August 2018

Detecting abnormalities in clinical data with Generative Adversarial Networks

Matthias Wright

Abstract. In this study, we employed *Generative Adversarial Networks* to examine whether they are capable of distinguishing between the data of patients who are suffering from respiratory failure or cardiac arrhythmia and the data of healthy patients. Specifically, we looked at the ventilation parameter PEEP and the blood gases pC02 and PO2. We trained the generative adversarial network with the data of patients who were not suffering from respiratory failure. In order to validate the results, we first fed test data of patients suffering from respiratory failure and then test data of patients not suffering from respiratory failure into the discriminative model and compared the cross entropy losses. The test data of patients suffering from respiratory failure caused an average cross entropy loss of 0.7508 and the test data of patients not suffering from respiratory failure caused an average cross entropy loss of 0.5034. The best performance was achieved with the PO2 data. However, for the test data the area under the ROC curve was 0.57, leading us to the conclusion that either *Generative Adversarial Networks* are not capable of diagnosing respiratory failure or the used data did not contain the necessary information to perform this distinction.

Furthermore, we trained a generative adversarial network with normal ECG recordings. We fed test data containing normal ECG recordings and test data containing arrhythmic ECG recordings into the discriminative model and compared the cross entropy losses. The test data containing normal ECG recordings caused an average cross entropy loss of 0.0021 and the test data containing arrhythmic ECG recordings caused an average cross entropy loss of 0.0807. The area under the ROC curve was 0.71, leading us to the conclusion that *Generative Adversarial Networks* are capable of distinguishing between the two groups.

Keywords. Machine Learning, Deep Learning, Generative Adversarial Networks, Healthcare

1. Introduction

Severely ill or injured patients are in need of intensive care. In an intensive care unit, the vital functions of a patient are constantly carefully monitored so that doctors can quickly and appropriately respond to life-threatening changes in the patient's condition. In 70% of the cases, the patient in intensive care only depends on monitoring and vital support for a couple of days with a high chance of survival. However, 30% of intensive care patients stay for longer periods, sometimes even months and the likelihood of survival decreases over time [1]. In the US, 39.5% of intensive care unit patients are receiving mechanical ventilation [2]. A study, conducted in 1995, examined 1426 patients, who were suffering from respiratory failure and received mechanical ventilation. While 55.6% survived their stay in intensive care unit, 44.4% died in the hospital [3]. Respiratory failure is also one of the two commonest causes of death after discharge from intensive care [4]. When it comes to mechanical ventilation, it is essential to take measures at the right time, but in order to do that we have to be able to detect changes in the patient's condition. Usually, this is done by a medical expert, who analyses the available clinical data. However, this data may contain much more information than a human can extract from it. Studies have shown that the human mind is only capable of dealing with seven parameters at the same time [5] but according to estimates, there are approximately 250 parameters for a intensive care unit patient [6].

Another frequent complication that occurs among hospitalized patients is cardiac arrest [7]. Especially unexpected cardiac arrests in the intensive care unit lead to high mortality rates, only 15% of the patients survive [8]. In order to detect cardiac complications, patients are monitored with the electrocardiogram (ECG) [9]. However, the interpretation of ECG signals is not a trivial task.

Employing machine learning models in order to leverage the massive amounts of available data could benefit the medical personnel as well as the patients.

The goal of this thesis is to develop a *Generative Adversarial Network* and examine whether it is capable of distinguishing between the data of patients suffering from respiratory failure or cardiac arrhythmia and the data of healthy patients. We will train our model only on the data of healthy patients because there is much more data of healthy patients available.

2. Background and Related Work

2.1. Discriminative Models for ECG Arrhythmia Detection

Discriminative models have been used in the past for classifying ECG Arrhythmia. Polat et al. [11] employed *Support Vector Machines* (Cortes et al., 1995) [12] for detecting arrhythmia in ECG recordings and achieved a classification accuracy of 96.86% on a dataset from the University of California. Vishwa et al. [13] employed artificial neural networks for classifying arrhythmia in ECG recordings. The model was tested on the *MIT-BIH Arrhythmia Database* [14] and achieved an accuracy of 96.77%.

However, both models were trained using normal ECG data as well as arrhythmic ECG data. As stated in section 1, our goal is to build a model for detecting cardiac arrhythmia that is trained only on normal ECG data. Thus, we will explore generative modelling.

2.2. Generative Adversarial Networks

A *Generative Adversarial Network* (Goodfellow, 2014) [15] is a framework for estimating generative models by employing an adversarial training process where a generative model G and a discriminative model D are simultaneously trained. The goal is for G to capture the distribution of the training data and for D to estimate the probability that a given sample originates from the training data and was not generated by G. In *Game theory* terms, this process can be described as a minimax two-player game. According to the framework, both G and D have to be differentiable functions. In practice, G and D are commonly represented by neural networks and can be trained using the *Backpropagation* algorithm.

2.2.1. Training Process

The generative model G takes in a random input vector z and outputs a sample $G(z)$ that has the same shape as the training examples in the dataset p_{data}. $D(x)$ denotes the probability that x came from the dataset. The goal of the discriminative model D is to correctly decide whether a given example came from the dataset or was generated by G. Simultaneously, G is trained to minimize the probability $\log(1-D(G(z)))$ that a generated sample did not come from the dataset, where z is a random input vector. This process corresponds to a two-player minimax game with the value function $V(D,G)$:

$$\min_G \max_D V(D,G) = E_{x\sim p_{data}(x)}[\log(D(x))] + E_{z\sim p_z(z)}[\log(1-D(G(z)))]. \tag{1}$$

$E_{x\sim p_{data}(x)}[\log(D(x))]$ denotes the expected value of $\log(D(x))$, averaged over the distribution p_{data} and $E_{z\sim p_z(z)}[\log(1-D(G(z)))]$ denotes the expected value of $\log(1-D(G(z)))$, averaged over the distribution p_z, where the random input vector z is sampled from. Early in the learning process, the probability $D(G(z))$ that a generated sample came from the training data is close to 0. As a result, the partial derivative $\frac{\partial \log(1-D(G(z)))}{\partial G(z)}$ of the cost function $\log(1-D(G(z)))$ with respect to the generated sample $G(z)$ is also close to 0. A small gradient leads to small step sizes for *Gradient descent*, which slows down the learning process. To prevent this, Goodfellow et al. have proposed to train the generative model G by maximizing $\log(D(G(z)))$ instead of minimizing $\log(1-D(G(z)))$. Furthermore, Goodfellow et al. proposed to alternate the training of D and G. One entire training iteration comprises j steps of optimizing D and one step of optimizing G. Algorithm 1 shows *Mini-batch stochastic gradient descent* training for a generative adversarial network.

Algorithm 1 Mini-batch gradient descent training for a generative adversarial network.

```
for number of training iterations do
    for $k$ steps do
        Sample minibatch of $m$ noise samples $\{z^{(1)},...,z^{(m)}\}$
        Sample minibatch of $m$ examples $\{x^{(1)},...,x^{(m)}\}$ from the training set
        Update the discriminative model with gradient descent:
            $\nabla \frac{1}{m} \sum_{i=1}^{k} [\log(D(x^{(i)})) + \log(1-D((G(z^{(i)}))))]$
    Sample minibatch of $m$ noise samples $\{z^{(1)},...,z^{(m)}\}$
    Update the generative model with gradient descent:
            $\nabla \frac{1}{m} \sum_{i=1}^{k} \log(D(G(z^{(i)})))$
```

3. Materials and Methods

3.1. Data

3.1.1. Ventilation Data

The data used in the experiments for detecting respiratory failure was extracted from the freely available *Medical Information Mart for Intensive Care III Database (MIMIC-III*

Database)[16] made available by the *PhysioNet*[17] project. The *MIMIC-III Database* comprises over forty thousand records of critical care patients of the *Beth Israel Deaconess Medical Center* that were collected over a period from 2001 to 2012. The database includes demographical information, vital sign measurements, laboratory test results, procedures, medications, caregiver notes, imaging reports, and mortality. The data was recorded during routine clinical care and was not particularly intended for the purpose of analysis.
We conducted separate experiments with three different measurements.

PEEP The Positive End-Expiratory Pressure (PEEP) is the positive pressure that remains in the airways at the end of expiration [18]. The *MIMIC-III Database* contains PEEP measurements for 12630 different patients. Of the 12630 patients, 6861 were suffering from respiratory failure and 5769 were not. On average, there are 6.88 recorded values per patient.

PO2 The Partial Pressure of Oxygen (PO2) indicates the amount of oxygen gas dissolved in the blood [19]. The *MIMIC-III Database* contains CO2 measurements for 31424 different patients, 9403 were suffering from respiratory failure and 22021 were not. On average, there are 15.61 recorded values per patient.

pCO2 The Partial Pressure of Carbon Dioxide (pCO2) indicates the amount of carbon dioxide gas dissolved in the blood [19]. The *MIMIC-III Database* contains pCO2 measurements for 31424 different patients, 9403 were suffering from respiratory failure and 22021 were not. On average, there are 41.72 recorded values per patient.

3.1.2. ECG Data

The data used for the ECG experiments was also provided by the *PhysioNet*[17] project. For the training of the *Generative Adversarial Network* we used the *Long-Term ST Database* [20]. It contains 86 ECG recordings obtained from 80 patients. The recordings originate from different hospitals over a period from 1995 to 2002. For testing we used the *MIT-BIH Arrhythmia Database* [14]. It contains 48 ECG recordings obtained from 47 patients, which were studied by the BIH Arrhythmia Laboratory between 1975 and 1979.

3.2. Experiments

3.2.1. Ventilation Data

For the discriminative model D we used a neural network with three layers, an input layer with 49 neurons, one hidden layer with 24 neurons, and a output layer with one neuron. We applied the sigmoid function to the output neuron because we want it to output a probability and the rectifier function to the hidden layer. For the generative model G we used a neural network with three layers, an input layer with 20 neurons, one hidden layer with 40 neurons, and one output layer with 49 neurons. We applied the rectifier function to the hidden layer and the output layer because the measurements take on only positive values. For the training process we used *Mini-batch gradient descent* where the size of the mini-batches was 25, in each iteration we performed one step of optimizing D and one step of optimizing G. In order to stabilize the training process we added noise ε to the input of D (proposed by Arjovsky and Bottou [21]), with $\varepsilon \sim \mathcal{N}(0, \sigma^2)$. Table 1 shows the sizes of the training and test sets.

	pCO2 Data	PO2 Data	PEEP Data
Training set	15525 examples	15530 examples	2638 examples
Test sets	2000 examples	2000 examples	500 examples

Table 1. Sizes of the training and test sets for ventilation data.

3.2.2. ECG Data

For the discriminative model D we used a neural network with five layers, an input layer with 101 neurons, three hidden layers with 50 neurons, and an output layer with one neuron. We applied the sigmoid function to the output neuron because we want it to output a probability and the tanh function to the hidden layers. For the generative model G we used a neural network with five layers, an input layer with 60 neurons, three hidden layers with 80 neurons, and an output layer with 101 neurons. We applied the tanh function to the hidden layers and no activation function to the output layer because the measurements can take on positive and negative values. For the training process we used *Mini-batch gradient descent* where the size of the mini-batches was 512, in each iteration we did one step of optimizing D and one step of optimizing G. As described in section 3.2.1, we stabilized the training process by adding noise to the input of D. In order to prevent the neural networks from overfitting, we employed *Dropout* regularization (Srivastava et al., 2014) [22]. When we apply dropout regularization to a layer, every neuron output in that layer is set to zero with a specified probability p. We set $p = 0.5$ for every hidden layer of D and every hidden layer of G.
The training set consisted 1.594.410 examples, the test set containing normal ECG recordings consisted of 66.641 examples, and the test set containing abnormal/arrhythmic ECG recordings consisted of 41.457 examples.

4. Results

As mentioned in section 1, we only used the measurements of healthy patients for training the models. In order to examine whether or not *Generative Adversarial Networks* can be used to diagnose respiratory diseases or cardiac arrhythmia, we used two test datasets for each model. One dataset containing data of healthy patients and one dataset containing data of patients with respiratory failure and cardiac arrhythmia, respectively. We separately fed both of the test datasets into the discriminator D and compared the resulting losses.
In order to test the difference of the means of the two test sets, we applied the t-test. We formulated the null hypothesis $H_0 : \mu_1 = \mu_2$ and the alternative hypothesis $H_a : \mu_1 > \mu_2$, where μ_1 is the mean of the test set of patients who were suffering from respiratory failure or cardiac arrhythmia and μ_2 is the mean of the test set containing data of healthy patients.
We plotted box plots to visualize the value range of the respective losses.
In order to examine the diagnostic ability of the models, we constructed *Receiver Operating Characteristic* (ROC) curves. A ROC curve represents the relationship between sensitivity and specificity [23]. The area under the ROC curve (AUC) gives measure to the accuracy of a test, where 1 is the maximum. For medical diagnosis, an AUC of 0.7 to 0.8 is considered acceptable, 0.8 to 0.9 is considered excellent, and greater than 0.9 is considered outstanding [24].

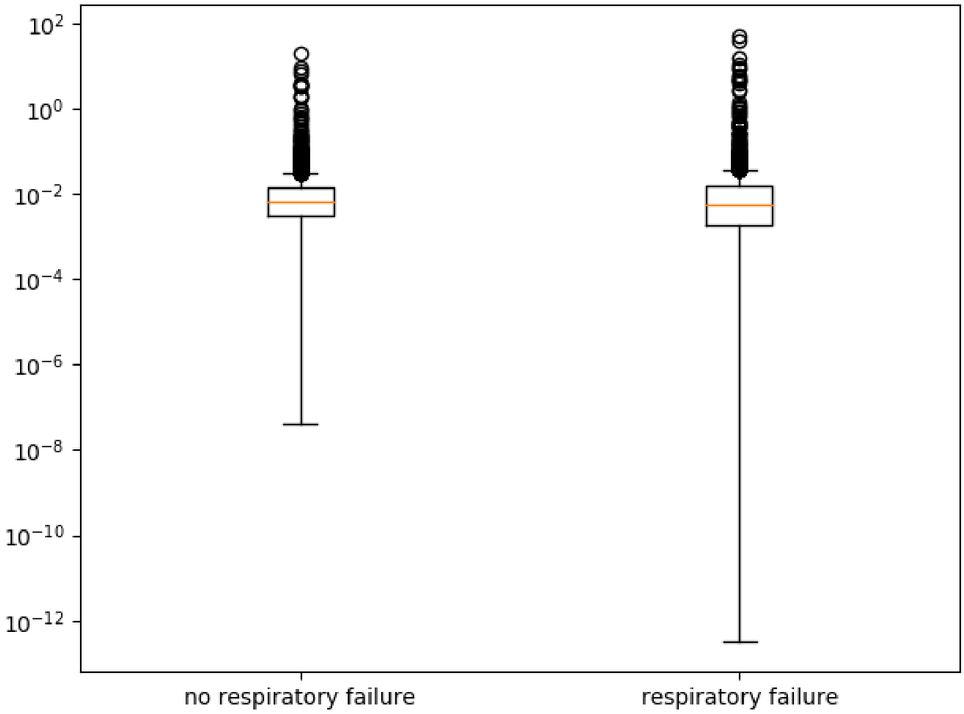

Figure 1. Box plot of the test cross entropy loss for pCO2 data that was fed into the discriminator *D*. Note that we used a log scale.

4.1. Ventilation Data

4.1.1. pCO2 Data

The test data of the patients who were not suffering from respiratory failure produced an average cross entropy loss of 0.0485 and the test data of the patients who were suffering from respiratory failure an average cross entropy loss of 0.1008. Figure 1 shows a box plot for the test dataset. The plot shows that the losses have nearly the same range. We were able to reject the hypothesis $H_0 : \mu_1 = \mu_2$ in favor of $H_a : \mu_1 > \mu_2$ with a p-value of 0.1. Figure 4 shows the ROC curve as well as the AUC.

4.1.2. PO2 Data

The test data of the patients who were not suffering from respiratory failure produced an average cross entropy loss of 0.6691 and the test data of the patients who were suffering from respiratory failure an average cross entropy loss of 1.2531. Figure 2 shows a box plot for the test dataset. The plot shows that although the means are varying, the losses have nearly the same range. We were able to reject the hypothesis $H_0 : \mu_1 = \mu_2$ in favor of $H_a : \mu_1 > \mu_2$ with a p-value of 0.1. Figure 4 shows the ROC curve as well as the AUC.

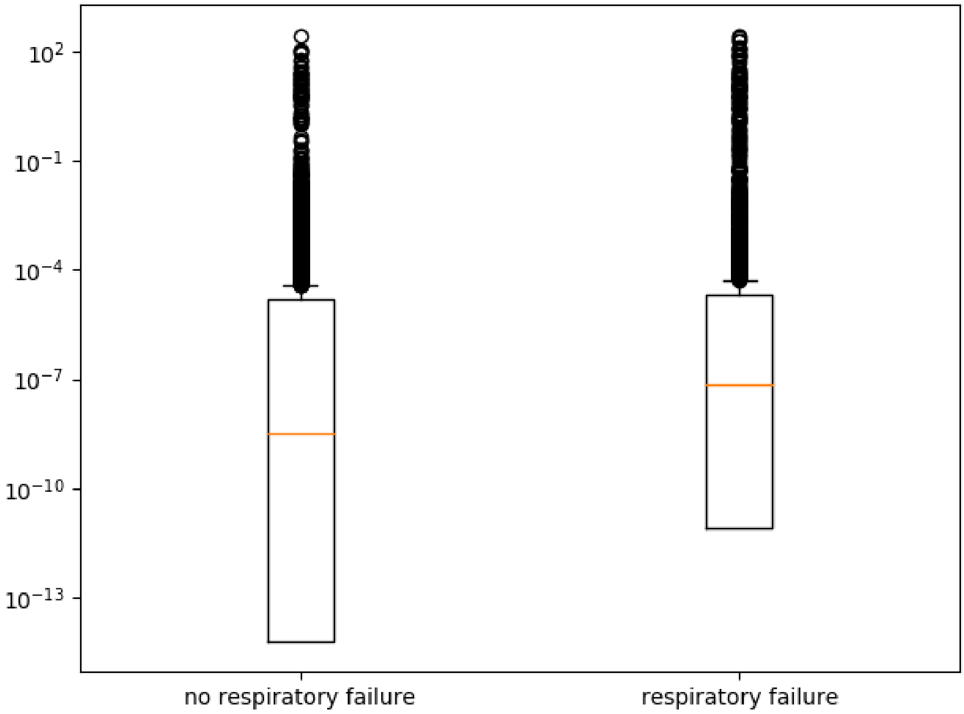

Figure 2. Box plots of the test cross entropy loss for PO2 data that was fed into the discriminator *D*. Note that we used a log scale.

4.1.3. PEEP Data

The test data of the patients who were not suffering from respiratory failure produced an average cross entropy loss of 0.7925 and the test data of the patients who were suffering from respiratory failure an average cross entropy loss of 0.8984. Figure 3 shows a box plot for the test dataset. The plot shows that the losses have approximately the same range. We accepted the hypothesis $H_0 : \mu_1 = \mu_2$ with a *p*-value of 0.1. Figure 4 shows the ROC curve as well as the AUC.

4.2. ECG Data

The test data of normal ECG recordings produced an average cross entropy loss of 0.0021 and the test data of abnormal ECG recordings an average cross entropy loss of 0.0807. Figure 5 shows a box plot for the test dataset. We were able to reject the hypothesis $H_0 : \mu_1 = \mu_2$ in favor of $H_a : \mu_1 > \mu_2$ with a *p*-value of 0.05. Figure 6 shows the ROC curve as well as the AUC.

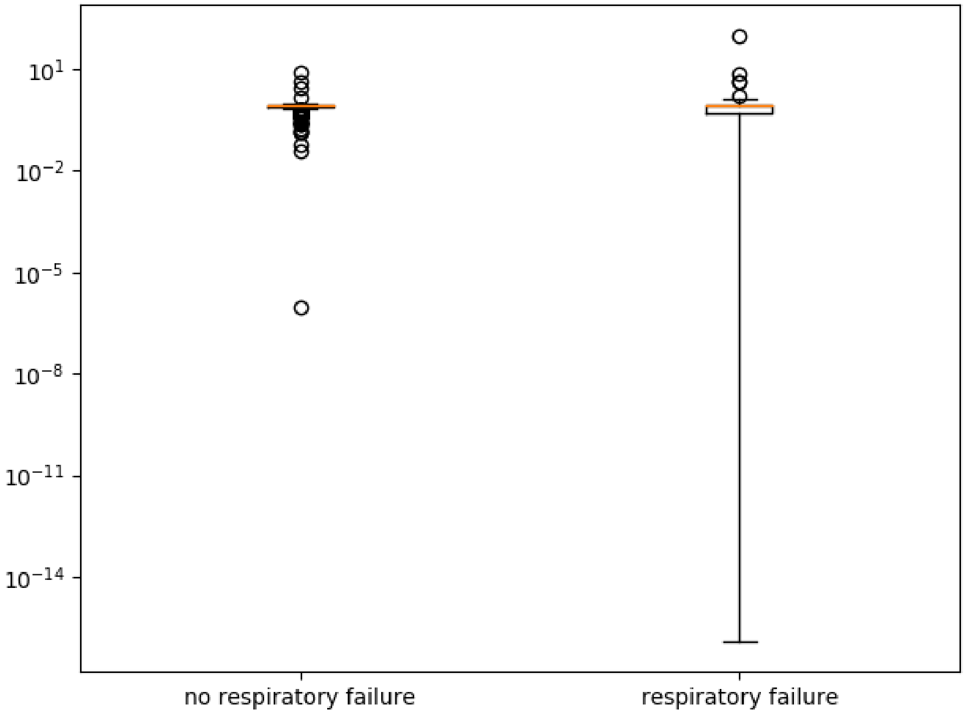

Figure 3. Box plots of the test cross entropy loss for PEEP data that was fed into the discriminator D. Note that we used a log scale.

5. Discussion

5.1. Ventilation Data

Looking at the results in section 4.1, we can observe that the average cross entropy loss was larger for the test data of the patients who were suffering from respiratory failure. This was also verified by means of the t-test with a p-value of 0.1. Only for PEEP data, we could not reject H_0, which could be explained by the significantly less amount of data that was available for PEEP measurements (section 3.2.1). However, as shown by the box plots in section 3.2.1, the losses for the test data of the patients who were suffering from respiratory failure and the test data of the patients who were not suffering from respiratory failure have approximately the same range. This makes it difficult to infer a threshold for the cross entropy loss, where surpassing that threshold could indicate respiratory failure. Furthermore, the ROC curves depicted in figure 4 indicate that this method is not suited to serve as a means of medical diagnosis, because the AUC is below 0.7 for all three cases. This means that either *Generative Adversarial Networks* are not capable of distinguishing between measurement data of the two patient groups or the used data does not contain the information that is needed to perform this distinction.
The generalization of the results may be restricted to some extend by the quality as well as the quantity of the data. Due to the relatively small amount of measurement values per patient and the variety in the number of values per patients.

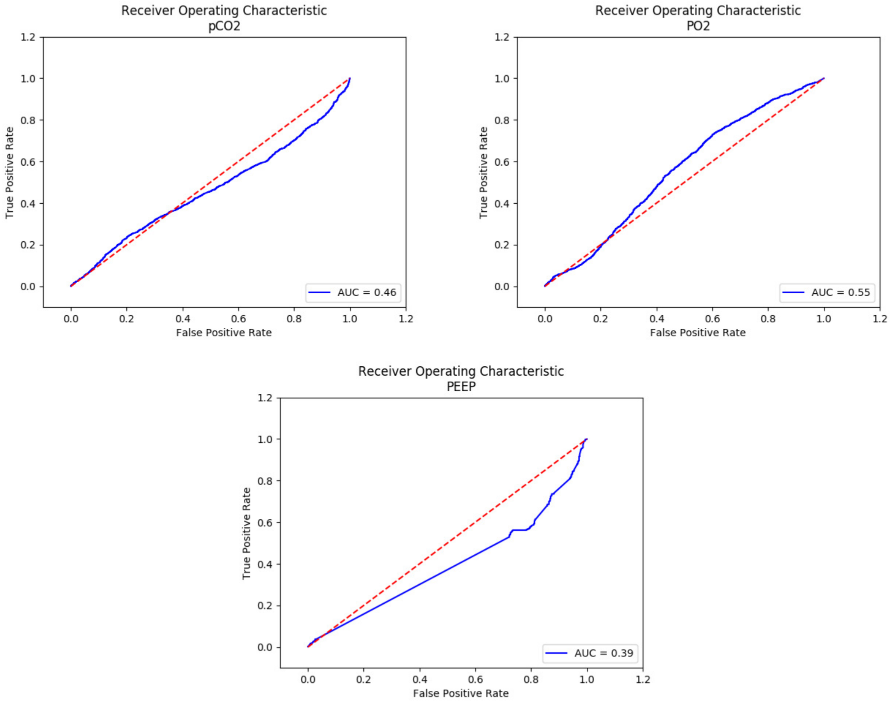

Figure 4. Receiver Operating Characteristic (ROC) curve for ventilation data.

5.2. ECG Data

The results in section 4.2 show that the average cross entropy loss was larger for the test dataset containing abnormal ECG recordings. This was also verified by means of the t-test with a p-value of 0.05. The ROC curve depicted in figure 6 shows an AUC of 0.71, which indicates that *Generative Adversarial Networks* are capable of distinguishing between normal ECG recordings and abnormal ECG recordings and also may be suited to be used as a means of medical diagnosis.

5.3. Future Work

Due to time constraints, some experiments have been left for the future. Since the introduction of *Generative Adversarial Networks* in 2014, different variations and extensions have been proposed. A prominent example, *Wasserstein GANs* (Arjovsky et al., 2017) [25], addresses the instability of the training procedure of *Generative Adversarial Networks*. During the training procedure of our experiments, one of the two models kept getting ahead of the other model (in training loss). When the gap became to big, usually the other model was not able to recover. *Wasserstein GANs* may reduce this problem.
Recurrent Neural Networks have proven to be well suited for classification of sequential data [26]. Recently, they have been applied to the classification of ECG recordings [27,28]. One could implement a generative adversarial network, where the discriminative model D is a recurrent neural network. D would still be trained solely on normal ECG recordings.

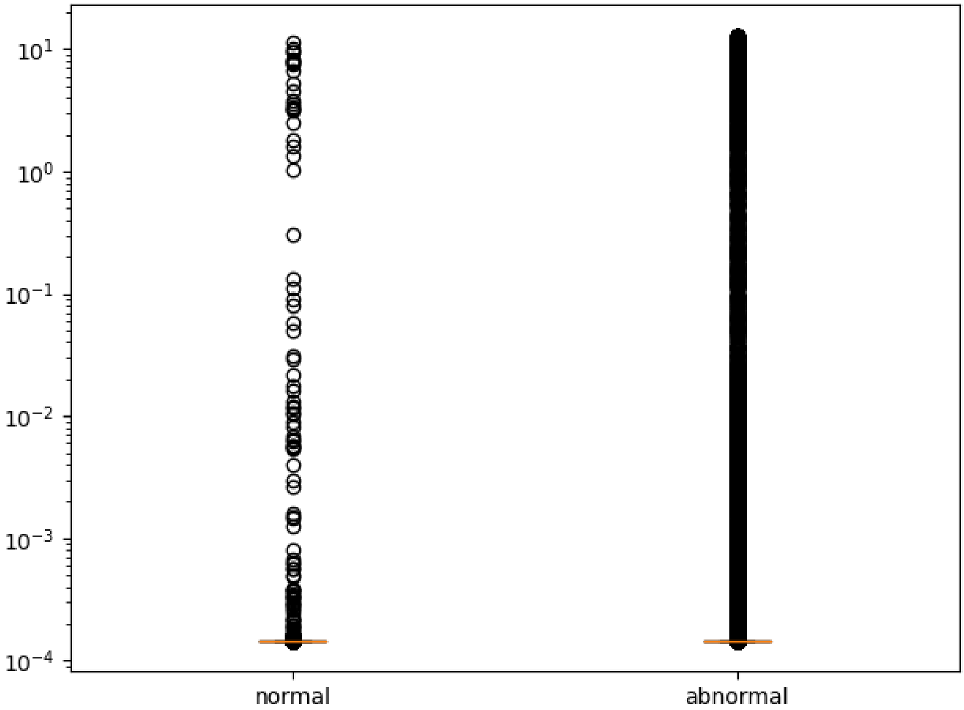

Figure 5. Box plots of the test cross entropy loss for ECG data that was fed into the discriminator *D*. Note that we used a log scale.

6. Conclusion

The objective of this thesis was to evaluate whether *Generative Adversarial Networks* are capable of distinguishing between the data of patients suffering from respiratory failure or cardiac arrhythmia and the data of healthy patients. We were able to show a significant difference between the overall means of the test set containing data of patients with respiratory failure and the test set containing data of patients without respiratory failure. However, the model fell short when it came to the medical diagnosis of individual cases. We attributed this to either the inability of *Generative Adversarial Networks* or to the quality of the used ventilation data.

For our other model, we were also able to show a significant difference between the overall means of the test set containing normal ECG recordings and the test set containing abnormal ECG recordings. Furthermore, the model proved to be suited to be used for medical diagnosis of ECG arrhythmia. This leads us to the overall conclusion that *Generative Adversarial Networks* may be suited for medical diagnosis. This is, however, highly dependent upon the amount and the quality of the available data.

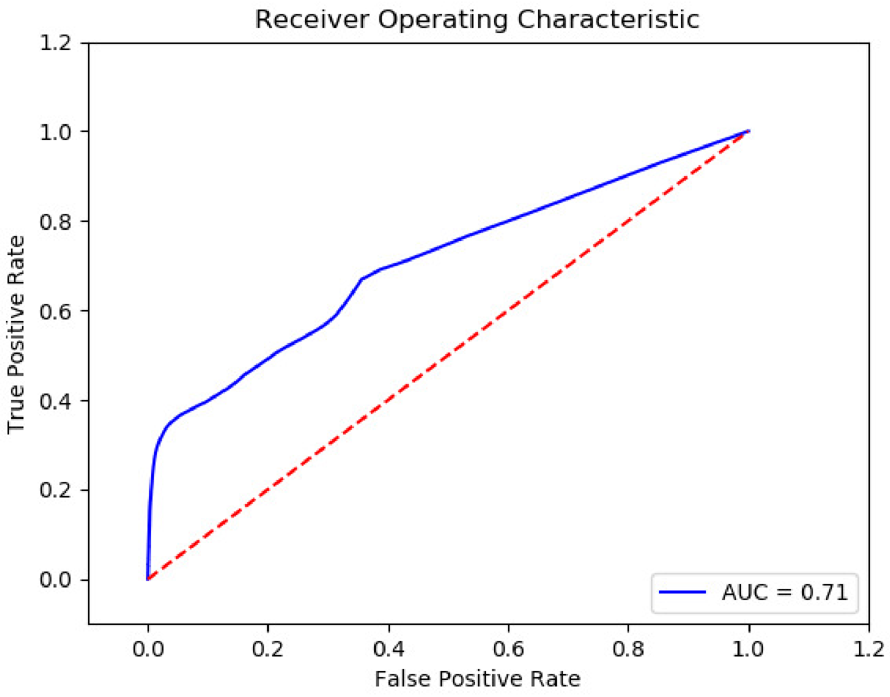

Figure 6. Receiver Operating Characteristic (ROC) curve for ECG data.

References

[1] J. Ramon, D. Fierens, F. Güiza, G. Meyfroidt, H. Blockeel, M. Bruynooghe, and G. Van Den Berghe. *Mining data from intensive care patients.* Advanced Engineering Informatics, 21 (3), p. 243-256, 2007.

[2] H. Wunsch, J. Wagner, M. Herlim, D.H. Chong, A.A. Kramer, and S.D. Halpern. *ICU occupancy and mechanical ventilator use in the United States.* Crit Care Med. 2013 Aug 19; PubMed, 2013.

[3] S. Vasilyev, R.N. Schaap, and J.D. Mortensen. *Hospital survival rates of patients with acute respiratory failure in modern respiratory intensive care units.* Chest 1995;107:1083-1088, 1995.

[4] S. Ridley and J. Purdie. *Causes of death after critical illness.* Anaesthesia 1992, 47, p. 116-119, 1992.

[5] G.A. Miller. *The magical number seven, plus or minus two: some limits on our capacity for processing information.* Psychological Review, 63 (1956), p. 81-97, 1956.

[6] J. Ramon. *Predicting Evolution of Critically Ill Patients.* In Proc. of the KDD 2006 workshop on Theory and Practice of Temporal Data Mining, p. 1-3, 2006.

[7] A. Cariou, D. Bracco, and A. Combes. *Sudden death in ICU: the Finnish experience.* Intensive Care Medicine 40(12), p. 1960-1962, 2014.

[8] O. Lesieur and L. Maxime. *Unexpected Cardiac Arrest in the Intensive Care Unit: State-of the-Art and Perspectives.* Réanimation 26(5), p. 411-424, 2017.

[9] B. Goldstein. *Intensive Care Unit ECG Monitoring.* Cardiac Electrophysiology Review 3, p. 308-310, 1997.

[10] M. AlGhatrif and J. Lindsay. *A brief review: history to understand fundamentals of electrocardiography.* J Community Hosp Intern Med Perspect 2(1), 2012.

[11] K. Polat and S. Güneş. *Detection of ECG Arrhythmia using a differential expert system approach based on principal component analysis and least square support vector machine.* Applied Mathematics and Computation 186(1), p 898-906, 2007.

[12] C. Cortes and V.N. Vapnik. *Support-Vector Networks.* Machine learning 20(3), p. 273-297, 1995.

[13] A. Vishwa, M.K. Lal, S. Dixit, and P. Vardwaj. *Clasification Of Arrhythmic ECG Data Using Machine Learning Techniques* [sic]. International Journal of Interactive Multimedia and Artificial Intelligence, 1 (4), p. 68-71, 2011.

[14] G.B. Moody and R.G. Mark. *The impact of the MIT-BIH Arrhythmia Database.* IEEE Eng in Med and Biol 20(3):45-50 (May-June 2001). (PMID: 11446209)

[15] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. *Generative Adversarial Networks*. ArXiv e-prints. 1406.2661, 2014.

[16] A.E.W. Johnson, T.J. Pollard, L. Shen, L. Lehman, M. Feng, M. Ghassemi, B. Moody, P. Szolovits, L.A. Celi, and R.G. Mark. *MIMIC-III, a freely accessible critical care database.* Scientific Data (2016). DOI: 10.1038/sdata.2016.35. Available at: `http://www.nature.com/articles/sdata201635`

[17] A.L. Goldberger, L.A.N. Amaral, L. Glass, J.M. Hausdorff, P.C. Ivanov, R.G. Mark, J.E. Mietus, G.B. Moody, C.K. Peng, and H.E. Stanley. *PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals.* Circulation 101 (23), e215-e220, 2000.

[18] A.L. Mora Carpio and J.I. Mora. *Positive End-Expiratory Pressure (PEEP).* In: StatPearls. Treasure Island (FL): StatPearls Publishing, 2018. `https://www.ncbi.nlm.nih.gov/books/NBK441904/`.

[19] Glowm. *A free resource for medical professionals.* ISSN: 1756-2228. `https://www.glowm.com/lab_text/item/3`. Accessed: 13.06.2018.

[20] F. Jager, A. Taddei, G.B. Moody, M. Emdin, G. Antolic, R. Dorn, A. Smrdel, C. Marchesi, and R.G. Mark. *Long-term ST database: a reference for the development and evaluation of automated ischaemia detectors and for the study of the dynamics of myocardial ischaemia.* Medical & Biological Engineering & Computing 41(2):172-183 (2003).

[21] M. Arjovsky and L. Bottou. *Towards principled methods for training generative adversarial networks.* International Conference on Learning Representations (ICLR 2017), 2017.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov. *Dropout: A Simple Way to Prevent Neural Networks from Overfitting.* Journal of Machine Learning Research 15, p. 1929-1958, 2014.

[23] A.P. Bradley. *The Use of the Area Under the ROC Curve in the Evaluation of Machine Learning Algorithms.* Pattern Recognition, 30(6), p. 1145-1159, 1997.

[24] J.N. Mandrekar. *Receiver operating characteristic curve in diagnostic test assessment.* J Thorac Oncol. 2010;5:1315-6, 2010.

[25] M. Arjovsky, S. Chintala, and L. Bottou. *Wasserstein GAN.* arXiv preprint arXiv:1701.07875, 2017.

[26] A. Graves. *Supervised Sequence Labelling with Recurrent Neural Networks.* Ph. D. thesis, Technical University of Munich, 2008.

[27] S. Singh, S.K. Pandey, U. Pawar, and R.R. Janghel. *Classification of ECG Arrhythmia using Recurrent Neural Networks.* Procedia Computer Science 132, p. 1290-1297, 2018.

[28] Z. Xiong, M.K. Stiles, and J. Zhao. *Robust ECG signal classification for detection of atrial fibrillation using a novel neural network.* 2017 Computing in Cardiology (CinC), p. 1-4, 2017.