

# An AI based digital-twin for prioritising pneumonia patient treatment

Neeraj Kavan Chakshu  and Perumal Nithiarasu

Proc IMechE Part H:  
J Engineering in Medicine  
2022, Vol. 236(11) 1662–1674

© IMechE 2022



Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/09544119221123431

journals.sagepub.com/home/pih



## Abstract

A digital-twin based three-tiered system is proposed to prioritise patients for urgent intensive care and ventilator support. The deep learning methods are used to build patient-specific digital-twins to identify and prioritise critical cases amongst severe pneumonia patients. The three-tiered strategy is proposed to generate severity indices to: (1) identify urgent cases, (2) assign critical care and mechanical ventilation, and (3) discontinue mechanical ventilation and critical care at the optimal time. The severity indices calculated in the present study are the probability of death and the probability of requiring mechanical ventilation. These enable the generation of patient prioritisation lists and facilitates the smooth flow of patients in and out of Intensive Therapy Units (ITUs). The proposed digital-twin is built on pre-trained deep learning models using data from more than 1895 pneumonia patients. The severity indices calculated in the present study are assessed using the standard benchmark of Area Under Receiving Operating Characteristic Curve (AUROC). The results indicate that the ITU and mechanical ventilation can be prioritised correctly to an AUROC value as high as 0.89. This model may be employed in its current form to COVID-19 patients, but transfer learning with COVID-19 patient data will improve the predictions. The digital-twin model developed and tested is available via accompanying Supplemental material.

## Keywords

COVID-19, pneumonia, digital-twin, artificial intelligence, ITU

Date received: 12 July 2021; accepted: 16 August 2022

## Introduction

The COVID-19 pandemic<sup>1</sup> has put an unprecedented stress on the already strained healthcare infrastructure.<sup>2,3</sup> This situation has forced the healthcare providers to prioritise patients in critical need to access Intensive Therapy Units (ITU) and mechanical ventilation. Some of the currently used scoring systems for patient prioritisation include SOFA (Sepsis-related Organ Failure Assessment),<sup>4</sup> APACHE (Acute Physiology and Chronic Health Evaluation),<sup>5</sup> and SAPS II (simplified acute physiology score).<sup>6</sup> These systems (see Table 1) have been validated over time for analysis of intensive care treatments.<sup>7–10</sup> In recent past, several works have been published on severity scoring using neural networks and other machine learning algorithms.<sup>11–13</sup> Majority of these algorithms were trained on large ITU datasets to calculate severity scores, covering a wide range of diseases and medical conditions. Such systems, though valuable during normal times, may not be sufficiently specific to address the current pandemic. In the case of COVID-19 (and in other similar forms of influenza), more precise and

dynamically evolving system may be necessary to address the sudden increase in severity and the need for mechanical ventilation. With more mutations of the virus being identified over time,<sup>14</sup> an evolving knowledge of the disease severity of each virus variant has become extremely important. A robust and dynamically adaptable model that takes into account the progression of severity over the course of care, which may be different for different variants of the disease, is therefore urgently needed.

A human digital-twin is one such model, widely used in other areas,<sup>23–25</sup> which is a digital replica of a human system or sub-system. This replica is a personalised digital representation, in terms of structure or

Biomedical Engineering Group, Zienkiewicz Centre for Computational Engineering, Faculty of Science and Engineering, Swansea University, Swansea, UK

### Corresponding author:

Perumal Nithiarasu, Biomedical Engineering Group, Zienkiewicz Centre for Computational Engineering, Faculty of Science and Engineering, Swansea University, Bay Campus, Swansea SA1 8EN, UK.

Email: p.nithiarasu@swansea.ac.uk

functioning or both, of an individual or patient's system. It can provide real-time feedback on how a patient's health is likely to vary based on their current known condition using periodic input data from the patient's vitals (such as heart rate, respiration rate). The severity scores calculated by these digital replicas (models), personalised to every patient using the individual's vitals and other readings, can form the basis for prioritising potential pneumonia patients for ITU and mechanical ventilation.

The three-tiered strategy of identify, assign and discontinue, as shown in Figure 1, is employed in the present work to produce a patient priority list. The three-tiers are: (1) identify urgent cases amongst those awaiting care, (2) assign mechanical ventilation to critical cases amongst patients receiving care, and (3) discontinue mechanical ventilations and other care at the optimal time, thereby freeing-up vital resources. In the present work, each of these steps are enabled by artificial neural network models, a type of artificial intelligence system, that are well established for complex diagnosis with unprecedented levels of accuracy.<sup>26-29</sup> These methods require significant amount of data for training and testing. However, acquiring, de-identifying and indexing huge amounts of COVID-19 patient data from ITU is currently challenging. Hence, to expedite progress, a transfer learning<sup>30</sup> approach is adopted. In such approaches, artificial neural networks are trained using large amounts of data from similar backgrounds to that of COVID-19. The trained model can be bolstered, known as transfer learning, with smaller amounts of COVID-19 data, when available, to improve accuracy. Since the background data used here is from pneumonia patients, the model proposed should be representative of severe COVID-19 patients. All the artificial neural networks used in the three tiers of Figure 1 together forms an individual patient's digital-twin system.

## Methodology

An interconnected system, comprising multiple independent neural networks, designed to assist a concerted decision making process, is proposed. Independent severity indices are calculated at the three tiers shown in Figure 1 to assist in identifying cases that are severe but with a high chance of survivability. The indices used in each of the tiers are as follows (see Figure 1):

- (i) Tier 1, Patients awaiting intensive care – (a) Difference between probabilities of death with and without mechanical ventilation support and (b) Probability of requiring mechanical ventilation.
- (ii) Tier 2, Intensive care inpatients awaiting mechanical ventilation support – (a) Probability of death (based on data from continuous monitoring of vitals), (b) Probability of requiring mechanical

**Table 1.** A brief list of proposed or existing severity scoring methods and their examples.

Severity scoring method	Name
Aggregate of weights subjectively assigned to variables.	APACHE <sup>15</sup> APACHE II <sup>16</sup> SAPS <sup>17</sup>
Aggregate of weights assigned to variables using statistically pre-defined value range.	APACHE IV, <sup>5</sup> SAPS II <sup>6</sup> MPM II (Mortality probability models) <sup>18</sup>
Neural networks	Various unnamed models <sup>19-21</sup>
Super learner algorithm	SICULA (Super ICU Learner algorithm) <sup>22</sup>

ventilation support (based on data from continuous monitoring of vitals), and (c) Severity indices used in Tier 1.

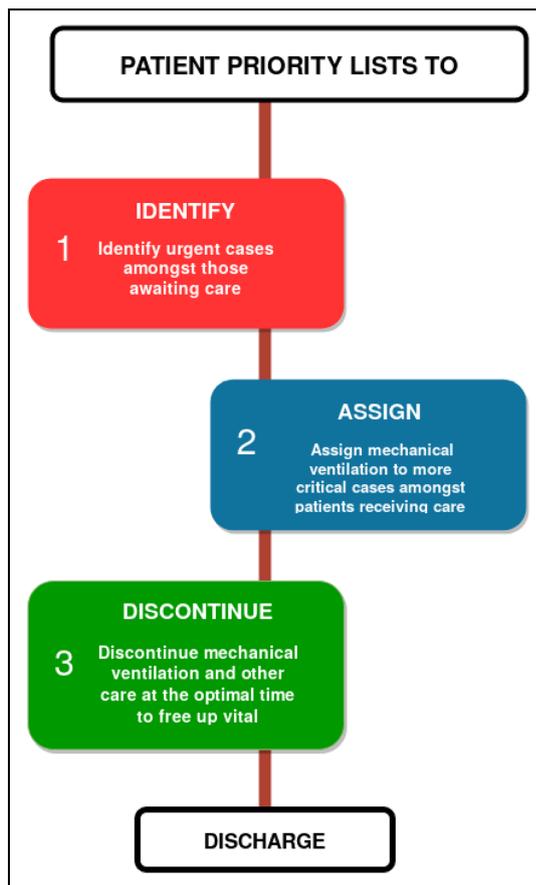
- (iii) Tier 3, Discontinuation of mechanical ventilation support amongst inpatients – (a) Probability of requiring mechanical ventilation support (based on data from continuous monitoring of vitals) and (b) Probability of death (with consideration of mechanical ventilation data).

Within each tier, based on probabilities calculated, thresholds can be set by the ITU professionals for decision making. Independent severity score thresholds can be set within each tier to allow situations in which different stages of ITU care may be needed at different healthcare units or settings (Example – A patient moving from one hospital to another).

## Data selection and pre-processing

Primary sources of data used in this work are from MIMIC-III<sup>22,31</sup> and eICU Collaborative Research Database v2.0,<sup>32,33</sup> obtained from PhysioNet.<sup>34</sup> MIMIC III is a publicly-available database comprising de-identified health-related data associated with approximately 60,000 admissions of patients who stayed in critical care units of the Beth Israel Deaconess Medical Centre between 2001 and 2012. eICU Collaborative Research Database v2.0 is multi-centre database comprising de-identified health data associated with over 200,000 admissions to ITUs across the United States between 2014 and 2015 (see Table 2). Both of these databases consist of various vital sign measurements, de-identified information on patient stay, diagnosis information, records of medical procedures carried out, drugs administered, and various other intensive care information.

Since severe COVID-19 patients suffer from pneumonia,<sup>35</sup> the selection and aggregation of data here is based on critical symptom of pneumonia. Therefore, various subsets of MIMIC-III and eICU databases with



**Figure 1.** Three tiered patient prioritisation strategy to reduce mortality rate in intensive care units.

**Table 2.** Baseline characteristics of subset selected from eICU Collaborative Research Database v2.0.

Overall population, $n = 1895$	
Age	68 [56–79]
Gender (Male)	994
On mechanical ventilation	1155 (60.94%)
HR (bpm)	112 [98–128]
MAP(mmHg)	62 [51–123]
RR (cpm)	33 [25–41]
Na (mmol/l)	138 [133–141]
Glucose (mmol/l)	146 [96–201]
WBC ( $10^3/\text{mm}^3$ )	9.7 [5.33–15.06]
$\text{PaO}_2/\text{FiO}_2$	1.66 [1.12–2.37]
Bilirubin (mg/dl)	0.6 [0.4–0.9]
Dead	230 (12.13%)

pneumonia are chosen for training. All patient data with missing vital information are ignored. The subsets chosen are:

- (a) Pneumonia Mortality subset – Patients who suffered from any form of pneumonia. This subset included both survived and died cases at the time of discharge.

- (b) Pneumonia Ventilator subset – This consists of all survived patients at the time of discharge, but suffered from some form of pneumonia.

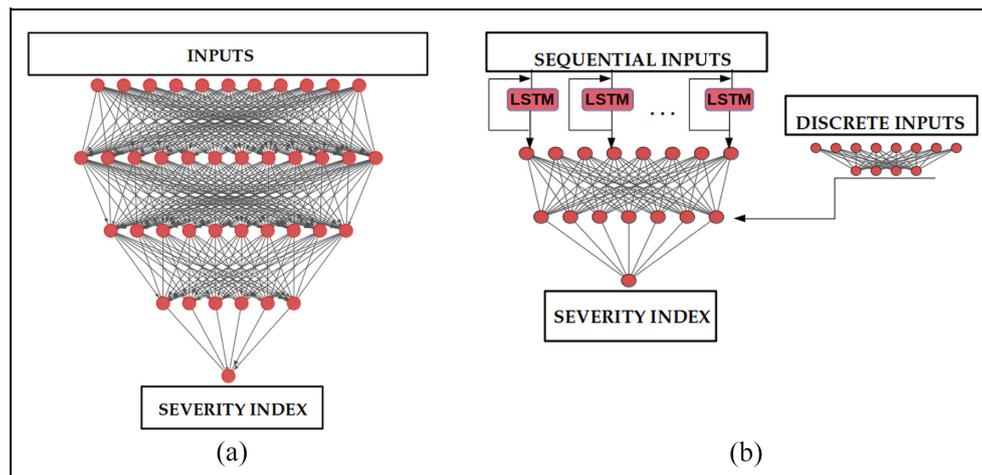
In the case of MIMIC-III database, a total of 493 ITU stays of pneumonia patients were considered. Patient data with available matching and admissible quality waveform data (collected from MIMIC-III matched waveform subset) were considered. Out of which, 211 (42.79%) stays required mechanical ventilation and 95 (19.26%) stays resulted in death.

### Neural networks and architecture

The proposed three-tiered system is constructed primarily using multiple independent neural networks. However, the type of neural networks used can be classified into two categories, Multilayer Perceptron (MLP) and Recurrent Neural Networks (RNN). Artificial neural networks may be used to perform functions such as classification and prediction. MLP models use a series of cascaded non-linear transformations of weighted coefficients, as shown in Figure 2(a), to perform these functions. The MLPs typically accept discrete values as inputs. In the case of patient data, some examples of discrete inputs are patient gender, blood glucose level, and sodium levels. These input values can be used to calculate the probability of death and the probability of a patient requiring mechanical ventilation.

An RNN, a type of neural network, is capable of handling time series or other sequential input and/or output data. In ITU, waveforms (time series) of body vitals measured are examples of sequential input. A special type of RNN is a Long Short-Term Memory (LSTM) cell.<sup>36–38</sup> In order to predict or classify patients, LSTM cells are designed to ‘retain’ and/or ‘forget’ parts of input data sequence(s). These abilities make them the appropriate choice of neural networks to calculate severity indices and assess patient criticality continuously or at regular time intervals. The RNN based models used in the present work use a combination of sequential and discrete value inputs. As seen in Figure 2(b), a combination of LSTM cell layer and additional MLP model is used to predict severity index of interest using fully connected neural layers. All models were built using Tensorflow library,<sup>39</sup> using Keras<sup>40</sup> library, on Python.

The MLP Models consisted of three dense hidden layers. These layers are non-linearly activated using either Rectified Linear Unit (ReLU) functions. Depending on the severity index to be predicted, the input layer consisted of 21–22 input parameters. As the final value being predicted in these models is a single value between 0 and 1, an output layer with single neuron and sigmoid activation is used. Detailed architecture of MLP models and training parameters are provided in Supplemental Appendix A.



**Figure 2.** Multilayer Perceptron (MLP) and Recurrent Neural Networks (RNN) for calculating severity indices: (a) MLP and (b) RNN.

In the case of RNN based models, sequential input data was processed by a network model with single or double LSTM layer(s). These layers use input arrays with eight input parameters. Discrete inputs are processed by an independent network model with a 21–26 input parameters, depending on the severity index. The outputs of these networks are then merged and fed to final model activated with ‘ReLU’ functions. Similar to the MLP models, the final model has an output neuron with sigmoid activation. Dropouts and activity regularisers, were used in these models to avoid over-fitting. A detailed description of RNN based model architectures and training parameters have been provided in the Supplemental Appendix B.

The discrete input data was standardised within a similar range. All MLP models were trained and cross-validated on data subsets chosen from eICU Collaborative Research Database v2.0. All RNN based models were trained on the MIMIC-III database, owing to the waveform datasets available in them. Datasets were split into training (80%) and testing dataset (20%). Cross-validation of models were carried out on the training to tune the hyper parameters and analyse model performance using K-fold method, for which the training dataset was divided into 10-folds.

### **Tier 1-“Identify”:** Identification of ill patients amongst those awaiting intensive care

Since this class of patients are not monitored continuously, data obtained within 24 h of hospital admission is used. A subsystem of two MLP models were trained on a subset containing pneumonia patient data from eICU Collaborative Research Database v2.0 database. One was used to predict the probability of death, and the other for calculating the probability of requiring mechanical ventilation.

The first model was trained on ‘Pneumonia mortality subset’, consisting of 1895 patients and the second model was trained on ‘Pneumonia ventilator subset’, consisting of 1665 patients. The 20 input fields chosen for the first MLP are age, gender, ventilation status (a binary value), intubation status (a binary value), dialysis status (a binary value), heart rate, respiration rate, Glasgow coma scale, white blood cells (WBC) count, blood glucose levels, partial pressure of oxygen in arterial blood ( $PaO_2$ ), partial pressure of carbon dioxide in arterial blood ( $PaCO_2$ ), mean blood pressure, body temperature, sodium, potassium, bicarbonate, bilirubin, concentration of oxygen in the inhaled gas mixture ( $FiO_2$ ), chronic diseases (Metastatic cancer, AIDS or Haematologic malignancy), and Blood Urea Nitrogen (BUN) levels. These inputs are also used by APACHE IV scale. The second MLP for predicting probability of requiring mechanical ventilation used the same input parameters as the first except for ventilation status, which is now added as additional parameter. To remind the reader, the data set used for this model consists of only patients who survived at the time of discharge. Thus, the training criteria should include mechanical ventilation support data to successfully discharge the patient alive.

Using the predicted severity indices, a patient priority list may be generated to admit patients with more urgent need for care into ITU. Since mechanical ventilation is the indication of severity in COVID-19 patients, probability of requiring mechanical ventilation is the primary severity index in preparing such a list. The patients with similar probabilities of requiring ventilation are prioritised using difference between probabilities of death, with and without ventilation. The probability of death with mechanical ventilation may also be used to further refine the prioritisation list. The prioritised patients in Tier1 may be admitted to

the ITU and the prioritisation list may be further refined in Tier 2 for providing mechanical ventilation.

### **Tier 2-“Assign”:** Identifying and providing mechanical ventilation support amongst intensive care patients

The patients selected from Tier 1 to go into ITU can now be continuously monitored. However, with limited number of mechanical ventilators it becomes a necessity to provide this support to the more urgent cases. To identify urgent cases amongst the inpatients in an ITU, a continuous monitoring based severity index is required. The primary severity indices used here are probability of requiring mechanical ventilation support (based on continuous monitoring) and probability of death (based on continuous monitoring). The RNN based model shown in Figure 2(b) is used to predict these severity indices. The sequential (continuous) inputs used here are heart rate, pulse, systolic and diastolic blood pressures, respiration rate, and spot oxygen saturation ( $SpO_2$ ), in addition to multiple discrete input fields,<sup>41</sup> AADO<sub>2</sub> (alveolar-arterial difference of oxygen partial pressure), bicarbonate (HCO<sub>3</sub>), carboxyhaemoglobin, chloride, calcium, base excess, glucose, haematocrit, haemoglobin, sodium, potassium, oxygen saturation in arterial blood (SO<sub>2</sub>),  $FiO_2$ ,  $PaO_2$ , methaemoglobin, temperature, age, and gender. The lowest measured discrete input value in a 24 h period is used in the training. A direct allocation of ventilation support at the first tier is also possible when required.

### **Tier 3-“Discontinue”:** Identifying patients who can discontinue mechanical ventilation support

The patients identified in Tier 2 to receive mechanical ventilation can now undergo continuous monitoring to decide when to discontinue treatment. The probability of requiring mechanical ventilation support (based on continuous monitoring of patient) and probability of death (considering continuous ventilation support) are used as the indices to safely remove the mechanical ventilation support. The RNN used in Tier 2 with similar inputs is continued here to observe the probability of still requiring mechanical ventilation support. The patients with probability of requiring mechanical ventilation lower than a set threshold may be removed from the ventilator support. To reduce the risk of false negatives, probability of death is also monitored in this tier. An additional RNN based model is trained to take ventilator settings along with the inputs used in Tier 2 to predict probability of death. The additional input fields include positive end-expiratory pressure (PEEP), required  $O_2$ , tidal volume and ventilation rate. The person with probability of requiring mechanical ventilation less than the set limit must also have a probability of death lower than the set threshold limit to discontinue treatment.

## **Results and discussion**

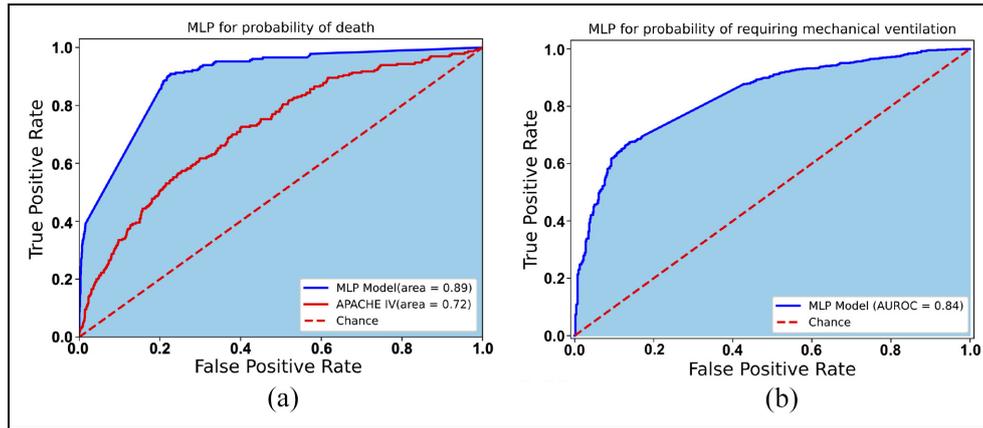
All three tiers discussed in the previous section together form the digital-twin of a patient. A total of five independently trained neural network models take the measured patient vitals and provide various severity indices as output. All models in the present work are designed with ease of use in mind. Many healthcare units lack the state of the art IOT (Internet of Things) based health monitoring systems or face issues with platform compatibility. Manual inputting of data, including time dependent sequential inputs, would be necessary in units where electronic recording capabilities are unavailable. Thus, all inputs are designed to allow manual intervention (see Supplemental Information). The MLP models are designed to take, the most severe value for each input field, within a 24 h period. All RNN based models require two types of inputs, discrete input fields once every 24 h period, that are most severe, and all sequential input fields such as heart rate,  $SpO_2$ , and blood pressures at regular time intervals. These models inherently take note of time elapsed from the start of care and use latest data for calculating severity indices. The two patient databases used for training the models are explained in the previous sections (see for example Table 2).

### **Performance measures**

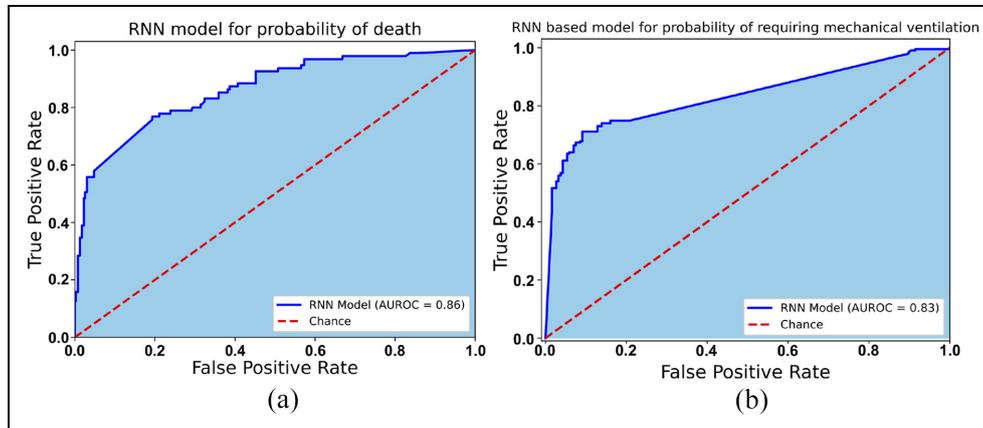
All models, performing classification to calculate probabilities, are evaluated using the Area Under Receiving Operating Characteristic Curve (AUROC), reported with corresponding 95% confidence interval. Receiving Operating Characteristic (ROC)<sup>42,43</sup> curve is a probability curve. Performance is measured using the AUROC. It represents the degree of separability between classes. In the context of the present work, this separability represents the difference between severe and not so severe cases, for each severity index. Higher AUROC represents better performance. The parameters used to estimate AUROC are true and false positive rates. The graphs representing ROC are plotted for false positive rate (one-specificity) versus true positive rate (sensitivity). At AUROC of 0.5, the model loses its discrimination capacity to distinguish between classes. It is represented by a dashed diagonal line in the figures below for reference.

### **Tier 1-“Identify”**

Here, two MLP models are used to predict the probability of death and the probability of requiring mechanical ventilation using patient vital data, measured over a period of 24 h. The AUROC obtained over the total dataset are respectively shown in Figure 3(a) and (b). An AUROC of 0.89 (95% CI: 0.88–0.91) is obtained for the probability of death and 0.84 (95% CI: 0.82–0.86) is obtained for the probability of requiring mechanical ventilation. Cross-validated accuracy



**Figure 3.** AUROC (shaded area) curve for MLP models used to predict the probability of death and the probability of requiring mechanical ventilation support: (a) probability of death and (b) probability of requiring mechanical ventilation.



**Figure 4.** AUROC (shaded area) curve for RNN based model results of the probability of death for patients and the probability of requiring mechanical ventilation support: (a) probability of death and (b) probability of requiring mechanical ventilation.

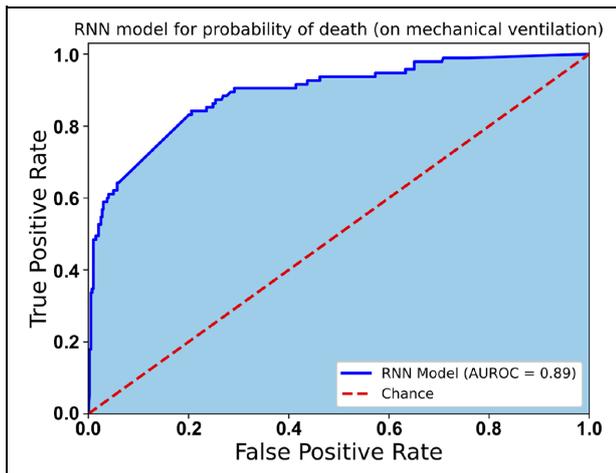
scores of 0.86 (SD:  $\pm 0.005$ ) and 0.70 (SD:  $\pm 0.055$ ), respectively, were obtained in these models on the training dataset. Furthermore, an accuracy score of 0.88 and 0.72 over the testing dataset. Figure 3(a) also shows APACHE IV score with a predicted probability of death with an AUROC of 0.72. Thus, the proposed approach appears to be a substantially better predictor. The model shows that the probability of death and the probability of requiring ventilation with respectively 86% and 85% chances of accurate prediction. From the data computed, a patient prioritisation list for ITU admission may be created by calculating

- (a) Difference in probabilities of death with and without mechanical ventilation.
- (b) Probability of death with mechanical ventilation.
- (c) Probability of requiring mechanical ventilation support.

**Tier 2-“Assign”**

The inpatients within ITUs, whose vitals are continuously monitored, can be assessed for severity of lung disease periodically using the RNN model. As seen in Figure 4(a), an AUROC of 0.86 (95% CI: 0.81–0.90) is obtained from the model for the probability of death, over the entire dataset. The probability of requiring mechanical ventilation is calculated with an AUROC of 0.83 (95% CI: 0.79–0.86) as shown in Figure 4(b). These models obtained a cross validated accuracy scores of 0.73 (SD:  $\pm 0.04$ ) and 0.74 (SD:  $\pm 0.12$ ) respectively. Furthermore, accuracy scores of 0.82 and 0.71 respectively were obtained over the testing dataset. This indicates that with continuous or regular inputs of measured vitals (heart rate, respiration rate,  $SpO_2$ , and blood pressures) and discrete inputs such as haemoglobin, WBC, and sodium, need for mechanical ventilation can be identified with 83% chance. This calculation can

be  
c-  
o-



**Figure 5.** ROC curve for RNN based model used to predict probability of death for patients in Tier Three.

continued into Tier 3 to discontinue ventilation as demonstrated below.

### Tier 3-“Discontinue”

For patients in ITU an approach for the probability of death must be calculated by including the mechanical ventilation. Unlike the RNN based model used in Tier 2, the model here includes mechanical ventilator settings along with the other inputs to calculate the probability of death. As seen in Figure 5, the model performed classification with an AUROC of 0.89 (95% CI: 0.85–0.93) over the entire dataset. This model obtained a cross validated accuracy score of 0.81 (SD:  $\pm 0.025$ ) and testing accuracy score of 0.80. This severity index is required to support the probability of still requiring mechanical ventilation support. A combined analysis of probability of requiring mechanical ventilation support and death is necessary to reduce the chances of erroneous decisions of early removal of support. If both indices are substantially low, the mechanical ventilation may be discontinued.

### Conclusions

The digital-twin system proposed in the present study to construct a subject-specific digital twin appears to provide better results than existing scoring methods. Although the data used to train the models is for non-COVID-19 patients with pneumonia, the similarity of the data to COVID-19 patients appears to be strong. Thus, the proposed digital-twin can be used as a starting point to further refine the scoring system. Overall, the accuracy of prediction is excellent with a minimum AUROC of 0.8 in all cases. By adapting transfer learning on emerging COVID-19 data, this accuracy may be substantially enhanced for severe COVID-19 patients.

Supplementary material provided with this manuscript makes the code open to everyone to download

and train on additional data. With further refinement by the community and our own research group, we believe that the model will serve the healthcare community to deal not only with the COVID-19 pandemic but any future influenza patient prioritisation.

### Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The second author acknowledges partial support from Ser Cymru III - Tackling Covid 19 fund, Welsh Government Project number 095.

### ORCID iD

Neeraj Kavan Chakshu  <https://orcid.org/0000-0002-6430-7939>

### Supplemental material

Supplemental material for this article is available online.

### References

- Adhanom T. WHO director-general’s opening remarks at the media briefing on COVID-19 - 11 march 2020, 2020. <https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020>
- Meares HD and Jones MP. When a system breaks: queueing theory model of intensive care bed needs during the COVID-19 pandemic. *Med J Aust* 2020; 212(10): 470–471.
- Nacoti M, Ciocca A, Giupponi A, et al. At the epicenter of the covid-19 pandemic and humanitarian crises in italy: changing perspectives on preparation and mitigation. *NEJM Catalyst* 2020; 1(2) DOI: 10.1056/CAT.20.0080.
- Vincent JL, Moreno R, Takala J, et al. The SOFA (sepsis-related organ failure assessment) score to describe organ dysfunction/failure. *Intensive Care Med* 1996; 22: 707–710.
- Zimmerman JE, Kramer AA, McNair DS, et al. Acute physiology and chronic health evaluation (APACHE) IV: hospital mortality assessment for today’s critically ill patients. *Crit Care Med* 2006; 34(5): 1297–1310.
- Le Gall JR, Lemeshow S and Saulnier F. A new simplified acute physiology score (SAPS II) based on a european/north american multicenter study. *JAMA* 1993; 270(24): 2957–2963.
- Ferreira FL, Bota DP, Bross A, et al. Serial evaluation of the SOFA score to predict outcome in critically ill patients. *JAMA* 2001; 286(14): 1754–1758.
- Aegerter P, Boumendil A, Retbi A, et al. SAPS II revisited. *Intensive Care Med* 2005; 31(3): 416–423.

9. Beck DH, Smith GB, Pappachan JV, et al. External validation of the SAPS II, APACHE II and APACHE III prognostic models in south England: a multicentre study. *Intensive Care Med* 2003; 29(2): 249–256.
10. Ledoux D, Canivet JL, Preiser JC, et al. SAPS 3 admission score: an external validation in a general intensive care population. *Intensive Care Med* 2008; 34(10): 1873–1877.
11. Gentimis T, Alnaser AJ, Durante A, et al. Predicting hospital length of stay using neural networks on mimic III data. In: *2017 IEEE 15th Intl Conf on Dependable, Autonomous and Secure computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/PiCom/DataCom/CyberSciTech)*, Orlando, FL, USA, 6–10 November 2017, pp.1194–1201. New York, NY: IEEE.
12. Barbieri S, Kemp J, Perez-Concha O, et al. Benchmarking deep learning architectures for predicting readmission to the ICU and describing patients-at-risk. *Sci Rep* 2020; 10(1): 1–10.
13. Pirracchio R, Petersen ML, Carone M, et al. Mortality prediction in intensive care units with the super ICU learner algorithm (sicula): a population-based study. *Lancet Respir Med* 2015; 3(1): 42–52.
14. Assessment RR. Detection of new SARS-CoV-2 variants related to mink. *Eur Cent Dis Prev Control* 2020.
15. Knaus WA, Zimmerman JE, Wagner DP, et al. Apache-acute physiology and chronic health evaluation: a physiologically based classification system. *Crit Care Med* 1981; 9(8): 591–597.
16. Knaus WA, Draper EA, Wagner DP, et al. Apache II: a severity of disease classification system. *Crit Care Med* 1985; 13(10): 818–829.
17. Le Gall JR, Loirat P, Alperovitch A, et al. A simplified acute physiology score for ICU patients. *Crit Care Med* 1984; 12(11): 975–977.
18. Lemeshow S, Teres D, Klar J, et al. Mortality probability models (MPM II) based on an international cohort of intensive care unit patients. *JAMA* 1993; 270(20): 2478–2486.
19. Xia H, Daley BJ, Petrie A, et al. A neural network model for mortality prediction in icu. In: *2012 Computing in cardiology*, Krakow, Poland, 9–12 September 2012, pp.261–264. New York, NY: IEEE.
20. Zahid MAH and Lee J. Mortality prediction with self normalizing neural networks in intensive care unit patients. In: *2018 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, Las Vegas, NV, USA, 4–7 March 2018, pp.226–229. New York, NY: IEEE.
21. Thorsen-Meyer HC, Nielsen AB, Nielsen AP, et al. Dynamic and explainable machine learning prediction of mortality in patients in the intensive care unit: a retrospective study of high-frequency data in electronic patient records. *Lancet Digit Health* 2020; 2(4): e179–e191.
22. Johnson AE, Pollard TJ, Shen L, et al. Mimic-III, a freely accessible critical care database. *Sci Data* 2016; 3: 160035.
23. Chakshu NK, Carson J, Sazonov I, et al. A semi-active human digital twin model for detecting severity of carotid stenoses from head vibration-a coupled computational mechanics and computer vision method. *Int J Numer Methods Biomed Eng* 2019; 35(5): e3180.
24. Kritzinger W, Karner M, Traar G, et al. Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine* 2018; 51(11): 1016–1022.
25. Batty M. Digital twins. *Environ Plan B Urban Anal City Sci* 2018; 45(5): 817–820.
26. Bakator M and Radosav D. Deep learning and medical diagnosis: a review of literature. *Multimodal Technol Interact* 2018; 2(3): 47.
27. Carson JM, Chakshu NK, Sazonov I, et al. Artificial intelligence approaches to predict coronary stenosis severity using non-invasive fractional flow reserve. *Proc IMechE, Part H: J Engineering in Medicine* 2020; 234(11): 1337–1350.
28. Amigo N, Valencia A, Wu W, et al. Cerebral aneurysm rupture status classification using statistical and machine learning methods. *Proc IMechE, Part H: J Engineering in Medicine* 2021; 235(6): 655–662.
29. Singh K, Singh S and Malhotra J. Spectral features based convolutional neural network for accurate and prompt identification of schizophrenic patients. *Proc IMechE, Part H: J Engineering in Medicine* 2021; 235(2): 167–184.
30. Torrey L and Shavlik J. Transfer learning. In: Olivás ES, Guerrero JDM, Sober MM, et al. (eds) *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. Hershey: IGI Global, 2010, pp.242–264.
31. Johnson A, Pollard T and Mark R. Mimic-III clinical database (version 1.4). *PhysioNet*, 2016.
32. Pollard TJ, Johnson AE, Raffa JD, et al. eICU collaborative research database (version 2.0). *PhysioNet*, 2019.
33. Pollard TJ, Johnson AEW, Raffa JD, et al. The eICU collaborative research database, a freely available multi-center database for critical care research. *Sci Data* 2018; 5: 180178.
34. Goldberger AL, Amaral LA, Glass L, et al. Physiobank, physiotoolkit, and physionet: components of a new research resource for complex physiologic signals. *Circulation* 2000; 101(23): e215–e220.
35. Shi H, Han X, Jiang N, et al. Radiological findings from 81 patients with COVID-19 pneumonia in Wuhan, China: a descriptive study. *Lancet Infect Dis* 2020; 20(4): 425–434.
36. Gers FA, Schmidhuber J and Cummins F. Learning to forget: continual prediction with lstm. In: *1999 Ninth International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470)*, Edinburgh, UK, 7–10 September 1999, pp.850–855. New York, NY: IEEE.
37. Hochreiter S and Schmidhuber J. Long short-term memory. *Neural Comput* 1997; 9(8): 1735–1780.
38. Graves A. Long short-term memory. In: *Supervised sequence labelling with recurrent neural networks*. Berlin, Heidelberg: Springer, 2012, pp.37–45.
39. Abadi M, Agarwal A, Barham P, et al. TensorFlow: largescale machine learning on heterogeneous systems. *arXiv preprint arXiv:1603.04467*, 2016.
40. Chollet F, et al. Keras, <https://keras.io>, 2015.
41. Johnson AEW, Stone DJ, Celi LA, et al. The mimic code repository: enabling reproducibility in critical care research. *J Am Med Inform Assoc* 2018; 25(1): 32–39.
42. Fawcett T. An introduction to ROC analysis. *Pattern Recognit Lett* 2006; 27(8): 861–874.
43. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res* 2011; 12: 2825–2830.

**Appendix A: MLP models**

The neural network architecture of the MLP models used in Tier one and the parameters employed during training are presented in this section. Table A describes the training parameters used along with the activation functions employed. Figure A1 are graphical representations of the two models used in the first tier.

In Figure A1, L1 ( $\lambda = 0.0003$ ) is an activity regulariser in the first hidden layer of MLP model used for predicting probability of death. L2 with  $\lambda = 0.0038$  and  $\lambda = 0.0018$  respectively are weight regularisation and activity regularisation, in the first hidden layer of MLP model to predict probability of requiring mechanical ventilation (See supplementary material). First and second dropout layers of MLP model for predicting probability of death has 50% and 40% dropouts respectively. Similarly, both dropout layers in MLP for predicting probability of requiring mechanical ventilation have a dropout of 50%.

The performance of the model is shown in Figure A2. This figure shows the loss seen over epochs, averaged over 10-folds, during training. The convergence of training and validation losses confirm an optimal fitting of the model. Figure A3 shows the calibration curve for the three models, showing close to a perfect calibration.

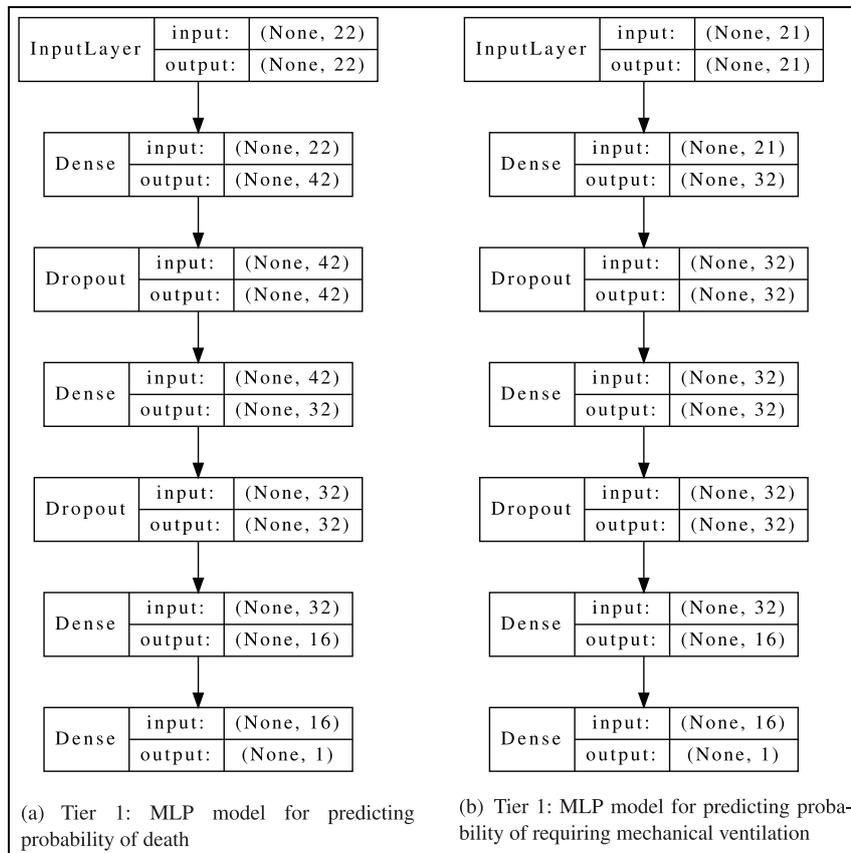
**Table A.** Neural network parameters used in architecture and training of MLP models.

Activation functions	
Hidden layers:	ReLU
Output layer:	Sigmoid
Number of epochs:	100
Batch size:	150
Cost function:	Binary cross-entropy
Optimizer:	Adam
Regularisers used:	L1 & L2

**Appendix B: RNN based models**

The neural network architecture of the MLP models used in tiers 2 and 3, and the parameters employed during training are presented in this section. Table B describes the training parameters used along with the activations functions employed. Figures B1 and B2 are graphical representations of the three RNN models used in the tiers 2 and 3.

The performance of the model is shown in Figure B3. This figure shows the loss seen over epochs, averaged over 10-folds, during training. The convergence of



**Figure A1.** Neural network architecture used for MLP models.

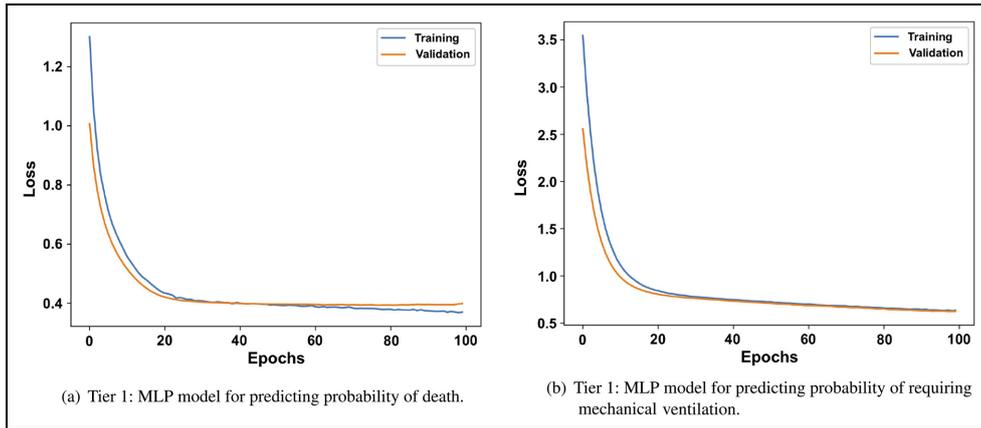


Figure A2. Loss observed over epochs during training for MLP models.

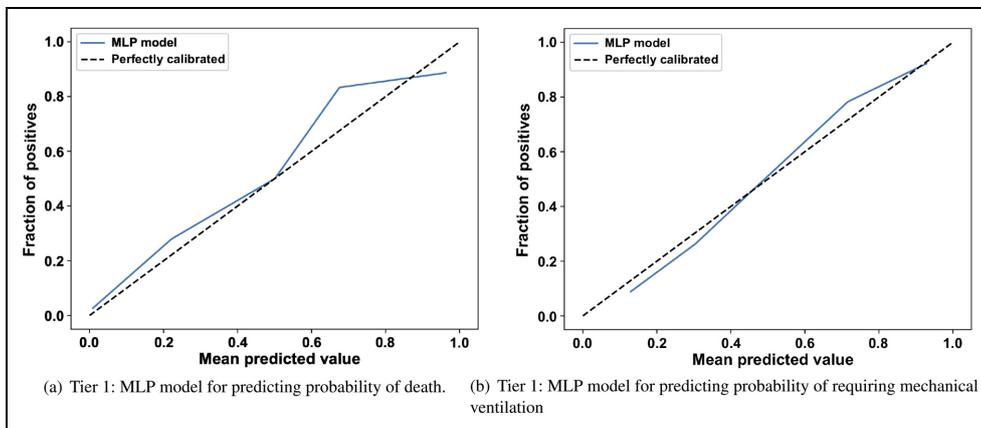


Figure A3. Calibration curves for MLP models.

Table B. Neural network parameters used in architecture and training of RNN based models.

---

Activation functions

---

Dense hidden layers: ReLU  
 Output layer: Sigmoid  
 LSTM: Tanh & Sigmoid (Recurrent activation)  
 (Default for CuDNNLSTM)  
 Number of epochs: 100  
 Batch size: 10  
 Cost function: Binary cross-entropy  
 Optimizer: Adam  
 Regularizers: L1 ( $\lambda = 0.02$ )(All LSTM layers)

---

training and validation losses confirm an optimal fitting of the model. Figure B4 shows the calibration curves for the three models, showing close to a perfect calibration.

In Figures B1 and B2, all dropout layers have 30% dropout except for the RNN model used in predicting probability of requiring mechanical ventilation in which the dropout is 50%. All weights were initialised, during training, with Glorot Normal function.

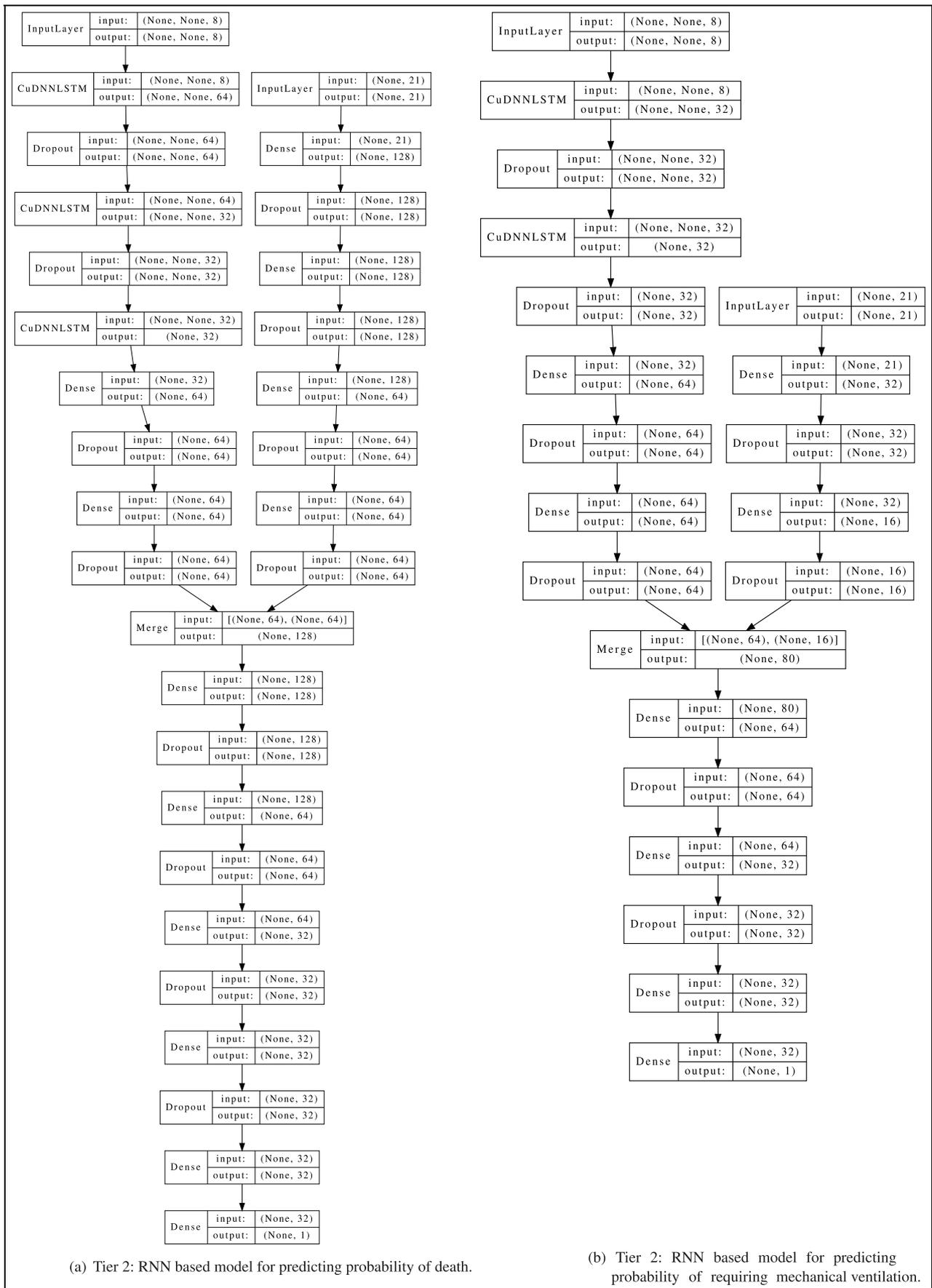


Figure B1. Neural network architecture used for RNN based models in Tier 2.

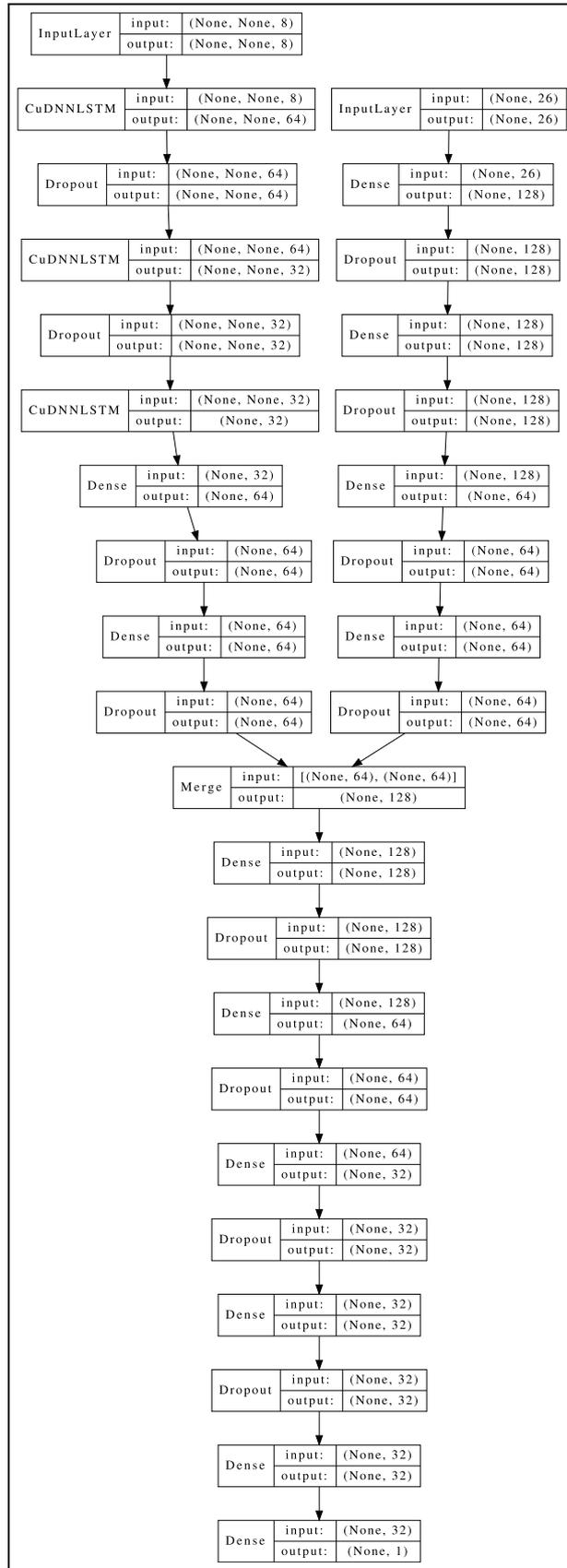


Figure B2. Neural network architecture used for RNN based models in Tier 3.

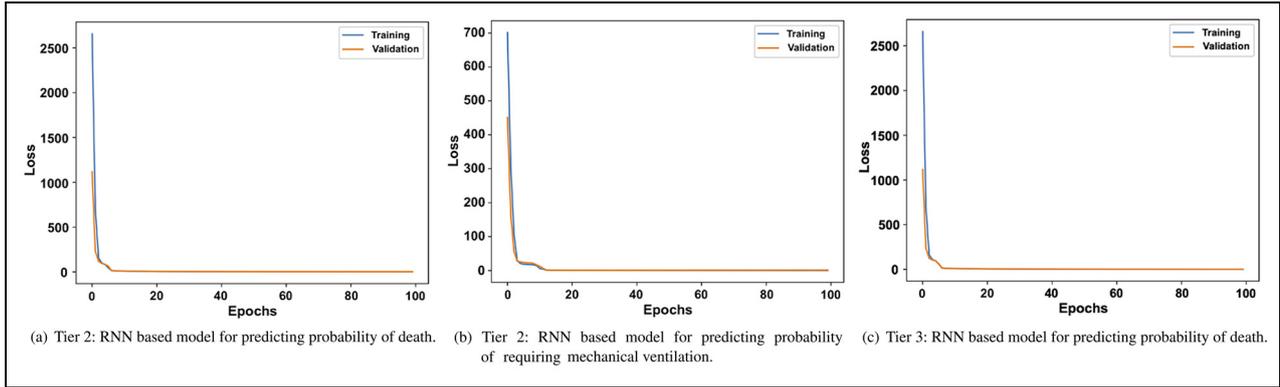


Figure B3. Loss observed over epochs during training for MLP models.

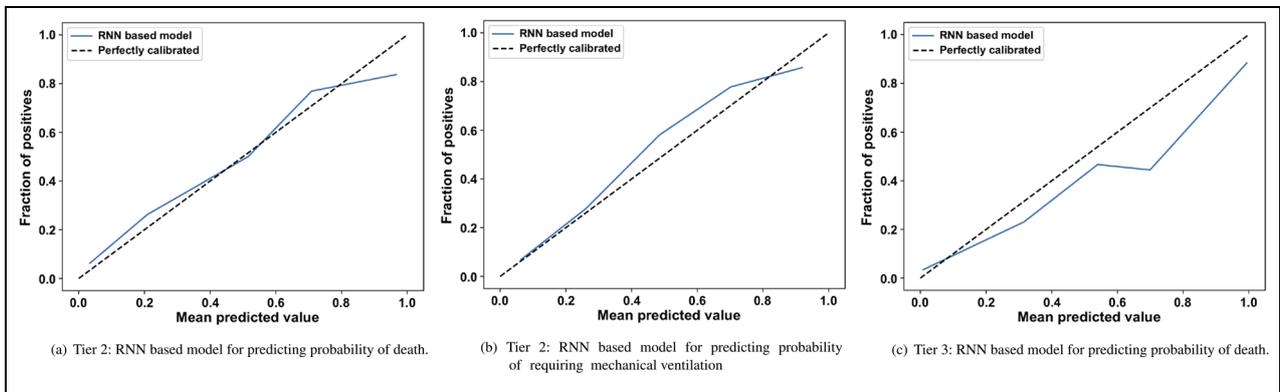


Figure B4. Calibration curves for RNN models.