University of Tartu
Faculty of Science and Technology
Institute of Bioengineering
Science and Technology Curriculum

Aleksandrs Rebriks

# Emotion Recognition using EEG signal data from EMO2018 Dataset

Master's Thesis (30 ECTS)

Supervisor(s): Egils Avots, PhD
Liina Juuse, MA

Tartu 2024

## Abstract

Emotion Recognition (ER) is developing area within the artificial intelligence field that is focused on comprehending and further interpreting of human emotions through various modalities. Despite that, these approaches are often not ubiquitous as they are affected by external factors. With recent physiology research connecting development of emotions to the central nervous system, usage of brain signals became a highly practical option for emotion recognition. One of the most promising methods of emotion recognition using brain signals for emotion recognition involves using Electroencephalography (EEG). Despite being more complex than classical machine learning or deep learning approaches, EEG-based emotion recognition is potentially more accurate and robust, with applications in mental health monitoring, researches in applied physiology or human-computer interactions. This thesis studies existing approaches of EEG-based emotion recognition methods for private EMO2018 dataset. We adopted methods of Fast Fourier Transform with additional processing for key features extraction and tested different Deep Learning models. Our results show performances of utilized Deep learning models with best accuracy of 88.6% from Hybrid Neural Network approach.

**Keywords:**

EEG, Emotion recognition, Machine learning, Deep learning, Hybrid Neural Networks, CNN, LSTM

**CERCS:**

T121 Signal processing, P176 Artificial intelligence, B640 Neurology, neuropsychology, neurophysiology

## Tüübituletus neljandat järku loogikavalemitele

### Lühikokkuvõte:

Emotsioonide tuvastamine (ER) on tehisintellekti valdkonnas arenev ala, mis keskendub inimemotsioonide mõistmisele ja tõlgendamisele erinevate modaliteetide kaudu. Vaatamata sellele ei ole need lähenemisviisid sageli kõikjal levinud, kuna neid mõjutavad välised tegurid. Viimaste füsioloogiauuringute käigus on emotsioonide arengut seostatud kesknärvisüsteemiga, mistõttu on ajusignaalide kasutamine muutunud emotsioonide tuvastamisel väga praktiliseks võimaluseks. Üks paljulubavamaid meetodeid emotsioonide tuvastamiseks ajusignaalide abil hõlmab elektroentsefalograafia (EEG) kasutamist. Kuigi EEG-põhine emotsioonide tuvastamine on keerulisem kui klassikalised masinõppe või süvaõppe lähenemisviisid, on see potentsiaalselt täpsem ja vastupidavam, pakkudes rakendusi vaimse tervise jälgimises, rakendusfüsioloogia uuringutes või inimese ja arvuti interaktsioonides. Käesolevas uuringus analüüsisime olemasolevaid EEG-põhiseid emotsioonide tuvastamise meetodeid EMO2018 privaatse andmekogu põhjal. Kasutasime kiire Fourier' teisenduse (FFT) meetodeid koos täiendava töötlemisega võtmetunnuste eraldamiseks ja testisime erinevaid süvaõppemudeleid. Meie tulemused näitavad kasutatud süvaõppemudelite sooritust, parima täpsusega 88,6%, saavutatud hübriidse närvivõrgu lähenemise abil.

### Võtmesõnad:

EEG, Emotsioonide tuvastamine, Masinõppimine, Sügavõpe, Hübriidsed närvivõrgud, CNN, LSTM

### CERCS:

T121 Signaalitöötlus, P176 Tehisintellekt, B640 Neuroloogia, neuropsühholoogia, neurofüsioloogia

# Contents

# Terms, abbreviations and notations

**ML** - Machine Learning

**NN** - Neural Network

**ANN** - Artificial Neural Network

**AI** - Artificial Intelligence

**EEG** - Electroencephalogram

**ER** - Emotion recognition

**PAD** - Pleasure, Arousal, Dominance

**P.A.D.A** - Positive activation - negative activation

**DFT** - Discrete Fourier Transform

**FFT** - Fast Fourier Transform

**CNN** - Convolutional Neural Network

**RNN** - Recurrent Neural Network

**LSTM** - Long Short-Term Memory

**SPWVD** - Smoothed pseudo Wigner–Ville distribution

# 1 Introduction

Emotion recognition is a process of identifying emotional state and, hereby, mental state of a given human. With recent developments in artificial intelligence field, Human-Computer Interaction systems have become more robust in automatic emotion recognition from various modalities, of which, the most common ones include facial expressions, gestures and speech.

These approaches, however proven not to be appropriately accurate [37], since the feature extraction on those can be affected by external factors such as age, general appearance, habits or language of a particular subject. Recent studies concluded that emotions are to be recalled in the brain and central nervous system is directly related to emotion development [18], bringing the argument to use brain signals as a novel approach in emotion recognition. Electroencephalogram (EEG) signals, that record electric impulses produced from neural activity, are a particular interest as they are directly connected to brain activity.

On the one hand, EGG-based emotion recognition systems could potentially be more reliable and robust than established classical approaches, and EEG has already proven to be successful in various fields related to central nervous system [8-13]. On the other hand, despite focused research on adopting EEG for emotion recognition, it still remains a challenge, due to convoluted nature of EEG signals data structure for feature extraction as well as complexity in measurements and definitions of human emotions [53].

The aim of this research is to explore relevant studies to identify the types of models used on publicly available datasets. This will help build understanding and develop and test a robust model for emotion recognition from EEG signals, using the private, self-acquired dataset EMO2018.

The results of this study would be most applicable to any related future work with EMO2018 dataset or its applications on emotion recognition, as well as general contribution to a field of research of EEG based emotion recognition. In addition, the implication may also be extended into reflecting activities of central nervous system. Inspired models could be adopted to improve mental state, in, for example, health check applications, and improve user experience, in interfaces those can generate responses based on user state of mind.

# 2 Literature review

Advances in the field of Neural Networks allowed for their rapid adoption in various classification tasks and data processing. [47] utilized generic Convolutional Neural Network (CNN) model (pre-trained on ImageNet dataset) with transfer learning approach to improve validation accuracy (paper report an increase from 39.13% to 55.6%). In addition, the following paper [33], while focusing on image classification, demonstrated capabilities of Deep Convolutional Neural Network to achieve impressive prediction accuracy (67.4% compared to 78.1% as of best reported paper [45]) using only supervised learning [9]. This yielded CNN models to become very popular method in facial expressions based emotion recognition.

With development in Recurrent Neural Network (RNN) approaches, speech based emotion recognition systems were introduced. In comparison to CNN models, it was reported that extracted speech based features carried more valuable information regarding specific emotion, despite facial expression features resulting in better performance [8].

Following the establishment of relations between emotion development and central neural system, EEG based emotion recognition could potentially yield to better results. The advantage of EEG data is to be minimally affected by human characterized external factors, such as appearance, age, language and culture. Many approaches adopting CNN, RNN variations and composite Neural Networks (NN) for emotion classification, although difficult to compare.

## 2.1 Emotions

An emotion is a spontaneous yet complex feeling that implies a reaction of a human subject to any particular event [52]. Even though a person could experience different types of affective phenomena, such as, moods, feelings or attitudes, to a particular event, those can simply be designated as emotions [43].

Different psychological studies have proposed various models for structuring emotions, most commonly used are presented in Table 1, those can be separated into two groups based on their dimensionality, particularly the two-dimensional models (2D) and three-dimensional models (3D). 2D models describes emotions with two values, namely valence, that is a measurement of positivity or negativity of an emotion, and arousal, that is a measurement of intensity of an emotion. 3D models use same valence and arousal dimensions with an addition of dominance, that is a measurement of influence (or degree of control) of an emotion, ranging from submissive to dominant.

Table 1. Most notable emotion models with model name, author and short summary

| Model name | Author | Core concept/Details |
|---|---|---|
| Ekman's model [15] | Paul Ekman | Basic emotions separation to happiness, fear, anger, disgust, surprise, sadness |
| Parrot's model [48] | Parrot, W. G | Basic emotions separation to joy, sadness, anger, surprise, fear, love |
| Plutchik's model [50] | Robert Plutchik | Basic emotions separation to fear, anger, disgust, sadness, surprise, joy, trust, anticipation |
| PAD model [1] | Albert Mehrabian and James A. Russell | 3D model based on pleasure, arousal and dominance |
| Vector model [28] | Bradley, M.M | Arousal and valence vectors |
| Positive activation - negative activation (P.A.D.A) model [62] | Watson and Tellegen | 2D model, y-axis representing low-to-high positive affect, x-axis low-to-high negative affect |

## 2.2 Electroencephalogram

Electroencephalogram (EEG) is a medical, non-invasive procedure that reads brain signals from electrical activity on the surface of the scalp. EEG works by measuring current flow between dendrites of neurons in cerebral cortex during synaptic excitation (e.g. brain cells activated).

In order to correctly identify and analyze EEG data it is important to have an understanding on brain structure and anatomy. The brain consists of three parts: cerebrum, brainstem and cerebellum, of which, cerebrum is the largest. Cerebrum is further separated into four distinct regions, namely lobes, that is frontal, pariental, occipital, temporal.

EEG data acquisition is typically done by placing a number of electrodes on subject scalp and providing a stimuli to induce emotional response. For emotion recognition the International 10/20 electrode placement (shown in Figure 1) is most commonly used [20]; this methodology was also employed in the EMO2018 dataset. A stimuli is often any audio-video piece of information (emotional photos, videos, audio recordings) or an emotional event that is used to record EEG signal data; EMO2018 dataset adopted a hybrid trigger in a form of video information (emotional words, faces and certain images) coupled with emotional tasks of how a subject should express the certain emotion.

Resulting EEG data consists of waveform signals, each corresponding to the electrode placed, ranging in 0.5 to 100 $\mu$V [58]. For EEG data analysis, a signal frequency spectrum (using Fourier transformations) is used as it can demonstrate a contribution for each frequency, that is very useful specifically for EEG-based Emotion Recognition [25].

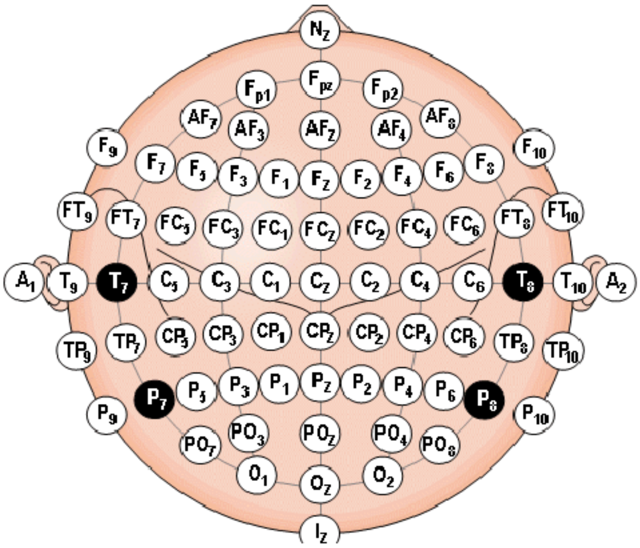


Figure 1. Visualisation of 10/20 system electrode placement from top head view [34]. Adjacent electrodes are in 10% or 20% distance of skull front-to-back and left-to-right measurements

## 2.3 Features of EEG signal

The most important step in developing any Human-Computer Interface (HCI) based system, especially emotion recognition systems, is to correctly identify key features. For an EEG signal, useful characteristics could be extracted from spatial information, spectral information and temporal information [39], those show relation of data points to some location, distribution of signal magnitudes across frequencies and signal changes over some range of time, respectively. Regarding EEG features, those could be utilized to acquire relation of each signal to specifically placed electrode it originates from, contribution of different frequencies to each signal and changes in brain electric activity over specific time interval.

Resulting EEG features could be generally separated into time domain, frequency domain and time-frequency domain. Even thought time-domain features are considered to be informative for emotion states classifying, those were utilized in many related studies, using different methods as statistical measurements [60], for example, Non-Stationary Index (NSI), Higher Order Crossings (HOC) [49], Hjorth features [6], Fractal Dimension (FD) [55] [38] and Event Related Potentials (ERP) [16].

On the other hand, frequency domain features was used by the majority of studies [4], as those can suggest on predominant frequencies in reaction to emotion stimuli, resulting more useful learning patters for classification. Raw data, that is represented in time domain, can be transformed into frequency domain by the means of Discrete Fourier Transform (DFT) [56] and those popular derivatives: Fast Fourier Transform (FFT) [14], highly optimized algorithm to perform DFT on data and Short-time Fourier Transform, that performs DFT on slices of times series of data.

As reported in the following paper [3], FFT may have performance problems when dealing with non-linear and non-stationary signal type as EEG. To overcome this, data can be transformed into time-frequency domain, using very popular method, namely Wavelet Transform, that has proved to be well-optimized for analysis in both time and frequency domains [46]. Wavelet Transform can be separated into two types, namely Continuous Wavelet Transform (CWT) and Discrete Wavelet Transform (DWT). In EEG based emotion recognition, utilizing Discrete Wavelet Transform (DWT) is a common choice, as it decomposes a signal into sub-bands, allowing for better feature extraction as well as removing noise.

## 2.4 Machine learning

Machine learning (ML) is a branch of artificial intelligence (AI) that focuses on developing autonomous computer algorithms capable of self-improving over a period of time on various forms of data, thus resembling human-like intelligence. At the core of any machine learning algorithm, typically referred to as a model, is minimization of a loss function that quantifies the difference between the predicted output and actual target values. An ML model is capable of enhancing its prediction accuracy of a given task through a learning process, which is done by changing values or structures of its inner components. Producing a model consists of running it on a training set, that will be used by the model to learn and testing model output on a separate testing set. In a machine learning approach, it is typical to adopt, test and compare multiple ML models with numerous parameters to achieve a state-of-art solution for a given task.

Traditional machine learning algorithms, characterized by their relatively simpler structure and improved interpretation, can be referred as shallow learning algorithms. Despite the fact of requiring extensive data and feature preparation, these are very commonly used as classifiers for emotion recognition, bearing in mind that feature extraction from EEG signal is complex task. It is to note that the majority of related studies adopted multiple classifiers, among which the most comprehensive ones were Support Vector Machine (SVM) and its variations and K-Nearest Neighbor (kNN) [4].

## 2.5 Deep learning

In a wide realm of machine learning algorithms some models are engineered as interconnected networks, with numerous nodes, namely neurons, and connections between them, namely edges, organized into separate groups, namely layers, resembling a structure similar to a human brain. A typical model of such type has an input layer for the input values, several hidden layers, where data processing is taking place and an output layer. Corresponding model is termed as an Artificial Neural Network.

In Artificial Neural Networks (ANN), each neuron is a function, namely an activation function, that takes an input data, some sort of a signal, represented as numerical value, and produces an output. Predominantly, the architecture of multilayered network implies a unidirectional flow of data [2] from input layer neurons to output layer neurons via connections between them. Each connection has an assigned weight to it, represented as numerical value, which acts as a scaling factor, determining the significance of corresponding neuron output. A particular output of a neuron is computed from the weighted sum of its connected input neurons, bias term associated with the neuron and its activation function:

$$z_j = \sum_{i=1}^{n} (w_i j \cdot a_i) \quad (1)$$

$$Output_j = Activation(z_j + b_j) \quad (2)$$

where:

- $n$ is number of neurons in the previous layer
- $w_i j$ is weighted sum of previous layer's neurons connected to neuron $j$
- $a_i$ is the output of activation function of neuron $i$ in previous layer
- $b_j$ is a bias represented as constant numerical value.
- $Activation$ is the activation function of current neuron.

The main configuration of ANN that defines how well it performs, namely accuracy, on a given dataset is its weights values. Unlike shallow learning algorithms, where data and features must be carefully engineered, the architecture of Artificial Neural Networks allows it to train itself by feeding on some training data with patterns and adjusting weights according to some loss function. In addition, weights could also be set explicitly, by using some pre-trained data.

## 2.6 Deep learning in EEG-based emotion recognition

With advancements in deep learning techniques, usage of neural networks of various types as data classifiers became a common way to produce robust and efficient models. This is particularly notable in the field of EEG-based emotion recognition, allowing more comprehensive ways to interpret neurological data and subsequent feature extraction. The usage of neural network models such as Convolutional Neural Networks and Recurrent Neural Networks leveraged leveraged extraction of subtle patterns from EEG signals with commendable accuracy. Acquired results contribute to developments of more sophisticated response generative application and Human-Computer Interfaces.

### 2.6.1 Convolutional Neural Networks

The working principle of Convolutional Neural Network (CNN) is fundamentally the same as with typical neural network, in terms of being a network of interconnected neurons with weighted connections and biases. However, the main difference, and fundamental building block of any Convolutional Neural Networks is the presence and usage of convolution layers.

Convolution layers use filters, namely kernels, matrices of small size (usually 3-by-3 or 5-by-5), to perform a convolution operation upon spatial input data. Performing a convolution operation on a given two-dimensional input consists of sliding kernel across the input and performing element-wise multiplication, which, in the end, results in convolution of input data with kernel values (Figure 2).

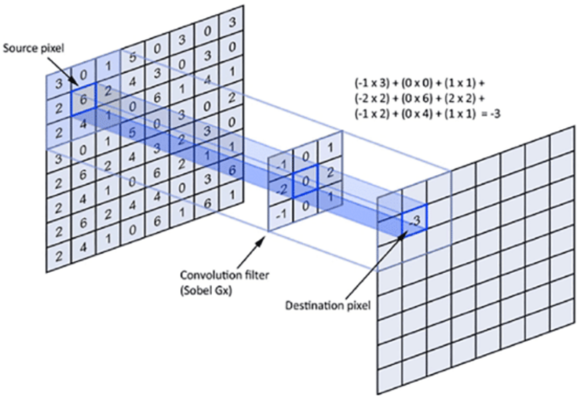


Figure 2. Visual example of convolution operation on spatial input with 3x3 kernel size [5]

It is to be noted that the resulting *width* and *height* of an output, compared to ones of input, are decreasing with each convolution operation. This is because the convolution operation is applied on the neighboring area, with the same size as kernel, of the target value, therefore excluding values of "vertical edges" and "horizontal edges" from calculations as those would be out of bounds. Expanding on this, convolution process can be controlled with parameters Stride $s$ and Padding $p$, those determine a step size for kernel movement and number of extra values, typically zeroes, to be added around an input data, respectively. The nature of convolution operations allows them to be used as robust method of feature extraction, which is the principle within the Convolution layers of CNN models

In comparison to traditional ANNs, the concept of a neuron (described in section 2.5), regarding connections, operations and output, differs significantly in CNN, where a neuron is organized into feature maps, produced by convolution with kernels. In addition, kernels act as weights [32] those values are adjusted during training, thus controlling the contribution level of neighbouring data at target input. The level of extracted features increases along the layers of CNN, with the first layers contributing to basic, low-level features (for example, edges, lines, contours in case of any image-based classification tasks) and highest-level features being extracted in the last convolution layer [51]. In this manner, the output of a particular feature map is computed from the weighted sum of its connected inputs, bias term and its activation function. With input data $I$ of shape $H \times W$ and kernel $K$ with of shape $k_H \times k_W$ a feature map $S$ at given position $(i, j)$ can be calculated as follows:

$$z = \sum_{m=0}^{k_H - 1} \sum_{n=0}^{k_W - 1} I(i + m, j + m) \cdot K(m, n) \tag{3}$$

$$Output(i, j) = Activation(z + b) \tag{4}$$

where

- $z$ is weighted sum, expressed for spatial dimensional data,
- $b$ is a bias term,
- $Activation$ is the activation function (typically, ReLU).

Apart from Convolutional layers, models built with CNN architecture utilizes Pooling layers for decreasing the number of parameters, hence computational power required, Fully Connected (FC) layers, that interconnect all neurons between the previous and next layers, therefore combining learned features, Dropout layers that are to prevent a model from overfitting by neglecting input of some fraction of neurons from the previous layer and Batch Normalization layers, that normalizes inputs with the aim of improving training efficiency (Figure 3).

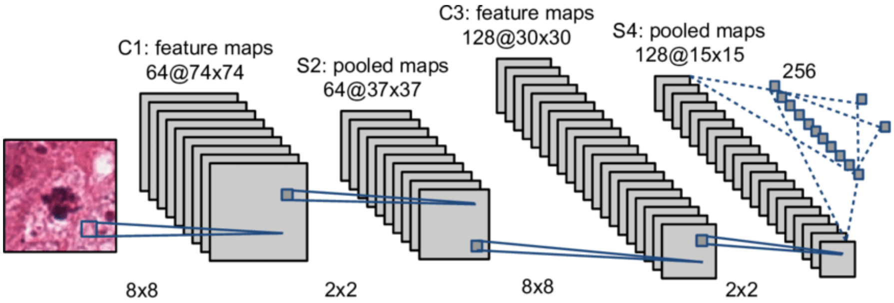


Figure 3. Architecture of CNN model layers. Adopted model was used to detect mitosis in breast cancer pathology images [59]

### CNN classifiers of EEG features

Despite CNNs being most commonly associated with image feature extractions [40], they are generally capable of processing any data that reflects spatial characteristics, including signal data. Pre-trained CNN model InceptionResnetV2 (modified with Global Average Pooling layer and extra Fully Connected layers) has been used [11] with four asymmetrical pairs of electrodes (AF1, F3, F4, F7, T7, AF2, F5, F8, and T8 from frontal and temporal lobes) as features. A Positive-Negative-Neutral and binary (Positive-Negative) classification was performed on arousal on public datasets DEAP [30] and SEED [66], respectively.

Several pre-trained CNN models with classic architectures were used in comparison to configure CNN [29] with SPWVD method [12] to transform raw signal data into image form. Classification was performed for 4 classes (fear, relax, happy, sad) with average validation accuracy of >92% among all models. Surprisingly, the confusion matrix reveals that the accuracy for classifying the ’happy’ emotion is 87%, which is notably lower than the average accuracy. Similar signal processing approach was also utilized in predictions for valence and arousal [23] on multiple datasets with different number of emotion classes, resulting in relatively good accuracy.

Although the accuracy of classic CNN and image pre-trained models might be sufficient, those struggle to perform on different datasets and classes, that implies engineering more dedicated and complex system for this task. Approaches utilizing modified CNN methods (DE-CNN [61], HCNN [35], MC-CNN [42], PCRNN [67], RA-CNN [24]) have demonstrated an improved performance.

### 2.6.2 Long Short-Term Memory networks

Long Short-Term Memory (LSTM) networks are a subset of Recurrent Neural Networks (RNNs). Unlike majority of deep learning networks encompassing unidirectional, forward flow of data, in Recurrent Neural Networks neurons are capable of sending feedback signals down the layers. This allows for RNNs to remember and utilize the information from previous inputs, making them very effective for tasks, where order and context of data is important, for example, EEG recordings analysis.

However, it has been discovered that RNNs could not train on long-sequence data, due to vanishing gradient problem [31]. Recurrent Neural Networks are trained using Backpropagation Through Time (BPTT), a variant of backpropagation, that computes gradient of the loss function with respect to the weights at each time step. With weights becoming smaller, gradients decrease exponentially, resulting in less significant weights updates until those become negligible, which is inevitable for long sequence data. In LSTM networks this was overcome by introducing a gate system that is to maintain the flow of data across Recurrent layers for gradient descend calculations. This has heavily contributed to various fields of RNNs, including speech recognition, text translations and voice assistant applications [63].

With this improvement, LSTM models have become a very popular choice for EEG signal feature extraction in emotion recognition. Following article [57] proposed a Bimodal LSTM model, consisting of one classification layer and two LSTM encoders for EEG feature extraction, using the SEED and DEAP datasets. Results demonstrated higher accuracy than similar papers [36], [65] proving LSTM to be a good option for classification. This paper [41] used LSTM model with attention mechanism on multichannel EEG data.

### 2.6.3 Hybrid Neural Networks

Hybrid Neural Networks (Hybrid NNs) is a sophisticated class of neural network architecture, that consists of layers from different types of neural networks (for example, convolution layers from CNN and recurrent layers from RNN), to leverage their unique strengths. This structure allows for Hybrid NN to outperform most traditional models (those that utilize a single type of their characteristic layer, for example, convolution layers in CNN) even with some degree of modification and configuration.

LSTM network approach introduced in [19] laid groundwork for combining LTSM with other network types, following with more dedicated hybrid network composition [17], [64]. Most of Hybrid NN applications in EEG emotion recognition, utilize CNN and LSTM types of layers [22, 21, 10].

# 3 Aims of the thesis

Preliminary to the field of EEG based automatic emotion recognition, this thesis aims to engineer a model for emotion recognition on electroencephalogram (EEG) data from a private dataset EMO2018. It was developed in Attention, Behavior and Cognition (ABC) lab in the Institute of Psychology and the Intelligent Computer Vision (iCV) lab in the Institute of Technology at the University of Tartu and described in [26]. EMO2018 consists multi modal data of reaction to emotion stimuli, namely EEG signals and facial expression recordings. The specific objectives of the researched are as follows:

- Review and evaluate existing approaches on EEG based emotion recognition;
- Design and develop thorough emotion recognition model that can be further used with EMO2018 dataset.

# 4 Methodology

The following chapter introduces selected methods and gives an overview of adopted resources. Section 4.1 overviews utilized EMO2018 dataset, process of data acquisition and structure of data records. Features extraction methods and reasoning are depicted in Section 4.2. Lastly, characteristics and architecture of used models as well as training process is presented in Section 4.3.

## 4.1 Dataset

Most of existing studies in the related field have predominantly relied on publicly available datasets for their experimentation. Among these, the most frequently utilized are described in the Table 2, highlighting eminent features.

This thesis, however, is focused on working with private, self-acquired dataset EMO2018 which consist of multi modal data of EEG signals and videos of facial expressions with key features for EEG listed in the table Table 3.

Table 2. Details of popular public datasets for EEG emotion recognition

| **Dataset** | **Labels** | **EEG characteristics** |
|---|---|---|
| DEAP [30] | Arousal, Valence, Dominance, Liking, Familiarity | 32-channel, 512 Hz |
| SEED [13] | Positive, Negative, Neutral | 62-channel, 200-1000 Hz |
| DREAMER [27] | Valence, Arousal, Dominance | 14-channel, 128Hz |
| MANHOB-HCI [54] | Arousal, Dominance, Valence, Predictability, Emotional keywords | 32-channel |

Table 3. Characteristics of EMO2018 EEG data features

| **Feature characteristics** | **Feature value** |
|---|---|
| Number of samples | 26432 |
| Number of channels | 80 |
| Frequency range | 0-104 Hz |
| Labels | Anger, disgust, fear, happiness, neutral, sadness, surprise |

### 4.1.1 Data acquisition

Data gathering was done in multiple sessions across several months and included a total number of 118 subjects. Each participant had a unique identifier assigned that enables a cross link between all experiments and modalities. The experiment consisted of four preliminaries:

1. **Mirroring** - participant is asked to mimic an emotion depicted using *Ekman's facial expressions* [44] with 7 emotions (anger, disgust, contempt, sad, joy, surprise, fear) and 6 repetitions, each describing same emotion but using different person facial expression. Resulting EEG signals from this condition were used for futher analysis.

2. **Expressing** - participant is asked to express an emotion depicted by words on the screen, namely *anger, disgust, happiness, neutral, sadness, surprise, fear* with 7 words and 6 repetitions.

3. **Over expressing** - participant is asked to express an emotion on their face as purely as possible. Emotions are depicted using 2 sets of images from *International Affective Picture System* (IAPS) [7] with 7 images in each set and 5 repetitions.

4. **Under expressing** - participant is asked to hide any emotional response on their face as best as they can, using IAPS images and same approach as in previous condition.

For each session the order of the four above mentioned preliminaries was randomized and in the middle of the session the order of last two conditions was switched for next participant's session. The reason for that is to prevent any possible misleading results that may occur from repetition of same sequence of trials over many times.

The experiment was conducted in an electrically shielded lab to minimize electromagnetic interference from equipment and surroundings and enhance signal-to-noise ratio, improving the data quality.

### 4.1.2 Data organisation

Data is organized in two sets consisting of EEG signal data for each participant (in *.bdf* format) and labeling data (in *.txt* format) containing timestamps and labels for when the particular emotion in any experiment was expressed. EEG data and annotation files linked through file name that is constructed with unique participant ID and some metadata. Each pair of corresponding signal and labeling files encompass the data of the whole experiment of a specific subject. Example EEG signal for one of the probes for one of the emotions can be seen on Figure 4.

Each line in a label file is associated with annotations of particular sub-experiment mentioned in 4.1.1 and can be linked with its corresponding condition. These annotations include experiment name, start time of expression, stop time of expression, emotion, emotion intensity and other meta data.

In this thesis, we concentrate on the analysis of data from Ekman's facial expressions experiment. The rationale behind using images of individuals displaying emotions is that such visual stimuli are likely to elicit stronger resonance with participants. Consequently, this approach is anticipated to yield EEG signals that more accurately reflect real-life emotional responses, thereby enhancing the reliability of emotion prediction.

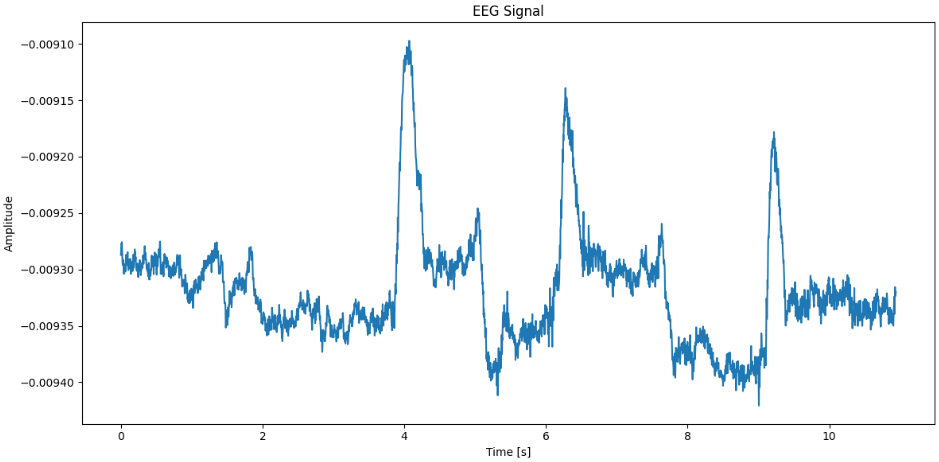
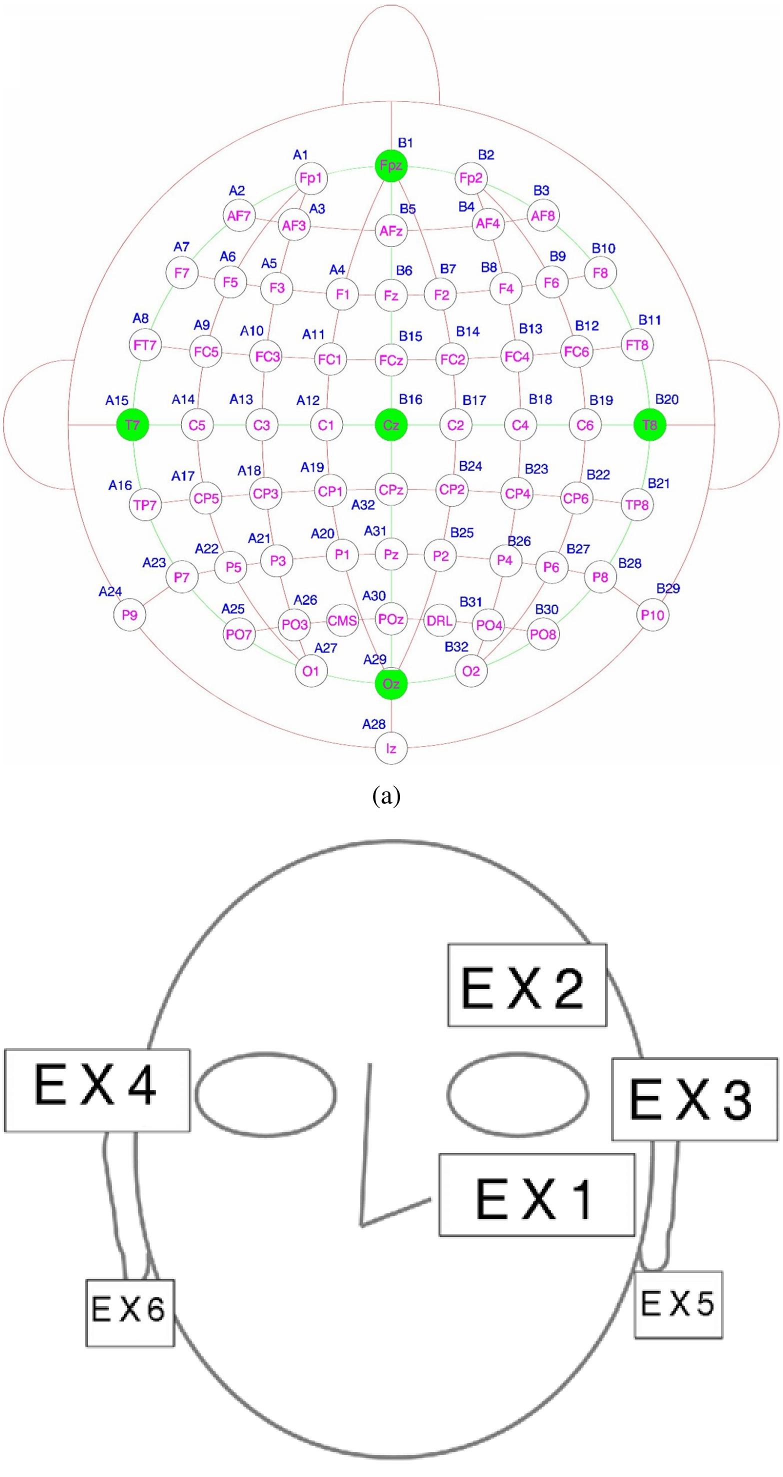


Figure 4. Example EEG signal from probe T7 for one the participants expressing happy emotion.

EMO2018 dataset International 10/20 electrode positioning encompasses signals spatial information totaling in 80 channels, consisting of 74 scalp-placed and 6 facial electrodes (Figure 5).



(a)

(b)

Figure 5. Electrodes placement for EEG signal reading: a) scalp electrodes with top head view, face is located the direction of the face is positioned upwards. b) facial electrodes.

## 4.2 Feature extraction

The feature extraction process is a critical step in analyzing EEG data for emotion prediction. Various approaches were employed to extract meaningful features from the raw EEG signals to enhance the accuracy of emotion prediction models.

Initially, five EEG signals were selected from four primary brain lobes (frontal, parietal, temporal, occipital) and the central region. This selection was based on the hypothesis that these regions play significant roles in emotional processing and response. To transform the time-domain EEG signals into the frequency domain, signal de-trending Fast Fourier Transform (FFT) was employed. This transformation aimed to capture the frequency components of the EEG signals, which are believed to carry essential information about brain activity associated with emotional states. The extracted frequency components were then used as input features for simple neural network models, including Support Vector Machines (SVM), Random Forest classifiers, and Fully Connected (FC) Neural Networks. Despite the theoretical advantages, these initial attempts did not yield satisfactory predictive performance. Example frequency spectrum generated by fft on one of the probes can be seen on Figure 6.

To explore alternative feature representations, raw EEG signals were subsequently fed directly into more complex neural network architectures, such as Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN). The rationale was that these models could potentially learn relevant features directly from the raw data without the need for explicit feature engineering. However, this approach also proved ineffective, with models exhibiting low accuracy in predicting emotional states.

Recognizing the need for a more comprehensive analysis, all 76 EEG signals were included in the feature extraction process. Both the initial FFT-based approach and the raw signal approach were reapplied to this extended signal set, yet neither method improved prediction accuracy substantially. Consequently, a more refined feature extraction process was implemented by applying FFT to all 76 EEG signals, and from each signal, the 10 most dominant frequencies within the range of 0-110 Hz were selected. This selection was based on the understanding that dominant frequencies could better capture the essential aspects of brain activity related to emotions. This process resulted in a comprehensive feature set comprising 760 inputs (76 signals × 10 frequencies per signal), which was then used as input for various advanced neural network architectures.

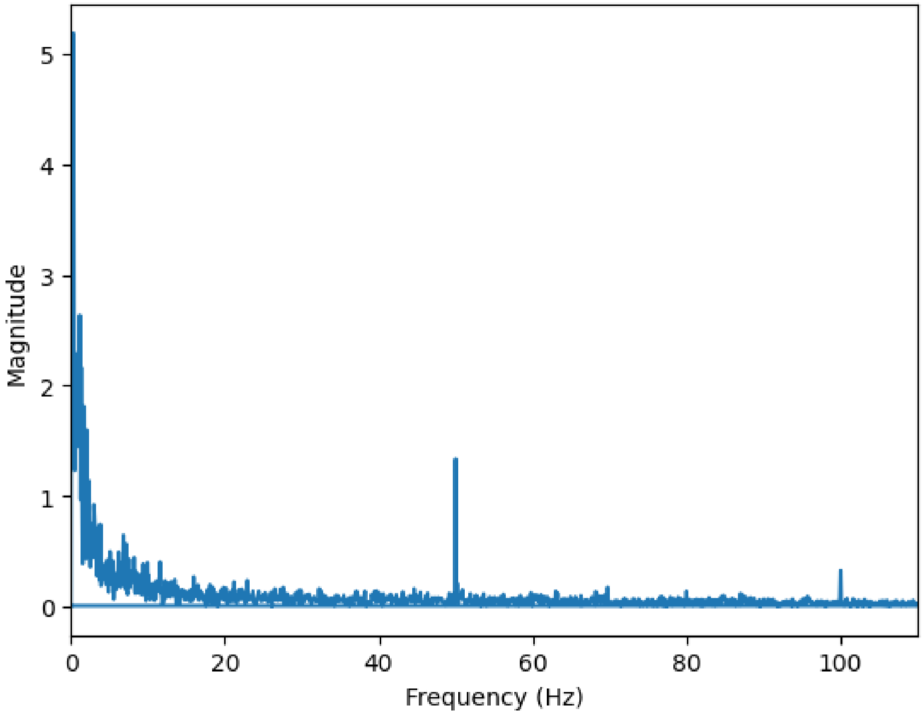


Figure 6. Example frequency spectrum of EEG signal from probe T7 for one the participants expressing happy emotion.

## 4.3 Model development

The model development process involved the construction and evaluation of three advanced neural network architectures to predict emotions from EEG data. These architectures were designed to leverage the extracted features from the 10 most dominant frequencies of each of the 76 EEG signals, resulting in a comprehensive feature set of 760 inputs. The following neural network models were employed:

### EmotionCNN

The first model, EmotionCNN, focused purely on convolutional layers to extract spatial features from the frequency domain. This model aimed to capture local patterns and hierarchies within the frequency features, leveraging the power of deep convolutional networks.

The architecture of EmotionCNN includes:

- Convolutional Layers: Three convolutional layers with batch normalization and max-pooling operations extract hierarchical spatial features from the input data.
- Fully Connected Layers: The output of the convolutional layers is flattened and

passed through fully connected layers with dropout for regularization, leading to the final emotion prediction.

### EmotionLSTM

The second model, EmotionLSTM, employed bidirectional LSTM layers to capture the temporal dependencies in the EEG data. This model was designed to utilize the sequential nature of the frequency features, providing a robust representation of the temporal patterns associated with different emotional states.

The architecture of EmotionLSTM is as follows:

- LSTM Layers: The input features are processed by bidirectional LSTM layers, which capture temporal dependencies from both directions.
- Batch Normalization and Dropout: The output of the LSTM layers undergoes batch normalization and dropout to improve generalization and prevent overfitting.
- Fully Connected Layer: The final output is passed through a fully connected layer to produce the emotion prediction.

### Advanced EmotionNet with CNN, Bidirectional LSTM, and Attention

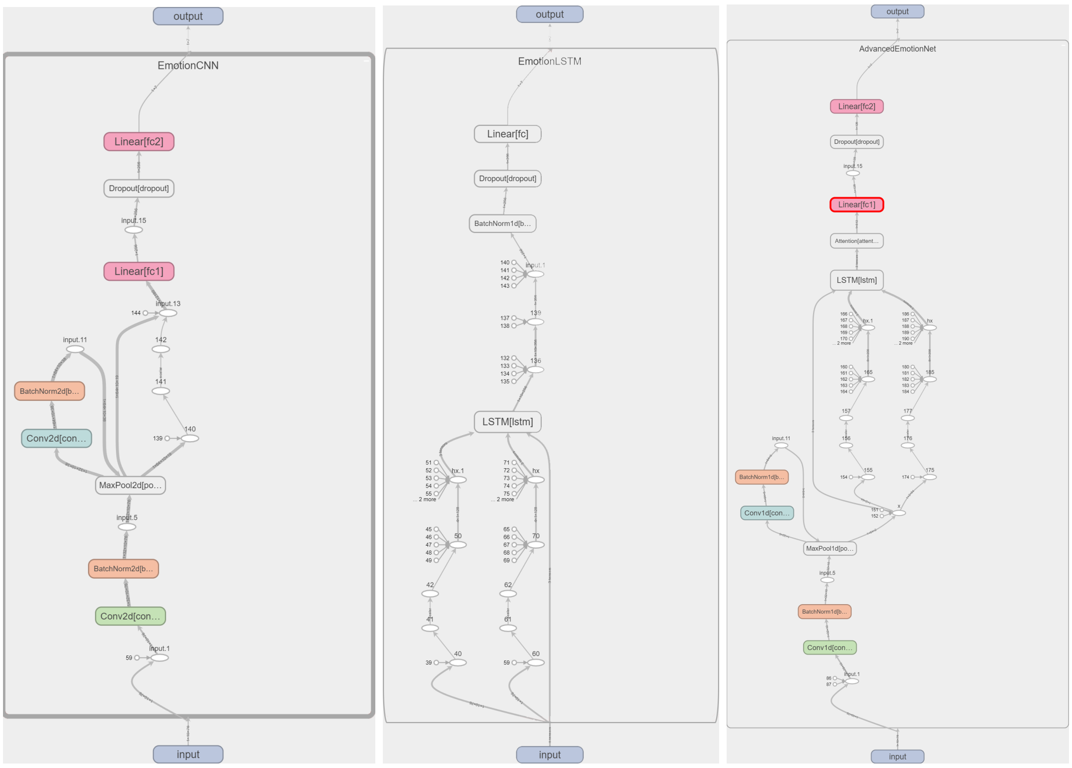
The third model, AdvancedEmotionNet, integrated convolutional layers, bidirectional LSTM layers, and an attention mechanism to capture both spatial and temporal dependencies in the EEG data. The convolutional layers extracted local features, which were then processed by the LSTM layers to capture temporal patterns. The attention mechanism was employed to focus on the most relevant parts of the sequence data, enhancing the model's ability to predict emotional states accurately.

The architecture of AdvancedEmotionNet is as follows:

- Convolutional Layers: The input features are passed through two convolutional layers with batch normalization and max-pooling operations to extract and downsample local features.
- LSTM Layers: The processed features are then fed into bidirectional LSTM layers, which capture the temporal dependencies from both forward and backward directions.

- Attention Layer: The output of the LSTM layers is processed by an attention mechanism, which helps the model focus on the most relevant parts of the hidden states.
- Fully Connected Layers: Finally, the output is passed through fully connected layers with dropout for regularization, leading to the final emotion prediction.

The graphical representations of all 3 networks can be seen on Figure 7.



(a) (b) (c)

Figure 7. Graphical representations of constructed networks: (a) CNN based network, (b) LSTM based network, (c) Advanced network with LSTM, CNN and attention mechanism. A full-size version of figures is available in Appendix II.

### 4.3.1 Training process

For all models, the training process involved standard practices such as data normalization, splitting the dataset into training and testing sets, and applying cross-validation to ensure robustness. The models were trained on a comodity graphics card using 'Adam' optimizer. Techniques such as early stopping and learning rate annealing were employed to prevent overfitting and ensure optimal convergence. Hyperparameters such as learning rate, batch size, number of layers, and number of hidden units were carefully tuned using grid search. All networks were trained for 150 epochs. The performance of the models was evaluated using standard metrics such as accuracy, precision, recall, and F1-score. Cross-validation was employed to ensure the robustness and generalizability of the models.

The iterative process of model training and refinement underscores the importance of combining advanced feature extraction techniques with sophisticated neural network architectures to achieve high-performance emotion prediction from EEG data. The successful application of these models highlights their potential in capturing the complex, multifaceted nature of emotional brain activity.

# 5 Results

The initial attempts at emotion prediction using EEG data were characterized by low accuracy rates, consistently falling below 20%. These early models, which included Support Vector Machines (SVM), Random Forest classifiers, and Fully Connected (FC) Neural Networks, were trained using features extracted from Fast Fourier Transform (FFT) of five selected EEG signals. Despite various efforts to optimize these models, including hyperparameter tuning and cross-validation, the performance remained unsatisfactory. Similarly, models trained on raw EEG signals using Long Short-Term Memory (LSTM) networks and Convolutional Neural Networks (CNN) also failed to produce significant improvements, with accuracy rates still below the 20%.

## 5.1 Advanced Model Training and Successful Results

A breakthrough was achieved with the implementation of more advanced neural network architectures and comprehensive feature extraction techniques. Three primary models were developed and tested: EmotionCNN, EmotionLSTM and AdvancedEmotionNet described in Section 4.3. These models were trained using the 10 most dominant frequencies from each of the 76 EEG signals, resulting in a feature set of 760 inputs. The following sections describe the training process and the successful results obtained.

The models were trained for 150 epochs with a batch size of 32. Both the LSTM-based models and the AdvancedEmotionNet employed two hidden layers, each with 256 hidden units. The training process involved standard optimization techniques such as Adam optimizer, early stopping, and learning rate annealing to ensure optimal convergence and prevent overfitting.

### 5.1.1 EmotionCNN Results

EmotionCNN, relying solely on convolutional layers, failed to improve previous unsuccessful results, falling bellow the accuracy of 20%. This highlighted that relying only on spacial features might not be enaugh to accurately capture all the necessary data. Training loss plot and confusion matrix for this network can be seen on Figure 8 and Figure 9, respectively.

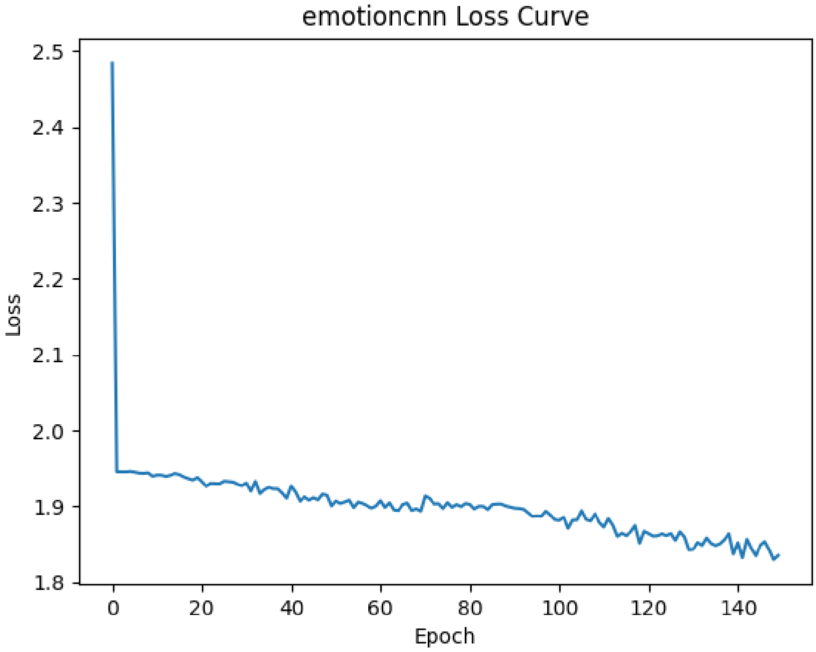


Figure 8. Loss plot for EmotionCNN network.

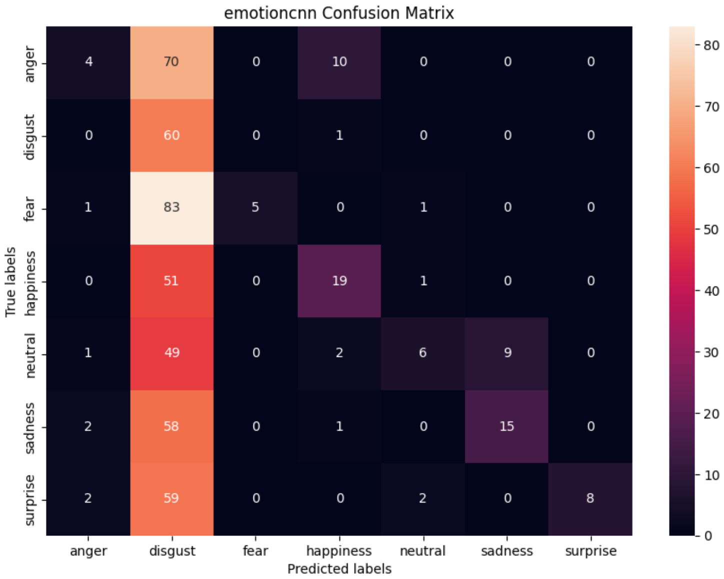


Figure 9. Confusion matrix for EmotionCNN network predictions.

### 5.1.2 EmotionLSTM Results

The EmotionLSTM model, utilizing bidirectional LSTM layers, effectively learned temporal dependencies in the EEG data, achieving training accuracy of 96.7% and validation accuracy of 87.4%. This model also performed greatly, showcasing the importance of capturing sequential patterns in the data. Training loss plot and confusion matrix for this network can be seen on Figure 10 and Figure 11, respectively.

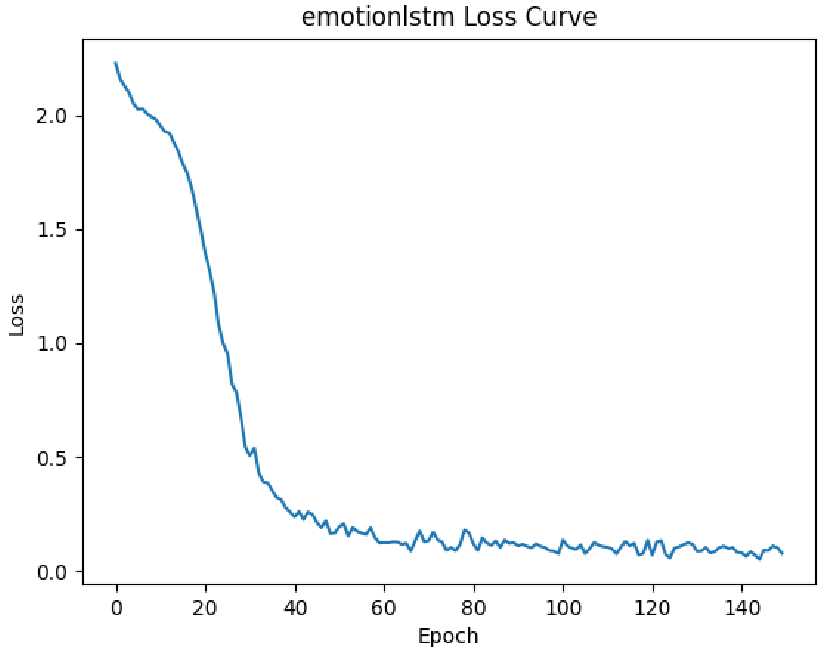


Figure 10. Loss plot for EmotionLSTM network.

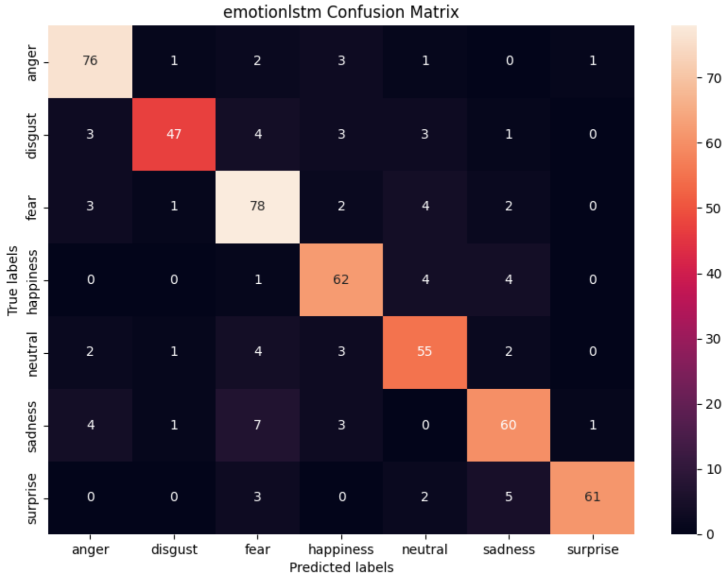


Figure 11. Confusion matrix for EmotionLSTM network predictions.

### 5.1.3 AdvancedEmotionNet Results

The AdvancedEmotionNet, which uses Hybrid Neural Network approach, integrates convolutional layers, bidirectional LSTM layers, and an attention mechanism, achieved the highest accuracy among the tested models scoring 98.5% on training and 88.6% on validation. This model successfully captured both spatial and temporal dependencies in the EEG data, resulting in robust emotion prediction. Training loss plot and confusion matrix for this network can be seen on Figure 12 and Figure 13 respectively.

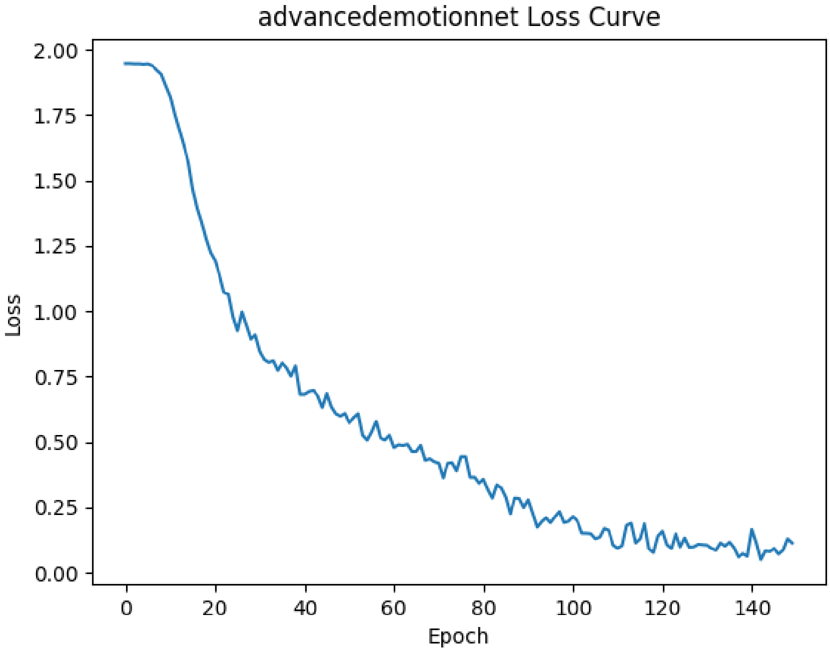


Figure 12. Loss plot for AdvancedEmotionNet network.

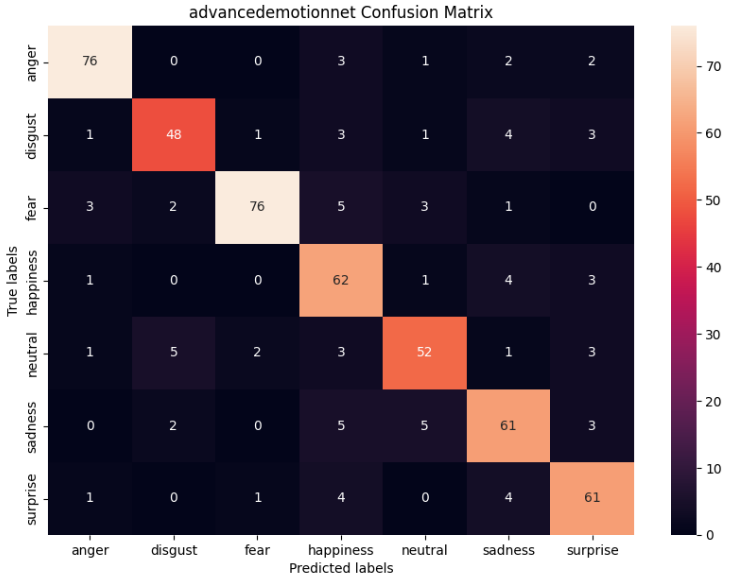


Figure 13. Confusion matrix for AdvancedEmotionNet network predictions.

# 6 Discussion

The results from the advanced neural network architectures demonstrate significant improvements over initial attempts. The AdvancedEmotionNet and EmotionLSTM models showed the highest accuracy, highlighting the importance of capturing both spatial and temporal dependencies in the EEG data for robust emotion prediction.

The AdvancedEmotionNet's integration of convolutional layers, bidirectional LSTM layers, and an attention mechanism proved particularly effective. The convolutional layers extracted essential local features, while the bidirectional LSTM layers captured the temporal patterns from both directions. The attention mechanism further enhanced the model by allowing it to focus on the most relevant parts of the input sequence. This combination resulted in the highest accuracy among the tested models, demonstrating its potential for practical applications in emotion recognition using EEG data.

The EmotionLSTM model also performed well, emphasizing the significance of temporal dependencies in EEG signals. The bidirectional LSTM layers effectively captured the sequential patterns, contributing to high prediction accuracy. This model's success further validates the importance of utilizing LSTM networks for time-series data like EEG.

On the other hand, the EmotionCNN model did not achieve the desired accuracy, indicating that spatial features alone are insufficient for accurate emotion prediction from EEG signals. This suggests that future work should focus on models that can capture both spatial and temporal features to enhance performance or further investigate other ways of feature extraction that can be suited for CNN networks.

These findings highlight the critical role of advanced feature extraction and neural network architectures in improving emotion prediction accuracy from EEG data. The successful application of these models underscores their potential in various domains, including mental health monitoring, human-computer interaction, and applied physiology research.

# 7 Conclusion

In this research we explored EEG signals for automatic emotion recognition, evaluating and improving existing approaches on related field on EMO2018 dataset, private collection of EEG recordings. We have compared the performance of Convolutional Neural Networks (CNNs), Long Short-Term Memory networks (LSTMs), with numerous configurations, and a Hybrid Neural Network (Hybrid NN) that integrated convolution layers, recurrent layers and attention mechanisms.

Our experiments demonstrated that Hybrid Neural Network, with above mentioned structure, had the best results with accuracy of 98.5% on training and 88.6% on validation, proving it to be competent of extracting both spatial and temporal information for feature extraction. On the other hand, the application of CNN model revealed that spatial feature extraction alone might not be sufficient.

Future work could explore further enhancements, such as combining additional physiological signals with EEG data, experimenting with different neural network architectures, and applying transfer learning techniques to leverage pre-trained models on larger, diverse datasets. Additionally, improving the interpretability of these models could provide deeper insights into the neural correlates of emotions, contributing to both scientific understanding and practical applications.

# Acknowledgements

I would like to express my deepest gratitude to my supervisors PhD. Egils Avots and MA. Liina Juuse for allowing me to work on this topic and providing great guidance and feedback along the way, and granting me exclusive opportunity to use the dataset. I am especially thankful to my closest people among friends and family members circles for support, motivation, not letting me discourage myself in the darkest hours and keeping my sanity.

# References

[1] A measure of arousal seeking tendency. *Environment and Behavior*, 5(3):315–333, 1973.

[2] Ajith Abraham. Artificial neural networks. *Handbook of measuring system design*, 2005.

[3] Amjed S Al-Fahoum and Ausilah A Al-Fraihat. Methods of eeg signal features extraction using linear analysis in frequency and time-frequency domains. *International Scholarly Research Notices*, 2014, 2014.

[4] Soraia M. Alarcão and Manuel J. Fonseca. Emotions recognition using eeg signals: A survey. *IEEE Transactions on Affective Computing*, 10(3):374–393, 2019.

[5] Ayad Almryad and Hakan Kutucu. Automatic identification for field butterflies by convolutional neural networks. *Engineering Science and Technology, an International Journal*, 23, 02 2020.

[6] Karim Ansari-Asl, Guillaume Chanel, and Thierry Pun. A channel selection method for eeg classification in emotion assessment based on synchronization likelihood. In *2007 15th European signal processing conference*, pages 1241–1245. IEEE, 2007.

[7] Margaret M. Bradley and Peter J. Lang. International affective picture system, 2017.

[8] Carlos Busso, Zhigang Deng, Serdar Yildirim, Murtaza Bulut, Chul Min Lee, Abe Kazemzadeh, Sungbok Lee, Ulrich Neumann, and Shrikanth Narayanan. Analysis of emotion recognition using facial expressions, speech and multimodal information. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 205–211, 2004.

[9] Rich Caruana and Alexandru Niculescu-Mizil. An empirical comparison of supervised learning algorithms. In *Proceedings of the 23rd International Conference on Machine Learning*, ICML '06, page 161–168, New York, NY, USA, 2006. Association for Computing Machinery.

[10] Bhuvaneshwari Chakravarthi, Sin-Chun Ng, MR Ezilarasan, and Man-Fai Leung. Eeg-based emotion recognition using hybrid cnn and lstm classification. *Frontiers in Computational Neuroscience*, 16:1019776, 2022.

[11] Yucel Cimtay and Erhan Ekmekcioglu. Investigating the use of pretrained convolutional neural network on cross-subject and cross-dataset eeg emotion recognition. *Sensors*, 20(7):2034, 2020.

[12] Edmundo Pereira de Souza Neto, Marc-Antoine Custaud, Jean Frutoso, Laurence Somody, Claude Gharib, and Jacques-Olivier Fortrat. Smoothed pseudo wigner–ville distribution as an alternative to fourier transform in rats. *Autonomic Neuroscience*, 87(2-3):258–267, 2001.

[13] Ruo-Nan Duan, Jia-Yi Zhu, and Bao-Liang Lu. Differential entropy feature for eeg-based emotion classification. In *2013 6th International IEEE/EMBS Conference on Neural Engineering (NER)*, pages 81–84, 2013.

[14] Pierre Duhamel and Martin Vetterli. Fast fourier transforms: a tutorial review and a state of the art. *Signal processing*, 19(4):259–299, 1990.

[15] Paul Ekman and Harriet Oster. Facial expressions of emotion. *Annual review of psychology*, 30(1):527–554, 1979.

[16] Christos A Frantzidis, Charalampos Bratsas, Christos L Papadelis, Evdokimos Konstantinidis, Costas Pappas, and Panagiotis D Bamidis. Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE transactions on Information Technology in Biomedicine*, 14(3):589–597, 2010.

[17] Alex Graves and Navdeep Jaitly. Towards end-to-end speech recognition with recurrent neural networks. In *International conference on machine learning*, pages 1764–1772. PMLR, 2014.

[18] Dirk Hagemann, Shari R. Waldstein, and Julian F. Thayer. Central and autonomic nervous system integration in emotion. *Brain and Cognition*, 52(1):79–87, 2003. Affective Neuroscience.

[19] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural computation*, 9(8):1735–1780, 1997.

[20] Richard W Homan, John Herman, and Phillip Purdy. Cerebral location of international 10–20 system electrode placement. *Electroencephalography and clinical neurophysiology*, 66(4):376–382, 1987.

[21] Mohammad Saleh Khajeh Hosseini, Seyed Mohammad Firoozabadi, Kambiz Badie, and Parviz Azadfallah. Personality-based emotion recognition using eeg signals with a cnn-lstm network. *Brain Sciences*, 13(6):947, 2023.

[22] Zhentao Huang, Yahong Ma, Rongrong Wang, Weisu Li, and Yongsheng Dai. A model for eeg-based emotion recognition: Cnn-bi-lstm with attention mechanism. *Electronics*, 12(14):3188, 2023.

[23] Md Rabiul Islam and Mohiuddin Ahmad. Virtual image from eeg to recognize appropriate emotion using convolutional neural network. In *2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT)*, pages 1–4. IEEE, 2019.

[24] Md Rabiul Islam, Mohammad Ali Moni, Md Milon Islam, Md Rashed-Al-Mahfuz, Md Saiful Islam, Md Kamrul Hasan, Md Sabir Hossain, Mohiuddin Ahmad, Shahadat Uddin, Akm Azad, et al. Emotion recognition from eeg signal focusing on deep learning and shallow learning techniques. *IEEE Access*, 9:94601–94624, 2021.

[25] Robert Jenke, Angelika Peer, and Martin Buss. Feature extraction and selection for emotion recognition from eeg. *IEEE Transactions on Affective Computing*, 5(3):327–339, 2014.

[26] Liina Juuse, Kairi Kreegipuu, Nele Põldver, Annika Kask, Tiit Mogom, Gholamreza Anbarjafari, and Jüri Allik. Processing emotions from faces and words measured by event-related brain potentials. *Cognition and Emotion*, 37(5):959–972, 2023. PMID: 37338015.

[27] Stamos Katsigiannis and Naeem Ramzan. Dreamer: A database for emotion recognition through eeg and ecg signals from wireless low-cost off-the-shelf devices. *IEEE journal of biomedical and health informatics*, 22(1):98–107, 2017.

[28] Andreas Keil, Margaret M Bradley, Olaf Hauk, Brigitte Rockstroh, Thomas Elbert, and Peter J Lang. Large-scale neural correlates of affective picture processing. *Psychophysiology*, 39(5):641–649, 2002.

[29] Smith K Khare and Varun Bajaj. Time–frequency representation and convolutional neural network-based emotion recognition. *IEEE transactions on neural networks and learning systems*, 32(7):2901–2909, 2020.

[30] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1):18–31, 2011.

[31] John F. Kolen and Stefan C. Kremer. *Gradient Flow in Recurrent Nets: The Difficulty of Learning LongTerm Dependencies*, pages 237–243. 2001.

[32] Scott Krig. *Computer Vision Metrics*. Springer International Publishing, 2016.

[33] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[34] Nilesh Kulkarni and Vinayak Bairagi. Electroencephalogram based diagnosis of alzheimer disease. 01 2015.

[35] Jinpeng Li, Zhaoxiang Zhang, and Huiguang He. Hierarchical convolutional neural networks for eeg-based emotion recognition. *Cognitive Computation*, 10:368–380, 2018.

[36] Wei Liu, Wei-Long Zheng, and Bao-Liang Lu. Emotion recognition using multimodal deep learning. In Akira Hirose, Seiichi Ozawa, Kenji Doya, Kazushi Ikeda, Minho Lee, and Derong Liu, editors, *Neural Information Processing*, pages 521–529, Cham, 2016. Springer International Publishing.

[37] Xucheng Liu, Ting Li, Cong Tang, Tao Xu, Peng Chen, Anastasios Bezerianos, and Hongtao Wang. Emotion recognition and dynamic functional connectivity analysis based on eeg. *IEEE Access*, 7:143293–143302, 2019.

[38] Yisi Liu and Olga Sourina. Real-time fractal-based valence level recognition from eeg. In *Transactions on computational science XVIII: special issue on Cyberworlds*, pages 101–120. Springer, 2013.

[39] Fabien Lotte and Marco Congedo. *EEG Feature Extraction*, chapter 7, pages 127–143. John Wiley & Sons, Ltd, 2016.

[40] D. Lu and Q. Weng. A survey of image classification methods and techniques for improving classification performance. *International Journal of Remote Sensing*, 28(5):823–870, 2007.

[41] Jiaxin Ma, Hao Tang, Wei-Long Zheng, and Bao-Liang Lu. Emotion recognition using multimodal residual lstm network. In *Proceedings of the 27th ACM International Conference on Multimedia*, MM '19, page 176–183, New York, NY, USA, 2019. Association for Computing Machinery.

[42] Daksh Maheshwari, Samit Kumar Ghosh, RK Tripathy, Manish Sharma, and U Rajendra Acharya. Automated accurate emotion recognition system using rhythm-specific deep convolutional neural network technique with multi-channel eeg signals. *Computers in Biology and Medicine*, 134:104428, 2021.

[43] Stacy Marsella and Jonathan Gratch. Computationally modeling human emotion. *Commun. ACM*, 57(12):56–67, nov 2014.

[44] David Ricky Matsumoto. *Japanese and Caucasian facial expressions of emotion (JACFEE)*. University of California, 1988.

[45] Thomas Mensink, Jakob Verbeek, Florent Perronnin, and Gabriela Csurka. Metric learning for large scale image classification: Generalizing to new classes at near-zero cost. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part II 12*, pages 488–501. Springer, 2012.

[46] Zeynab Mohammadi, Javad Frounchi, and Mahmood Amiri. Wavelet-based emotion recognition system using eeg signal. *Neural Computing and Applications*, 28:1985–1990, 2017.

[47] Hong-Wei Ng, Viet Dung Nguyen, Vassilios Vonikakis, and Stefan Winkler. Deep learning for emotion recognition on small datasets using transfer learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, pages 443–449, 2015.

[48] W. Gerrod Parrott, editor. *Emotions in Social Psychology: Key Readings*. Psychology Press, 2000.

[49] Panagiotis C. Petrantonakis and Leontios J. Hadjileontiadis. Emotion recognition from eeg using higher order crossings. *IEEE Transactions on Information Technology in Biomedicine*, 14(2):186–197, 2010.

[50] Robert Plutchik. The nature of emotions: Human emotions have deep evolutionary roots, a fact that may explain their complexity and provide tools for clinical practice. *American Scientist*, 89(4):344–350, 2001.

[51] Jimmy SJ. Ren, Wei Wang, Jiawei Wang, and Stephen Liao. An unsupervised feature learning approach to improve automatic incident detection. In *2012 15th International IEEE Conference on Intelligent Transportation Systems*, pages 172–177, 2012.

[52] Klaus R. Scherer. What are emotions? and how can they be measured? *Social Science Information*, 44(4):695–729, 2005.

[53] Jan N Schneider, Magdalena Matyjek, Anne Weigand, Isabel Dziobek, and Timothy R Brick. Subjective and objective difficulty of emotional facial expression perception from dynamic stimuli. *Plos one*, 17(6):e0269156, 2022.

[54] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, 2012.

[55] Olga Sourina and Yisi Liu. A fractal-based algorithm of emotion recognition from eeg using arousal-valence model. In *Biosignals*, pages 209–214, 2011.

[56] Duraisamy Sundararajan. *The discrete Fourier transform: theory, algorithms and applications*. World Scientific, 2001.

[57] Hao Tang, Wei Liu, Wei-Long Zheng, and Bao-Liang Lu. Multimodal emotion recognition using deep neural networks. pages 811–819, 10 2017.

[58] Michal Teplan et al. Fundamentals of eeg measurement. *Measurement science review*, 2(2):1–11, 2002.

[59] Haibo Wang, Angel Cruz-Roa, Ajay Basavanhally, Hannah Gilmore, Natalie Shih, Mike Feldman, John Tomaszewski, Fabio González, and Anant Madabhushi. Mitosis detection in breast cancer pathology images by combining handcrafted and convolutional neural network features. *Journal of Medical Imaging*, 1:1–8, 12 2014.

[60] Xiao-Wei Wang, Dan Nie, and Bao-Liang Lu. Eeg-based emotion recognition using frequency domain features and support vector machines. In *Neural Information Processing: 18th International Conference, ICONIP 2011, Shanghai, China, November 13-17, 2011, Proceedings, Part I 18*, pages 734–743. Springer, 2011.

[61] Yingdong Wang, Qingfeng Wu, Chen Wang, and Qunsheng Ruan. De-cnn: An improved identity recognition algorithm based on the emotional electroencephalography. *Computational & Mathematical Methods in Medicine*, 2020.

[62] David Watson, Lee Anna Clark, and Auke Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063, 1988.

[63] Yonghui Wu, Mike Schuster, Zhifeng Chen, Quoc V. Le, Mohammad Norouzi, Wolfgang Macherey, Maxim Krikun, Yuan Cao, Qin Gao, Klaus Macherey, Jeff Klingner, Apurva Shah, Melvin Johnson, Xiaobing Liu, Łukasz Kaiser, Stephan Gouws, Yoshikiyo Kato, Taku Kudo, Hideto Kazawa, Keith Stevens, George Kurian, Nishant Patil, Wei Wang, Cliff Young, Jason Smith, Jason Riesa, Alex Rudnick, Oriol Vinyals, Greg Corrado, Macduff Hughes, and Jeffrey Dean. Google's neural machine translation system: Bridging the gap between human and machine translation, 2016.

[64] Kelvin Xu, Jimmy Ba, Ryan Kiros, Kyunghyun Cho, Aaron Courville, Ruslan Salakhudinov, Rich Zemel, and Yoshua Bengio. Show, attend and tell: Neural

image caption generation with visual attention. In *International conference on machine learning*, pages 2048–2057. PMLR, 2015.

[65] Zhong Yin, Mengyuan Zhao, Yongxiong Wang, Jingdong Yang, and Jianhua Zhang. Recognition of emotions using multimodal physiological signals and an ensemble deep learning model. *Computer Methods and Programs in Biomedicine*, 140:93–110, 2017.

[66] Wei-Long Zheng and Bao-Liang Lu. Investigating critical frequency bands and channels for eeg-based emotion recognition with deep neural networks. *IEEE Transactions on autonomous mental development*, 7(3):162–175, 2015.

[67] Siyu Zhu, Jin Qi, Jie Hu, and Sheng Hao. A new approach for product evaluation based on integration of eeg and eye-tracking. *Advanced Engineering Informatics*, 52:101601, 2022.

# Appendix

## I. Figures

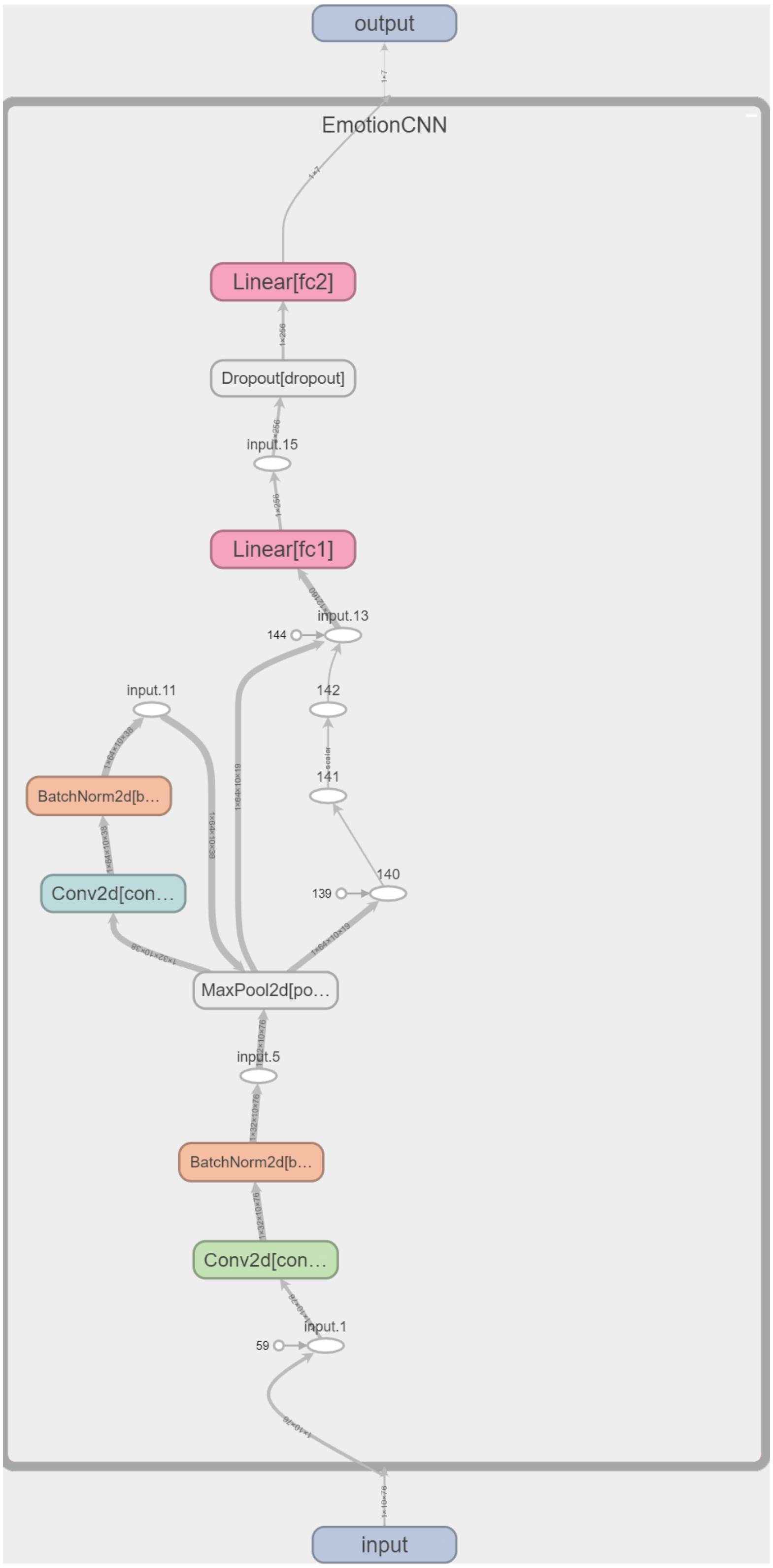


Figure 14. Graphical visualization of EmotionCNN network structure in full-size

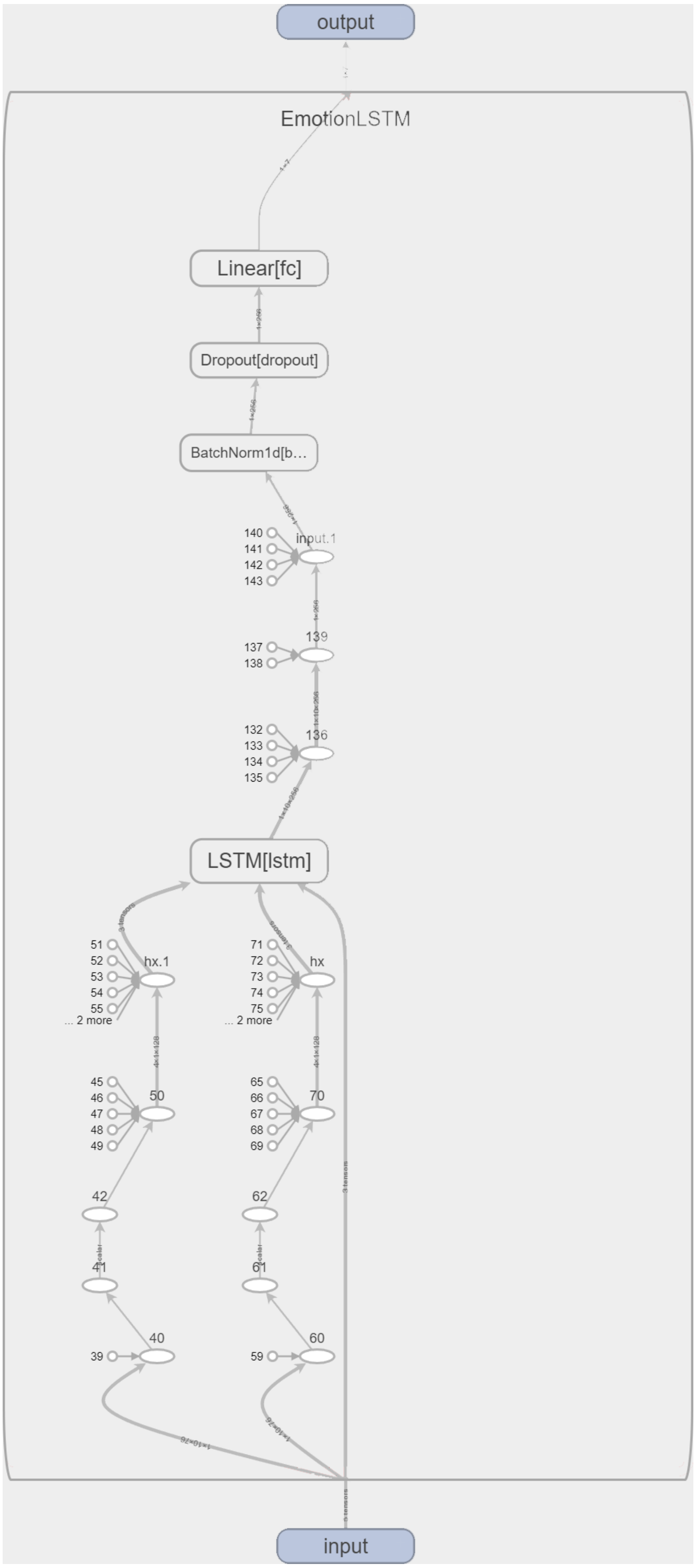


Figure 15. Graphical visualization of EmotionLSTM network structure in full-size

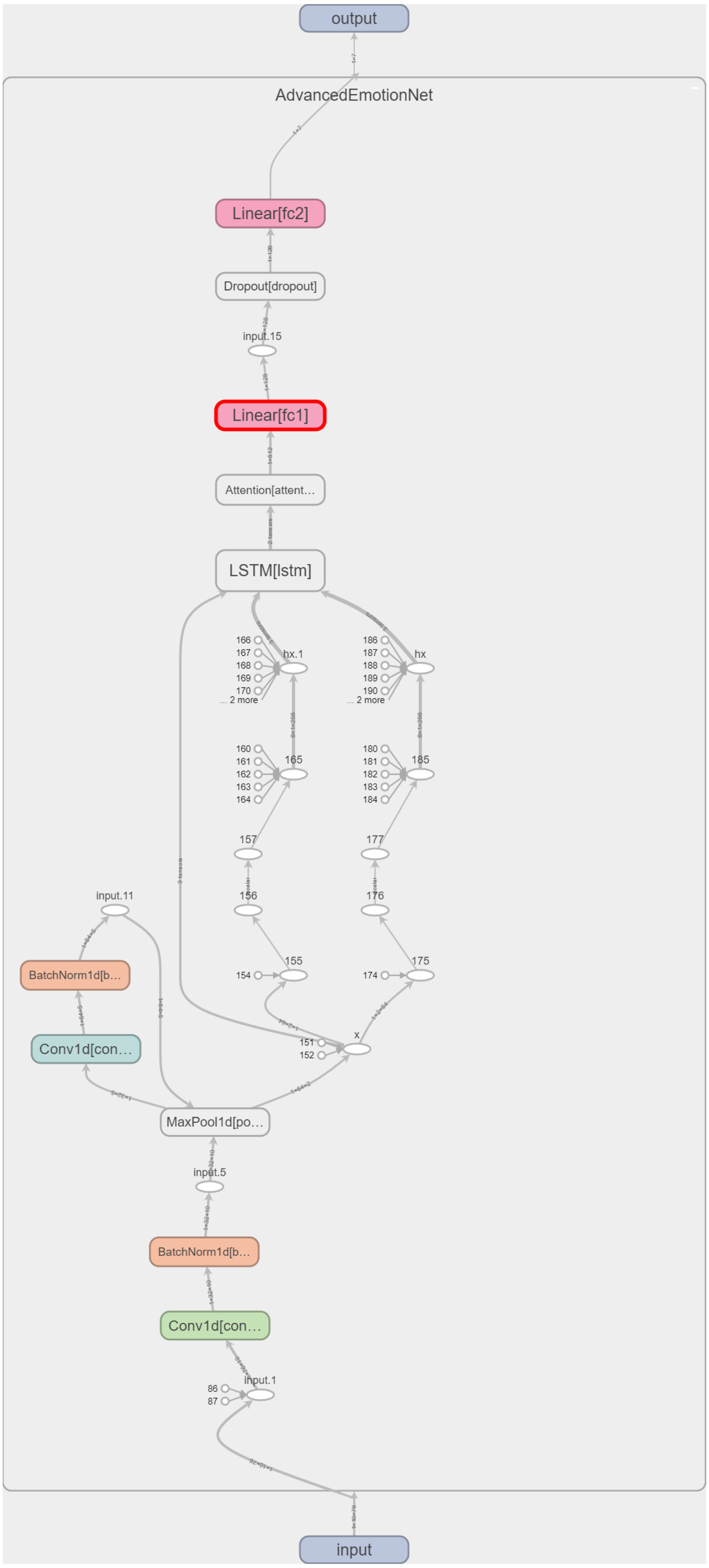


Figure 16. Graphical visualization of AdvancedEmotionNet network structure in full-size

# Licence

## Non-exclusive licence to reproduce thesis and make thesis public

I, **Aleksandrs Rebriks**,

(author's name)

1. herewith grant the University of Tartu a free permit (non-exclusive licence) to

   reproduce, for the purpose of preservation, including for adding to the DSpace digital archives until the expiry of the term of copyright,

   **Emotion Recognition using EEG signal data from EMO2018 Dataset** ,

   (title of thesis)

   supervised by Egils Avots and Liina Juuse.

   (supervisor's name)

2. I grant the University of Tartu a permit to make the work specified in p. 1 available to the public via the web environment of the University of Tartu, including via the DSpace digital archives, under the Creative Commons licence CC BY NC ND 4.0, which allows, by giving appropriate credit to the author, to reproduce, distribute the work and communicate it to the public, and prohibits the creation of derivative works and any commercial use of the work until the expiry of the term of copyright.

3. I am aware of the fact that the author retains the rights specified in p. 1 and 2.

4. I certify that granting the non-exclusive licence does not infringe other persons' intellectual property rights or rights arising from the personal data protection legislation.

Aleksandrs Rebriks

***03/06/2024***