



AUTOMATED DATA-COLLECTION FOR PERSONALIZED FACIAL EXPRESSION RECOGNITION IN HUMAN-ROBOT INTERACTION

FOUZIA ADJAILIA

Technical University of Kosice
Department of Cybernetics and Artificial intelligence
fouzia.adjailiak@tuke.sk

PETER SINCAK

University of Miskolc, Hungary
Department of Information Engineering
Peter.Sincak@tuke.sk

Abstract. Face recognition systems, which attempt to identify the emotions that a person is feeling, have been around for quite some time. Facial expression recognition is the technique of detecting facial expressions based on interpretations of patterns in a picture. Because every person's face is unique, when we apply these methods to pictures of people, we are able to identify their facial expressions as being unique. In this research, we build a web-based data collecting application that is completely automated and includes a virtual avatar to guide users through the procedure. The input data we dealt with included written input in the form of six emotions (anger, disgust, fear, happiness, surprise, and sorrow) plus neutral, as well as video footage with a length of 20 seconds for each. With the use of the data, a customized face expression recognition method based on deep learning architecture known as MobileNets would be developed.

Keywords: Personalized facial expression recognition; human-robot interaction; MobileNet; data collection; virtual avatar.

1. Introduction

The ability to recognize facial expressions is important for human-robot interaction, because the ability to perceive and act upon human emotions allows robots to better understand their users' needs and to react accordingly. While some robots are capable of recognizing specific emotional states, such as happiness or sadness, it is difficult for them to recognize unanticipated emotional states. The proposed method uses machine learning to create a personalized

model of how a user's face changes when they feel different emotions, so that the robot can better match an appropriate response.

Macro personalization is a broad term that refers to the process of pleasing a large number of individuals who have similar requirements and desires. It is based on research that demonstrates that various groups of individuals express and interpret emotions differently. On a larger scale, it's a comparison of cultures - the Japanese place a greater emphasis on the eyes when reading emotions [1], while Americans place a greater emphasis on the mouth region. Other group-based instances include individuals with autism and their emotional manifestations, as well as groups of individuals suffering from different physical or mental diseases.

At the micro-scale, personalization entails concentrating on a single individual and customizing systems to meet their unique requirements and desires; as a result, system responses vary by individual. A prominent example of this kind of customization are the customized ads that are presently being shown to individuals based on their past purchases, search history, and likes/dislikes.

There are currently very few studies focusing on micro-scale personalized emotion recognition from images; the primary focus is on personalization via speech recognition, heart rate, skin conductivity, and other physiological signals; however, because classic emotion recognition from images is a well-researched topic, it is only a matter of time before all the new studies resurface.

1.1. Macro Personalization

Hammami in [2] developed a novel technology called AMD (Autistic Meltdown Detector) that alerts caretakers of autistic individuals when they are experiencing Meltdown Crisis based on their facial expressions. It is important to recognize Meltdown crises because they may be hazardous for autistic individuals - particularly children - since they impair a person's ability to regulate their behaviors and increase the likelihood of their injuring themselves or others. To perform their experiment, scientists needed to build their own dataset of real-world situations involving autistic children in both their normal and meltdown states. They used the Kinect for Windows v2 as a camera to capture numerous movies in an autistic health facility. This study focuses on compound emotions - those that are formed by the mixing of fundamental emotions (happily surprised, angrily surprised, ...). They evaluated different techniques for feature selection, including filtering, wrapping, and NCA (Neighborhood Component Analysis). Multiple deep learning algorithms were used to assess the outcomes of these methods: FF (Feed Forward), CFF (Cascade Feed Forward), RNN (Recurrent Neural Network), and LSTM (Long Short

Term Memory). The greatest accuracy - 85.5 percent - was obtained using the Information Gain feature selection technique and RNN classification.

Psychopathology is diagnosed via a verbal examination of a person using organized conversation and questions. Integrating behavioral patterns into the assessment process may expedite the process of providing assistance to people in need. Detecting et al in [3] concentrated on detecting serious depression via facial expressions, speech rhythm, stress, and intonation. They were conducting interviews with patients in a clinic study for depression therapy. Their facial expressions were recorded manually and automatically using the Facial Action Coding System (FACS) and active appearance modeling (AAM). Support Vector Machine (SVM) was utilized as a classifier in each of these methods. It is a binary classifier that has been shown to be effective in pattern recognition tasks as well as in the identification of faces and facial expressions.

1.2. Micro Personalization

With the use of commercial sensors, Subramanian et al in [4] developed ASCERTAIN - a multimodal database for implicit personality and affect recognition. It establishes a link between personality characteristics and emotional states through physiological reactions. It includes personality scores and emotional self-ratings from 58 individuals, as well as electroencephalogram, electrocardiogram, galvanic skin reaction, and facial activity data gathered while watching affective film clips utilizing commercial and wearable sensors and a webcam. Binary recognition was used to assess personality and emotional characteristics. Arousal and valence were the output classes for emotion. They trained naive Bayes and linear SVM on user-specific characteristics and compared their findings to those of other researchers (DEAP). Their F1-scores were higher. They attributed it to the use of film clips rather than music videos as emotional triggers and to the increased participation, which resulted in a larger training set.

In [5], the author suggested a method for customized emotion recognition by pre-classifying emotions using an identification phase. It was built upon the author's earlier studies, in which he addressed identity and emotion detection as distinct tasks. The technology was designed with resilience and processing speed in mind as a component of a social robot. The FaceTracker was used to identify and track faces. It is an application that does these objectives by fitting deformable 3D models of a human face iteratively. It adjusts to the tracked face by taking the position of the brows, lips, jaw, and other facial characteristics into consideration. The outcome is a collection of points in three-dimensional space that accurately represent the face. This geometric data is used only by the identifying system.

2. Contribution of the article

The research in this paper has made several contributions to the development of facial expression recognition, For example:

- Design and implement a fully automated web application interface for crowd-sourcing and collecting our own data for the learning algorithm for facial expression recognition.
- Propose a strategy for human-robot interaction by using the technologies related to virtual avatars to assist in the process of data collection.
- Organize, analyze, and evaluate the data collected, as well as creating our own dataset for facial expression recognition.
- Propose a cutting-edge deep learning-based system for personalized facial expression recognition.

3. Experimental Setup

The rapid prototyping process generally starts with the creation of an experimental setup. An experimental setup gives researchers a way to identify, and work around, sources of error often encountered during experimentation. In particular, we will cover our architecture decisions. System architecture for our experiment consists of five steps: data collection, data acquisition, data pre-processing, facial recognition, and facial expression recognition, figure 1.

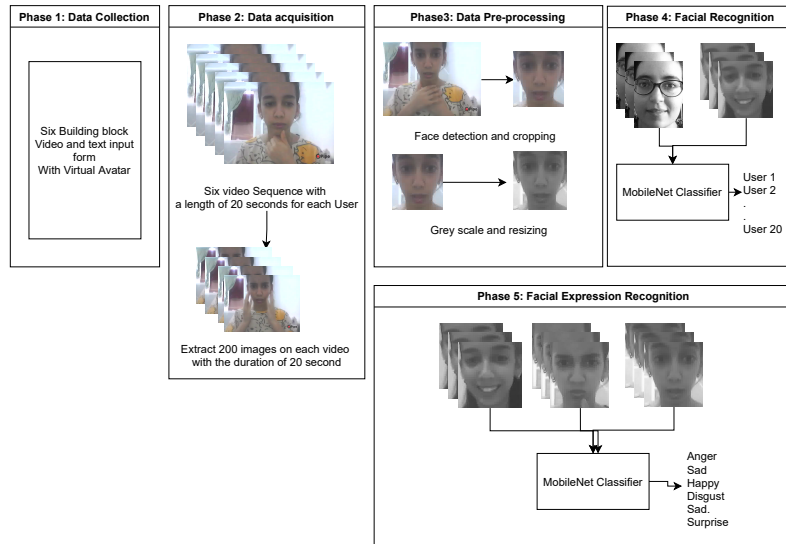


Figure 1. The proposed System Architecture.

4. Data Collection

Our experiment has been designed to collect data that will enable us to answer our research question. Our research design can be broken down into three parts: the implementation of the web application, the process of recording data, and user experience. We will go over our use of tools to collect data and track user experience. We used some technologies for the web application such as wordpress to design the web application and we hosted our web application in EasyWp. Further technologies chosen for the development of the final solution are described in the following sections.

4.1. Technologies for Web Application

The website was developed using WordPress (4.1.1) and it was hosted on EasyWP (4.1.2).

4.1.1. *WordPress*

WordPress [6] is a free and open source blogging tool and a content management system based on PHP and MySQL. It was designed for web designers, developers, and users to build and manage websites and blogs. WordPress utilizes a template processor to power its web template system. It is a front controller design, with all requests for non-static URIs being routed to a single PHP file that parses the URI and determines the destination page. This enables more human-readable permalinks to be supported. Users of WordPress may install and switch between several themes. Themes enable users to customize the appearance and functionality of a WordPress website without modifying the website's fundamental code or content. Each WordPress website must have at least one theme, and each theme must adhere to WordPress standards by utilizing structured PHP, valid HTML (HyperText Markup Language), and Cascading Style Sheets (CSS). Themes may be installed straight from the WordPress dashboard's "Appearance" management feature. WordPress' plugin architecture enables users to enhance a website's or blog's features and functionality by adding custom functions and features that enable users to adapt their sites to their unique needs.

4.1.2. *EasyWP*

EasyWP [7] is a cloud-based WordPress hosting solution that simplifies the process of hosting WordPress websites by offering a fully managed service that includes automated backups, malware scanning, and security upgrades.

4.2. Technologies for Data Collection

We utilized two plugins for data collection: Pip video recorder (4.2.1) and Contact form 7 (4.2.2). The first plugin is responsible for capturing videos and storing them in the cloud, while the second plugin allows users to submit their data after agreeing to the terms and conditions.

4.2.1. Pip Video Recorder

Pip Video Recorder [8] Provide a Framework for recording, processing, organizing, and even playing back. It takes care of recording from desktop and mobile web browsers, as well as native apps. All recordings are transcoded to (.mp4) format, rotated, and watermarked as required. Store: it takes care of storage, but recordings can be pushed to our own storage. Their (.mp4) conversion profile (H.264+ AAC) ensures cross-browser and cross-device playback.

4.2.2. Contact Form 7

Contact Form 7 [9] is a plugin developed by Takayuki Miyoshi that manages numerous contact forms and allows for flexible customization of the form and letter contents using simple markup. The form has Ajax-powered submission, CAPTCHA, and Akismet spam screening, among other features.

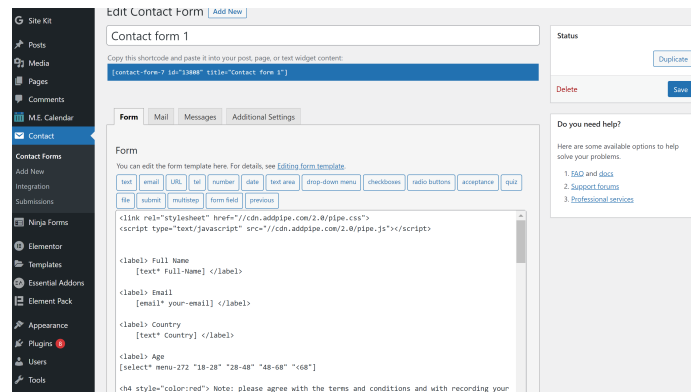


Figure 2. Contact form plugin screenshot [9].

4.3. Technologies for Virtual Avatar

Human-like virtual avatars have the potential to influence people, and to guarantee a smooth and well established huma robot interaction, we used a virtual avatar (4.3.1) that used a human-like voice(4.3.2).

4.3.1. Graphics Interchange Format Image

We utilized an animated (GIF) moving image of a young lady waving her arms as the virtual avatar.



Figure 3. Animated figure with synchronized lips and arms [10].

4.3.2. Voice Maker

We attempted to make the speech delivered by the virtual avatar sound human-like in order to reassure users that they are not speaking to a robot. To do this, we used VoiceMaker [11] which is an artificial intelligence-powered online text-to-speech converter. They have over 600 standard and natural-sounding AI voices available in over 70 languages. We may utilize our voices in videos that you post on YouTube, Vimeo, Facebook, Instagram, or your own personal website. They leverage artificial intelligence and machine learning to push the envelope and deliver a highly human-like Text to Speech experience complete with a configurable audio style, voice speed, pitch, volume, pause, emphasis, audio format, and audio profile options.

5. Data Acquisition

Data Acquisition is a very specific process. It involves expanding the data collection efforts by automating some process that was previously performed manually. The videos were collected from 16 participants, each video was 20 seconds long, and the data acquisition procedure comprised the extraction of 200 pictures from each video for each subject. Because several movies lasted less than 20 seconds, the pictures retrieved varied from the 200. The step of data acquisition consists also of emotion labelling which is an automatic process where a user can label each frame of the video with its emotion using a set of predefined options.

6. Data pre-processing

In the data pre-processing phase, we start with facial detection, which is accomplished using a simple Python script that takes use of opencv; once the faces are identified, the images are cropped and scaled to 96x96px in preparation for classification using MobileNet.

7. Classification

G. Howard et al. [12] introduced the term "MobileNets" to refer to a class of efficient models for mobile and embedded vision applications. MobileNets are built on a simplified design that uses light deep neural networks using depthwise separable convolutions. The MobileNet architecture is defined in Table 1.

Table 1. MobileNet Body Architecture [12].

Type / Stride	Filter Shape	Input Size
Conv / s2	3 X 3 X 3 X 32	224 X 224 X 3
Conv dw / s1	3 X 3 X 32 dw	112 X 112 X 32
Conv / s1	1 X 1 X 32 X 64	112 X 112 X 32
Conv dw / s2	3 X 3 X 64 dw	112 X 112 X 64
Conv / s1	1 X 1 X 64 X 128	56 X 56 X 64
Conv dw / s1	3 X 3 X 128 dw	56 X 56 X 128
Conv / s1	1 X 1 X 128 X 128	56 X 56 X 128
Conv dw / s2	3 X 3 X 128 dw	56 X 56 X 128
Conv / s1	1 X 1 X 128 X 256	28 X 28 X 128
Conv dw / s1	3 X 3 X 256 dw	28 X 28 X 256
Conv / s1	1 X 1 X 256 X 256	28 X 28 X 256
Conv dw / s2	3 X 3 X 256 dw	28 X 28 X 256
Conv / s1	1 X 1 X 256 X 512	14 X 14 X 256
Conv dw / s 1 5x Conv / s1	3 X 3 X 512 dw 1 X 1 X 512 X 512	14 X 14 X 512 14 X 14 X 512
Conv dw / s2	3 X 3 X 512 dw	14 X 14 X 512
Conv / s1	1 X 1 X 512 X 1024	7 X 7 X 512
Conv dw / s2	3 X 3 X 1024 dw	7 X 7 X 1024
Conv / s1	1 X 1 X 1024 X 1024	7 X 7 X 1024
Avg Pool / s1	Pool 7 X 7	7 X 7 X 1024
FC / s1	1024 X 1000	1 X 1 X 1024
Softmax / s1	Classifier	1 X 1 X 1000

8. Evaluation and Discussion

We will present in the subsections a series of procedures that follow the implementation of our experiment for the purpose of evaluation.

8.1. Data Collection Evaluation

Data Collection Evaluation is the act of gathering and analyzing data that helps us understand how users interact with a product. User research can be used to inform design or influence the direction of development, which may help achieve our goals.

8.1.1. Design and User Experience

To ensure a successful data collection evaluation we need to ensure that the design of our experiment through questioning has been effective. In order to do this effectively it will be necessary for us to develop an effective interview protocol both for the researcher and receiver to be aware of during data collection in order to produce reliable and valid results that can support our findings.

Our experiment's landing page is simple and well-designed; the use of typefaces and a virtual robot aided users in navigating and submitting their application swiftly and simply, figure 4.



Figure 4. Landing page of our data collection web application [10].

At the beginning of the form, users are prompted to provide their name, username, country, and age interval in a text input box, Figure 12. We provided two messages to users prior to initiating the submission process, and these messages are as follows:

- Note: please agree with the terms and conditions and with recording your face after starting the experiment
- caution: Some videos may not be very pleasing and comfortable for some people, we apologize if some videos caused uncomfortable emotions

As seen in Figure 13, the form is divided by six blocks that correspond to the six emotions (anger, fear, disgust, happy, sad, surprise). Figure 13, depicts the form's first block, which has a title, a video recording field, a video display field, and a radio list containing the six emotions plus neutral.

8.1.2. Emotion Triggers Database

The emotion triggers database is a collection of videos arranged to provoke an emotion from the viewer. Each video contains one or more of the following human emotions: (anger, fear, disgust, happy, sad, surprise). While compiling the testing database, We were astounded at how difficult it was to locate even a small sample of short videos capable of eliciting any response in a person. Of course, there are several videos that make people laugh or at the very least smile, but this alone is insufficient. There is a dataset accessible that contains videos and graphics designed to elicit emotional responses in autistic children, but those for adults are available upon request. However, we believe that developing one's own can be advantageous, particularly if the expert or target group can participate.

The video selected for our experiment are included in Table 3, along with their associated URLs.

To define and evaluate the choice of the videos, that fall under each emotion, and based on the results tabulated for convenience in Table 2 of the emotion triggers database and the corresponding users submission, the effectiveness of each video could be reviewed.

Taking in consideration that user 20, made a mistake of not submitting the input text of the emotion corresponding to each video. According to the table 2, the success rate of sad and disgust videos was 78.94% each, while happy was 73.68%, anger 57.63% surprise 52.63% and fear with 36.84% of success rate.

8.1.3. Quantitative Analysis

Quantitative analysis is the process of analysing data which contains numerical information.

As indicated in Table 4, our data collection process resulted in 15 submissions; 16 individuals out of 20 provided both video and text input; however, four participants submitted only text because they did not want their faces filmed; and one participant contributed only video input.

Table 2. Users respond to each video.

User	Intended emotion						Input emotion
	Fear	Happiness	Disgust	Anger	Surprise	Sadness	
1	Surprise	Surprise	Disgust	Neutral	Neutral	Sadness	
2	Disgust	Happiness	Disgust	Neutral	Fear	Sadness	
3	Disgust	Happiness	Disgust	Fear	Surprise	Sadness	
4	Neutral	Happiness	Disgust	Anger	Neutral	Sadness	
5	Neutral	Surprise	Disgust	Anger	Happiness	Sadness	
6	Neutral	Happiness	Disgust	Sadness	Neutral	Sadness	
7	Fear	Happiness	Disgust	Anger	Happiness	Sadness	
8	Fear	Happiness	Disgust	Anger/sadness	Surprise	Sadness	
9	Neutral	Neutral	Neutral	Anger	Neutral	Anger	
10	Surprise	Neutral	Disgust	Anger	Fear	Sadness	
11	Sadness	Happiness	Neutral	Neutral	Surprise	Neutral	
12	Disgust	Happiness	Happiness	Fear	Fear	Sadness	
13	Disgust	Happiness	Neutral	Sadness	Happiness/surprise	Neutral	
14	Fear	Happiness	Disgust	Anger	Surprise	Sadness	
15	Fear	Happiness	Disgust	Anger	Surprise	Sadness	
16	Fear	Happiness	Disgust	Anger	Surprise	Sadness	
17	Fear	Happiness	Disgust	Anger	Surprise	Sadness	
18	Neutral	Disgust	Disgust	Surprise	Surprise	Sadness	
19	Fear	Happiness	Disgust	Anger	Surprise	Happiness	
20	Happiness	Happiness	Happiness	Happiness	Happiness	Happiness	

8.1.4. Qualitative Analysis

We did a more in-depth analysis of the data based on four qualitative sub sections: age, gender, country, and place.

First, we plotted the distribution of ages. This gives us a picture of how our data is distributed with respect to the variable age. Figure 5a, shows that 25% of participants are aged between 18 and 28. However, 30% were in the age interval of 28-48. 35% were between 48-68 years old, and 10% were above 68 years old.

In our data, figure 5b, we notice that we have two outlier equale groups; women 50% and men 50%.

We saw that while around 65% of participants were from Slovakia, 30% from Algeria and 15% from Colombia, figure 5c.

Figure 5d describes the distribution of respondents by their place, in terms of the organizational status of their workplaces (workplace, home or unknown) and geographical. 60% of our participants submitted their applications from home, while 35% were at work. One participant was unknown.

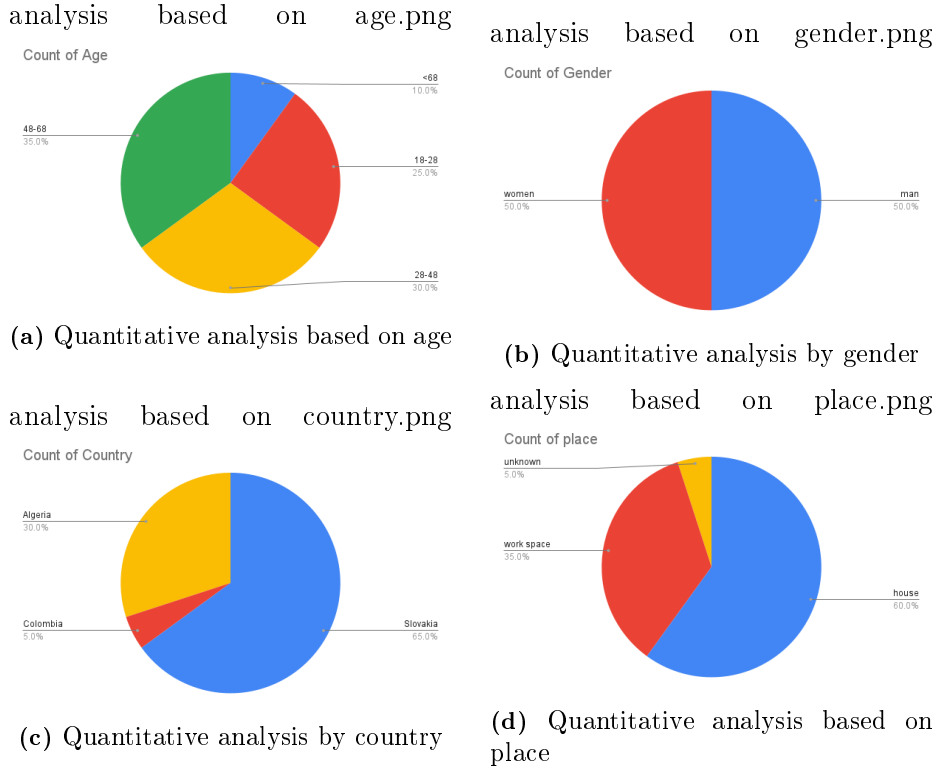


Figure 5. Quantitative analysis

8.2. Experiment Evaluation

In this section, we will interpret the results obtained from the experimental design and implementation of Facial recognition and facial expression recognition for the 16 users.

8.2.1. Facial Recognition

Prior to training, photos were sorted by person ID; in all, the dataset for face recognition contained 17627 images.

For training, we used 85% of the data, while only 15% was used for testing. The input size was 224x224px, the learning rate was 0.001, and the patch size was 16.

Graphs are used to explain the performance of models; the graph in Figure 6a illustrates the performance of facial recognition; the model performed well on both training and testing data. where the training and test accuracy were

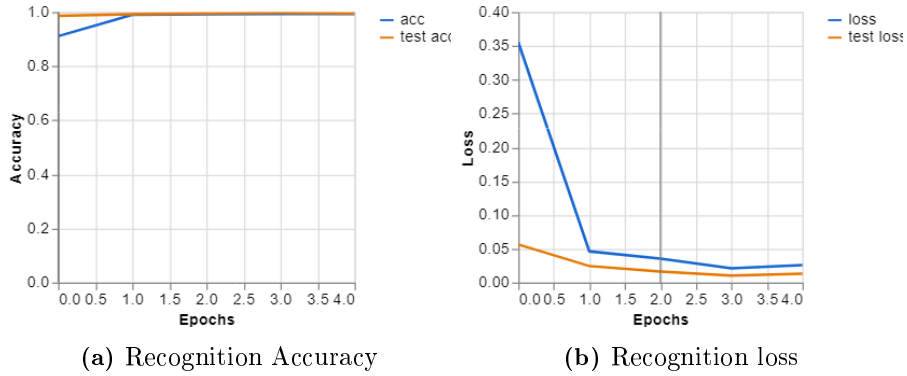


Figure 6. 5 epochs: Recognition and Accuracy.

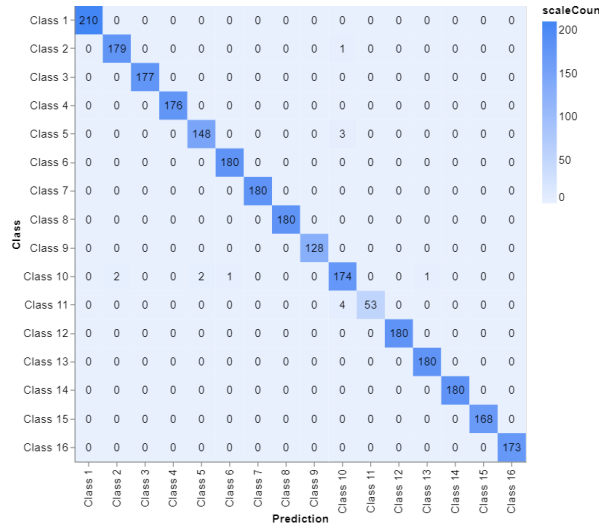


Figure 7. Confusion matrix.

both 100%, but the recognition loss, figure 6b was 0.3 and the test loss was 0.2.

Confusion matrices can be seen in Figure 7. Facial Recognition Confusion Matrix shows, that the model was well performing

8.2.2. Facial Expression Recognition

To determine if the personalization aspect was effective, I selected two subjects (User 4 and User 14) with performing networks from the learning process

- both trained in five and ten epochs. where the training data was 85% and the test data 15%.

User 4: a Slovak man user in the age between 28-48 years old, with beard and no glasses, in a workplace.

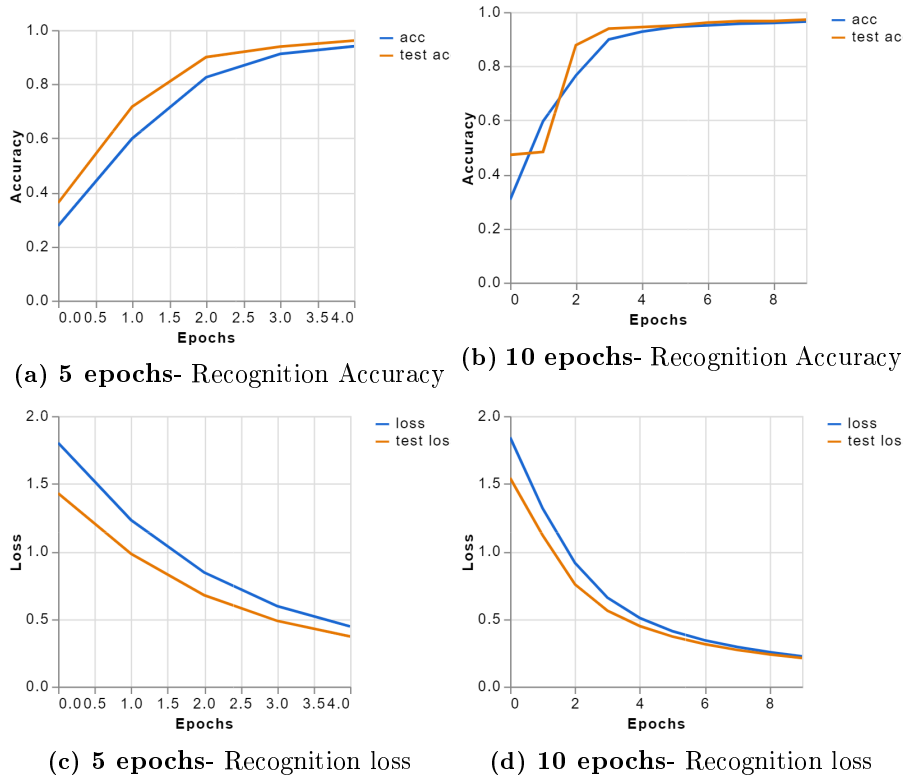


Figure 8. User 4 results.

The findings shown in figures above demonstrate that the model performs well for user 4.

User 14: Algerian women user in the age between 18-28 years old, with glasses, at home.

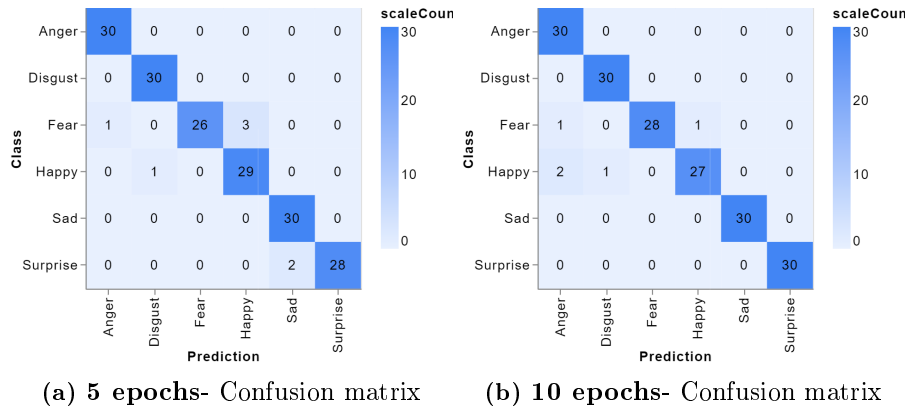


Figure 9. User 4 - Confusion matrix.

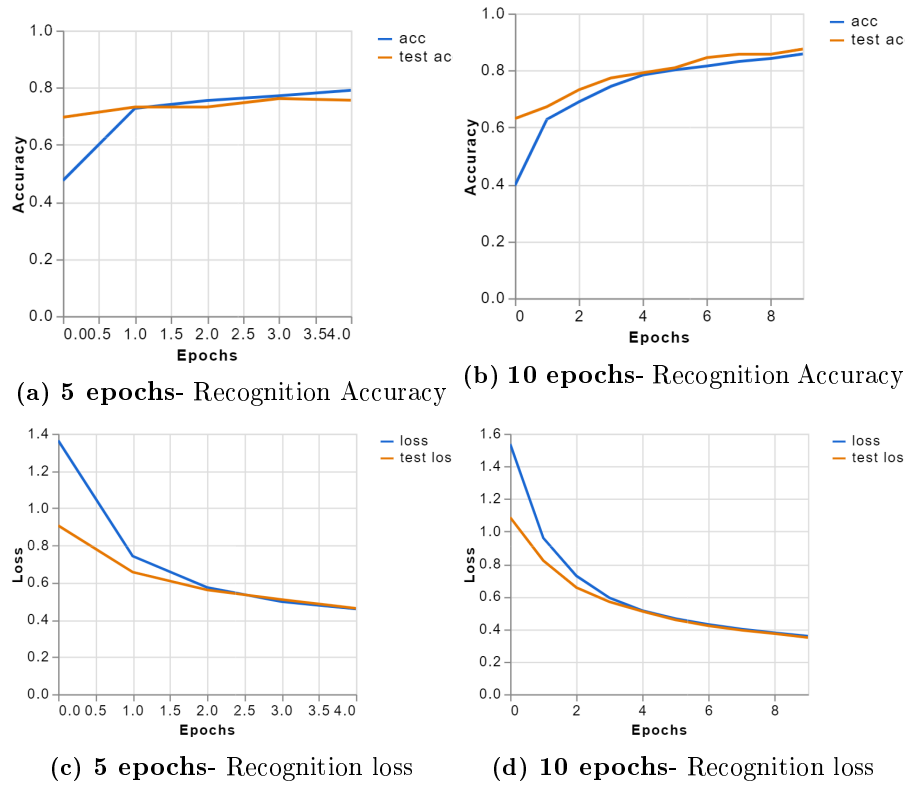


Figure 10. User 14 - Accuracy and loss.

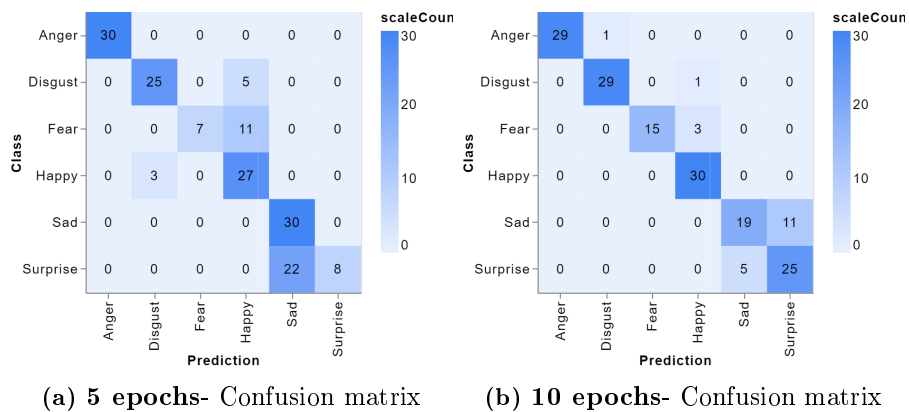


Figure 11. User 14- Confusion matrix

The results shown in figures above show that user 14 had slightly less accuracy than user 4, due to the fact that the user is wearing a glasses.

According to the results in Appendix 8.2.2, the results of facial expression personalization vary according to age, gender, and work; individuals over the age of 68 appear to exhibit less facial expression; we also observed a slight decrease in accuracy among users who wear glasses; and it is observed that women are better at expressing their emotion than men. However, Nationality didn't seem to be a valid factor to judge the results of facial expression recognition.

9. Conclusion

The data labeling and collecting for this experiment was critical. Twenty subjects responded to stimuli in a variety of ways; some expressed emotions that were more or less consistent with the emotions assigned to the stimuli; others had a few reactions that did not match the labels, causing slight confusion in the system and potentially resulting in ambiguous labels.

Given the dataset's size and available resources, MobileNet proved an excellent fit for this purpose. When models were trained for varying numbers of epochs, substantial differences in accuracy and loss were observed. The network's testing revealed that the models do indeed meet the requirements for Personalizing. We believe that the experiment established and the underlying concept was solid and that further effort should be made. To get the greatest outcomes, it is likely that many rounds of improvement will be necessary. Future research should focus on increasing the number of examples used in training as well as tackling more diverse examples and environments Future research should focus on increasing the number of the dataset and tackling more diverse users. As well as designing an embedded web application for the personalisation of Facial expression recognition.

Acknowledgement. The described article was carried out as part of the 2020-1.1.2-PIACI-KFI-2020-00165 "ERPA - Development of Robotic Process Automation solution for heavily overloaded customer services" project implemented with the support provided from the National Research, Development and Innovation Fund of Hungary, financed under the 2020-1.1.2-PIACI KFI funding scheme.

References

- [1] YUKI, M., MADDUX, W. W., and MASUDA, T.: Are the windows to the soul the same in the east and west? cultural differences in using the eyes and

- mouth as cues to recognize emotions in japan and the united states. *Journal of Experimental Social Psychology*, **43**(2), (2007), 303–311, URL <https://doi.org/10.1016/j.jesp.2006.02.004>.
- [2] JARRAYA, S. K., MASMOUDI, M., and HAMMAMI, M.: Compound emotion recognition of autistic children during meltdown crisis based on deep spatio-temporal analysis of facial geometric features. *IEEE Access*, **8**, (2020), 69311–69326, URL <https://doi.org/10.1109/access.2020.2986654>.
- [3] COHN, J. F., KRUEZ, T. S., MATTHEWS, I., YANG, Y., NGUYEN, M. H., PADILLA, M. T., ZHOU, F., and DE LA TORRE, F.: Detecting depression from facial actions and vocal prosody. In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, IEEE, 2009, pp. 1–7, URL <https://doi.org/10.1109/acii.2009.5349358>.
- [4] SUBRAMANIAN, R., WACHE, J., ABADI, M. K., VIERIU, R. L., WINKLER, S., and SEBE, N.: Ascertain: Emotion and personality recognition using commercial sensors. *IEEE Transactions on Affective Computing*, **9**(2), (2016), 147–160, URL <https://doi.org/10.1109/taffc.2016.2625250>.
- [5] ŻARKOWSKI, M.: Identification-driven emotion recognition system for a social robot. In *2013 18th International Conference on Methods & Models in Automation & Robotics (MMAR)*, IEEE, 2013, pp. 138–143, URL <https://doi.org/10.1109/mmar.2013.6669895>.
- [6] Wordpress.com: Create a free website or blog. <https://wordpress.com/>. Accessed: 2021-08-19.
- [7] Managed wordpress hosting - publish in minutes. <https://www.easywp.com/>. Accessed: 2021-08-19.
- [8] Pipe video recorder | addpipe.com. <https://addpipe.com/>. Accessed: 2021-08-19.
- [9] Contact form 7. <https://wordpress.org/plugins/contact-form-7/>. Accessed: 2021-08-20.
- [10] Fouzia adjailia. <https://fouziaadjailia.com/>. Accessed: 2021-08-20.
- [11] Voice maker. <https://voicemaker.in/>. Accessed: 2021-08-20.
- [12] HOWARD, A. G., ZHU, M., CHEN, B., KALENICHENKO, D., WANG, W., WEYAND, T., ANDREETTO, M., and ADAM, H.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Appendix 1

The following section represents the findings and further experiment results regarding the data collection and the classification results of personalization of facial expression recognition.

Let's begin

Full Name

Email

Country

Age
18-28

Note: please agree with the terms and conditions and with recording your face after starting the experiment


caution: Some videos may not be very pleasing and comfortable for some people, we apologize if some videos caused uncomfortable emotions

Agree

Figure 12. Beginning of the form [?].

First video

Add video recording to your website using Pipe



Do NOT Watch This Video AT NIGHT (SCARIEST VIDEO ON INTERNET)

Watch later Share

Watch on YouTube

What emotion did the video trigger?

Happiness Anger Fear Disgust Surprise Sadness Neutral

Figure 13. First block of the form [?].

Table 3. Emotion triggers database

Emotion	Video chosen to trigger the emotion
Fear	https://www.youtube.com/watch?v=CufJDBvrECA&ab_channel=AkshayRathee
Happy	https://www.youtube.com/watch?v=v-Q5pYIipZg&ab_channel=FunnyPets
Disgust	https://www.youtube.com/watch?v=Juyv2es0i7o&ab_channel=SatisfyingASMRhd
Anger	https://www.youtube.com/watch?v=JGqkXcWcqlg&ab_channel=WPLGLocal10
Surprise	https://www.youtube.com/watch?v=CsWiZpMdfLs&ab_channel=Newsflare
Sad	https://www.youtube.com/watch?v=Fq7I0ooU3Ps&ab_channel=EmotionalCulture

Table 4. Input description.

User	Text input					Video input						
	Fear	Happiness	Disgust	Anger	Surprise	Sadness	Fear	Happiness	Disgust	Anger	Surprise	Sadness
1	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
2	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
3	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
4	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
5	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
6	yes	yes	yes	yes	yes	yes	no	no	no	no	no	no
7	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
8	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
9	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
10	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
11	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
12	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
13	yes	yes	yes	yes	yes	yes	no	no	no	no	no	no
14	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
15	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
16	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
17	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes	yes
18	yes	yes	yes	yes	yes	yes	no	no	no	no	no	no
19	yes	yes	yes	yes	yes	yes	no	no	no	no	no	no
20	no	no	no	no	no	no	yes	yes	yes	yes	yes	no

User 1

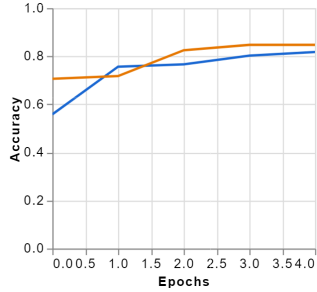


Figure 14.5 epochs- Recognition Accuracy

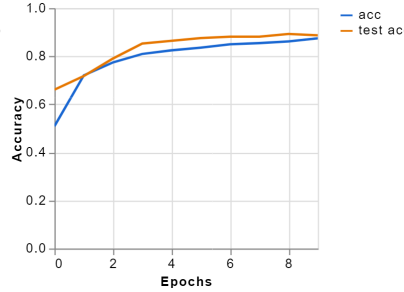


Figure 15.10 epochs- Recognition Accuracy

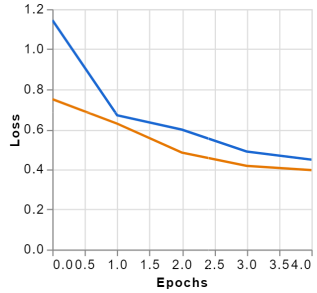


Figure 16.5 epochs- Recognition loss

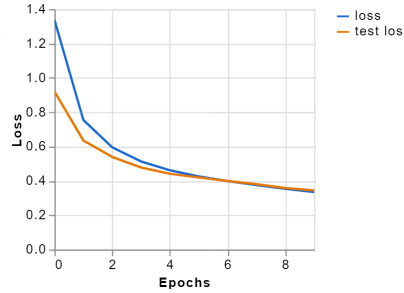


Figure 17.10 epochs- Recognition loss

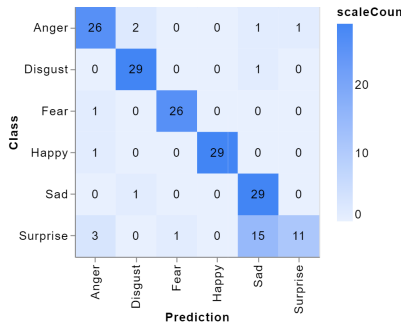


Figure 18.5 epochs- Confusion matrix

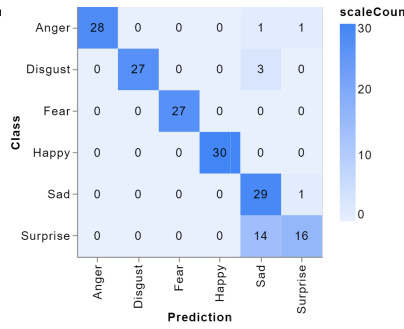


Figure 19.10 epochs- Confusion matrix

User 2

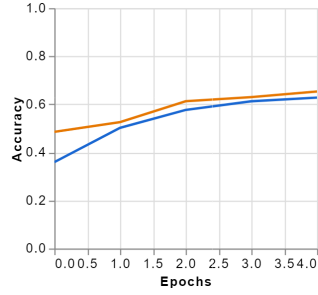


Figure 20. 5
epochs- Recognition Accuracy

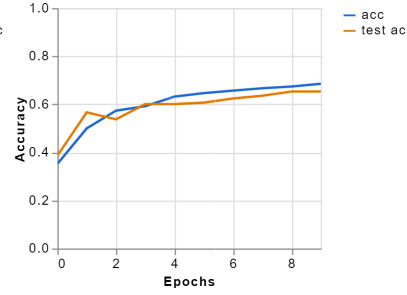


Figure 21. 10
epochs- Recognition Accuracy

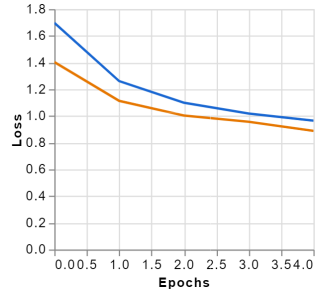


Figure 22. 5
epochs- Recognition loss

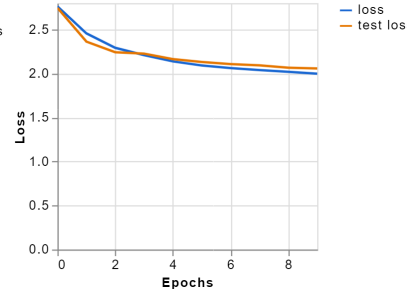


Figure 23. 10
epochs- Recognition loss

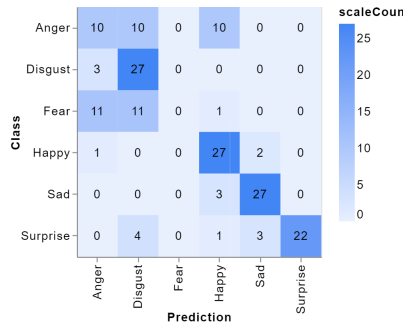


Figure 24. 5
epochs- Confusion matrix

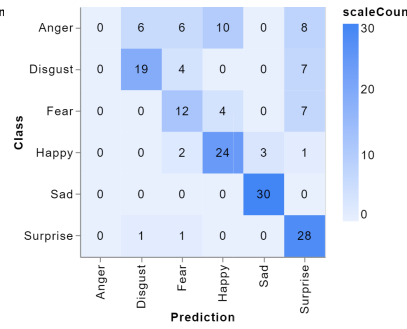


Figure 25. 10
epochs- Confusion matrix

User 3

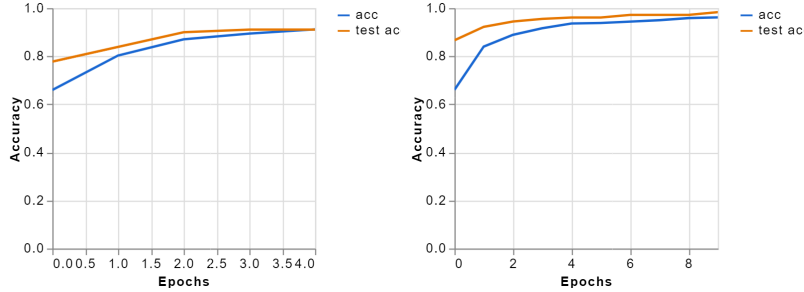


Figure 26. 5 epochs- Recognition Accuracy

Figure 27. 10 epochs- Recognition Accuracy

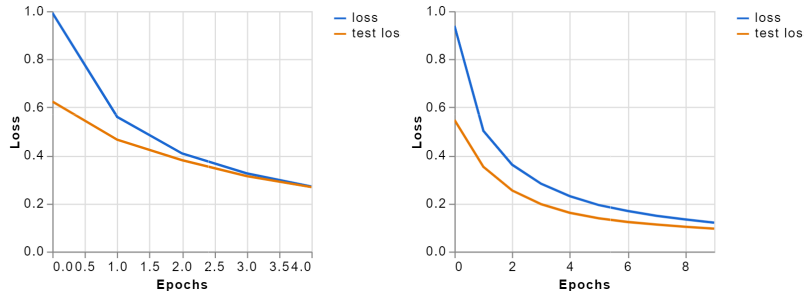


Figure 28. 5 epochs- Recognition loss

Figure 29. 10 epochs- Recognition loss

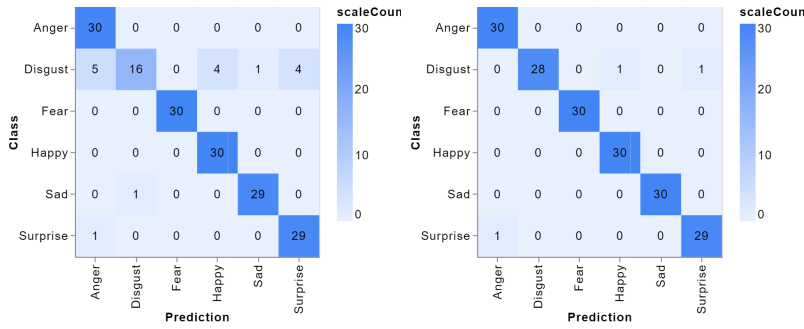


Figure 30. 5 epochs- Confusion matrix

Figure 31. 10 epochs- Confusion matrix

User 4

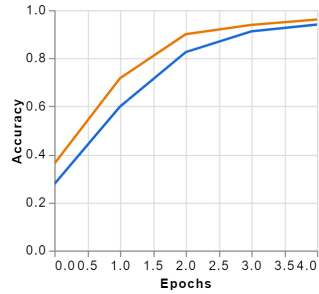


Figure 32. 5
epochs- Recognition Accuracy

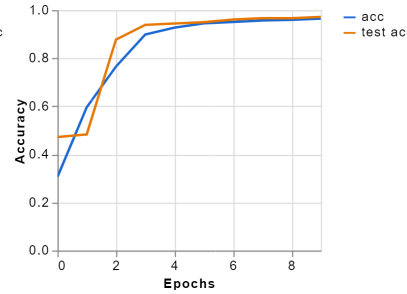


Figure 33. 10
epochs- Recognition Accuracy

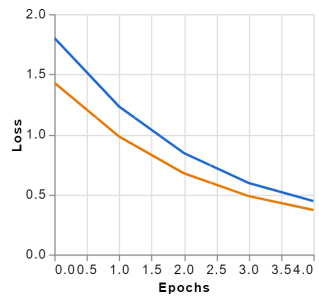


Figure 34. 5
epochs- Recognition loss

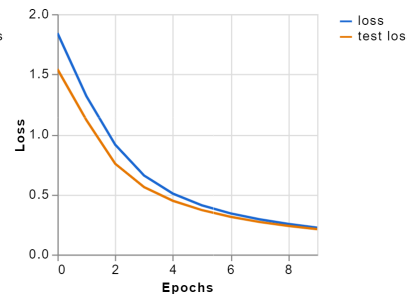


Figure 35. 10
epochs- Recognition loss

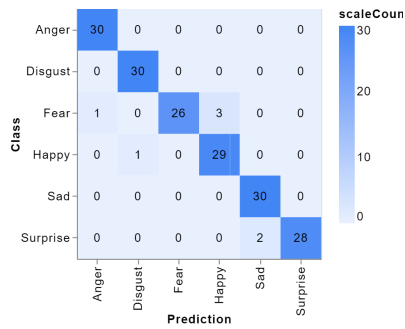


Figure 36. 5
epochs- Confusion matrix

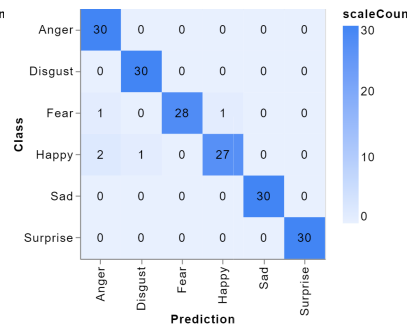


Figure 37. 10
epochs- Confusion matrix

User 5

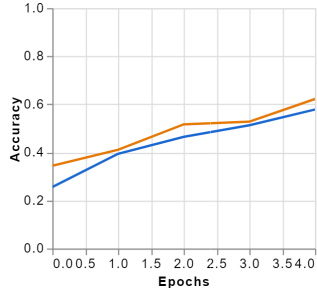


Figure 38. 5
epochs- Recognition Accuracy

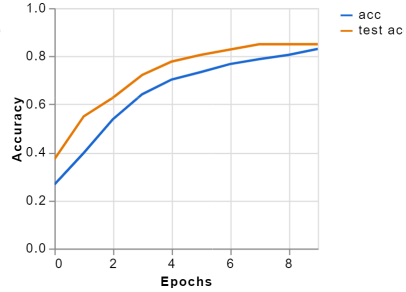


Figure 39. 10
epochs- Recognition Accuracy

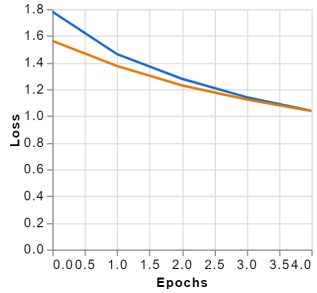


Figure 40. 5
epochs- Recognition loss

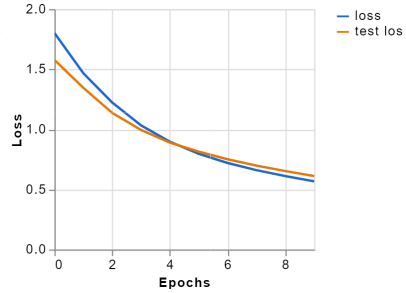


Figure 41. 10
epochs- Recognition loss

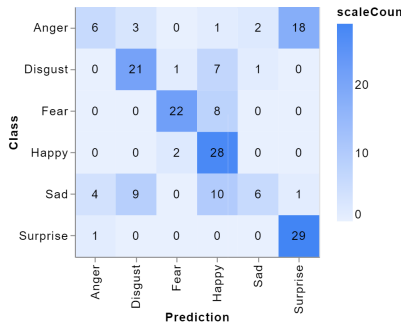


Figure 42. 5
epochs- Confusion matrix

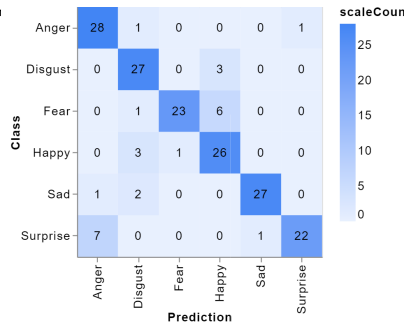


Figure 43. 10
epochs- Confusion matrix

User 6

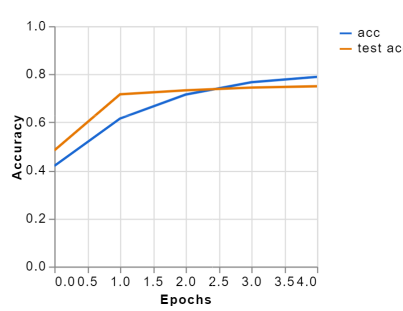


Figure 44.5
epochs- Recognition Accuracy

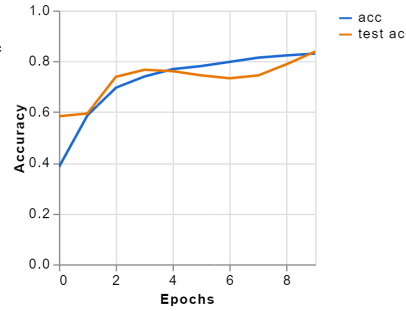


Figure 45.10
epochs- Recognition Accuracy

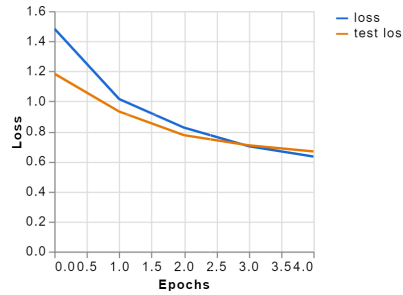


Figure 46.5
epochs- Recognition loss

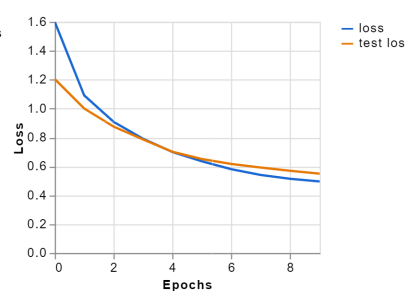


Figure 47.10
epochs- Recognition loss

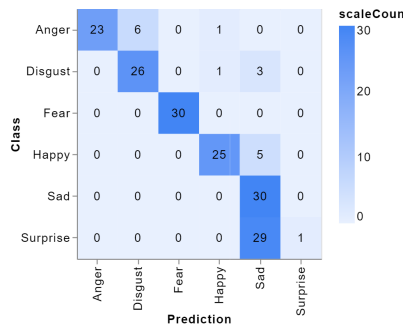


Figure 48.5
epochs- Confusion matrix

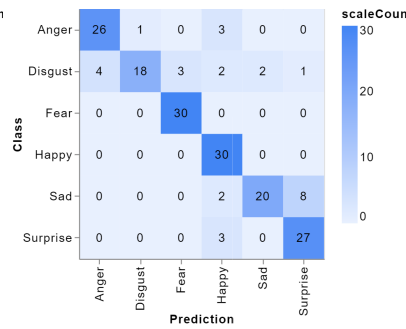


Figure 49.10
epochs- Confusion matrix

User 8

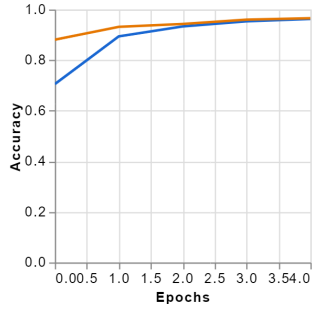


Figure 50. 5 epochs- Recognition Accuracy

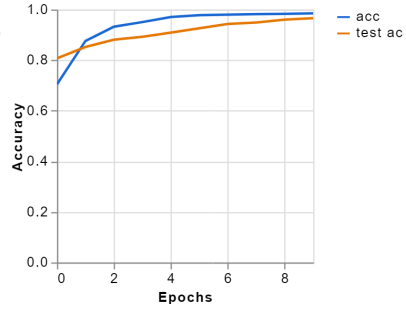


Figure 51. 10 epochs- Recognition Accuracy

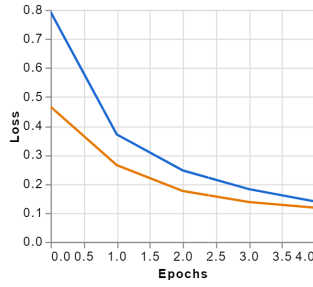


Figure 52. 5 epochs- Recognition loss

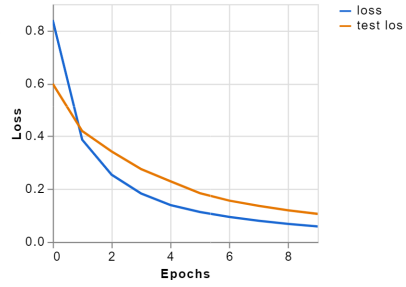


Figure 53. 10 epochs- Recognition loss

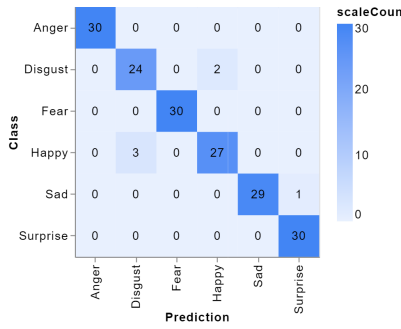


Figure 54. 5 epochs- Confusion matrix

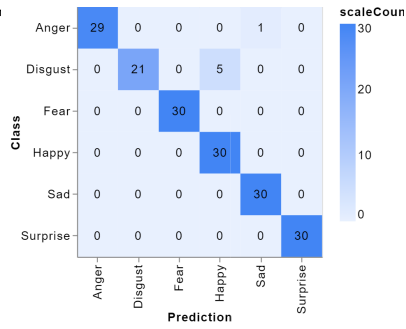


Figure 55. 10 epochs- Confusion matrix

User 9

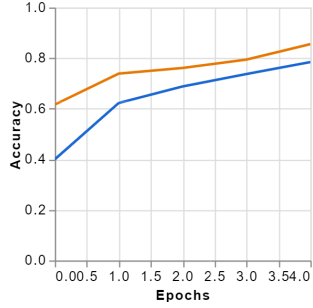


Figure 56.5 epochs- Recognition Accuracy

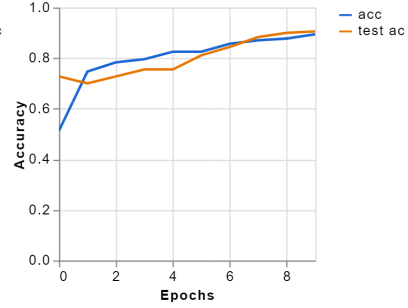


Figure 57.10 epochs- Recognition Accuracy

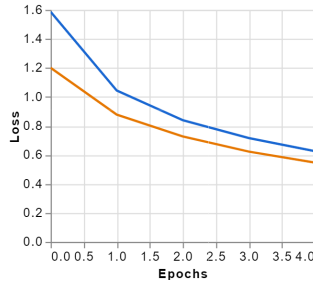


Figure 58.5 epochs- Recognition loss

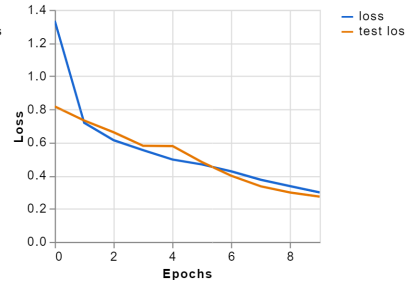


Figure 59.10 epochs- Recognition loss

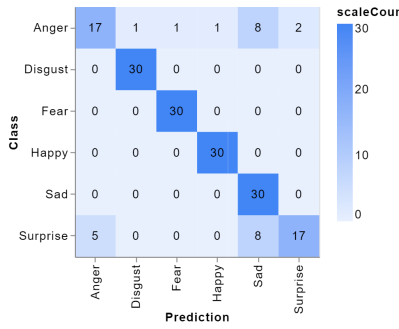


Figure 60.5 epochs- Confusion matrix

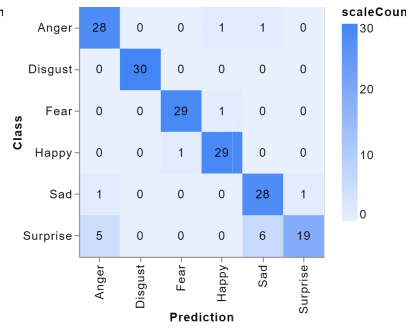


Figure 61.10 epochs- Confusion matrix

User 10

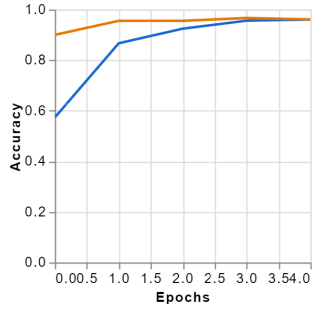


Figure 62.5 epochs- Recognition Accuracy

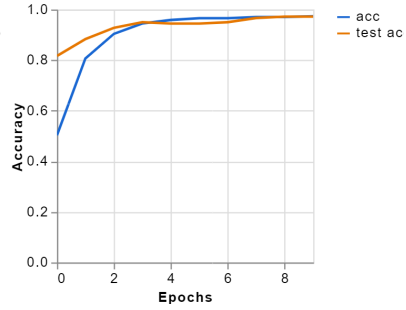


Figure 63.10 epochs- Recognition Accuracy

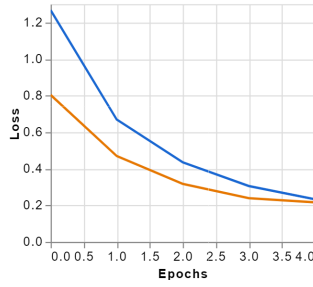


Figure 64.5 epochs- Recognition loss

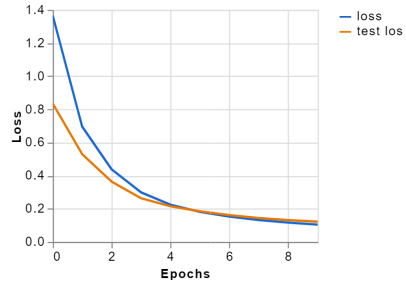


Figure 65.10 epochs- Recognition loss

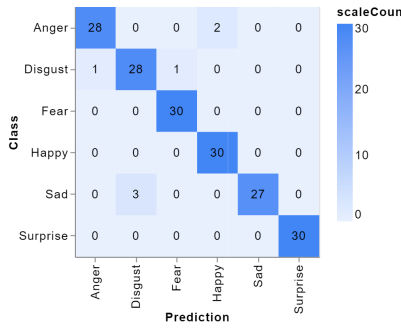


Figure 66.5 epochs- Confusion matrix

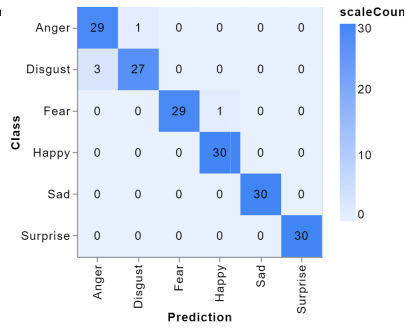


Figure 67.10 epochs- Confusion matrix

User 11

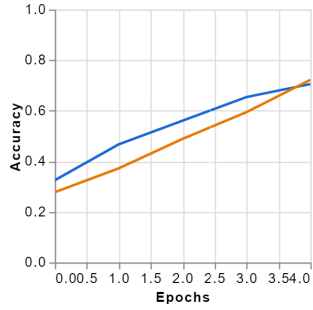


Figure 68. 5
epochs- Recognition Accuracy

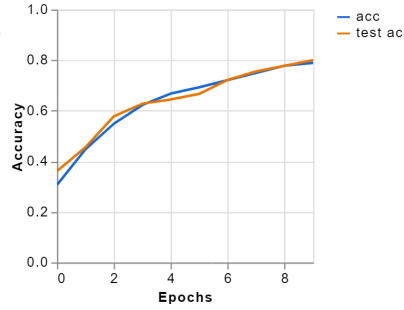


Figure 69. 10
epochs- Recognition Accuracy

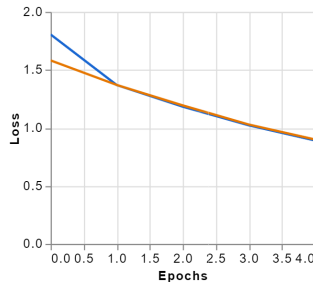


Figure 70. 5
epochs- Recognition loss

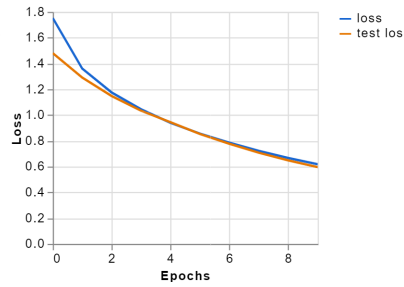


Figure 71. 10
epochs- Recognition loss

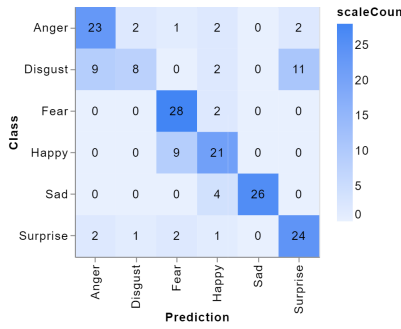


Figure 72. 5
epochs- Confusion matrix

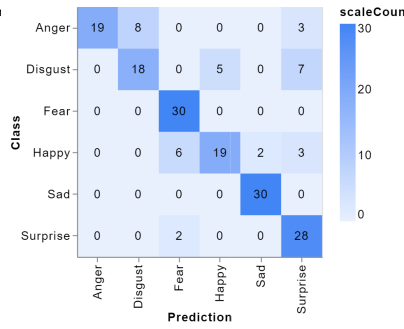


Figure 73. 10
epochs- Confusion matrix

User 12

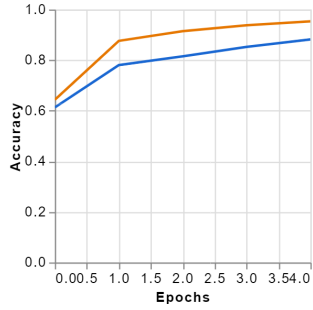


Figure 74.5 epochs- Recognition Accuracy

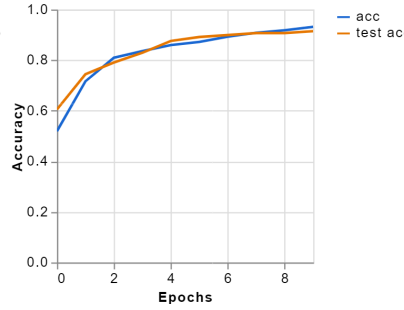


Figure 75.10 epochs- Recognition Accuracy

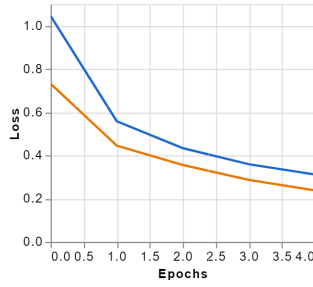


Figure 76.5 epochs- Recognition loss

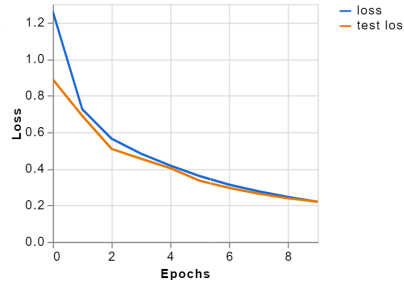


Figure 77.10 epochs- Recognition loss

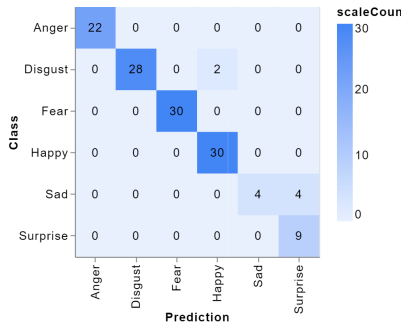


Figure 78.5 epochs- Confusion matrix

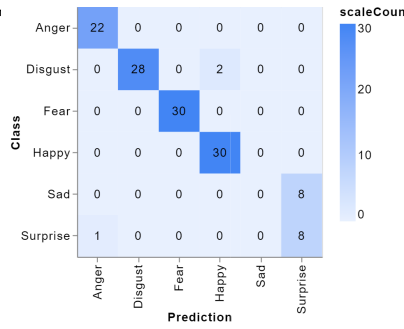


Figure 79.10 epochs- Confusion matrix

User 14

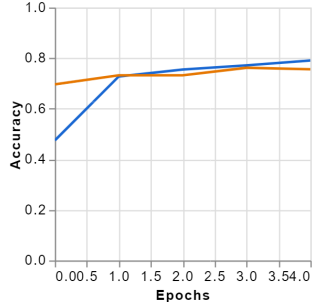


Figure 80.5
epochs- Recognition Accuracy

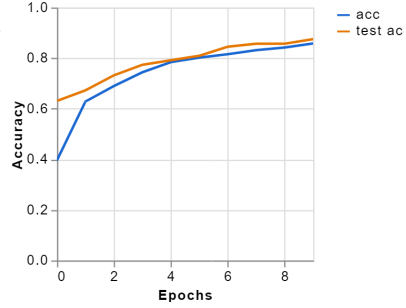


Figure 81.10
epochs- Recognition Accuracy

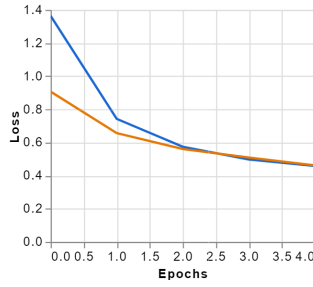


Figure 82.5
epochs- Recognition loss

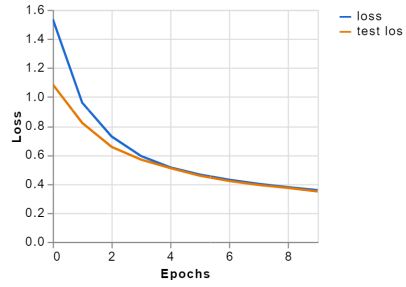


Figure 83.10
epochs- Recognition loss

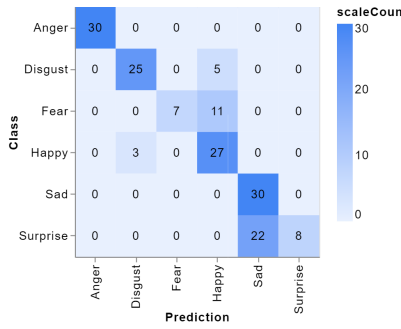


Figure 84.5
epochs- Confusion matrix

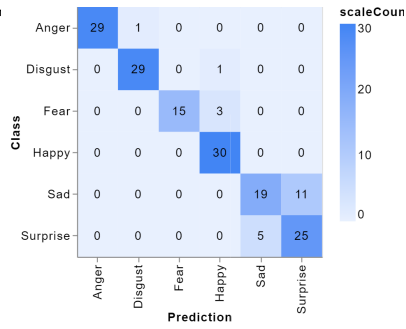


Figure 85.10
epochs- Confusion matrix

User 15

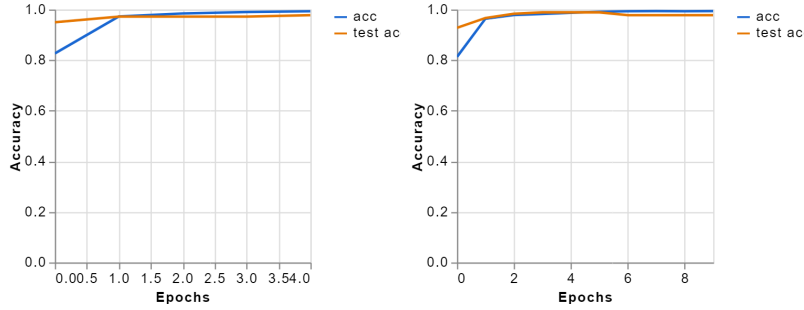


Figure 86.5
epochs- Recognition Accuracy

Figure 87.10
epochs- Recognition Accuracy

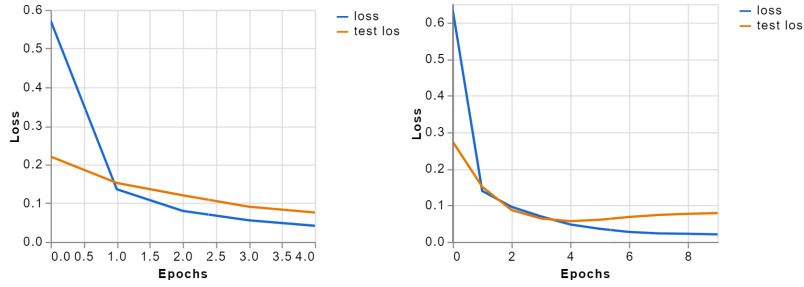


Figure 88.5
epochs- Recognition loss

Figure 89.10
epochs- Recognition loss

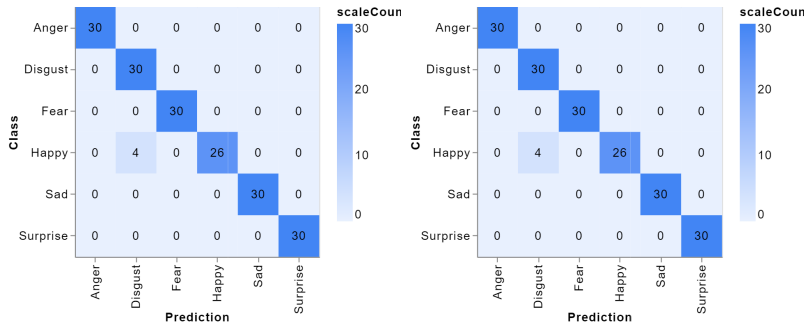


Figure 90.5
epochs- Confusion matrix

Figure 91.10
epochs- Confusion matrix

User 16

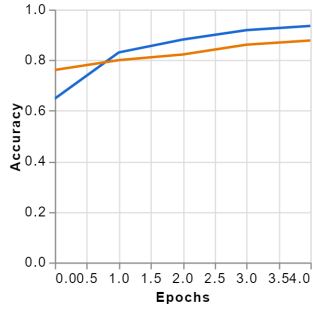


Figure 92.5
epochs- Recognition Accuracy

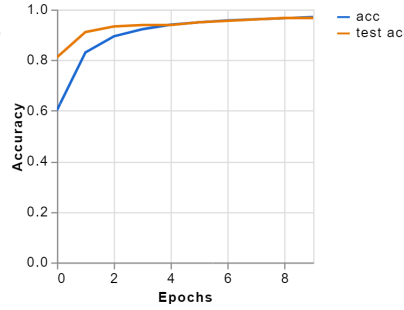


Figure 93.10
epochs- Recognition Accuracy

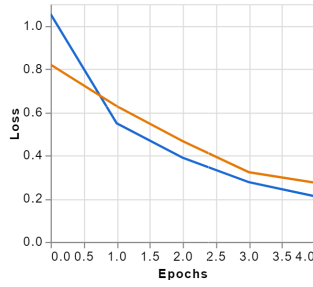


Figure 94.5
epochs- Recognition loss

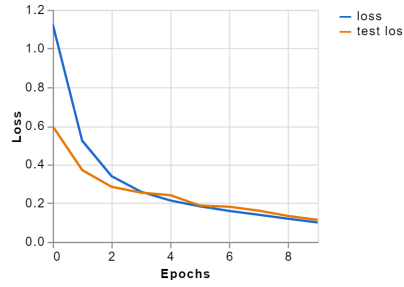


Figure 95.10
epochs- Recognition loss

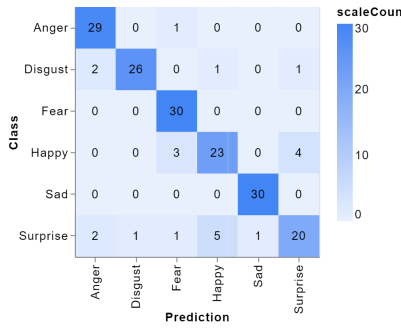


Figure 96.5
epochs- Confusion matrix

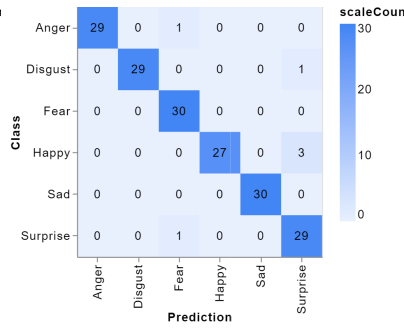


Figure 97.10
epochs- Confusion matrix

User 17

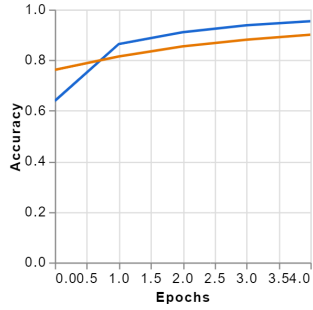


Figure 98.5 epochs- Recognition Accuracy

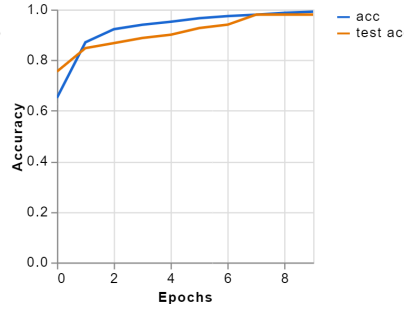


Figure 99.10 epochs- Recognition Accuracy

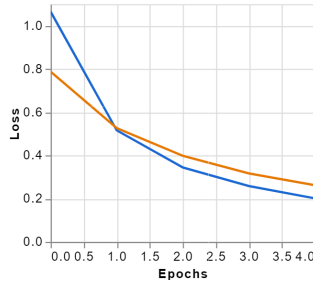


Figure 100.5 epochs- Recognition loss

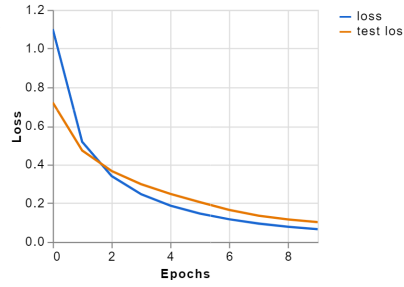


Figure 101.10 epochs- Recognition loss

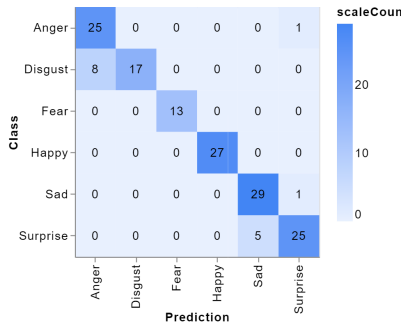


Figure 102.5 epochs- Confusion matrix

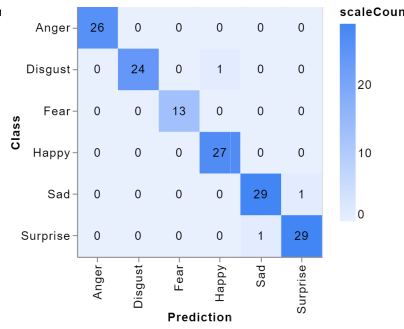


Figure 103.10 epochs- Confusion matrix

User 20

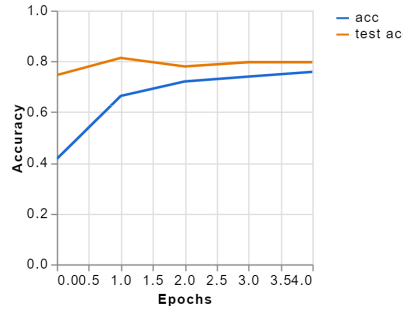


Figure 104.5
5 epochs- Recognition Accuracy

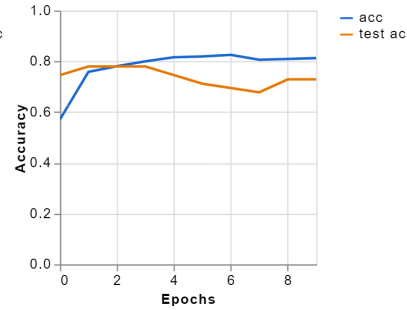


Figure 105.10
10 epochs- Recognition Accuracy

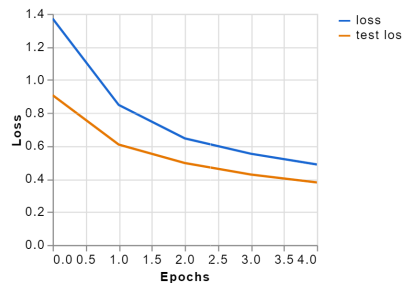


Figure 106.5
5 epochs- Recognition loss

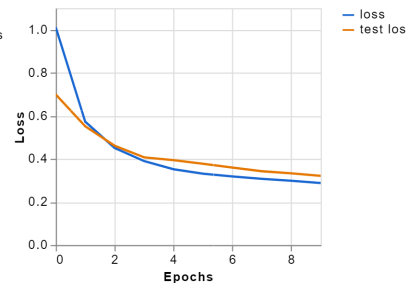


Figure 107.10
10 epochs- Recognition loss

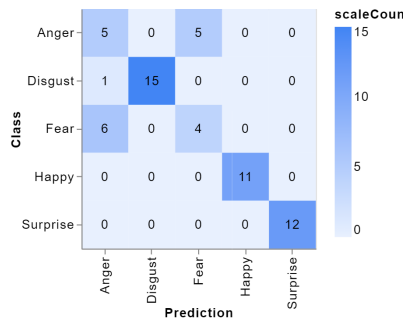


Figure 108.5
5 epochs- Confusion matrix

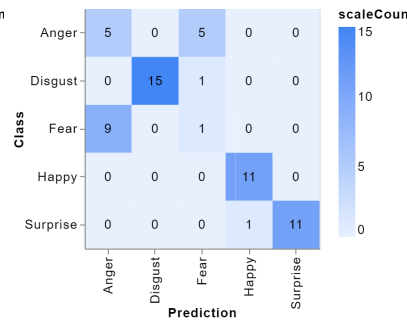


Figure 109.10
10 epochs- Confusion matrix