Jerzy Rudolf Tchórzewski[1]
ORCID: 0000-0003-2198-7185

Jakub Rosiński[2]
ORCID: 0009-0003-8686-2524

[1] University of Siedlce
Faculty of Exact and Natural Sciences
Institute of Computer Science
ul. 3 Maja 54, 08-110 Siedlce, Poland

[2] Computer Science Graduate
(Engineer of Computer Science)
University of Siedlce
Faculty of Exact and Natural Sciences
Institute of Computer Science
ul. 3 Maja 54, 08-110 Siedlce, Poland

[1]jerzy.tchorzewski@uws.edu.pl,[2]jr86210@stud.uws.edu.pl

Searching form the Model of Human Facial Expressions and its Implementation in MATLAB and Simulink Environments

DOI: 10.34739/si.2025.32.07

Abstract. The publication contains selected research results on the creation of a neural model of human facial expressions along with its implementation in the MATLAB and Simulink environments using the library called Deep Learning Toolbox. The model was generated as an object to recognize selected human faces and emotional states based on visual data recorded on the human face in real time. The study was placed against the background of available literature on the analysis of facial expressions and emotion classification methods. In addition to the concept of the original solution, the assumptions of the research experiment were given, a method for measuring facial expressions was selected for the experimental conditions, and a set of data was developed for training a neural model of the facial expression system with their preparation for ANN learning. Ultimately, various artificial neural networks were trained to model the facial expression system and sensitivity and comparative tests were performed in Simulink to, among others: assess the quality of the model in relation to real data. Very high results of ANN training of the facial expression system model were obtained (MSE on the order of 10^{-14}, R close to 1), as well as relatively high model quality relative the facial expression system, measured, among others, average relative error, the value of which was several percent. However, in terms of prediction effectiveness, because of the obtained results, which are not very high for the assumed high measurement accuracies, the research that has been started is continued.

Keywords: Artificial Neural Networks, Cluster Analysis, Facial expression recognition, Image analysis, MATLAB and Simulink

1 Introduction

In the current times of intensive development of artificial intelligence methods and humanoid robotics, an extremely current challenge is, on the one hand, the search for systems that can imitate human thinking, and on the other hand, human body movements, including body facial expressions. This publication does not address the issues related to the generation of signals and the field related to them in the field of artificial thinking, i.e. the cognitive system, but does address the issue of human facial movements called facial expressions.

The solution to this issue is of interest to scientists from all over the world who are looking for opportunities to create artificial systems capable of interacting with humans in a natural and human-like way, which involves, among others, with visual, sensory and vocal contact. In particular, there is growing interest in solutions that enable machines to recognize, interpret and respond appropriately to human emotions, gestures and unusual behaviors. One of the most important channels for transmitting emotions are gestures expressed, among others: facial expressions, which are a basic element of non-verbal communication and play a key role in human social processes. Facial expressions provide information about the interlocutor's emotional state, intentions and reactions to specific situations. For this reason, precise analysis of facial expressions becomes not only a fascinating scientific challenge, but also an extremely useful tool in many areas of life of a group of people, society, and even the whole of humanity.

Without exhausting the wealth of applications, it is worth adding that the integration of facial models in social robotics allows for the creation of more empathetic and responsive machines that better understand and reproduce human emotions [21-22, 32].

In the context of intelligent engineering, a particular challenge remains developing an effective model of a human facial expression system that can analyze facial images, extract important features, and then classify the expressed emotions and interpret their internal states, such as sadness, joy, or interruption. This model should be accurate and optimized for performance, so it can be an be used in real time. An equally important aspect is ensuring the ability to test and validate the model in a controlled environment, allowing for analysis of its performance in various scenarios [5, 12, 31].

The article presents, among other things, selected results of a critical review of the subject literature, the assumptions and design of the experiment on measuring and analyzing facial expressions during conversation, the method of preparing and conducting the analysis and processing of measurement data, the choice of neural network learning methods and the implementation of a model of human facial expressions, as well as the course and selected simulation experiments of the human facial expression model and the results obtained along with their discussion and conclusions.

The research was carried out on a Lenovo Legion 5-15 Ryzen 5/16GB/512 RTX2060 120Hz computer with a Ryzen 5 3600 processor and RAM 16GB 3200MHz CL16. An HD camera with a resolution of 1,280 x 720 pixels (9.92 MP) was used to measure facial expressions.

2 Research on facial expressions

Facial expressions are one of the most natural and inherent ways to expressions and nonverbal communication in interpersonal relationships. From a biological perspective, they are a complex process involving, among other things, facial muscle movements that are the result of neural impulses generated in response to internal mental states or external stimuli.

In the context of psychology, facial expressions are not only a form of communication but also serve an adaptive function – they facilitate the formation of social bonds, help regulate interpersonal relationships, and enable quick responses to social situations [21]. According to Paul Ekman's research, certain facial expressions are universal and recognized in people across cultures, regardless of ethnicity or upbringing [6-7].

From the perspective of cognitive systems, facial expressions constitute a valuable source of knowledge for analysis in human emotion recognition systems. Artificial neural networks, when properly trained, can recognize and interpret changes in facial expressions with high accuracy [10-11, 15-17, 29], and various tools and models are used to analyze facial expressions. One of the most popular methods is the Facial Action Coding System (FACS), which breaks down facial expressions into basic movement units (Action Units - AUs) and analyzes them in detail [6-7].

In AI systems, facial expressions are analyzed using computer vision techniques, utilizing facial landmarks. This analysis includes, among other things, distances between landmarks, tilt angles, deformations and changes over time, skin tone, etc. [27]. From the perspective of biomedical engineering, cognitive science, and artificial intelligence, facial muscle movement analysis is based on detecting changes in the geometric arrangement of facial points, as well as using techniques for tracking skin texture deformations [2,14,24,37]. The most well-known classification was proposed by Ekman, who identified six basic emotions: joy, sadness, fear, anger, surprise, and disgust [6-7].

Each of these emotions has a set of characteristic facial features, such as raising the corners of the mouth and furrowing the eyelids in joy, or widening the eyes and opening the mouth in surprise. Modern emotion recognition systems rely on the detection of these patterns using machine learning, neural networks, and analysis of geometric and textural features [1,3-4,13,18-19,25-27].

3 Data acquisition methods for facial expression analysis

Effective facial expression modeling requires high-quality input data. The process of acquiring this data plays a key role in the subsequent stages of emotion analysis and classification. Modern technologies offer numerous methods for acquiring facial expression data, ranging from simple video cameras, through advanced 3D imaging systems, to biomechanical measurement sensors [4,18]. Various types of cameras are used, such as the following:

- **Video cameras**

  2D cameras allow for the analysis of geometric facial features, such as the distance between the eyes, the width of a smile, the movement of the eyebrows, and changes in the position of the mouth. Images captured by the cameras can then be processed using feature detection algorithms such as Active Appearance Models (AAM) or Constrained Local Models (CLM) [24].

- **3D Cameras**

  3D cameras capture spatial data, allowing for the measurement of depth, facial contours, and actual skin surface deformations. The most commonly used technologies include laser scanners, stereo cameras, and devices based on light structure, such as Microsoft Kinect and Intel RealSense [16].

- **Optical systems (Motion Capture)**

  These involve the use of cameras and special markers placed on the subject's face, enabling highly accurate mapping of the movement trajectories of characteristic points. Infrared light is often used, which increases detection precision and minimizes environmental interference [24].

4 Intelligent Methods in Facial Expression Modeling

Machine learning and artificial intelligence methods are used to learn facial expressions [1, 4, 12, 14, 16, 29, 32]. These methods leverage their capabilities for knowledge generalization, automatic pattern detection, and learning from large datasets [12]. Depending on the available data, the required accuracy, and the available computational resources, various machine learning methods are used, sometimes in combination with various artificial intelligence methods, such as [3, 14, 32]:

- **Artificial Neural Networks**
  Convolutional Artificial Neural Networks (CNNs) are most commonly used for facial expression recognition, as they perform similarly to other image recognition techniques, as they can automatically extract relevant features from facial images and assign them to appropriate emotional categories. Recurrent Artificial Neural Networks (RNNs), including LSTM-type ANNs, excel at analyzing temporal sequences, making them ideal for video processing [23]. Hybrid ANNs are also frequently used, combining various types of neural networks into ensembles, such as a CNN + LSTM combination, which allows for the simultaneous analysis of spatial features (face shape and texture) and temporal features (changes in expression over time). Such solutions are used in emotional diagnostic systems, humanoid robotics, and adaptive interfaces, among others [32].

- **Classification algorithms**
  More classical approaches use classification algorithms such as [14,33]:
  - SVM – effective for linearly separable problems, often used for binary classification of emotions (e.g. positive vs. negative),
  - k-NN a simple algorithm that works well for small data sets,
  - decision trees and random forests – allow for intuitive interpretation of results and show greater resistance to overfitting, provided that the parameters are properly selected.

- **Principle Component Analysis (PCA)**
  PCA is a dimensionality reduction technique used to simplify data analysis, especially in cases where the number of variables is very high. PCA is used to extract key facial features that enable effective facial expression recognition and classification [32-33].

- **Unsupervised learning algorithms**
  These types of algorithms enable the detection of hidden patterns in data without the need for supervision. One example is clustering, which allows for the grouping of similar facial expressions and emotions into a single set, which can aid in further data processing and analysis [14,30,33].

5 Facial Expression Recording and Analysis during Online Conversation

An experiment was proposed to record and analyze facial expressions during a simulated online conversation in which the participant expressed three basic emotions: anger, happiness, and sadness [6-7]. The approach was aimed at assessing the effectiveness of facial expression recognition systems in detecting and classifying emotions based on data collected in real time. The MediaPipe library was used to achieve this goal, enabling the detection and tracking of facial points with high accuracy.

5.1 Setup and Recording Procedure

The experiment was conducted in a home environment, near a desk, where the participant's face was positioned stably within the camera frame. This was important because changes in the frame could disrupt the software's facial detection.

The following components were used to conduct the experiment:

- **Laptop Camera** – Cameras built into laptops, while often less advanced than external recording devices, can be sufficient to capture facial expression details in HD resolution;
- **Frontal Illumination** — To ensure even illumination of the face, an LED light was used, positioned directly behind the camera, in line with the participant's face. This arrangement was intended to minimize shadows that could interfere with the accuracy of facial point detection;
- **Background and environment control** — this was another crucial element of the experiment. The background was designed to be uniform, free from unnecessary distractions.

Participant Positioning

The participant sat in a chair approximately 40–50 cm from the laptop camera, which allowed for full-face capture and provided adequate workspace for the facial expression recognition system. The camera was positioned at the participant's eye level, ensuring image stability and an appropriate perspective for the facial expression analysis algorithm.

Recording Procedure The experimental recordings consisted of three separate video series, each 3 minutes long. Each series depicted a different emotion, and the participant consciously controlled their facial expressions according to specific instructions [19].

Description of emotions presented in the recordings

In the experiment, the participant was asked to express emotions such as anger, happiness, and sadness [22]. Each emotion was precisely defined by a set of specific changes in facial expressions. A detailed description of the facial expressions associated with each emotion is provided below.

1. **Anger** – this emotion was expressed through pursed lips, furrowed eyebrows, and tense forehead muscles. Additionally, anger manifested itself through an intense facial expression, with a tendency to wrinkle the forehead and increased muscle tension around the eyes.

2. **Happiness** - the emotion of happiness was expressed by raised corners of the mouth and relaxed facial features. It was also characterized by wrinkles around the eyes, which were a natural expression of joy.

3. **Sadness** - sadness was manifested by the corners of the mouth falling, the eyelids slightly lowering, and the inner parts of the eyebrows rising. Additionally, the entire face took on an expression of reduced muscle tension, which was intended to reflect the emotion of sadness.

Each of these emotional states was presented by the participant for the full 3 minutes of the recording, maintaining as much naturalness and consistency with the actual emotions as possible.

Test Recording

In addition to the three main recordings, in which the participant displayed a single emotion, one test recording was also made. This recording was intended to test the effectiveness of the facial point detection system under dynamic conditions of changing facial expressions.

The test recording also had a calibration function. This allowed for a preliminary assessment of the analytical software's effectiveness, identification of any errors in facial point registration, and adjustment of the detection system parameters. This allowed for optimization of the registration process in subsequent recordings.

6 Measurement Data Analysis and Processing

6.1 Film Frame Extraction and Analysis Process

After completing the video recording process, a key step in the project was to convert the collected raw recordings into data suitable for further quantitative analysis.

The first step involved segmenting the video recordings, i.e., dividing the footage into individual image frames. This process was performed continuously, without selecting key frames, meaning that each captured frame was processed. This method, although computationally demanding, allows for the preservation of full information about the dynamics of facial movements in real time – even subtle facial changes that might be missed with a traditional key-moment approach.

The sampling rate corresponded to the camera's 30 frames per second (fps) setting. As a result, each three-minute recording contained approximately 5,400 individual frames, resulting in a total of over 21,000 frames over four recording sessions. Additionally, approximately 1,800 frames from a test recording were included, which contained more complex and dynamic emotional transitions.

6.2 Numerical Data Extraction Using MediaPipe Library

After the video frame extraction stage was completed, we proceeded to obtain numerical data representing the arrangement of facial landmarks in each frame. For this purpose, an advanced library was used **MediaPipe**, developed by Google, which offers a wide range of functions in the field of real-time video and image analysis.

Specifically, the module used *MediaPipe Face Mesh* [9], enabling the detection of 468 3D landmarks on the surface of the human face. These landmarks describe, among other things, the contours of the eyes, eyebrows, nose, mouth, and the shape of the jaw and cheeks. Each landmark contains a set of spatial coordinates (X,Y,Z), allowing for a highly detailed representation of the facial configuration at a given moment in time.

All acquired data was saved in tabular format in Excel files. The data structure was as follows: the first column corresponded to the frame number (recording time), while the subsequent columns contained sequences of coordinate values of the feature points. One row corresponded to one video frame and contained data in the form:

$$\text{frame: } point1_X, point1_Y, point1_Z, \ldots, point468_X, point468_Y, point468_Z \qquad (1)$$

The use of such a format enabled easy processing of data in computational environments and their use in machine learning models.

6.3 Data pre-processing and preparation for statistical analysis

After completing the video recording stage and extracting numerical data from each frame, the obtained information was pre-processed and prepared for further statistical analysis. The primary goal of this stage was to properly organize, clean, and format the data so that it could be used in an analytical environment [1].

First, data from the **MediaPipe Face Mesh** library were exported as raw 3D spatial coordinates (X,Y,Z) axis for each of the 468 facial landmarks. These data represented the positions of the facial landmarks in each frame of the video (assuming a frame rate of approximately 30 frames per second, which translated into a large number of observations for each recording). Each dataset for a single recording therefore contained a matrix with dimensions corresponding to the number of frames and the number of facial landmarks, with separate columns for each spatial coordinate.

All data was stored in spreadsheet (.csv) files, facilitating further processing and importing into the analytical environment **Python** programming language and its data analysis libraries, such as `Pandas`, `NumPy` or `Matplotlib` were used for this purpose. This enabled not only convenient data management but also data visualization and preliminary exploration.

6.4 Features of MediaPipe

MediaPipe incorporates a point tracking system that maintains the position of facial points even when partially occluded or the viewing angle changes, as long as the face remains in the frame. This resulted in relatively high quality and stability of the acquired data, which was beneficial for the planned analysis of facial expressions.

At this stage, no manual data corrections, such as interpolation of missing points or removal of outliers, were performed. It was assumed that the built-in mechanisms of MediaPipe were sufficient to ensure an adequate level of data consistency and completeness for the experimental purposes.

6.5 Spatial Coordinate Normalization

To ensure data comparability between different recordings and points, a simple but effective coordinate **normalization mechanism** was used, which transformed all values to a range between 0 and 1. Normalization was performed according to the following formula. Each coordinate (X,Y,Z) was therefore:

- first scaled by dividing by 1000 – which was intended to reduce the units to decimal fractions,
- then scaled by 2 – which brought the values to the range <-0.5,0.5>
- and finally shifted by 0.5 – which finally moved the data to the range <0.1>.

All normalized data were re-saved in new Excel spreadsheets, preserving the original data structure (i.e.: frame number, point number, X, Y, Z coordinates), which enabled their seamless integration with the subsequent stages of analysis.

6.6 Data Validation

After the normalization process was completed, an initial data validation check was performed. Among other things, it was verified that all values fell within the assumed range of <0.1> and that there were no missing values (blank or NaN values). Additionally, preview graphs of selected facial points over time were generated, allowing for a visual assessment of detection stability and scaling accuracy.

7 Statistical Analysis of Spatial Facial Data

After preparing the numerical data, statistical analysis was undertaken, the aim of which was to detect recurring spatial patterns of facial points corresponding to different emotional states. Instead of using classical exploratory methods, the analysis focused on the use of the k-means clustering algorithm in the environment [8,20].

Specifically, the input data was analyzed in the form of a 40 × 40 matrix, where each row (or column, depending on the data organization in MATLAB) represented a single feature vector from a given video frame. This vector contained selected spatial coordinates of facial landmarks (or transformed features based on these coordinates), which served as a description of facial expressions at a given moment in time [30, 33].

In particular, the input data was analyzed in the form of a 40 × 40 matrix, where each row (or column, depending on the data organization in MATLAB) represented one feature vector of a given video frame. This vector contained selected spatial coordinates of facial landmarks (or transformed features based on these coordinates), which served as a description of facial expressions at a given moment in time [30, 33].

The k-means algorithm allowed for grouping similar facial expression configurations into clusters, which could be interpreted as characteristic "patterns" of facial expression. This process not only allowed for the identification of structures corresponding to specific emotions, but also for the detection of so-called transition states – fragments of the recording in which facial expression changed dynamically, not clearly belonging to a single emotional category. The clustering results are presented in Fig. 1-3 [28].

8 Training and Test Dataset

The collected and processed dataset was divided into two primary sets: training and testing. The training set comprised 90% of the available samples and was used to train models

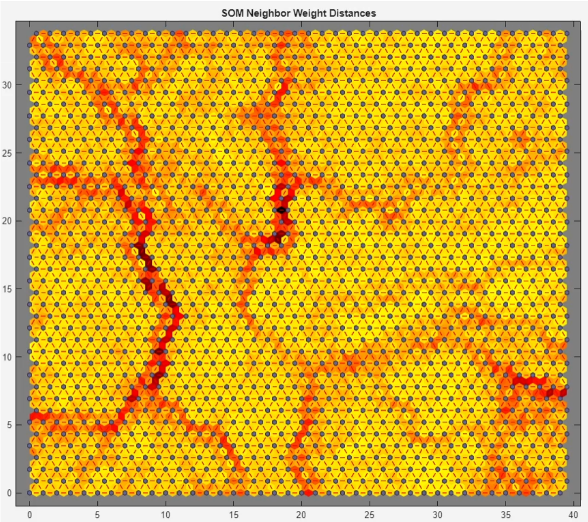

Figure 1. Neighbor connections in the SOM neural network. Source: Own study [28,33]]

.

classifying emotions based on spatial data. The testing set (the remaining 10%) was used to evaluate the model's performance and its ability to generalize to previously unknown data [28,30].

The division of samples was made while maintaining an even distribution of emotions in both sets, which was crucial for obtaining reliable results.

9 Selection of learning method and model implementation

9.1 Selecting a Learning Algorithm

Choosing the appropriate learning method plays a key role in building an emotion recognition system based on facial expression. This task falls into the class of classification problems, where the goal is to assign a given feature vector (in this case, the spatial arrangement of facial points) to one of three emotion classes: happiness, anger, or sadness.

Several alternative approaches were considered during the design phase:

– Classic statistical algorithms such as:
  - **Support Vector Machines (SVM)** - effective in linear and non-linear classifications with small sample numbers.
  - **Decision trees and Random Forest** –Decision trees create transparent models that make decisions based on sequential logical conditions, making them easy to interpret. Random Forest, as an ensemble version, improves prediction accuracy by reducing the risk of overfitting through random selection of data and features.

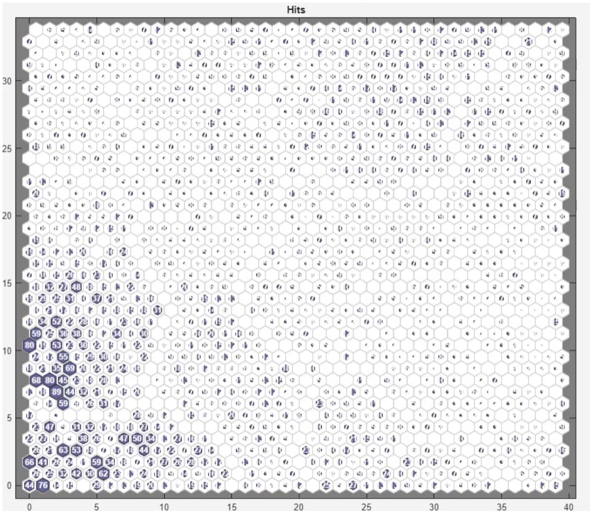

Figure 2. Map of effective hits of the model results according to the results of the human facial expression system. Source: Own study [28,33]

  - **K-Nearest Neighbors (k-NN)** – This is an intuitive algorithm that classifies new cases based on their similarity to the closest training examples. Its effectiveness depends on selecting the appropriate number of neighbors and the distance metric.
- Modern deep learning methods:
  - **Multilayer Neural Networks (MLP)**,
  - **Recurrent Neural Networks (RNN)** – for the analysis of temporal sequences,
  - **Convolutional Neural Networks (CNN)** – useful in image analysis, but requiring large data sets.

Due to the nature of the data – a numerical description of the coordinates of facial points over time – and the limited number of training samples, it was decided to use to **Single-Layer Feedforward Neural Network**. This solution provides a compromise between the simplicity of the model, its training speed, and its ability to recognize nonlinear patterns.

An additional, yet important, requirement was the ability to integrate with **MATLAB and Simulink Environment**, which enabled seamless implementation and testing of the network.

10 Implementation of the model in the MATLAB environment

The model implementation used input data in the form of time series representing changes in the position of facial landmarks in subsequent frames of the video recording. Each landmark was described by three spatial coordinates (x, y, z).

After extraction, the data was converted into a format suitable for further processing in MATLAB. A matrix structure was adopted, where each row corresponds to a single video frame, and the columns represent the spatial coordinates of all facial points in that frame. This

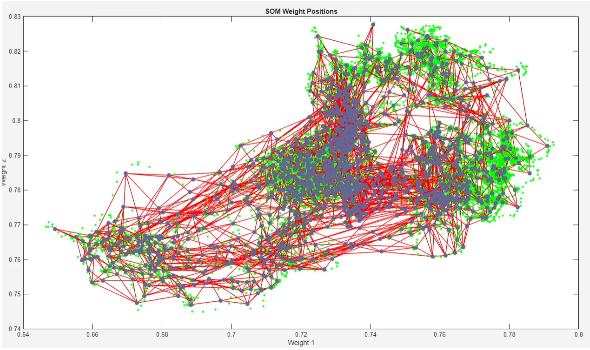

Figure 3. Projection of neuron weight positions in the data space onto a plane. Source: Own study [28,33]

means that a single row contains $3N$ numerical values, where N is the number of detected points (e.g., 478) and 3 denotes the number of spatial coordinates for each point. For a sequence of T frames, the resulting input matrix X is $T \times 3N$.

Formally, the matrix X takes the form

$$\begin{bmatrix} x_{1,1} & y_{1,1} & z_{1,1} & x_{2,1} & y_{2,1} & z_{2,1} & \cdots & x_{N,1} & y_{N,1} & z_{N,1} \\ x_{1,2} & y_{1,2} & z_{1,2} & x_{2,2} & y_{2,2} & z_{2,2} & \cdots & x_{N,2} & y_{N,2} & z_{N,2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ x_{1,T} & y_{1,T} & z_{1,T} & x_{2,T} & y_{2,T} & z_{2,T} & \cdots & x_{N,T} & y_{N,T} & z_{N,T} \end{bmatrix} \tag{2}$$

Each row of this matrix describes the complete spatial configuration of facial points at a given time frame $t \in [1, T]$. In turn, each triplet of columns $(x_{i,t}, y_{i,t}, z_{i,t})$ refers to a specific point $i \in [1, N]$ at time t, defining its position in three-dimensional space. This data representation allows for the analysis of both static facial geometric features and their dynamic changes over time.

The matrix prepared in this way provides direct input to the learning model implemented in the MATLAB environment. Its structure allows for the flexible application of various techniques for analyzing spatiotemporal data, including sequential models such as Recurrent Neural Networks (RNNs) and 1D Convolutional Networks operating along the temporal axis. Data preprocessing in MATLAB included, among other things, loading the matrix, standardizing the values, and preparing emotion class labels assigned to individual sequences based on the source data.

11 Architecture and training of the ANN model of the facial expression system

11.1 ANN architecture

As part of the research, a total of 12 artificial neural networks were designed as models of the facial expression system, differing in the number of neurons in the hidden layer, the training algorithm used, and the range of input data. Due to the very long training time of the ANN, the architecture was limited to a single hidden layer with a varying number of neurons, with the ANN output designed to generate three emotion classes.

The models can be divided into two main groups based on the type of input data:

- **Models based on the full set of facial points**:
  - Numbers of neurons: 3, 6, 9, 12 — learning using algorithm *Bayesian Regularization (BR)*,
  - numbers of neurons: 12 — learning using algorithm *Levenberg–Marquardt (LM)*,
  - numbers of neurons: 12 — learning using algorithm *Scaled Conjugate Gradient (SCG)*.

  The number of neurons was limited to a maximum of 12 to reduce the demand on working memory and computation time resulting from the large number of input features.
- **Models based on key facial features (eyes, eyebrows, nose, mouth)**:
  - numbers of neurons: 12 i 22,
  - Learning algorithm: *Levenberg–Marquardt (LM)*, *Bayesian Regularization (BR)*, *Scaled Conjugate Gradient (SCG)*.

  The number of 12 neurons was chosen to allow for direct comparison with models trained on the full set of points, while the configuration with 22 neurons was chosen based on the principle of approximating the number of neurons as the square root of the number of input features (in this case approximately $\sqrt{447} \approx 22$), taking into account the greater flexibility of the model.

11.2 The ANN training process for the facial expression system model

All models were trained using common parameters:

- the total number of observations is 12 000, with 4 000 for emotions
- maximum number of epochs: 1000,
- error tolerance: 0.001,
- data division: 90% data learning, 5% data validation, 5% data testing,
- monitored metrics: MSE error, R determination index

Table 1 presents selected results comparing all 12 models tested. The first six models used all measured human facial points as input, while the next six used only selected points. The number of neurons in the hidden layer ranged from 3 to 22, depending on the ANN architecture. Three learning algorithms were used for training, testing, and validation: Bayesian Regularization (BR), Levenberg-Marquandt (LM), and Scaled Conjugate Gradient (SCG). In most cases, the training resulted in an R-score of 1 or very close to 1 ($R = 0.999$), and the MSE error was also very low, ranging from 10^{-14} to 10^{-4}.

11.3 Model quality assessment method

To assess the quality of neural learning on the one hand, and the quality of the obtained neural models of the facial expression system on the other, attention was paid to the very large

Table 1. Comparison of the MSE error and the coefficient of determination R of the trained 12 ANN facial expression system model. Source: Own study [28,30].

Model	Input data	Neurons	Algorythm	MSE	R
1	all points	3	BR	$9{,}1669 \times 10^{-14}$	1
2	all points	6	BR	$2{,}1541 \times 10^{-13}$	1
3	all points	9	BR	$1{,}8028 \times 10^{-14}$	1
4	all points	12	BR	$7{,}7938 \times 10^{-14}$	1
5	all points	12	LM	$2{,}8821 \times 10^{-14}$	1
6	all points	12	SCG	$6{,}1685 \times 10^{-5}$	0,9999
7	main features	12	LM	$4{,}9525 \times 10^{-11}$	1
8	main features	12	BR	$6{,}1336 \times 10^{-11}$	1
9	main features	12	SCG	$1{,}4590 \times 10^{-4}$	0,9997
10	main features	22	LM	$1{,}1028 \times 10^{-10}$	1
11	main features	22	BR	$3{,}1879 \times 10^{-10}$	1
12	main features	22	SCG	$2{,}4748 \times 10^{-4}$	0,9994

decrease in MSE error during ANN training, even to the order of $10^{-4} - 10^{-14}$, and the very high R-value, close to 1 or close to 1.

However, in assessing the quality of the obtained models, the absolute errors, relative errors, and mean relative errors were examined. In this case, a simulation model was built in Simulink for comparative studies, sensitivity testing, and other quality indicators, including the accuracy of human facial expression prediction. For this purpose, the outputs from the neural models were compared with the outputs used in ANN training of the system models, obtaining discrepancies, absolute errors, relative errors, and mean relative errors, which were used to assess the quality of the model-system fit [28, 30]. Furthermore, an attempt was made to evaluate the model's effectiveness in predictive situations. The obtained output results were then compared to the expected values, the accuracy of which was set to the order of magnitude of 10^{-3}. For this purpose, both the absolute error and the relative error were used, with the model's effectiveness being determined by assessing the number of cases in which facial expressions could be assessed using the model in accordance with the system for all facial behaviors [28]. Individual errors were therefore determined using the following formulas:

- **absolute error**calculated as the difference between the predicted and expected values:

$$ar = |y_{\text{pred}} - y_{\text{true}}|;$$

- **relative error** expressed as the ratio of the absolute error to the expected value of the output signal from the facial expression system model:

$$re = \frac{|y_{\text{pred}} - y_{\text{true}}|}{|y_{\text{true}}|};$$

- **effectiveness** was expressed as the number of correct predictions divided by the total number of test cases, assuming the degree of accuracy of the experiment being conducted, which was related to the so-called **threshold value**, which in this study was set at 0.001) and used to determine whether a given prediction falls within the acceptable margin of such established measurement accuracy. If the result fell within this accuracy range, such a prediction was considered accurate.

This approach was intended to assess the performance of neural models in terms of assigning obtained results to emotion classes. However, it turned out that the accuracy assumption was too high, and therefore the effectiveness was only about $(7 - 9)$

11.4 Results and discussion

For each of the 12 resulting models presented in Table 1, we determined, among other things, the learning quality, the model's quality with respect to the facial expression system (so-called fitness), and the model's prediction accuracy based on, among other things, emotion accuracy—that is, the percentage of correct predictions regarding membership in a specific class. The results regarding model effectiveness are summarized in Table 2. From the analysis of these results, it can be noted that in all tested cases, due to the established high accuracy of the results generated by the model as predictions of the actual facial expression system (on the order of 10^{-3}), the achieved model learning efficiency was $(7.81 - 8.66)\%$.

Table 2. Prediction effectiveness for 12 neural network models. Own study [28,30]

Model	Neurons	Input data	Consistency of results[%]
1	3	all points	8,341
2	6	all points	8,339
3	9	all points	8,341
4	12	all points	8,335
5	12	all points	8,293
6	12	all points	8,001
7	12	main features	8,325
8	12	main features	8,307
9	12	main features	7,812
10	22	main features	8,291
11	22	main features	8,306
12	22	main features	8,656

As can be seen from Table 2, despite the very good regression indicators presented earlier (low MSE and high correlation coefficient R), the neural networks achieved a relatively low classification efficiency - accuracy at the level of 7–9%, which indicates the need for further research to determine what should be the acceptable level of forecast accuracy in relation to the accuracy of the ANN model, and this in relation to the measurement data, especially since the average relative errors in the field of testing the model-to-system fit were very low (consistency at the level of 100%). Furthermore, it is worthwhile to conduct further research on increasing the flexibility of weight adjustment in neural models, for example, using an evolutionary algorithm, or on the possibility of including a layer of neurons in the ANN in a spatial form representing the human face, which involves the use of hybrid ANNs with a cellular ANN as the output layer. Such an approach would also eliminate the overlap of characteristic features between emotion classes—some emotions in the experiment exhibited similar facial point arrangements, which could also have contributed to the lack of unambiguous classification.

12 Conclusions and directions for further research

A neural model of human facial expressions was designed and implemented in MATLAB and Simulink, using data analysis from the MediaPipe Face Mesh system to enable simulation and further integration with human-machine interaction systems. This involved, among other things:

- Obtaining a proprietary dataset of facial expressions by recording under controlled conditions,
- Extracting significant facial landmarks using the MediaPipe Face Mesh system, including eyebrows, eyes, and mouth,
- Developing a method for processing and analyzing the coordinates of these points to define basic facial movements,
- Designing a functional facial model that maps the relationships between input data and expressed emotions,
- Implementation of neural models of the facial expression system in MATLAB and Simulink, which were characterized by a low ANN learning error of 10^{-14} and a high coefficient of determination R close to 1.0.
- Building a simulation model in Simulink, enabling simulations, sensitivity testing, and comparative studies, thus preparing the model for potential real-time application.
- Conducting a series of tests and experiments on the simulation model in Simulink, confirming the model's correctness and relatively high performance compared to the facial expression system, with an average relative error close to several percent.
- One issue not fully resolved by the conducted research is the method for determining the model's effectiveness measurement, and perhaps also for measuring the effectiveness of the human facial expression system. This involves, among other things, the need to examine the relationship between the accuracy of the effectiveness measurement and the accuracy of, among other things, the accuracy of the human facial expression system. measurement data used in the research.
- Therefore, in further research experiments, it is worth using at least dedicated classification mechanisms, which may also increase the accuracy of emotion prediction based on facial features.

The obtained neural models of the human facial expression system have practical and research value, as they are characterized by a relatively low learning error (MSE) and a high coefficient of determination (R), as well as very low mean relative errors between the model and the system. However, high prediction accuracy was not achieved in the model's recognition of actual emotional states based on facial expressions. Therefore, research in this area should be continued.

Due to the fact that the article presents the results of modeling and extracting human facial expressions using artificial neural networks and the MATLAB/Simulink environment, these studies should therefore be continued in order to improve, among other things, the quality of the obtained models and classification results [38-39], including

1) a more in-depth analysis of the obtained results and interpretation of the presented research results in the figures included in the publication, which involves the need to obtain

appropriate quality results of the recognized images and may provide interesting descriptive comments,

2) a broader and more critically oriented review of the subject literature (including in the area of using deep learning in facial expression recognition), which may provide interesting conclusions in the context of the studied facial expression problem, indicating both the positive obtained research results as well as the limitations of the obtained facial expression model,

3) to examine the performance of the model, including, among other things, considering the confusion matrix and evaluation metrics such as the error rate (ERR), accuracy (ACC), sensitivity, and precision, etc., in order to provide a quantitative assessment of the effectiveness, efficiency, and potential robustness of the model and the actual system based on the obtained measurements, etc.

In further research experiments, it is worthwhile to employ at least dedicated classification mechanisms, which may also improve the accuracy of emotion prediction based on facial expressions. Furthermore, a separate research topic may concern the generalization of knowledge regarding the measurement of facial expressions across different people's faces, which was beyond the capabilities of the intended research and computational environment.

References

1. M. Adamiak, Wykorzystanie technik uczenia maszynowego i teledetekcji do wspomagania analiz, Uniwersytet Łódzki. 2022. URL: https:www.bip.uni.lodz.pl/fileadmin/user_upload/Adamiak_M._-_prof._S.Mikrut.pdf [access: 2024-25].
2. A. Bulat and G. Tzimiropoulos, „How far are we from solving the 2 D & 3 D face alignment problem ?", Proceedings of the IEEE International Conference on Computer Vision (ICCV). 2017, pp. 1021–1030.
3. M. Białas and A. Kowalska, „Analiza ekspresji twarzy w czasie rzeczywistym z użyciem biblioteki Open CV i uczenia maszynowego". Informatyka i Automatyka, Vol. 30.4(2021), pp. 45–53.
4. I. Cohen and others, „Facial expression recognition from video sequences: Temporal and static modeling". Computer Vision and Image Understanding 91.1–2 (2003), pp. 160–187
5. T. F. Cootes, G. J. Edwards and C. J. Taylor, „Active appearance models". IEEE Transactions on Pattern Analysis and Machine Intelligence 23.6 (2001), pp. 681–685.
6. P. Ekman, „Facial expressions of emotion: New findings, new questions", Psychological Science, Jan. 1992, Vol. 3, No. 1, pp. 33-38.
7. P. Ekman, Telling Lies: Clues to deceit in the marketplace, politics, and marriage, W. W. Norton & Company, 2009.
8. Guide for MATLAB, Guide for Simulink, Guide for System Identification Toolbox, Guide for Control System Toolbox, Guide for Neural Network Toolbox, Guide for Deep Learning Toolbox. Getting Started Guide, versions 2019b–2024b. The MathWorks®.
9. Google Research, MediaPipe – biblioteka do przetwarzania sygnałów wizualnych w czasie rzeczywistym.https://mediapipe.dev/.. 2023 [access: 2024-25].
10. A. Horzyk, Sztuczne systemy skojarzeniowe i asocjacyjna sztuczna inteligencja, Exit, Warszawa 2013, pages 277.
11. H. Jia, Z. Xiao and P. Ji, "Fatigue Driving Detection Based on Deep Learning and Multi-Index Fusion", IEEE Access, Vol. 9, pp. 147054-147062, 2021.

12. M. Janik, „Rozpoznawanie emocji na podstawie analizy mimiki twarzy z wykorzystaniem głębokich sieci neuronowych", Praca doktorska, Politechnika Warszawska, Wydział Elektroniki i Technik Informacyjnych, Warszawa 2020.
13. V. Kazemi and J. Sullivan, „One millisecond face alignment with an ensemble of regression trees", Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE. 2014, pp. 1867–1874.
14. B. J. Kowalski, „Zastosowanie algorytmów klasyfikacyjnych w rozpoznawaniu emocji na podstawie mimiki twarzy", Informatyka i Automatyka Vol. 27.2(2018), pp. 89–96.
15. Z. Liu, Y. Peng, W. Hu, Driver fatigue detection based on deeply-learned facial expression representation, Journal of Visual Communication and Image Representation, Volume 71, August 2020.
16. S. Li and W. Deng, "Deep Facial Expression Recognition: A Survey," in IEEE Transactions on Affective Computing, Vol. 13, No. 3, pp. 1195-1215, July-Sept, 2022.
17. M. J. Lyons, J. Budynek and S. Akamatsu. „Automatic classification of facial expressions using Gabor wavelets". Proceedings of the Third IEEE International Conference on Automatic Face and Gesture Recognition. IEEE. 1999, pp. 200–205.
18. G. Littlewort and others, „Dynamics of facial expression extracted automatically from video", Image and Vision Computing, Vol. 24.6(2006), pp. 615–625.
19. M. Laszczak, „Umiejętność deszyfracji mimiki twarzy – studium przypadku", TBSP UJ, Kraków 2015, https:tbsp.wpia.uj.edu.pl/documents/4137545/108606169/Laszczak_5%20_2015.pdf [access: 2024-2025].
20. MATLAB and Simulink Documentation. Online documentation for MATLAB R2022b-R2024b. The MathWorks®. 2023-2025.
21. „Mimikra emocje w kontekście społecznym. Znaczenie mimicznego naśladowania emocji", SWPS Projekty Naukowe (2023), SWPS Uniwersytet Humanistyczno-Społeczny. https://swps.pl/nauka-i-badania/projekty/42-projekty-naukowe/20712-mimikraemocji-w-kontekscie-spolecznym, [acces: 2024-25].
22. A. Nowakowski, Emocje i ich rozpoznawanie na podstawie ekspresji mimicznej. Wydawnictwo Uniwersytetu Śląskiego, Katowice 2018.
23. S. Osowski, Sieci neuronowe do przetwarzania informacji. OW Politechniki Warszawskiej, Warszawa 2013, pages 422.
24. M. Pantic i L. J. M. Rothkrantz, „Automatic analysis of facial expressions: The state of the art", IEEE Transactions on Pattern Analysis and Machine Intelligence (2006).
25. T. Pfister and others, „Recognising spontaneous facial micro-expressions", IEEE International Conference on Computer Vision (ICCV), IEEE, 2011, pp. 1449–1456.
26. Polsko-Japońska Akademia Technik Komputerowych. „Systemy wykrywania i rozpoznawania twarzy na obrazach". Przegląd Policyjny 4(140)(2020). URL: https://pja.edu.pl/wpcontent/uploads/2023/02/Systemy-wykrywania-i-rozpoznawania-twarzy-na-obrazach.pdf [access: 2024-25].
27. B. Paradowski, Sztuczna inteligencja w analizie zachowań ludzkich, Warszawa, WN PWN, Warszawa 2019.
28. J. Rosiński, Model mimiki twarzy człowieka i jego implementacja w środowsku MATLABA i Simulinka, praca inżynierska napisana w Instytucie Informatyki na Wydziale Nauk Ścisłych i Przyrodniczych pod kierunkiem dr hab. inż. J. Tchórzewskiego, prof. uczelni, Uniwersytet w Siedlcach, Siedlce 2025, pages 51.
29. C. Shan, S. Gong and P. W. McOwan, „Facial expression recognition based on Local Binary Patterns: A comprehensive study", Image and Vision Computing, Vol. 27.6(2009), pp. 803–816.
30. J. Tchórzewski, Metody sztucznej inteligencji i informatyki kwantowej w ujęciu teorii sterowania i systemów, WN UPH, Siedlce 2021, pages 343.
31. J. Tchórzewski, A. Wielgo, Mobile humanoid robots and the possibility of neural modeling human motion, Chapter 3 [in:] Intelligent Systems and Information Technologies. Modeling and Analysis of

Intelligent Information Systems [ed.] J. Tchórzewski and P. Świtalski, No. 2, University of Siedlce, Siedlce 2023, pp. 67-98.

32. D. Wojciechowski, „Integracja sieci neuronowych w analizie sekwencji mimiki twarzy w czasie rzeczywistym”, Techniki Obliczeniowe, Vol. 34.1(2020), pp. 45–53.
33. S. Wierzchoń and M. Kłopotek, Algorytmy analizy skupień. Warszawa, WNT, Warszawa 2015, pages 395.
34. M. Valstar and M. Pantic, „Fully automatic recognition of the temporal phases of facial actions”. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), Vol. 36.1(2006), pp. 61–75.
35. L. Yin and others, „A 3D facial expression database for facial behavior research”, 8th IEEE International Conference on Automatic Face & Gesture Recognition, IEEE, 2008, pp. 211–216.
36. X. Zhu and oters, „Face alignment across large poses: A 3D solution”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 146–155.
37. Y. Zhang, Q. Ji, S. Gong, „Real-time facial expression recognition based on hierarchical probabilistic models”, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2005, pp. 118–125.
38. A. Talukder, S. Ghosh, Facial Image expression recognition and prediction system. Sci Rep 14, 27760. 2024.
39. M. Fatchan, M. Hery Purnomo, Affandy, A. Zainul Fanani and L. Marlinda, "Support Vector Machine and Neural Network Algorithm Approach to Classifying Facial Expression Recognition," 2020 Fifth International Conference on Informatics and Computing (ICIC), Gorontalo, Indonesia, 2020, pp. 1-5.