2021 43rd Annual International Conference of the
IEEE Engineering in Medicine & Biology Society (EMBC)
Oct 31 - Nov 4, 2021. Virtual Conference

Deep Learning End-to-End Approach for the Prediction of Tinnitus based on EEG Data*

Johannes Allgaier[1] and Patrick Neff[2] and Winfried Schlee[2] and Stefan Schoisswohl[2] and Rüdiger Pryss[1]

Abstract— **Tinnitus is attributed by the perception of a sound without any physical source causing the symptom. Symptom profiles of tinnitus patients are characterized by a large heterogeneity, which is a major obstacle in developing general treatments for this chronic disorder. As tinnitus patients often report severe constraints in their daily life, the lack of general treatments constitutes such a challenge that patients crave for any kind of promising method to cope with their tinnitus, even if it is not based on evidence. Another drawback constitutes the lack of objective measurements to determine the individual symptoms of patients. Many data sources are therefore investigated to learn more about the heterogeneity of tinnitus patients in order to develop methods to measure the individual situation of patients more objectively. As research assumes that tinnitus is caused by processes in the brain, electroencephalography (EEG) data are heavily investigated by researchers. Following this, we address the question whether EEG data can be used to classify tinnitus using a deep neural network. For this purpose, we analyzed 16,780 raw EEG samples from 42 subjects (divided into tinnitus patients and control group), with a duration of one second per sample. Four different procedures (with or without *noise reduction* and *down-sampling* or *up-sampling*) for automated preprocessing were used and compared. Subsequently, a neural network was trained to classify whether a sample refers to a tinnitus patient or the control group. We obtain a maximum accuracy in the test set of 75.6% using noise reduction and down-sampling. Our findings highlight the potential of deep learning approaches to detect EEG patterns for tinnitus patients as they are difficult to be recognized by humans.**

I. Introduction

Many individuals experience a constant noise in their ears, which is widely known as tinnitus, further described as a whistling or ringing sound in the ears [1]. In terms of prevalence, about 10 - 15% of the worldwide population report this kind of symptoms [2], [3]. Although many people perceiving tinnitus do not experience a considerable burden, about 2.4% of the worldwide population severely suffers from tinnitus on a daily basis [4]. In most of these cases, tinnitus is a subjective perception that can only be perceived by the affected person. The physiological mechanisms of tinnitus are still being subject to research. At present, it is known that tinnitus is a heterogeneous phenomenon with many possible causes and subtypes [5], [6]. These subtypes are not necessarily disjoint and can therefore be combined. However, one issue unites all tinnitus subtypes: The perception of a phantom sound. Our general hypothesis is that this perception must be encoded in brain activities. In general, recent research shows that tinnitus is highly relevant in the context of clinical neuroscience [7], [8].

Brain activity, in turn, can be measured using electroencephalography (EEG) devices, but the recognition of patterns within the collected signal data is very complex, even for experts. Experts, in turn, constitute a rare resource and have only limited time. We therefore want to investigate whether a deep learning based end-to-end approach can distinguish the patterns of tinnitus patients from non-tinnitus patients using a small sample of 42 individuals. `End-to-end` means that the system itself can learn the necessary features to solve the task. This contrasts with feature engineering in classical machine learning approaches. In the first step, automated pre-analyses on resting-state data could be successfully accomplished, which can also accelerate processes in the daily clinical routine. It was effectively revealed that deep learning offers a methodology with high potential for predictive accuracy, efficiency, robustness, automation, and generalizability for the question at hand. In particular, our results unveil the following:

- In order to gain insights relevant for tinnitus researchers in their daily practice, the task to learn more about EEG-Data in the context of tinnitus using machine learning constitutes a highly interdisciplinary venture.
- It is still a debate to which extent and level preprocessing should be accomplished.
- Our results indicate that based on machine learning and EEG data, tinnitus can be classified to a level which justifies further investigations.
- We have decided to use an end-to-end approach for the investigations, accepting the potential limitations [9]. However, extant research in the context of EEG data has shown that end-to-end learning approaches can provide meaningful results [10].

The remainder of this paper is structured as follows. Section II discusses related work, while Section III presents the applied methods and used materials. In Section IV, the results are presented, whereas Section V discusses them. The paper concludes with a summary and outlook in Section VI.

II. Related Work

Various studies have investigated tinnitus in the context of EEG resting state data [11]. In [12], the EEG signals of people with various disorders (depression, tinnitus, or Parkinson's disease) were compared with healthy subjects.

*This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement Number 848261.

[1]J. Allgaier (johannes.allgaier@uni-wuerzburg.de) and R. Pryss (ruediger.pryss@uni-wuerzburg.de) are with the Institute of Clinical Epidemiology and Biometry, University of Würzburg, Würzburg, Germany.

[2]P. Neff (patrick.neff@ukr.de), S. Schoisswohl (stefan.schoisswohl@ukr.de), and W. Schlee (winfried.schlee@gmail.com) are with the Department of Psychiatry and Psychotherapy, University of Regensburg, Regensburg, Germany.

Thereby, an increased low-frequency rhythmicity was found. The authors of [13], in turn, used one second of raw EEG data and trained a convolutional neural network to automatically differentiate sub-anesthetic states and depths of anesthesia. Deep learning models have been used to decode or visualize EEG data [14]. Furthermore, EEG data were used to distinguish sleeping from non-sleeping individuals [15]. In the context of tinnitus, it was shown that prediction-specific neural information of auditory encoding seems to strongly differ between tinnitus patients and a control group [16]. Furthermore, support vector machines have been used to classify tinnitus patients on preprocessed data [17]–[19]. It has also been shown that tinnitus individuals can be distinguished from non-tinnitus individuals by means of EEG in the gamma and alpha wave ranges [20]–[22]. In the field of end-to-end learning, extant research has shown that the application of end-to-end deep learning approaches to EEG data can reveal tangible insights [10], [23], and is therefore used also in the work at hand. Furthermore, to the best of our knowledge, no deep learning end-to-end approach has been applied to classify tinnitus patients using EEG data.

III. Materials and Methods

The following chapter is divided into three parts. In the first part, patient data are presented, while in the second part, data preprocessing is discussed. Finally, details about the deep learning approach are presented, in particular, the architecture of the neural network as well as its output generating process.

A. Data Set

The data set consists of a total of 29 tinnitus patients and 13 individuals in the control group. The data were collected using hardware from Brainproducts GmbH, Gilching, Germany. The recordings with a sampling frequency of 500 Hz were recorded in the resting state. The data were stored in a BrainVision data format [24]: One electroencephalography (.eeg) file, containing binary data from 62 EEG and 2 ECG channels (i.e., the voltage values), a text header file (.vhdr), containing meta data, and a text marker file (.vmrk), containing information about the events in the data. Note that project is carried out within the scope of the UNITI project, which aims to unify treatment and interventions for tinnitus patients in a pan-European setting [25].

B. Preprocessing

The used neural network only accepts matrices containing numbers as an input. Therefore, the signals from the binary .eeg files must be converted into matrices. We solved this problem using libraries from Python, namely MNE [26], NumPy, and Pandas. MNE reads .vhdr files in a raw object, which have $n\_channels$ as height and $n\_measurements$ as width. We chose one second as the sample length. Since the data are recorded at 500 Hz, one sample is sized at 62 rows x 500 columns. Following this, we had about 320,000 measurements per individual, which resulted in about 642 samples per individual for the feeding procedure of the neural network. Further note that the 642 samples are all labeled with tinnitus (=1) or no-tinnitus (=0), respectively.

To reduce noise in the data, we used the Autoreject library [27], which is a library for automatically deleting bad trials and repairing bad sensors in EEG data. We used Autoreject due to the following reasons: It is completely automated, could potentially be developed in real time and is therefore workable in any clinical setting. Data coming from one individual can finally be summarized to the following shape: $(n\_samples, n\_channels, n\_measurements)$ = (642, 62, 501). The corresponding label vector is denoted with $(n\_samples, 1)$. The neural network was then given one second of a non-overlapping 62-channel EEG signal. Similar approaches can be found in [11], [28], [29]. After converting the raw data into slices of NumPy arrays, we split the data into two sets: 70% are stored for training and hyperparameter tuning (validation), and 30% are preserved in a test set. As another information, we used a hierarchic data file structure format (.hdf5).

To get a better performance when training the neural network, we further normalized the used data. More specifically, we subtracted the mean for each channel and divided it by its standard deviation. We further created a balanced dataset. As we have 29 individuals with tinnitus, but only 13 controls, a naive classifier would get roughly 69% accuracy in both training and testing if always classifying *tinnitus*. To prevent this, we randomly drew samples from the tinnitus group until we had a balanced data set, which we call the down-sampling approach. In the up-sampling approach, we randomly drew samples with replacement from the control group until the dataset was balanced again. Finally, we pseudo-shuffled the data using Python's `random_seed = 0`, and saved it as a new .hdf5 dataset. All computations were done with Python3, Keras, and TensorFlow, on a Core i7 9th gen. CPU. GPU computing would have been faster, but could bias the overall reproducibility [30].

C. Deep Learning Architecture

We used Keras for the neural network architecture and the actual machine learning process. 20% of the training data were used to validate the current architectural approach, the remaining 80% were used for the training. We set `random_seeds` for NumPy, TensorFlow and the OS environment to ensure the model is deterministic and changes in architecture do not *randomly* improve the model. The training process then was configured as follows: We defined a callback class to force the model stopping earlier when the accuracy gets above a certain threshold. This approach saved time in training, helped us to speed up the iterations in the development process, and prevented overfitting. The architecture of the model is described in Fig. 1. The model has a total of 76,208,129 trainable parameters, 5 dense layers, one dropout layer, and one flatten layer. As this is a binary classification task, a sigmoid layer is used at the end. The batch size was set to 32 and the neural network was trained for one epoch at a time. A larger batch size led to memory issues, but with a smaller batch size, it is

TABLE I

DEMOGRAPHIC CHARACTERISTICS OF THE DATA SET. UNPAIRED, TWO-SIDED T-TESTS AND TWO-TAILED CHI-SQUARE TESTS WERE USED TO PROBE THE GROUPS FOR DIFFERENCES. P(STAT): P-VALUE OF STATISTICAL TEST. SD: STANDARD DEVIATION.

	Tinnitus		**Controls**		
Variable	**Mean**	**SD**	**Mean**	**SD**	**p(stat)**
Age	53.11	9.26	56.46	13.75	0.36
Hearing loss	29.35	10.78	35.24	18.45	0.2
Sex (f/m)	4	25	2	11	0.89

difficult to ensure that the target remains balanced within one iteration of gradient descent. The model was compiled using a `binary_crossentropy` and RMSprop with a learning rate of 0.001 as an optimizer [31]. For validation, the accuracy score has been used. Every change in the architecture was finally evaluated in the hold-out test set. If the model improved by more than 1 percent point, it was saved in the history folder.

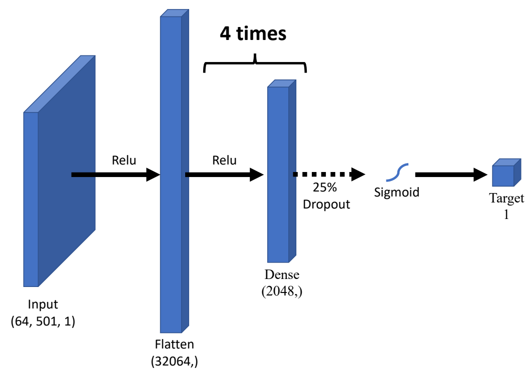

Fig. 1. Architecture of the fully connected neural network. Each sample gets rolled out in a flatten layer with shape 32,064. This layer is passed to a total of 4 fully connected layers, which then process the input to features.

IV. RESULTS

In this section, we present four variants of our machine learning approach. Using Autoreject, noise reduction can be (1) *true* or (2) *false* and the data can be (3) *up-* or (4) *down-sampled*, as described in the preparation subsection.

TABLE II

ACCURACY SCORES FOR THE FOUR DIFFERENT MACHINE LEARNING APPROACHES. ALL HYPER-PARAMETERS OF THE NEURAL NETWORK WERE FIXED. THE NUMBER IN (BRACKETS) DENOTES THE TEST SIZE.

	Noise reduction true	**Noise reduction false**
Down-sampled	0.76 (5,140)	0.66 (5,140)
Up-sampled	0.72 (11,576)	0.70 (11,576)

We achieved the best accuracy score with noise reduction using Autoreject and the down-sampling approach. The final test data set contained 5,140 samples, each with one second of EEG recordings. The use of Autoreject improved the accuracy score in the down-sampling approach by 10 percentage points and in the up-sampling approach by 2 percentage points. Due to the increased amount of data in the up-sampling approach, the neural network might have learned to distinguish the noise from the tinnitus pattern. This assumption is supported by the larger difference of Autoreject *true* or *false* between the two approaches down- or up-sampling.

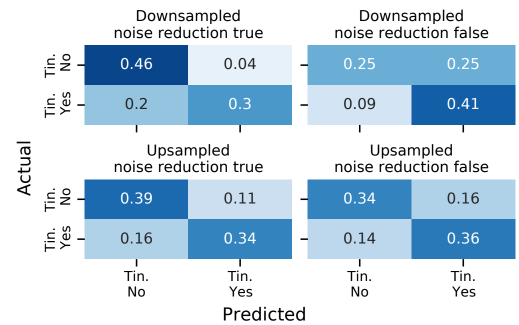

Fig. 2. Confusion matrices for all four approaches. *Tin.* = Tinnitus. For each of the four confusion matrices, the values sum up to 100%. Within each matrix, 0.5 for top left and bottom right would be the optimum.

The confusion matrix in Fig. 2 shows that the down-sampled approach with noise reduction `true` has a better score for the control group (top left matrix with `0.46`). Setting noise reduction on `false`, the score in the down-sampled approach is higher for tinnitus individuals (top right matrix with `0.41`). In the up-sampled approach, the neural network identifies samples from the control group several times. From Fig. 2, we can see that up-sampling does not necessarily lead to an improvement of the values, although all available data from the tinnitus group were used for training and testing.

V. DISCUSSION

It is questionable why the neural network is not better in the up-sampling approach than in the down-sampling approach. According to the general understanding, more data from the same distribution should lead to an improved model, given the fact that the test data also come from this distribution. By adding more data, however, we would expect the performance gap between Autoreject true or false to shrink further, meaning the model would better learn to distinguish artifacts from tinnitus patterns. When it comes to classification, a balanced data set is rarely available. A class in the target is usually (strongly) underrepresented. We made our dataset sufficiently balanced to work without the need of additional balancing. However, one could have also trained the data in a 29 to 13 ratio (29 tinnitus and 13 controls), and then set the baseline accuracy to 69%. A naive classifier would then always have classified *tinnitus*. Or, one could have set a high penalty term for false positives in the loss function. We decided against both variants, because the accuracy score is better comparable and interpretable than other scores, especially beyond this paper. The approach presented in this work also poses several limitations. First, the amount of used EGG data might be not enough in order to be representative. In addition, it is not clear whether the included participants are representative for a broad tinnitus patient audience. Furthermore, a more fine-grained approach compared to an end-to-end approach might produce better results. Finally, preprocessing is always a subject to debate. However, we used an off-the-shelf mechanism in this context, which should limit potential drawbacks to a minimum.

VI. Summary and Outlook

In this paper, we used raw EEG data from 42 individuals (29 tinnitus, 13 controls) in an end-to-end approach to distinguish subjects with tinnitus from those without tinnitus. The neural network was given one second (500 observations at 500 Hertz) of 62 EEG channels as input and learned a feature space in a fully connected 5-layer neural network with over 76 million parameters. We therefore varied four approaches: Up- and down-sampling and with or without Autoreject. The best result in the test set was 75.6% accuracy for noise reduced data (Autoreject true) and with a down-sampling approach. To further improve the accuracy score, the feature space could be increased. So, one could attach more sensors like ECG data, put them into a separate neural network, and classify them with an ensemble method by majority vote similar to a random forest approach. Note that other works have already shown the feasibility of ensemble methods in the given context [32], [33]. However, an increase of the feature space also entails an increase of the data volume to avoid overfitting. By adding more data from different distributions, we expect better model performance. With a better classification rate, the model also has the potential to be deployed as a real-time classifier. To conclude, our end-to-end deep neural network approach on EEG data justifies further investigations for tinnitus in particular and other medical questions in general.

Acknowledgment

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme, Grant Agreement Number 848261. The studies were approved by the Ethics Committee of the University Clinic of Regensburg (ethical approval No. 17-819-101 and 18-1054-101). All users read and approved the informed consent before participating in the study. The study was carried out in accordance with relevant guidelines and regulations.

References

[1] N. Kiang *et al.*, "Auditory-nerve activity in cats with normal and abnormal cochleas," *Sensorineural hearing loss*, pp. 241–273, 1970.
[2] A. Davis and E. A. Rafaie, "Epidemiology of tinnitus," *Tinnitus handbook*, vol. 1, p. 23, 2000.
[3] B. Langguth, "A review of tinnitus symptoms beyond 'ringing in the ears': a call to action," *Current medical research and opinion*, vol. 27, no. 8, pp. 1635–1643, 2011.
[4] J. B. Halford and S. D. Anderson, "Anxiety and depression in tinnitus sufferers," *Journal of psychosomatic research*, vol. 35, no. 4-5, pp. 383–390, 1991.
[5] B. Langguth *et al.*, "Different patterns of hearing loss among tinnitus patients: a latent class analysis of a large sample," *Frontiers in neurology*, vol. 8, p. 46, 2017.
[6] C. Cederroth *et al.*, "Towards an understanding of tinnitus heterogeneity," *Frontiers in aging neuroscience*, vol. 11, p. 53, 2019.
[7] P. Neff *et al.*, "The impact of tinnitus distress on cognition," *Scientific reports*, vol. 11, no. 1, pp. 1–9, 2021.
[8] W. Schlee *et al.*, "Using big data to develop a clinical decision support system for tinnitus treatment." *Current Topics in Behavioral Neurosciences*, 2021.
[9] T. Glasmachers, "Limits of end-to-end learning," in *Asian Conference on Machine Learning*. PMLR, 2017, pp. 17–32.
[10] H. Dose *et al.*, "An end-to-end deep learning approach to mi-eeg signal classification for bcis," *Expert Systems with Applications*, vol. 114, pp. 532–542, 2018.
[11] Y. Roy *et al.*, "Deep learning-based electroencephalography analysis: a systematic review," *Journal of neural engineering*, vol. 16, no. 5, p. 051001, 2019.
[12] R. Llinas *et al.*, "Thalamocortical dysrhythmia: a neurological and neuropsychiatric syndrome characterized by magnetoencephalography," *Proc of the National Academy of Sciences*, vol. 96, no. 26, pp. 15 222–15 227, 1999.
[13] K. Patlatzoglou *et al.*, "Generalized prediction of unconsciousness during propofol anesthesia using 3d convolutional neural networks," in *42nd Annual Intl Conf of the IEEE Engineering in Medicine & Biology Society*. IEEE, 2020, pp. 134–137.
[14] R. Schirrmeister *et al.*, "Deep learning with convolutional neural networks for eeg decoding and visualization," *Human brain mapping*, vol. 38, no. 11, pp. 5391–5420, 2017.
[15] A. Patanaik *et al.*, "An end-to-end framework for real-time automatic sleep stage classification," *Sleep*, vol. 41, no. 5, p. zsy041, 2018.
[16] M. Partyka *et al.*, "Phantom auditory perception (tinnitus) is characterised by stronger anticipatory auditory predictions," *bioRxiv*, p. 869842, 2019.
[17] F. Mohagheghian *et al.*, "Computer-aided tinnitus detection based on brain network analysis of eeg functional connectivity," *Journal of biomedical physics & engineering*, vol. 9, no. 6, p. 687, 2019.
[18] S. Vanneste *et al.*, "Thalamocortical dysrhythmia detected by machine learning," *Nature communications*, vol. 9, no. 1, pp. 1–13, 2018.
[19] S. Wang *et al.*, "Tinnitus eeg classification based on multi-frequency bands," in *Intl Conf on Neural Information Processing*. Springer, 2017, pp. 788–797.
[20] N. Weisz *et al.*, "Tinnitus perception and distress is related to abnormal spontaneous brain activity as measured by magnetoencephalography," *PLoS Med*, vol. 2, no. 6, p. e153, 2005.
[21] ——, "The neural code of auditory phantom perception," *Journal of Neuroscience*, vol. 27, no. 6, pp. 1479–1484, 2007.
[22] W. Schlee *et al.*, "Reduced variability of auditory alpha activity in chronic tinnitus," *Neural plasticity*, vol. 2014, 2014.
[23] . Miladinović *et al.*, "Spindle: End-to-end learning from eeg/emg to extrapolate animal sleep scoring across experimental settings, labs and species," *PLoS computational biology*, vol. 15, no. 4, p. e1006968, 2019.
[24] R. Oostenveld *et al.*, "Fieldtrip: open source software for advanced analysis of meg, eeg, and invasive electrophysiological data," *Computational intelligence and neuroscience*, vol. 2011, 2011.
[25] W. Schlee *et al.*, "Towards a unification of treatments and interventions for tinnitus patients: The eu research and innovation action uniti." *Progress in brain research*, vol. 260, pp. 441–451, 2021.
[26] A. Gramfort *et al.*, "Meg and eeg data analysis with mne-python," *Frontiers in neuroscience*, vol. 7, p. 267, 2013.
[27] M. Jas *et al.*, "Autoreject: Automated artifact rejection for meg and eeg data," *NeuroImage*, vol. 159, pp. 417–429, 2017.
[28] T. Wilaiprasitporn *et al.*, "Affective eeg-based person identification using the deep learning approach," *IEEE Transactions on Cognitive and Developmental Systems*, vol. 12, no. 3, pp. 486–496, 2019.
[29] H. Cecotti and G. Jha, "3d convolutional neural networks for event-related potential detection," in *41st Annual Intl Conf of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2019, pp. 4160–4163.
[30] F. Jézéquel *et al.*, "Estimation of numerical reproducibility on cpu and gpu," in *Federated Conf on Computer Science and Information Systems*. IEEE, 2015, pp. 675–680.
[31] G. Hinton *et al.*, "Neural networks for machine learning lecture 6a overview of mini-batch gradient descent," *Cited on*, vol. 14, no. 8, 2012.
[32] M. Hosseini *et al.*, "Random ensemble learning for eeg classification," *Artificial intelligence in medicine*, vol. 84, pp. 146–158, 2018.
[33] X. Zheng *et al.*, "Ensemble deep learning for automated visual classification using eeg signals," *Pattern Recognition*, vol. 102, p. 107147, 2020.