Facial Expression Recognition using Facial Mask with EMG Sensors

Ivana Kiprijanovska[1,*], Borjan Sazdov[1,2], Martin Majstoroski[2], Simon Stankoski[1], Martin Gjoreski[1,3], Charles Nduka[1], Hristijan Gjoreski[1,2]

1 Emteq Ltd., Brighton BN1 9SB, UK.

2 Faculty of Electrical Engineering and Information Technologies, Ss. Cyril and Methodius University in Skopje, North Macedonia.

3 Faculty of Informatics, Università della Svizzera Italiana, 6900 Lugano, Switzerland.

* Correspondence: ivana.kiprijanovska@emteqlabs.com

ABSTRACT

In this study, we examine the relationship between surface electromyography (sEMG) and facial expressions using a novel Virtual Reality multi-sensor mask insert – emteqPRO^{tm}, equipped with seven sEMG sensors. We designed a dataset collection scenario to analyze the effects of expression intensity, expression duration, and head movements. Using data from 30 participants, we developed a machine learning pipeline that included preprocessing of the sensor data, de-noising, filtering, segmentation, feature engineering, and training a classification model. The experimental results indicate that the mask is suitable for recognizing five posed facial expressions (smile, frown, eyebrows raise, squeezed eyes, and neutral expression). The best-performing model achieved an F1-Macro score of 0.86. Head movement decreased the results to an F1-macro score of 0.82. The facial expressions that activate the same muscles were the most challenging to differentiate. We also present results on the influence of different scaling and oversampling techniques. Finally, expression duration, intensity, and head movements influence the performance of the models for expression recognition and should be considered in the development of recognition algorithms.

KEYWORDS

Facial expressions, Surface EMG, Wearable sensors, Machine learning, emteqPRO

1 Introduction

The book “The Expression of the Emotions in Man and Animals” by Charles Darwin reports on the first studies on human emotions [1] arguing that the emotions are a universal language. These findings were later supported by Ekman’s groundbreaking work on emotions and their relation to facial expressions [2]. The human face is considered as one of the primary affect expression mediators, and as such, it has been explored as the primary marker of human affect. Generally, facial expressions result from the contraction of a set of facial muscles, from which affective states can be inferred [3]. Besides their relation to the affective states, facial expressions account for a large proportion of nonverbal communication [4].

Affective states can lead to different physiological and behavioral responses [5]. Measuring these responses is a key factor in understanding human behavior [6] and how these behaviors affect one’s mental health. Mental health monitoring is a growing scientific field striving to help people in need. An important goal of the field is to detect the first signs of mental health problems so that they can be identified and acted upon to reduce risks, build resilience and establish supportive environments [7].

Virtual reality (VR) has been a growing trend in the past decade. It enables the simulation of ecologically validated scenarios, which are ideal for studying behaviour in controllable conditions. Physiological measures captured in such conditions provide a deeper insight into how an individual responds to a given stimuli, making the VR tools suitable for diagnosis, intervention, and monitoring of mental health and wellbeing outcomes. Such solutions for improved emotion tracking will positively impact the lives of over one hundred million people in the EU alone who experience mental health problems.

Automatic facial expression recognition has been an active scientific subject since the early 1990s [8]. Recent studies have considered EMG sensing for facial expressions and emotion recognition, and classification methods have seen significant improvements in recent years. Mithbavkar et al. [9] focused on the recognition of emotions through facial expressions using data collected in a musical environment. They trained several neural networks to classify four emotions: joy, anger, sadness, and pleasure, and achieved the highest accuracy of 99.1% using a Nonlinear autoregressive exogenous network (NARX). A comparison between an EMG-based facial expression detection model and an image processing model was made by Kulke et al. [10]. Affectiva iMotions software was compared with EMG measurements of the zygomaticus major and corrugator supercilii muscles in identifying happy, angry, and neutral faces. They concluded that the outputs from both systems were highly correlated, showing that EMG-based model can identify facial expressions produce results comparable to an image processing-based model. Chen et al. [11] intended to recognize facial emotions from sEMG data in a human–computer interaction scenario. They used a specially designed headband to record sEMG signals from the frontalis and corrugator supercilii muscles of six participants who were instructed to pose the facial expressions of anger, fear, sadness, surprise, and disgust. They achieved 95% accuracy using an Elman neural network (ENN).

Our study aims to explore the usage of a novel VR facial mask equipped with seven surface electromyography (sEMG) sensors to monitor facial muscle activity and classify five different facial

expressions. Our approach is based on signal-processing and machine learning (ML) techniques to detect smiles, frowns, eyebrows raise, squeezed eyes, and neutral facial expressions with different intensities (high and low) and duration (short and long). We chose these five facial expressions because of their relation to specific affective states: smiles are related to positive affect and happiness; frowns are related to negative affect, depression, and anxiety; eyebrows raise is related to surprise, which can be positive and negative in terms of affective valence; and squeezed eyes is a facial expression generally related to negative affective states like fear and disgust.

2 Data

The experiment was done on 30 participants aged 16 - 23 (20.8 ± 1.4), eighteen males and twelve females. All participants were healthy and had no family history of facial neuromuscular and nervous disorders or heart problems. The data were recorded using the emteqPRO^{tm} mask [7][12]. It is a face-mounted mask that can be combined with a VR head-mounted display, or it can be used as an open-face mask. The EMG sensors in the mask are positioned to overlap the zygomaticus muscles (which spread from the cheekbones to the corners of the lips), the frontalis muscles (which cover parts of the forehead above the eyebrows), the orbicularis muscles (which are close to the outside of the eyes), and the corrugator (a small muscle between the eyebrows). The sensor mask mounted on a VR device and the sensor positions are depicted in Figure 1.

Figure 1: emteqPRO^{tm} multi-sensor face mask

The participants were asked to perform two tasks (Task A and Task B) that included five posed expressions: smile, frown, eyebrows raise, squeezed eyes, and neutral expression. Task A contains the five posed expressions with different durations (short and long) and intensities (low and high), with three repetitions of each expression. Task B includes the same posed expressions as those in Task A. The main difference was the inclusion of head movement in a specific direction (left, right, up, down) while doing the expressions. Also, as a difference from task A, the expressions in task B were only of high intensity and long duration. The data collection process was uninterrupted. The participants had a neutral expression on their faces between the posed expressions, making the neutral class the most common in the dataset, with 55.6%, while the rest of the classes comprised 11.1% each.

3 Methodology

3.1 Sensor data preprocessing and feature extraction

During the data collection procedure, the sEMG data were continuously recorded at a fixed rate of 1000 Hz. These data underwent a data preparation process, including data filtering, segmentation, and feature engineering. To increase the data quality, we performed signal de-noising and filtering. The EMG signals were initially filtered with a Hampel filter to remove sudden peaks in the signals that appear because of rapid movements. Additionally, to reduce the noise caused by electromagnetic interference, which has visible components at 50 Hz and its harmonics, we utilized a frequency-based filtering method based on spectrum interpolation [13]. A sliding window technique was utilized for the sensor data segmentation. The signals were segmented using a 0.5-second window and a 0.1-second stride. Eventually, we extracted 34 features per EMG channel, resulting in a total of 238 features. The features included various amplitude-based features (e.g., average amplitude change and mean absolute value), amplitude derivatives, auto-regressive coefficients, cepstral coefficients, frequency-based features (e.g., main frequency), and statistical features (e.g., statistical moments).

3.2 Modeling

Due to the class imbalance, we experimented with three different data resampling techniques to achieve a balanced class distribution. These were: (i) Random Undersampling – instances from the majority class are randomly chosen and removed from the training dataset; (ii) Synthetic Minority Oversampling Technique (SMOTE) – an oversampling technique that creates a synthetic example of the minority class based on the features of K-nearest neighbors; and (iii) One-Sided Selection Undersampling (OSS) – an undersampling technique that combines Tomek Links and the Condensed Nearest Neighbor (CNN) Rule to remove ambiguous points on the class boundary and to eliminate redundant examples from the majority class that are far from the decision boundary.
For feature scaling, we implemented standardization and normalization. Both techniques were done participant-wise, i.e., for each participant data separately.

The data resampling and feature scaling techniques were combined, and such processed data were used as input to several ML algorithms, including Decision Tree Classifier [14], Random Forest, and Extreme Gradient Boost (XGBoost) [15]. Eventually, the best performing classifiers were combined with a Hidden Markov Model (HMM) [16].

The dataset was divided into three disjoint subsets: a validation set (five randomly selected participants), a test set (five randomly selected participants), and a training set consisting of the remaining 20 participants' data. The validation set was used for tunning, and all the models were evaluated on the test set. As performance metrics, accuracy and F1-Macro scores were used.

4 Experimental results

A) Task A – Long and short-duration expressions with different intensity

The results obtained when data from Task A were used for training and testing are shown in Table 1. All the results presented in the table were achieved with the Random Forest algorithm, which proved to be the most effective one out of the ML algorithms used in our experiments (on the validation set). Depending on the scaling and the over/undersampling technique, the accuracy values range from 84.2% to 89.48%, while F1-Macro score values are between 0.75 and 0.86.

Table 1: Evaluation on the Task A dataset.

Approach	*Accuracy*	*F1-Macro*
Default	84.90%	0.77
Standardization	87.62%	0.82
Normalization	85.04%	0.78
Random Undersampling	84.20%	0.76
OSS Undersampling	85.50%	0.80
SMOTE Oversampling	86.34%	0.80
Standardization + Random Undersampling	87.69%	0.83
Standardization + SMOTE	88.59%	0.84
Standardization + SMOTE + HMM	**89.48%**	**0.86**

Regarding the feature scaling, both the standardization and the normalization improved the model's performance compared to the default (no scaling and no additional data sampling). The improvement was greater in the case where feature standardization was used. We believe there are two reasons for the improvement: (i) the scaling was performed for each participant's data separately, thus it acts as an unsupervised personalization technique reducing the inter-participants differences; (ii) besides scaling the feature ranges (e.g., in the range -3 to 3), the standardization also shifts the data distribution for each participant separately. Whereas the normalization only scales the feature ranges (e.g., 0 to 1). Thus, the additional distribution shift that the standardization causes for each feature may be why standardization is better than normalization.

Regarding the data subsampling technique, both OSS undersampling and SMOTE oversampling performed better than the random undersampling. The SMOTE oversampling technique was the best performing one, achieving an accuracy of 86.34% and an F1-score of 0.8. By combining feature scaling (standardization) and SMOTE, we achieved the highest results (an F1-score of 0.84). Finally, Hidden Markov Method was applied in combination with Standardization and SMOTE, achieving an F1-Macro score of 0.86 and an accuracy of 89.48%.

To further inspect the best-performing model, we present the confusion matrix in Figure 2. It indicates that the model struggles to correctly predict the smiling expressions. We speculate that this may be the case because smiling activates only the zygomatic face muscles. A high intense smile is easily distinguishable from neutral expressions as the zygomatic muscle activity is high. However, the low-intensity smile leads to low activation of the zygomatic muscles. A low-intensity smile resembles a neutral expression, and it doesn't affect the sEMG sensors enough to notice the difference between these two expressions.

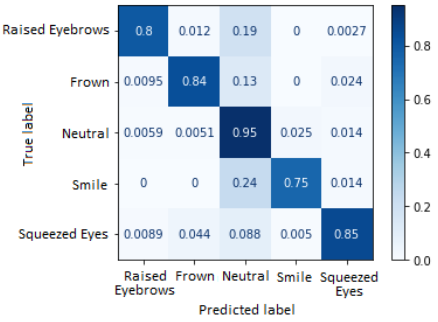

Figure 2: Confusion Matrix of the best performing model for Task A.

B) Task B – Long-duration expressions with high intensity and head movements

The results obtained when data from Task B were used for training and testing are shown in Table 2. All the results presented in the table were achieved with Extreme Gradient Boost (XGBoost) with gbtree, which proved to be the most effective one on the validation set.

Table 2: Evaluation on the Task B dataset.

Approach	*Accuracy*	*F1-Macro*
Default	86.24%	0.81
Standardization	86.46%	0.82
Normalization	82.65%	0.76
Random Undersampling	83.51%	0.77
OSS Undersampling	84.86%	0.79
SMOTE Oversampling	86.20%	0.81
Standardization + Random Undersampling	85.88%	**0.82**
Standardization + SMOTE	85.47%	0.81

The accuracy values range from 82.6% to 86.4%, and the F1-Macro scores are between 0.76 and 0.82. Feature scaling and resampling are not as critical preprocessing steps as in Task A. The method with the highest accuracy is the one where only participant-wise standardization was performed. It achieved 86.46% accuracy and an F1-Macro score of 0.82. However, the method that combines standardization and random undersampling has the highest F1-Macro score of 0.82 and 85.88% accuracy. Although this method's accuracy is lower, we consider it the best performing one since the F1-Macro score is more suitable for evaluations on an unbalanced dataset.

Figure 3 presents the confusion matrix for the best-performing model. We can see from the confusion matrix that the model can differentiate between all the classes with the neutral class, which was not the case in task A. This is because, in Task B, only high-intensity expressions with a long duration were examined. In this case, the muscles were highly activated, making the expressions more distinguishable from the neutral expression.

The problem with this model is that it struggles to detect frowns and eyes squeezed. Both expressions activate the same facial muscles: mainly the frontalis and corrugator muscles are activated, and this leads to wrongly predicting these expressions.

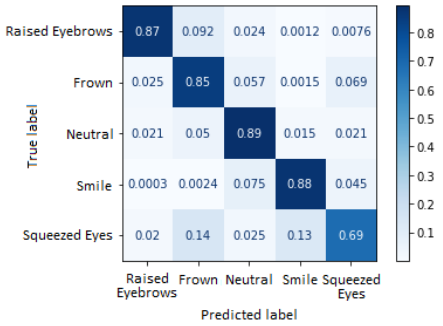

Figure 3: Confusion matrix of the best-performing model for Task B.

Overall, it seems that the head movements had a minor influence, i.e., the best- performing model achieved an F1-score of 0.82, which is on par with the best-performing model from Task A (Table 1), where the method without HMM achieved and F1-score of 0.84. We excluded the HMM-based method from this analysis and the following one because it adds a layer of complexity to the training process.

C) Task A and Task B combined – Long and short-duration expressions with head movements

Table 3 shows the performance of the methods for six train-test combinations: (i) training on Task A data and testing on Task A data; (ii) training on Task B data and testing on Task A data; (iii) training on both Task A and Task B data and testing on Task A data; (iv) training on Task B data and testing on Task B data; (v) training on Task A data and testing on Task B data; (vi) training on both Task A and Task B data and testing on Task B data. With this, we want to examine whether mixing no-movement and movement data for training and testing will substantially influence the method's performance. For evaluating methods' performance on Task A and Task B data, we used the best-performing methods from Table 1 and Table 2, respectively.

Table 3: Accuracy and F1-Macro by combining Task A and Task B

Approach	*Accuracy*	*F1-Macro*
Trained on Task A, tested on Task A	88.59%	0.84
Trained on Task B, tested on Task A	83.86%	0.78
Trained on AB, tested od Task A	85.48%	0.78
Trained on Task B, tested on Task B	85.88%	0.82
Trained on Task A, tested on Task B	83.61%	0.78
Trained on AB, tested on Task B	86.00%	0.80

From Table 3, we can see that for Task A, the best results are achieved when only no-movement data is used in the training set (i.e., Task A is used), and the inclusion of movement data (Task B data) reduces the method's accuracy by 3 percentage points and the F1-score drops by 0.06. On the other hand, when Task B data are used for testing, the inclusion of no-movement data (Task A data) in the training set has a lower influence on the results for Task B, as the F1-score drops by 0.02. By mixing the training sets, we did not observe any improvements in the results for both Task A and Task B data. These results indicate that scenario specificity is important for the model's accuracy, i.e., if we expect movement during the usage of the models, then it is better to include training data that involves movement.

5 Conclusion

This study examined the relationship between sEMG sensor data from facial muscles and posed facial expressions using the novel emteqPRO^{tm} VR multi-sensor facial mask. We analyzed sEMG data from 30 participants that performed five facial expressions while wearing the device. The data collection scenario was specifically designed to inspect several aspects of facial expression recognition - duration (short vs. long), intensity (low vs. high), and head movements. The collected data was then used to develop models that recognize smiles, frowns, eyebrows raise, squeezed eyes, and neutral facial expressions. We explicitly inspected the influence of normalization techniques and data oversampling and undersampling techniques. On the test data of five unseen participants, the best-performing model achieved an accuracy of 89.48% and an F1-Macro score of 0.86. The approach is based on Random Forest in combination with standardization and oversampling (SMOTE) steps, and the Hidden Markov Method as a prediction-smoothing technique [17]. The best-performing model evaluated on the data that includes head movement achieved an F1-Macro score of 0.82 (a decrease from 0.84). These results indicate that there is an influence of the head movement on the detection of facial expressions. The main weakness of the models was observed in distinguishing between frown and squeezed eyes. Both expressions activate forehead muscles closely placed to each other (corrugator and frontalis muscles). In the future, we plan to investigate feature selection, model personalization, and end-to-end deep learning to overcome this weakness.

It should also be noted that in some cases, the differences in the results may have been due to the different random steps in the processing steps and in the learning ensembles that have built-in random steps. Nonetheless, this does not diminish the main findings of the study:(i) The novel VR-mask equipped with sEMG sensors in combination with ML is suitable for recognizing facial expressions (smile, frown, eyebrows raise, squeezed eyes, and neutral); (ii) facial expressions that activate the same muscles are the most challenging to differentiate; (iii) feature scaling is an important step which enables for minimizing inter-participant feature differences; (iv) the results regarding data oversampling or undersampling were inconclusive, as it improved the results in some cases (Table 1), but not in other cases (and Table 2); Finally, (v) expression duration, intensity, and head movements influence the performance of the models for expression recognition and should be taken into account in the development of facial expression recognition algorithms. These conclusions contribute to affect sensing in VR, which has potential in symptom monitoring during VR-delivered therapy for mental health disorders.

ACKNOWLEDGMENTS

This study was partially supported by the WideHealth project (European Horizon 2020) under grant agreement No. 952279.

REFERENCES

[1] Darwin, C. The Expression of the Emotions in Man and Animals. London. In The expression of the emotions in man and animals. University of Chicago press. (2015).

[2] Ekman, P., Friesen, W.V., & Ellsworth, P. Emotion in the Human Face in Studies in Emotion and Social Interaction, (1972).

[3] Van Boxtel, A. Facial EMG as a tool for inferring affective states. Proc. Meas. Behav. 7, 104–108, (2010).

[4] Oh Kruzic, C., Kruzic, D., Herrera, F., & Bailenson, J. (2020). Facial expressions contribute more than body movements to conversational outcomes in avatar-mediated virtual environments. Scientific reports, 10(1), 1-23.

[5] Myers, D. G. Theories of Emotion in Psychology: Seventh Edition, (2004).

[6] World Health Organization. 2022. Mental health: strengthening our response (June 2022). Retrieved July 30, 2022, from https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response.

[7] Gnacek, Michal & Broulidakis, John & Mavridou, Ifigeneia & Fatoorechi, Mohsen & Seiss, Ellen & Kostoulas, Theodoros & Balaguer-Ballester, Emili & Kiprijanovska, Ivana & Rosten, Claire & Nduka, Charles. 2022. emteqPro-Fully Integrated Biometric Sensing Array for Non-Invasive Biomedical Research in Virtual Reality. Frontiers in Virtual Reality. 3. (Mar. 2022) DOI: https://doi.org/10.3389/frvir.2022.781218.

[8] Vinay Bettadapura. 2012. Face Expression Recognition and Analysis: The State of the Art. DOI: https://doi.org/10.48550/arXiv.1203.6722.

[9] S. A. Mithbavkar and M. S. Shah. 2019. Recognition of Emotion Through Facial Expressions Using EMG Signal. 2019 International Conference on Nascent Technologies in Engineering (ICNTE), 2019, pp. 1-6, DOI: https://doi.org/10.1109/ICNTE44896.2019.8945843.

[10] Kulke, Louisa & Feyerabend, Dennis & Schacht, Annekathrin. (2018). Comparing the Affectiva iMotions Facial Expression Analysis Software with EMG. DOI: https://doi.org/10.31234/osf.io/6c58y.

[11] Chen, Yumiao, Yang, Zhongliang & Wang, Jiangping. 2015. Eyebrow Emotional Expression Recognition Using Surface EMG Signals. Neurocomputing. 168. (May. 2015) DOI: https://doi.org/10.1016/j.neucom.2015.05.037. (2015).

[12] Hristijan Gjoreski, Ifigeneia I. Mavridou, Mohsen Fatoorechi, Ivana Kiprijanovska, Martin Gjoreski, Graeme Cox, and Charles Nduka. 2021. emteqPro: Face-mounted Mask for Emotion Recognition and Affective Computing. 2021. In Adjunct Proceedings of the 2021 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2021 ACM International Symposium on Wearable Computers (UbiComp '21). Association for Computing Machinery, New York, NY, USA, 23–25. https://doi.org/10.1145/3460418.3479276.

[13] Mewett, D. T., Reynolds, K. J., & Nazeran, H. Reducing power line interference in digitised electromyogram recordings by spectrum interpolation. Medical and Biological Engineering and Computing, 42(4), 524-531, (2004).

[14] Ali, Jehad & Khan, Rehanullah & Ahmad, Nasir & Maqsood, Imran. 2012. Random Forests and Decision Trees. International Journal of Computer Science Issues (IJCSI). 9. (Sept. 2012).

[15] Bentéjac, Candice & Csörgő, Anna & Martínez-Muñoz, Gonzalo. (2019). A Comparative Analysis of XGBoost.

[16] Khanna, Rahul & Awad, Mariette. 2015. Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers. DOI: https://doi.org/10.1007/978-1-4302-5990-9.

[17] Gjoreski, M., Janko, V., Slapničar, G., Mlakar, M., Reščič, N., Bizjak, J., Drobnič, V., Marinko, M., Mlakar, N., Luštrek, M. and Gams, M., 2020. Classical and deep learning methods for recognizing human activities and modes of transportation with smartphone sensors. Information Fusion, 62, pp.47-62.