

Video Assessment to Detect Amyotrophic Lateral Sclerosis

Guilherme Camargo Oliveira^{a,b} Quoc Cuong Ngo^b

Leandro Aparecido Passos^a Leonardo Silva Oliveira^a Stella Stylianou^c

João Paulo Papa^a Dinesh Kumar^b

^aSchool of Science, São Paulo State University, São Paulo, Brazil; ^bSchool of Engineering, Royal Melbourne Institute of Technology, Melbourne, VIC, Australia; ^cSchool of Science, Royal Melbourne Institute of Technology, Melbourne, VIC, Australia

Keywords

Amyotrophic lateral sclerosis · Facial expression · Facial action units · Machine learning · Logistic regression

Abstract

Introduction: Weakened facial movements are early-stage symptoms of amyotrophic lateral sclerosis (ALS). ALS is generally detected based on changes in facial expressions, but large differences between individuals can lead to subjectivity in the diagnosis. We have proposed a computerized analysis of facial expression videos to detect ALS. **Methods:** This study investigated the action units obtained from facial expression videos to differentiate between ALS patients and healthy individuals, identifying the specific action units and facial expressions that give the best results. We utilized the Toronto NeuroFace Dataset, which includes nine facial expression tasks for healthy individuals and ALS patients.

Results: The best classification accuracy was 0.91 obtained for the pretending to smile with tight lips expression.

Conclusion: This pilot study shows the potential of using computerized facial expression analysis based on action units to identify facial weakness symptoms in ALS.

© 2024 The Author(s).

Published by S. Karger AG, Basel

Introduction

Amyotrophic lateral sclerosis (ALS) is a highly dilapidating motor neuron disease. It is characterized by progressive muscle weakness and the inability to perform voluntary muscle contractions which impacts on many essential functions such as chewing, walking, talking, and facial expressions. Early symptoms of ALS include reduced facial expressiveness [1, 2]. However, with large differences in facial expressions between individuals, recognizing these changes, especially in the early stages, can be subjective with potential misdiagnosis [3–5]. Thus, there is an opportunity for computerized analysis of facial expressions to improve the reliability in the diagnosis of ALS.

Computerized identification of facial expressions has been proposed for a number of applications, such as biometrics, emotion recognition, and neurological identification [6–10]. While different approaches have been proposed, recognizing facial expressions using the Facial Action Coding System [11] has the advantage because it provides a physical description of the feature in terms of the facial expression. This is a systematic approach that analyzes **facial actions based on a set of 46 individual actions, namely, action units (AUs)**, each corresponding to specific facial muscle movement. The method has been

Table 1. Demographic and clinical information, including duration in months from ALS symptom onset

Group	Age, years	Duration, months	ALSFRS-R
HC	63.2±14.3	–	–
ALS	61.5±8.0	49.6±31.6	34.8±5.0

Table 2. Toronto NeuroFace Dataset ALS distribution

Task	ALS	HC
SPREAD	11	11
KISS	11	11
OPEN	11	11
PA	10	11
PATAKA	10	11
BBP	9	11
BLOW	6	7
BROW	4	4
BIGSMILE	4	3

widely used, improved, and adapted for number of applications [12]. Hamm et al. [13] estimated using the Facial Action Coding System from temporal AU profiles for emotion analysis. Lucey et al. [14] described a method to detect pain in video through facial AUs. Barrios et al. [15] developed an approach to detect muscle activations and intensities as users perform facial exercises in front of a mobile device camera using AU intensity. Oliveira et al. [16, 17] used a similar approach for hypomimia detection in Parkinson’s disease and for identifying signs of stroke.

This work proposes using AUs to detect the differences in the facial expressions of ALS patients and healthy people; AUs have the advantage because these describe the actions. The Toronto NeuroFace Dataset [18] was utilized, comprising nine video sequences that represent various facial expressions captured during routine orofacial examinations of both healthy controls (HCs) and patients with ALS. The AUs were computed for each of the video recordings and classified against the disease label. The primary contribution of this work is the demonstration of the use of AUs for differentiating between the facial expressions of ALS and healthy individuals. It has the potential to assist the neurologists to detect symptoms of ALS in their patients.

Methods

This section outlines the dataset utilized in the study, the methods applied for extracting the AUs, and the grid search space used for classification. It also details the metrics employed in assessing the proposed model.

Dataset

This study used the Toronto NeuroFace Dataset [18], which has facial expression videos of 46 participants: 11 healthy individuals (7 male, 4 female), 24 poststroke patients (10 male, 4 female), and 11 people (4 male, 7 female) with ALS. In this study, the 24 stroke patients were not considered, and videos of the other 22 participants were analyzed. These videos were captured as the participants engaged in displaying different speech- and nonspeech-related facial expressions. To the best of our knowledge, it is the only public dataset for studying facial expressions in patients with neurological disorders.

All study participants were cognitively unimpaired, each achieving a Montreal Cognitive Assessment score of at least 26 and passing a hearing screening. ALS patients were diagnosed based on the El Escorial Criteria from the World Federation of Neurology. The severity of their condition was quantified using the ALS Functional Rating Scale-Revised (ALSFRS-R), where scores averaged 34.8 ± 5.0 . This score range highlights slight and mild severity cases across ALS patients. Table 1 provides demographic and clinical details, including the number of months since ALS symptoms began.

The dataset comprises nine facial expression activities used to assess oro-motor capabilities. While “Buy Bobby a Puppy” (BBP) at a comfortable speaking rate and intensity had ten repetitions, the syllable /pa/ (PA) and word /pataka/ (PATAKA) were repeated as many times as possible on a single breath. There were five repetitions each of pretending to blow a candle (BLOW), pretending to kiss a baby (KISS), the maximum opening of the jaw (OPEN), pretending to smile with tight lips (SPREAD), a big smile (BIGSMILE), and raising the eyebrows (BROW). Participants were recommended to take breaks between activities to prevent fatigue. However, not everyone was able to finish all the exercises. The distribution of tasks and the number of ALS and HC participants who completed each task are detailed in Table 2. The dataset comprises 76 videos from the ALS group and 80 from the HC group.

Feature Extraction

The extraction of AUs was performed through the Python Facial Expression Analysis Toolbox (Py-Feat) [19]. Cheong et al. [19] developed an XGBoost [20] classifier model, which was trained using Histogram of Oriented Gradients derived from five separate datasets: BP4D [21], DISFA [22], CK+ [23], UNBC-McMaster shoulder pain [24], and AFF-Wild2 [25]. Figure 1 displays the facial landmarks and AUs of a facial image, as analyzed using Py-Feat. Figure 2 illustrates the extraction flow.

Four statistical measures of AUs were computed across the video frames: mean, maximum, minimum, and standard deviation. The extreme values of minimum and maximum are important for identifying the range of facial muscle movements, essential for detecting muscle weakness or hyperactivity due to ALS. Variance is for identifying irregularities in muscle control, and the mean value serves as a standard for contrasting typical and atypical muscle activity. The standard deviation is useful in differentiating normal variations in facial expressions, providing a different scale from variance. Although variance and standard deviation are related, their application as features in logistic regression models can lead to diverse outcomes. These statistical methods provide a comprehensive view of facial muscle behavior, facilitating the detection of facial weakness in ALS through minor changes in facial expressions.

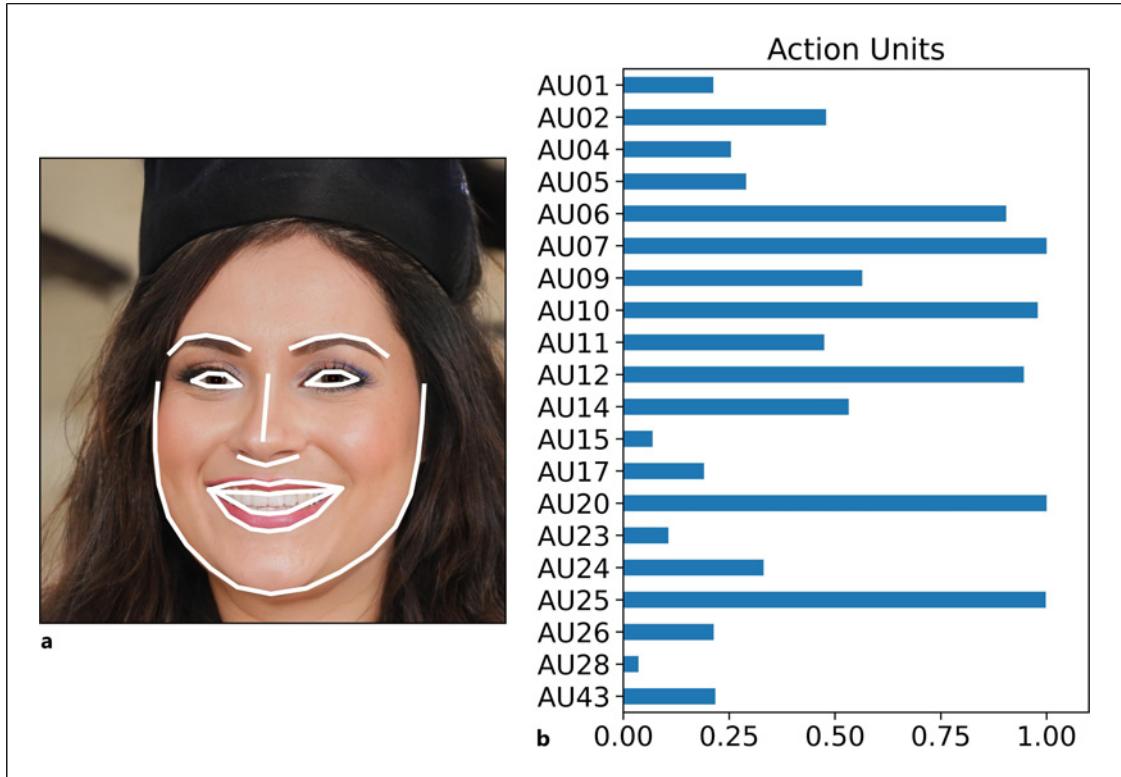


Fig. 1. Example of Py-Feat extraction: facial landmarks (a), AUs (b).

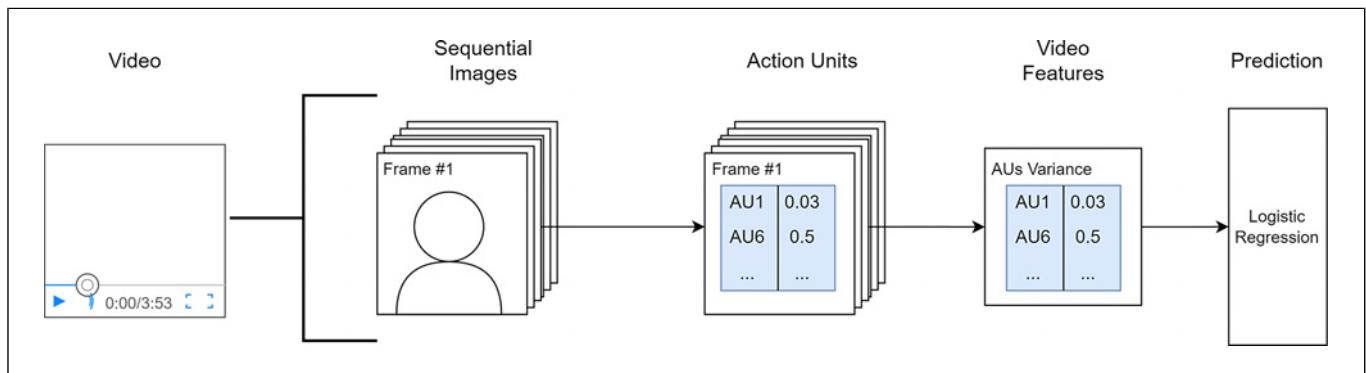


Fig. 2. Feature extraction flow: video frames are analyzed with Py-Feat to extract AUs. In this example, variance is calculated as the statistical measure across frames to represent the video feature, which is then input into a logistic regression for classification.

Classification

The sample size of BROW and BIG SMILE were very small and were not included in the study. Binary classification was performed to distinguish between HC individuals and ALS patients from each of the seven videos: KISS, OPEN, SPREAD, PA, PATAKA, BBP, and BLOW.

The classification pipeline used in this study consists of three steps: (i) standardization of features, where each feature is adjusted

to have a mean of zero and a standard deviation equal one; (ii) removal of correlated features with Pearson correlation coefficients above 0.7. In this step, the first identified feature was retained, and all subsequent correlated features were removed; and (iii) employing a machine learning approach for classification using logistic regression [26].

Logistic regression is a method used for clinical prediction modeling, as highlighted by Christodoulou et al. [27]. They noted

Table 3. Configuration of grid search space

Parameters	Values
solver	[<i>liblinear</i>]
penalty	[<i>l1</i> , <i>l2</i> , <i>elasticnet</i>]
C	[0.01, 0.1, 1, 10, 100]

that despite the increasing popularity of machine learning algorithms, such as artificial neural networks, support vector machines, and random forests, logistic regression remains highly effective, especially in the medical domain.

The classification pipeline and grid search were carried out using *Scikit-learn* [28]. However, *Scikit-learn* lacks comprehensive statistical analysis capabilities for logistic regression, leading to the use of the *Statsmodels* [29] library for assessing feature importance. The search space and parameters used in *Scikit-learn* for logistic regression are detailed in Table 3. In the analysis with *Statsmodels*, the *alpha* parameter was set to 1, employing the *l1* regularization method.

The assessment was carried out on a task-by-task basis, employing a “Leave-One-Out” k-fold cross-validation method, where k is equal to the total number of entries in the database. A distinct model was created for each instance in the training set and then evaluated. This method involves evaluating a model for every single instance, offering a more thorough and stronger evaluation of the model’s effectiveness.

The performance of the system was assessed based on accuracy, sensitivity, and specificity. Furthermore, the area under the curve was calculated as the area under the *sensitivity* versus (1-*specificity*) curve.

Results

This section presents the results obtained in the classification tasks. The experiments were conducted over the statistical information of 20 AUs computed for each task using the logistic regression classifier with hyper-parameters optimized through a grid search approach to find the best model, i.e., the one that maximizes the accuracy.

Table 4 shows the results for HC versus ALS for each of the seven facial expression tasks. The highest accuracy was 0.91 for the SPREAD task when the min of AUs was used. Additionally, this table shows significant variations in performance across various tasks and features.

Feature Importance

AUs’ descriptive power can be expressed by their respective coefficients, obtained through the logistic regression. Consider, for instance, the SPREAD task, in which the grid search procedure optimized the logistic regression by selecting the following parameters: $C = 1$, $penalty = l1$, $random\ state = 42$, $solver = liblinear$.

Under such configuration, Figure 3 displays the corresponding coefficients. Notice that AU09, AU17, and AU11 have large coefficients for HC, suggesting that the features have a strong predictive power for HC in the given model. These features are correlated with HC and are able to contribute significantly to the overall performance of the model in making accurate predictions. However, it is important to note that a high coefficient alone does not necessarily imply causality or that the feature is the most important in all contexts as the significance of a feature may vary depending on the specific problem and dataset.

Table 5 shows the result of logistic regression extracted from the *Statsmodels* library with a high Pseudo R-squared value of 0.8007, indicating good model fit. The model, based on 22 observations, is statistically significant with a low LLR *p* value of 0.0001804. Among the predictors, AU09 (Levator labii superioris alaqueae nasi) is notably significant with a coefficient of -2.0042 and a *p* value of 0.027, suggesting a negative relationship with the dependent variable. Other variables, including AU04 (Depressor Glabellae, Depressor Supercilli, Currugator), AU11 (Zygomatic Minor), AU14 (Buccinator), AU17 (Mentalis), and AU25 (Depressor Labii, Relaxation of Mentalis, Orbicularis Oris), are not statistically significant, indicating they have a lesser impact on the dependent variable using logistic regression.

Finally, Table 6 compares the accuracy values against the results reported in literature, i.e., reported by Bandini et al. [30] and Gomes et al. [10] for the HC versus ALS classification assignment considering distinct tasks, i.e., BBP, PATAKA, PA, KISS, OPEN, SPREAD, and BLOW. For each task, we calculated the mean accuracy which is shown in Figure 4.

Gomes et al. [10] employed Delaunay triangulation with graph neural networks [31]. Bandini et al. [30] and Gomes et al. [10] combined kinematic features with Logistic regression and Support Vector Machine (SVM) [32]. Our method used AUs as the feature with classification using logistic regression, likely for facial expression analysis through muscle movement classification. Each approach varies in accuracy, reflecting their effectiveness in specific classification or prediction tasks.

One advantage of our method is that it does not require manual video segmentation nor the need for normalization using REST subtask. While Gomes et al. [10] adopted an approach similar to Bandini et al. [30], their results were distinct. The difference in results could have been attributed to the unavailability of videos from the REST subtask, manually cropping the frames, and because they did not incorporate three-dimensional depth features, which adds to the computational and imaging complexity.

Table 4. Logistic regression – comparative performance in HC versus ALS

Task	Measure between frames	Acc	Sens	Spec	AUC
BBP	var	0.45	1.00	0.00	0.50
	mean	0.45	1.00	0.00	0.50
	min	0.45	1.00	0.00	0.50
	max	0.65	0.78	0.55	0.67
	std	0.45	1.00	0.00	0.50
PATAKA	var	0.48	1.00	0.00	0.50
	mean	0.57	0.60	0.55	0.47
	min	0.57	0.40	0.73	0.56
	max	0.62	0.50	0.73	0.57
	std	0.62	0.80	0.45	0.59
PA	var	0.48	1.00	0.00	0.50
	mean	0.48	1.00	0.00	0.50
	min	0.71	0.80	0.64	0.71
	max	0.48	1.00	0.00	0.50
	std	0.62	0.60	0.64	0.42
KISS	var	0.82	0.73	0.91	0.84
	mean	0.73	0.73	0.73	0.71
	min	0.73	0.73	0.73	0.79
	max	0.82	0.73	0.91	0.75
	std	0.77	0.64	0.91	0.82
OPEN	var	0.59	0.45	0.73	0.50
	mean	0.77	0.82	0.73	0.77
	min	0.82	0.91	0.73	0.86
	max	0.73	0.64	0.82	0.74
	std	0.59	0.45	0.73	0.45
SPREAD	var	0.82	0.73	0.91	0.69
	mean	0.77	0.73	0.82	0.77
	min	0.91	1.00	0.82	0.97
	max	0.68	0.73	0.64	0.60
	std	0.82	0.73	0.91	0.74
BLOW	var	0.62	0.50	0.71	0.50
	mean	0.46	1.00	0.00	0.50
	min	0.69	0.67	0.71	0.64
	max	0.62	0.67	0.57	0.62
	std	0.46	1.00	0.00	0.50

AUC, area under the curve.

Discussion

Previous research [16] has shown that it is possible to distinguish between healthy individuals and PD patients using the variance of AUs. On the other hand, the present study extends the earlier work and shows the effectiveness of using AUs from facial videos to identify ALS individuals. Unlike the earlier studies, it has investigated different facial expression tasks and statistical measures to represent the facial expressions.

The results have important implications, particularly for individuals with neurological diseases who commonly suffer from reduced facial expressions due to stiffness in facial muscles. The data indicate that AUs are capable of differentiating between healthy persons and those affected by ALS. This underscores the value of studying facial expressions and their associated symptoms in people with neurological conditions. It also highlights the facial expressions that are more affected and the AUs that detect these the best.

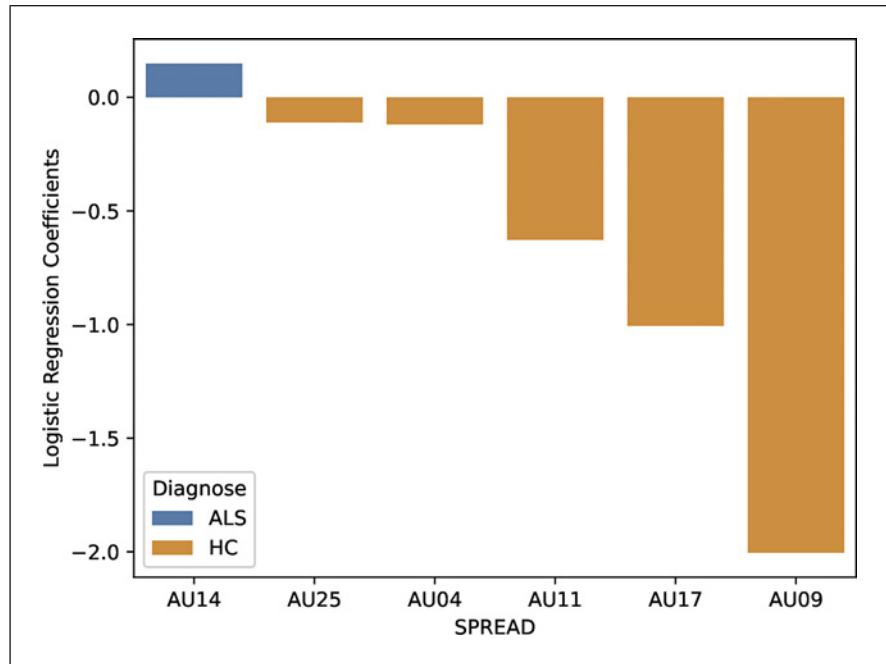


Fig. 3. Logistic regression coefficients.

Table 5. Logistic regression analysis

Dep. Variable	Y	No. observations	22
Model	Logit	df Residuals	16
Method	MLE	df Model	5
Date	Tue, Jun 13, 2023	Pseudo R-squ	0.8007
Time	17:09:49	Log-Likelihood	-3.0399
Converged	True	LL-Null	-15.249
Covariance type	Nonrobust	LLR p value	0.0001804
coef	std err	z	p> z
AU04 -0.1191	0.886	-0.135	0.893
AU09 -2.0042	0.903	-2.219	0.027
AU11 -0.6289	1.231	-0.511	0.609
AU14 0.1488	0.808	0.184	0.854
AU17 -1.0058	0.999	-1.007	0.314
AU25 -0.1091	1.078	-0.101	0.919

The model considered for AU extraction, i.e., Py-Feat, was not trained or fine-tuned on a dataset that contains faces of people with neurological diseases. As a result, the prediction of the AUs is only estimated.

A significant constraint of this research is the limited number of participants and their exclusive association with one hospital, potentially resulting in biases and limiting the model's broad applicability. This is a pilot study with a small sample size. The small sample size can lead to higher variability and reduced confidence in the stability of the reported metrics. Additionally, because of the absence of longitudinal data, this cannot be evaluated

for the long-term consistency and repeatability of the results.

The tasks OPEN and SPREAD show high mean accuracy in Figure 4. This suggests that these tasks might be better at differentiating between ALS and HC. The variation in performance across tasks indicates that certain facial movements might be more indicative of ALS-related changes, and the classifier's performance is task-dependent. This warrants further investigation with larger datasets to validate these findings and explore the underlying reasons for these differences.

Table 6. Comparison with prior works

Task	Study	Approach	Accuracy
BBP	Gomes et al. [10]	Delaunay triangulation + graph neural networks	50.0%
	Gomes et al. [10]	Kinematics + SVM-lin	45.0%
	Bandini et al. [30]	Kinematics + logistic regression	89.0%
	Our method	AUs + logistic regression	65.0%
PATAKA	Gomes et al. [10]	Delaunay triangulation + graph neural networks	66.6%
	Gomes et al. [10]	Kinematics + logistic regression	42.0%
	Bandini et al. [30]	Kinematics + SVM-lin	82.0%
	Our method	AUs + logistic regression	62.0%
PA	Gomes et al. [10]	Delaunay triangulation + graph neural networks	57.1%
	Gomes et al. [10]	Kinematics + SVM-lin	33.0%
	Bandini et al. [30]	Kinematics + logistic regression	77.0%
	Our method	AUs + logistic regression	71.0%
KISS	Gomes et al. [10]	Delaunay triangulation + graph neural networks	68.1%
	Gomes et al. [10]	Kinematics + logistic regression	50.0%
	Bandini et al. [30]	Kinematics + SVM-lin	55.0%
	Our method	AUs + logistic regression	82.0%
OPEN	Gomes et al. [10]	Delaunay triangulation + graph neural networks	81.8%
	Gomes et al. [10]	Kinematics + SVM-lin	77.0%
	Bandini et al. [30]	Kinematics + SVM-RBF	72.0%
	Our method	AUs + logistic regression	82.0%
SPREAD	Gomes et al. [10]	Delaunay triangulation + graph neural networks	81.8%
	Gomes et al. [10]	Kinematics + logistic regression	68.0%
	Bandini et al. [30]	Kinematics + SVM-lin	82.0%
	Our method	AUs + logistic regression	91.0%
BLOW	Gomes et al. [10]	Delaunay triangulation + graph neural networks	38.4%
	Gomes et al. [10]	Kinematics + SVM-RBF	45.0%
	Bandini et al. [30]	Kinematics + SVM-lin	65.0%
	Our method	AUs + logistic regression	69.0%

It is important to note that the study has only considered the loss of facial expressions, which is only one of the symptoms of these complex, multi-symptom disease conditions. This should be considered to be the basis to assist the clinician and improved by adding other clinical observations, signs, and symptoms.

Using facial videos in healthcare raises privacy issues and requires careful data management. This limits the scalability of such a study, and factors such as diversity in the datasets become difficult to address. We realize that addressing these is crucial for the success of AI-driven diagnostics to ensure they are unbiased and accurate, and we are considering these factors for future studies.

Future research can strengthen the current findings by conducting a comparative study of different disease stages versus HCs. Such an approach would involve examining both facial expressions and other motor functions to better understand disease progression in

comparison to a normal baseline. Utilizing advanced imaging techniques or motion capture may also help quantify differences in muscle activity and offer deeper insights into the onset and progression of disease-specific motor impairments.

It would be advantageous to conduct longitudinal studies, where the same person can be monitored repeatedly over the progression of the disease. This would track the evolution of symptoms over time and validate the accuracy of predictive models, particularly in diseases where symptoms evolve gradually.

Another extension of this work could be the inclusion of Patient-Reported Outcome (PRO) scores in future studies. Incorporating these scores, which reflect patients' perceptions of their health status and quality of life, alongside clinical assessments of facial muscle function, could provide a correlation between objective measures and subjective patient experiences. This approach would offer a more holistic view of the disease's impact on patients' daily

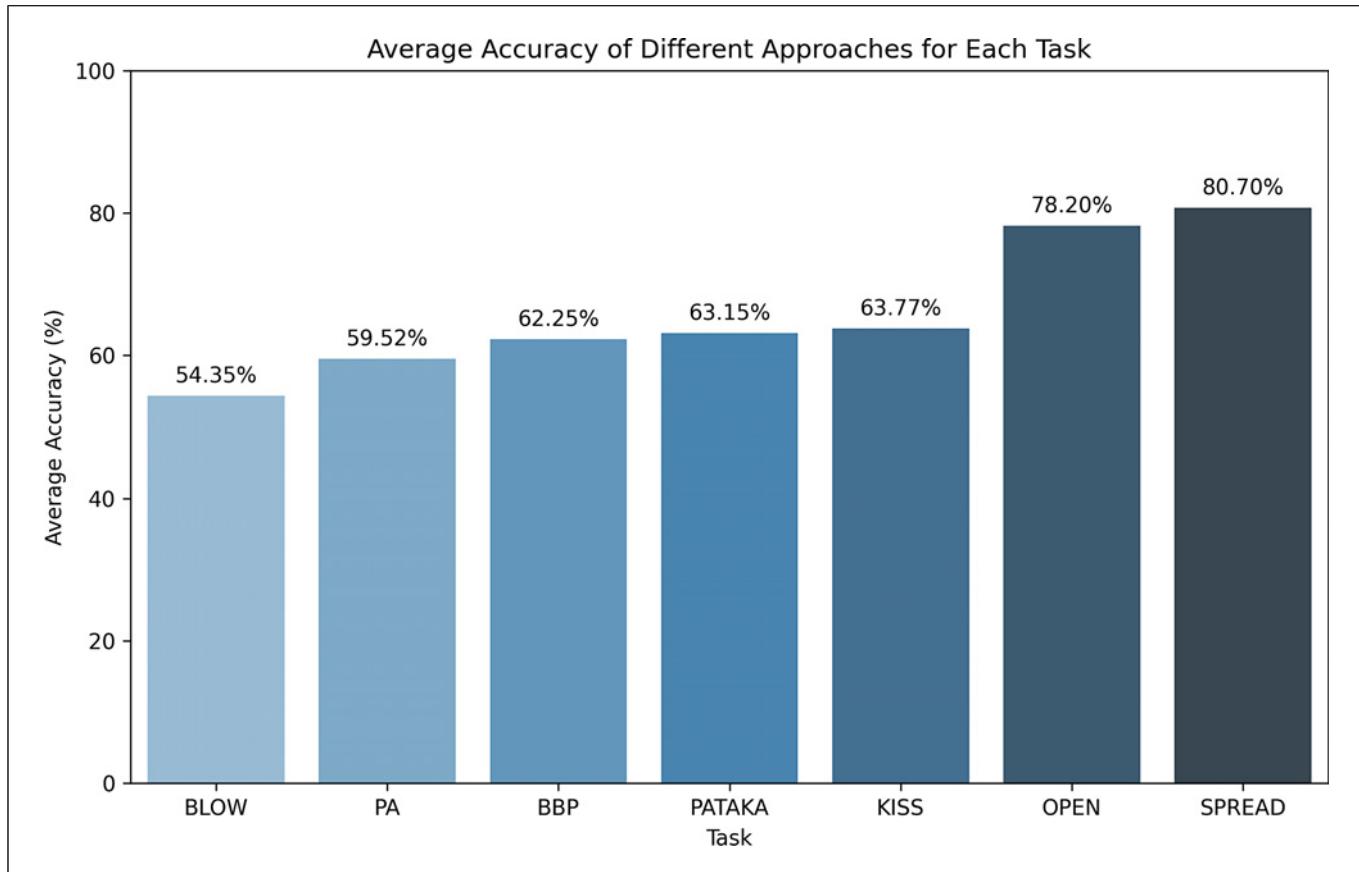


Fig. 4. Comparison of the mean accuracy of different approaches for each task.

lives and could significantly improve the effectiveness of clinical assessments and treatment strategies.

Finally, there is a need to develop remote smartphone-based video assessments for ALS. Abbas et al. [33] have shown that smartphone-based video assessments can provide an objective evaluation of motor abnormalities, and they have shown success in schizophrenia. The potential of using this technology for remote monitoring could make it suitable for patients living in remote or underserved areas where access to specialized ALS care is limited.

Conclusion

This manuscript presents a preliminary investigation into the potential of using AUs obtained from facial videos to identify facial weakness symptoms in ALS patients. The model was cross-validated and the most suitable facial expressions, the corresponding AUs, and the statistical measures were identified, and the highest accuracy was 0.91 when the SPREAD facial expression task was per-

formed. This shows that computerized analysis of videos of people performing facial expression tasks is a promising approach to assist clinicians in detecting ALS symptoms.

The experiments were conducted over the Toronto NeuroFace Dataset [18], which comprises facial videos of ALS and HC group participants while performing nine predefined facial expression tasks. This study is based on a small dataset; additional research is necessary before its findings can be applied widely in clinical settings. Further investigations should focus on examining the impact of factors like ethnicity and age and optimizing these methods for more diverse patient populations, employing larger datasets to fill these gaps. Additionally, distinct techniques can be explored for extracting and correlating clinically significant information from facial AUs and specific clinical symptoms.

Acknowledgment

We express our gratitude to Dr. Yana Yunusova for granting us permission to utilize the Toronto NeuroFace Dataset.

Statement of Ethics

As stated in Bandini et al. [18]: “The study was approved by the Research Ethics Boards at the Sunnybrook Research Institute and UHN: Toronto Rehabilitation Institute. All participants signed informed consent according to the requirements of the Declaration of Helsinki, allowing inclusion into a shareable database.” The face in Figure 1 is not real. It was synthetically generated by StyleGAN2 for illustration purposes. StyleGAN2 is a deep learning model capable of producing fake facial images.

Conflict of Interest Statement

The authors have no conflicts of interest to declare.

Funding Sources

We acknowledge the scholarship for G. Oliveira from RMIT University. We also acknowledge the financial support from Promobilia Foundation (Sweden). J. Papa is grateful to the São Paulo Research Foundation (FAPESP) grants 2013/07375-0, 2019/

07665-4, 2023/14427-8, and 2023/14197-2, as well as to the Brazilian National Council for Scientific and Technological Development grant 308529/2021-9.

L. Passos is also grateful to the FAPESP grant 2023/10823-6. This study was also financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brazil (CAPES) – Finance Code 001.

Author Contributions

G.O. and L.O. performed the analysis. J.P. and D.K. conceptualized and designed the project. S.S. evaluated statistically analyses. D.K., G.O., Q.N., L.P., and J.P. contributed to the manuscript.

Data Availability Statement

The Toronto NeuroFace Dataset can be accessed through an appropriate application procedure at <https://slp.utoronto.ca/faculty/yana-yunusova/speech-production-lab/datasets/>. Further inquiries can be directed to the corresponding author.

References

- 1 Wijesekera LC, Leigh PN. Amyotrophic lateral sclerosis. *Orphanet J Rare Dis.* 2009; 4:1–22.
- 2 Kiernan MC, Vucic S, Cheah BC, Turner MR, Eisen A, Hardiman O, et al. Amyotrophic lateral sclerosis. *The lancet.* 2011;377(9769):942–55. [https://doi.org/10.1016/S0140-6736\(10\)61156-7](https://doi.org/10.1016/S0140-6736(10)61156-7)
- 3 Oh S, Oh K-W, Kim H-J, Park J-S, Kim SH. Impaired perception of emotional expression in amyotrophic lateral sclerosis. *J Clin Neurol.* 2016;12(3):295–300. <https://doi.org/10.3988/jcn.2016.12.3.295>
- 4 Aho-Ozhan HE, Rgen Keller J, Heimrath J, Uttner I, Jan K, Birbaumer N, et al. Perception of emotional facial expressions in amyotrophic lateral sclerosis (ALS) at behavioural and brain metabolic level. *PLoS One.* 2016;e0164655.
- 5 Carelli L, Solca F, Tagini S, Torre S, Verde F, Ticozzi N, et al. Emotional processing and experience in amyotrophic lateral sclerosis: a systematic and critical review. *Brain Sci.* 2021;11(10):1356. <https://doi.org/10.3390/brainsci11101356>
- 6 Li SZ, Jain AK, Tian Y-L, Kanade T, Cohn JF. Facial expression analysis. *Handbook of face recognition;* 2005. p. 247–75.
- 7 Givens GH, Beveridge JR, Lui YM, Bolme DS, Draper BA, Phillips PJ. Biometric face recognition: from classical statistics to future challenges. *WIREs Comput Stats.* 2013;5(4): 288–308. <https://doi.org/10.1002/wics.1262>
- 8 Flynn M, Effraimidis D, Angelopoulou A, Kapetanios E, Williams D, Hemanth J, et al. Assessing the effectiveness of automated emotion recognition in adults and children for clinical investigation. *Front Hum Neurosci.* 2020;14:70. <https://doi.org/10.3389/fnhum.2020.00070>
- 9 Gomes NB, Yoshida A, Camargo de Oliveira G, Roder M, Papa JP. Facial point graphs for stroke identification. *Iberoamerican congress on pattern recognition.* Springer; 2023. p. 685–99.
- 10 Gomes NB, Yoshida A, Roder M, Camargo de Oliveira G, Papa JP. Facial point graphs for amyotrophic lateral sclerosis identification. In: *Proceedings of the 19th international joint Conference on computer vision, Imaging and computer graphics Theory and applications - volume 3: VISAPP.* SciTePress, 2024, pp. 207–14.
- 11 Friesen E, Paul E. Facial action coding system: a technique for the measurement of facial movement. *Palo Alto.* 1978;3:25.
- 12 Zhi R, Liu M, Zhang D. A comprehensive survey on automatic facial action unit analysis. *Vis Comput.* 2020;36(5):1067–93. <https://doi.org/10.1007/s00371-019-01707-5>
- 13 Hamm J, Kohler CG, Gur RC, Verma R. Automated facial action coding system for dynamic analysis of facial expressions in neuropsychiatric disorders. *J Neurosci Methods.* 2011;200(2):237–56. <https://doi.org/10.1016/j.jneumeth.2011.06.023>
- 14 Lucey P, Cohn JF, Matthews I, Lucey S, Sridharan S, Howlett J, et al. Automatically detecting pain in video through facial action units. *IEEE Trans Syst Man Cybern B Cybern.* 2011;41(3):664–74. <https://doi.org/10.1109/TSMCB.2010.2082525>
- 15 Barrios Dell’Olio G, Sra M. FaraPy: an augmented reality feedback system for facial paralysis using action unit intensity estimation. *The 34th annual ACM symposium on user interface software and technology;* 2021. p. 1027–38.
- 16 Oliveira GC, Ngo QC, Passos LA, Papa JP, Jodas DS, Kumar D. Tabular data augmentation for video-based detection of hypomimia in Parkinson’s disease. In: *Comput Methods Programs Biomed,* *Comput Methods Programs Biomed.* 2023;107713. <https://doi.org/10.1016/j.cmpb.2023.107713>
- 17 Oliveira GC, Ngo QC, Passos LA, Oliveira LS, Papa JP, Kumar D. Facial expressions to identify post-stroke: a pilot study. In: *Comput Methods Programs Biomed,* *Comput Methods Programs Biomed.* 2024;250, p. 108195. <https://doi.org/10.1016/j.cmpb.2024.108195>
- 18 Bandini A, Rezaei S, Guarin DL, Kulkarni M, Lim D, Boulos MI, et al. A new dataset for facial motion analysis in individuals with neurological disorders. *IEEE J Biomed Health Inform.* 2021;4:1111–9. <https://doi.org/10.1109/JBHI.2020.3019242>
- 19 Cheong JH, Xie T, Byrne S, Chang LJ. Py-feat: Python facial expression analysis toolbox. *arXiv preprint arXiv:2104.* 2021:03509.
- 20 Chen T, Guestrin C. XGBoost: a scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD international Conference on knowledge Discovery and data mining, KDD ’16.* San Francisco, California, USA: ACM; 2016. p. 785–94. <https://doi.org/10.1145/2939672.2939785>

- 21 Zhang X, Yin L, Cohn JF, Canavan S, Reale M, Horowitz A, et al. Bp4d-spontaneous: a high-resolution spontaneous 3d dynamic facial expression database. *Image Vis Comput.* 2014;32(10):692–706. <https://doi.org/10.1016/j.imavis.2014.06.002>
- 22 Mavadati SM, Mahoor MH, Bartlett K, Trinh P, Cohn JF. Disfa: a spontaneous facial action intensity database. *IEEE Trans Affect Comput.* 2013;4(2):151–60. <https://doi.org/10.1109/t-affc.2013.4>
- 23 Lucey P, Cohn JF, Kanade T, Saragih J, Ambadar Z, Matthews I. The extended cohn-kanade dataset (ck+): a complete dataset for action unit and emotion-specified expression. 2010 ieee computer society conference on computer vision and pattern recognition workshops. IEEE; 2010. p. 94–101.
- 24 Lucey P, Cohn JF, Prkachin KM, Solomon PE, Matthews I. Painful data: the UNBC-McMaster shoulder pain expression archive database. 2011 IEEE international conference on automatic face gesture recognition (FG); 2011. p. 57–64.
- 25 Kollias D, Zafeiriou S. Aff-wild2: extending the aff-wild database for affect recognition. arXiv preprint arXiv:1811; 2018.
- 26 Cox DR. The regression analysis of binary sequences. *J Roy Stat Soc B.* 1958;20(2): 215–32. <https://doi.org/10.1111/j.2517-6161.1958.tb00292.x>
- 27 Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. *J Clin Epidemiol.* 2019;110:12–22. <https://doi.org/10.1016/j.jclinepi.2019.02.004>
- 28 Pedregosa F, Varoquaux G, Gramfort A, Vincent M, Bertrand T, Grisel O, et al. Scikit-learn: machine learning in Python. *J Machine Learn Res.* 2011;12:2825–30.
- 29 Seabold S, Perktold J. Statsmodels: econometric and statistical modeling with python. Austin, TX; 2010. Vol. 57. 61; p. 10–25080. Proceedings of the 9th Python in Science Conference.
- 30 Bandini A, Green JR, Taati B, Orlandi S, Zinman L, Yunusova Y. Automatic detection of amyotrophic lateral sclerosis (ALS) from video-based analysis of facial movements: speech and non-speech tasks. 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) IEEE; 2018. p. 150–7.
- 31 Wu Z, Pan S, Chen F, Long G, Zhang C, Yu PS. A comprehensive survey on graph neural networks. *IEEE Trans Neural Netw Learn Syst.* 2021;32(1):4–24. <https://doi.org/10.1109/tnnls.2020.2978386>
- 32 Cortes C, Vapnik V. Support-vector networks. *Mach Learn.* 1995;20(3):273–97. <https://doi.org/10.1007/bf00994018>
- 33 Abbas A, Yadav V, Smith E, Ramjas E, Rutter SB, Benavidez C, et al. Computer vision-based assessment of motor functioning in schizophrenia: use of smartphones for remote measurement of schizophrenia symptomatology. *Digit Biomark.* 2021;5(1):29–36. <https://doi.org/10.1159/000512383>