.38, 0.24
	12,563	98.1%	93.42%	0.50, 0.36, 0.29	76.01%	78.57%	0.50, 0.36, 0.29

Table 6. Results for various network topologies and stopping criteria.

7. Results

Artificial Neural Networks provide a wide range of potential applications in predicting development of conditions, complications and health risk. In this paper, we have presented the use of AI to predict the health risk for patients infected with Influenza or human Coronavirus using data from April 2016 to March 2019.

The purpose of predicting the health risk after infection with the Coronavirus or Influenza is to provide health economies with a way of prioritising admitted patients and to enable people to check their risks in order to inform the level of quarantine required.

No prediction system is error free, so users should avoid exposure to pathogens under any circumstances even if their health risk is low. The 255-80-3 neural network trained with 9,843 iterations has a sensitivity of 80.1% across all three categories resulting in a misclassification of 19.9%.

The PCA identified 244 pre-existing conditions that provided sufficient accuracy for predicting the health risk. Because the health risk for patients without pre-existing conditions cannot be predicted by AI, this does not mean it is zero and that quarantine would avoid infection.

Patients with medium to high risk should consider being tested if early Influenza or human Coronavirus symptoms are detected to increase chances of treatment. Since human Coronavirus and SARS-CoV-19 are in the same family, causing similar symptoms, it can be assumed that these risk categories also apply to Covid-19, which will be confirmed once data for SARS-Cov-19 becomes available.

The AI system described in this study is available online for free on https:\\www.coronavirusrisk.org for individuals accessible with an easy to use user interface. For processing large population datasets a Webservice endpoint is available from https://www.i5analytics.com/free-healthcare-ai.

Individuals, as well as Public Health authorities, Governments and Clinicians, can use this prediction tool to assess the morbidity of people, with or without Covid-19 for:

- Individuals without SARS-CoV-2, where the health risk helps in deciding the level of quarantine needed to

adjust their lifestyle and minimise the risk of infection.

- Clinicians and Hospital Management, where the health risk for admitted patients with Covid-19 can help prioritising care and improve bed management.
- Public Health authorities and Governments, where the health risk of populations within a geography supports planning for disease control and levels of quarantine.
- Infection Control Units, where the health risk helps classification of people into risk bands that were in touch with an infected individual to facilitate disease control.
- Telephone Assessment or Remote Consultations, where support staff or clinicians enter the patients' information and advice on risk.

By estimating the health risk for Coronavirus, people at higher risk can take extra precautions so that pressures on hospitals and clinicians can be lessened and to reduce the overall mortality rate.

8. References

[1] Coronavirus Covid-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU) https://www.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf6

[2] Stylianou N, Fackrell R, Vasilakis C; Are medical outliers associated with worse patient outcomes?; (2017) https://doi.org/10.1136/bmjopen-2016-015676

[3] Disease Control Priorities: Improving Health and Reducing Poverty. 3rd edition; (2017) https://doi.org/10.1596/978-1-4648-0527-1

[4] Tognotti E. Lessons from the history of quarantine, from plague to influenza A. Emerg Infect Dis; (2013) https://doi.org/doi:10.3201/eid1902.120312

[5] Guo L, Wei D, Zhang X, et al. Clinical Features Predicting Mortality Risk in Patients With Viral Pneumonia; (2019) https://doi.org/10.3389/fmicb.2019.02752

[6] World Health Organisation; Coronavirus disease (Covid-19) outbreak; (2020) https://www.who.int/emergencies/diseases/novel-coronavirus-2019

[7] Chung, J., Hsu, C., Chen, J. et al. Geriatric influenza death (GID) score; (2018). https://doi.org/10.1038/s41598-018-27694-6

[8] Shu Jing Shi Hui Li et al. Mortality prediction to hospitalized patients with influenza pneumonia (2015) https://doi.org/10.1111/crj.12346

[9] Ethical and Legal Considerations in Mitigating Pandemic Disease (2007); doi: 10.17226/11917 https://www.ncbi.nlm.nih.gov/books/NBK54163/

[10] Preventing the Spread of Coronavirus Disease 2019 in Homes and Residential Communities (2020); https://www.cdc.gov/coronavirus/SARS-CoV-2/hcp/guidance-prevent-spread.html

[11] Chan KH, Peiris JS, Lam SY, Poon LL, Yuen KY, Seto WH. The Effects of Temperature and Relative Humidity on the Viability of the SARS Coronavirus. (2011) https://doi.org/10.1155/2011/734690

[12] Casanova LM, Jeon S, Rutala WA, Weber DJ, Sobsey MD. Effects of air temperature and relative humidity on coronavirus survival on surfaces (2010) https://doi.org/10.1128/AEM.02291-09

[13] Prussin AJ 2nd, Schwake DO, et al. Survival of the Enveloped Virus Phi6 in Droplets as a Function of Relative Humidity, Absolute Humidity, and Temperature; (2018) https://doi.org/10.1128/AEM.00551-18

[14] Deng SQ1, Peng HJ1. Characteristics of and Public Health Responses to the Coronavirus Disease 2019 Outbreak in China. (2020) https://doi.org/10.3390/jcm9020575.

[15] Charleen Yeo, Sanghvi Kaushal, Danson Yeo; Enteric involvement of coronaviruses: is faecal–oral transmission of SARS-CoV-2 possible? (2020) https://www.thelancet.com/journals/langas/article/PIIS2468-1253(20)30048-0/fulltext

[16] Geller C, Varbanov M, Duval RE. Human coronaviruses: insights into environmental resistance and its influence on the development of new antiseptic strategies. (2012) https://doi.org/1doi:10.3390/v4113044

[17] Zou S, Guo J, Gao R, et al. Inactivation of the novel avian influenza A (H7N9) virus under physical conditions or chemical agents treatment. (2013) https://doi.org/10.1186/1743-422X-10-289

[18] Collins AS. Preventing Health Care–Associated Infections. Agency for Healthcare Research and Quality; (2008) https://www.ncbi.nlm.nih.gov/books/NBK2683/

[19] Gaio V, Nunes B, Pechirra P, et al. Hospitalization Risk Due to Respiratory Illness (H1N1) (2016) https://doi.org/10.1371/journal.pone.0158181

[20] Delgado-Sanz C, Mazagatos-Ateca C, et al. Illness

Severity in Hospitalized Influenza Patients by Virus Type and Subtype, (2020)
https://doi.org/10.3201/eid2602.181732

[21] World Health Organisation; Coronavirus disease (Covid-2019) situation reports
https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports

[22] del Rio C, Malani PN. Covid-19—New Insights on a Rapidly Changing Epidemic. February 28 (2020).
https://doi.org/10.1001/jama.2020.3072

[23] Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China; (2020)
https://www.thelancet.com/journals/lancet/article/PIIS0140-6736%2820%2930211-7/fulltext

[24] World Health Organisation; International Statistical Classification of Diseases and Related Health Problems (ICD)
https://apps.who.int/classifications/apps/icd/icd10online2003/fr-icd.htm

[25] Braun, Harald & Mclean, Jeanetta & Davies, Keith; Artificial Neural Networks for Population Health Management to support Atrial Fibrillation Screening Programmes. (2017).
https://tinyurl.com/y9qjhg55

[26] Khan, N. F., Harrison, S. E., & Rose, P. W. (2010). Validity of diagnostic coding within the General Practice Research Database: a systematic review. The British Journal of General Practice, 60(572), e128–e136.
https://doi.org/10.3399/bjgp10X483562.

[27] Herbert, A., Wijlaars, L., Zylbersztejn, A., Cromwell, D., & Hardelid, P. (2017). Data Resource Profile: Hospital Episode Statistics Admitted Patient Care (HES APC). International Journal of Epidemiology, 46(4), 1093–1093i.
https://doi.org/10.1093/ije/dyx015.

[28] NHS Digital; National data opt-out programme;
https://digital.nhs.uk/services/national-data-opt-out-programme

[29] Zhang, Z., & Castelló, A. (2017). Principal components analysis in clinical studies. Annals of Translational Medicine, 5(17), 351.
https://doi.org/10.21037/atm.2017.07.12

[30] Wikipedia; Binary classification
https://en.wikipedia.org/wiki/Binary_classification

[31] Hermundstad, A. M., Brown, K. S., Bassett, D. S., & Carlson, J. M. (2011). Learning, Memory, and the Role of Neural Network Architecture. PLoS Computational Biology, 7(6), e1002063.
https://doi.org/10.1371/journal.pcbi.1002063.

[32] Zhang, Z. (2016). Neural networks: further insights into error function, generalized weights and others. Annals of Translational Medicine, 4(16), 300.
https://doi.org/10.21037/atm.2016.05.37.

[33] J. Horn, O. De Jesus and M. T. Hagan, "Spurious Valleys in the Error Surface of Recurrent Networks—Analysis and Avoidance," in IEEE Transactions on Neural Networks, vol. 20, no. 4, pp. 686-700, April 2009.
https://ieeexplore.ieee.org/abstract/document/4798196