



Real-Time Prediction for Neonatal Endotracheal Intubation Using Multimodal Transformer Network

Jueng-Eun Im, Shin-Ae Yoon, Yoon Mi Shin , and Seung Park 

Abstract—Neonates admitted to neonatal intensive care units (NICUs) are at risk for respiratory decompensation and may require endotracheal intubation. Delayed intubation is associated with increased morbidity and mortality, particularly in urgent unplanned intubation. By accurately predicting the need for intubation in real-time, additional time can be made available for preparation, thereby increasing the safety margins by avoiding high-risk late intubation. In this study, the probability of intubation in neonatal patients with respiratory problems was predicted using a deep neural network. A multimodal transformer model was developed to simultaneously analyze time-series data (1–3 h of vital signs and FiO_2 setting value) and numeric data including initial clinical information. Over a dataset including information of 128 neonatal patients who underwent noninvasive ventilation, the proposed model successfully predicted the need for intubation 3 h in advance (area under the receiver operator characteristic curve = 0.880 ± 0.051 , F1-score = 0.864 ± 0.031 , sensitivity = 0.886 ± 0.041 , specificity = 0.849 ± 0.035 , and accuracy = 0.857 ± 0.032). Moreover, the proposed model showed high generalization ability by achieving AUROC 0.890, F1-score 0.893, specificity 0.871, sensitivity 0.745, and accuracy 0.864 with an additional 91 dataset for testing.

Index Terms—Endotracheal intubation, neonatal intensive care units, multimodal transformer network, deep neural network.

Manuscript received 2 September 2022; revised 30 March 2023; accepted 11 April 2023. Date of publication 17 April 2023; date of current version 6 June 2023. This work was supported by the Korea Health Industry Development Institute through Korea Health Technology R&D Project, funded by the Ministry of Health & Welfare, Republic of Korea under Grant HI21C1074070021. (Jueng-Eun Im and Shin-Ae Yoon are co-first authors.) (Corresponding authors: Yoon Mi Shin; Seung Park).

Jueng-Eun Im is with the Medical AI Research Team, Chungbuk National University Hospital, Cheongju-si 28644, Republic of Korea (e-mail: tnqkr125@naver.com).

Shin-Ae Yoon is with the Department of Pediatrics, Chungbuk National University Hospital, Chungbuk National University College of Medicine, Cheongju-si 28644, Republic of Korea (e-mail: mercede@naver.com).

Yoon Mi Shin is with the Division of Pulmonary and Critical Care Medicine, Department of Internal Medicine, Chungbuk National University Hospital, Chungbuk National University College of Medicine, Cheongju-si 28644, Republic of Korea (e-mail: anees94@hanmail.net).

Seung Park is with the Biomedical Engineering, Chungbuk National University Hospital, Cheongju-si 28644, Republic of Korea (e-mail: spark.cbnuh@gmail.com).

This article has supplementary downloadable material available at <https://doi.org/10.1109/JBHI.2023.3267521>, provided by the authors.

Digital Object Identifier 10.1109/JBHI.2023.3267521

I. INTRODUCTION

ENDOTRACHEAL intubation is used for sustaining life in premature and term neonates with respiratory difficulty in the neonatal intensive care units (NICUs). However, intubation is accompanied by serious adverse events, including increased morbidity, reintubation, subglottic stenosis, laryngeal injury, ventilator-associated pneumonia events, swallowing and speech impairment, and tracheobronchitis [1], [2], [3], [4]. In particular, for neonates, attempting endotracheal intubation is dangerous compared to adults because they have a small airway [3], hence, it needs to be cautious when attempting endotracheal intubation in neonates. Meanwhile, predicting endotracheal intubation is challenging in neonates because they are in a physiologically unstable period with underdeveloped immunity and fluctuating cardiopulmonary status. Even though several studies on adult intubation have been already performed [1], [5], [6], [7], [8], considerable efforts are still required to develop models for neonates. Because clinical characteristics for adults have a very different range and pattern from those for neonates [9], [10], [11], applying an adult endotracheal intubation model to neonates has limitations. Clark et al.'s study [12] on the prediction of endotracheal intubation in neonates focused on predicting intubation within a relatively broad period (> 24 h), which did not provide precise information on when to perform intubation. Therefore, the purpose of this study is to provide real-time prediction on neonatal intubation within the next 3 h. By accurately predicting the need for intubation in real-time, additional time may be made available for preparation, which can help increase the safety margins by avoiding late intubation for high-risk neonates [1].

To develop real-time prediction model, the multimodal transformer based on deep neural networks was utilized to analyze both 1–3 h of time-series data and numeric data including risk factors for RDS [12], [13], [14] such as prematurity, cesarean section, pregnancy-induced hypertension, male sex, maternal diabetes mellitus, and multiple births. The deep neural networks have been widely applied for prognosis prediction in NICUs with great success [15], [16], [17], [18]. Feng et al. [15] used long short-term memory (LSTM) networks to predict the real-time mortality risk of preterm neonates in the initial stage of NICU hospitalization. Jia et al. [16] proposed a convolutional neural network (CNN)-based model that used routinely recorded neonatal patient information to predict weaning from

mechanical ventilation. Recently, the transformer with the self-attention mechanism, introduced by Vaswani et al. [19], enables the model to compute contextualized representations of the input sequence, while the feed forward neural network is used to apply non-linear transformations to the representations. The transformer has achieved state-of-the-art results in a variety of natural language processing tasks, and has been particularly useful for time-series analysis tasks such as forecasting, anomaly detection, and classification. The transformer networks successfully utilized to efficiently capture disease-medication associations within their temporal context, and outperformed the CNN and LSTM in forecasting time series problems [20], [21], [22], [23], [24], [25].

In addition, the multimodal approach was used to enhance the model performance by providing a comprehensive understanding of clinical data [26], [27], [28], [29], [30]. Shamout et al. [26] proposed a multimodal approach to predict the deterioration of COVID-19 patients in the emergency department. The multimodal used a DNN that learns from chest X-ray images and a gradient boosting model that learns from routine clinical variables, and this model outperformed each single-modal prediction. Sano et al. [31] developed a sleep pattern detector that uses multimodal data acquired from smartphone and wearable technologies to detect sleep patterns with the bidirectional LSTM, and the results were statistically superior to those of non-temporal machine learning algorithms and state-of-the-art actigraphy software.

In this paper, we conducted a retrospective study with datasets of 128 neonates in NICU who experienced respiratory distress and developed a multimodal network to distinguish neonates who require intubation 3 h in advance. The multimodal network consisted of two subnetworks: a multilayer perceptron (MLP) block for numeric data analysis, and a transformer block for time-series data analysis. The feature vectors obtained by two blocks were concatenated and fed to the fully connected layer to calculate the intubation probability. The experimental results demonstrated that the average metrics for the proposed model were as follows: area under the receiver operating curve (AUROC) = 0.880, F1-score = 0.864, sensitivity = 0.886, specificity = 0.849, and accuracy = 0.857. The contributions of this research can be summarized as follows:

- A novel DNN architecture was developed to predict neonatal endotracheal intubation 3 h in advance, resulting in an excellent performance with AUROC 0.880 on a total of 128 neonatal patients in NICU.
- The transformer architecture was successfully integrated with the MLP block for multimodal analysis.
- We established a dataset for intubated neonates in the NICU, consisting of 128 patients.

II. MATERIALS AND METHODS

A. Dataset

A dataset was obtained from Chungbuk National University Hospital (Cheong-Ju, Korea) between June 1, 2020, and March 11, 2023. This retrospective study was conducted in accordance with the Declaration of Helsinki and approved by the

Institutional Review Board of Chungbuk National University Hospital (IRB no. CBNUH 2021-02-034-001), and informed consent was waived. The dataset collected in this study includes the personal information of neonates and their mothers. This dataset will only be used if approved by the corresponding author. The dataset on neonatal patients was manually collected from electronic medical records (EMRs). As described in Fig. 1, we found a list of 577 neonatal patients who were admitted to the NICU at Chungbuk National University Hospital. Patients without respiratory distress (288), patients who were admitted more than 48 hours after birth (53), and patients who had already undergone intubation at the time of admission (73) were excluded from the list. Then, we collected initial clinical and time-series data for 163 patients on the list, and excluded five patients who had attempted intubation after 12 h of admission and those with missing data, defined as more than 2 numeric data or 10% of time-series data, resulting in a total of 128 patients' data included. Because the initial tabular data such as body temperature and blood gas analysis results (pH, PCO₂, PO₂, BE, and lactate) could gradually recover or worsen over time, and may not be representative of the patient's condition in the long term (> 12 h). Accordingly, of the 128 patients, 36 neonates with intubation records were classified as intubated (or positive) patients, and the remaining 92 data were classified as non-intubated (or negative) patients. Before analysis, all positive and negative data were randomly shuffled and subsequently utilized for the analysis of the prediction model. The results of the statistical comparison between intubated and non-intubated neonates showed that the intubated patients had lower pH and pulse oximetry (SpO₂), higher PCO₂ and fraction of inspired oxygen settings (FiO₂), and faster heart rates (Table I).

The dataset includes 19 numeric features (gestational age (GA), birth weight, Apgar scores at 1 and 5 min, sex, cesarean delivery, antenatal steroid use, pregnancy-induced hypertension, gestational diabetes mellitus, premature membrane rupture, out-born delivery, multiple births, initial body temperature, clinical risk index for babies (CRIB-II) score and parameters in initial capillary blood gas analysis (CBGA) including PO₂, PCO₂, base excess (BE), lactate, and pH) and four time-series features (HR, RR, FiO₂, and SpO₂). The time-series data points were recorded in 1 h intervals and included records before the intubation attempts (or the last vital sign record). If intubation was performed less than 3 h after admission, the data recorded immediately after the admission were considered. In the control case, we randomly selected the total observation time and gathered time-series data as in the intubation case.

The numeric and time-series data were preprocessed as follows: The missing values in the numeric data and time-series data were filled with the average values and the most recent data, respectively. Then, the data were normalized via (1), where \hat{x} is the normalized value, and μ_x and σ_x denote the average value and standard deviation of x , respectively.

$$\hat{x} = \frac{x - \mu_x}{\sigma_x} \quad (1)$$

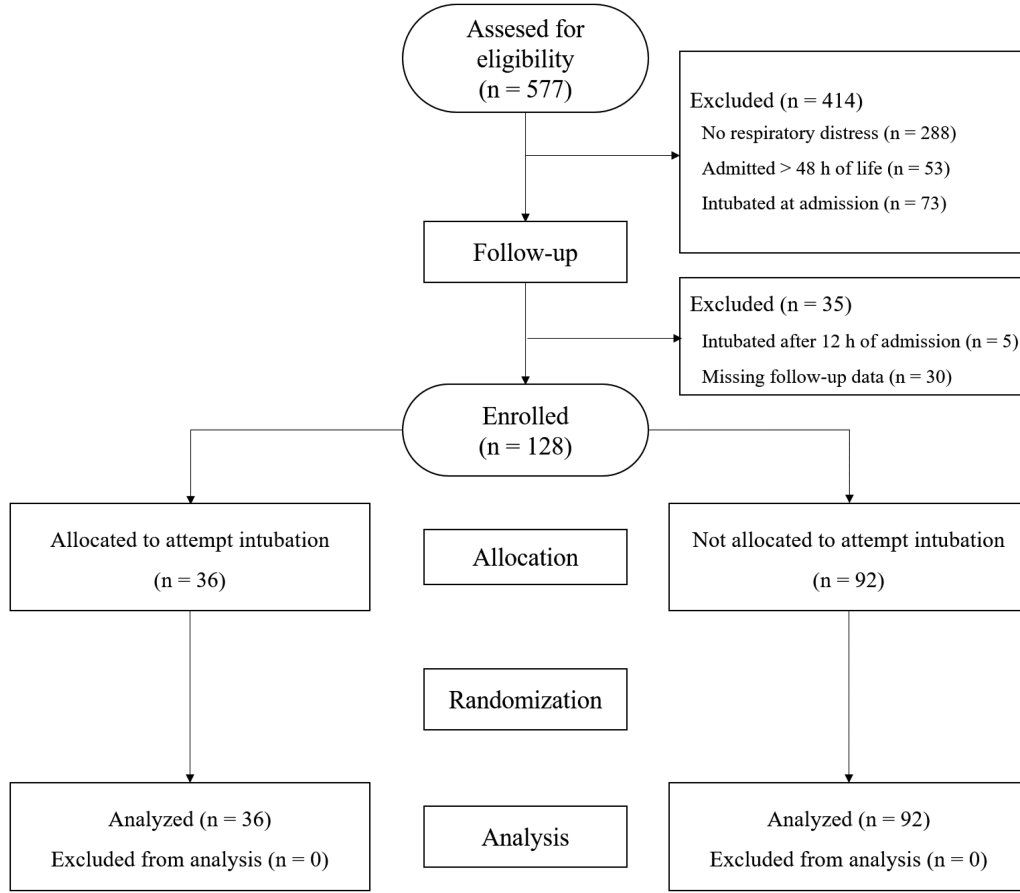


Fig. 1. Flow diagram indicating the selected and excluded clinical studies. Among 577 patients in the neonatal intensive care unit (NICU), 128 patients (36 intubated and 92 non-intubated) were filtered by the exclusion criteria associated with patient selection and missing data.

B. Model Architecture

1) *Multimodal Transformer*: This section describes the multimodal transformer to predict the intubation probability of NICU patients. The proposed model uses two types of data: numeric data including initial clinical information, and time-series data including the vital signs and FiO_2 setting value of a patient. Fig. 2 schematically illustrates the multimodal transformer. The model has an MLP block and a transformer block to analyze numeric data and time-series data, respectively. The feature vectors obtained through the MLP and transformer blocks are concatenated and fed to the additional MLP block to calculate the intubation probability.

Each MLP block consists of a dense layer ($D_{\text{MLP}} = 32$), a batch normalization layer, an activation layer (ReLU for numerical analysis and sigmoid activation for outcome prediction), and a dropout (0.2) layer. The transformer block consists of a transformer encoder including the input embedding network, multi-head attention network, and feedforward network, as illustrated in Fig. 2. In the input embedding network, the time-series data (input_i) are processed by the input embedding network to encode the time sequence of each data point. Specifically, input_i is multiplied by the positional encoding function (PE_i). PE_i is represented as sine and cosine functions with different frequencies, where i and d denote the position and dimension

of the time-series data, respectively, and the embedding model dimension is $D_{\text{embed}} = 32$ (2) and (3).

$$PE_{(i,2k)} = \sin\left(i/10000^{(2k/D_{\text{embed}})}\right), \quad 2k \in d \quad (2)$$

$$PE_{(i,2k+1)} = \cos\left(i/10000^{(2k/D_{\text{embed}})}\right), \quad 2k+1 \in d \quad (3)$$

The positional-encoded feature vector and time-series data processed via the dense layer followed by Gaussian error linear unit (GELU) activation are added at the end of the stack in the input embedding network. In the multi-head attention network, the embedded feature vector is copied as the query (Q), key (K), and value (V). The dot product of Q and K is obtained and scaled to determine the attention weight via the softmax function. The value is multiplied by V (4) to obtain a new sequence (z), where α_{ij} is the weight calculated by softmax, and $W^q, W^k, W^v \in \mathbb{R}^{d \times d}$ are the layer-specific weights for Q, K , and V , respectively. Each attention head processes an input sequence $x = (x_1, \dots, x_n)$ of n elements, with $x_i \in \mathbb{R}^d$, and calculates the output sequence $z = (z_1, \dots, z_n)$ of the same length, where $z_i \in \mathbb{R}^d$ (5).

$$\alpha_{ij} = \text{softmax}\left(\frac{(x_i W^q)(x_j W^k)^T}{\sqrt{D_{\text{embed}}}}\right) \quad (4)$$

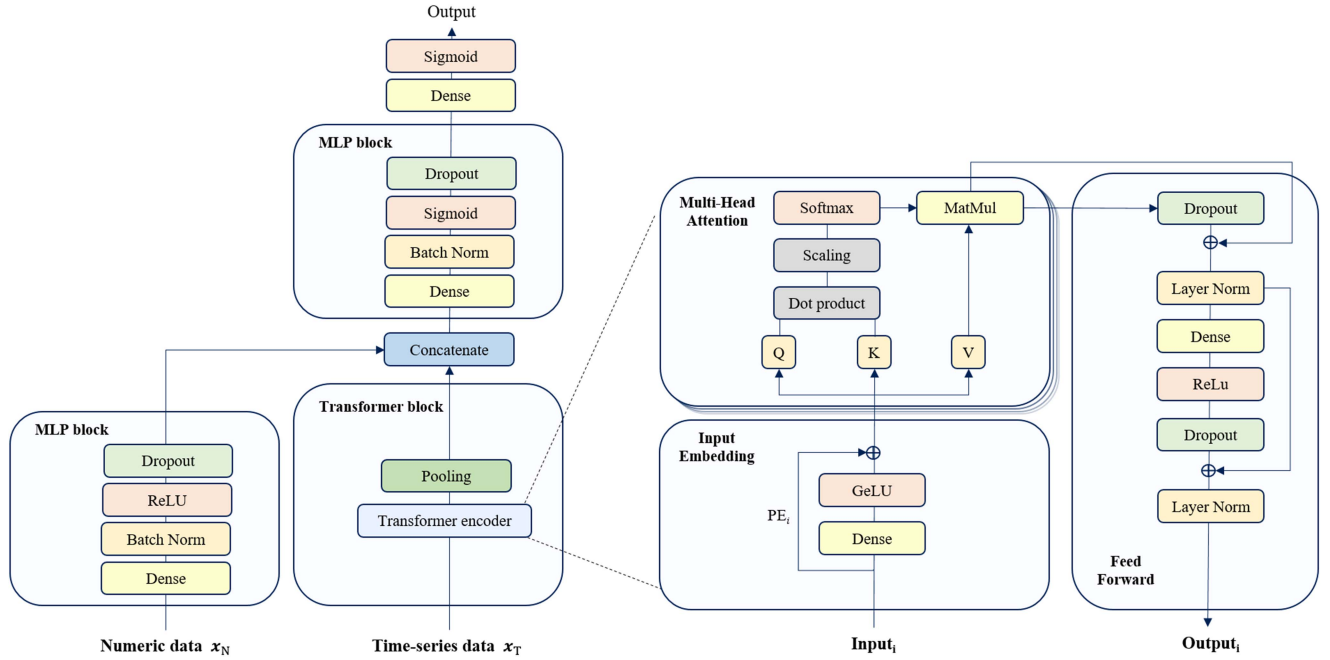


Fig. 2. Multimodal model architecture. The model uses two types of inputs: (i) the numeric data pass through a multilayer perceptron (MLP) block, and (ii) the time-series input passes through the transformer encoder consisting of the input embedding, multi-head attention, and feedforward subnetworks. The subscript i indicates the position of the time-series data. The representations extracted by the two branches are concatenated into a feature vector, which is passed through the MLP model.

$$z = \sum_{j=1}^n \alpha_{ij} (x_j W^v) \quad (5)$$

C. Model Training

1) *Model Parameters*: In a binary classification problem, if one class has significantly more instances than the other, then the model may be biased toward the majority class. To address the problem of data imbalance, one approach is to use loss weighting factors [32], [33], [34]. The idea is to assign different weights to the loss function for each class based on the class frequency in the dataset. Since the imbalanced dataset including 36 positive cases and 92 control cases was utilized in this study, we adjusted the loss weighting factor to prevent learning bias towards the more frequent class. Accordingly, we assigned a weight of 0.8 to the positive class and 0.2 to the negative class during training based on Bayesian Optimization. In addition, we used the Adam optimizer with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an exponential learning rate decay with 100 steps. The initial learning rate was 5×10^{-5} , and the decay rate was set as 0.96. The model was trained for 2,000 epochs.

2) *Implementation Details*: The proposed model was implemented through Python 3.8 (<https://www.python.org/>) with PyCharm (<https://www.jetbrains.com/>) and Windows 10 as the operating system. The transformer and MLP blocks were implemented using the Keras library (<https://keras.io/>) and TensorFlow 2.4 (<https://www.tensorflow.org/>). All models were trained on an NVIDIA RTX 3090 GPU with CUDA v.11.0 (NVIDIA, Santa Clara, CA, US) and Intel core i5-11500 CPU (Intel, Santa Clara, CA, US).

3) *Performance Evaluation*: We performed 4-fold cross-validation to ensure the reliability and consistency of the model performance. In each fold, the dataset was split into training (75%) and validation (25%). In the training and validation datasets, the positive/negative ratios were maintained as approximately 1:4. We evaluated the AUROC, f1-score, sensitivity, specificity, accuracy, ROC curve, and confusion matrix over the validation datasets. The statistical calculation was performed using a scikit-learn library (<https://scikit-learn.org/>).

III. RESULT

A. Data Optimization

1) *Optimization of Time-Series Data*: The proposed model is trained to predict the need for intubation a certain period cutoff time (t_c) in advance. For constructing the dataset, data at a certain time point were labeled as either intubation (positive) or non-intubation (negative). The criterion for the labeling was whether the data were included in t_c before each neonatal intubation in NICU or not. If data were included in t_c , the data were labeled as positive (Fig. 3). To evaluate the effect of t_c on the model performance, we conducted extensive experiments in which t_c was set as 1, 3, 5, and 12 h. In our experiment, the model performance was improved as t_c was set to shorter. The t_c with 1 h corresponded to the best performance (AUROC = 0.892, F1-score = 0.875, sensitivity = 0.858, specificity = 0.853, and accuracy = 0.853), followed by the t_c with 3 h (AUROC = 0.880, F1-score = 0.864, sensitivity = 0.886, specificity = 0.849, and accuracy = 0.857) (Fig. 4). The model exhibited a poor

TABLE I
BASELINE CHARACTERISTICS FOR INTUBATED AND NON-INTUBATED NEONATAL PATIENTS

Variables		Whole cohort (n=128)	Intubated (n=36)	Non-intubated (n=92)	p-value
Perinatal factors	Pregnancy-induced hypertension, No. (%)	9 (7.7)	3 (8)	6 (7)	0.71
	Maternal diabetes mellitus, No. (%)	13 (11.1)	4 (11)	9 (10)	0.248
	Cesarean delivery, No. (%)	88 (75.2)	26 (72)	62 (67)	0.675
	Premature membrane rupture, No. (%)	29 (24.8)	9 (25)	20 (21)	0.823
	Multiple birth, No. (%)	8 (23.9)	9 (25)	19 (21)	0.637
	Outborn delivery, No. (%)	40 (34.2)	14 (38.9)	26 (28.3)	0.34
	Antenatal steroids use, No. (%)	44 (37.6)	11 (31)	33 (38)	0.724
	Neonatal factors	Gestational age at delivery, week (SD)	35.8 (2.8)	35.2 (2.7)	36.1 (2.9)
Birth weight, mean (SD), g		2602 (807.2)	2563.3 (746.1)	2617.2 (833.3)	0.736
Male, No. (%)		72 (56.3)	23 (64)	49 (53)	0.325
Apgar score at 1 min, median (range)		8.7 (4-10)	8.6 (4-10)	8.8 (5-10)	0.225
Apgar score at 5 min, median (range)		9.7 (6-10)	9.5 (6-10)	9.7 (8-10)	0.201
CRIB II score, mean (SD)		2.9 (1.2)	3.1 (1.1)	2.9 (1.2)	0.27
pH, mean (SD)		7.3 (0.1)	7.2 (0.1)	7.3 (0.1)	0.004*
PCO ₂ , mean (SD)		59 (10.2)	63.4 (11.7)	57.2 (9)	<0.001*
PO ₂ , mean (SD)		39.6 (16)	37.1 (10)	40.5 (17.8)	0.273
Base excess, mean (SD)		-4.7 (2.6)	-5.3 (2.2)	-4.4 (2.7)	0.075
Lactate, mean (SD)		3.2 (1.8)	2.7 (1.4)	3.4 (2)	0.081
Heart rate, mean (SD)		134.6 (14.2)	143.5 (15.4)	133.9 (13.9)	<0.001*
Respiratory rate, mean (SD)		51.7 (19.1)	50.8 (18.9)	51.7 (19.1)	0.651
Pulse oximetry at admission, mean (SD)		96.9 (3.7)	93.3 (6.8)	97.2 (3.1)	<0.001*
Fraction of inspired oxygen, mean (SD)		23 (4.1)	27.3 (8)	22.6 (3.3)	<0.001*

performance in extremely long-term prediction $t_c = 12$ (AUROC = 0.805, F1-score = 0.750, sensitivity = 0.816, specificity = 0.707, and accuracy = 0.739). The model performance at $t_c = 1$ h and 3 h was comparable, indicating that the appropriate t_c of model was 3 h because it allowed for accurate predictions at an earlier time (Table II).

2) Multimodality Data: We used two subnetworks to simultaneously process the numeric and time-series data. To analyze the effect of each subnetwork on the model performance, we performed an ablation study, the results of which are summarized in Table III. The model using only the numeric data without CBGA scores exhibited a poor performance (AUROC = 0.710, F1-score = 0.673, sensitivity = 0.826, specificity = 0.643, and accuracy

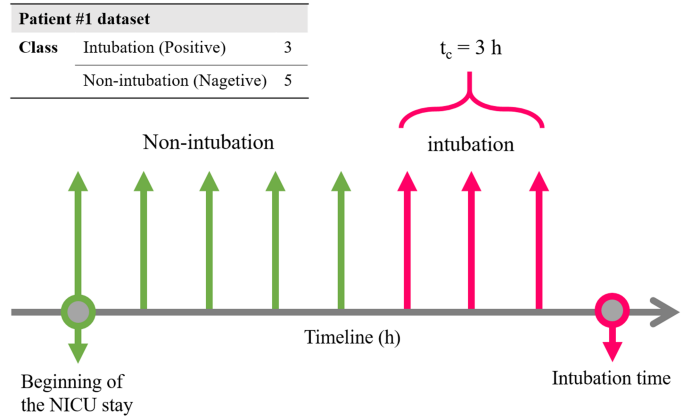


Fig. 3. An illustration of an intubated patient timeline. Multiple samples are gathered over time from an intubated patient. The data within the cutoff time ($t_c = 3$ h) is classified as intubation (positive), and the others are classified as non-intubation (negative).

TABLE II
COMPARISON OF MODEL PERFORMANCE UNDER VARIOUS CUTOFF TIME (t_c) VALUES

t_c	AUC	F1-score	Sensitivity	Specificity	Accuracy
1h	0.892 ± 0.096	0.875 ± 0.058	0.858 ± 0.139	0.853 ± 0.062	0.853 ± 0.068
3h	<u>0.880 ± 0.051</u>	<u>0.864 ± 0.031</u>	0.886 ± 0.041	<u>0.849 ± 0.035</u>	0.857 ± 0.032
5h	0.823 ± 0.057	0.790 ± 0.051	0.846 ± 0.087	0.756 ± 0.088	0.778 ± 0.054
12h	0.805 ± 0.020	0.750 ± 0.045	0.816 ± 0.037	0.707 ± 0.075	0.739 ± 0.046

The best and second-best results in each category are in bold and underlined, respectively.

TABLE III
COMPARISON OF MODEL PERFORMANCE USING DIFFERENT DATA MODALITIES

Data type	AUC	F1-score	Sensitivity	Specificity	Accuracy
Numeric ^a	0.773 ± 0.045	0.784 ± 0.060	0.776 ± 0.070	0.768 ± 0.078	0.770 ± 0.065
Numeric ^b	0.710 ± 0.073	0.673 ± 0.132	0.826 ± 0.039	0.643 ± 0.023	0.648 ± 0.066
Series	<u>0.866 ± 0.088</u>	0.831 ± 0.082	<u>0.871 ± 0.046</u>	0.807 ± 0.106	0.815 ± 0.078
Numeric ^a +Series	0.880 ± 0.051	0.864 ± 0.031	0.886 ± 0.041	0.849 ± 0.035	0.857 ± 0.032
Numeric ^b +Series	0.862 ± 0.066	<u>0.837 ± 0.081</u>	0.848 ± 0.058	<u>0.820 ± 0.110</u>	<u>0.827 ± 0.091</u>

The best and second-best results in each category are in bold and underlined, respectively.

^aNumeric, numeric data with CBGA; ^bNumeric, numeric data without CBGA

= 0.648), followed by the model using numeric data (AUROC = 0.773, F1-score = 0.784, sensitivity = 0.776, specificity = 0.768, and accuracy = 0.770). The results demonstrated that the numeric data initially collected at NICU admission are insufficient to predict the need for intubation. The model using the time-series data exhibited the following average values: AUROC = 0.866, F1-score = 0.815, sensitivity = 0.854, specificity = 0.856, and accuracy = 0.811, which indicated that the changes in the vital signs and FiO₂ setting value plays an essential role in predicting the need for intubation. The multimodal model achieved the best performance (AUROC = 0.880, F1-score = 0.864, sensitivity = 0.886, specificity = 0.849, and accuracy = 0.857). The performance of the multimodal model without

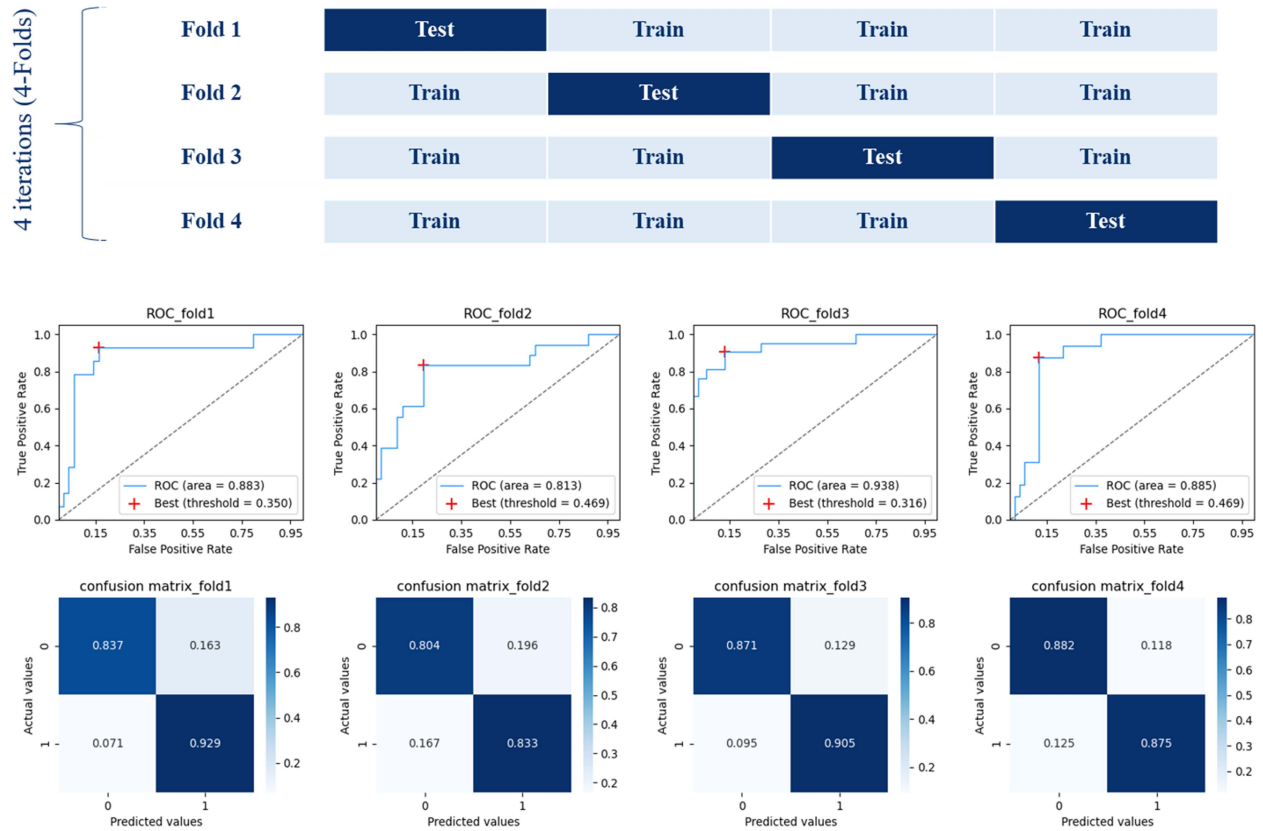


Fig. 4. Performance analysis of the proposed model with $t_c = 3$ h. The model was evaluated through 4-fold cross-validation. The receiver operating curve (ROC) and the normalized confusion matrix for each fold were obtained. Rows of the confusion matrix represent the actual values, and the columns represent the model prediction (0: non-intubation case; 1: intubation case).

CBGA scores was comparable to that of the model using CBGA scores (AUROC = 0.862, F1-score = 0.818, sensitivity = 0.902, specificity = 0.780, and accuracy = 0.807).

B. Model Comparison

The proposed model based on the multimodal transformer was compared with other models used in the recent prognosis studies, specifically, LR [35], [36], [37], extreme gradient boosting (XGBoost) [38], [39], [40], SVM [36], [41], [42], MLP [36], [43], [44], [45] and LSTM [46], [47], [48], [49]. We evaluated the AUROC, sensitivity, specificity, and accuracy of each model over the validation datasets.

The LR and SVM models with a Gaussian radial basis function (RBF) were performed using scikit-learn library. The XGBoost regressor was implemented using XGBoost library. The subsample ratio of columns and max-depth were set as 0.8 and 8, respectively. For the model training, the learning rate was set as 10^{-4} , and the tree estimator was set to 100. To address the problems of data imbalance and overfitting issues, the weights of the positive class and gamma were set as 2 and 1, respectively. These hyperparameters were optimized through a grid search method with 4-fold cross-validation. Before being fed to the LR, SVM, and XGBoost models, the time-series data were flattened and concatenated with the numeric data. For the MLP and LSTM models, the transformer block was replaced

TABLE IV
COMPARISON OF THE PROPOSED MODEL WITH OTHER MODELS

	AUC	F1-score	Sensitivity	Specificity	Accuracy
LR	0.790 ± 0.122	0.633 ± 0.128	0.727 ± 0.414	0.627 ± 0.282	0.637 ± 0.134
XGBoost	0.857 ± 0.093	0.818 ± 0.081	<u>0.853 ± 0.111</u>	0.791 ± 0.096	0.807 ± 0.086
SVM	0.786 ± 0.117	0.797 ± 0.041	0.798 ± 0.252	0.785 ± 0.124	0.788 ± 0.052
MLP	0.863 ± 0.073	0.812 ± 0.083	0.820 ± 0.113	0.797 ± 0.135	0.801 ± 0.092
LSTM	<u>0.863 ± 0.045</u>	<u>0.820 ± 0.045</u>	0.820 ± 0.058	<u>0.807 ± 0.055</u>	<u>0.809 ± 0.049</u>
Proposed	0.880 ± 0.051	0.864 ± 0.031	0.886 ± 0.041	0.849 ± 0.035	0.857 ± 0.032

The best and second-best results in each category are in bold and underlined, respectively.

with an MLP block and LSTM block, respectively (Appendix I, Fig. 6). These models were implemented with the Keras library. The MLP block consisted of a hidden dense layer ($D_{MLP} = 128$) with rectified linear unit (ReLU) activation. The time-series data were flattened and fed to the input MLP block used instead of the transformer block. The LSTM block included an LSTM layer ($D_{LSTM} = 128$).

Table IV highlighted that the proposed model achieved the best performance (AUROC = 0.880, F1-score = 0.864, sensitivity = 0.886, specificity = 0.849, and accuracy = 0.857). This

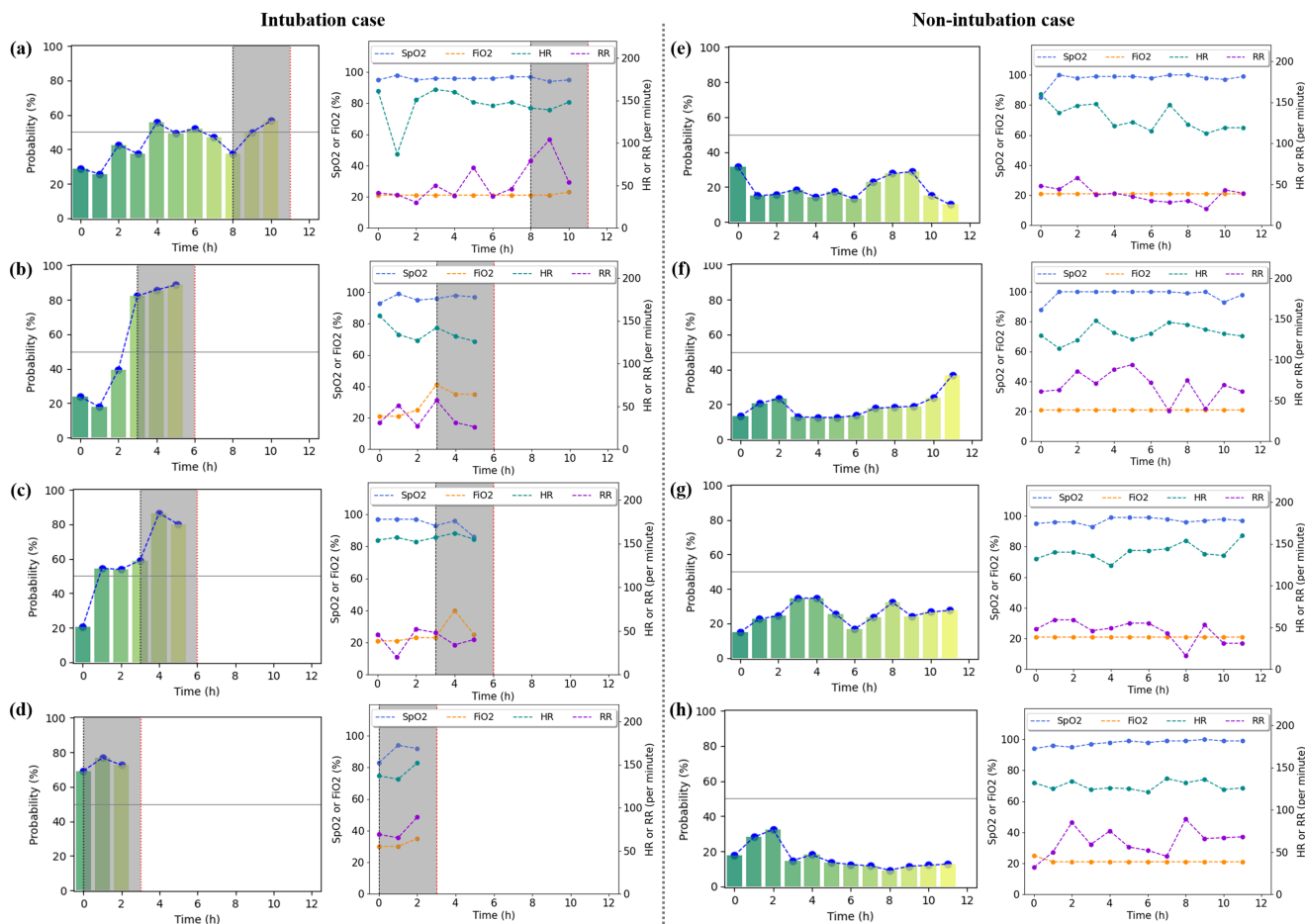


Fig. 5. Real-time prediction for eight patients, obtained using the proposed model. Intubated and non-intubated patients are represented at (a)–(d), and (e)–(h), respectively. Graphs indicating the predicted intubation probability (left) and time-series data (right) in real-time. The predicted probabilities were classified as intubated and non-intubated classes according to the best threshold (horizontal dotted lines). The dotted black line represents the intubation time derived 3 h in advance ($t_c = 3$ h), and the dotted red line represents the intubation time. t_c : cutoff time; HR: heart rate (beats/min); RR: respiratory rate (breath/min); SpO₂: oxygen saturation; FIO₂ (%): fraction of inspired oxygen (%).

result showed that the proposed multimodal transformer outperformed machine learning models such as LR (AUROC = 0.790, F1-score = 0.633, sensitivity = 0.727, specificity = 0.627, and accuracy = 0.630), XGBoost (AUROC = 0.857, F1-score = 0.818, sensitivity = 0.853, specificity = 0.791, and accuracy = 0.807), and SVM (AUROC = 0.786, F1-score = 0.797, sensitivity = 0.798, specificity = 0.785, accuracy = 0.788). In addition, the proposed model outperformed other DNN models including LSTM model (AUROC = 0.863, F1-score = 0.820, sensitivity = 0.820, specificity = 0.807, and accuracy = 0.809), and the MLP model (AUROC = 0.863, F1-score = 0.812, sensitivity = 0.820, specificity = 0.797, accuracy = 0.801).

IV. DISCUSSION

As coronavirus disease 2019 (COVID-19) pandemic spreads around the world, respiratory-related studies have been introduced using machine learning and deep learning. To be specific, Varzaneh et al. [8] developed a decision tree-based model to predict the intubation risk of hospitalized patients with an accuracy of 0.93. Bolourani et al. [50] suggested the XGBoost

model that predicts respiratory failure in admitted patients within 48 h of admission using a dataset of the emergency department with the highest mean accuracy of 0.919. Siu et al. [1] also successfully predicted adult intubation using RF model with a freely available Medical Information Mart for Intensive Care (MIMIC) dataset with an AUROC of 0.87. However, only scarce literature targeting neonates or infants has been reported. Previous studies [12], [51] proposed simple logistic models to predict intubation (an AUROC of 0.84) or late respiratory support (AUROCs of 0.801–0.881) for very low birth weight infants. Since the recent studies have focused only on whether to attempt intubation or respiratory support, developing a real-time model was needed to identify the timing of intubation, which requires a high burden and cost for medical staff. Furthermore, the Clark et al. [12] requires vital sign collected at 2-second intervals requiring automatic extraction features from electronic medical records (EMRs), but the different clinical data formats (CRFs) between hospitals and countries makes it impractical for use in the majority of hospitals. Additionally, our study targets all term and preterm neonates with respiratory diseases, thus it is inevitable to collect additional data and develop an optimized

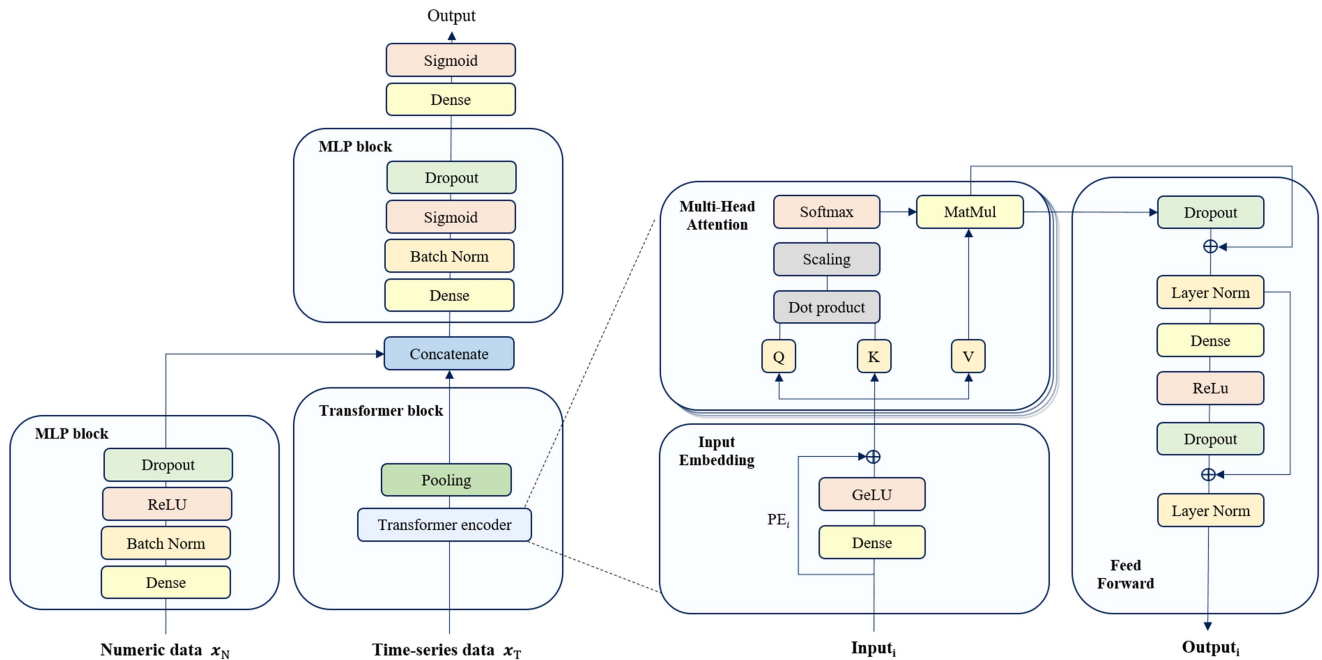


Fig. 6. Multimodal MLP and LSTM architectures.

model accordingly. To develop a real-time prediction model for intubation, the advanced model architecture needs to be applied. Compared to the traditional machine learning methods such as logistic regression, SVM, and RF, the transformers have shown a great ability for interactions in sequential data, and have widely adapted to a variety of time-series tasks such as forecasting [52], [53], [54], anomaly detection [55], [56], and classification [57], [58]. Therefore, in this study, a transformer architecture [19] was adapted for analyzing time-series data and an MLP block was added for comprehensive analysis with numeric data, resulting in the development of a multimodal transformer capable of effectively analyzing both time-series data and numeric data simultaneously.

In this study, the proposed model provides prediction of intubation probabilities using multi-modal information including 1–3 h of vital signs with FiO_2 setting value and initial clinical variables. To develop and evaluate this model, we established a dataset including the information of 128 neonatal patients under NIV support and validate the multimodal model that could classify neonates requiring intubation 3 h in advance. We focused on developing a real-time system capable of predicting short-term prediction because our dataset included 75% of intubated neonatal patients who attempted intubation 3 hours after NICU admission. Fig. 5 demonstrates that the predicted intubation probability reflects the historical trends of time-series data. In the Fig. 5(a), the respiratory rate (RR) increased to 104 beats/min (normal range of RR: 30–60 beats/min) between 9–10 h after NICU admission. During this period, the predicted intubation probabilities increased (50%–56.5%). In Fig. 5(b), the FiO_2 drastically increased to 35–41% (FiO_2 is typically maintained at 21%) between 3–5 h after NICU admission, and the predicted intubation probability approached 85.4%. Fig. 5(c) represents the probability increased slightly to 54.5% at 1 h,

and sharply increased to 86.4% at 4 h after admission to the NICU, which is similar to the trend of time-series data. The RR was temporarily decreased by 20 beats/min at 1 h, and the FiO_2 increased rapidly to 40% after 4 hours of admission to the NICU. Fig. 5(d) shows that the predicted intubation probability was maintained at 70.7% or higher. During this period, the RR (69–89 beats/min) and FiO_2 (30–35%) were out of the normal range. In addition, we also present the prediction results of non-intubated patients (Fig. 5(e)–(h)). Fig. 5(e) shows fluctuation of the HR (112–160 beats/min) (normal range of HR: 120–160 breaths/min). Fig. 5(f)–(h) shows the RR values were often appeared to deviate normal range. However, we observe stable trends in the FiO_2 and SpO_2 , and confirm that the predictive probability of intubation was also stable.

We compared the multimodal transformer network model with traditional machine learning models, and the proposed model outperformed the LR, XGBoost, SVM, MLP, and LSTM. The transformer network could effectively learn sequence data through parallel computation and a structured memory for managing long-term dependencies [19]. Recently, transformer-based networks have exhibited outperforming performance in various research domains [19], [59], [60], [61], [62], [63], [64].

The ablation study was conducted to optimize the model architecture, and the model using both numeric and time-series data achieved the highest performance. The model with 18 clinical data without CBGA scores (PO_2 , PCO_2 , BE, lactate, and pH) also showed comparable predictive performance (AUROC = 0.862) with one using CBGA scores (AUROC = 0.880). This model could be also practically used in institutions and hospitals because the CBGA requires expensive analyzing instruments and trained technicians.

We did further study to test the proposed model using a recently obtained test dataset from 91 neonatal patients, including

21 intubated patients and 70 non-intubated patients, in addition to the existing data of 128 patients used for model development. The proposed model that achieved the highest performance was utilized for testing. Using this model, we performed endotracheal intubation predictions for 91 neonatal patients at hourly intervals. This result was validated by checking whether the neonates had a record of endotracheal intubation within the next 3 hours. As a result, the model achieved a performance of AUROC 0.890, F1-score 0.893, specificity 0.871, sensitivity 0.745, and accuracy 0.864. These results, as compared by the cross-validation results, demonstrate that the proposed model has high generality ability, though additional research may be needed to improve sensitivity.

For future work, ensemble learning [65], [66] could be considered to improve the accuracy and robustness of a decision-making system by combining the outputs of multiple models. In addition, state-of-art feature selection technologies [67], [68], [69] could be applied to reduce the dimensionality of the dataset, simplify the model, and improve its accuracy and generalization ability.

V. CONCLUSION

We have developed a real-time prediction system using a multimodal transformer network to predict neonatal endotracheal intubation 3 hours in advance. The proposed model has outperformed traditional machine learning models with AUROC = 0.880, F1-score = 0.864, sensitivity = 0.886, specificity = 0.849, and accuracy = 0.857, and reflected the historical trends of vital signs and FiO₂ setting value well. Furthermore, the proposed model demonstrated high generalization ability by showing AUROC 0.890, F1-score 0.893, specificity 0.871, sensitivity 0.745, and accuracy 0.864 with an additional 91 collected data for testing. To find the optimal model architecture, we have conducted the ablation study. It has been desirable to simultaneously process the numeric and time-series data compared to the single-modal network. In addition, we have performed extensive experiments to select the most appropriate cutoff time among $t_c = 1, 3, 5,$ and 12 h, and the appropriate t_c was selected as 3 h. We expect that the proposed model could be practically applied to neonates at high risk for respiratory failure by using 23 factors readily available in hospitals.

APPENDIX A

ARCHITECTURE OF MULTIMODAL MLP AND LSTM

The multimodal MLP (Fig. 6(a)) and the multimodal LSTM (Fig. 6(b)) utilize two-type inputs: (i) the numeric data passes through a multilayer perceptron (MLP) block, and (ii) the time-series data passes through MLP and LSTM block, respectively. The representations extracted by the two branches are concatenated into a feature vector, which is passed through an MLP model.

REFERENCES

- [1] B. M. K. Siu et al., "Predicting the need for intubation in the first 24 h after critical care admission using machine learning approaches," *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020.
- [2] A. Lamoshi et al., "Association of anesthesia type with prolonged post-operative intubation in neonates undergoing inguinal hernia repair," *J. Perinatol.*, vol. 41, no. 3, pp. 571–576, 2021.
- [3] J. L. Wei and J. Bond, "Management and prevention of endotracheal intubation injury in neonates," *Curr. Opin. Otolaryngol. Head Neck Surg.*, vol. 19, no. 6, pp. 474–477, 2011.
- [4] V. Venkatesh et al., "Endotracheal intubation in a neonatal population remains associated with a high risk of adverse events," *Eur. J. Pediatrics*, vol. 170, no. 2, pp. 223–227, 2011.
- [5] I. Ucgun, M. Metintas, H. Moral, F. Alatas, H. Yildirim, and S. Erginel, "Predictors of hospital outcome and intubation in COPD patients admitted to the respiratory ICU for acute hypercapnic respiratory failure," *Respir. Med.*, vol. 100, no. 1, pp. 66–74, 2006.
- [6] M. Hua, J. Brady, and G. Li, "A scoring system to predict unplanned intubation in patients having undergone major surgical procedures," *Anesth. Analg.*, vol. 115, no. 1, pp. 88–94, 2012.
- [7] V. Arvind, J. S. Kim, B. H. Cho, E. Geng, and S. K. Cho, "Development of a machine learning algorithm to predict intubation among hospitalized patients with COVID-19," *J. Crit. Care*, vol. 62, pp. 25–30, 2021.
- [8] Z. A. Varzaneh, A. Orooji, L. Erfannia, and M. Shanbehzadeh, "A new COVID-19 intubation prediction strategy using an intelligent feature selection and K-NN method," *Inform. Med. Unlocked*, vol. 28, 2022, Art. no. 100825.
- [9] C.-C. Wang et al., "Differences in clinical and laboratory characteristics and disease severity between children and adults with dengue virus infection in Taiwan 2002," *Trans. Roy. Soc. Trop. Med. Hyg.*, vol. 103, no. 9, pp. 871–877, 2009.
- [10] J. S. Tregoning and J. Schwarze, "Respiratory viral infections in infants: Causes, clinical symptoms, virology, and immunology," *Clin. Microbiol. Rev.*, vol. 23, no. 1, pp. 74–98, 2010.
- [11] L. C. Ku and P. B. Smith, "Dosing in neonates: Special considerations in physiology and trial design," *Pediatr. Res.*, vol. 77, no. 1, pp. 2–9, 2015.
- [12] M. T. Clark et al., "Predictive monitoring for respiratory decompensation leading to urgent unplanned intubation in the neonatal intensive care unit," *Pediatr. Res.*, vol. 73, no. 1, pp. 104–110, 2013.
- [13] J. Liu, N. Yang, and Y. Liu, "High-risk factors of respiratory distress syndrome in term neonates: A retrospective case-control study," *Balkan Med. J.*, vol. 31, no. 1, pp. 64–68, 2014.
- [14] N. Brix et al., "Predictors for an unsuccessful intubation-surfactant-extubation procedure: A cohort study," *BMC Pediatrics*, vol. 14, no. 1, pp. 1–8, 2014.
- [15] J. Feng et al., "Predicting mortality risk for preterm infants using deep learning models with time-series vital sign data," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–8, 2021.
- [16] Y. Jia et al., "Prediction of weaning from mechanical ventilation using convolutional neural networks," *Artif. Intell. Med.*, vol. 117, 2021, Art. no. 102087.
- [17] A. Jalali et al., "Deep learning for improved risk prediction in surgical outcomes," *Sci. Rep.*, vol. 10, no. 1, pp. 1–13, 2020.
- [18] X. P. Burgos-Artizzu et al., "Evaluation of an improved tool for non-invasive prediction of neonatal respiratory morbidity based on fully automated fetal lung ultrasound analysis," *Sci. Rep.*, vol. 9, no. 1, pp. 1–7, 2019.
- [19] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Conf. Neural Inf. Process. Syst.*, 2017, pp. 1–11.
- [20] O. Nitski et al., "Long-term mortality risk stratification of liver transplant recipients: Real-time application of deep learning algorithms on longitudinal data," *Lancet Digit. Health*, vol. 3, no. 5, pp. e295–e305, 2021.
- [21] S. Rao et al., "An explainable transformer-based deep learning model for the prediction of incident heart failure," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 7, pp. 3362–3372, Jul. 2022.
- [22] M. Mahbub et al., "Unstructured clinical notes within the 24 hours since admission predict short, mid & long-term mortality in adult ICU patients," *Plos One*, vol. 17, no. 1, 2022, Art. no. e0262182.
- [23] Y. S. Dosso, D. Kyrollos, K. J. Greenwood, J. Harrold, and J. R. Green, "NICUface: Robust neonatal face detection in complex NICU scenes," *IEEE Access*, vol. 10, pp. 62893–62909, 2022.
- [24] Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. S. Torr, and G. Zhao, "PhysFormer: Facial video-based physiological measurement with temporal difference transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4186–4196.
- [25] B. Fu et al., "Attention-based recurrent multi-channel neural network for influenza epidemic prediction," in *Proc. IEEE Int. Conf. Bioinf. Biomed.*, 2018, pp. 1245–1248.

- [26] F. E. Shamout et al., "An artificial intelligence system for predicting the deterioration of COVID-19 patients in the emergency department," *NPJ Digit. Med.*, vol. 4, no. 1, pp. 1–11, 2021.
- [27] S.-C. Huang et al., "Fusion of medical imaging and electronic health records using deep learning: A systematic review and implementation guidelines," *NPJ Digit. Med.*, vol. 3, no. 1, pp. 1–9, 2020.
- [28] X. Zheng et al., "Deep learning radiomics can predict axillary lymph node status in early-stage breast cancer," *Nature Commun.*, vol. 11, no. 1, pp. 1–9, 2020.
- [29] I. I. Karipidis et al., "Simulating reading acquisition: The link between reading outcome and multimodal brain signatures of letter–speech sound learning in prereaders," *Sci. Rep.*, vol. 8, no. 1, pp. 1–13, 2018.
- [30] A. E. Fetit et al., "A multimodal approach to cardiovascular risk stratification in patients with type 2 diabetes incorporating retinal, genomic and clinical features," *Sci. Rep.*, vol. 9, no. 1, pp. 1–10, 2019.
- [31] A. Sano, W. Chen, D. Lopez-Martinez, S. Taylor, and R. W. Picard, "Multimodal ambulatory sleep detection using LSTM recurrent neural networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 4, pp. 1607–1617, Jul. 2018.
- [32] H. He, Y. Bai, E. A. Garcia, and S. Li, "ADASYN: Adaptive synthetic sampling approach for imbalanced learning," in *Proc. IEEE Int. Joint Conf. Neural Netw. (World Congr. Comput. Intell.)*, 2008, pp. 1322–1328.
- [33] M. Kozłowski, M. Woźniak, and B. Krawczyk, "Combined cleaning and resampling algorithm for multi-class imbalanced data with label noise," *Knowl.-Based Syst.*, vol. 204, 2020, Art. no. 106223.
- [34] K. R. M. Fernando and C. P. Tsokos, "Dynamically weighted balanced loss: Class imbalanced learning and confidence calibration of deep neural networks," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 7, pp. 2940–2951, Jul. 2021.
- [35] H. Caly et al., "Machine learning analysis of pregnancy data enables early identification of a subpopulation of newborns with ASD," *Sci. Rep.*, vol. 11, no. 1, pp. 1–14, 2021.
- [36] M. Roberts et al., "Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans," *Nature Mach. Intell.*, vol. 3, no. 3, pp. 199–217, 2021.
- [37] S. Valpione et al., "The T cell receptor repertoire of tumor infiltrating T cells is predictive and prognostic for cancer survival," *Nature Commun.*, vol. 12, no. 1, pp. 1–8, 2021.
- [38] D. R. Giacobbe, "Clinical interpretation of an interpretable prognostic model for patients with COVID-19," *Nature Mach. Intell.*, vol. 3, no. 1, pp. 16–16, 2021.
- [39] Z. Zhong et al., "Machine learning prediction models for prognosis of critically ill patients after open-heart surgery," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, 2021.
- [40] J. Jiang et al., "Predictive model for the 5-year survival status of osteosarcoma patients based on the SEER database and XGBoost algorithm," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021.
- [41] Y. Essam et al., "Predicting suspended sediment load in Peninsular Malaysia using support vector machine and deep learning algorithms," *Sci. Rep.*, vol. 12, no. 1, pp. 1–29, 2022.
- [42] X. Su et al., "Application of DBN and GWO-SVM in analog circuit fault diagnosis," *Sci. Rep.*, vol. 11, no. 1, pp. 1–14, 2021.
- [43] S. El-Sappagh et al., "A multilayer multimodal detection and prediction model based on explainable artificial intelligence for Alzheimer's disease," *Sci. Rep.*, vol. 11, no. 1, pp. 1–26, 2021.
- [44] S.-K. Lin et al., "Classification of patients with Alzheimer's disease using the arterial pulse spectrum and a multilayer-perceptron analysis," *Sci. Rep.*, vol. 11, no. 1, pp. 1–14, 2021.
- [45] M. Mourad et al., "Machine learning and feature selection applied to SEER data to reliably assess thyroid cancer prognosis," *Sci. Rep.*, vol. 10, no. 1, pp. 1–11, 2020.
- [46] M. Nasser et al., "Ambulatory seizure forecasting with a wrist-worn device using long-short term memory deep learning," *Sci. Rep.*, vol. 11, no. 1, pp. 1–9, 2021.
- [47] T. D. Pham, "Time-frequency time-space long short-term memory networks for image classification of histopathological tissue," *Sci. Rep.*, vol. 11, no. 1, pp. 1–12, 2021.
- [48] R. Nakamura et al., "Prognostic prediction by hypermetabolism varies depending on the nutritional status in early amyotrophic lateral sclerosis," *Sci. Rep.*, vol. 11, no. 1, pp. 1–10, 2021.
- [49] D. Bychkov et al., "Deep learning based tissue analysis predicts outcome in colorectal cancer," *Sci. Rep.*, vol. 8, no. 1, pp. 1–11, 2018.
- [50] S. Bolourani et al., "A machine learning prediction model of respiratory failure within 48 hours of patient admission for COVID-19: Model development and validation," *J. Med. Internet Res.*, vol. 23, no. 2, 2021, Art. no. e24246.
- [51] H. Wu et al., "The value of oxygen index and base excess in predicting the outcome of neonatal acute respiratory distress syndrome," *Jornal de Pediatria*, vol. 97, pp. 409–413, 2021.
- [52] L. Shen and Y. Wang, "TCCT: Tightly-coupled convolutional transformer on time series forecasting," *Neurocomputing*, vol. 480, pp. 131–145, 2022.
- [53] V. Shankar, E. Yousefi, A. Manashty, D. Blair, and D. Teegapuram, "Clinical-GAN: Trajectory forecasting of clinical events using transformer and generative adversarial networks," *Artif. Intell. Med.*, vol. 138, 2023, Art. no. 102507.
- [54] D. Du, B. Su, and Z. Wei, "Preformer: Predictive transformer with multi-scale segment-wise correlations for long-term time series forecasting," 2022, *arXiv:2202.11356*.
- [55] X. Wang, D. Pi, X. Zhang, H. Liu, and C. Guo, "Variational transformer-based anomaly detection approach for multivariate time series," *Measurement*, vol. 191, 2022, Art. no. 110791.
- [56] S. Tuli, G. Casale, and N. R. Jennings, "TranAD: Deep transformer networks for anomaly detection in multivariate time series data," 2022, *arXiv:2201.07284*.
- [57] R. Fan, J. Li, W. Song, W. Han, J. Yan, and L. Wang, "Urban informal settlements classification via a transformer-based spatial-temporal fusion network using multimodal remote sensing and time-series human activity data," *Int. J. Appl. Earth Observ. Geoinf.*, vol. 111, 2022, Art. no. 102831.
- [58] M. Liu et al., "Gated transformer networks for multivariate time series classification," 2021, *arXiv:2103.14438*.
- [59] A. Dosovitskiy et al., "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proc. 9th Int. Conf. Learn. Represent.*, 2021.
- [60] L. Dong, S. Xu, and B. Xu, "Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2018, pp. 5884–5888.
- [61] K. Zeng and I. Paik, "A lightweight transformer with convolutional attention," in *Proc. IEEE 11th Int. Conf. Awareness Sci. Technol.*, 2020, pp. 1–6.
- [62] N. Moritz, T. Hori, and J. Le, "Streaming automatic speech recognition with the transformer model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, 2020, pp. 6074–6078.
- [63] S. Huang et al., "HitAnomaly: Hierarchical transformers for anomaly detection in system log," *IEEE Trans. Netw. Service Manage.*, vol. 17, no. 4, pp. 2064–2076, Dec. 2020.
- [64] L. Ilias and D. Askounis, "Explainable identification of dementia from transcripts using transformer networks," *IEEE J. Biomed. Health Inform.*, vol. 26, no. 8, pp. 4153–4164, Aug. 2022.
- [65] R. Pramanik, S. Dey, S. Malakar, S. Mirjalili, and R. Sarkar, "Topsis aided ensemble of CNN models for screening COVID-19 in chest X-ray images," *Sci. Rep.*, vol. 12, no. 1, 2022, Art. no. 15409.
- [66] K. Murugadoss et al., "Building a best-in-class automated de-identification tool for electronic health records through ensemble learning," *Patterns*, vol. 2, no. 6, 2021, Art. no. 100255.
- [67] V. Simic, A. E. Torkayesh, and A. I. Maghsoodi, "Locating a disinfection facility for hazardous healthcare waste in the COVID-19 era: A novel approach based on Fermatean fuzzy ITARA-MARCOS and random forest recursive feature elimination algorithm," *Ann. Operations Res.*, pp. 1–46, 2022.
- [68] F. Saberi-Movahed et al., "Decoding clinical biomarker space of COVID-19: Exploring matrix factorization-based feature selection methods," *Comput. Biol. Med.*, vol. 146, 2022, Art. no. 105426.
- [69] R. Pramanik, S. Sarkar, and R. Sarkar, "An adaptive and altruistic PSO-based deep feature selection method for pneumonia detection from chest X-rays," *Appl. Soft Comput.*, vol. 128, 2022, Art. no. 109464.