

1 **Tell me what I told you Mini-Me: Constructing and providing two-layer feedforward**
2 **networks for classification of continuous data in (and) a virtual T-maze**

3

4

Johannes Rodrigues

5

Julius-Maximilians-Universität Würzburg

6

7

Co-authors:

8

Philipp Ziebell

9

Mathias Müller

10

Johannes Hewig

11

Julius-Maximilians-Universität Würzburg

12

13

Author Note

14

Johannes Rodrigues, Department of Psychology I: Differential Psychology, Personality

15

Psychology and Psychological Diagnostics, University of Würzburg

16

ORCID ID: <https://orcid.org/0000-0001-8471-0816>

17

Philipp Ziebell, Department of Psychology I: Differential Psychology, Personality Psychology

18

and Psychological Diagnostics, University of Würzburg

19

ORCID ID: <https://orcid.org/0000-0003-1591-3075>

1 Mathias Müller 1) VTplus GmbH, An den Breiten 4, 97078 Würzburg 2) Department of
2 Psychology I: Biological Psychology, Clinical Psychology and Psychotherapy, University of
3 Würzburg

4 ORCID ID: <https://orcid.org/0000-0002-8525-993X>

5 Johannes Hewig, Department of Psychology I: Differential Psychology, Personality Psychology
6 and Psychological Diagnostics, University of Würzburg

7 ORCID ID: <https://orcid.org/0000-0002-8400-189X>

8

9 This research was founded by the faculty of human sciences of the University of Würzburg

10

11 Conflict of interests:

12 Mathias Müller is a shareholder and executive officer of a commercial company (VTplus GmbH)
13 that develops medical devices for virtual reality therapy applications and virtual environment
14 research systems for empirical studies in the field of psychology, psychiatry, and psychotherapy.

15 No further potential conflicting interests exist.

16

17 This research was conducted from 2017 – 2019

18 The data and paradigm are available on the following repositories:

19 FTP:

20 <http://update.cybersession.info/T-Maze/VrProject.T-Maze.zip> (282 MB, 2019-10-15)

21 http://update.cybersession.info/T-Maze/TwoLayerFeedForward_Data.zip (330 MB)

1
2
3
4
5
6
7
8
9
10
11
12

SVN Repos:

<https://crm.vtplus.eu/svn/open-access.t-maze/trunk>

<https://crm.vtplus.eu/svn/open-access.t-maze/trunk/Paradigm>

https://crm.vtplus.eu/svn/open-access.t-maze/trunk/sourcemods/VrSessionMod06_T-maze

Correspondence concerning this article should be addressed to

Johannes Rodrigues, Department of Psychology I,

Julius-Maximilians-Universität Würzburg,

Marcusstraße 9-11, 97070 Würzburg, Germany

Tel: +49 931 31 81771.

Email: johannes.rodrigues@uni-wuerzburg.de

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24

Introduction

Since Ioannidis (Ioannidis, 2005) as well as the open science collaboration under Nosek and colleagues (2015) pointed out that psychology is facing a massive problem concerning replication of research studies, the focus on reliable and replicable analysis as well as replicable methodological approaches grows stronger. The open access documentation of methods and analysis are becoming more and more common these days in Psychology, which is of key importance in order to get to reliable and replicable results. However, as the focus lies primarily on the documentation and standardization of the data analysis, one may forget about the problems arising in complex data patterns classification. Examples for those kinds of data would be continuous movement recordings (e.g. mouse trajectory, joystick movement, virtual movement in virtual reality, driving simulation, head movement, animal movement trajectories) but also continuous physiological recordings (e.g. human and animal EEG, EMG, ECG, EDA, MRI) and complex combinations of distinct variables (e.g. digital picture recognition). In rather simple cases, the needed replicable classification can be achieved using simple logical and algebraic expressions (e.g. in a force choice task where one knows the alternatives, in movement recordings where one distinct feature such as “going up” is identifying a behavioral class “going to floor 2” for all participants, because they all start on floor 1 and there is no floor 3). Since this simple structure is not always given, complex behavioral patterns are often classified by human experts. These experts may introduce subjective and non-replicable biases into the data, as they are not available to everyone and possibly also sometimes having a different detection threshold concerning underlying patterns in the data due to decreased performance because of stress or other obstacles and states (e.g. Kipnis & Schmidt, 1988, Pilcher & Huffcutt, 1996). In order to overcome this problem, some researchers started to use neural network pattern classifier in order to get to a reliable and replicable result (e.g. classification of sleep stages: Barschdorff, Gerhardt

1 & Trowitzsch, 1996, Baumgart-Schmitt, Eilers & Herrmann, 1997, selection of artifact ICs in
2 EEG research: Winkler, Haufe & Tangermann, 2011).

3 However, neural networks are not generally “better” than humans in decision making.
4 The idea of neural networks especially pattern classifier being an entire new and better level of
5 decision making and classification is sometimes propagated, but very often this is not the case
6 (see e.g. He, Zhang, Ren & Sun, 2015, for one of the neural nets that outperforms humans). The
7 major reason for humans outperforming neural networks, at least in case of the classical pattern
8 classifiers, is that the classifier has to learn from data and must be trained by data (Rey &
9 Wender, 2011). Hence one already has to know what is right or wrong, ergo one has to
10 previously classify the items of interest and train this classification to the pattern classifier. This
11 may lead to the result that the neural network is not performing better than the human decision
12 maker or expert normally. However, it is more consistent, as the underlying decision rules of the
13 classification will hopefully be discovered by the pattern classifier and used accordingly for the
14 new data. Hence, they are more consistent and therefore they may introduce less variability in the
15 data. Also, they can be made available for everyone, which is hard to be achieved for or by
16 human experts. Still, many decisions are made by the experts” and the replacement by automatic
17 detection algorithms is not yet implemented in cases of complex behavioral decisions, because it
18 is rather hard to achieve sufficient performance with simple automated algorithms. Nevertheless
19 as mentioned above, some fields successfully implemented neuronal nets in order automatize
20 decisions that were previously made by hand (Barschdorff, et al., 1996; Baumgart-Schmitt et al.,
21 1997; Winkler, et al., 2011) .If the decisions are still made by hand, even by experts, this leads to
22 the problem that only the researcher themselves or the expert might reproduce the decisions.
23 Concerning the reliability of the decisions, this leads to disastrous effects, especially if the
24 decision is not be that easy for a given task set (e.g. McFarland & Ashton, 1978, Anderson,

1 Wright & Immink, 1998) and one has to decide between categories that may not be separated
2 easily. For example, this has been shown in a categorization task involving fish outlines that
3 varied across four different shape dimensions, namely “dorsal fin”, “tail”, “ventral fins”, and
4 “mouth”, which might theoretically seem easy to separate without difficulties, but empirically
5 were not (Sigala, Gabbiani & Logothetis, 2002). Here, a neural network might provide a good
6 solution for reproducible and reliable effects, although it is not necessarily more accurate than a
7 human decision maker (see He et al., 2015), but still it might be more consistent. Another great
8 feature of neural networks is the generalization of the decisions even to problems that were not
9 present before (Rey & Wender, 2011). This feature can lead to a good classification of data that
10 are not learned but similar in their structure to the training data. Accordingly, neural networks
11 provide a mechanism that will generalize the classification to new stimuli, although one has to
12 admit, that we humans tend also to be able do that. However, this neat feature of learning a
13 decision pattern that can generalize comes at a price. If the data is not similar, the neural network
14 will still classify it according to its internal rules and propagate a data classification, although it
15 might not be correct. So, if the data is not similar to the data that was used to train the neural net,
16 there can be problems with the classification and we will need a new net in order to classify the
17 data correctly. On the other hand, as there are many similar problems or similar outcome
18 classifications, one might use one net for different classification tasks as long as the output and
19 input is similar. Nevertheless, for different tasks, different neural networks are needed. So, if we
20 need many different neuronal networks to classify our data, how do we construct the neural
21 networks in order to solve the classification problems? Here we can get very different answers
22 (see e.g. Minin, 2012), but we would like to concentrate on work by Huang (2003), that showed
23 that every classification problem can be solved with two-layer feedforward networks. Also,
24 Huang provided a formula how to construct these networks and this formula was used to create

1 the network presented here. In the following, we will show which criteria were used to create the
2 neural network for classification and which training and performance parameters were extracted.
3 Additionally, we will briefly introduce the solution and the overall fit of the solution that was
4 tested with other data. Here see that the fit of the neural network is not the only indicator that is
5 worth having a look at, but also the performance on a new data set, because some algorithms tend
6 to find an overly specific solution to the problem (overfitting) and tend not to generalize to the
7 problem as well as other networks (e.g. Lawrence & Gilles, 2000).

8 Before we start to explain the problems that we used as examples for the classification of
9 the two-layer feedforward network in the method section, we take a brief look on what this type
10 of network is. A neural network basically consists of different units or neurons, which are put in
11 different layers and will eventually be linked according to different learning rules and
12 propagation constraints in order to propagate the input information to the output layer (for details
13 see e.g. Rey & Wender, 2011). The weights, the strengths of the links between neurons in one
14 layer to the next layer, store the information and the learned feature association in the net. Every
15 net has an input layer and an output layer, but between them can be a variable amount of neuron
16 layers. These layers between the input and output layers are called hidden layers. Depending on
17 the number of layers and the amount of neurons in the different layers compared to the following
18 layer, the net reacts systematically different to the input. In general, if few layers are used, a very
19 high amount of neurons are needed per layer as the different features and information of the input
20 need to be classified by these layers and their interconnection using the provided learning rule
21 and the propagation constraints. In our case, we are using a two-layer feedforward network. This
22 means, that we are using two hidden layers between the input and the output layer, additionally to
23 the constraint that the data is only propagated from the input layer to the output layer through the
24 two hidden layers without any loop as one would have with for example recurrent or

1 backpropagation of the data (e.g. Rey & Wender, 2011; Zhang & Gupta, 2000). A graphical
2 display of the two-layer feedforward networks that are used as examples are given in Figure 4
3 and Figure 5 in the method section below.

4 Subsequently, we want to shed light on our motivation to provide examples for the use of
5 these networks. The reason why we want to provide the community with those examples is that
6 these solutions solve the problem of replicable results as well as accurate classification of
7 complex problems, because the neural networks that are constructed mirror the decisions of the
8 experts. Hence, one will be able to create different other neural networks for other tasks,
9 mirroring expert decisions concerning this task. Therefore, a new level of standardization in
10 complex tasks as well as a new level of scientific discussion may arise based on different neural
11 networks that can be applied to datasets of different working groups. We want to provide other
12 researchers with the tools for the classification of complex tasks that normally need expert
13 decisions and with the necessary knowledge on how to build a new automatized classification
14 network for their own specific research questions. Accordingly, the present study is an
15 application of the procedure suggested by Huang (2003) and just one example of its applicability
16 to behavioral classification and physiological data classification (for details see method section
17 below). Of course, as mentioned above, one may use such networks to categorize any complex
18 data where an expert is normally needed and given, as well as a reasonable input and output for
19 the neural network can be defined.

20 In order to further share our knowledge with the scientific community, we additionally
21 provide the environment of our current examples of complex research problems that we address
22 here as an example. Thus, we offer other working groups the opportunity to use and further
23 develop the behavioral classification solution or the physiological data classification along with
24 the paradigm. The environment is a virtual T-maze developed by Rodrigues (2016). It was used

1 for recording psychophysiological signals such as the electroencephalogram, electrodermal
2 activity and heart rate while showing different virtual free-choice behavior patterns via joystick in
3 different situations. These behavioral decisions and performances were previously categorized by
4 hand according to standardized criteria (e.g. Rodrigues, 2016, Rodrigues et al., 2018) into six
5 categories. However, as small changes in the paradigm led also to changes in the resulting
6 behavior, six new categories were proposed and a previously trained neural network to classify
7 the behavior was not of good use anymore (see e.g. Rodrigues, Müller & Hewig in press). A
8 detailed description of the different categories is mentioned below in the methods section and
9 Figure 2. Importantly, based on these classifications or any other behavioral or physiological
10 categorizations, one is able to provide a dataset that can be used to train a neural network in order
11 to automatize and standardize the classification of behavioral or physiological responses (here as
12 an example in the virtual T-maze). In order to provide other researchers with the necessary
13 tracking devices for the joystick movement in the present example of behavioral responses in the
14 virtual T-maze , a tutorial will be attached in the supplementary materials, describing how to set
15 up the paradigm correctly and how to use the free trial version of the tracking software that was
16 used to record the behavioral responses in the virtual T-maze. With this work we hope to support
17 and encourage other researchers to share their paradigms as well as to develop standardized ways
18 for analyzing even difficult phenomena or data. Furthermore, we try to provide the scientific
19 community with our paradigm and the necessary tool to analyze data from it. It may be used as an
20 example to share and it can be modified to the needs of the researchers and the solution for
21 automatized classification (for example in this paradigm) may lead to replication, along with
22 further development of research paradigms.

1 **Methods**

2 **Ethical statement**

3 The data collection was carried out in accordance with the recommendations of the “Ethical
4 guidelines, The Association of German Professional Psychologists” (“Berufsethische Richtlinien,
5 Berufsverband Deutscher Psychologinnen und Psychologen”) with written informed consent
6 from all subjects. All subjects gave written informed consent in accordance with the Declaration
7 of Helsinki before they participated in the experiment. The protocol was approved with the file
8 reference GZEK-2017-18 by the local ethics committee of the department of psychology of the
9 Julius-Maximilians-University of Würzburg (Ethikkommission des Institutes für Psychologie der
10 Humanwissenschaftlichen Fakultät der Julius-Maximilians-Universität Würzburg).

11 **Classification problem example 1: Behavior in virtual T-maze paradigm**

12 The first problem we wanted to address as an example for the application of a two-layer
13 feedforward network to solve classification problems categorized by experts in the current study
14 is the behavioral classification in a virtual T-maze (Rodrigues, 2016). In the virtual T-maze,
15 humans perform free virtual movement behavior patterns. These patterns used to be categorized
16 into six categories: “fleeing from a stimulus”, “approach safety from a stimulus via left turn”,
17 “approach safety from a stimulus via right turn”, “reaching out for a stimulus via left turn”,
18 “reaching out for a stimulus via right turn”, and “behavioral conflict”. This was performed semi-
19 automatically according to simple rules and was controlled and corrected by hand for every trial
20 (for classification details see Rodrigues, Müller, Mühlberger & Hewig, 2018, Rodrigues, 2016, or
21 Figure 2 and method section below). However, as different raters with different experience
22 started to try this classification task, it quickly became obvious that additional supervision and
23 training of these raters would be necessary, in order to achieve the same results and accuracy as

1 expert raters. In addition, as more complex behavioral categories arose in some modifications of
2 the paradigm (e.g. conflict, classification 6 in Figure 2 and method section below), such a simple
3 classification approach was no longer suitable. Nevertheless, having categorized the behavioral
4 responses of the modified conditions of the virtual T-maze (see Figure 1 and method section
5 below) by an expert rater for this task (first author), there was ideal training-, test- and validation-
6 data in order to start the transfer of this problem into an automated classification via two-layer
7 feedforward networks.

8 The problem that was solved by the two-layer feedforward network was the classification
9 of different behavioral response types in a virtual T maze. In this paradigm that is similar to the
10 paradigm used by Rodrigues et al. (2018), the participants experienced five different conditions.
11 It was a desktop virtual reality approach, where participants were able to navigate through a
12 virtual T-maze via joystick. The conditions were linked to credits and cued by color bars at the
13 walls of the T-maze. The participants were instructed to go into the maze and earn as many
14 credits as possible. They received immediate visual as well as auditory feedback when they
15 gained or lost credits (a pleasant harmonic chord and textual information on credit gain or
16 avoidance of credit loss versus an unpleasant disharmonic chord and textual information on credit
17 loss or miss of credit gain). Also, they experienced a training phase before the experiment, so
18 they were able to react to the different conditions appropriately. The duration of one trial was 13
19 seconds and 100 Trials (20 per condition) were performed by every participant. In each trial, the
20 participants started in a passage, looking in the direction of the T-arms of the maze and seeing a
21 color cue on the wall of the T-arms of the maze. The cue indicated the type of outcomes possible
22 in a trial. Participants were able to choose to move further forward towards the T-arms or move
23 away from the T-arms, either by moving backwards or by turning around to move away from the
24 T-arms. After the cue had appeared for four seconds, a specific entity (positive trial entity: sheep,

1 negative trial entity: monster, for examples see Figure 1 or repositories) appeared in the maze.
2 The entity was only visible if the participant moved in the appropriate direction in the case of a
3 positive entity running away from the participant, or kept the VR view to the direction where the
4 entity was coming from in case of a negative trial entity.

5 The different types of trials will be described in the following paragraphs and consisted of
6 negative events, positive events, approach-avoidance conflicts, approach-approach conflicts and
7 ambiguous trials.

8 Negative event trials had a monster, chasing the subject and threatening to cause a loss of
9 credits if one could not escape. In response to negative events, one had to escape by not going
10 forward into the T-arms of the maze, but by returning down the starting passage and reaching the
11 starting end of the passage before the monster did. Although one had to go back down to the end
12 of the passage, one could still do so in two ways: moving backwards or turning around and
13 moving forward towards the end of the passage. Negative events were cued with a red color cue
14 on the wall with the position of the cue next to the T-arm of the maze from where the monster
15 approached.

16 Positive event trials had a sheep as entity, providing the opportunity to raise the credits, if
17 one could reach the sheep by moving forward in one of the T- arms of the maze before the sheep
18 reached the end of this passage. Positive events were cued with a green color cue on the wall next
19 to the T-maze arm in which the sheep was running away.

20 The third event category was an approach-avoidance conflict, consisting of a negative
21 and a positive event at the same time, being also cued the same time, but the negative and
22 positive event never being in the same T-maze arm. Having to choose whether they wanted to go
23 forward for the sheep and get caught by the monster, or whether they wanted to flee the monster

1 and miss the sheep, the participants had the same expected outcome value in this trial for both
2 behavioral options.

3 The fourth event was an approach-approach conflict, consisting of two positive events at
4 the same time on different T-maze arms. As the time was not sufficient to get both sheep, the
5 participant had to decide which sheep to follow and which to let go.

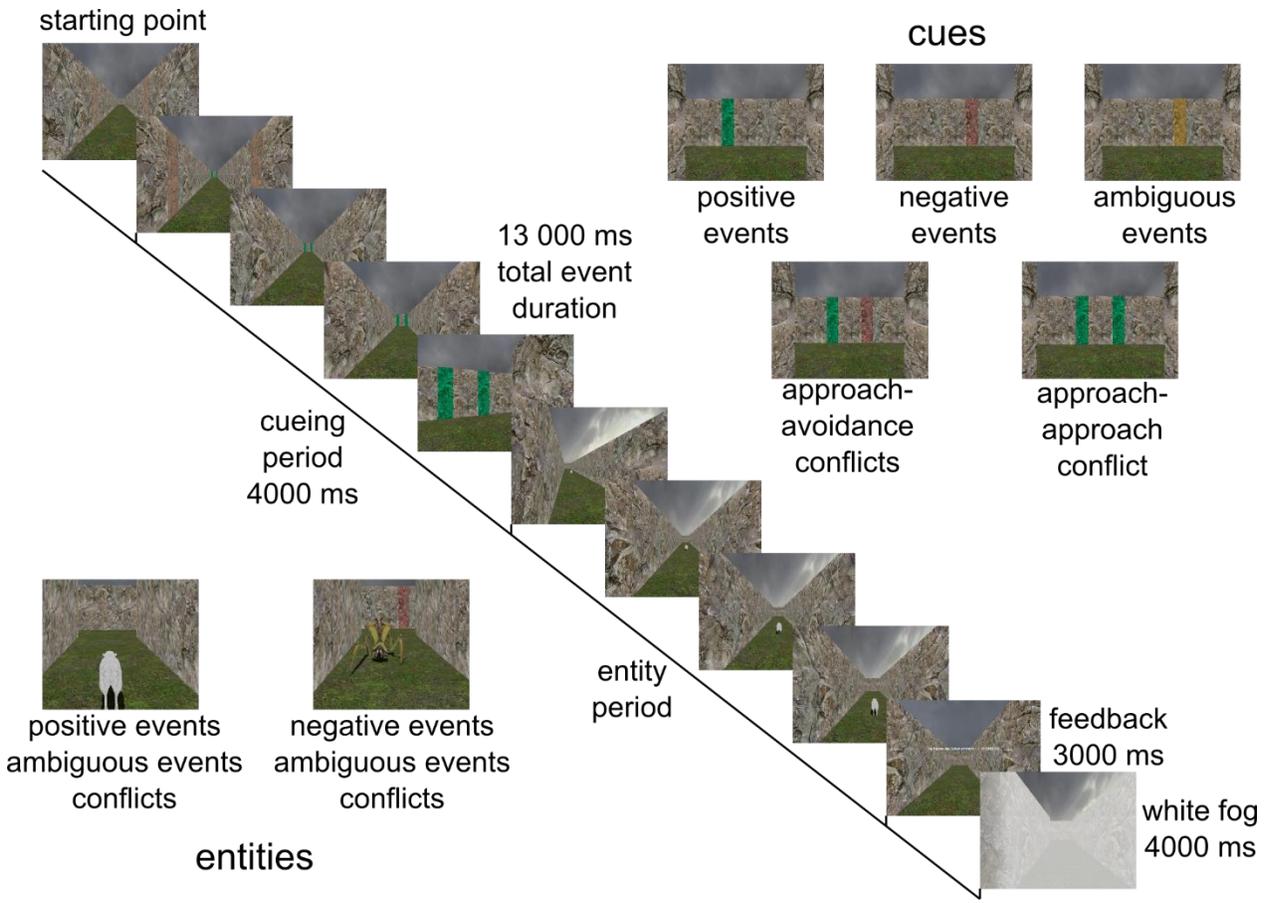
6 The fifth event was an ambiguous trial, where the participants saw a yellow cue at one
7 side of the T-maze. This cue indicated that either a sheep or a monster would occur, but the
8 participants were not informed which of those entities they would engage with eventually. Hence,
9 the participants had to decide whether they would go forward in order to possibly catch the sheep,
10 or whether they would run away to escape the monster. Of course as they did not know whether it
11 would be a sheep or monster, a decision to go after the sheep if the monster was actually present
12 resulted in being caught by the monster and losing credits, as missing the sheep while running
13 away from a monster that was not present led to a stagnation of the credits.

14 In order to strengthen the impact of negative and positive outcomes of the events,
15 harmonic and disharmonic chords were presented via headphones together with visual feedback
16 of the credit gain and loss. If one reached the sheep that was running away, in addition to the
17 raise of credits, a harmonic chord was played. If one did not reach the sheep in time, in addition
18 to the message that the credits remain constant, a disharmonic chord was played. If a monster was
19 present and the participant managed to escape the monster, in addition to the message that the
20 credits remain constant a different harmonic chord was played. Finally, if a monster was present
21 and the participant was caught by the monster, in addition to the message of the loss of credits a
22 different disharmonic chord was played.

1 At the end of each trial, the participants experienced a white fog, “beaming” them to the
2 starting position for the next trial once more. A schematic display of the trials and examples for
3 the cueing as well as the entities used in the trials can be seen in Figure 1.

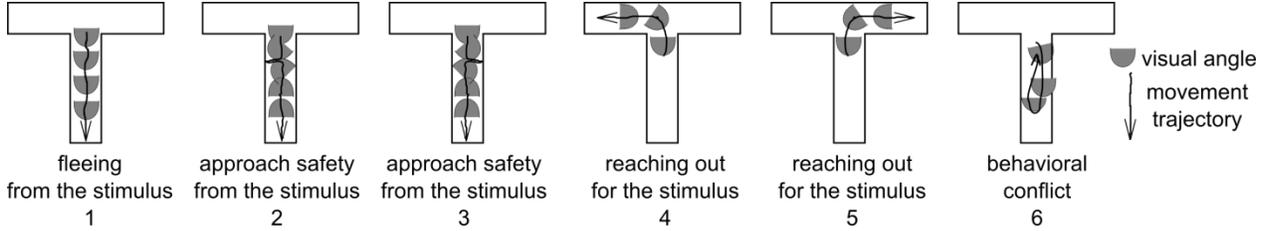
4 The different behavioral patterns that the participants showed in this kind of paradigm had
5 been categorized manually into different categories already by Rodrigues et al. (2018). However,
6 as small changes in the paradigm led also to changes in the resulting behavior, six new categories
7 were proposed and a previously trained neuronal network to classify the behavior was not of
8 good use anymore (see e.g. Rodrigues et al., 2019, in prep). All categories were based on the
9 movement and acceleration/deceleration trajectories of the joystick in the virtual T-maze and the
10 time scale, as well as the overall performance of the participant. The manual criteria for the
11 training set that have been selected based on successful reliable empirical classification
12 experience of the expert from the mentioned earlier studies are as follows: The first category was
13 fleeing, with the backward movement speed being higher than 75 % of the maximum backward
14 movement speed and the turning of the participant being indicated by the joystick excursion to
15 the left or right sides being smaller than 50 % of the possible excursion in the first 8 seconds
16 (Figure 2). Note that it is important for the first category, that the person is moving backwards
17 and the joystick is not moved further than 50% sideways. As the participants tend to not perform
18 only a straight line while moving backwards, a rather broad margin of error is provided by the
19 50% yet indicating that there is still no turning around in the relevant time of behavioral
20 classification, because the backward movement is still present. The second and third categories
21 were the approach safety behavior that was characterized by the turning around of the participant
22 indicated by the joystick excursion to the left or right side being higher than 15 % and having a
23 change from backward to forward movement in the first 8 seconds (Figure 2). Here the specific
24 behavior of changing from backward to forward movement is important and that this is

1 accompanied by a sideways motion of the joystick indicating a turning motion. As the virtual
2 space that the participants may cover during the relevant time window is rather large, a sideways
3 turning of 15 % already provides the opportunity to change movement directions from backwards
4 to forwards movement. The reason for two categories was the direction of the turning movement
5 with a distinction between right and left turning. Depending on the specific research question, it
6 might either make sense to combine those two categories eventually as an approach safety
7 behavior category independent of right or left turning or to analyze the two categories separately.
8 It might for example be of interest, whether individual participant characteristics such as
9 handedness influence the preferred turning direction when approach safety behavior is executed
10 or whether the turning movement direction when executing approach safety behavior is related to
11 the occurrence location of the negative stimulus. The fourth and fifth categories were the
12 reaching out movements. They were characterized by the forward movement, indicated by the
13 joystick excursion in forward direction being higher than 50 % of the maximum excursion in the
14 first 8 seconds and the backward movement, indicated by the joystick excursion in the backward
15 direction being smaller than 10 % in the first 4 seconds (Figure 2). The reason for 2 categories
16 was again the direction of turning. The last category was the most complicated to classify, the
17 behavioral conflict category. This category was picked if the behavioral reaction onset latency
18 was higher than 1.5 seconds or if there was a change in the behavior during the trial. For instance
19 if the participants started to flee but then decided to reach out for the stimulus after a short period
20 of time (Figure 2). Also, this category was chosen if the behavioral performance of the trial was
21 very bad in comparison to other trials executed by the participant, indicating a less accurate
22 execution of the joystick behavior (Figure 2). These behavioral categories were used to train the
23 two-layer pattern classifier in order to automatize the classification for other experiments that are
24 to come in this virtual T-maze.



1

2 **Figure 1: Schematic display of a trial in the VR paradigm. In the lower left corner examples**
 3 **of the entities used in the VR paradigm, in the upper right corner examples of the cues used**
 4 **in the trials are shown.**

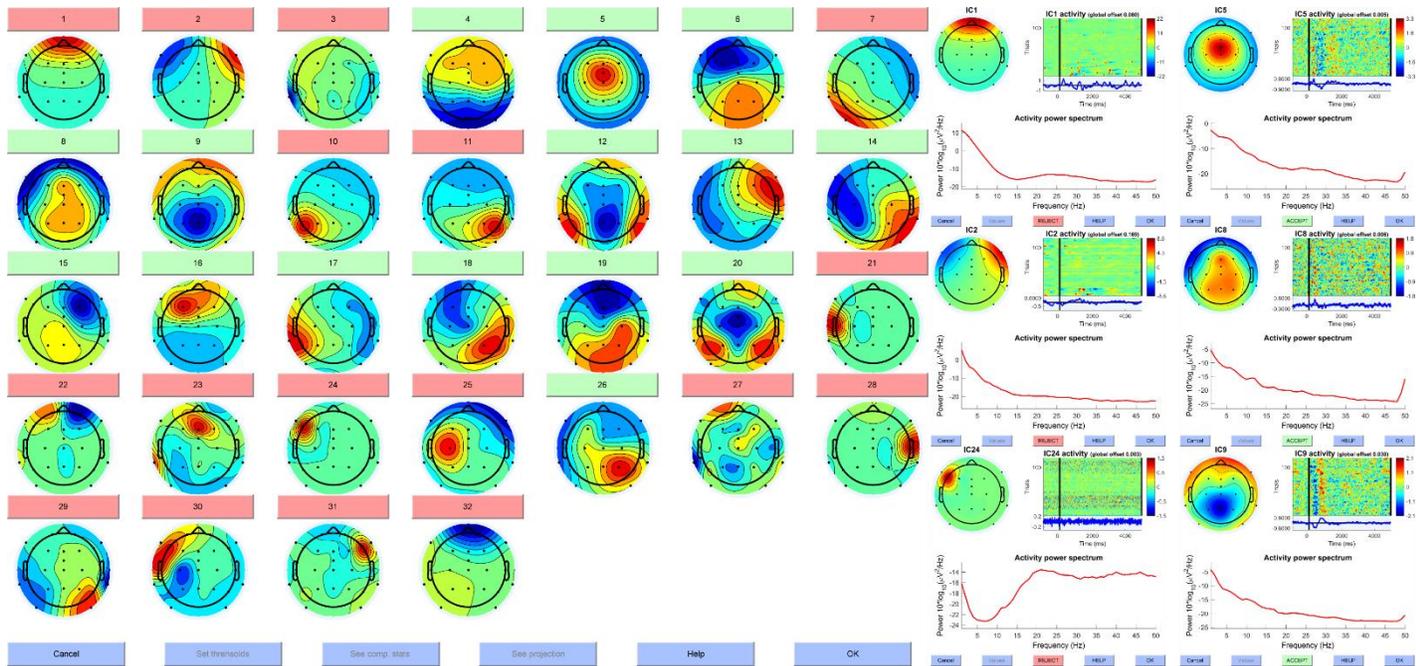


5

6 **Figure 2: Behavioral categories in the virtual T-maze. The numbers show the different**
 7 **classes that are later used for classification. Note that the behavioral conflict category is not**
 8 **only the start of different behaviors in one trial but also very bad performance compared to**
 9 **the other trials of the participant. Note also, that the category 2 and 3 are both approach**
 10 **safety behaviors, yet their difference is in the turning direction. This may be interesting**
 11 **dependent on the question on is to answer.**

1 **Classification problem example 2: Automatic IC artifact detection in EEG research**

2 The second problem solved with the two-layer feedforward network was the automatic
 3 classification of independent component artifact components in the experimental task of the T-
 4 maze that produced rather specific artifacts. In EEG research, artifact detection, as exemplarily
 5 the detection of muscular signals, pulse related confounds (e.g. from templar structures), eye
 6 blinks, eye-movements, electrode site disconnections and other non-brain-related electrical
 7 artifacts, is very important to get to a valid and reliable interpretation of the electro-cortical
 8 signal. One approach to detect these artifacts is the independent component decomposition (ICA:
 9 Makeig, Debner, Onton & Delorme, 2004) of the signal, in order to identify underlying signal
 10 patterns that are related to artifacts. As the task of detecting an artifact component is a rather
 11 complex procedure and not easily to be explained and performed, researchers needed to be
 12 trained for a long time in this process by experts, in order to identify artifact components by the
 13 complex combination of topography, time-course and frequency distribution of the component
 14 (see example Figure 3).



15 **Figure 3: Independent components classified in artifact components (red) and signal**
 16 **components (green).**

1 **components (green). Note in the right panel the frequency distribution, topography and**
2 **time courses of different artifact and signal components.**
3

4 Early endeavors to achieve a standardization in this work was undertaken by Flexer and
5 colleagues (Flexer, Bauer, Pripfl & Dorffner, 2005), who correlated the ocular activity with
6 independent components (ICs) and excluded the component that correlated the most with the
7 ocular activation. However as the ocular activation is able to show up in many components due to
8 the very significant structure of this muscular activation and dipole induction, this approach was
9 soon followed by other ideas of standardized artifact IC detection, based automated algorithms
10 including the temporal and spatial distribution of ICs (e.g. ADJUST: Mognon, Jovicich,
11 Bruzzone & Buiatti, 2011) but with the caveat of still having to supervise the detection by human
12 experts (see e.g. SASICA: Chaumon, Bishop & Busch, 2015). One approach was also to use
13 machine learning algorithms to identify an automatic solution for a specific preprocessing of the
14 data, including the frequency spectrum as an additional information, but excluding high
15 frequency noise above 50 Hz (MARA: Winkler, et al., 2011). As these higher frequencies
16 however are rather important to detect several types of artifacts such as pulse artifacts and
17 muscular activity, the provided solution is not ideal for every task, especially if more muscular
18 artifacts are to be expected from behavior that is not only executed by simple button presses, but
19 by other behavioral options including virtual or real movements.

20 A common theme in all these approaches to classify artifacts via ICs is the idea, that one
21 solution may be used for all tasks at hand. For experimental tasks that are rather similar
22 concerning the appearing artifacts, a good performance of these solutions could also be shown
23 (see e.g. Chaumon et al., 2015, Mognon et al., 2011, Winkler, et al., 2011). On the other hand, if
24 a task produces other types of artifacts due to the response opportunities for example or the
25 sensual input for the participant, different solutions that are more specific for this type of task will

1 be needed. As the first author of this manuscript has been extensively trained in detecting artifact
2 ICs and applied this knowledge to his research by manually detecting such components in the
3 first place, we used the EEG-activation in the virtual T-maze task explained above as an
4 additional example for the creation of a specific automatic detection mechanism based on expert
5 knowledge. Although different types of artifacts can be detected (see Figure 3), the prime goal
6 was to detect whether an artifact is present or not. This is the reason why the neural net only has
7 two resulting categories: Signal and Artifact.

8 **Structure of the neural net: two-layer feedforward network**

9 The structure of the nets was determined according to the problem at hand and due to the
10 formula given by Huang et al. (2003) and can be seen in Figure 4 and Figure 5. The formula
11 given for layer one is $\sqrt{(m + 2)N} + 2\sqrt{N/(m + 2)}$. The formula for layer two is
12 $m\sqrt{N/(m + 2)}$. Here, N is the number of distinct input samples and m is the number of output
13 neurons.

14 **Classification problem example 1: Behavior in virtual T-maze.**

15 The input vector structure of the neuronal net for the first example consisted of 6003
16 elements, with 3000 elements being the coordinates in the x direction, 3000 elements being the
17 coordinates in the y direction, the id number of the participant, the trial of the condition and
18 condition number. For the output vector structure, we used six binary elements, classifying all
19 behavioral options specified above. The neural networks were trained with 54 participants with
20 100 trials each (20 per condition), leading to 5400 different input vectors and a 6003 x 5400 input
21 matrix for the training and determination of the performance. The input vectors and the output
22 vector structure were used to compute the necessary two-layer feedforward network to solve the
23 classification problem. According to the formula given by Huang et al. (2003), the first layer

1 consisted of $\sqrt{((7 + 2) * 5400) + 2 * \frac{5400}{7+2}} = 259.81 \approx 260$ neurons. The second layer

2 consisted of $7 * \sqrt{\frac{5400}{7+2}} = 155.88 \approx 156$ neurons. The differentiation of the training set, the

3 validation set and the test set was always kept at 70 % training data, 15 % validation data and 15

4 % test data. The necessary validation trials for the different neural networks were set from 100 to

5 150 to 200 trials. This means that either 100, 150 or 200 trials had to be classified correctly in a

6 row to validate the neural network being finished with its internal adjustments of weights. As

7 more validation trials may lead to overfitting of the neural network, the restriction to a maximum

8 of 200 consecutive correct classified trials was used. Two different training algorithms were used

9 in order to compare their performance. The first training algorithm was the “trainscg” that

10 updates the weights and bias values according to the scaled conjugate gradient method (Møller,

11 1993), the second algorithm that was used was the “trainrp” that updates the weights and bias

12 values according to the resilient backpropagation algorithm (Rprop, e.g. Riedmiller & Braun,

13 1993, Igel & Hüsken, 2003). The best nets were detected with regard to the overall fit of the

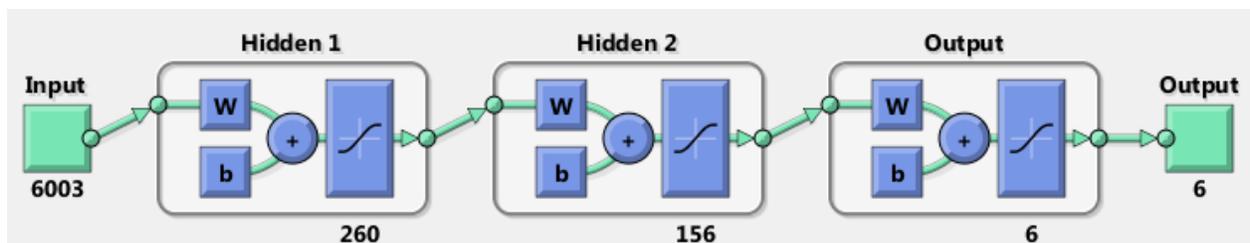
14 classifications in the test and validation part of the set. Additionally, a new dataset of 111

15 participants plus another new dataset with 62 participants and 100 trials each was categorized by

16 the selected nets and the net with the best performance in this new data set concerning the

17 classification above chance was finally selected (Table 1, chance level for 6 categories: $1/6 =$

18 16.66%).

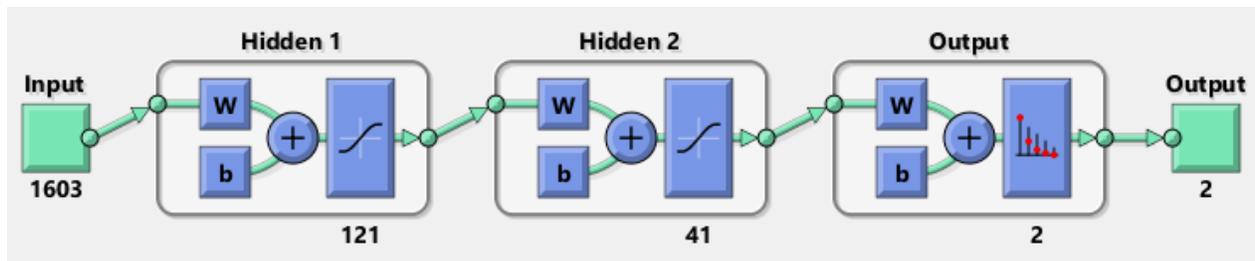


19

20 **Figure 4: Structure of the two-layer pattern classifier of the first example.**

1 **Classification problem example 2: Automatic IC artifact detection in EEG research.**

2 The input vector structure of the neuronal net for the second example consisted of 1603
3 elements, with 1500 elements being the mean time-course of the component, 32 elements being
4 the topographical information of the component (here for a 32 electrode EEG montage), 70
5 elements being the frequency band information of the component (from 1-70 Hz) and one
6 element being the participant number. For the output vector structure, we used two binary
7 elements, classifying whether an artifact or clean signal is given in the component. The neural
8 networks were trained with 51 participants with 32 components each, leading to 1632 different
9 input vectors and a 1603 x 1632 input matrix for the training and determination of the
10 performance. The input vectors and the output vector structure were used to compute the
11 necessary two-layer feedforward network to solve the classification problem. According to the
12 formula given by Huang et al. (2003), the first layer consisted of $\sqrt{(2 + 2) * 1632} + 2 *$
13 $\sqrt{\frac{1632}{2+2}} = 120.19 \approx 121$ neurons. The second layer consisted of $2 * \sqrt{\frac{1632}{2+2}} = 40.39 \approx 41$
14 neurons. As in the first example, the differentiation of the training set, the validation set and the
15 test set was always kept at 70 % training data, 15 % validation data and 15 % test data. Also, the
16 necessary validation trials for the different neural networks were set from 100 to 150 to 200 trials
17 and the two different training algorithms “trainscg” and “trainrp” were used in order to compare
18 their performance. The best nets were detected with regard to the overall fit of the classifications
19 in the test and validation part of the set. Additionally, a new dataset of 52 participants and 32
20 components of resting EEG was categorized by the selected nets.



1

2 **Figure 5: Structure of the two-layer pattern classifier of the second example.**3 **Software**

4 The neural networks were made with matlab 2011b and the neural network toolbox
 5 Version 7.0.2 for the first example and with matlab R2015b and the neural network toolbox
 6 Version 8.4 for the second example. The virtual T-maze paradigm was designed with a freely
 7 available game engine (Source SDK, Valve, Bellevue, Washington, USA), and the participants'
 8 movement and joystick input were acquired using the VR-experimentation software
 9 CyberSession CS-Research 5.6 (VTplus, Würzburg, Germany; see www.cybersession.info for
 10 detailed information). EEG was recorded from 32 electrode positions using BrainVision
 11 BrainAmp Standard and Brain Vision Recorder software. For further computation, MATLAB
 12 and EEGLAB toolbox (Delorme & Makeig, 2004) were used.

13

Results14 **Classification problem example 1: Behavior in virtual T-maze**

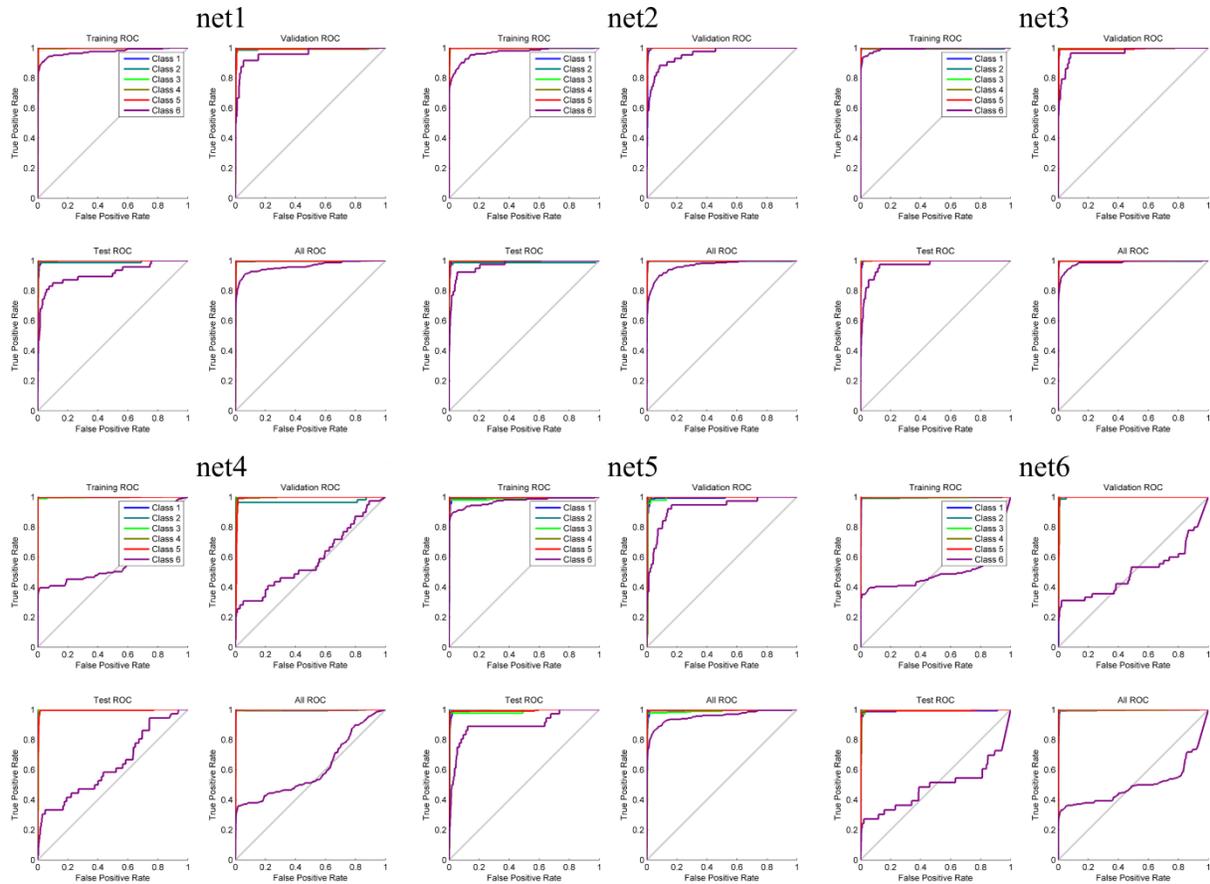
15 The different two-layer feedforward networks were compared using their overall
 16 classification performance. All pattern classifiers performed very well and achieved classification
 17 accuracy from 96.2% to 98.4% (Table 1).

18 **Table 1: Overall performance accuracy, validation trials, algorithms and trials which could**
 19 **not be categorized in the two new datasets for different two-layer feedforward networks.**
 20 **The networks were composed of 260 neurons in the first layer and 156 in the second layer.**

Two-layer pattern classifiers	overall performance: classification accuracy	validation trials	algorithm	Trials missed in new dataset1 (chance level .17)	Trials missed in new dataset2 (chance level .17)
net1	98.2	100	trainscg	3	2
net2	97.9	150	trainscg	16	4
net3	98.4	200	trainscg	10	5
net4	96.4	100	trainrp	239	153
net5	97.1	150	trainrp	85	56
net6	96.2	200	trainrp	258	162

1 **Note:** The net that has been finally selected has been highlighted. The behavioral categories were
2 selected against the threshold of $p = .17$, meaning that if the indicated probability of a behavioral
3 category exceeded this threshold, this category was selected. This specific threshold was selected
4 because of the chance level of selecting a category against 6 possible categories was $p = .1667$ and
5 rounded upto $p = .17$.

6 However, there were substantial differences in the classification of the conflict category
7 (class 6, Figure 6) and a subsequent classification of a new dataset from the same virtual T-maze
8 (Table 1). The ROC graphs (e.g. Fawcett, 2006) that can be seen in Figure 6 have the
9 "true positive rate" $\approx \frac{\text{positives correctly classified as positives}}{\text{total true positives}}$ on the y axis and the
10 "false positive rate" $\approx \frac{\text{negatives incorrectly classified as positives}}{\text{total true negatives}}$ on the x axis. Hence the ROC graph
11 depicts the relative tradeoffs between benefits (true positive) and costs (false positive). As a
12 perfect classification is reached if only true positives are classified, curves that tend towards the
13 upper left corner of the ROC graphs depict more accurate classifications and therefore better
14 classification performance, as can be seen for net1, net2, net3 and net 5. If a curve is pointing
15 towards the lower right corner of the graph, it indicates that the misclassification is prevalent, as
16 it can be seen in net 4 and especially in net 6.



1

2 **Figure 6: Receiver operating characteristic for the different neural networks. The class**
 3 **number of the different behavioral categories can be seen in Figure 2.**

4

5 Based on their overall classification performance

6 χ^2 McNemar performance of two best networks compared to others $> 9.529, p < .01$ and the very low count of missed
 7 trials in the new dataset, the two neural networks net1 and net3 were selected for a more detailed
 8 comparison.

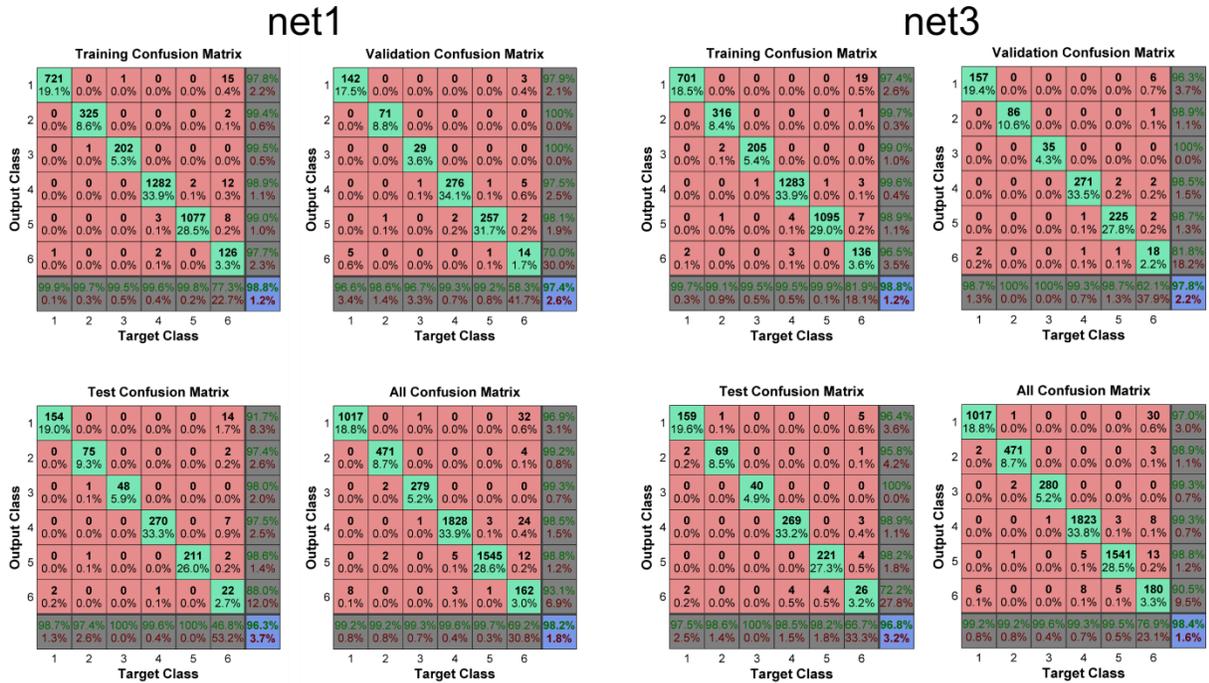
9 The confusion matrices and additional information for net1 and net3 are displayed below

10 in Figure 7 and Table 2. The confusion matrices depict the "precision" \approx

11 $\frac{\text{positives correctly classified as positives}}{\text{total classification as positives}}$ (e.g. Fawcett, 2006) in the right column and the true positive

12 rate in the bottom row for the training set, the validation set, the test set and overall for the two

1 different two-layer feedforward networks. Table 2 depicts the true positives, false positives, false
 2 negatives and true negatives of every behavioral category along with the false positive rate, the
 3 true positive rate, precision and "accuracy" \approx
 4 $\frac{\text{positives correctly classified and negatives correctly classified}}{\text{all cases}}$ (e.g. Fawcett, 2006).



5
 6 **Figure 7: Confusion matrices of the different two-layer feedforward networks. The class**
 7 **number of the different behavioral categories can be seen in Figure 2.**
 8

9 **Table 2: Common performance metrics calculated from confusion matrix.**

neural network	behavioral category	true positive	false positive	false negative	true negative	false positive rate:	true positive rate	precision	accuracy
net1	1	1017	33	8	4338	0.8%	99.2%	96.9%	99.2%
	2	471	4	4	4917	0.1%	99.2%	99.2%	99.9%
	3	279	2	2	5113	0.0%	99.3%	99.3%	99.9%
	4	1828	28	8	3532	0.8%	99.6%	98.5%	99.3%
	5	1545	19	4	3828	0.5%	99.7%	98.8%	99.6%
	6	162	12	72	5150	0.2%	69.2%	93.1%	98.4%
net3	1	1017	31	8	4340	0.7%	99.2%	97.0%	99.3%
	2	471	5	3	4917	0.1%	99.4%	98.9%	99.9%

3	280	2	1	5113	0.0%	99.6%	99.3%	99.9%
4	1823	12	13	3548	0.3%	99.3%	99.3%	99.5%
5	1541	19	8	3828	0.5%	99.5%	98.8%	99.5%
6	180	19	54	5143	0.4%	76.9%	90.5%	98.6%

1

2 Taken together all the information and comparisons of the two neural networks suggested

3 by Dougherty (2013), we decided to use the net1 for further classification of the behavior in the

4 paradigm described above, although no clear decision was indicated by the statistical tests: As the

5 overall classification accuracies of the two neural networks were only marginally different

6 (net1: 98.2 %, net3: 98,4 % $X^2_{McNemar}$ performance net1 vs. net3 = 2.778, $p = .096$), we also had a look on

7 the missing in classification of an entirely new and similar dataset from the paradigm (see Table

8 1). Still, this comparison was not conclusive for a decision between the two solutions

9 (net1: 3 missed trials, net3: 10 missed trials, $X^2_{McNemar}$ net1 vs. net3 = 3.769, $p = .052$). Finally, we

10 decided to use net1 because of the more pronounced marginally significant effect of less missed

11 trials in the classification of the new dataset compared to the marginal effect of overall

12 performance, along with the less validation trials boundary that was set for net1 (Table 1). As the

13 final selection of the neural network of interest was more based on the very low count of missed

14 trials in the new dataset than the overall performance, this decision may of course be debated and

15 net3 may also be selected as best fitting neural network, although it may also be slightly more

16 prone to overfitting because of more validation trials.

17 **Classification problem example 2: Automatic IC artifact detection in EEG research**

18 The different two-layer feedforward networks were compared using their overall

19 classification performance. All pattern classifiers performed rather well and achieved

20 classification accuracy from 88.8% to 94.6% (Table 3). The resting EEG was successfully

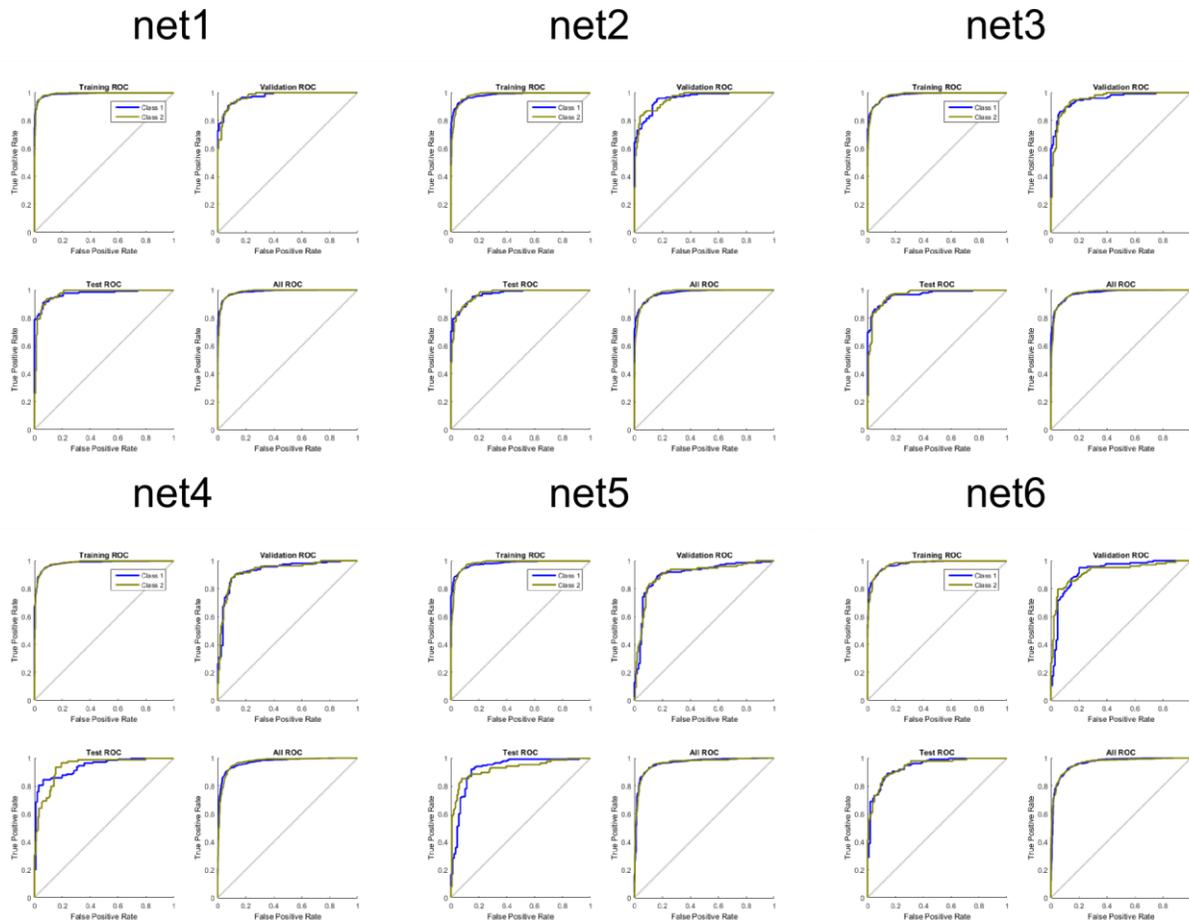
21 categorized by all selected nets.

1 **Table 3: Overall performance accuracy, validation trials, algorithms and trials which could**
 2 **not be categorized in the resting dataset for different two-layer feedforward networks. The**
 3 **networks were composed of 121 neurons in the first layer and 41 in the second layer.**
 4

Two-layer pattern classifiers	overall performance: classification accuracy	validation trials	algorithm	Trials missed in resting data set (chance level .50)
net1B	94.6	100	trainscg	0
net2B	92.1	150	trainscg	0
net3B	91.8	200	trainscg	0
net4B	92.0	100	trainrp	0
net5B	91.4	150	trainrp	0
net6B	88.8	200	trainrp	0

5 **Note:** The net that has been finally selected has been highlighted.

6 The ROC graphs can be seen in Figure 8. All nets perform rather well, although the
 7 backpropagation algorithm networks perform slightly worse than the other networks.



1
 2 **Figure 8: Receiver operating characteristic for the different neural networks. Class 1 means**
 3 **artifacts, Class 2 means signal.**

4 Based on the overall classification performance
 5 $X^2_{McNemar}$ performance of net1B to others > 41.00 , $ps < .001$, net1B was chosen, although the easy
 6 classification task of the resting EEG was solved by all neural networks. Details of all neural
 7 networks are depicted in Table 4.

8 **Table 4: Common performance metrics calculated from confusion matrix.**

neural network	component type	true positive	false positive	false Negative	true negative	false positive rate:	true positive rate	precision	accuracy
net1B	artifact	967	49	39	577	7.8%	96.1%	95.2%	94.6%
	signal	577	39	49	967	3.9%	92.2%	93.7%	94.6%
net2B	artifact	945	68	61	558	10.9%	93.9%	93.3%	92.1%
	signal	558	61	68	945	6.1%	89.1%	90.1%	92.1%

net3B	artifact	937	65	69	561	10.4%	93.1%	93.5%	91.8%
	signal	561	69	65	937	6.9%	89.6%	89.0%	91.8%
net4B	artifact	938	62	68	564	9.9%	93.2%	93.8%	92.0%
	signal	564	68	62	938	6.8%	90.1%	89.2%	92.0%
net5B	artifact	929	64	77	562	10.2%	92.3%	93.6%	91.4%
	signal	562	77	64	929	7.7%	89.8%	87.9%	91.4%
net6B	artifact	876	53	130	573	8.5%	87.1%	94.3%	88.8%
	signal	573	130	53	876	12.9%	91.5%	81.5%	88.8%

1

2 To sum it up, in this example the net1B performed very well with the best accuracy,
3 although all other networks created here were also able to classify the task of resting EEG
4 without any omissions.

5

Discussion

6 In this work, we showed that it is possible to create neural networks to classify continuous
7 behavior task as well as ICA artifact classification as examples for manual tasks that were
8 previously categorized by experts. We used two-layer feedforward networks to classify our
9 examples for any expert categorization decision that can be solved with a reasonable input vector
10 structure. The structure of the net was computed using the formula given by Huang (2003) and
11 this formula led to effective results concerning the fitting of the different neural networks.
12 Nevertheless, it could be seen that it is important to test the different neural networks for their
13 performance on new data in order to see whether this data can be categorized in meaningful ways
14 and no overfitting is given. As we tested two different training algorithms and therefore
15 propagation rules, we found that the resilient backpropagation algorithm (net4, net 5, net 6) led to
16 less favorable results than the scaled conjugate gradient method (net1, net2, net3), as one would

1 expect from benchmark tests (The MathWorks, 2019) as well as that the architecture given by
2 Huang (2003) was intended for feedforward propagation. This leads to the assumption that the
3 resilient backpropagation as it is implemented in matlab 2011b and the nntoolbox Version 7.0.2
4 might not be the optimal algorithm to be used in the two-layer feedforward networks, or at least
5 in this specific problem. However, this result has also been replicated by our workgroup
6 (Rodrigues et al., in press), so we believe this finding to be fairly robust. Also, the benchmark
7 tests suggest that the scaled conjugate gradient algorithm performs well over a variety of
8 classification problems (The MathWorks, 2019). Another interesting result was that the network
9 with the best ROC curve and confusion matrix concerning the input vectors (net3) did not
10 perform best on the new data, but a network that was very close to this performance (net1). This
11 may indicate that the best performing net can already be prone to overfitting, leading to a too
12 specific solution of the classification and therefore to problems with the new data. This makes the
13 testing with new data a very important step of the validation process of the automatic
14 classification. However, the difference in performance on the trained and new dataset between
15 the two neural networks is rather small and one might argue whether the preference of net1 vs.
16 net3 is a necessary choice. Nevertheless, we chose net1 as we were interested in the classification
17 of as many behavioral trials as possible with a reasonable solution, also on new data that is still to
18 come. As the importance of losing trials or having a potential overfitting might be very different
19 in other cases, other researchers might prefer solutions that are similar to net3 in their specific
20 case. Nevertheless, one conclusion can be drawn from these results: it is still important to
21 perform tests of the neural networks on new data and it is important to check whether the network
22 that was chosen is able to categorize the data sufficiently accurate.

23 For our second example, there was no difference in the performance of the created neural
24 networks, in a slightly easier task than the nets were created for. This indicates that the difficulty

1 of the problem used to create the net should match the difficulty of the new data sets if possible,
2 if no other criteria than the classification omission can be used (fortunately, there is a significant
3 performance difference that we could use to select net1B as explained above). Having created
4 neural networks for behavioral responses as well as for IC artifact detection in this virtual T-maze
5 task, we have achieved solutions for two problems that used to be solved by expert decisions.
6 Now these solutions can be used by other researchers in order to get a standardized method that
7 can be applied by anyone to these problems. Thus, we can get more reliable and replicable results
8 in this kind of behavioral classification as well as in this specific IC artifact detection task.

9 **Seven questions about the two-layer feedforward networks and their impact on the field**

10 Having reiterated the findings for the examples in detail, we may ask important questions:

11 **1) How to create a neural network for the given problem?**

12 Huang (2003) provides a detailed instruction to create neural network for classification than can
13 be used for every task where a meaningful input vector can be created and where an expert has
14 already taken an approach on the classification. This does not imply that the entire task is already
15 categorized, but a sufficient learning set for the task should be given. The architecture of the net
16 is dependent on the sample size and output vector, as mentioned in the method section above.

17 **2) Is the neural network “better” than the expert that was used to create it?**

18 This question cannot be answered easily, but there are different perspectives on that
19 question. First, one may say that the neural net can be shared with other researchers and therefore
20 this expert opinion can be shared without having to interpret or even trying to teach this kind of
21 expertise. This enables many researchers to use the expertise without the expert having to
22 perform every time. Also, the time used to classify the behavior by hand would be several days or
23 weeks, while the neural net performs it in seconds or minutes. Hence, we get a faster and widely
24 shared version of the expert’s opinion. Nevertheless, every bias that was consistently present in

1 the expert is now transferred to the neural network. The net merely reproduces the classification
2 without variable biases that may have been introduced by different states, but the consistent
3 biases of the expert are still present. This is very important to keep in mind, as the network does
4 not guarantee a correct classification (as is often seen by neural networks being “racists” in case
5 of threat detection or income- and work-related questions). It is only a generalization of the
6 learned classification that can be used to reliably and consistently classify the data. Also, the
7 experts are able to adapt their judgements to new types of classes that may arise in new datasets
8 (for example a new type of behavior showing up or a new class of IC artifact due to a different
9 EEG-setup), while the net will not be able to do so. Hence new types of classes will be not
10 detected and classified in a wrong manner (see e.g. pulse related artifacts from templar structures
11 in MARA for EEG IC artifact classification.)

12 **3) Is one neural network enough to classify a task?**

13 While the architecture provided by Huang (2003) is very promising to create neuronal
14 networks to replicate every classification solution by experts, it is still important to check what
15 differences are given by the differences of the learning rule and the propagation restrictions, as
16 well as the training and validation sampling size. Still, not only the validation in the sense of the
17 machine learning sample error needs to be evaluated, but also the applicability to new datasets in
18 order to identify well-fitting networks that are not over-fitted to the respective training data.
19 Hence, it is important to create different networks with slightly different training samples in order
20 to select the neural network that provides a fitting combination of the least overfitting and the
21 best classification performance.

22 **4) Are different neural networks needed for every single task?**

23 If the tasks are similar, with identical output and input categories, then a new neuronal net
24 will not be needed. Otherwise one needs to create a different neural network or at least check the

1 results concerning plausibility as well as for classification accuracy and uncategorized cases. If
2 the input vector size is different, a new net will be needed.

3 **5) How to create the input vector and output vector?**

4 This question is essential in order to create a successful neural network and is not easy to
5 be answered. The output vector should be your classification results as logical/binary vectors.
6 The count of the output nodes is the count of your classes. However, for the input vector, the
7 structure is not that easy to determine. You have to identify necessary information that the expert
8 uses to create the classification and exclude information that is irrelevant in order to get a good
9 performing neural network for the problem. In our first example, this meant not including all data
10 from the entire behavioral performance, but only the relevant time point that we also used for
11 manual classification. For the second example, this was even more restrictive, as an input vector
12 of $120 \text{ (trials)} * 1500 \text{ (datapoints EEG segment)} = 180\,000$ datapoints would have been possible,
13 but most of the data would not have given any information about the classification. As the
14 manual task performed by the expert is also based on the frequency distribution, the topography
15 and the mean time course of the component, we identified these features as a meaningful input
16 vector in this case.

17 This task to identify the input and output vector is very important and it takes an expert to
18 identify the needed features of your task for a successful automatic classification.

19 **6) Will experts become “worthless” as their opinion can be shared efficiently?**

20 No! As explained above, experts provide meaningful classification sets, the evaluation of
21 features that can be used to create the architecture of the neural network as well as to assess
22 whether the neuronal networks are able to perform adequately and without any performance
23 errors in the classification of data. Their expertise will be the ground for other researchers to
24 build their neural networks on and therefore their work is very important and should be valued.

1 **7) How may this standardization of data classification have impact on our field?**

2 Hopefully, the standardization and availability of these neural networks to complex
3 decisions can lead to a more replicable and reliable research in our field. Complex tasks such as
4 behavioral categorization in animal research, automatic picture classification, movement decision
5 categorization according to continuous data or complex physiological patterns that were only
6 categorized by experts may now be solved by different experts via trained neural networks. As
7 every expert may train a different net or even more experts may train a net on their consent,
8 different replicable classifications for complex data may arise and can be compared. This does
9 not only provide the basis of reliable and comprehensible research, but also may provide insights
10 into the differences in expert opinions as one may choose to apply different neural networks to
11 the data and compare those resulting classifications. A productive discussion can arise from these
12 comparisons, as perspectives of these different experts may not only be debated on the theoretical
13 level, but also on different concrete datasets. Therefore, not only a general reliability and more
14 replicable science may rise, but also the experts may further discuss the specificity of their
15 classifications and the appropriate expert solution may be applied to the data at hand.

16 As mentioned above, a similar approach has already been taken for other problems too
17 (e.g. classification of sleep stages: Barschdorff et al., 1996; Baumgart-Schmitt et al., 1997,) and
18 in theory, one might be able to construct a two-layer feedforward neural network for virtually
19 every classification problem (Huang, 2003). Neural networks that can be provided to other
20 researchers do not only lead to a quick solution for problems at hand, but they also provide a
21 standardized solution instead of introducing correction by eye or other methods that are hard to
22 replicate. Huang (2003) provides a very easy formula to construct two-layer feedforward neural
23 networks to address virtually every categorization problem and with the present work we try to
24 encourage other researchers to also work on neural networks solving tasks that are yet done by

1 hand, in order to standardize the results. Although experts might get sometimes the impression of
2 “losing their expertise to others” by publishing and providing their expertise to other via these
3 neural networks, it is important to convince them that they can help the field to advance if they
4 contribute their knowledge in training and creating neural networks for the whole scientific
5 community.

6 In summary, in order to get to a good and reliable science, it is important to not only
7 publish interesting and marvelous results, but also to provide the scientific community with the
8 necessary tools to solve problems in standardized ways (e.g. Munafò et al., 2017). Using and
9 applying the work of Huang (2003) we may build standardized classification tools for many
10 problems as shown here for the examples of the behavioral classification of continuous
11 movements in the virtual T-maze and the IC-artifact detection in the same task. Hence, we may
12 be able to gradually work our way to a reliable and sharing scientific community, tackling
13 replication problems and opening a new level of expert discussions as more datasets may be
14 classified by different expert approaches.

15

References

- 1
2 Anderson, T., Wright, D. L., & Immink, M. A. (1998). Contextual Dependencies during
3 Perceptual-motor Skill Performance: Influence of Task Difficulty. *Memory*, 6(2), 207–221.
4 <https://doi.org/10.1080/741942069>
- 5 Barschdorff, D., Gerhardt, D., & Trowitzsch, E. (1996). Rapid eye movement detection in infants
6 using a neural network. In Proceedings of 18th Annual International Conference of the
7 IEEE Engineering in Medicine and Biology Society (Vol. 3, pp. 935–936). IEEE.
8 <https://doi.org/10.1109/IEMBS.1996.652648>
- 9 Baumgart-Schmitt, R., Eilers, R., & Herrmann, W. M. (1997). On the use of neural network
10 techniques to analyze sleep EEG data. *Somnologie - Zeitschrift Für Schlafforschung Und*
11 *Schlafmedizin*, 1(4), 171–183. <https://doi.org/10.1007/s11818-997-0031-3>
- 12 Chaumon, M., Bishop, D. V. M., & Busch, N. A. (2015). A practical guide to the selection of
13 independent components of the electroencephalogram for artifact correction. *Journal of*
14 *Neuroscience Methods*, 250, 47–63. <https://doi.org/10.1016/j.jneumeth.2015.02.025>
- 15 Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial
16 EEG dynamics including independent component analysis. *Journal of Neuroscience*
17 *Methods*, 134(1), 9–21. <https://doi.org/10.1016/j.jneumeth.2003.10.009>
- 18 Dougherty, G. (2013). *Pattern Recognition and Classification*. New York, NY: Springer New
19 York. <https://doi.org/10.1007/978-1-4614-5323-9>
- 20 Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8), 861–
21 874. <https://doi.org/10.1016/J.PATREC.2005.10.010>

- 1 Flexer, A., Bauer, H., Pripfl, J., & Dorffner, G. (2005). Using ICA for removal of ocular artifacts
2 in EEG recorded from blind subjects. *Neural Networks*, 18(7), 998–1005.
3 <https://doi.org/10.1016/j.neunet.2005.03.012>
- 4 Guang-Bin Huang. (2003). Learning capability and storage capacity of two-hidden-layer
5 feedforward networks. *IEEE Transactions on Neural Networks*, 14(2), 274–281.
6 <https://doi.org/10.1109/TNN.2003.809401>
- 7 He, K., Zhang, X., Ren, S., & Sun, J. (2015). Delving Deep into Rectifiers: Surpassing Human-
8 Level Performance on ImageNet Classification. Retrieved from [https://www.cv-
foundation.org/openaccess/content_iccv_2015/html/He_Delving_Deep_into_ICCV_2015_
paper.html](https://www.cv-
9 foundation.org/openaccess/content_iccv_2015/html/He_Delving_Deep_into_ICCV_2015_
10 paper.html)
- 11 Igel, C., & Hüsken, M. (2003). Empirical evaluation of the improved Rprop learning algorithms.
12 *Neurocomputing*, 50, 105–123. [https://doi.org/10.1016/S0925-2312\(01\)00700-7](https://doi.org/10.1016/S0925-2312(01)00700-7)
- 13 Ioannidis, J. P. A. (2005). Why Most Published Research Findings Are False. *PLoS Medicine*,
14 2(8), e124. <https://doi.org/10.1371/journal.pmed.0020124>
- 15 Kipnis, D., & Schmidt, S. M. (1988). Upward-Influence Styles: Relationship with Performance
16 Evaluations, Salary, and Stress. *Administrative Science Quarterly*, 33(4), 528.
17 <https://doi.org/10.2307/2392642>
- 18 Lawrence, S., & Giles, C. L. (2000). Overfitting and neural networks: conjugate gradient and
19 backpropagation. In *Proceedings of the IEEE-INNS-ENNS International Joint Conference
20 on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives
21 for the New Millennium* (pp. 114–119 vol.1). IEEE.
22 <https://doi.org/10.1109/IJCNN.2000.857823>

- 1 Makeig, S., Debener, S., Onton, J., & Delorme, A. (2004). Mining event-related brain dynamics.
2 *Trends in Cognitive Sciences*, 8(5), 204–210. <https://doi.org/10.1016/j.tics.2004.03.008>
- 3 McFarland, K., & Ashton, R. (1978). The influence of concurrent task difficulty on manual
4 performance. *Neuropsychologia*, 16(6), 735–741. [https://doi.org/10.1016/0028-](https://doi.org/10.1016/0028-3932(78)90008-8)
5 [3932\(78\)90008-8](https://doi.org/10.1016/0028-3932(78)90008-8)
- 6 Minin, A. S. (2012). Modeling of Dynamical Systems with Complex-Valued Recurrent
7 Neural Networks. Technical University of Munich, Germany.
- 8 Møller, M. F. (1993). A scaled conjugate gradient algorithm for fast supervised learning. *Neural*
9 *Networks*, 6(4), 525–533. [https://doi.org/10.1016/S0893-6080\(05\)80056-5](https://doi.org/10.1016/S0893-6080(05)80056-5)
- 10 Mognon, A., Jovicich, J., Bruzzone, L., & Buiatti, M. (2011). ADJUST: An automatic EEG
11 artifact detector based on the joint use of spatial and temporal features. *Psychophysiology*,
12 48(2), 229–240. <https://doi.org/10.1111/j.1469-8986.2010.01061.x>
- 13 Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie du Sert,
14 N., ... Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human*
15 *Behaviour*, 1(1), 0021. <https://doi.org/10.1038/s41562-016-0021>
- 16 Open Science Collaboration (2015). Estimating the reproducibility of psychological science.
17 *Science*, 349(6251), aac4716–aac4716. <https://doi.org/10.1126/science.aac4716>
- 18 Pilcher, J. J., & Huffcutt, A. I. (1996). Effects of Sleep Deprivation on Performance: A Meta-
19 Analysis. *Sleep*, 19(4), 318–326. <https://doi.org/10.1093/sleep/19.4.318>

- 1 Rey, G. D., & Wender, K. F. (2011). *Neuronale Netze : eine Einführung in die Grundlagen,*
2 *Anwendungen und Datenauswertung.* Huber. Retrieved from
3 <https://www.goodreads.com/book/show/12720412-neuronale-netze>
- 4 Riedmiller, M., & Braun, H. (1993). A Direct Adaptive Method for Faster Backpropagation
5 Learning: The RPROP Algorithm. *IEEE INTERNATIONAL CONFERENCE ON*
6 *NEURAL NETWORKS*, 16, 586--591. Retrieved from
7 <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.21.1417>
- 8 Rodrigues, J. (2016). Let me change your mind... Frontal brain activity in a virtual T-maze.
9 Retrieved from [https://opus.bibliothek.uni-](https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/14328)
10 [wuerzburg.de/frontdoor/index/index/docId/14328](https://opus.bibliothek.uni-wuerzburg.de/frontdoor/index/index/docId/14328)
- 11 Rodrigues, J., Müller, M., & Hewig, J. (in press). Cardiac defensive reactions and orienting
12 responses correspond to virtual withdrawal behavior choices in a virtual T-maze.
13 *International Journal of Psychophysiology.*
- 14 Rodrigues, J., Müller, M., Mühlberger, A., & Hewig, J. (2018). Mind the movement: Frontal
15 asymmetry stands for behavioral motivation, bilateral frontal activation for behavior.
16 *Psychophysiology*, 55(1), e12908. <https://doi.org/10.1111/psyp.12908>
- 17 Sigala, N., Gabbiani, F., & Logothetis, N. K. (2002). Visual Categorization and Object
18 Representation in Monkeys and Humans. *Journal of Cognitive Neuroscience*, 14(2), 187–
19 198. <https://doi.org/10.1162/089892902317236830>
- 20 The MathWorks, I. (2011). *MATLAB Version 7.13.0.564 release 2011b.* Natick, Massachusetts,
21 United States. Retrieved from www.mathworks.com

- 1 The MathWorks, I. (2011). Neural Network Toolbox Version 7.0.2 release 2011b. Natick,
2 Massachusetts, United States. Retrieved from www.mathworks.com
- 3 The MathWorks, I. (2019). Choose a Multilayer Neural Network Training Function. Retrieved
4 from [https://de.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-](https://de.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html#responsive_offcanvas)
5 [training-function.html#responsive_offcanvas](https://de.mathworks.com/help/deeplearning/ug/choose-a-multilayer-neural-network-training-function.html#responsive_offcanvas)
- 6 Valve. (2017). Source SDK. Bellevue, WA. Retrieved from
7 https://developer.valvesoftware.com/wiki/Source_SDK:de
- 8 VTplus. (2017). CyberSession CS.Research 5.6. Würzburg, Germany. Retrieved from
9 www.cybersession.info
- 10 Winkler, I., Haufe, S., & Tangermann, M. (2011). Automatic Classification of Artifactual ICA-
11 Components for Artifact Removal in EEG Signals. *Behavioral and Brain Functions*, 7(1),
12 30. <https://doi.org/10.1186/1744-9081-7-30>
- 13 Zhang, Q. J., & Gupta, K. C. (2000). *Neural networks for RF and microwave design*. Artech
14 House.

15 **Repositories: FTP:**

- 16 <http://update.cybersession.info/T-Maze/VrProject.T-Maze.zip> (282 MB, 2019-10-15)
17 http://update.cybersession.info/T-Maze/TwoLayerFeedForward_Data.zip (330 MB)

18 **Repositories: SVN:**

- 19 <https://crm.vtplus.eu/svn/open-access.t-maze/trunk>
20 <https://crm.vtplus.eu/svn/open-access.t-maze/trunk/Paradigm>

- 1 https://crm.vtplus.eu/svn/open-access.t-maze/trunk/sourcemos/VrSessionMod06_T-maze