UNIVERZA NA PRIMORSKEM
FAKULTETA ZA MATEMATIKO, NARAVOSLOVJE IN INFORMACIJSKE TEHNOLOGIJE

Zaključna naloga
(Final project paper)

Naslov zaključne naloge

(Title of final project paper in English)

Ime in priimek: Marko Jankovic
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 Jankovic

Naslov zaključne naloge:

Kraj:

Leto: Koper

Število listov: Število slik: Število tabel:

Število prilog: Število strani prilog: Število referenc:

Mentor:

Somentor:

Ključne besede:

Math. Subj. Class. (2010):

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:

Title of final project paper:

Place: Koper

Year:

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:

Math. Subj. Class. (2010):

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

1.1 Importance of detecting EEG signals

1.2 Purpose of the thesis

2 EEG and ICA

We add some connecting text.

2.1 EEG signals

We add some connecting text.

2.2 ICA label

ICA was first created to deal with the cocktail party problem, upon which you attempt to isolate a pertinent conversation from the noise of other conversations at a cocktail party (Hyvarinen and Oja 2000). 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 (Anemuller et al. 2003).

2.3 Alternative Methods

3 Methodology

3.1 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.

3.2 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.

6. Run ICA

   In this step, Independent Component Analysis (ICA) is applied to separate independent sources (e.g., eye blinks, muscle noise, and brain activity) in the EEG signals. This helps in identifying and later removing artifacts, which is a common step after the basic preprocessing (filtering, re-referencing, etc.) has been completed.

3.3 Model architecture

3.4 Training and validation

4 Experiments and Results

5 Discussion

6 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.

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).

8 Bibliography

[1] A. Blum, G. Konjevod, and R. Ravi, Semidefinite relaxations for minimum bandwidth and other vertex-ordering problems. *Theor. Comp. Sci.* 235 (2000) 25–42. *(Not cited.)*

Appendices

A Title of First Appendix

Here we add the first appendix.

B Title of Second Appendix

Here we add the second appendix.