# Noise Masking Recurrent Neural Network for Respiratory Sound Classification: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4–7, 2018, Proceed…

**Chapter** · October 2018

DOI: 10.1007/978-3-030-01424-7_21

CITATIONS
7

READS
895

**5 authors**, including:

Kirill Kochetov
ITMO University
**17** PUBLICATIONS **93** CITATIONS
SEE PROFILE

Evgeny Putin
ITMO University
**37** PUBLICATIONS **971** CITATIONS
SEE PROFILE

Andrey Filchenkov
ITMO University
**119** PUBLICATIONS **257** CITATIONS
SEE PROFILE

# Noise Masking Recurrent Neural Network for Respiratory Sound Classification

Kirill Kochetov(✉), Evgeny Putin, Maksim Balashov, Andrey Filchenkov, and Anatoly Shalyto

Computer Technologies Lab, ITMO University,
49 Kronverksky Pr, 197101 St. Petersburg, Russia
{kskochetov,eoputin,balashov,afilchenkov,shalyto}@corp.ifmo.ru

**Abstract.** In this paper, we propose a novel architecture called noise masking recurrent neural network (NMRNN) for lung sound classification. The model jointly learns to extract only important respiratory-like frames without redundant noise and then by exploiting this information is trained to classify lung sounds into four categories: normal, containing wheezes, crackles and both wheezes and crackles. We compare the performance of our model with machine learning based models. As a result, the NMRNN model reaches state-of-the-art performance on recently introduced publicly available respiratory sound database.

**Keywords:** Respiratory sound classification
Recurrent neural networks · Deep learning

## 1 Introduction

In the last decades many machine learning (ML) approaches have been introduced to analyze respiratory cycle sounds including crackles, coughs, wheezes [1–6]. However almost all conventional ML models solely rely on hand-crafted features. Furthermore, highly complex preprocessing steps are required to make use of designed features [4–6]. Thus, merely ML-based models may not be robust to external/internal noises in lung sounds and may not generalize their performance across different softwares and measuring devices. However, to be used in clinics respiratory tracking systems have to reach high classification accuracy.

From that perspective deep learning (DL) models [7] have gained a lot of attention in the community. DL-based models primary rely on high abstract representation of data that are learned through the training of models. Due to this fact, DL models reach state-of-the-art performance on the range of tasks including image recognition [8], speech recognition [9], time series forecasting [10].

In this work, we propose an architecture of recurrent neural network (RNN) called NMRNN that is trained in end-to-end manner to simultaneously detect noise in respiratory cycles and to classify lung sounds into several categories such as: normal, wheezes, crackles or wheezes and crackles. In other words, our model


V. Kůrková et al. (Eds.): ICANN 2018, LNCS 11141, pp. 208–217, 2018.
https://doi.org/10.1007/978-3-030-01424-7_21

itself decides what information and from what time points it should use to make effective prediction of respiratory sounds. The crucial feature of the model is that it is trained without applying any hand preprocessing stages like slicing of individual respiratory cycles. Through extensive testing, the proposed model has reached state-of-the-art performance on recently published large open database of lung sound records [11].

The rest of the paper is organized as follows. In Sect. 2, we review several notable works in respiratory sounds classification using ML and DL based models. Detailed description of NMRNN is given in Sect. 3. Sections 4 and 5 presents results and comparative study with solely ML-based models. Conclusions are presented in Sect. 6.

## 2 Related Work

Recently a comprehensive comparative study of applying different ML models to automatic wheeze detection was done in [4]. Authors used a lot of models including feed-forward neural network, random forest (RF), support vector machine (SVM) and trained them on two datasets: phonopneumogram samples and the Dubrovnik General Hospital (DGH) dataset. To reduce the influence of cardiovascular and muscular noise, they applied Yule-Walker filter followed by STFT procedure. Then, two types of features were extracted from the lung sounds: MFCC (Mel-frequency cepstral coefficients) features and some statistical features. The authors reported that their best model with statistical features got 93.62% and 91.77% accuracy on phonopneumograms and DGH datasets, accordingly. Meanwhile, based on MFCC features SVM model reached 99% accuracy on both datasets.

In [12], authors proposed to use hidden Markov models (HMM) coupled with Gaussian mixture models (GMM) for classification of respiratory sounds into four categories: normal, containing wheezes, crackles and both crackles and wheezes. The main idea behind applying HMM was that it is able to take into account frame position in a sequence which leads to better accuracy comparing to GMM. To tackle with noise in sound records, they applied spectral subtraction technique [13]. MFCC extracted from the records were used as input features to the model. In addition to MFCC features obtained in range from 50 Hz to 2000 Hz, the first time derivatives of MFCCs were used to track feature dynamics and to decorrelate feature vectors resulting in feature set with size 30. As a result, the ensemble model of 28 HMMs with 5 states and 1 Gaussian per state achieved 0.495 and 0.396 scores on the cross-validation and second evaluation score respectively. In both experiments different patients were used for training and testing, so it was honest validation, and we can compare these results with ours.

One of the most successful attempts of applying DL models to the field of respiratory sound classification was done in [14]. Authors used convolutional neural networks (CNN) to detect wheezes in lung sound records. Firstly, respiratory records were augmented by biasing sound sample in several time frames.

Then, STFT features were computed followed by standard normalization. Lastly, obtained normalized spectrograms of lung sounds were used to train 2D CNN. The final model received 99% accuracy and 0.96 AUC on the dataset.

## 3 Method

RNNs are a class of artificial neural networks (ANNs), which are able to process temporal data, such as sound and text. RNNs can use their internal state (memory) and feedback to process sequences of inputs.

LSTM (Long short-term memory) and GRU (gated recurrent unit) networks [15,16] are popular variants of RNN. They are show unprecedented performance on sequence-related tasks such as NLP (Natural Language Processing) [17] and speech recognition [18].

We use both LSTM and GRU units for our experiments. NMRNN is based on three main ideas:

1. Adapt RNNs, which are designed for time-scale data and can consider all information from sequential frames of input signal.
2. Distinguish noise and content automatically during training.
3. Make predictions using only breath (without noise), because noise can include biased anomalies similar to wheezes or crackles.

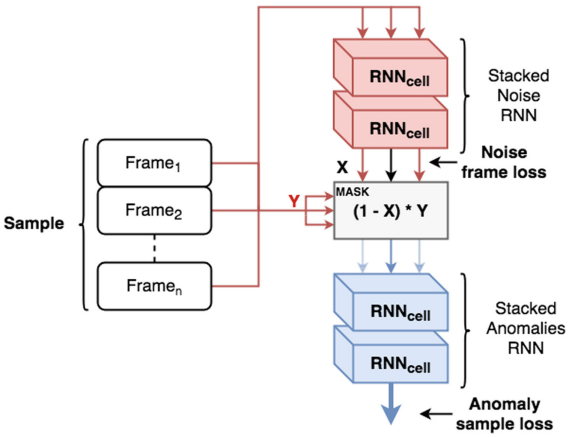


**Fig. 1.** MNRNN architecture. Stacked Noise RNN predicts one noise label per frame using original MFCC data. MASK block adds attention mechanism of the most important frames with respiratory cycles. Stacked Anomalies RNN predicts one anomaly label per sample using highlighted data from the MASK block.

The MNRNN model consists of three parts: noise classifier, respiratory (or anomaly) classifier and some kind of attention called MASK. Schematic overview of the model is shown in Fig. 1.

First of all, before model training each sound sample was split on frames with equal length. There is only one anomaly label for sound sample and one noise label for each frame.

Noise classifier is a stacked RNN called NRNN, which predicts noise label for every frame from the sample. NRNN optimizes a cross-entropy loss calculated for each output during training

$$L_{CE}(p, q) = -\sum p(x) \times \log(q(x)). \tag{1}$$

Then predicted noise labels propagates through masking layer called MASK, where original frames multiplies with masking coefficient $(1 - X) \times Y$, where $X$ is the predicted noise label ($X = 1$ for noise frame) and $Y$ is a frame.

Anomaly classifier is a stacked RNN called ARNN, which predicts one anomaly label for one sample (all frames). ARNN takes highlighted frames from MASK block as input data and optimize cross-entropy loss for one label per sample.

The final loss of the proposed architecture is following:

$$L_{model} = a_1 \times L_{CEnoise} + a_2 \times L_{CEanom}. \tag{2}$$

Values of coefficients $a_1$ and $a_2$ are based on the idea that the main goal of the model is anomaly classification, not noise classification.

The proposed MASK mechanism is simple and efficient and was inspired by gating technique used in GRU cell, where memory needs to be rewritten on each time step using only important information from the input. NRNN parameters were optimized using both NRNN and ARNN losses, so together NRNN and MASK mechanisms allow not only to mask noise frames, but to highlight useful subsamples with respiratory-like content. Attention mechanism used in current model is not the same as usually used for seq2seq models [19]. The main difference is that seq2seq attention mechanism commonly create context vector with weighted sum of encoder hidden states and maps it with current decoder hidden state. So attention in seq2seq extends sight of decoder during sequence prediction. Our MASK layer relies on both predicted noise and anomaly labels, because it receives gradients from both RNN blocks. We conducted additional experiment to show that model with MASK mechanism outperforms model without it in terms of classification metrics.

The main feature of MNRNN method is the ability to perform end-to-end classification without using any manual preprocessing steps like slicing breath on separate cycles. The only commonly used preprocessing step that we did was splitting data to equal frames. The amount of frames does not affect on model training and testing too.

## 4 Experiments

In the study, logistic regression (LR), random forest (RF), gradient boosting machine (GBM), SVM-based classifier [20] and standard RNN were used as

baselines for comparison with the NMRNN model. For baseline experiments, we used the same preprocessing as provided in [4].

### 4.1 Database

For training and evaluation the ICBHI Scientific Challenge database was used [11]. The database contains audio samples, collected independently by two research teams in two different countries over several years. The database consists of 920 annotated audio samples from 126 patients. It includes 6898 different respiratory cycles with 1864 crackles, 886 wheezes and 506 crackles and wheezes. The database summary is presented in Table 1.

There are a lot of noise in sounds: 1840 noise cycles in all data and 1366 in AKGC417L data. It simulates real life conditions and made the classification algorithm more robust and stable for noise attack.

**Table 1.** Database summary. Recordings columns includes statistics about separate sound recordings data. Cycles columns includes statistics about individual respiratory cycles

| Num of | Recordings | | Cycles | |
|---|---|---|---|---|
| | All equipment | AKGC417L | All equipment | AKGC417L |
| Patients | 126 | 56 | 126 | 56 |
| Samples | 920 | 683 | 6898 | 4697 |
| Normal breath | 287 | 196 | 3642 | 2226 |
| Wheezes | 134 | 77 | 886 | 512 |
| Crackles | 297 | 252 | 1864 | 1578 |
| Wheezes and Crackles | 202 | 158 | 506 | 381 |

### 4.2 Experiments Setup

In this work, we conducted several experiments. Different data and preprocessing steps were used for them. The key idea of all experiments is to compare proposed approach with other machine learning models in different situations in terms of performance and robustness.

1. Simple noise binary classification experiment for initial model checking.
2. 4-class anomalies classification using individual respiratory cycle as input.
3. 4-class anomalies classification using sound samples with several respiratory cycles in each (end-to-end classification).

The aim of the first experiment is to check RNN and NMRNN ability to learn respiratory and noise cycle interval lengths and frequencies. The second experiment should compare our baseline models with recently proposed method [12].

The second experiment is demonstrative, but it has one critical limitation: it is not end-to-end experiment, because first of all we need to split lung sounds on respiratory cycles, but there is no automatic universal solution for this task yet. So, for each new lung sound record we need to manually split it into respiratory cycles.

For this reason, the third experiment was conducted. The aim of this experiment is to check the abilities of the models to find what input information is important and where it is located in multidimensional feature space. Model as end-to-end classifier needs to find respiratory-dependent features in the data by itself.

Also, there are two variations of data for each experiment. We use all available data and data recorded only on AKGC417L microphone. The main idea of using second data type is to show that the models can achieve better performance using only one unbiased data source.

All experiments were conducted on a computer with Intel Core i7-6900 CPU with 128GB of RAM and NVIDIA GTX 1080Ti GPU.

### 4.3 Result Evaluation

Due to the unbalanced data set, we used sensitivity and specificity as statistical indicators of the models performance. Sensitivity, specificity and overall score were proposed in the original data set paper [11,12].

Overall evaluation score can be formulated as:

$$Score = \frac{Sensitivity + Specificity}{2}. \tag{3}$$

We used 5-fold cross-validation over patients to evaluate the results and it is important to note that there is no patients from the train set in the test set on each split. So, we used honest real-oriented division of the data for validation.

### 4.4 Preprocessing

To remove sounds caused by heartbeats, the signal components at low frequencies have to be suppressed. We use the high pass finite impulse response (FIR) filter with cutoff frequency $fc = 100$ Hz for remove sounds caused by heartbeat [12].

In this work, MFCC was used as feature extractor. The lower and upper frequencies of processed content were cut to 50 and 2000 Hz respectively, because wheezes and crackles are in this interval [12]. Parameters frame length and frame step were both chosen equal to 0.05 s using grid search optimization [21].

Every sound sample from original database was sliced on pieces called frames with length of 0.5 s each. Every frame was split on 10 non-overlapping frames. Both frame length and frame step are 0.05 s. One MFCC set (13 values) was extracted from each frame. So, every piece is described by 130 MFCC features. Each frame and sample corresponds to a breathing (presence of anomaly) and noise label. There are four breathing classes in the database: normal breathing, breathing with wheezes, crackles and with both wheezes and crackles.

During anomaly classification using all frames (one label per sound) or subset of frames (one label per respiratory cycle) we want to predict existence of anomalies in the overall sound sample or in the only one respiratory cycle respectively. So, for baseline models each sound sample or respiratory cycle was reshaped into a single flattened array. Taking into account different audio lengths, final data samples were cut or filled using standard padding technique. Also, augmentation technique (was proposed in [14]) with shifting was used for solving the problem of respiratory cycles localization. PCA (Principal Component Analysis) was used for dimensionality reduction (only for baseline models).

## 5 Results

For noise binary classification task NMRNN achieved 0.89 evaluation score compared with the best baseline model GBM, which reached only 0.53 score. It can be explained by the ability of RNN to learn cycle and noise intervals length and frequency and use this additional information during prediction.

**Table 2.** Results of 4-class classification of each respiratory cycle. Metrics of Jakovljevic HMM was not provided with AKGC417L data

| | All equipment | | | AKGC417L | | |
|---|---|---|---|---|---|---|
| Model | Sens | Spec | Score | Sens | Spec | Score |
| GBM | 0.476 | 0.554 | 0.515 | 0.534 | 0.568 | 0.551 |
| LR | 0.425 | 0.508 | 0.466 | 0.426 | 0.51 | 0.468 |
| RF | 0.438 | 0.538 | 0.488 | 0.483 | 0.521 | 0.502 |
| SVM | 0.49 | 0.502 | 0.496 | 0.502 | 0.518 | 0.51 |
| Jakovljevic [12] | 0.423 | 0.567 | 0.495 | - | - | - |
| RNN (ours) | 0.584 | 0.73 | **0.657** | 0.617 | 0.741 | 0.679 |

Results of 4-class classification of each respiratory cycle are presented in Table 2. There is a comparison of our baseline and NMRNN models with HMM-based method proposed by Jakovljevic. All models were trained on MFCC features. Performance of our models is similar with performance of Jakovljevic HMM [12], except for NMRNN, which outperforms competitors. So, it is correct to compare presented baseline models with proposed RNN-based approach in the next experiment. Also, models trained only on AKGC417L data show better scores as expected due to reduced bias of data distribution. The second experiment is less complex than the third one, because of data manually sliced on respiratory cycles before training.

Results of end-to-end classification are provided in Table 3. NMRNN definitely outperforms other methods with respect to the chosen criterion. The main reason is that RNN was designed to process such kind of data with temporal dependencies. Another models face with problems of large dimensionality

**Table 3.** Results of 4-class classification of each sound sample

| | All equipment | | | AKGC417L | | |
|---|---|---|---|---|---|---|
| Model | Sens | Spec | Score | Sens | Spec | Score |
| GBM | 0.362 | 0.142 | 0.252 | 0.348 | 0.174 | 0.261 |
| LR | 0.348 | 0.184 | 0.266 | 0.366 | 0.236 | 0.301 |
| RF | 0.433 | 0.054 | 0.244 | 0.451 | 0.079 | 0.265 |
| SVM | 0.313 | 0.251 | 0.282 | 0.278 | 0.256 | 0.267 |
| RNN (ours) | 0.511 | 0.717 | 0.614 | 0.572 | 0.728 | 0.65 |
| NMRNN (ours) | 0.56 | 0.736 | **0.648** | 0.62 | 0.75 | **0.685** |

and localization of respiratory cycles. So, neither PCA or augmentation do not help to solve these problems, because the baseline models are not adapted for unstable data with floating content such as sound with several respiratory cycles.

MASK block with noise classification increases performance on about 0.035 in terms of score. It can be explained by ability of the final model to concentrate only on frames with respiratory cycles, not with noise. Also MASK block helps to distinguish false positive anomalies (biased noise) with real anomalies (crackles or wheezes) as justified on Fig. 2.

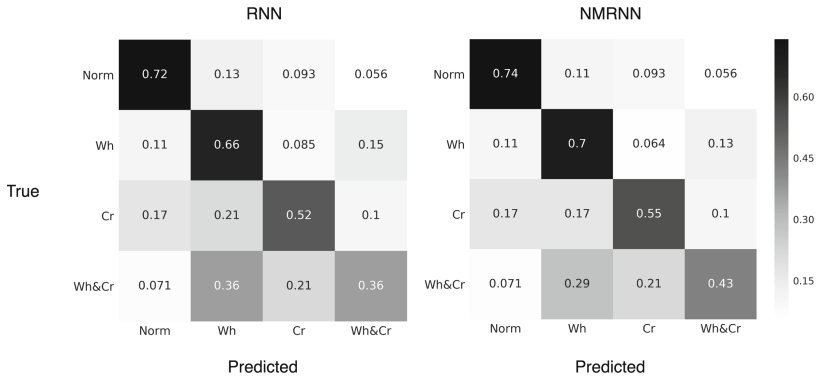


**Fig. 2.** Confusion matrices of RNN and NMRNN. MASK block helps to clarify some samples similarity by masking false positive anomalies detected in noise frames. Due to that both sensitivity and specificity was improved.

Models trained only on AKGC417L data show performance as in the previous experiments. This proves that the model can be adapted for single source and can in theory boost performance with increasing of amount of unbiased data for training.

We used grid search [21] as optimization algorithm for finding best hyper-parameters for baseline and RNN-based models. So the best RNN-based model

with MASK block consists of 2-layer RNNs as both NRNN and ARNN parts with GRU cells with 256 units in each. Coefficients $a_1$ and $a_2$ from Eq. 2 are 0.3 and 0.7 respectively, which corresponds to the main task of the model (anomaly classification). Overall model architecture was trained using Adam [22] optimizer with $learning\_rate = 0.0001$.

## 6 Conclusion

In this paper, we proposed RNN-based end-to-end model architecture called NMRNN to detect different anomalies in lung sound data with masking of noise. MASK block is very powerful, so it allows the model to consider only relevant frames during classification. We assume, that the trained MASK mechanism is a superior direction of further improvement.

The main contribution of this approach is that it is trained without applying any manual preprocessing steps using respiratory records of any lengths. NMRNN reaches state-of-the-art performance in comparison with another ML models on respiratory sound classification task and, including recently proposed [12], on individual respiratory cycle classification task.

Also, this study shows the ability of the model to learn cycle and the lengths of noise intervals and frequencies. Experiments with AKGC417L microphone motivate to concentrate on single data source during creation of approach applicable in real life conditions.

**Acknowledgements.** This work was financially supported by the Government of the Russian Federation, Grant 08-08.

## References

1. Bahoura, M., Pelletier, C.: Respiratory sounds classification using cepstral analysis and Gaussian mixture models. In: 26th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, IEMBS 2004, vol. 1, pp. 9–12. IEEE (2004)
2. Mayorga, P., Druzgalski, C., Morelos, R.L., Gonzalez, O.H., Vidales, J.: Acoustics based assessment of respiratory diseases using GMM classification. In: 2010 Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 6312–6316. IEEE (2010)
3. Palaniappan, R., Sundaraj, K., Sundaraj, S.: A comparative study of the SVM and K-NN machine learning algorithms for the diagnosis of respiratory pathologies using pulmonary acoustic signals. BMC Bioinform. **15**(1), 223 (2014)
4. Milicevic, M., Mazic, I., Bonkovic, M.: Classification accuracy comparison of asthmatic wheezing sounds recorded under ideal and real-world conditions. In: 15th International Conference on Artificial Intelligence, Knowledge Engineering and Databases (AIKED 2016), Venice (2016)
5. Rocha, B.M., Mendes, L., Chouvarda, I., Carvalho, P., Paiva, R.P.: Detection of cough and adventitious respiratory sounds in audio recordings by internal sound analysis. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds.) Precision Medicine Powered by pHealth and Connected Health. IP, vol. 66, pp. 51–55. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7419-6_9

6. Serbes, G., Ulukaya, S., Kahya, Y.P.: An automated lung sound preprocessing and classification system based onspectral analysis methods. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds.) Precision Medicine Powered by pHealth and Connected Health. IP, vol. 66, pp. 45–49. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7419-6_8
7. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436 (2015)
8. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: AAAI, vol. 4, p. 12 (2017)
9. Palaz, D., Magimai-Doss, M., Collobert, R.: Analysis of CNN-based speech recognition system using raw speech as input. Technical report, Idiap (2015)
10. Weigend, A.S.: Time Series Prediction: Forecasting the Future and Understanding the Past. Routledge, New York (2018)
11. Rocha, B.M., et al.: A respiratory sound database for the development of automated classification. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds.) Precision Medicine Powered by pHealth and Connected Health. IP, vol. 66, pp. 33–37. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7419-6_6
12. Jakovljević, N., Lončar-Turukalo, T.: Hidden Markov model based respiratory sound classification. In: Maglaveras, N., Chouvarda, I., de Carvalho, P. (eds.) Precision Medicine Powered by pHealth and Connected Health. IP, vol. 66, pp. 39–43. Springer, Singapore (2018). https://doi.org/10.1007/978-981-10-7419-6_7
13. Berouti, M., Schwartz, R., Makhoul, J.: Enhancement of speech corrupted by acoustic noise. In: IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 1979, vol. 4, pp. 208–211. IEEE (1979)
14. Kochetov, K., Putin, E., Azizov, S., Skorobogatov, I., Filchenkov, A.: Wheeze detection using convolutional neural networks. In: Oliveira, E., Gama, J., Vale, Z., Lopes Cardoso, H. (eds.) EPIA 2017. LNCS (LNAI), vol. 10423, pp. 162–173. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-65340-2_14
15. Hochreiter, S., Schmidhuber, J.: Long short-term memory. Neural Comput. **9**(8), 1735–1780 (1997)
16. Cho, K., Van Merriënboer, B., Bahdanau, D., Bengio, Y.: On the properties of neural machine translation: encoder-decoder approaches. arXiv preprint arXiv:1409.1259 (2014)
17. Sundermeyer, M., Schlüter, R., Ney, H.: LSTM neural networks for language modeling. In: Thirteenth Annual Conference of the International Speech Communication Association (2012)
18. Graves, A., Mohamed, A., Hinton, G.: Speech recognition with deep recurrent neural networks. In: 2013 IEEE international conference on Acoustics, speech and signal processing (ICASSP), pp. 6645–6649. IEEE (2013)
19. Luong, M.-T., Pham, H., Manning, C.D.: Effective approaches to attention-based neural machine translation. arXiv preprint arXiv:1508.04025 (2015)
20. Hastie, T., Tibshirani, R., Friedman, J.: The Elements of Statistical Learning. SSS. Springer, New York (2009). https://doi.org/10.1007/978-0-387-84858-7
21. Bergstra, J., Bengio, Y.: Random search for hyper-parameter optimization. J. Mach. Learn. Res. **13**, 281–305 (2012)
22. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)