UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga
(Final project paper)

Naslov zaključne naloge

(Convolutional neural network for eyeblink artifact detection in EEG data)

Ime in priimek: Marko Janković
Študijski program: (name of the study program, in Slovene)
Mentor: (with all titles, in Slovene)
Somentor: (with all titles, in Slovene)

Koper, mesec leto

Ključna dokumentacijska informacija

Ime in PRIIMEK: Marko Janković

Naslov zaključne naloge:

Kraj: Koper

Leto: 2025

Število listov: Število slik: Število tabel:

Število prilog: Število strani prilog: Število referenc:

Mentor:

Somentor:

Ključne besede:

Izvleček:

Izvleček predstavlja kratek, a jedrnat prikaz vsebine naloge. V največ 250 besedah nakažemo problem, metode, rezultate, ključne ugotovitve in njihov pomen.

Key words documentation

Name and SURNAME: Marko Janković

Title of final project paper: Convolutional neural network for eyeblink artifact detection in EEG data

Place: Koper

Year: 2025

Number of pages: Number of figures: Number of tables:

Number of appendices: Number of appendix pages: Number of references:

Mentor: title First Name Last Name, PhD

Co-Mentor:

Keywords:

Abstract:

Acknowledgement

Here we thank all involved with our final project paper, that is, persons or institutions that helped us in our work and/or made it possible. We can also thank the mentor and the co-mentor (if there is one).

Contents

List of Tables

List of Figures

Appendices

List of Abbreviations

i.e.	that is
e.g.	for example

1 Introduction

Electroencephalogram (EEG) is a noninvasive neuroimaging technique that involves the placement of electrodes on the scalp to record the electrical activity of the brain. This enables researchers to measure and analyze the electrical signals generated by the brain. These signals offer valuable information on the operating mechanisms of the brain, covering the identification of various neurological disorders and the exploration of cognitive processes such as perception, attention, and memory. However, EEG signals are prone to various artifacts, one of the most frequent being eye blinks, which for adults typically occur 15-20 times per minute. Eye closure changes brain activity, so eye-blink tracking of subjects undergoing EEG analysis is relevant for identifying when a subject blinks, falls asleep, or keeps their eyes closed. Eye blink and muscle movement artifacts are usually unwanted signals in the brain that can contaminate EEG data significantly and affect the accuracy of subsequent analysis of brain activity that is being measured with an electroencephalogram (EEG). Traditional methods for detecting and removing eye blink artifacts involve manual inspection, filtering, or statistical techniques like Independent Component Analysis (ICA). While these methods can work, they are very time-consuming and prone to human error, especially in real-time applications. They also fail to capture the complex spatial and temporal relationships in EEG data, hence they vary for different subjects and recording conditions.

1.1 Purpose of the thesis

In recent years CNNs have shown to be very effective in various domains as they can learn hierarchical features from raw data. This thesis proposes a CNN-based approach to detect signals in EEG data automatically, by demonstrating this ability on detection of eye blink artifacts. The aim is to design a model to detect signals in EEG recordings and with that prove that neural networks can be used to enhance EEG recordings and make EEG-based systems perform better.

2 Background and context

In this section we lay the ground work for the thesis by explaining some of the core components that will be used in proving our thesis.

2.1 Electroencephalography (EEG)

Electroencephalography (EEG) is the measurement of the electric potentials on the scalp surface generated (in part) by neural activity originating from the brain. The sensitivity of EEG to changes in brain activity on such a millisecond time scale is the major advantage of EEG over other brain imaging modalities such as functional magnetic resonance imaging (fMRI) or near-infrared spectroscopy (NIRS) that operate on time scales in the seconds to minutes range. Over the past 100 years, neuroscientists and clinical neurologists have made use of EEG to obtain insight into cognitive or clinical disease state by applying a variety of signal processing and statistical analyses to EEG time series. More recently there has been growing interest in making use of statistical modeling of EEG signals to directly control physical devices in Brain-Computer Interfaces. EEG power is typically split up into bands which correspond to different spectral peaks that relate to behavior or cognitive state. These bands are typically defined as the delta (1-4 Hz), theta (4-8 Hz), alpha (8-13 Hz), beta (13-20 Hz), and gamma (>20 Hz). [1]

2.2 Neural Networks

Convolutional networks have been applied to sequences for decades. They were used prominently for speech recognition in the 80s and 90s. ConvNets were subsequently applied to NLP tasks such as part-of-speech tagging and semantic role labelling. More recently, convolutional networks were applied to sentence classification and document classification. Recurrent networks are dedicated sequence models that maintain a vector of hidden activations that are propagated through time. This family of architectures has gained tremendous popularity due to prominent applications to language modeling and machine translation. The intuitive appeal of recurrent modeling is that the hidden state can act as a representation of everything that has been seen so far in the sequence. Basic RNN architectures are notoriously difficult to train and more elaborate architectures are commonly used instead, such as the LSTM and the GRU. Many other architectural innovations and training techniques for recurrent networks have been introduced and continue to be actively explored. Other recent works have aimed to combine aspects of RNN and CNN architectures. This includes the Convolutional LSTM, which replaces the fully-connected layers in an LSTM with convolutional layers to allow for additional structure in the recurrent layers; the Quasi-RNN model that interleaves convolutional layers with simple recurrent layers; and the dilated RNN, which adds dilations to recurrent architectures.

First, we describe a generic architecture for convolutional sequence prediction. The aim is to distill the best practices in convolutional network design into a simple architecture that can serve as a convenient but powerful starting point. The presented architecture is referred to as a temporal convolutional network (TCN), emphasizing that this term serves not as a label for a truly new architecture, but as a simple descriptive term for a family of architectures. The distinguishing characteristics of TCNs are:

1. the convolutions in the architecture are causal, meaning that there is no information "leakage" from future to past;

2. the architecture can take a sequence of any length and map it to an output sequence of the same length, just as with an RNN.

Beyond this, it is emphasized how to build very long effective history sizes (i.e., the ability for the networks to look very far into the past to make a prediction) using a combination of very deep networks (augmented with residual layers) and dilated convolutions. The architecture is informed by recent convolutional architectures for sequential data, but is distinct from all of them and was designed from first principles to combine simplicity, autoregressive prediction, and very long memory. For example, the TCN is much simpler than WaveNet (no skip connections across layers, conditioning, context stacking, or gated activations). Compared to the other language modeling architectures, TCNs do not use gating mechanisms and have much longer memory. [2]

2.3 ICA label

ICA was firstly created for dealing with the cocktail party problem, upon which you attempt to isolate a pertinent conversation from the noise of other conversations in a cocktail party. Applying the ICA to the EEG data involves the decomposition of

EEG time series data into a set of components. More specifically, EEG data are transformed to a collection of simultaneously recorded outputs of spatial filters applied to the whole multi-channel data, instead of a collection of simultaneously recorded single-channel data records. Thus, ICA is also a source separation technique that attempts to identify independent sources of variance in the EEG data. In the original EEG data collected at single channels, each row of the recording data matrix represents the time course of summed in voltage differences between the respective channel and the references channels. After ICA decomposition, each row of the transformed data matrix represents the time course of the activity of one independent component that is spatially filtered from the channel data. The outputs of ICA procedure are statistically independent component (IC) waveforms, as well as matrix that transforms EEG data to IC data, and its inverse matrix to transform IC data back to EEG data. These outputs provide information about an IC's temporal and spatial properties. Concerning that ICA assumes an instantaneous relationship (e.g., common volume conduction) and that any relationship between EEG and EMG signals should involve propagation delays, it is recommended to only select EEG channels for ICA decomposition. [3]

3 Methodology

In this section, the proposed methodology is described stepwise. First, a description of the data is given, and then the data pre-processing is explained, followed by the architecture of the model.

3.1 Dataset

We used dataset from ... (have to find the exact article). https://www.cell.com/heliyon/fulltext/S2405-8440(24)17502-0

3.1.1 Dataset description

This dataset examines the effects of caffeine on brain stimulation and consists of EEG recordings of two categories of participants: the first group which had an intake of regular coffee, and the other which had decaffeinated coffee. The primary difference between the two types of participants was the drinks and hence caffeine content. Participants from both groups were given the chance to add different amounts of sugar to their drinks, which causes an issue with the data but this is not the concern of this analysis.

In relation to our project, this dataset will be used for dealing with artifact identification and detection in EEG data, and while the original research was focused on caffeine stimulation of the brain, we are interested only in artifacts present in EEG signals, such as blinking, movements of muscles, and other activities nearby that interfere with the EEG readings.

3.1.2 Data preprocessing

While EEG recordings tend to contain noise and artifacts such as eye blinking or movement, EEG signals measured from the scalp do not necessarily accurately represent signals originating from the brain. Therefore, it is very essential to apply preprocessing and denoising to the recorded EEG data. Generally, preprocessing steps include the transformations or reorganizations of the recorded EEG data by removing bad or artifact-ridden data without changing clean data (transformation) and segmenting continuous raw signals without change of the data (reorganizations).

Our preprocessing of the data comes in 6 steps:

1. Filter 1 to 50Hz:

   We apply a bandpass filter to the EEG data, keeping frequencies between 1 Hz (lower limit) and 50 Hz (upper limit).

   1 Hz: This high-pass filter removes very slow components (below 1 Hz) that may correspond to artifacts like slow drifts in signals.

   50 Hz: This low-pass filter removes fast components above 50 Hz, such as muscle artifacts and electrical noise.

2. Cutaway first 1000 samples

   We remove the first 1000 samples, which might correspond to initial noise or artifacts at the start of the recording.

3. Re-reference to average

   This step re-references the EEG signals by subtracting the average of all electrodes from each electrode. This is commonly done to minimize noise and make the signals more comparable across channels.

   The assumption of average reference is: the sum of the electric field values recorded at all scalp electrodes (sufficiently dense and evenly distributed) is always 0, and the current passing through the base of the skull to the neck and body is negligible. Since our EEG recording system has enough even channels using average reference makes sense as the overall activity averages to 0.

4. Resample from 600 to 300 Hz

   We resample the data from a sampling rate of 600 Hz to 300 Hz. Resampling reduces the data size while retaining enough frequency resolution for EEG analysis.

5. Check dataset integrity

   This function checks the integrity of the EEG dataset after the preprocessing steps to ensure there are no inconsistencies or issues.

3.2 ICA label based artifact removal

The EEG data recorded from scalp electrodes can be considered summations of real EEG signals and artifacts, which are independent of each other. ICA is therefore potentially a useful methodology to separate artifacts from EEG signals. To removal artifacts embedded in EEG recordings, the computed ICs are firstly classified as either artefactual or neural related components. If detected and flagged as artifact-related ICs, they can be subtracted from the recorded data, and the remaining data can be remixed. In the artifact correction, ICA is used to separate components in order to identify artifacts relevant with eye movements or heartbeats. These relevant ICs have characteristic shapes (topographies, time courses, and frequency spectra) and can often be identified automatically. That is, artifact-relevant components generally can be identified according to the topographies, across-trial temporal distributions, and frequency distributions of the components.

Abnormal topographies can be appeared as (1) power concentrated only in the frontal lobe in topography (ocular artifacts); (2) discontinued topography (noise artifacts); and (3) topography constrained within single electrode (electrode artifacts). Abnormal across-trial temporal distributions can be appeared as (1) inconsistent between epochs (without obvious peaks in average waveforms); (2) periodic waveform (power line interference); and (3) noisy pattern (similar to Gaussian noise). In addition, the frequency of artifact relevant components is in higher-frequency band (e.g., >30 Hz), while the frequency content of neural signals is in lower-frequency band (e.g., 5–20 Hz). Particularly, for components relevant with blink artifacts (as shown in Fig. 1), they have an anterior distribution, and their time courses are largely flat with occasional very high-amplitude spikes indicating artifacts of the eye muscles as they close and open.

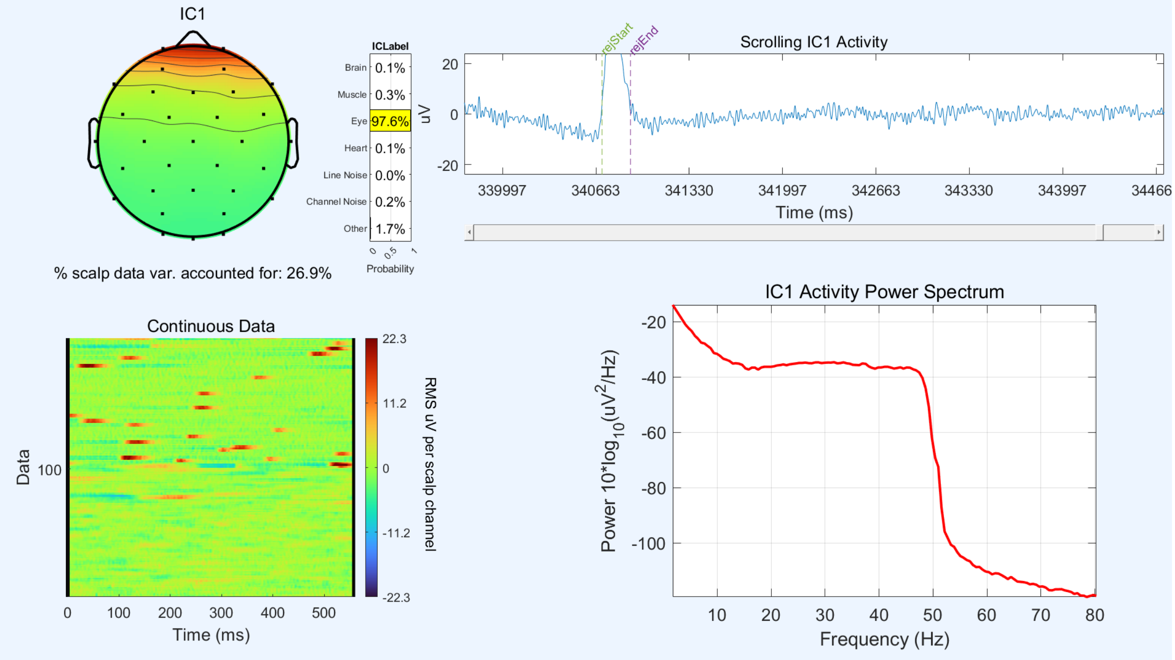


Figure 1: Examples of components relevant with ocular artifacts.

Particularly, the application of ICA seems to be particularly useful in removing blinks and other oculomotor artifacts. Using ICA to correct artifacts is generally considered the best, since it does not assume orthogonal or gaussian behavior of the individual signals. In contrast, PCA (principal component analysis) assumes that all signals are orthogonal and creates a succession of orthogonal base vectors, where each vector will account for as much variance as possible. As a result, the first vector from PCA is significantly larger in magnitude than all the subsequent vectors. When the signal to noise ratio is low, important information in these subsequent vectors can get lost. [3]

3.3 Model selection and description

For the model, we chose Temporal Convolutional Network (TCN) architecture. While sequence-to-sequence tasks are commonly solved with recurrent neural network architectures, as it was previously shown that convolutional neural networks can match the performance of recurrent networks on typical sequence modeling tasks or even outperform them.

We begin by describing our architecture for convolutional sequence prediction. The main building block of a TCN is a dilated causal convolution layer and is based upon two principles: the fact that the network produces an output of the same length as the input, and the fact that there can be no leakage from the future into the past. To accomplish the first point, the TCN uses a 1D fully-convolutional network (FCN) architecture, where each hidden layer is the same length as the input layer, and zero padding of length (kernel size - 1) is added to keep subsequent layers the same length as previous ones. To achieve the second point, the TCN uses causal convolutions, convolutions where an output at time t is convolved only with elements from time t and earlier in the previous layer. In this context, "causal" means that the activations computed for a particular time step cannot depend on activations from future time steps.

To put it simply: TCN = 1D FCN + causal convolution

A simple causal convolution is only able to look back at a history with size linear in the depth of the network. This makes it challenging to apply the aforementioned causal convolution on sequence tasks, especially those requiring longer history like in our case.

This gives us two ways to increase the receptive field of the TCN: choosing larger filter sizes k and increasing the dilation factor d, where the effective history of one such layer is $(k-1)d$. As is common when using dilated convolutions, we increase d exponentially with the depth of the network (i.e., $d = O(2^i)$ at level i of the network).

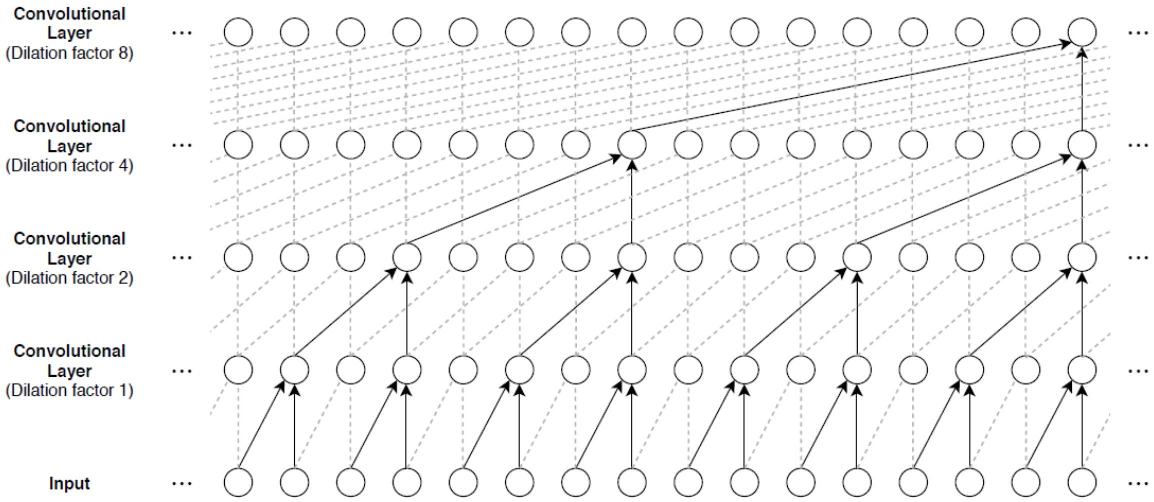

Convolutional Layer (Dilation factor 8)
Convolutional Layer (Dilation factor 4)
Convolutional Layer (Dilation factor 2)
Convolutional Layer (Dilation factor 1)
Input

Figure 2: A dilated causal convolution with dilation factors d = 1, 2, 4, 8 and filter size k = 2.

This ensures that there is some filter that hits each input within the effective history, while also allowing for an extremely large effective history using deep networks.

The general TCN architecture consists of multiple residual blocks, and a residual block contains a branch leading out to a series of transformations $\mathcal{F}$, whose outputs are added to the input x of the block:

$$o = Activation(x + \mathcal{F}(x)) \tag{3.1}$$

This effectively allows layers to learn modifications to the identity mapping rather than the entire transformation, which has repeatedly been shown to benefit very deep networks.

The residual block for our TCN is shown in Figure **??**, where each of the residual blocks, contains two sets of dilated causal convolution layers with the same dilation factor, followed by normalization, ReLU activation, and spatial dropout layers. The network adds the input of each block to the output of the block (including a 1-by-1 convolution on the input when the number of channels between the input and output do not match) and applies a final activation function. As our number of channels in input and output are the same and always match, the network does not need to apply 1-by-1 convolution to make them match.

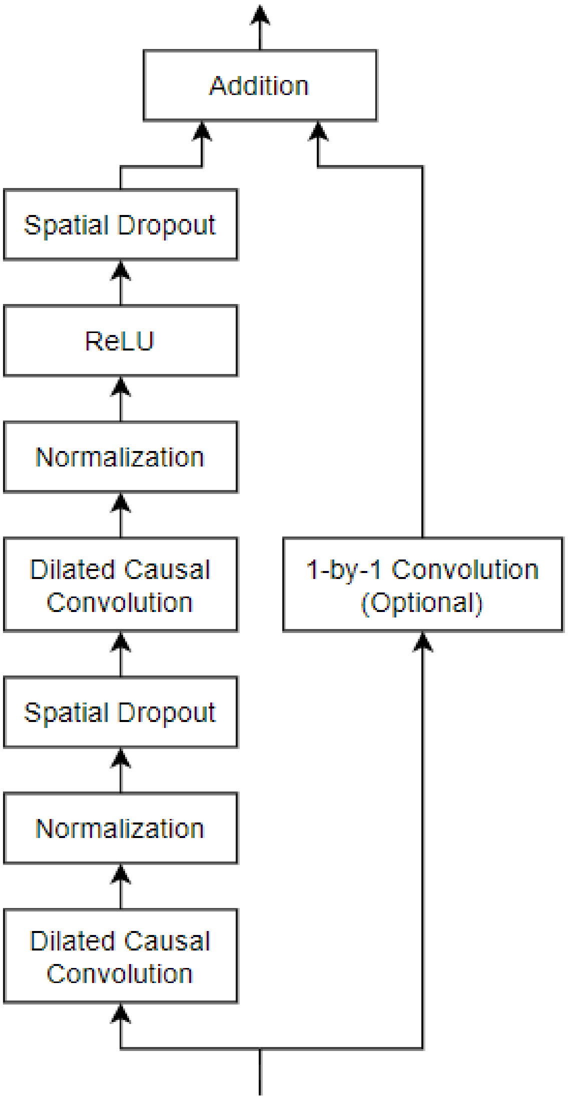

Addition
Spatial Dropout
ReLU
Normalization
Dilated Causal Convolution
1-by-1 Convolution (Optional)
Spatial Dropout
Normalization
Dilated Causal Convolution

Figure 3: TCN residual block. An 1-by-1 convolution is added when residual input and output have different dimensions.

4 Results

In this chapter, we report the performance of our model using key metrics such as accuracy, precision, and recall. These results show how well the model identifies and classifies artifacts in our test dataset. We will examine the results of our thorough testing process. The outcomes of our model's tests, both qualitative and quantitative, are presented here in our analysis. In addition to highlighting the model's potential, this assessment points out the drawbacks or areas in need of more development.

4.1 Results of training

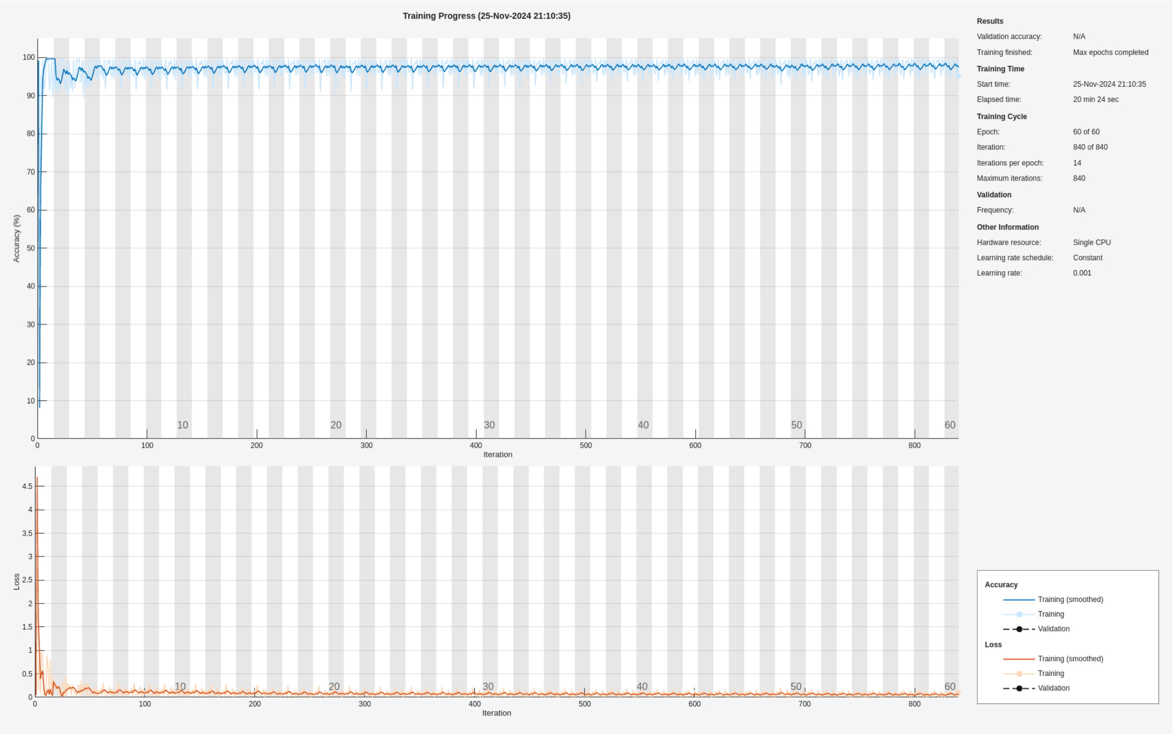
The training process was done on two datasets, with an already established model architecture without any modifications to its structure or hyperparameters. The focus of this training was replicating the performance suggested in previous studies, and using it to prove our hypothesis. As a result, no validation dataset was included in the training phase, and the choice was made based on lack of changes to the model that would require additional validation.

In Fig. 2 we display the accuracy and loss training measures. Within the first few epochs, the model achieves nearly perfect training accuracy, indicating rapid convergence of the accuracy curve. Also we can see this in the training loss curve, where we observe steep decline during the initial iterations and in the further training it stabilizes at minimal values. This demonstrates the model's ability to successfully fit the training data and is consistent with our expectations of it.

4.2 Results of testing the model

Here we present the results of testing our model.

4.3 Alternative model comparison



Results

Validation accuracy: N/A

Training finished: Max epochs completed

Training Time

Start time: 25-Nov-2024 21:10:35

Elapsed time: 20 min 24 sec

Training Cycle

Epoch: 60 of 60

Iteration: 840 of 840

Iterations per epoch: 14

Maximum iterations: 840

Validation

Frequency: N/A

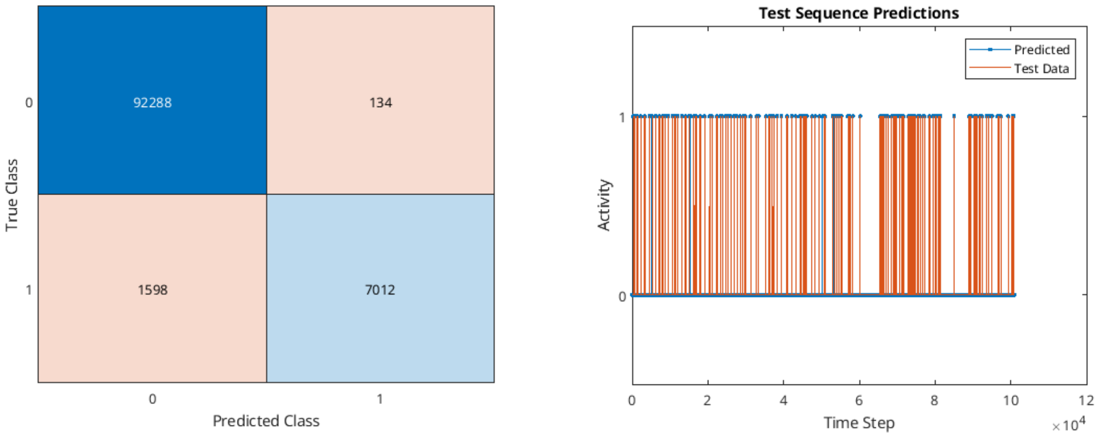
Other Information

Hardware resource: Single CPU

Learning rate schedule: Constant

Learning rate: 0.001

Figure 4: Results of training



(a) Confusion matrix

(b) Time-series line chart

Figure 5: Results of testing

5 Discussion and conclusion

In few sentences we briefly summarize the content of the project paper. This is also the place where we can add some further references for the interested reader.

6 Bibliography and Sources

[1] Michael Nunez, Paul Nunez, and Ramesh Srinivasan, *Electroencephalography (EEG): Neurophysics, Experimental Methods, and Signal Processing*, Jan. 1, 2016, pp. 175–197, ISBN: 978-1-4822-2097-1, DOI: `https://doi.org/10.13140/RG.2.2.12706.63687`. *(Cited on page 2.)*

[2] Bai, Shaojie, J. Zico Kolter, and Vladlen Koltun. *"An Empirical Evaluation of Generic Convolutional and Recurrent Networks for Sequence Modeling."* Preprint, submitted April 19, 2018. Available at: `https://arxiv.org/abs/1803.01271`. *(Cited on page 3.)*

[3] Makkar, Krish, Bisen, Anvikshaa, and Ijmtst, Editor, *EEG Signal Processing and Feature Extraction*, International Journal for Modern Trends in Science and Technology, vol. 9, pp. 45–50, 2023, DOI: `https://doi.org/10.1007/978-981-13-9113-2`. *(Cited on pages 4 and 8.)*

7 Povzetek naloge v slovenskem jeziku

This chapter contains a longer summary of the final project paper in Slovene, in total length between 4.000 and 10.000 characters (spaces included).

Appendices

A Title of First Appendix

Here we add the first appendix.

B Title of Second Appendix

Here we add the second appendix.