Universiteit van Amsterdam

UNIVERSITY OF AMSTERDAM

Amsterdam UMC
University Medical Centers

Master Thesis

Early prediction of bronchopulmonary dysplasia in preterm infants using temporal clinical data

Author:	Joris Pieter van der Vorst
Student number:	10210717
Email:	joris@jvandervorst.nl
Telephone:	06 17506399

daily supervisor:	dr. F.C. Bennis
email:	f.c.bennis@amsterdamumc.nl
telephone:	020 56664322
senior tutor:	prof. dr. J Oosterlaan
email:	j.oosterlaan@amsterdamumc.nl
telephone:	020 5662008

Emma Neuroscience Group
Amsterdam UMC — Department of Pediatrics

September 29, 2023

Abstract

Introduction. Bronchopulmonairy dysplasia (BPD) is one of the most common adverse outcomes in extreme preterm infants. Treatment with postnatal corticosteroids may be benificial, but is associated with a number of side effects, such as gastrointestinal perforation, cerebral palsy and neurodevelopmental impairment and is therefore only indicated in infacts with a high risk of developing BPD. The use of temporal clinical data in combination with machine learning model may be able to predict BPD at an early stage.

Methods. This retrospective cohort study included preterm infants (GA $<$ 30) weeks admitted to the NICU of an academic hospital between 2009 and 2015. Temporal data for the first 7 days (SpO2 and FiO2 values recorded every minute) were recorded in combination with clinical and demographic data (gestational age, birth weight, Apgar score, antenetal use of corticosteroids, sex and multiple births). Outcome of interest was the diagnosis of BPD at 36 weeks PMA.

Three types of predictive models were compared, a traditional machine learning model using only the data recorded around birth, a machine learning model that was able to access this static data in combination with predetermined clinically relevant features and a third neural network model (ConvAE-LSTM) that was able to process the temporal data directly.

Results. The best performing model using the static dataset achieved an AUC of 0,77 (SD 0,05), an accuracy of 0,80 (SD 0,04) and an F1-score of 0,21 (SD 0,15). The model that was able to use the temporal features attained an AUC of 0,81 (SD 0,07), an accuracy of 0,83 (SD 0,02) and an F1-score of 0,44 (SD 0,12). The best performing neural network model was able to predict BPD with an AUC of 0,83, and accuracy and f1-score of 0,85 and 0,46 respectively. The neural network model performed significantly higher than the logistic regression model using only static data (AUC + 0.063 p = 0,014), but did not perform better than the model using manually created temporal features (AUC + 0,022 $p = 0.549$).

Conclusions. In this thesis, the use of a neural network model on temporal data improved the prediction of bronchopulmonairy dysplasia compared to a logistic model using only static data, however this difference loses its statistical significance when the intermediate step of manually created temporal features is taken into account. This model in this study was able to achieve comparable results to other predictive models, while lacking the ability to access information on the type of respiratory support, showing the potential of the use of temporal data.

Contents

1

Introduction

1.1 Bronchopulmonairy dysplasia

Bronchopulmonairy dysplasia BPD is one of the most common adverse outcomes in extreme preterm infants [1]. Although the pathology is not yet fully understood, it appears to be caused by a combination of immature lung development and injury due to long term respiratory support. [1] The incidence of BPD increases with lower gestational age (GA) and is estimated to be around 48 - 68 % for all infants born before 28 weeks GA. Infants with BPD often have long term respiratory morbidity and BPD is associated with neurodevelopmental impairment, such as a lower cognitive ability and motor delay [2, 3]. It has been postulated that lowering the incidence of BPD could be the key to improving cognitive outcomes in very and extremely preterm infants [4].

BPD is defined by the need for supplemental oxygen at 36 weeks postmenstral age (PMA) and is graded by the mode of respiratory support [5]. And although, by this definition, a definite diagnosis is not set until 36 weeks PMA, there is a need for early identification of the risk of BPD in order to promptly start targeted interventions. Two Cochrane meta-reviews showed that systemic postnatal corticosteroids reduced the risk for the development of BPD, with a slightly larger effect for early ($<$ 7 days) start of treatment RR 0.89, 95% CI 0.84 to 0.94) compared to a later ($\geq$ 7 days) start (R 0.85, 95% CI 0.79 to 0.92) [6]. However, treatment with postnatal corticosteroids has a number of side effects, such as gastrointestinal perforation, cerebral palsy and neurodevelopmental impairment [6, 7]. Identifying which preterm infants are most at risk may aid to only treat the most high risk infants and not expose others to these side effects.

For this aim, a number of predictive models have been developed. A recent systematic review looked at 65 studies that aimed to predict BPD[8]. The 11 externally validated models that included data up to 7 days after birth achieved a median AUC of 0,81 (range 0,76 – 0,85). These models mainly used some form of multivariable logistic regression of routinely collected data around birth, such as gestational age, birthweight, sex, and information on the type of respiratory support at day 7.

1.2 Temporal data in clinical prediction models

Besides these features collected at birth, temporal data of vital parameters, such as heart rate (HR), blood pressure (RR), Fraction of inspired oxygen (FiO2), Peripheral oxygen saturation (SpO2) and settings for respiratory support are often recorded during admission to the neonatal intensive care unit (NICU). Due to the high amount of datapoints, it is difficult to make sense of this data without transforming it in some way. Aggregating these data points over time provides a rich dataset that may provide insight into the pulmonary function of these infants. The use of these type of temporal features have already been used in adult ICU settings to predict clinical outcomes using both machine learning and deep learning models. For example, extracted temporal variables such as last FiO2 and mean Tidal Volume had a high predictive value for successful extubation in patients with COVID-19 pneumonia [9] and a neural network model that is able to handle temporal data was able to predict sepsis and myocardial infarction in a public ICU dataset (MIMIC-III) with an AUC of 0.876 and 0.823 respectively [10]. Currently no known model has used this type of temporal data for the prediction of BPD.

Research question

The aim of this thesis is to predict the development of bronchopulmonary dysplasia in a cohort of preterm infants admitted to the neonatal intensive care unit (NICU) and to investigate whether the use of temporal data of the first week of admission will improve this prediction over static data present at birth.

The research question of this project will be: *What is the added value of temporal data for the prediction of bronchopulmonary dysplasia in preterm infants?*

In order to answer this question, the following subquestions will be investigated:

- *What is the accuracy of a machine learning model for prediction of BPD that only uses data that is collected at birth (baseline)?*
- *To what extend does the accuracy of the prediction improve when statistical summaries of temporal clinical data of the first week after birth are added to the baseline data?*

- *How does the performance a neural network that is used to directly process the temporal clinical data compare to a model that uses a summary of this data?*

2

Methodology

2.1 Data collection

This thesis uses a retrospective cohort of preterm infants with GA $<$ 30 weeks admitted to the NICU of an academic hospital (AMC, Amsterdam) in the period of 2009 to 2015. Exclusion criteria were no parental consent, mayor congenital anomalies, admission to the research center $>$ 24 hours after birth and death before 36 weeks PMA. Collection of this data was approved by the local medical ethics board, reference number W21_516#21.569. Static data on these patients includes characteristics recorded just after birth (gestational age, birth weight, Apgar score, antenetal use of corticosteroids, sex and multiple births). The temporal data consists of two respiratory variables, Peripheral oxygen saturation (SpO2) and fraction of inspired oxygen (FiO2)) recorded every minute.

2.2 Data preparation

In order to ensure a fair comparison between the different models, only patients that did not have any missing data for the data recorded at birth and did have at least some temporal data available in the first week after birth were used in this analysis. The continuous features recorded at birth (gestational age, birth weight, Apgar score) were normalized by subtracting the mean of the data and dividing by the standard deviation. Categorical data was kept as is. The temporal data for the first week was extracted and transformed in a number of ways: Firstly, FiO2 is recorded manually in the system every hour. In order to get a values for each minute, these values were imputed forward to the next value or 180 minutes into the future, allowing for a maximum of two subsequent missed entries.

Secondly, because very low SpO2 values are not reliable, all values below 50 % were set at a minimum of 50%.

After cleaning of the data, the temporal data was further transformed in two different ways for the two temporal models:

For the models using a statistical summary of the temporal data, a number of summary statistics for the SpO2 and FiO2 variables were developed in collaboration with neonatologists of the Amsteram UMC hospital. These included the mean and variance of SpO2 and FiO2, $SpO2/FiO2$ [11], hypoxia [12] and hyperoxia. Hypoxia was defined as the percentage points of SpO2 below the minimum target oxygenation of 90 % and hyperoxia was defined as the percentage points of SpO2 above the maximum target oxygenation of 95%, if the faction of inspired oxygen was higher than room air (21%).

$$Hypoxia = \frac{\sum_{i=0}^{i=N_{timesteps}} 90(\%) - SpO2_{[i]} \ \text{ if } \ SpO2_{[i]} < 90\% \ \text{ else } \ 0}{N_{timesteps}}$$

$$Hyperoxia = \frac{\sum_{i=0}^{i=N_{timesteps}} SpO2_{[i]} - 95(\%) \ \text{ if } \ SpO2_{[i]} > 95\% \ \& \ FiO2 > 21\% \text{ else } \ 0}{N_{timesteps}}$$

These summary statistics were calculated for each day separately, resulting in 8 summary features for each day. However, this methods results in a high dimensionality of the data in comparison with the total number of patients (8 features * 7 days = 56 total features for around 500 patients). The number of total features per day was reduced using principle component analysis (PCA). A scree plot was used to determine the optimal amount of components for use in the final model.

For the neural network model, all missing temporal values were imputed by 0, the data was normalized using the mean and standard deviation over all temporal data of all patients and the data was divided up into chunks of 120 minutes. Smaller sets of temporal data were padded with zero's at the end in order to ensure an equal size of data for each patient.

2.3 Models

For each research question we developed a separate model. These models then used the temporal data up to 7 days after birth in order to give a prediction of the development of BPD for each day.

2.3.1 Baseline model

For the baseline model two types of traditional machine learning models were used: A logistic regression (LR) and a random forest (RF) model. These were both solely trained on the data recorded at birth and therefore did not change for the various days.

2.3.2 Summary statistics model

The models with summary statistics also used the same traditional machine learning models. These models used the same data from the baseline model, combined with the temporal features for each available day. This means that the model for day 1 is able to use the data of the first day, the model for day 2 the temporal data of the first two days continuing up to the model for day 7 that is able to use all available features. In development of these models both the dataset with manually transformed features and the PCA were used to find the best performing model. Recursive Feature Elimination (RFE) was used to find the optimal number of predictive features [13].

2.3.3 Convolutional Autoencoder Long Short-Term Memory Network model

The neural network model consists of a variant of an Long Short-Term Memory (LSTM). Long short-term memory models[14] are a type of Recurrent Neural Network that is used for the classification of time series data. These types of models are able to take in a number of time steps one at a time and output a classification of the entire time series after the final time step. However, one downside of these models is that they suffer from vanishing gradients [15]. Due to these vanishing gradients the model tends to "forget" the beginning of a temporal sequence if the input data is too long. Therefore it is not possible to load in the entire data set of 10.080 time steps (7 days * 24 hours * 60 minutes) at once. In order to reduce the total number of time steps a CNN autoencoder is used to form a Convolutional Autoencoder Long Short-Term Memory Network (ConvAE-LSTM) [16]. This model consists of an encoder and decoder. The encoder takes in a certain number of time steps (in our case 120 minutes) and reduces this to to a smaller latent vector (embedding). The decoder will then take in this latent vector and try to reproduce to original time series of 120 minutes as well as possible. After this model has been trained, the embedding from the encoder will be a smaller representation of the original data. A visualisation of the model is shown in figure 2.1 and samples of the reproductions can be found in appendix A. Using this approach, the number of time steps in the LSTM classification model can be reduced to a maximum of 84.

Training of this model was done in three steps: Firstly, the autoencoder was trained on the chunks of 120 minutes. Secondly, the embeddings from this autoencoder were fed to the LSTM classification model that predicted the occurrence of BPD. Diagram 3.1 shows how the temporal data flows through the model. Finally, this LSTM classification model was combined with a fully connected neural network that took in the data recorded at birth as input. This resulted in one combined model, compared to the 7 seperate summary statistics models. In order to for example predict the risk at day 1, this one model is given access to only the temporal data of day 1.

2.4 Outcome

Outcome of interest is the diagnosis of BPD at 36 weeks PMA, using the diagnostic criteria of the time [17]. This outcome was recorded as a binary scale, grade of BPD was not used. The primary classification metric to evaluate the accuracy of the different models was the Area under the Receiver Operating Curve (AUC)). Secondarily accuracy and F1-score were also recorded. F1-score was used this better reflects model performance in unbalanced cohorts because it accounts for both false positives and false negatives, whereas accuracy will be high for a model that predicts no disease for all cases in a cohort with a low incidence.

$$F1\ Score = 2 * \frac{Precision * Recall}{Precision + Recall} = \frac{2 * True\ Positive}{2 * True\ Positive + False\ Postive + False\ Negative}$$

2.5 Validation and comparison of model

In order to give an accurate assessment of the accuracy on unseen data, the traditional machine learning models were trained using 5-fold stratified cross validation. Using cross validation, models are trained a number times, each time leaving out a part of the data for validation. The performance on these validation sets is recorded and mean and standard deviation over the five different training runs is reported.

The design and training of the neural network models is an iterative process. The model is trained a number of times with a different architecture each time in order to see what type of model design suits the data the best. It is therefore not possible to perform an automated training and validation loop such as cross validation. In order to still have

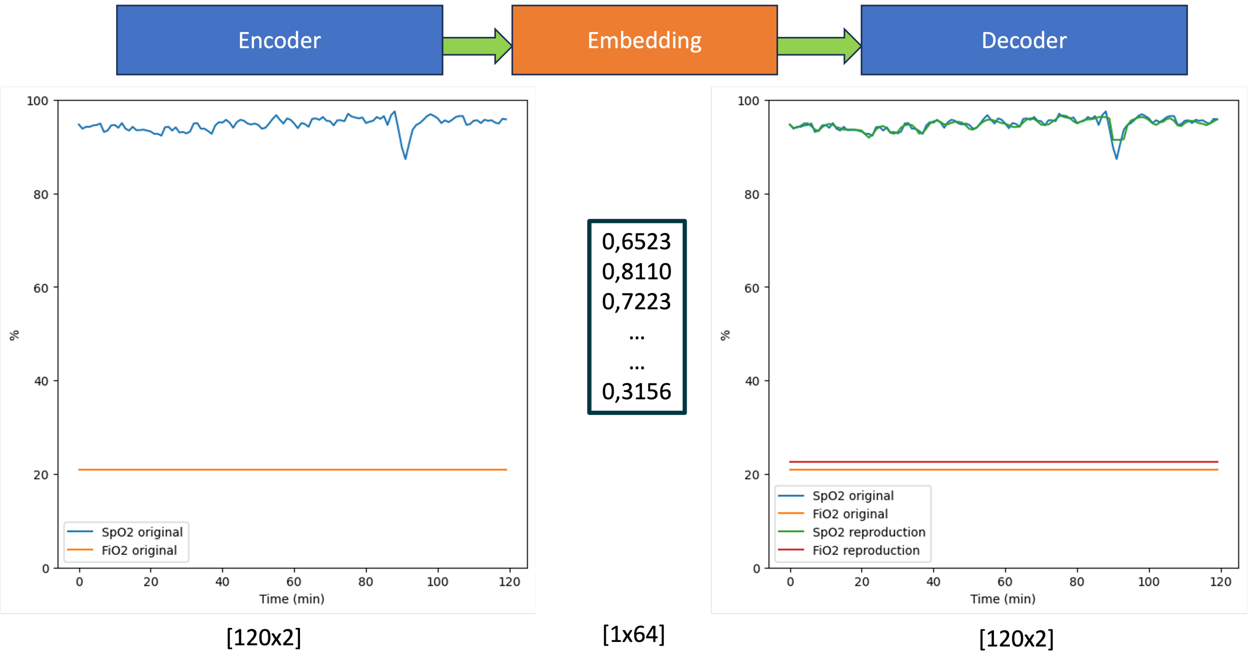


Figure 2.1: Visualisation of autoencoder model:
The original SpO2 is in blue and the original FiO2 is in orange. The orignal data of 120 minutes and two features $[120 * 2]$ is compressed by the encoder in order to create an embedding of size $[64 * 1]$. The decoder then uses this compressed embedding to reproduce the original signal of shape $[120 * 2]$. The signals in green and red represent the reproduced SpO2 and FiO2 after the compression step and are well fitted to the orignal data in this sample.

an valid estimation of the model performance on unseen data, the dataset is split into a training, validation and test set, consisting of 60%, 20% and 20% of the data respectively. The training and validation set are used for this iterative process of improving the model. The test set is kept aside and only used once after the model design is finalized to quantify the model performance.

Models that both use cross validation will be compared using a paired t-test of the AUC values. If one model used cross validation and one the test set, the 5 values of the cross validation are used to calculate a mean and standard deviation and perform a one sample t-test.

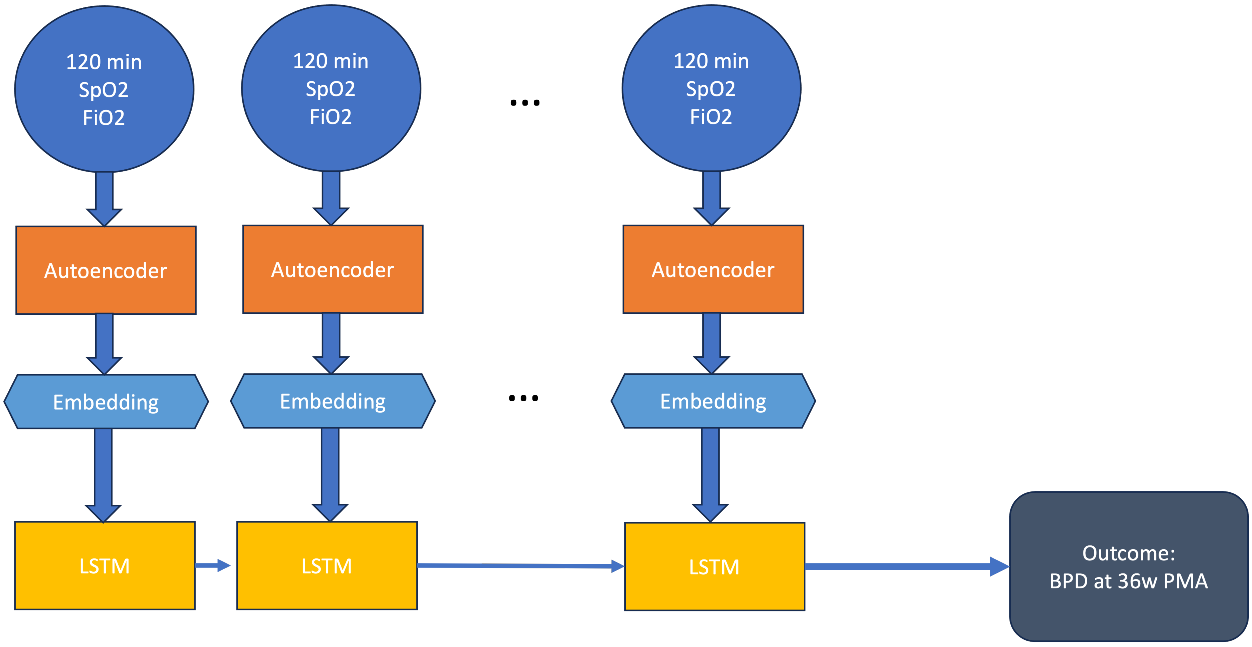


Figure 2.2: Diagram of temporal data flow through the autoencoder and LSTM model. Temporal data of 120 minutes is compressed using the encoder and each time step is fed sequentially to the LSTM model. After all time steps are given to the LSTM model, the last output of the model is used to predict the occurence of BPD at 36 weeks GA.

3

Results

3.1 Cohort characteristics

Of the 584 patients in the cohort that were alive at 36 weeks PMA, 7 did not have any data available for this thesis and 129 did have data recorded at birth but lacked temporal data for the first 7 days after birth. In order to fairly compare all models and to keep the same training and test splits, both of these previously mentioned groups were excluded, resulting in a final cohort size of 455. A full count of all in- and exclusions can be found in figure 3.1.

The mean gestational age (GA) was 27,9 (SD 1,5) weeks and mean birth weight was 1069,8 (SD 265,1) gram. In this cohort 87 patients (19,1 %) were diagnosed with BPD at 36 weeks PMA. gestational age, birth weight, small for gestional age and Apgar score differed significantly between the patients with and without BPD. Patients with BPD were born 1,3 weeks earlier (26.9 (SD 1.3) vs 28.2 (SD 1.4), p=<0.001) and weighed 228,7 grams less at birth (884.9 (SD 203.8) vs 1113.6 (SD 259.2), p=<0.001). In the cohort of patients with BPD there were more children born small for gestational age (24,1 % vs 12,5% p=0,01) and a higher proportion received antenatal steroids (89,9% in the group without BPD and 97,7 in the group with BPD, p=0,0335). A summary of all demographic and clinical characteristics can be found in table 3.1

3.2 Baseline model

Using the clinical and demographic features recorded at birth, a the best performing model was able to predict BPD with an AUC of 0,77 (SD 0,05), an accuracy of 0,80 (SD 0,04) and an F1-score of 0,21 (SD 0,15). The best model used logistic regression on all available

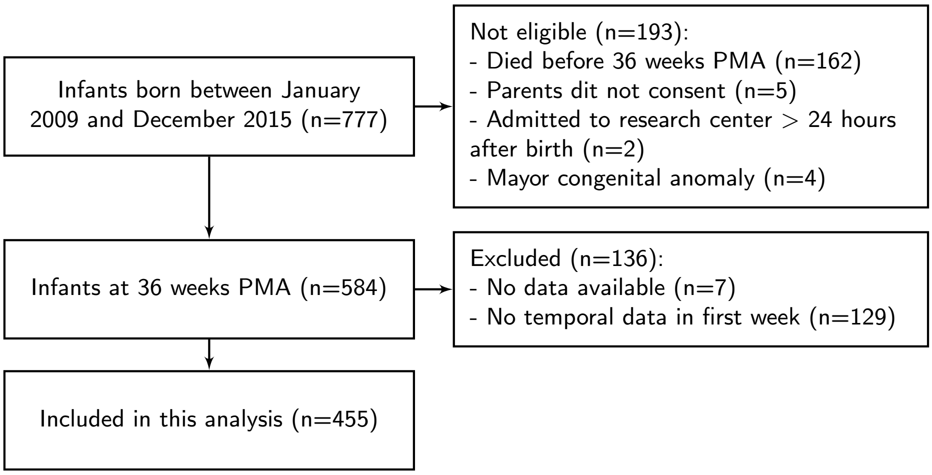

Infants born between January 2009 and December 2015 (n=777)
Not eligible (n=193):
- Died before 36 weeks PMA (n=162)
- Parents dit not consent (n=5)
- Admitted to research center > 24 hours after birth (n=2)
- Mayor congenital anomaly (n=4)
Infants at 36 weeks PMA (n=584)
Excluded (n=136):
- No data available (n=7)
- No temporal data in first week (n=129)
Included in this analysis (n=455)

Figure 3.1: Flow diagram of the cohort selection

		Overall	No BPD	BPD	P-Value
n		455	368	87	
Sex, n (%)	Male	253 (55.6)	204 (55.4)	49 (56.3)	0.976
	Female	202 (44.4)	164 (44.6)	38 (43.7)	
Gestational age (weeks), mean (SD)		27.9 (1.5)	28.2 (1.4)	26.9 (1.3)	<0.001
Birth weight (g), mean (SD)		1069.8 (265.1)	1113.6 (259.2)	884.9 (203.8)	<0.001
Number of births, n (%)	1	322 (70.8)	255 (69.3)	67 (77.0)	0.317
	2	122 (26.8)	103 (28.0)	19 (21.8)	
	3	11 (2.4)	10 (2.7)	1 (1.1)	
Small for gestational age, n (%)	No	388 (85.3)	322 (87.5)	66 (75.9)	0.010
	Yes	67 (14.7)	46 (12.5)	21 (24.1)	
Antenatal steroid use, n (%)	No	39 (8.6)	37 (10.1)	2 (2.3)	0.035
	Yes	416 (91.4)	331 (89.9)	85 (97.7)	
Apgar score at 5 minutes, mean (SD)		7.6 (1.6)	7.7 (1.5)	7.2 (2.0)	0.067
BPD, n (%)	0	368 (80.9)	368 (100.0)		
	1	87 (19.1)		87 (100.0)	

Table 3.1: Baseline characteristics of the study cohort. The summary statistics (mean and standard deviation) or the frequency distribution of each variable are shown in the overall population and by the presence/absence of BPD. The statistical tests and p-values relative to BPD are also reported.

Static data

	Logistic Regression	**Random Forest**
Mean AUC-ROC (SD)	**0,77 (0,05)**	**0,75 (0,03)**
Mean Accuracy (SD)	0,80 (0,04)	0,80 (0,03)
Mean F1 Score (SD)	0,21 (0,15)	0,31 (0,11)

Table 3.2: Machine learning models on static data available at birth . Mean and Standard Deviation (SD) are calculated using 5-fold cross validation

parameters. A curve of the mean reciever operator characteristics (ROC) of the best performing model can be seen in figure 3.2. Performance characteristics of the best performing logistic regression and random forest models are shown in table 3.2.

3.3 Manually extracted features

For each patient in the cohort, their SpO2 and FiO2 values of the first 7 days were transformed in the method described in 2.2. Using a scree plot, it was determined that 2 principle components for each day had the best balance between discriminatory ability and number of features. Summary statistics for each individual feature along with univariate t-tests between the cohorts with and without BPD can be found in supplementary table B.4. Similar analysis on the principal components for each day are shown in supplementary table B.1.

After the manually extracted temporal features were added to the static data, the best performing model achieved a mean AUC of 0,81 (SD 0,07), accuracy of 0,83 (SD 0,02) and a F1-score of 0,44 (0,12). The best performing model was a logistic regression model that used recursive feature elimination on the PCA transformed temporal features, the reciever operator characteristics of this model is displayed in figure 3.2.

Due to the random nature of the cross-validation training en evaluation loops, each loop selected different features in the final model. Counting the occurrences in each model (see supplementairy table B.2), it can be noted that both of the principle components for the final days (day 5, 6 and 7) were selected in every fold. In order to provide a more interpretable insight, the best performing model that did not PCA for data compression was selected and selected features for that model are shown in table B.3. This model achieved an AUC of 0.79 (SD 0.09) and was most commonly driven by the use of antenatal steroids, birth weight and the daily mean SpO2 of day 6 en 7 (all mentioned features were selected in 2 of the 5 folds).

Static and temporal data

	Logistic Regression	**Random Forest**
Mean AUC-ROC (SD)	**0,81 (0,07)**	**0,78 (0,07)**
Mean Accuracy (SD)	0,83 (0,02)	0,83 (0,03)
Mean F1 Score (SD)	0,44 (0,12)	0,42 (0,15)

Table 3.3: Machine learning models on static data and temporal data at day 7 . Mean and Standard Deviation (SD) are calculated using 5-fold cross validation

Temporal data only

	Logistic Regression	**Random Forest**
Mean AUC-ROC (SD)	**0,79 (0,09)**	**0,76 (0,07)**
Mean Accuracy (SD)	0,82 (0,03)	0,82 (0,03)
Mean F1 Score (SD)	0,42 (0,12)	0,41 (0,16)

Table 3.4: Machine learning models using only temporal data at day 7, without the use of static data . Mean and Standard Deviation (SD) are calculated using 5-fold cross validation

Figure 3.3 shows a slight increase in performance of the logistic regression model over time, although none of the differences between the days were statistically significant.

Finally, a logistic regression and random forest model was trained on only the temporal data up to day 7, so without static features such as gestational age and birth weight. This resulted in a comperable AUC of 0,79 (SD 0,09) (additional performance characteristics in table 3.4).

3.4 Neural network models

For the neural network models that used temporal data (the ConvAE-LSTM and combined model) the temporal data of each patient was first compressed by feeding the encoder part of the autoencoder chunks of 120 minutes of SpO2 and FiO2 data at a time. A number of sample reproductions of this temporal data by the autoencoder model can be found in appendix A. The model that used only static data attained a AUC of 0,75 on the test set, the model using only temporal data (ConvAE-LSTM) bettered this with an AUC of 0,80 and the best performance was achieved by the neural network model that combined data from the static and temporal neural networks, resulting in an AUC of 0,83, with an accuracy and f1-score of 0,85 and 0,46 respectively. The performance metrics of all neural network models can be found in table 3.5 and the reciever operator characteristics (ROC) curves of all models is shown in figure 3.2. Figure 3.3 shows the performance of the LSTM and combined model for each day of data that is added.

Neural network models

	Static	**ConvAE-LSTM**	**Combined**
AUC-ROC	**0,75**	**0,81**	**0,83**
Accuracy	0,80	0,82	0,85
F1 Score	0,31	0,46	0,46

Table 3.5: Accuracy of neural network models on temporal data at day 7 (ConvAE-LSTM) or a combination of static and temporal data (Combined). Performance was measured on the hold out test set

3.5 Comparison of models

When using a paired t-test, the model using static and temporal data up to day 7 does not perform significantly better than the model using only static features (difference in AUC: + 0,042 p = 0,2363). Using a one sample t-test the AUC value of the best neural network model was significantly higher than the logistic regression model using only static data (absolute difference + 0.063 p = 0,014). Compared to the logistic regression that used static data and temporal data up to 7 days, the neural network model did not achieve a significantly better result (difference in AUC + 0,022 p = 0.549).

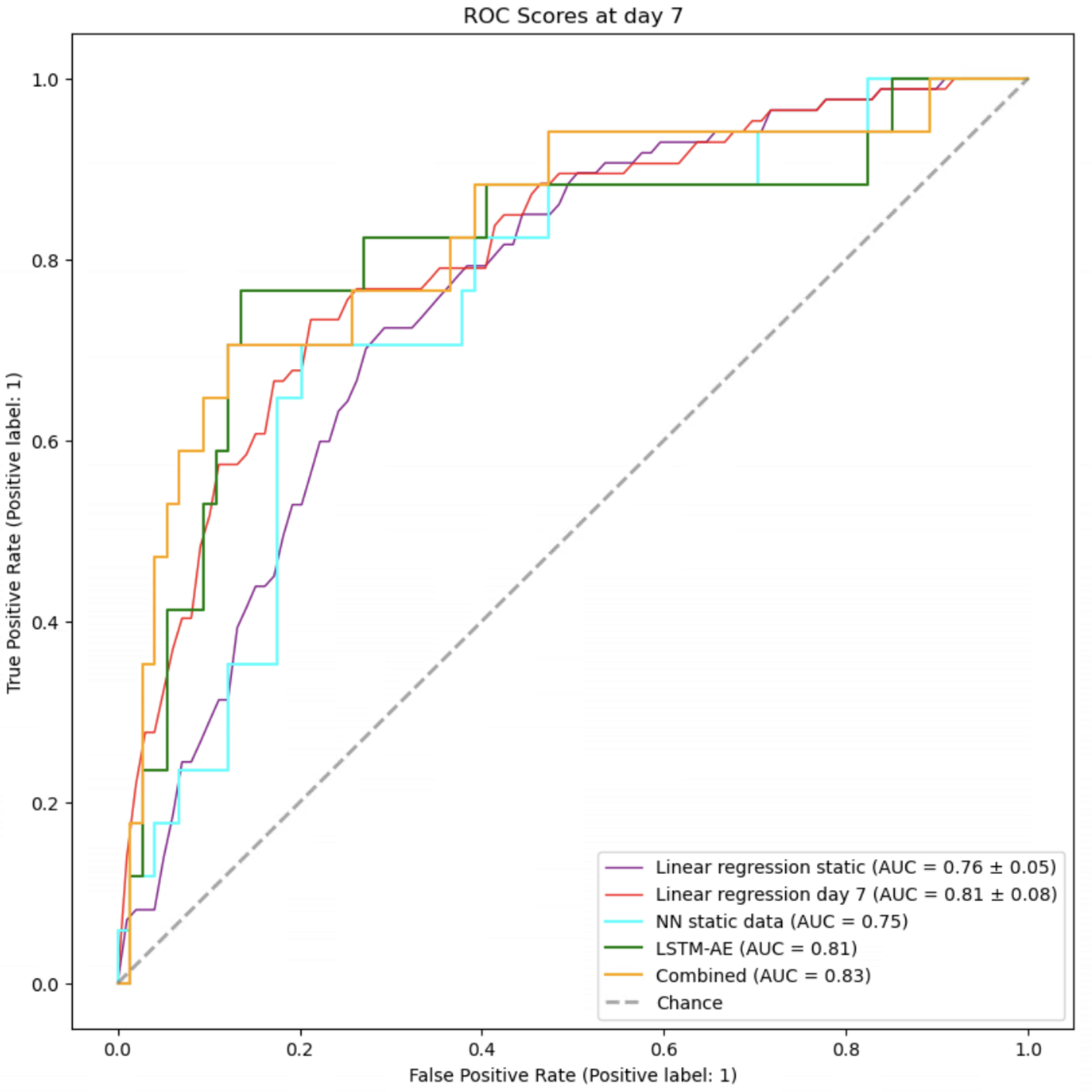


Figure 3.2: ROC score of all model at day 7. When reported with SD, performance was calculated using using 5-fold cross validation, when reported without SD, performance was measured on the hold out test set

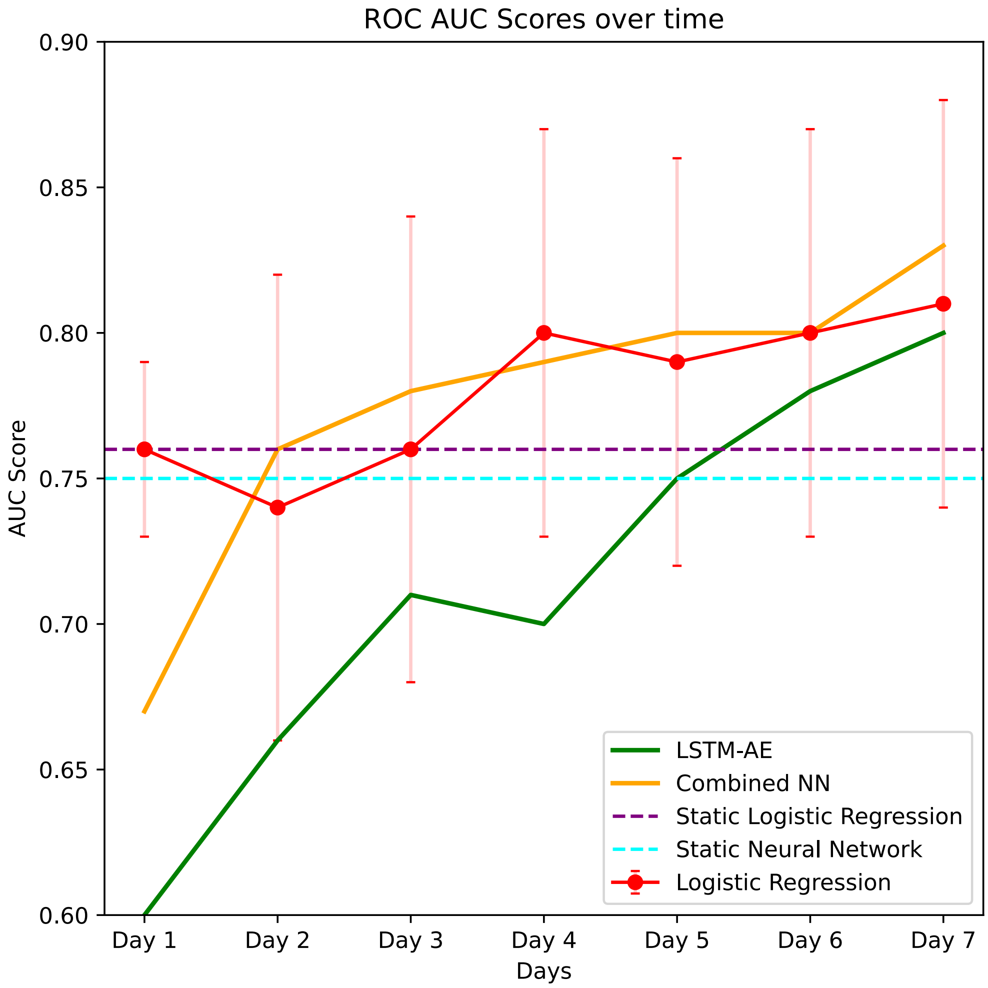


Figure 3.3: ROC score of all models for each day

4

Discussion and Conclusion

In this thesis, the addition of temporal data improved the prediction of bronchopulmonairy dysplasia. A neural network that used a weeks worth of SpO2 and FiO2 values recorded every minute to the clinical and demographic data recorded at birth was able to significantly outperform a logistic regression model that only used this static data (AUC 0,83 vs 0,77 (SD 0,05), p = 0,014).

However, this difference loses its statistical significance when the intermediate step of manually extracted temporal features is take into account. The logistic regression with manually extracted features did not improve in accuracy when compared to the logistic regression on the static data (0,81 (SD 0,07) vs 0,77 (SD 0,05) p = 0,236) and the Convolutional Autoencoder Long Short-Term Memory Network neural network that included static data was not better than the logistic regression model using static and temporal data at day 7 (0,83 vs 0,81 (SD 0,07), p = 0,549). Moreover, the models in this study were not superior to existing predicitve models that did not use temporal clinical data. The 2011 and 2021 NICHD predictive models at 7 days had an AUC of 0,771 and 0,692 respectively and a review of 65 studies that predict BPD found a median AUC of 0,81 (range 0,76 – 0,85).

Fortunately, there are still a number of indications that the use of temporal clincal data can lead to models with a better performance than currently exist. Firstly, in this study, the temporal only models preformed similarly to the models that also included static data. Demographic factors such as gestational age and birth weight are mayor risk factors for the development of BPD [8, 18–20]. To illustrate, in a NICHD multicenter cohort of 9575 extremely preterm infants born between 2003 and 2007, the overall incidence of BPD according to GA rose from 23% for infants born at 28 weeks GA to 69% for infants born at 24 weeks. In this study the best performing logistic regression model that used both static and temporal data did not outperform the model using temporal data only (AUC 0,81 (0,07) vs 0,79 (0,09), p= 0,71) and the result from the Convolutional Autoencoder Long Short-Term Memory Network (ConvAE-LSTM) model that only used temporal data was similar to the neural network model that used temporal and static data (AUC 0,80 and 0,83 on the test set respectively). The fact that the temporal only models were able to perform to such an extend without using these risk factors is an indication that this temporal data carriers highly usefull information. Additionally, the cohort in this study had a much lower proportion of BPD than earlier cohorts that are used to make predictive models. The proportion on BPD in this cohort was 19,1%, while the 2011 and 2021 NICHD models were created and evaluated on cohort with an incidence of BPD of 61% and 60% respectively. This much lower percentage of positive cases has a reducing effect on a models discriminatory ability due to the imbalance of the dataset and the tendency for a model to attain a bias of predicting no disease. Furthermore, the cohort in this study was an order of magnitude much smaller. The 2011 NICHD cohort included 3,636 infants and the 2021 included 9,181, while this thesis had 455 patient to to train and evaluate the models. Use of temporal data may be an efficient way to create predictive models in terms of number of patients that are needed. Finally, the models in this study were able to have comparative results without access to features such as the modality of respiratory support. In a disease that, later in the life of the infant, graded on the type of respiratory support[21], information on the use of respiratory support is highly informative on the state of the patient. In both the 2011 and 2021 predictive models, the type of respiratory support at 7 days was the most contributing factory in the prediction of BPD. In a review of 65 studies spanning 158 predictive models for BPD, type of respiratory support was included as a predictive feature in 94 of the models [8]. In conclusion, there were a number of hindrances for good performance in the development of the models in this study and it is encouraging that the model that used temporal data were still able to provide a decent discriminatory ability.

When analysing why the neural network model ConvAE-LSTM did not outperform the logistic regression model with manually created temporal features, the samples of the reproduction of the temporal data by the Convolutional Neural Network autoencoder in appendix A may provide insight. The reproductions in figure A and B show the way the autoencoder is suppose to work. It should be able to capture the complexity of the signal (in this case the variation in SpO2) and filter out the very sharp downward peaks to zero where data is missing. However, this type of exact reproduction is not always achieved. Figure C and D show examples were the autoencoder is not able to capture the complexity of the data. When the SpO2 signal get too variable, the model embedding is not able to capture this and defaults to reproducing the mean SpO2 over that time period. As a result, the signal that is send to the Long Short-Term Memory model does not contain all these periods of saturation drops. Unfortunately, this may lead to clinically relevant information not being passed along from the autoencoder model to the LSTM model. A higher incidence and longer duration of intermittent hypoxemia events are associated with the diagnosis of BPD at 36 week PMA[12] and when information on these events, which present a variations in the SpO2 signal, does not flow through the model, the discriminatory ability of these neural networks are decreased.

There are a number of ways to overcome this problem in future research. Firstly, these types of neural network models improve greatly when provided with larger datasets. The highest perfoming model in this study had a total of 146.563 parameters that can be adjusted when training, of which 128.482 are specific for the autoencoder. Because all these individual parameters have to be adjusted in training, it is likely that a model with more training data is able to learn to capture those complex signals as well. Secondly, it is possible to use the reconstruction loss of the autoencoder as a predictive feature. The autoencoder is most often not able to correctly reproduce the temporal signal at times it is highly variable. Because each signal is compared to it's own reproduction, the measure of how well the autoencoder performs is available when predicting on new unseen data. In anomaly detection models this reproduction loss is often used [22] and physiologically this can also be translated to the prediction of BPD: the reconstruction loss will be high when the SpO2 values are variable, this is most often at times the infant has a high number of hypoxemia events. Therefore the reproduction loss may be able to be used as a proxy measure of the amount of hypoximia events and could be correlated with the occurence of BPD at 36 weeks PMA. Thirdly, the current CNN autoencoder could be replaced by a better architecture. For example, a Temporal convolutional autoencoder for unsupervised anomaly detection[23] has been show to be highly effective in capturing patterns in electrocardiogram (ECG) recordings and was able to detect a number of different cardiac arrhythmias. Furthermore, if an attention layer is used, it may be able to remove the autoencoder architecture altogether,removing the limitation of inaccurate reproductions from the prediction loop entirely. Use of an attention layer may help a Recurrent Neural Network such as an Long Short-Term Memory to learn which parts of the incoming signal needs to be passed onward, instead of focusing on the entire signal. A LSTM model with an attention layer has been shown to predict sepsis and myocardial infarction in an adult intensive care unit with an AUC of 0,876 and 0,823 on the public MIMIC-III dataset [10].

A final way to improve future model performance is to increase the frequency of the recorded SpO2 values. The data is now recorded every minute, but this may lack to capture a high number of hypoxic events, as they are often much shorter that one minute and may already be resolved at the time of the next recording interval [12].

Looking back on the broader context of the prediction of BPD, this study shows potential for the use of temporal data. This thesis was able to show comparable performance to the current literature while using a relatively small cohort with a low incidence and a lack of access to data on the type of respiratory support. There are a number of predictive models that are able to achieve a higher performance with the help of lung ultrasound and or biomarkers [8]. However, most of these have not been validated in external cohort. Additionally, use of temporal data for prediction models match well to the clinical setting, providing information to the care team without the need to perform extra steps, Temporal models can also be constantly updated instead of only at the time these additional diagnostic test are performed.

In conclusion, this thesis showed the potential of the use of temporal data for the prediction of bronchopulmonairy dysplasia. A neural network model on temporal data improved the prediction of bronchopulmonairy dysplasia compared to a logistic model using only static data, however this difference loses its statistical significance when the intermediate step of manually created temporal features is taken into account. This thesis was able to achieve comparable results to the current models predicting BPD at day 7, using a cohort that was an order of magnitude smaller in size and having no available data on the type of respiratory support. Performance of the neural network was hindered by inaccuracies in the reproductions of the autoencoder model. Future studies using a different neural network architecture, larger cohort size and data recorded at a higher frequency may produce superior results and provide physicians with a highly accurate, frequently updated prediction model for bronchopulmonairy dysplasia, without the burden of needing additional tests.

References

1. Jobe, A. Mechanisms of Lung Injury and Bronchopulmonary Dysplasia. *American Journal of Perinatology* **33,** 1076–1078. ISSN: 0735-1631. `http://www.ncbi.nlm.nih.gov/pubmed/27603539` (Sept. 7, 2016).
2. Gilfillan, M., Bhandari, A. & Bhandari, V. Diagnosis and management of bronchopulmonary dysplasia. *BMJ,* n1974. ISSN: 1756-1833. `https://www.bmj.com/lookup/doi/10.1136/bmj.n1974` (Oct. 20, 2021).
3. Singer, L., Yamashita, T., Lilien, L., Collin, M. & Baley, J. A Longitudinal Study of Developmental Outcome of Infants With Bronchopulmonary Dysplasia and Very Low Birth Weight. *Pediatrics* **100,** 987–993. ISSN: 0031-4005. `https://doi.org/10.1542/peds.100.6.987` (2023) (Dec. 1, 1997).
4. Twilhaar, E. S. *et al.* Cognitive Outcomes of Children Born Extremely or Very Preterm Since the 1990s and Associated Risk Factors: A Meta-analysis and Meta-regression. *JAMA Pediatrics* **172,** 361–367. ISSN: 2168-6203. `https://doi.org/10.1001/jamapediatrics.2017.5323` (Apr. 1, 2018).
5. Dysart, K. *et al.* The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants An Evidence-based Approach. *American Journal of Respiratory and Critical Care Medicine* **200,** 751–759. ISSN: 15354970 (2019).
6. Doyle, L. W., Cheong, J. L., Ehrenkranz, R. A. & Halliday, H. L. Early ($<$ 7 days) systemic postnatal corticosteroids for prevention of bronchopulmonary dysplasia in preterm infants. *Cochrane Database of Systematic Reviews* **2018,** CD001145. ISSN: 14651858. `http://doi.wiley.com/10.1002/14651858.CD001145.pub5` (Oct. 24, 2017).
7. Kwok, T. C., Batey, N., Luu, K. L., Prayle, A. & Sharkey, D. Bronchopulmonary dysplasia prediction models: a systematic review and meta-analysis with validation. *Pediatric Research.* Publisher: Springer US, 1–12. ISSN: 0031-3998. `https://www.nature.com/articles/s41390-022-02451-8` (December 2022 Jan. 9, 2023).

8. Romijn, M. *et al.* Prediction Models for Bronchopulmonary Dysplasia in Preterm Infants: A Systematic Review and Meta-Analysis. *The Journal of Pediatrics* **10.** Publisher: The Author(s), 113370. ISSN: 00223476. https://linkinghub.elsevier.com/retrieve/pii/S0022347623000513 (Apr. 2023).

9. Fleuren, L. M. *et al.* Predictors for extubation failure in COVID-19 patients using a machine learning approach. *Critical Care* **25,** 448. ISSN: 1364-8535. https://doi.org/10.1186/s13054-021-03864-3 (2021).

10. Kaji, D. A. *et al.* An attention based deep learning model of clinical events in the intensive care unit. *PLoS ONE* **14.** ISBN: 1111111111, 1–17. ISSN: 19326203 (2019).

11. Lu, X. *et al.* Continuously available ratio of SpO2/FiO2serves as a noninvasive prognostic marker for intensive care patients with COVID-19. *Respiratory Research* **21.** Publisher: Respiratory Research, 1–4. ISSN: 1465993X (2020).

12. Thomas M. Raffay *et al.* Neonatal intermittent hypoxemia events are associated with diagnosis of bronchopulmonary dysplasia at 36 weeks postmenstrual age. *Pediatric Research* **85,** 318–323. https://doi.org/10.1038/s41390-018-0253-z (2023) (Dec. 12, 2018).

13. Guyon, I., Weston, J., Barnhill, S. & Vapnik, V. Gene Selection for Cancer Classification using Support Vector Machines. *Machine Learning* **46,** 389–422. ISSN: 1573-0565. https://doi.org/10.1023/A:1012487302797 (Jan. 1, 2002).

14. Hochreiter, S. & Schmidhuber, J. Long Short-Term Memory. *Neural Computation* **9,** 1735–1780. ISSN: 0899-7667. https://doi.org/10.1162/neco.1997.9.8.1735 (Nov. 15, 1997).

15. Le, P. & Zuidema, W. Quantifying the vanishing gradient and long distance dependency problem in recursive neural networks and recursive lstms. *Proceedings of the Annual Meeting of the Association for Computational Linguistics.* ISBN: 9781945626043, 87–93. ISSN: 0736587X. arXiv: 1603.00423 (2016).

16. Thakur, D., Biswas, S., Ho, E. S. L. & Chattopadhyay, S. ConvAE-LSTM: Convolutional Autoencoder Long Short-Term Memory Network for Smartphone-Based Human Activity Recognition. *IEEE Access* **10,** 4137–4156. ISSN: 2169-3536. https://ieeexplore.ieee.org/document/9668957/ (2023) (2022).

17. Jobe, A. H. & Bancalari, E. NICHD / NHLBI / ORD Workshop Summary. *American Journal of Respiratory and Critical Care Medicine* **163.** ISBN: 1073-449X (Print) 1073-449X (Linking), 1723–1729. ISSN: 1073-449X (2001).

18. Trembath, A. & Laughon, M. M. Predictors of Bronchopulmonary Dysplasia. *Clinics in Perinatology* **39,** 585–601. ISSN: 0095-5108. https://www.sciencedirect.com/science/article/pii/S0095510812000644 (2012).

19. Laughon, M. M. *et al.* Prediction of Bronchopulmonary Dysplasia by Postnatal Age in Extremely Premature Infants. *American Journal of Respiratory and Critical Care Medicine* **183,** 1715–1722. ISSN: 1073-449X, 1535-4970. https://www.atsjournals.org/doi/10.1164/rccm.201101-0055OC (2023) (June 15, 2011).

20. Greenberg, R. G. *et al.* Online clinical tool to estimate risk of bronchopulmonary dysplasia in extremely preterm infants. *Archives of Disease in Childhood - Fetal and Neonatal Edition* **107,** 638–643. ISSN: 1359-2998, 1468-2052. https://fn.bmj.com/lookup/doi/10.1136/archdischild-2021-323573 (2023) (Nov. 2022).

21. Jensen, E. A. *et al.* The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants. An Evidence-based Approach. *American Journal of Respiratory and Critical Care Medicine* **200,** 751–759. ISSN: 1073-449X, 1535-4970. https://www.atsjournals.org/doi/10.1164/rccm.201812-2348OC (2023) (Sept. 15, 2019).

22. Cheng, Z. *et al.* Improved autoencoder for unsupervised anomaly detection. *International Journal of Intelligent Systems* **36.** _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/in 7103–7125. https://onlinelibrary.wiley.com/doi/abs/10.1002/int.22582 (2021).

23. Thill, M., Konen, W., Wang, H. & Bäck, T. Temporal convolutional autoencoder for unsupervised anomaly detection in time series. *Applied Soft Computing* **112,** 107751. ISSN: 15684946. https://linkinghub.elsevier.com/retrieve/pii/S1568494621006724 (2023) (Nov. 2021).

Appendix A

Sample autoencoder reproductions of the temporal data

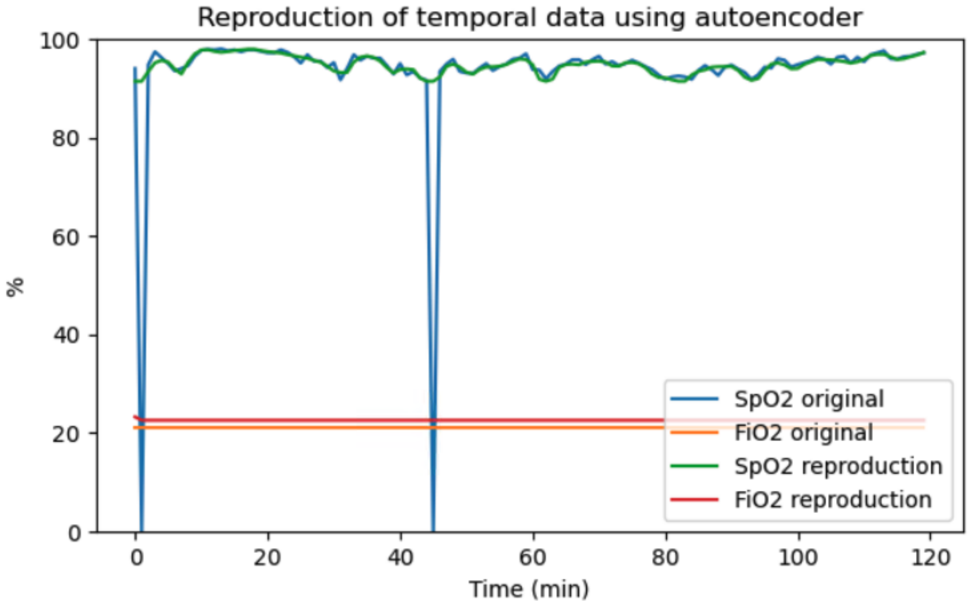


Figure A.1: A

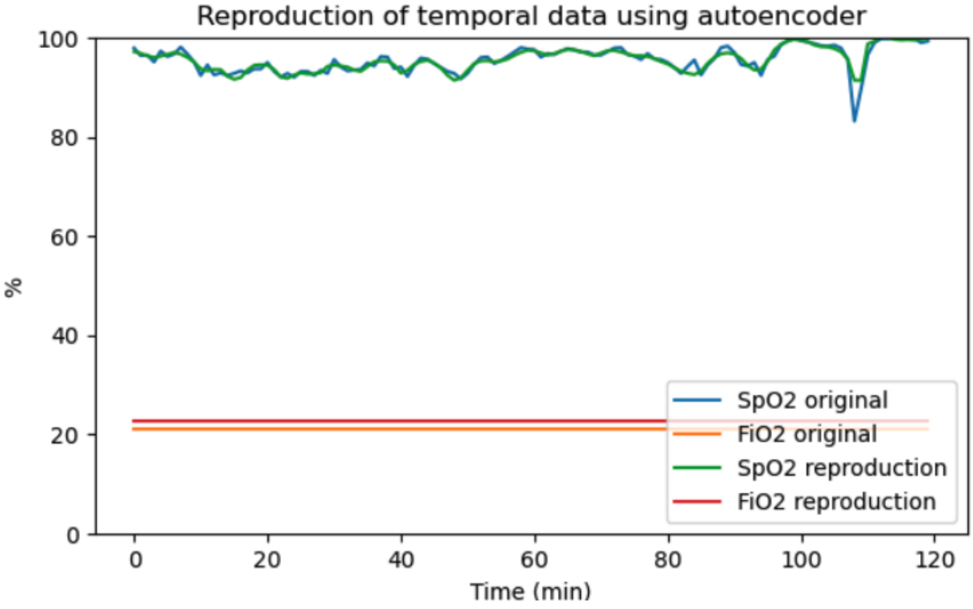


Figure A.2: B

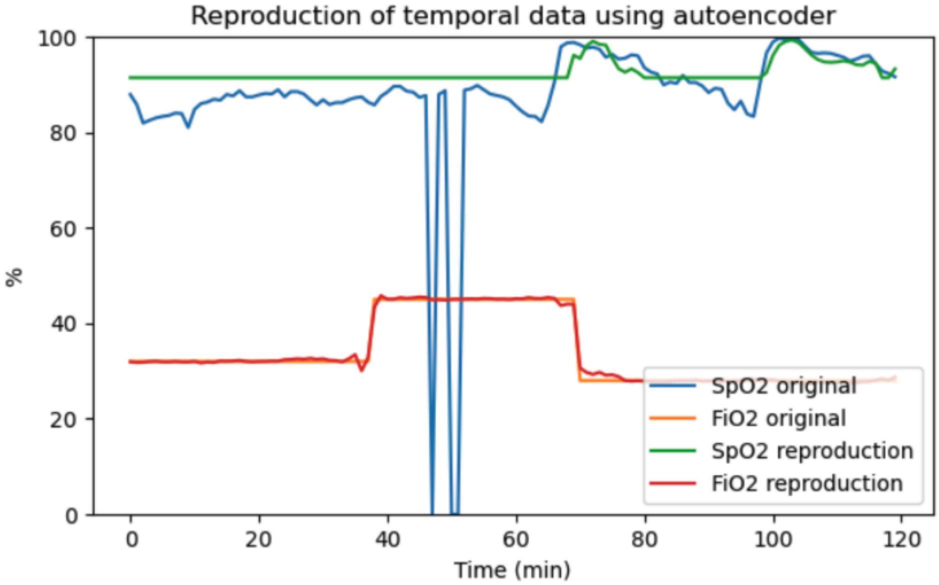


Figure A.2: C

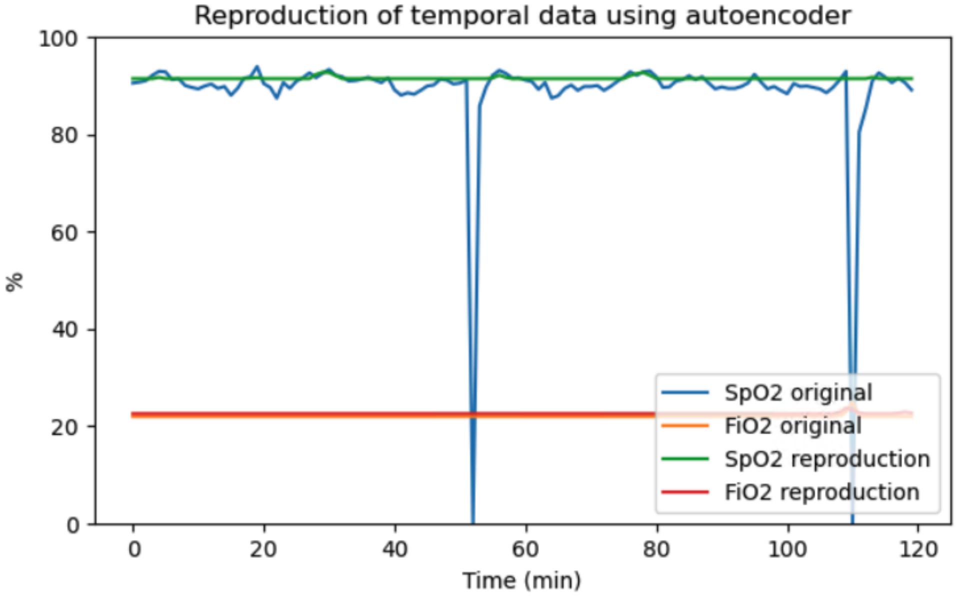


Figure A.3: D

Figure A.4: Examples of reproduction of the temporal data using the autoencoder model The original SpO2 is in blue and the original FiO2 is in orange. The signals in green and red represent the reproduced SpO2 and FiO2 after the compression step. The first two images (A,B) show a good representation of the data by the model. The last two images (C,D) show examples of the model falling back to predicting the mean value at times the signal is highly variable.

Appendix B

Supplementary tables

			Grouped by BPD	
	Overall	No BPD	BPD	P-Value
FiO2_filled_mean_01, mean (SD)	25.3 (6.1)	24.8 (5.6)	27.5 (7.4)	0.002
FiO2_filled_mean_02, mean (SD)	24.4 (5.9)	23.5 (4.8)	28.0 (8.2)	<0.001
FiO2_filled_mean_03, mean (SD)	24.2 (6.3)	23.1 (4.3)	28.5 (10.4)	<0.001
FiO2_filled_mean_04, mean (SD)	23.6 (5.6)	22.5 (3.4)	28.2 (9.4)	<0.001
FiO2_filled_mean_05, mean (SD)	23.4 (5.5)	22.2 (3.4)	28.3 (8.8)	<0.001
FiO2_filled_mean_06, mean (SD)	23.0 (4.8)	22.0 (3.1)	27.1 (7.4)	<0.001
FiO2_filled_mean_07, mean (SD)	22.7 (4.3)	21.9 (3.1)	26.0 (6.4)	<0.001
FiO2_filled_var_01, mean (SD)	67.3 (178.1)	59.6 (183.3)	99.0 (151.0)	0.038
FiO2_filled_var_02, mean (SD)	26.7 (102.2)	19.9 (101.8)	54.8 (99.3)	0.004
FiO2_filled_var_03, mean (SD)	46.7 (540.6)	18.8 (114.7)	161.9 (1201.6)	0.270
FiO2_filled_var_04, mean (SD)	13.1 (46.8)	6.9 (27.7)	38.9 (85.3)	0.001
FiO2_filled_var_05, mean (SD)	29.2 (174.3)	18.8 (139.0)	72.2 (273.0)	0.080
FiO2_filled_var_06, mean (SD)	25.1 (153.7)	18.5 (141.7)	52.1 (193.8)	0.131
FiO2_filled_var_07, mean (SD)	12.0 (75.4)	9.8 (81.2)	21.1 (44.5)	0.081
SpO2FiO2dev_mean_01, mean (SD)	4.0 (0.8)	4.1 (0.8)	3.8 (0.7)	<0.001
SpO2FiO2dev_mean_02, mean (SD)	4.1 (0.7)	4.2 (0.6)	3.7 (1.0)	<0.001
SpO2FiO2dev_mean_03, mean (SD)	4.2 (0.7)	4.3 (0.6)	3.7 (0.9)	<0.001
SpO2FiO2dev_mean_04, mean (SD)	4.2 (0.7)	4.3 (0.5)	3.7 (0.9)	<0.001
SpO2FiO2dev_mean_05, mean (SD)	4.3 (0.7)	4.4 (0.5)	3.7 (0.9)	<0.001
SpO2FiO2dev_mean_06, mean (SD)	4.4 (1.0)	4.5 (0.9)	3.8 (0.8)	<0.001
SpO2FiO2dev_mean_07, mean (SD)	4.6 (4.4)	4.8 (4.8)	3.8 (0.8)	0.001
SpO2FiO2dev_var_01, mean (SD)	1.8 (29.9)	2.2 (33.3)	0.5 (0.6)	0.342
SpO2FiO2dev_var_02, mean (SD)	1.0 (15.8)	0.2 (1.5)	4.1 (35.8)	0.313
SpO2FiO2dev_var_03, mean (SD)	1.3 (19.3)	1.3 (21.0)	1.1 (8.8)	0.872
SpO2FiO2dev_var_04, mean (SD)	0.3 (4.8)	0.4 (5.3)	0.2 (0.2)	0.516
SpO2FiO2dev_var_05, mean (SD)	0.5 (4.9)	0.5 (5.4)	0.3 (1.0)	0.607
SpO2FiO2dev_var_06, mean (SD)	4.3 (57.3)	5.3 (63.9)	0.2 (0.7)	0.139
SpO2FiO2dev_var_07, mean (SD)	2.8 (53.9)	3.4 (60.2)	0.2 (0.2)	0.310

Table B.4 – continued from previous page

				Grouped by BPD
	Overall	No BPD	BPD	P-Value
SpO2_filled_mean_01, mean (SD)	94.4 (2.7)	94.7 (2.7)	93.3 (2.4)	<0.001
SpO2_filled_mean_02, mean (SD)	94.5 (2.8)	94.8 (2.7)	93.0 (2.5)	<0.001
SpO2_filled_mean_03, mean (SD)	94.7 (2.7)	95.0 (2.7)	93.2 (2.6)	<0.001
SpO2_filled_mean_04, mean (SD)	94.9 (2.8)	95.3 (2.6)	93.2 (2.6)	<0.001
SpO2_filled_mean_05, mean (SD)	95.3 (2.7)	95.8 (2.5)	93.4 (2.5)	<0.001
SpO2_filled_mean_06, mean (SD)	95.7 (2.6)	96.2 (2.4)	93.6 (2.5)	<0.001
SpO2_filled_mean_07, mean (SD)	95.7 (2.6)	96.2 (2.4)	93.6 (2.5)	<0.001
SpO2_filled_var_01, mean (SD)	12.6 (9.6)	11.8 (8.7)	16.1 (12.1)	0.002
SpO2_filled_var_02, mean (SD)	10.1 (6.6)	9.5 (6.0)	12.8 (8.1)	0.001
SpO2_filled_var_03, mean (SD)	8.8 (6.1)	8.4 (6.1)	10.8 (5.7)	0.001
SpO2_filled_var_04, mean (SD)	8.3 (7.6)	7.3 (6.1)	12.4 (10.9)	<0.001
SpO2_filled_var_05, mean (SD)	7.5 (6.0)	6.6 (5.2)	11.6 (7.4)	<0.001
SpO2_filled_var_06, mean (SD)	7.6 (7.1)	6.5 (6.3)	12.2 (8.5)	<0.001
SpO2_filled_var_07, mean (SD)	7.5 (6.7)	6.5 (6.0)	12.0 (7.4)	<0.001
hyperoxic_burden_01, mean (SD)	0.1 (0.3)	0.1 (0.3)	0.2 (0.4)	0.254
hyperoxic_burden_02, mean (SD)	0.1 (0.3)	0.1 (0.3)	0.2 (0.4)	0.129
hyperoxic_burden_03, mean (SD)	0.1 (0.2)	0.1 (0.2)	0.2 (0.3)	0.004
hyperoxic_burden_04, mean (SD)	0.1 (0.2)	0.1 (0.2)	0.2 (0.3)	0.002
hyperoxic_burden_05, mean (SD)	0.1 (0.2)	0.1 (0.2)	0.2 (0.2)	<0.001
hyperoxic_burden_06, mean (SD)	0.1 (0.2)	0.1 (0.2)	0.2 (0.2)	<0.001
hyperoxic_burden_07, mean (SD)	0.1 (0.3)	0.1 (0.3)	0.2 (0.2)	0.001
hypoxic_burden_01, mean (SD)	0.5 (0.5)	0.5 (0.5)	0.7 (0.6)	0.001
hypoxic_burden_02, mean (SD)	0.4 (0.5)	0.4 (0.4)	0.6 (0.6)	<0.001
hypoxic_burden_03, mean (SD)	0.4 (0.4)	0.3 (0.4)	0.6 (0.5)	<0.001
hypoxic_burden_04, mean (SD)	0.3 (0.5)	0.3 (0.4)	0.6 (0.7)	<0.001
hypoxic_burden_05, mean (SD)	0.3 (0.4)	0.2 (0.3)	0.6 (0.4)	<0.001
hypoxic_burden_06, mean (SD)	0.3 (0.4)	0.2 (0.3)	0.6 (0.5)	<0.001
hypoxic_burden_07, mean (SD)	0.2 (0.4)	0.2 (0.3)	0.5 (0.5)	<0.001

Table B.4: Univariate analysis of each temporal featue for the 7 days

	Overall	Grouped by BPD No BPD	BPD	P-Value
Day1_PC0, mean (SD)	-0.7 (2.1)	-0.5 (2.0)	-1.6 (2.4)	<0.001
Day1_PC1, mean (SD)	0.1 (0.9)	0.1 (1.0)	0.1 (0.4)	0.362
Day2_PC0, mean (SD)	-0.3 (1.8)	-0.1 (1.6)	-1.4 (2.1)	<0.001
Day2_PC1, mean (SD)	0.0 (0.5)	-0.0 (0.3)	0.1 (1.0)	0.236
Day3_PC0, mean (SD)	-0.1 (1.8)	0.2 (1.5)	-1.3 (2.3)	<0.001
Day3_PC1, mean (SD)	-0.0 (0.6)	-0.0 (0.6)	-0.0 (0.4)	0.992
Day4_PC0, mean (SD)	0.1 (1.8)	0.4 (1.5)	-1.4 (2.4)	<0.001
Day4_PC1, mean (SD)	-0.1 (0.3)	-0.1 (0.3)	-0.0 (0.4)	0.248
Day5_PC0, mean (SD)	0.3 (1.7)	0.6 (1.4)	-1.2 (2.0)	<0.001
Day5_PC1, mean (SD)	-0.1 (0.3)	-0.1 (0.2)	-0.1 (0.3)	0.331
Day6_PC0, mean (SD)	0.4 (1.7)	0.7 (1.4)	-1.1 (2.1)	<0.001
Day6_PC1, mean (SD)	0.0 (1.6)	0.0 (1.7)	-0.0 (0.3)	0.661
Day7_PC0, mean (SD)	0.5 (1.7)	0.8 (1.5)	-1.0 (1.9)	<0.001
Day7_PC1, mean (SD)	0.1 (2.0)	0.1 (2.2)	0.0 (0.3)	0.586

Table B.1: Univariate analysis of each principle component for the 7 days

Feature	Count
Day7_PC0	5
Day7_PC1	5
Day6_PC1	5
Day5_PC0	5
Day5_PC1	5
Day4_PC1	5
Day2_PC1	5
meerlaantal	5
smlga	5
ANS	5
GA_exact	4
Day3_PC0	3
Day3_PC1	2
PT_geslacht	2
Day6_PC0	1
Day4_PC0	1
Day2_PC0	1
Day1_PC1	1

Table B.2: Count of how many times each feature is selected in the best performing logistic regression model after 5 fold recursive feature elimination

Feature	Count
ANS	2
SpO2_filled_mean_07	2
Gebgew	2
SpO2_filled_mean_06	2
SpO2FiO2dev_mean_05	1
FiO2_filled_mean_05	1
FiO2_filled_mean_04	1
hyperoxic_burden_03	1
FiO2_filled_mean_06	1

Table B.3: Count of how many times each feature is selected in the best performing logistic regression model without PCA after 5 fold recursive feature elimination

Model performance over time

Day / Model	**Logistic Regression**	**Random Forest**	**LSTM-AE**	**Combined Model**
1	0,76 (SD 0,03)	0,72 (SD 0,08)	0,60	0,66
2	0,74 (SD 0,08)	0,76 (SD 0,04)	0,65	0,68
3	0,76 (SD 0,08)	0,74 (SD 0,08)	0,69	0,71
4	0,80 (SD 0,07)	0,77 (SD 0,08)	0,68	0,75
5	0,79 (SD 0,07)	0,78 (SD 0,06)	0,71	0,79
6	0,80 (SD 0,07)	0,78 (SD 0,07)	0,78	0,80
7	0,81 (SD 0,07)	0,78 (SD 0,07)	0,81	0,82

Table B.5: Accuracy of neural network models on temporal data for the first 7 days (LSTM-AE) or a combination of static and temporal data (Combined). When reported with SD, performance was calculated using using 5-fold cross validation, when reported without SD, performance was measured on the hold out test set.

Appendix C

Glossary and list of abbreviations

C.1 List of abbreviations

AE autoencoder. 18–20

AUC area under the receiver operating curve. 10

BPD bronchopulmonairy dysplasia. 1, 2, 7, 9–11, 17, 19, 20, 34

CNN Convolutional Neural Network. 6, 18, 19

ConvAE-LSTM Convolutional Autoencoder Long Short-Term Memory Network. 6, 13, 14, 17, 18

FiO2 fraction of inspired oxygen. 4, 5, 12, 13, 17

GA gestational age. 1, 4, 9, 10, 13, 17

LR logistic regression. 6, 10, 12, 13

LSTM Long Short-Term Memory. 6, 7, 9, 19

NICU neonatal intensive care unit. 2, 4

PCA principle component analysis. 5, 6, 12, 30

PMA postmenstral age. 1, 4, 10, 19

RF random forest. 6, 12, 13

RFE Recursive Feature Elimination. 6

ROC reciever operator characteristics. 12, 13

SpO2 Peripheral oxygen saturation. 4, 5, 12, 13, 17–20

C.2 Glossary of terms

autoencoder A neural network consisting of an encoder and decoder that has the goal of reproducing the orginal data through a bottleneck (also known as an embedding or latent vector) that is smaller than the original data. It is maily used for data compression and noise reduction, common applications are anomaly detection and feature learning.. 8, 9, 13, 18–20, 26, 31

cross validation A technique to evaluate model performance by dividing a dataset into a number of subsets. The model is subsequently trained on all but one of these subset and the remaining subset is used to evalute the model. This process is repeated a number of times in order to estimate the model performance on external data.. 7, 12, 13, 15, 30, 32

hyperparameters Settings or configurations that dictate a machine learning model's behavior and performance, adjusted before training to optimize results.. 32

Long Short-Term Memory An advanced RNN that captures long-term dependencies in sequence data.. 6, 19, 31

machine learning A method where computers learn patterns from data to make decisions or predictions, in contrast to models that require explicit programming and rule-setting for each specific task.. 6, 12, 13, 34

nested cross validation A technique to train model hyperparameters and estimate model performance using two loops of cross validation. The inner loop is used to iterate over all possible model parameters and select the best performing parameters. The outer loop with data that is not used in this model tuning is then used to determine the model performance. 34

neural network A machine learning network that consists of different layers of neurons. These neuron are trained on a dataset in order to identify patterns and make predictions.. 5, 14, 30, 34

random forest Random forest is a machine learning model that consists of a large number of individual decision trees. Each tree is trained on a random subset of the data and makes its own predictions. The random forest model combines the predictions of all trees to produce a final result.. 6, 12, 13, 31

Recurrent Neural Network A neural network designed for processing sequences of data, capable of remembering previous information.. 6, 19

Recursive Feature Elimination A feature selection method that iteratively removes the least important features, training the model multiple times to identify the optimal subset of features for improving model accuracy and efficiency.. 6, 32

stratified Ensuring proportional representation of different data subsets or categories during sampling or splitting of data. 7

Appendix D

Student's contribution(s) and learning goals

During the scientific internship, the student provided the following input:

The student participated in meetings with neonatologists in order to develop relevant clinical features that can be used in the predictive models. He used the results from this meeting to develop a data cleaning pipeline that was able to match the static data from each patient with the temporal datasets, clear up outliers, handle missing data and extract features from the temporal dataset.

The student created an automated machine learning training loop that was able to iterate over different subsets of the data and various models in order to find the best performing algorithm and provide an estimate of the performance on unseen data using nested cross validation.

Using an earlier developed proof-of-concept, the student constructed a neural network that is able to predict the occurrence of BPD using a combination of static and temporal data. The student refined this model by experimenting with different training settings, model layers and architectures.

Finally, student also attended the bi-weekly meetings of the Emma Neuroscience research group.

Appendix E

Approved project plan

PROJECTBESCHRIJVING WETENSCHAPPELIJKE STAGE / PROJECT PLAN RESEARCH INTERNSHIP

Invulinstructie aan de student

- Het opstellen van de projectbeschrijving is meer dan een formaliteit: het legt de basis voor een succesvolle stage. Neem er de tijd voor en wees niet te snel tevreden.
- Nadat u met uw stagebegeleider (in het AMC of daarbuiten) voorlopige afspraken heeft gemaakt over de inhoud van uw stage schrijft u in dit formulier de projectbeschrijving.
- Het verdient sterk de voorkeur dat u de projectbeschrijving pas in Onstage uploadt als de AMC senior tutor en de dagelijks begeleider de projectbeschrijving via informeel contact hebben goedgekeurd.
- U schrijft de Projectbeschrijving in het Engels omdat hij dan de basis kan vormen voor de introductie en methodesectie van het wetenschappelijk stageverslag.
- Iedere projectbeschrijving is uniek voor één student. Als twee of meer studenten gelijktijdig de stage verrichten binnen één onderzoekslijn moet uit beide projectbeschrijvingen duidelijk blijken dat de stages onafhankelijk van elkaar worden uitgevoerd.
- De projectbeschrijving moet uiterlijk 2 weken na de start van de wetenschappelijke stage zijn goedgekeurd door:
    - de AMC senior tutor,
    - de dagelijks begeleider en
    - de coördinator.

  Formele accordering door alle drie vindt plaats via Onstage. U kunt de status van de beoordeling op ieder moment nakijken in Onstage (https://onstage2.xebic.com)
- Bewaar het ingevulde formulier zorgvuldig voor het geval om een aanvulling of wijziging wordt gevraagd.

Instructions to the student
After informally discussing your internship proposal with your tutors (both in the Amsterdam UMC, and possibly in a hosting institution) you can fill out this form. The project proposal should be approved by the coordinator prior to or within two weeks after the start of the internship. Formal approval is managed via Onstage. However we advice the student to have informal approval of the senior tutor in the AMC and the daily tutor/senior tutor in the hosting institution before uploading the project description to Onstage.

Naam / name student: Joris van der Vorst	Student nummer: 10210717
Project titel / project title: Early prediction of bronchopulmonary dysplasia in preterm infants using temporal clinical data	

Onderzoekslijn (circa 250 woorden): Geef een korte beschrijving van de onderzoekslijn van de stageverlenende onderzoeksgroep, waarbinnen het stage onderzoeksproject valt. Op basis van de hier geboden informatie zal worden ingeschat of de stage voldoende is ingebed en de kans op succesvolle afronding voldoende is. Hierbij moeten enkele referenties naar publicaties die vanuit de onderzoekslijn zijn verschenen worden toegevoegd.
Research line (250 words): Provide a short description of the research line in which the current project will be embedded. Add some references to publications from the hosting group.

The Emma neuroscience group is working on the development of machine learning models for the prediction of neurodevelopmental outcomes after preterm birth (see https://www.amsterdamumc.org/en/research/highlights/development-of-machine-learning-prediction-models-on-long-term-cognitive-and-motor-outcome-of-preterm-born-children.htm). They have recently published a review where current models were assessed for quality using an evaluation framework that encompasses different best practices in machine learning (van Boven et al., 2022). The research line aims to extend these predictive models to a variety of clinical problems in neonatal and childrens healthcare, which the current research is a part of.

One of the members of the Emma neuroscience group is a post doc researcher with extensive experience in machine learning in medicine, having published his doctoral thesis (Franciscus Cornelis Bennis, 2020) on the subject in addition to a number of published predictive models (Frank C Bennis, Hoogendoorn, Aussems, & Korevaar, 2022; Frank C Bennis et al., 2020; Schinkel et al., 2022). Current research of him focuses on the prediction of multiple outcomes after NICU admission (including BPD), and he has thus the expertise to supervise the project.

Achtergrond en probleemstelling (circa 500 woorden): Beschrijf de achtergrond van het door de student zelf uit te voeren onderzoek en hoe dat aansluit bij eerder onderzoek uit de stage-verlenende onderzoeksgroep. Vat samen wat al bekend is over het onderwerp, welke lacunes in de bestaande kennis zitten en welke probleemstelling hieruit volgt.

Background and research questions (500 words): Provide the background of the research that will be carried out by the student. Summarize the current body of knowledge on the subject as a background to the aim of the project.

Bronchopulmonairy dysplasia (BPD) is a chronic lung disease in preterm infants characterized by a disruption of pulmonary development and injury due to long term respiratory support.
Infants with BPD often have long term respiratory morbidity and BPD is associated with a lower cognitive ability (Gilfillan, Bhandari, & Bhandari, 2021; Twilhaar et al., 2018).
Early identification of the risk of BPD is important in order to promptly target interventions and give insight into long term prognosis (Romijn et al., 2023). For example, postnatal corticosteroids may be useful in the prevention of BPD but are associated with side effects such as gastrointestinal perforation and neurodevelopmental impairment (Kwok et al., 2023). Identifying which preterm infants are most at risk may aid to target high-risk infants and not expose others to these side effects.
BPD is classified at 36 weeks postmenstrual age by the necessity for supplemental oxygen and is graded by the mode of respiratory support (Dysart et al., 2019).

Traditional prediction models mainly use multivariable linear regression with routinely collected data such as gestational age, birthweight, sex and information about early respiratory status as predictive features (Romijn et al., 2023). A recent review of 65 models reported a median c-statistic of 0.84 (range 0.43-1.00), although it was noted that the better performing models had a high risk of bias, which corresponds with a lower c-statistic of 0.77 (range 0.41-0.97) at external validation (Romijn et al., 2023).

Although these models have a good diagnostic accuracy, the main predictive drivers were static items such as birth weight and gestational age or used data beyond the normal clinical records, such as lung ultrasound scores. If respiratory variables were used, these models mainly focused on the type of respiratory support (94/119 models that used respiratory values (Romijn et al., 2023)). However, temporal data itself is not used.

Temporal clinical data may be beneficial for use in a prediction model. This data is already collected in the routine monitoring of patients admitted in the NICU and can give detailed insight into the pulmonary function of these infants. This type of data has already been used in adult ICU settings. For example extracted temporal variables such as last FiO2 and mean Tidal Volume had a high predictive value for successful extubation in patients with COVID-19 pneumonia (Fleuren et al., 2021) and a LSTM model was able to predict sepsis and myocardial infarction in a public ICU dataset (MIMIC-III) with an AUC of 0.876 and 0.823 respectively (Kaji et al., 2019).
Currently no known model has used this type of temporal data for the prediction of BPD.

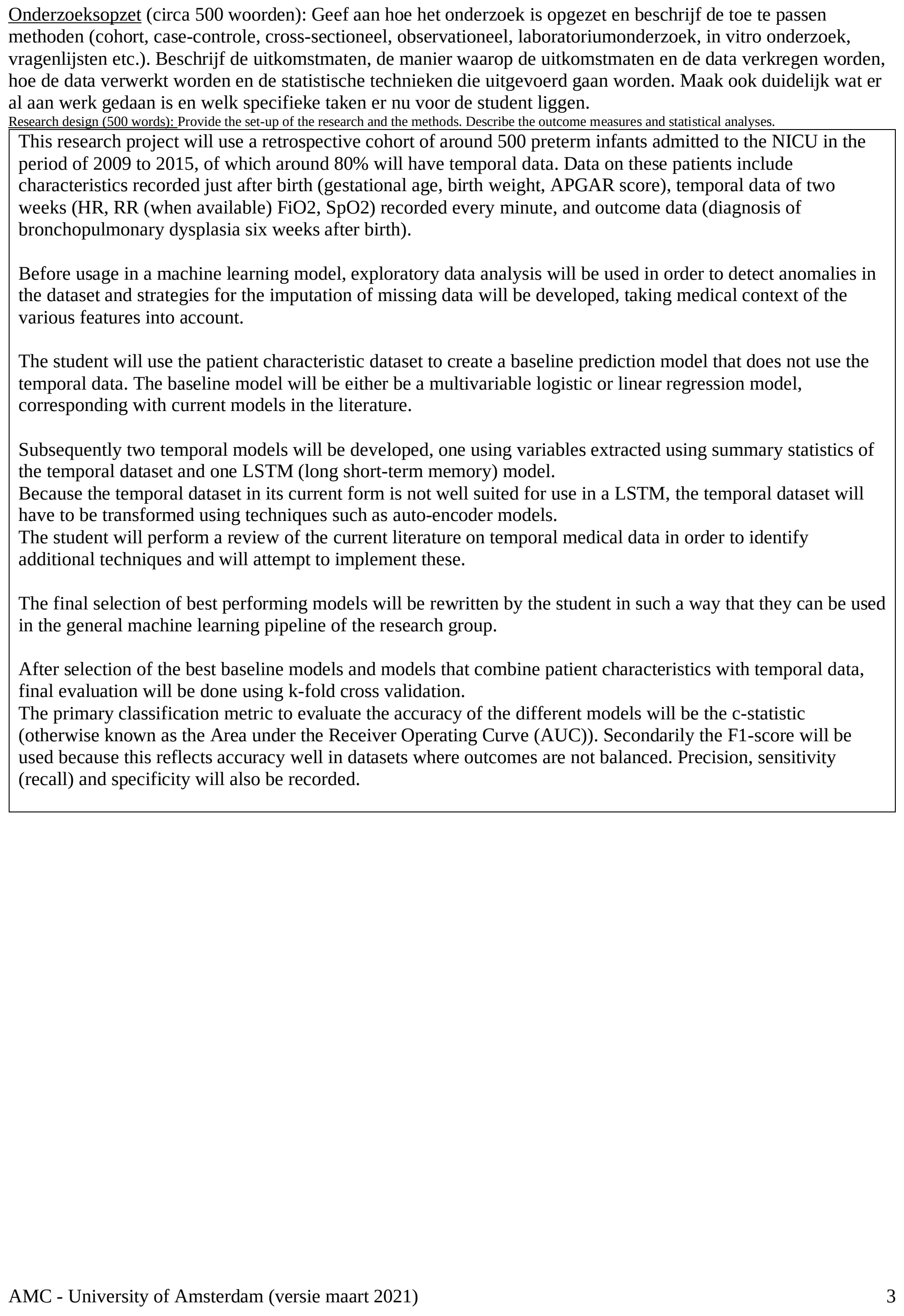
Vraagstelling en/of hypothese (circa 150 woorden)

Research question and/or hypothesis (150 words)

The aim of this project is to predict the development of bronchopulmonary dysplasia in a cohort of preterm infants admitted to the NICU and to investigate whether the usage of temporal data of the first two weeks of admission will improve this prediction over static data present at birth.

The research question of this project will be:
What is the added value of temporal data for the prediction of bronchopulmonary dysplasia in preterm infants?

In order to answer this question, the following subquestions will be investigated:
What is the accuracy of a machine learning model for prediction of BPD that only uses data that is collected at birth (baseline)?
Does the accuracy of the prediction improve if statistical summary from the clinical data in the first two weeks after birth are added?
How does the accuracy of the model improve when a LSTM is used to take vital parameters from the first two weeks is taken into account?



Onderzoeksopzet (circa 500 woorden): Geef aan hoe het onderzoek is opgezet en beschrijf de toe te passen methoden (cohort, case-controle, cross-sectioneel, observationeel, laboratoriumonderzoek, in vitro onderzoek, vragenlijsten etc.). Beschrijf de uitkomstmaten, de manier waarop de uitkomstmaten en de data verkregen worden, hoe de data verwerkt worden en de statistische technieken die uitgevoerd gaan worden. Maak ook duidelijk wat er al aan werk gedaan is en welk specifieke taken er nu voor de student liggen.

Research design (500 words): Provide the set-up of the research and the methods. Describe the outcome measures and statistical analyses.

This research project will use a retrospective cohort of around 500 preterm infants admitted to the NICU in the period of 2009 to 2015, of which around 80% will have temporal data. Data on these patients include characteristics recorded just after birth (gestational age, birth weight, APGAR score), temporal data of two weeks (HR, RR (when available) FiO2, SpO2) recorded every minute, and outcome data (diagnosis of bronchopulmonary dysplasia six weeks after birth).

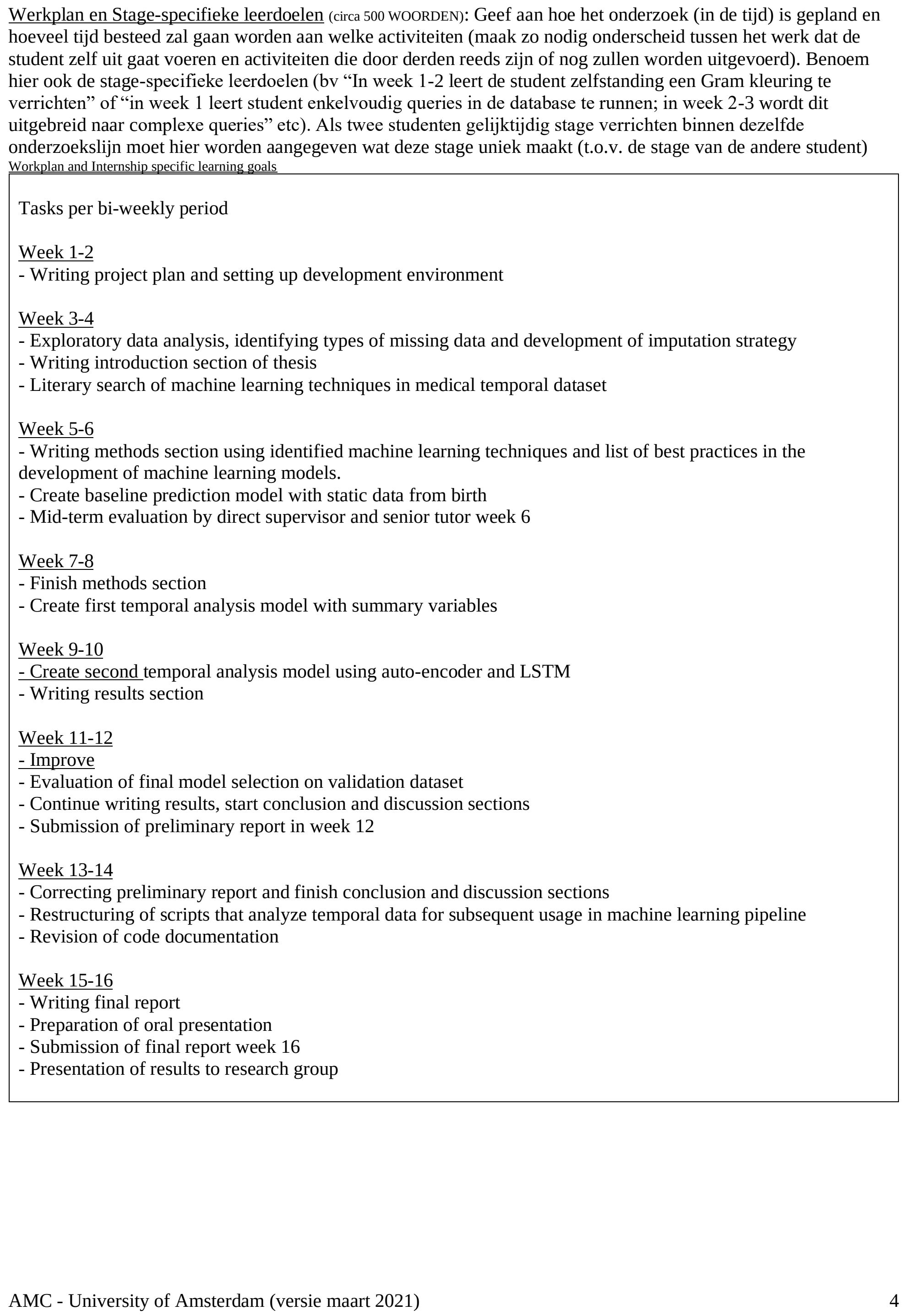
Before usage in a machine learning model, exploratory data analysis will be used in order to detect anomalies in the dataset and strategies for the imputation of missing data will be developed, taking medical context of the various features into account.

The student will use the patient characteristic dataset to create a baseline prediction model that does not use the temporal data. The baseline model will be either be a multivariable logistic or linear regression model, corresponding with current models in the literature.

Subsequently two temporal models will be developed, one using variables extracted using summary statistics of the temporal dataset and one LSTM (long short-term memory) model.
Because the temporal dataset in its current form is not well suited for use in a LSTM, the temporal dataset will have to be transformed using techniques such as auto-encoder models.
The student will perform a review of the current literature on temporal medical data in order to identify additional techniques and will attempt to implement these.

The final selection of best performing models will be rewritten by the student in such a way that they can be used in the general machine learning pipeline of the research group.

After selection of the best baseline models and models that combine patient characteristics with temporal data, final evaluation will be done using k-fold cross validation.
The primary classification metric to evaluate the accuracy of the different models will be the c-statistic (otherwise known as the Area under the Receiver Operating Curve (AUC)). Secondarily the F1-score will be used because this reflects accuracy well in datasets where outcomes are not balanced. Precision, sensitivity (recall) and specificity will also be recorded.



Werkplan en Stage-specifieke leerdoelen (circa 500 WOORDEN): Geef aan hoe het onderzoek (in de tijd) is gepland en hoeveel tijd besteed zal gaan worden aan welke activiteiten (maak zo nodig onderscheid tussen het werk dat de student zelf uit gaat voeren en activiteiten die door derden reeds zijn of nog zullen worden uitgevoerd). Benoem hier ook de stage-specifieke leerdoelen (bv "In week 1-2 leert de student zelfstanding een Gram kleuring te verrichten" of "in week 1 leert student enkelvoudig queries in de database te runnen; in week 2-3 wordt dit uitgebreid naar complexe queries" etc). Als twee studenten gelijktijdig stage verrichten binnen dezelfde onderzoekslijn moet hier worden aangegeven wat deze stage uniek maakt (t.o.v. de stage van de andere student)
Workplan and Internship specific learning goals

Tasks per bi-weekly period

Week 1-2
- Writing project plan and setting up development environment

Week 3-4
- Exploratory data analysis, identifying types of missing data and development of imputation strategy
- Writing introduction section of thesis
- Literary search of machine learning techniques in medical temporal dataset

Week 5-6
- Writing methods section using identified machine learning techniques and list of best practices in the development of machine learning models.
- Create baseline prediction model with static data from birth
- Mid-term evaluation by direct supervisor and senior tutor week 6

Week 7-8
- Finish methods section
- Create first temporal analysis model with summary variables

Week 9-10
- Create second temporal analysis model using auto-encoder and LSTM
- Writing results section

Week 11-12
- Improve
- Evaluation of final model selection on validation dataset
- Continue writing results, start conclusion and discussion sections
- Submission of preliminary report in week 12

Week 13-14
- Correcting preliminary report and finish conclusion and discussion sections
- Restructuring of scripts that analyze temporal data for subsequent usage in machine learning pipeline
- Revision of code documentation

Week 15-16
- Writing final report
- Preparation of oral presentation
- Submission of final report week 16
- Presentation of results to research group

AMC - University of Amsterdam (versie maart 2021) 4

Keuze voor een 16 of 24 weken wetenschappelijke stage: Geef aan of er een 16 of 24 weken stage uitgevoerd gaat worden.
Alleen wanneer de 24 weken variant wordt gekozen: Motiveer waarom er 24 weken nodig zijn voor het onderzoek. Er moet duidelijk worden waarom het onderzoek niet in 16 weken kan worden uitgevoerd.
De motivatie kan liggen op één of meer van de volgende domeinen:
Onderzoeksopzet:
- aantoonbaar grote eigen inbreng bij beschrijven probleem, achtergrond, hypothese en onderzoeksvraag en opzet van het onderzoek.

Kwantiteit:
- aantal patiënten of experimenten.

Kwaliteit:
- prospectief onderzoek; veel data per patiënt; complexe/tijdrovende experimenten

Analyse:
- complexe statistiek.

Discussie:
- uitgebreid literatuuronderzoek om resultaten te relateren aan relevante studies.

Een stage van het type meta-analyse, systematische review en status-onderzoek mag alleen in een 16 weken stage gedaan worden.

Choice and Motivation for 16 or 24 weeks internship

This project will be a 16 week internship. Only the criterium of complex statistics applies to this project, but because the analysis is the sole focus of the project, this can be achieved in 16 weeks.

Faciliteiten (circa 250 WOORDEN): Geef aan welke faciliteiten (b.v. toegang tot elelktronische dossiers, bepaalde gepatenteerde vragenlijsten, bepaalde monsters of reagentia) voor de student noodzakelijk zijn voor het succesvol verrichten van deze stage en welke afspraken hierover gemaakt zijn.

Which specific facilities are needed for this internship research and what has been arranged

Only a limited number of facilities are needed for this project. The datasets containing the various patient characteristics, temporal features and outcomes are already extracted into separate files and therefore no access to Epic or other EHR records are needed.
The student will need to have access to the share folders where this dataset and scripts of the research pipeline are stored. Additionally, the user account of the student will need to be allowed to run Python and Tensorflow and a software IDE needs to be installed. Finally, the student will need to have access to compute resources.
Above mentioned facilities have already been arranged.

METC, DEC, GGO, AVG: Is er voor de onderzoekswerkzaamheden van de student goedkeuring vereist van de Medisch Ethische Toetsingscommissie en/of van de Dieren Experimenten Commissie en/of is er een vergunning vereist om met Genetisch Gemodificeerde Organismen te werken en/of het onderzoek is aangemeld bij de functionaris gegevensbescherming? Vul de titel van de goedkeuringsaanvraag in, de status en datum van goedkeuring.
Het projectvoorstel kan alleen worden goedgekeurd als ten tijde van het indienen hiervan deze goedkeuring aanwezig is.

Ethics Review Approval: If for the research activities that the student will be performing approval of any Ethics Review Board (or related) is required, then this approval should have been obtained at the time of submission of the current proposal, otherwise the proposal can not be approved by the coordinator. Fill in the title of the approval request, status and date of approval.

This research is approved by the METC, reference number W21_516 # 21.569.

Professionele ontwikkeling student (circa 250 woorden) Hoe past deze specifieke onderzoeksstage, en de specifieke leerdoelen in de professionele ontwikkeling van deze student?
Professional development: How does this internship fit in the professional development of this student?

This internship provides the student with a number of professional skills that can be developed.

Data Analysis and Interpretation. Medical research datasets often present challenges due to their incomplete and sometimes inaccurate nature. The student will learn how to scrutinize and cleanse these datasets, detecting anomalies and applying their medical knowledge in this analysis. This may involve merging diverse features or devising strategies for interpolating missing data.

Programming and machine learning skills. In order to develop the predictive model the student will need to program in Python and use different libraries, such as TensorFlow, Scikit-Learn and Matplotlib.

Collaborating on software development in research. This project offers experience in collaborative research and software development. The student will work in tandem with fellow researchers on diverse predictive machine learning models, aiming for a reusable code in the research pipeline. They will learn to effectively coordinate tasks, ensuring scripts from various researchers work together. The student will also write comprehensive code documentation and use version control tools like Git.

Communication skills. The student will need to convey his methods, findings, and implications to audiences of different backgrounds. The student will therefore need to tailor his message to each audience, for example to insure that the medical specialists can understand the clinical relevance and that data scientist can focus on the technical aspects of the model.

References

Bennis, F. C. (2020). *Machine learning in medicine: big pictures require small, but crucial strokes* [Maastricht University]. https://doi.org/10.26481/dis.20201113fb

Bennis, F. C., Hoogendoorn, M., Aussems, C., & Korevaar, J. C. (2022). Prediction of heart failure 1 year before diagnosis in general practitioner patients using machine learning algorithms: a retrospective case–control study. *BMJ Open*, *12*(8), e060458. https://doi.org/10.1136/bmjopen-2021-060458

Bennis, F. C., Teeuwen, B., Zeiler, F. A., Elting, J. W., van der Naalt, J., Bonizzi, P., Delhaas, T., & Aries, M. J. (2020). Improving Prediction of Favourable Outcome After 6 Months in Patients with Severe Traumatic Brain Injury Using Physiological Cerebral Parameters in a Multivariable Logistic Regression Model. *Neurocritical Care*, *33*(2), 542–551. https://doi.org/10.1007/s12028-020-00930-6

Dysart, K., Gantz, M. G., McDonald, S., Bamat, N. A., Keszler, M., Kirpalani, H., Laughon, M. M., Poindexter, B. B., Duncan, A. F., Yoder, B. A., Eichenwald, E. C., DeMauro, S. B., & Jensen, E. A. (2019). The Diagnosis of Bronchopulmonary Dysplasia in Very Preterm Infants An Evidence-based Approach. *American Journal of Respiratory and Critical Care Medicine*, *200*(6), 751–759. https://doi.org/10.1164/rccm.201812-2348OC

Fleuren, L. M., Dam, T. A., Tonutti, M., de Bruin, D. P., Lalisang, R. C. A., Gommers, D., Cremer, O. L., Bosman, R. J., Rigter, S., Wils, E.-J., Frenzel, T., Dongelmans, D. A., de Jong, R., Peters, M., Kamps, M. J. A., Ramnarain, D., Nowitzky, R., Nooteboom, F. G. C. A., de Ruijter, W., … Collaborators, the D. I. C. U. D. S. A. C.-19. (2021). Predictors for extubation failure in COVID-19 patients using a machine learning approach. *Critical Care*, *25*(1), 448. https://doi.org/10.1186/s13054-021-03864-3

Gilfillan, M., Bhandari, A., & Bhandari, V. (2021). Diagnosis and management of bronchopulmonary dysplasia. *BMJ*, n1974. https://doi.org/10.1136/bmj.n1974

Kaji, D. A., Zech, J. R., Kim, J. S., Cho, S. K., Dangayach, N. S., Costa, A. B., & Oermann, E. K. (2019). An attention based deep learning model of clinical events in the intensive care unit. *PLoS ONE*, *14*(2), 1–17. https://doi.org/10.1371/journal.pone.0211057

Kwok, T. C., Batey, N., Luu, K. L., Prayle, A., & Sharkey, D. (2023). Bronchopulmonary dysplasia prediction models: a systematic review and meta-analysis with validation. *Pediatric Research*, *December 2022*, 1–12. https://doi.org/10.1038/s41390-022-02451-8

Romijn, M., Dhiman, P., Finken, M. J. J., van Kaam, A. H., Katz, T. A., Rotteveel, J., Schuit, E., Collins, G. S., Onland, W., & Torchin, H. (2023). Prediction Models for Bronchopulmonary Dysplasia in Preterm Infants: A Systematic Review and Meta-Analysis. *The Journal of Pediatrics*, *10*, 113370. https://doi.org/10.1016/j.jpeds.2023.01.024

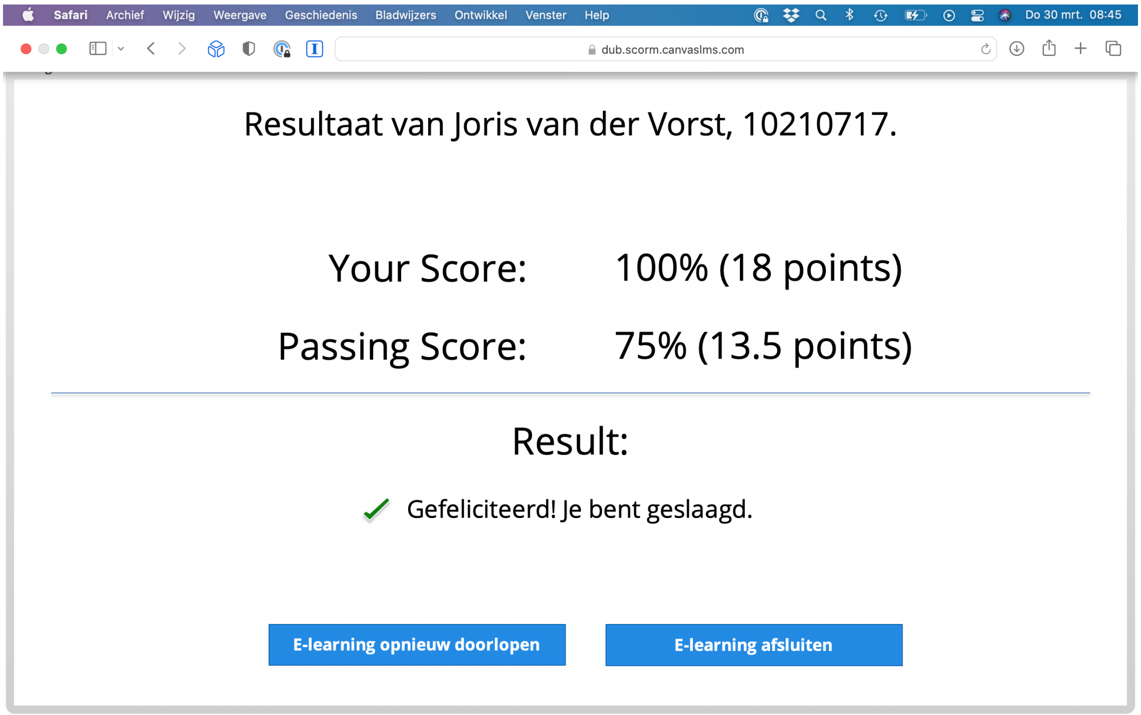
Schinkel, M., Boerman, A. W., Bennis, F. C., Minderhoud, T. C., Lie, M., Peters-Sengers, H., Holleman, F., Schade, R. P., de Jonge, R., Wiersinga, W. J., & Nanayakkara, P. W. B. (2022). Diagnostic stewardship for blood cultures in the emergency department: A multicenter validation and prospective evaluation of a machine learning prediction tool. *EBioMedicine*, *82*. https://doi.org/10.1016/j.ebiom.2022.104176

Twilhaar, E. S., Wade, R. M., de Kieviet, J. F., van Goudoever, J. B., van Elburg, R. M., & Oosterlaan, J. (2018). Cognitive Outcomes of Children Born Extremely or Very Preterm Since the 1990s and Associated Risk Factors: A Meta-analysis and Meta-regression. *JAMA Pediatrics*, *172*(4), 361–367. https://doi.org/10.1001/jamapediatrics.2017.5323

van Boven, M. R., Henke, C. E., Leemhuis, A. G., Hoogendoorn, M., van Kaam, A. H., Königs, M., & Oosterlaan, J. (2022). Machine Learning Prediction Models for Neurodevelopmental Outcome After Preterm Birth: A Scoping Review and New Machine Learning Evaluation Framework. *Pediatrics*, *150*(1), 1–15. https://doi.org/10.1542/peds.2021-056052

Appendix F

E-learning scientific Integrity



Safari Archief Wijzig Weergave Geschiedenis Bladwijzers Ontwikkel Venster Help
Do 30 mrt. 08:45
dub.scorm.canvaslms.com

Resultaat van Joris van der Vorst, 10210717.

Your Score: 100% (18 points)

Passing Score: 75% (13.5 points)

Result:

✓ Gefeliciteerd! Je bent geslaagd.

E-learning opnieuw doorlopen
E-learning afsluiten