MULTIMODAL SYSTEM FOR FACIAL EMOTION RECOGNITION BASED ON DEEP LEARNING

Atanas Atanassov, Dimitar Pilev, Fani Tomova

University of Chemical Technology and Metallurgy,
8 Kliment Ohridski Blvd., Sofia 1797, Bulgaria
E-mail: naso@uctm.edu

Received: 09 July 2023
Accepted: 24 October 2023
DOI: 10.59957/jctm.v59.i3.2024.29

ABSTRACT

Emotions are one of the main ways of communication between people and of expressing attitudes towards objects, products, services, etc. They are divided to verbal and non-verbal classes. Human speech and intonation belong to the first class, and to the second (non-verbal) facial and body emotions, known as body language. The subject of this report is the development of multimodal deep learning system intended to recognize facial and body emotions and their relationship with the scene (weather) in which they occur. It is based on three deep learning neural networks (DNN) each one for recognition of facial emotion, body emotion and weather. Combining their results, we improve significantly the final facial emotion recognition (FER) results.

Keywords: facial emotion recognition, deep learning neural network, body gesture recognition, weather recognition.

INTRODUCTION

The relationship between facial emotions, body emotions, and the weather conditions is complex and interconnected. Each component plays a role in shaping and influencing the others, and together they contribute to our overall emotional experiences. Facial emotions are a primary means of nonverbal communication, and they play a crucial role in expressing and conveying emotions. When we experience an emotion, such as happiness, sadness, anger, or fear, our facial muscles respond by forming specific expressions associated with those emotions. Body Emotions are not solely expressed through the face but also through the body. Our body posture, gestures, movements, and overall bodily sensations contribute to the expression and experience of emotions. For example, when we feel confident, we may stand tall, with an upright posture and open gestures. Conversely, when we feel sad or defeated, we may slump our shoulders and adopt a closed-off posture. Bodily sensations, such as a racing heart or tense muscles, can also accompany emotions and provide feedback to our brain about our emotional state. Weather conditions can also influence our emotions and interact with facial and body expressions in various ways. For example, the lack of sunlight can affect our mood, leading to symptoms such as sadness, low energy, and a decreased interest in activities. Conversely, during sunny and warm weather, people may feel more energetic, cheerful, and motivated.

According to Ekman Facial Action Coding System as given in Fig. 1, there are seven basic emotions (fear, sadness, happiness, surprise, anger, disgust and neutral) [1]. Proposed by Russell 3D Valence Arousal Dominance model (VAD) given in Fig. 2, extends them to 26 continues emotions [2, 3].

The body language is related to the position of the head, positions of the arms, hands and fingers, and positions of the legs, as well as the position of the torso. For example (Fig. 3), the body gestures corresponding to the facial emotion surprise are related to lifting one or both hands towards the head or touching the head, mouth or face with one or both hands.

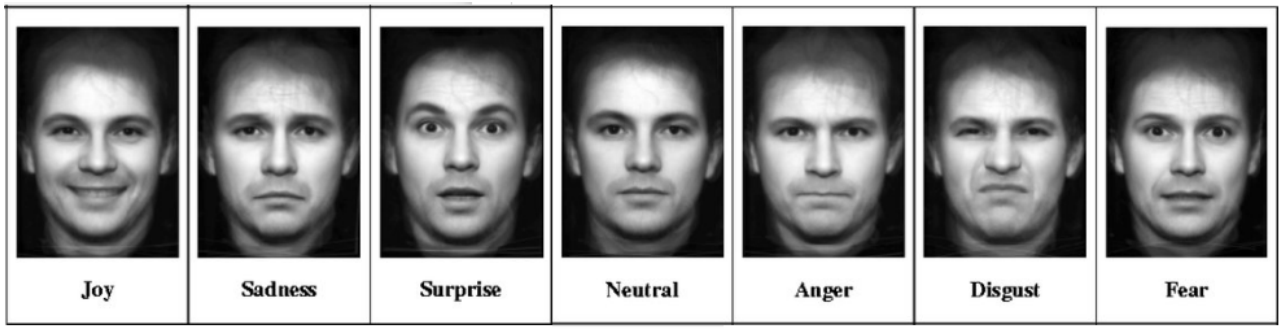

Fig. 1. Seven basic facial emotions identified by P. Ekman.

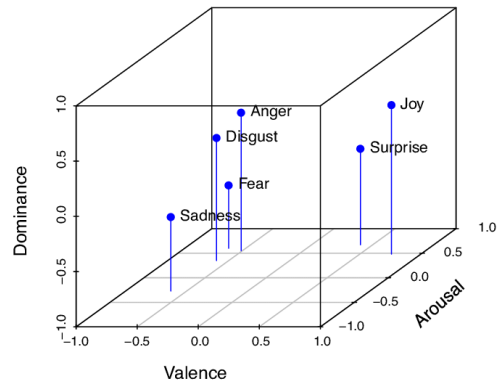

Fig. 2. Russell's 3D Valence Arousal Dominance model.

Fig. 3. Body emotion surprise.

A number of studies have found that there is a strong correlation between a person's emotions and the place, weather conditions (weather) and other objects such as other people, landscape, etc. [4 - 7]. Weather conditions are closely related to geographic regions and vary in some cases from sunny to cloudy, or in other cases include some subset of sunny, foggy, snowy, rainy, hot, etc.

In this paper, we propose a multimodal system for Facial Emotion Recognition /FER/ based on 3 deep learning neural networks /DNN/. First DNN recognizes facial emotions, second recognizes body emotion and the third weather conditions recognition. Combination of the results produced by the three DNNs improves significantly the final FER results.

EXPERIMENTAL

Choosing a DNN for facial emotion recognition

The selection of DNNs for facial emotion recognition is presented in our ICAI 2020 publication [8]. In it, we have analyzed many multiple neural networks such as DeepFace, OpenFace, VGG16, VGG19, Deepid, Resnet, Facenet, etc in order to select appropriate pre-trained neural network for FER. Mentioned DNNs are used for face recognition and provide very high accuracy up to 99.63 % as given in Table 1. For facial emotion recognition, these DNNs are trained with emotions datasets as FER-2013, KDEF, MUG, RAfD, CK+, etc. and obtained maximum accuracy is up to 70 %. Selected by us FER model has an architecture, which is close to DeepFace and Resnet18 and has five 2D convolutional groups, given in Fig. 4, with a total of 17 convolutional layers, each group ending with pooling and dropout layers.

Between the convolutional layers of each group there is a batch normalization layer (Batch normalization). The input data is a black and white image in 48x48 pixel format, and the output of the network is the recognized 7 emotions (angry, disgust, fear, joy, sadness, surprise and neutral). The number of pre-trained weight coefficients of the selected DNN model is 13111367, and it is trained with the FER2013 dataset. Its accuracy for this data set is 69.85 %.

Choosing a DNN for body emotion recognition

In our paper presented in ICAI21 conference, we analyzed most existing DNN models for body's gestures recognition, again for finding suitable pre-trained BER DNN model [9]. Most of analyzed models were based on ResNet18 and ResNet50 architectures. Some of them used bimodal DNN models combining FER and BER recognition. The analysis shows that the fusion of facial and body emotion recognition increases the total FER accuracy (from 5 to 15 %) of bimodal DNNs for emotions recognition. Mentioned DNNs are well described but only for few of them the pre-trained weights parameters were available.

Selected by us pre-trained DNN model used for body emotion recognition is based on adapted version of ResNet 18 [10]. It contains one input layer providing

Table 1. The Accuracy of most used Deep Learning Networks for FER.

Model	Parameters	Image size	Output	Accuracy, %
VGG-Face	145,002,878	224×224×3	2622	90.5
Facenet	22,808,144	160×160×3	128	99.63
OpenFace	3,743,280	96×96×3	128	97.3
DeepFace	137,774,071	152×152×3	8631	97.35
DeepId	400,000	55×47×3	160	96.05

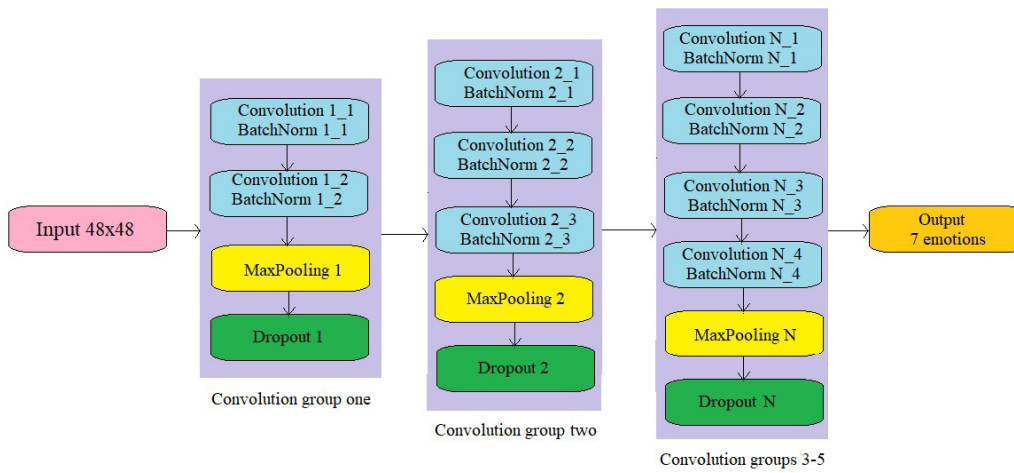

Fig. 4. Architecture of DNN for Facial Emotion Recognition.

112x112 pixel RGB images, 21 convolutional groups composed of 2D convolutional layer with Batch normalization and ReLU layers. Last layer is Adaptive Average Pooling. Only the first convolutional group includes at the end additional Max-pooling layer. The output of the FC layer are 26 emotions and 3 continuous VAD dimensions. The model weight parameters are 11363757. It is trained with EMOTIC dataset and produces 77.6 % accuracy.

Choosing a DNN for weather recognition

In our paper, published in ICAI22 we analyzed suitable pre-trained DNN models for recognition of weather conditions and especially conditions that affect human facial emotions. As mentioned above, conditions such as sunny, rainy, snowy, cloudy or hot weather are strongly related to facial emotions. Many of the meteorological DNNs discussed below are oriented towards recognizing conditions that are important for urban or road traffic regulation, for driving autonomous or semi-autonomous vehicles, for smart homes and cities, for agriculture, etc. These deep neural networks recognize many different weather phenomena (fog, smog, dew, sandstorm, icing, snow, hail, rain, frost, lightning, rainbow, clouds, etc.) [11, 12]. These DNNs have been analyzed in terms of their accuracy in recognizing weather phenomena related to facial emotions, the availability of the datasets used for their training, the memory size of the models and their respective weights, also what computing power (CPU, GPU, memory, etc.) is required to use them. Some of the most common deep neural networks for weather recognition and classification are EfficientNet, MobileNet, Xception, InceptionResNetV2, ResNet50, ResNet101, DenseNet201, VGG16, VGG19, GoogLeNet, AlexNet, etc. Their accuracy in recognizing different classes of weather conditions varies according to the number of classes [12]. When recognizing 3-4 classes, their accuracy reaches from 92 % to 98.20 %, and with 6 or 9 classes, it decreases to 81.20 %, which is understandable - as the recognition classes increase, the accuracy decreases. It also depends on the datasets used for their training and the specific meteorological phenomena belonging to the respective classes. Table 2 presents the accuracy of some of the aforementioned DNN models, trained with the WEAPD dataset containing 6877 images of weather phenomena as rain, hail, rainbow, snow, thunder, dew, sandstorm, frost, smog, frost and icing [13].

Table 2. The accuracy of most used DNN models for weather recognition.

DNN model	Accuracy, %	Coefficients
DenseNet201	84	74850304
ResNet50	83	94781440
MobileNet	83	17235968
ResNet101	81	171458560
VGG19	79	80150528
EffcientNetB7	77	258080768
Xception	76	83697664
InceptionResNetV2	73	219070464

From the table, it can be seen that the DenseNet201 model has the highest accuracy, followed by ResNet50, MobileNet and ResNet101, and the accuracy does not directly depend on the number of weight coefficients of the corresponding neural networks.

Several DNN models presented in are based on ResNet18, ResNet15 or combination of ResNet50, EfficientNet-b0 and SqueezeNet models are used to recognize four weather conditions (cloudy, rainy, sunny, and sunrise) obtain up to 98.22 % accuracy [14 - 16]. The accuracy of MeteCNN [17] which was designed to classify 11 weather conditions is 92.68 % and it is comparable to the results of ResNet50, MobileNet, and DenseNet given in Table I MeteCNN is a modified and optimized version of the convolutional neural network VGG16, which was discussed in [8] and was used by us for facial emotion recognition.

As can be seen from the analysis, one of the most used pre-trained DNN models is ResNet and its versions 50, 15, 18, 101, etc., because they have very high accuracy in recognizing different numbers of meteorological classes. Pre-trained versions of some of these models, as well as some of the datasets on which they are trained, are freely available for research.

Our choice of pre-trained weather recognition model came down to a version of ResNet50. It was selected as the third DNN model in our multimodal system. This version of ResNet50 has almost the same architecture as that of Fig. 5. The differences between our version of ResNet50 and the original are small and are related to the

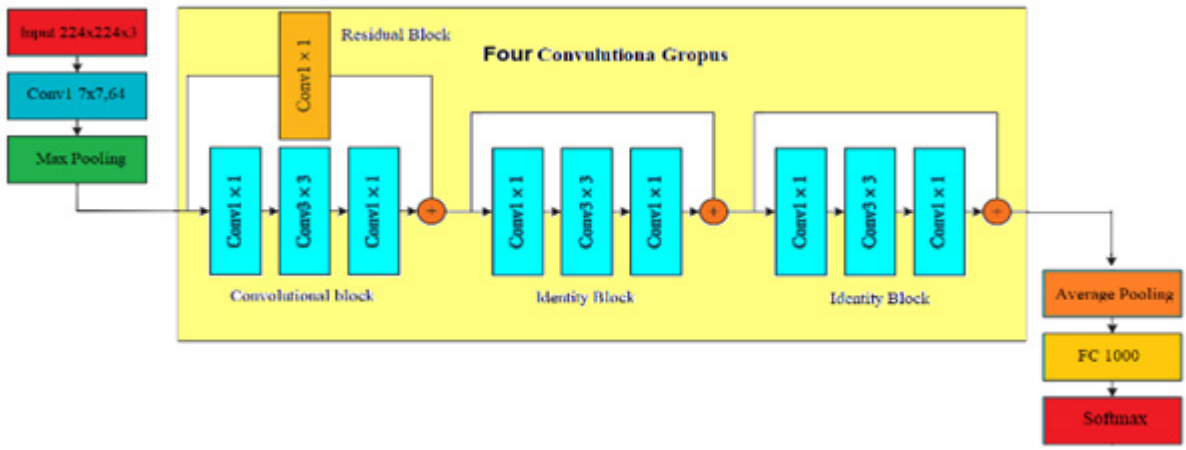

Fig. 5. The architecture of ResNet 50.

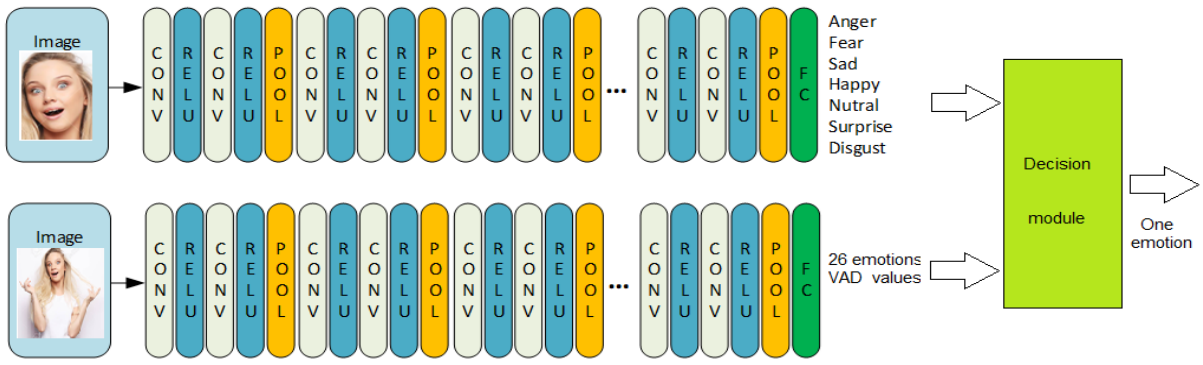

Fig. 6. Bimodal FER and BER models.

size (shape) of the input image and the recognized output classes, as well as in the additional normalization layers (Batch normalization) included after each convolutional layer in all convolutional groups and identity groups. The input images are in RGB format with a size of 100x100 pixels, and the output weather classes are snowy, rainy, sunny, cloudy, and hot weather. The two fully connected layers at the end of the network transform the output after the last convolutional layer from 2048 to 100 classes, and then from 100 to 5 meteorological classes.

Architecture of multimodal system

On the base of selected FER and BER DNN models, we constructed following bimodal system presented in Fig. 6. It combines the results from both DNNs. The results from both models are compared in a decision module and when they are similar or equal with more confidence we can state they are true. Our FER model provides as result seven emotions, but selected BER model recognizes 26 emotions, as well as VAD values. First, the decision module searches for equality or similarity between the emotions produced by the two models.

For example, to happy emotion from the seven emotions of the FER group correspond happiness, pleasure and peace of the BER group. Alternatively, of disgust emotion, correspond annoyance, disapproval, etc. If there is a match between the FER and BER emotions, the dominant facial emotion is taken as a final result. If there is no match as a final result the FER result is taken. Second, the decision module uses values of Valence, Arousal, and Dominance (see Table 3, row Happy) to detect if there is a match for the BER emotion Happy within the interval of plus/minus 0.05 % of the mentioned VAD values. The pseudocode of the solution module for FER emotion Happy has the following structure:

IF (FER_Happy IN (BER_Happiness, BER_Pleasure, BER_Peace))
{ Final_Result = FER_Happy}
ELSE
IF ((BER_V >= 0.95*0.76) AND (BER_V<= 1.05*0.76) AND
(BER_A>=0.95*0.48) AND (BER_A<=1.05*0.48) AND
(BER_D>=0.95*0.35) AND (BER_D<=1.05*0.35))
{ Final_Result = FER_Happy}
ELSE { Final_Result = FER_Happy}

Next the result produced from the bimodal system is taken as input to third DNN model for weather recognition. The architecture of the proposed multimodal system is given in Fig. 7. It includes the selected deep neural networks for facial emotion recognition (FER DNN) and for weather conditions (ResNet50). The first neural network FER DNN receives as input a black and white face image in 48x48 pixel format and predicts each of the seven facial emotions mentioned above with a certain accuracy. The second BER DNN takes 112x112 pixel RGB mage and provides as result 26 body emotions as well VAD results. Third DNN - ResNet50 takes a 100x100 pixel RGB image of people in a natural setting and predicts each of the five weather conditions in the same way.

On the basis of the two vectors with the predicted emotions and the weather, and on the basis of the

Table 3. Relationship between Ekman's seven basic emotions and their counterparts in Russell's VAD system.

Emotion	Valence	Arousal	Dominance
Angry	-0.43	0.67	0.34
Happy	0.76	0.48	0.35
Surprise	0.4	0.67	-0.13
Digust	-0.6	0.35	0.11
Fear	-0.64	0.6	-0.43
Sad	-0.63	0.27	-0.33
Neutral	0.0	0.0	0.0

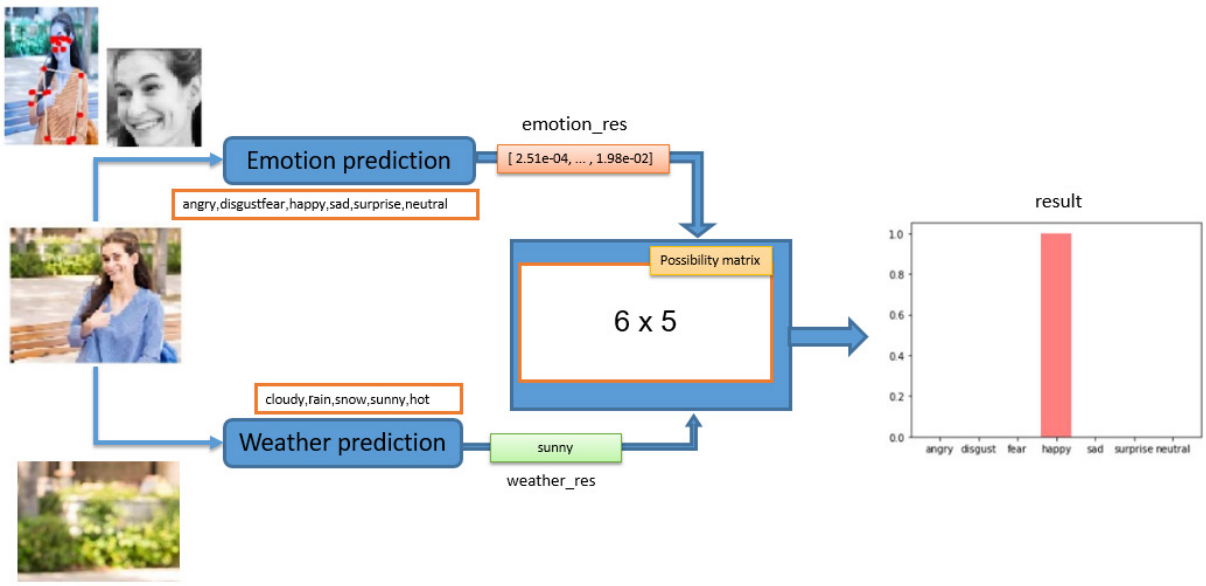

Fig. 7. Architecture of multimodal FER system.

previously formed probability matrix of (Possibility Matrix /PM/), the generalized result for the recognized facial emotions is formed. The probability matrix, presented in Fig. 8 has five rows corresponding to the five weather conditions and seven columns corresponding to seven facial emotions.

The Resnet50 model returns a number (the predicted weather) that is used as an index defining the row of the probability matrix. This row is multiplied by the vector returned by the FER DNN model. In this way, the weather-dependent weighting coefficients are added to the FER coefficients, which increases the accuracy of the model. The coefficients of each row of the matrix depend on the recognized weather. The sum of the weight coefficients for each row is 1. The coefficients themselves were determined empirically, using facial emotion and weather data from several thousand labeled photos from family albums, photos from our private data set of events held at UCTM-Sofia (conferences, awarding of diplomas, opening of the academic year, etc.) in the last 4 years. For example, row one of the probability matrix shows that in cloudy weather, 20 % of the people in the processed photos are angry, 3 % are disgusted, 1 % are feared, 20 % are happy, 25 % are sad, 3 % are surprised, and 28 % are in neutral condition.

Including the weather results improves the final emotion recognition results. For example in Fig. 9, the accuracy of the happy facial emotion increases from 60 % to 73 % after applying recognized (snowy) weather from the second model, and the other emotions of the FER model are suppressed by the second model, so the final result is the emotion happiness.

The ResNet50 model for weather recognition was further trained with the WEAPD dataset and then tested and verified with images from our private dataset (presented in Fig. 11), discussed in the next section of the report. The accuracy of the model is given on the left side of Fig. 10. Training accuracy is 91.4 % and validation accuracy is 82.5 %. The confusion matrix (right side of Fig. 10.) shows how accurately the weather conditions were predicted. As can be seen, hot, snowy, and rainy weather are predicted with 100 %, 98 %, and 95 %, while sunny and cloudy are predicted with 88 % and 75 %.

```
[
  [0.18, 0.03, 0.01, 0.22, 0.25, 0.03, 0.28],
  [0.09, 0.03, 0.05, 0.08, 0.2, 0.25, 0.3],
  [0.14, 0.01, 0.05, 0.30, 0.1, 0.18, 0.22],
  [0.07, 0.02, 0.03, 0.4, 0.12, 0.13, 0.23],
  [0.40, 0.12, 0.27, 0.03, 0.1, 0.02, 0.1]
]
```

Fig. 8. Possibility Matrix.

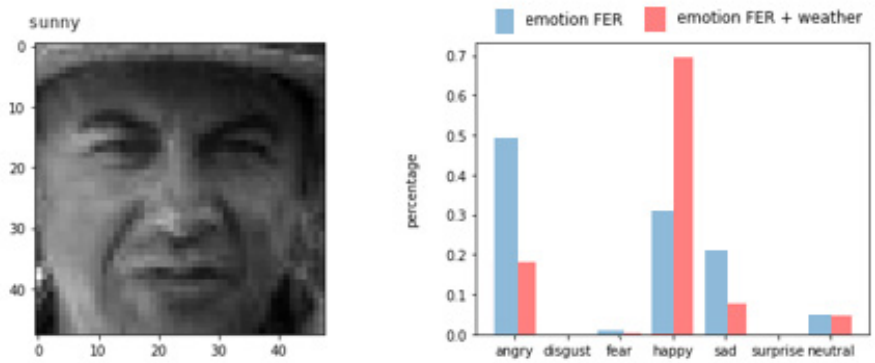

Fig. 9. Confirmation of the emotion of happiness based on recognized snowy weather.

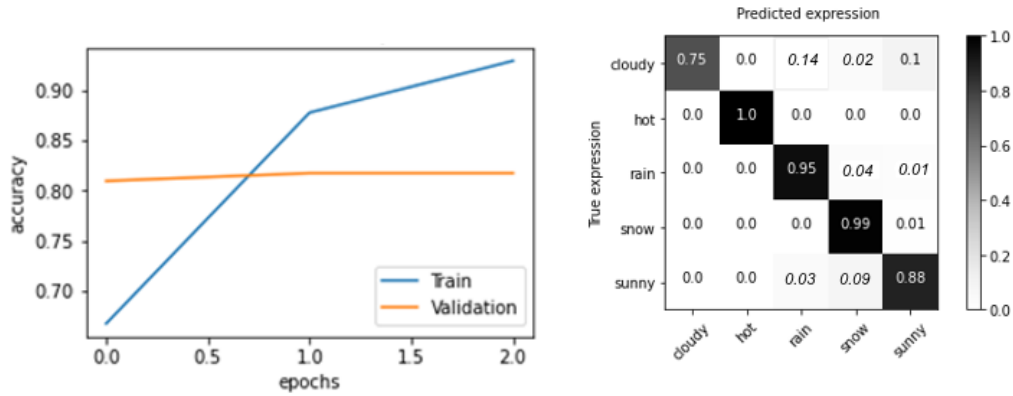

Fig. 10. Model accuracy and Confusion Matrix.

Fig 11. Our Private Dataset with facial, body and weather images.

RESULTS AND DISCUSSION

Testing of multimodal system

The multimodal system tested with images from our private dataset created in 2020 and reported in [8]. In the beginning, it was oriented to facial emotion recognition, and then in 2021 it was extended with images for body emotion recognition. Now it is additionally extended with more than 250 images with mentioned 5 weather conditions Fig. 11. Only the images with hot weather were taken from Internet. All other images were selected from private photo albums.

Following are some test results of our multimodal system:

On following pictures are given the base photo (Fig. 12) and the input images extracted from it for the three DNN models, as well the results produced by the models.

As can be seen on Fig. 13 and Fig. 14 there is a match between FER result Happy and BER results Happiness,

Fig. 12. Base photo.

so the dominant emotion Happy is confirmed by the decision module and it is used by Possibility Matrix together with sunny weather, recognized by the Weather model to form final result Happy emotion.

In Fig. 15 the use of a weather model based on the

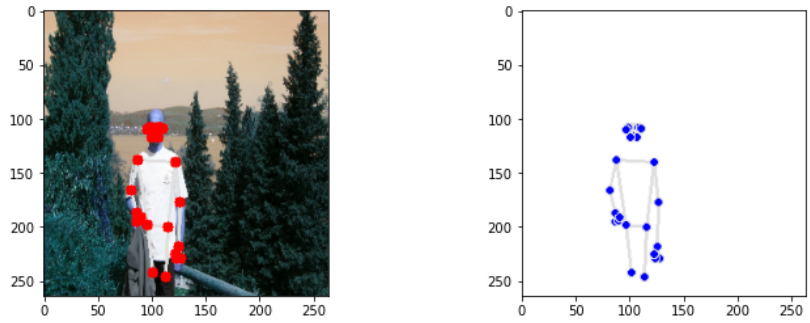

Fig. 13. Input image for BER model and BER emotion predictions (affection, confidence, esteem, excitement, happiness,

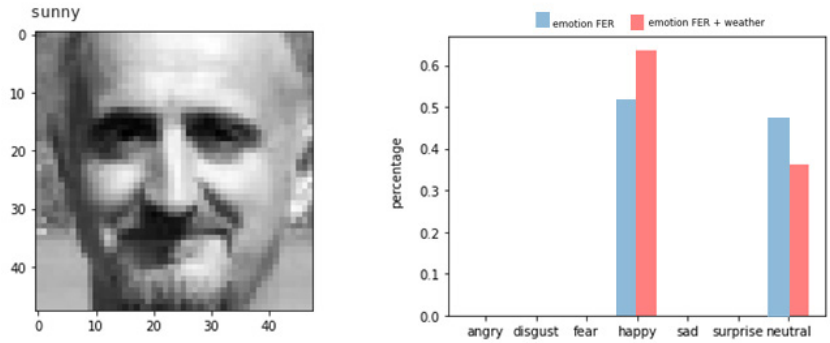

Fig. 14. FER model input image and FER and Weather models results.

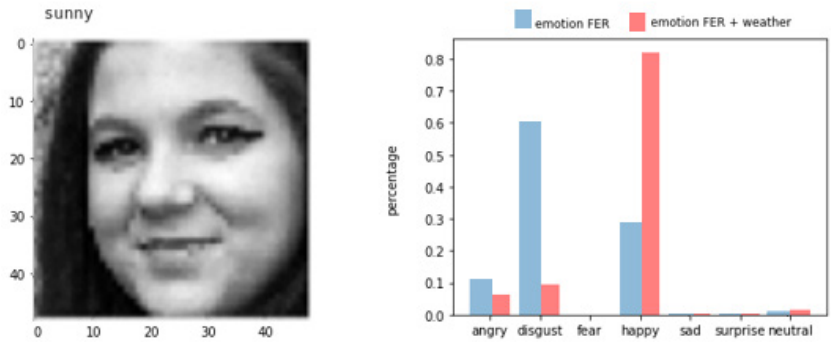

Fig. 15. Happiness facial emotion recognition based on recognized sunny weather.

probability matrix leads to the suppression of the disgust emotion and increases the role of the happy emotion to 80 %. So as a result, the emotion happiness is dominant.

In Fig. 16, unlike the previous figure, the snow weather information confirms the result of the FER and BER models, so the result is the emotion angry.

In the next Fig. 17, the use of the weather model confirms the last fear emotion recognized by the FER and BER models. The possibility matrix suppresses neutral emotions and increases the emotions of anger and fear.

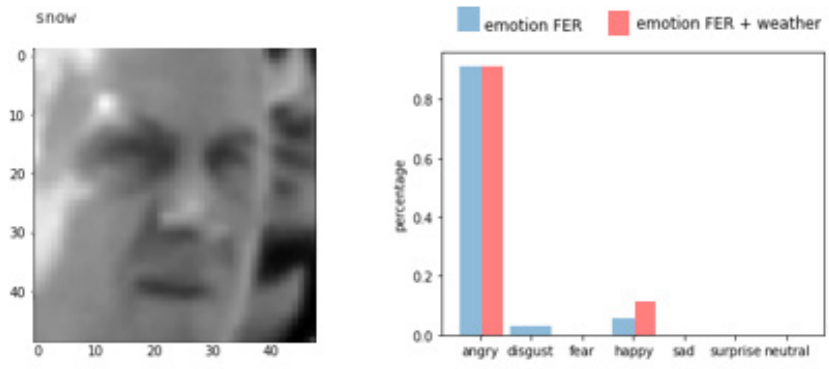

Fig. 16. Recognition of facial emotion angry, based on snowy weather.

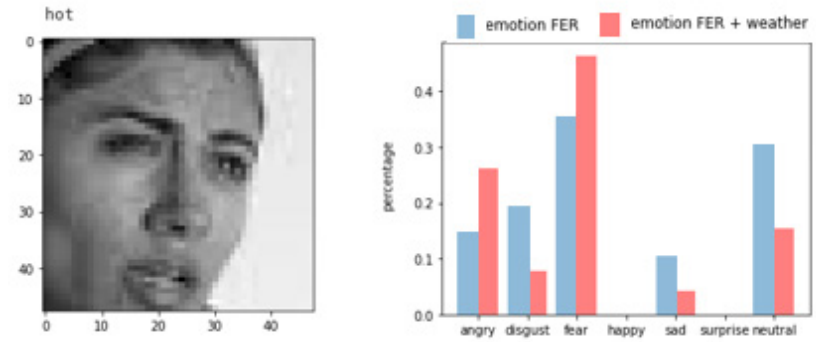

Fig. 17. Recognizing the emotion of fear based on hot weather.

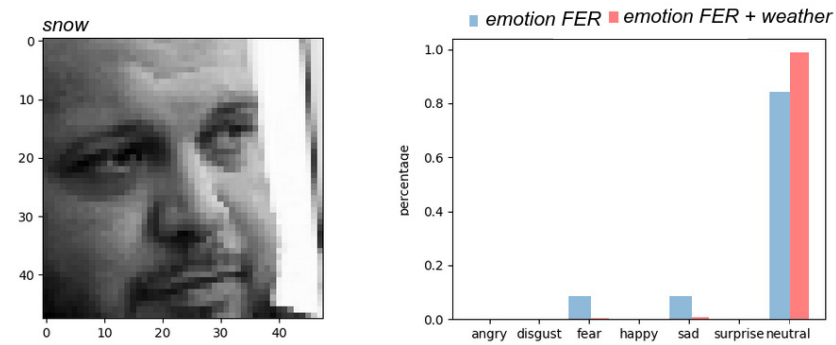

Fig. 18. Recognition of a Neutral emotion based on snowy weather.

Therefore, the result of the multimodal system is an increase of up to 45 % of the fear emotion.

In next case given in Fig. 18 the third model accounting for snow weather confirms and increases to 85 % the emotion recognized by the FER and BER models. And the result is a neutral emotion.

As can be seen in Fig. 19, even in rainy weather there are identical results (happy emotion) which are recognized by all DNN models and then the final emotion is happiness.

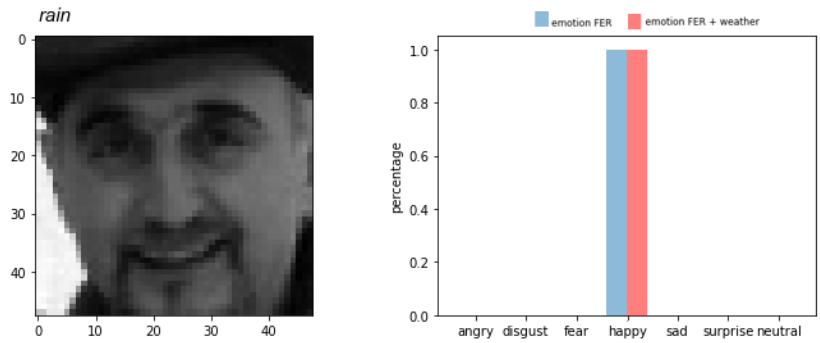

Fig. 19. Confirmation of happy emotion from all models in rainy weather.

CONCLUSIONS

The developed multimodal facial emotion recognition system based on FER, BER and weather neural networks increases the overall recognition of facial emotions from 69.85 % to more than 80 - 83 %. The weather model increases the accuracy especially in cases where the FER and BER models give fluctuating results, i.e. cannot unambiguously determine the facial emotion. Classified by ResNet, weather conditions are applied to human emotions using the empirically defined coefficients of the Possibility Matrix. As an advantage of the multimodal system, it can be stated that these coefficients can be changed depending on the type of event being held, for example, a sports competition, conference, wedding, celebration, vacation, etc., which can further increase the accuracy of the combined model.

A private weather set of more than 250 images belonging to the mentioned five classes is applied to verify the system. Our observation, using this dataset as well as images from the WEAPD dataset, is that the ResNet model misclassifies images from the sunny subset and recognizes them as belonging to the snowy subset. This can be explained by the fact that many white shiny parts or regions of the solar dataset images are recognized as snow.

In our future work, we intend to improve the decision module at the end of FER and BER models by using optimal solutions and Case-Based Reasoning methodology finding similarity between the seven discrete FER emotions and 26 VAD continues emotions [18].

Facial emotions recognized by the multimodal system can be used for various purposes. For example, to detect an early stage of some diseases or recognize the attitude of students to some events happening outdoors, as well as the influence of weather on people's emotional states [19].

REFERENCES

1. P. Ekman, W. Friesen, Facial action coding system: a technique for the measurement of facial movement, Palo Alto, Calif: Consulting Psychologists Press, 1978.
2. J. Russell, A Circumplex Model of Affect. Journal of Personality and Social Psychology, 1980.
3. J. Russell, Core affect and the psychological construction of emotion, Psychological Review, 110, 1, 2003, 145-172.
4. P. Pihkala, Toward a Taxonomy of Climate, Emotions, Front. Clim., Sec. Climate Risk Management, Volume 3, 2021, 1-22.
5. Scarantino, Handbook of Emotions, York, NY, Guilford Press, 2016.
6. J. Mizgajski, M. Morzy, Affective recommender systems in online news industry: how e motions influence reading choices, User Modeling and User-Adapted Interaction, 29, 2019, 345-379.
7. D. Keltner, J. Tracy, D. Sauter, D. Cordaro, G. McNeil, Handbook of Emotions, New York, NY, Guilford Press, 2016.
8. A. Atanassov, D. Pilev, Pre-trained Deep Learning Models for Facial Emotions Recognition, International Conference Automatics and Informatics ICAI2020, Varna, Bulgaria, 2020, 1-6.
9. A. Atanassov, D. Pilev, F. Tomova, V. D. Kuzmanova, Hybrid System for Emotion Recognition Based on Facial Expressions and Body Gesture Recognition,

International Conference Automatics and Informatics ICAI 201, 2021, 135-140.
10. R. Kosti, J. Alvarez, A. Recasens, A. Lapedriza, Context Based Emotion Recognition using EMOTIC Dataset, IEEE Transactions on Pattern Analysis And Machine Intelligence, arXiv:2003.13401v1 [cs.CV], 2020, 1-12.
11. J. Xia, D. Xuan, L. Tan, L. Xing, ResNet15: Weather Recognition on Traffic Road with Deep Convolutional Neural Network Advances in Meteorology, 2020, 2020, Article ID 6972826.
12. Q. Al-Haija, M. Smad, S. Zein-Sabatto, Multi-Class Weather Classification Using ResNet-18 CNN for Autonomous IoT and CPS Applications, International Conference on Computational Science and Computational Intelligence (CSCI), 2020, 1586-1591.
13. https://github.com/haixiaxiao/A-database-WEAPD, Available 10.04.2024.
14. J. Deng, W. Dong, R. Socher, L. Kai-Li, L. Fei-Fei, ImageNet: A large-scale hierarchical image database, 2009 IEEE Conference on Computer Vision and Pattern Recognition, 2009, 248-255.
15. G. Ajayi, Multi-class Weather Dataset for Image Classification", Mendeley Data, V1, 2018. doi: 10.17632/4drtyfjtfy.1
16. Q. Al-Haija, M. Gharaibeh, A. Odeh, Detection in Adverse Weather Conditions for Autonomous Vehicles via Deep Learning, AI, 3, 2022, 303-317.
17. H. Xiao, F. Zhang, Z. Shen, K. Wu, J. Zhang, Classification of weather phenomenon from images by using deep convolutional neural network, Earth and Space Science, 8, 2021, 1-9.
18. A. Avdzhieva, G. Nikolov, Asymptotically optimal definite quadrature formulae of 4th order, Journal of Computational and Applied Mathematics, 311, 2017, 565-582.
19. A. Ruseva, D. Tochev, Z. Boneva,Y. Assyov, L. Georgieva, D. Nikolovska, Marital status and education as risk factors for colorectal cancer, Trakia Journal of Sciences, 3, 2019, 224-228.