Amyotrophic Lateral Sclerosis and Post-Stroke Orofacial Impairment Video-based Multi-class Classification

Allan Magno E. Pecundo*
allan.pecundo@obf.ateneo.edu
Ateneo Laboratory for Intelligent Visual Environment, Ateneo de Manila University
Quezon City, National Capital Region (NCR), Philippines

Patricia Angela R. Abu*
pabu@ateneo.edu
Ateneo Laboratory for Intelligent Visual Environment, Ateneo de Manila University
Quezon City, National Capital Region (NCR), Philippines

Raphael B. Alampay*
ralampay@ateneo.edu
Ateneo Laboratory for Intelligent Visual Environment, Ateneo de Manila University
Quezon City, National Capital Region (NCR), Philippines

ABSTRACT

Neurological diseases, such as ALS and Stroke, that affect the brain including the nerves found throughout the body including the spinal cord generally require various forms of testing and clinical diagnosis in order to detect. These current forms of diagnosis, however, present a limitation in the form of being either expensive or subjective. Research has been done in the area of automated medical assessment via machine learning with the goal of offering cheaper and more objective alternatives for aiding diagnosis. For the case of ALS and orofacial impairment in stroke, it has been shown that using features derived from facial movement in videos, it is possible to detect the presence of these neurological diseases among healthy patients, separately. Research in this area, however, is still relatively few and allows for exploration of improvements in the overall model, especially with the emergence of newer algorithms for detecting facial landmarks.

For this research, the improvements to be explored in the model will come in the form of exploring how the model can be trained to detect both (multi-class) ALS and orofacial impairment in post-stroke among a healthy population. Results show that features calculated from facial landmarks in videos, it is possible to develop a single muti-class detection model ALS, and orofacial impairment in stroke among a healthy population with accuracy as high as 86%.

CCS CONCEPTS

• **Computing methodologies** → Machine learning.

KEYWORDS

classification, neurological disease, video, facial alignment, machine learning

ACM Reference Format:

Allan Magno E. Pecundo, Patricia Angela R. Abu, and Raphael B. Alampay. 2022. Amyotrophic Lateral Sclerosis and Post-Stroke Orofacial Impairment Video-based Multi-class Classification. In *2022 5th Artificial Intelligence and Cloud Computing Conference (AICCC) (AICCC 2022), December 17–19, 2022, Osaka, Japan.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3582099.3582123

*All authors contributed equally to this research.

AICCC 2022, December 17–19, 2022, Osaka, Japan

ACM ISBN 978-1-4503-9874-9/22/12…$15.00
https://doi.org/10.1145/3582099.3582123

1 INTRODUCTION

Neurological disease generally refer to disorders that affect the brain including the nerves found throughout the body including the spinal cord [15]. Neurological disease may be caused by various abnormalities found in the brain and spinal cord ranging from structural to functional [15]. Neurological disease include disorders such as Parkinson's Disease, Stroke, Amyotrophic Lateral Sclerosis (ALS), Cerebral Palsy, and many more [15].These various diseases which damage the nervous system may cause lasting impact on the individual senses and functions such as communication, hearing, and movement; one case in particular are the movements within the orofacial muscles such as the mouth.

ALS is a neurological disease that is characterized by degeneration of both upper (from cortex to the brain stem and spinal cord) and lower (from brain stem and spinal cord to the muscles) motor neurons that lead to motor symptoms such as difficulty in speech and swallowing [6] while Stroke is a neurological disease attributed to an acute focal injury of the central nervous system with causes that are vascular in nature, such as hemorrhage [18].

Both forms of disease are neurological in nature and may lead to degradation of functionality of orofacial muscles in particular as the disease progresses even during the post-stage.

Diagnosis of neurological disorders help provide early diagnosis of the disorder and help the rehabilitation process, especially when functionality of muscles are affected. Diagnosis of these neurological disorders, however, may be done through clinical diagnosis which will require either an evaluation of medical professionals who perform tests that require subjective analysis, or through tests that require instrumentation or tools that are costly such as MRI, EMG, EEG, and more [15].

ALS is generally diagnosed and deduced via clinical diagnosis conducted through a series of processes to narrow down and eliminate possible causes of symptoms as there is no single specific diagnostic test for this condition [6, 15]. EMG test may be used to find evidence of denervation within the muscles such as indicators of reduced motor unit firing rate which is an indicator of ALS [6, 15].

Stroke is diagnosed through various methods, one of which is clinical diagnosis wherein current symptoms as well as the medical history of the patient are considered to narrow down the disease and/or rule out possibilities of other diseases [18]. Various forms of testing can also be done to help diagnose stroke such as CT and MRI can help provide images that localizes regions in the brain affected by hemorrhaging, if any, and helps confirm the presence of stroke through identifying possible vascular damage [18].

As mentioned above, diagnosis may require multiple steps which includes various consultations with experts, and undergoing laboratory tests which might be expensive or invasive.

Computer Vision has been used as a tool to aid diagnosis of multiple medical conditions through different media such as digital images or videos. Standard methods used by doctors and other medical practitioners may require the use of various laboratory tests; many of which may be expensive, time-consuming, relatively subjective, or relatively invasive to the patient [15]. Some research within the area of Computer Vision has been conducted with the aim of providing cheaper and more convenient alternatives of aiding diagnosis of medical conditions providing objective measures that are indicative of disease presence [2, 3, 16].

Studies have tackled automatic detection of neurological disease via various methods, with some research moving in the direction of image or video-based detection using facial alignment models (for extracting important facial features) coupled with classifier models. In recent periods, there have been studies focusing on the development of automated methods for detection of ALS and orofacial impairment in stroke with the goal of providing alternatives for diagnosis, however research within the area is still relatively few in number.

Bandini et al., have worked on the automatic detection of ALS, and Orofacial Impairment in Post-Stroke [2, 3]. These studies have tackled models that can detect each of the neurological disorders separately. This however presents a possible disadvantage wherein knowing which model to use on a patient exhibiting atypical facial structure and movement will require some prior knowledge or hypothesis regarding the current condition of the patient which might takeaway from the value of the detection.

This study proposes to build on this topic by creating a single end-to-end multi-class model that can detect both ALS, and Orofacial Impairment in Post Stroke, among healthy patients without needing prior hypothesis which among the two disorders the patient might have. To the researchers' knowledge, there has been no research tackling this potential challenge using facial videos as inputs.

2 REVIEW OF RELATED LITERATURE

2.1 Overview of Neurological Disease Detection

There have been various methods in which medical practitioners have tried diagnosing neurological diseases. Standard practices of diagnosing neurological disease in the medical field require various kinds of tests or instruments [15]. Neurological tests are usually done to help establish the diagnosis of a patient such as providing functional, or structural abnormalities [15]. The tests conducted help provide more detail on the diagnosis such as the location of areas of the body where the disease is present or is currently affecting [15].

Functional diagnosis of neurological disease examines how well the neurological aspects of the patient functions [15]. Neuropsychological tests are a type of functional diagnosis that aims at evaluating the higher cortical function [15]. Some of these tests, for example, help distinguish patient dementia from other psychological illnesses such as depression [15]. This type of test, while inexpensive, might generally last a long time, between 1 to 4 hours, due to various sets of tests needed to be conducted to the patient to narrow down diagnosis [15].

Other tests conducted for functional diagnosis of neurological disorders are the Electrocephalogram (EEG) wherein electrical activity of found in the superficial layers of the cerebral cortex are measured [15]. In this test, electrodes are usually placed in the scalp to record the brain's electrical activity when awake or asleep [15]. An Electromyogram (EMG) is another functional test which evaluates the electrical function of muscle motor units at rest and during contraction. This test measures electrical activity from the muscle fibers which are interpreted through an oscilloscope wherein the record helps distinguish normal muscle activity from muscle affected by disease and nerve damage [15]. Both tests, which provide a way to measure activity which can present indicators of neurological disease, however still require the judgement and interpretation of a trained medical practitioner.

Structural diagnosis examines where there are any structural damage or defects in the organs affected by neurological disease such as brain tissue. Computed topography (CT) and magnetic resonance imaging (MRI) are one of the most widely used imaging tests that provide high resolution of the brain and surrounding structures [15]. Single Photon/Positron Emission Computed Tomography (SPECT or PET) is another form of testing wherein radiolabeled compounds are injected in tracer amounts which allows photon emissions to be detected [15]. The compounds reflect diagnostics such as blood flow, oxygen and glucose metabolism, or even concentration of specific neurotransmitter receptors [15].

Overall, there are a wide variety of tests that can be conducted to help diagnoses neurological disease detection each with its own objectives and required instruments involved. The challenge with having to diagnose through various tests along with some required judgement of medical practitioners is that it can sometimes be expensive, and in some cases be subjective in terms of judgement as well [14]. With this, exploring the opportunity for the use of automated detection of neurological disease through less specialized method of data collection, such as facial videos, allows for a more accessible alternative for aiding the diagnosis of patients with neurological disorders, such as the case of ALS and orofacial impairment in post-stroke.

2.2 Review of Computer-Aided Diagnosis of Neurological Diseases

There have been studies focusing on computer aided diagnosis and assessment of neurological disease and its conditions, mostly with the goal of providing an automated diagnosis.

One popular approach would be the use of the various outputs of standard clinical testing as inputs and building models around that for automating the diagnosis. One study by Chawla et al. made use of the CT images of stroke patients as inputs and developed a machine learning model for classifying which patients had stroke [7]. Another study by Acharya et al. used certain features in MRI images for detecting ischemic stroke [1].

Another approach made us of collected speech data for automating diagnosis. A study by Wang et al. collected speech samples, along with tongue and lips movement data via attached sensors, and extracted features to try and create models that would automatically classify which patient would had ALS with an accuracy of as high as 92% for acoustic data features and 97% for acoustic and muscle movement data. [23]

With various computer-aided, certain methods still show a limitation due to the certain types of equipment still needed to collect the required data for the analysis or diagnosis, despite the convenience of an automated model present. With this in mind, it is also worth reviewing studies have made use of video-based methods for the assessment of neurological disease.

2.3 Review of Video-Based Automatic Detection of Neurological Disease

Other studies have made use of video-based methods for assessment neurological disease which did not require the use of markers to be attached to the body of the subject, or sophisticated machines to produce the outputs. The studies mostly relying on video-based assessments made use of face data as inputs due to the various indicators of neurological diseases being present in facial data.

Facial assessment for various symptoms have been conducted when diagnosing specific medical conditions as facial muscle activity can provide important information on the health condition of a subject [22]. For example, involuntary movement of facial muscles is a characteristic found in multiple forms of diseases such as hemifacial spams, Meige syndrome, and dyskenisia [21]. For ALS, some symptoms which manifest from the loss of bulbar lower motor neurons cause lower facial muscle weakness and atrophy [15]. For stroke, difficulty in swallowing and are some of the symptoms manifested through the face [18].

Noting the importance of facial features in the diagnosis of some diseases, research has been done in the area of computer-vision aided diagnosis of neurological disease using facial data.

A study by Guarin et al. developed an automatic facial landmark localization model specific to patients with facial palsy [11].

A study by Bandini et al. made use of kinematic measurements of orofacial movements which have been seen to help distinguish the phase in which the disorder progresses, pre-symptomatic or symptomatic, specifically in the form of bulbar decline in ALS [4]. It was shown in the study that by capturing the movement of lips and jaw muscles in the face via motion-capture systems with infrared markers, and expressing the the recorded movements as features, such as velocity and range of motion, a machine learning model can use these inputs to determine whether patients are in pre-symptomatic or symptomatic phase of bulbar decline [4].

In a study by Schimmel et al., facial muscle impairment in patients with hemispheric stroke was assessed via 3D video system wherein various facial motions and asymmetries were measured and compared against a control group of healthy patients; video captured results that showed that lower facial muscles were more affected by hemispheric stroke [20].

In a study by Bandini et al., a 3D-camera was used to capture various participants doing different facial exercises, the videos of which were used to classify exercises/activities and using the facial movements as inputs by extracting features such as facial asymmetry, speed of motion, range of motion, facial asymmetry, and facial eccentricity which was used to create a model that classified which patients had ALS among patients in the healthy control group with accuracy as high as 89%[3]. In another study by Bandini et al., another set of features extracted, such as facial asymmetry and range of motion, from different videos of participants doing various facial exercises was done to create a model that would automatically detect whether the patient would have orofacial impairment due to post-stroke among healthy patients doing the same set of exercises [2]. The study was able to create a model that would automatically detect orofacial impairment in post-stroke with accuracy as high as 87% [2].

In a study by Parra-Dominguez et al., a model was built to identify facial paralysis (class defined by having either ALS, Facial Palsy, or Orofacial Impairment in Post-stroke), through photographs of patients from the Massachussets Eye and Ear Infirmary (MEEI) database and Toronto Neurofaces dataset (TNF) [17]. Unlike the mentioned studies of Bandini et al. [2, 3], this study made use only of the facial photographs and relying only of facial asymmetry as input features, extracted via facial landmarks, for the final classifier model.

3 METHODOLOGY

3.1 Background of the Dataset

The dataset proposed for the study will be the Toronto Neurofaces dataset, a public dataset of 261 videos of individuals that with health conditions such as (1) healthy, (2) orofacial impairment through post-stroke, and (3) ALS performing various facial tasks [5]. The dataset also comes with 3306 representative frames from the various videos with accompanying facial landmark annotation, following the MULTI-PIE 2D 68-point format, a commonly used format in facial landmarking datasets such as the known 300W (300 Faces-in-the-Wild) dataset [9, 19].

For the data collection, 36 participants were recruited with 11 patients with ALS, 14 patients with post-stroke, and 11 healthy subjects to act as a control group. Participants were asked to be seated in front of the camera within a distance between 30 cm and 60 cm with a light source used to illuminate the face.

Each participant was then asked to perform various set of tasks, both speech and non-speech, which are commonly asked during clinical examination [8, 24]. The following tasks are summarized in the Table 1.

In total, there are around 261 videos recorded in the dataset: 80 representing Healthy subjects, 76 ALS patients, and 105 post-stroke patients.

From the videos recorded of the participants, 3306 frames were extracted on the basis of being representative of the movements conducted during the video. The frames were then manually annotated using the Multi-PIE 2D format, which uses 68 points to represent facial landmarks. The facial landmarks were provided to allow researchers to further train or tune facial alignment models to account for any differences in facial landmark placements or

Table 1: Description of Facial Tasks Performed by Participants

Subtask	Description
BBP	Repetition of the phrase "Buy Bobby a Puppy"
PA	Repetition of the syllable /pa/ as fast as possible in a single breath
PATAKA	Repetition of the syllables /pataka/ as fast as possible in a single breath
BLOW	Puckering lips in a manner pretending to blow a candle
KISS	Puckering lips in a manner pretending to kiss a baby
OPEN	Open the jaw as large as possible
SPREAD	Pretending to smile with lips tightened
BIGSMILE	Making a big smile
BROW	Raising the eybrows

behavior in patients affected with orofacial impairments due to neurological disorders, accounting for possible bias of pre-trained models to healthy faces.

3.2 Preparation of the Dataset Split

For the experiment, each of the videos were manually segmented given that tasks were performed multiple times in a single video. This allowed each video to provide multiple repetitions of each task, allowing for more observations to be used in the modelling which was also done by Bandini et. al when conducting their research using this dataset[10]. The subtasks used for modelling were also limited to just 7 out of the 9 subtasks, with "PA" and "PATAKA" subtasks excluded.

After manually segmenting the videos, the dataset was further split into a total of 1105 repetitions. The following number of repetitions per subtask are as follows on Table 2:

Table 2: Number of Repetitions (by Train and Test Set)

Subtask	Total	Training Set	Test Set
BBP	**331**	261	70
BLOW	**116**	99	17
KISS	**181**	145	36
OPEN	**181**	147	34
SPREAD	**174**	138	36
BIGSMILE	**49**	34	15
BROW	**73**	53	20

3.3 Facial Alignment

In this study, features that will be used for classifying will be derived from various aspects of facial structure and movement. This proposed flow will follow the same set of steps in the previous studies done on automatic neurological disease detection, namely (1) facial alignment for detection of facial landmarks, and (2) calculation of relevant features from the detected facial landmarks [2, 3, 10].

The proposed facial alignment model to be used will be Ensemble Regression Tree by Kazemi and Sullivan [12] implemented on the dlib library [13].

3.4 Creating Final Set of Input Features

For this study, the input features will be calculated from the identified 68 point facial landmarks via facial alignment model; with 9 specific points identified as relevant for the calculation of all the features due to the nature of the neurological disorder affecting mostly the jaw and lip movement [6, 18]. These specific features were also taken from previous studies done on the detection of ALS and orofacial impairment in Stroke [2, 3]. The specific points are illustrated here on Figure 1.

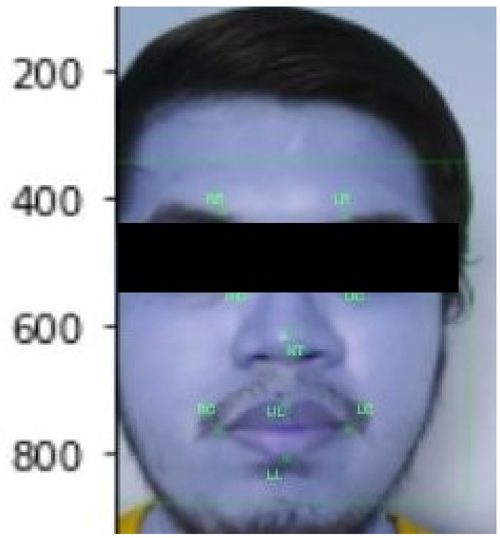

Figure 1: Illustration of 9 Facial Landmarks used for Input Feature Calculation

The 9 points represent the location and movement of the eyebrows, eyes, nose, and mouth. Specific details of each of the 9 points are described on Table 3.

Table 3: Description of 9 Facial Landmarks

Point	Description	Body Part
RE	Right Eyebrow	Eyebrows
LE	Left Eyebrow	
RIC	Right Inner Canthus of the Eye	Inner Canthus of Eyes
LIC	Left Inner Canthus of the Eye	
NT	Nose Tip	Nose
UL	Upper Lip	Lips
LL	Lower Lip	
RC	Right Mouth Corner	Mouth Corners
LC	Left Mouth Corner	

The study makes use of a combination of 14 features taken from the studies of Bandini et al, representing facial range of motion, facial asymmetry, and face shape [2, 3]. The set of features proposed, following the notations of the related literature, are presented on Table 4 below [2, 3].

After the features are created, standard (z-score) normalization will be applied to the final input features before it is passed on to the model similar to the previous studies [2, 3].

Table 4: List of Features Used

Representation	Feature	Description
Facial Asymmetry	A_{diff}	Mean of differences in areas formed by (1) left mouth corner, upper lips, and lower lips, and (2) right mouth corner, upper lips, and lower lips
	d_0	Mean of differences in distance formed by Eyebrows and Inner Canthus of the Eyes on left and right sides of the face
	d_1	Mean of differences in distance formed by Mouth Corners and Inner Canthus of the Eyes on left and right sides of the face
	d_2	Mean of differences in distance formed by the Eyebrows and Upper Lip on left and right sides of the face
	d_3	Mean of differences in distance formed by the Mouth Corners and Upper Lip on left and right sides of the face
	r_{LCRC}	Correlation between trajectory (previous frame vs current frame) of Left Cheek and Right Cheek
Range of Motion (ROM)	W_{min}	Minimum mouth opening found
	W_{max}	Maximum mouth opening found
	O_{min}	Minimum lip opening found
	O_{max}	Maximum lip opening found
	A'_{mean}	Mean value of area formed by the mouth with respect to area formed by mouth during at-rest position
	$\Delta A'$	Range (max - min) of area formed by the mouth with respect to area formed by mouth during at-rest position
Mouth Shape and Geometry	e_{mean}	Mean value of mouth eccentricity (calculated as an index involving mouth opening and lip opening)
	Δe	Range (max - min) value of mouth eccentricity (calculated as an index involving mouth opening and lip opening)

3.5 Classification and Evaluation

After creating the different features from the identified facial landmarks, the final step of the process is classifying the different observations as either being part of the (1) ALS, (2) Post-stroke, or (3) Healthy class. The classifier models that will be tested in making the prediction based on the input features are the Random Forest model, SVM model, and K Nearest Neighbors model. The Random Forest model was trained on n = 100 trees. K Nearest Neighbors used n = 5 neighbors. Lastly, the SVM model uses a radial basis function (RBF) kernel with regularization parameter c = 1. There will be three classifier models trained and evaluated per each of the 7 subtasks given the different types of movement done by the mouth for each of the subtasks.

After the various models are run, the end-to-end model will be evaluated using a train and test set, split by subject or participant similar to the evaluation in the studies of Bandini et al. with accuracy, precision, and recall as the metrics to be measured [2, 3]. For each subtask, the best performing model among the 3 mentioned will be highlighted.

The metrics will be evaluated on a subject basis (e.g., entire video counted as one observation) rather than a repetition basis (e.g., one repetition or video-split counts as one observation). Since videos of the subject are composed of multiple repetitions, the final prediction of a video (e.g., whether the subject participant performing a certain task) class will be the majority prediction of the model on the repetitions segmented from the video (e.g., if a video has 10 repetitions with predictions being 4 "healthy", 3 "ALS", and 3 "Post-stroke", then the prediction is that the subject is "healthy").

In summary, the general flow follows the flowchart illustrated below on Figure 2

4 RESULTS

The end-to-end model for detecting neurological disease performance was measured in terms of its multi-class accuracy, precision, and recall to allow for balanced and easy interpretation of performance. Each model was trained and evaluated on a per subtask basis.

The classification accuracy of each model per subtask is as follows on Figure 3 with breakdown of the best performing model of each subtask, including other metrics, shown on Table 5 as follows:

From the different subtasks performed, we see that the model trained and evaluated on the "SPREAD" subtask performed the

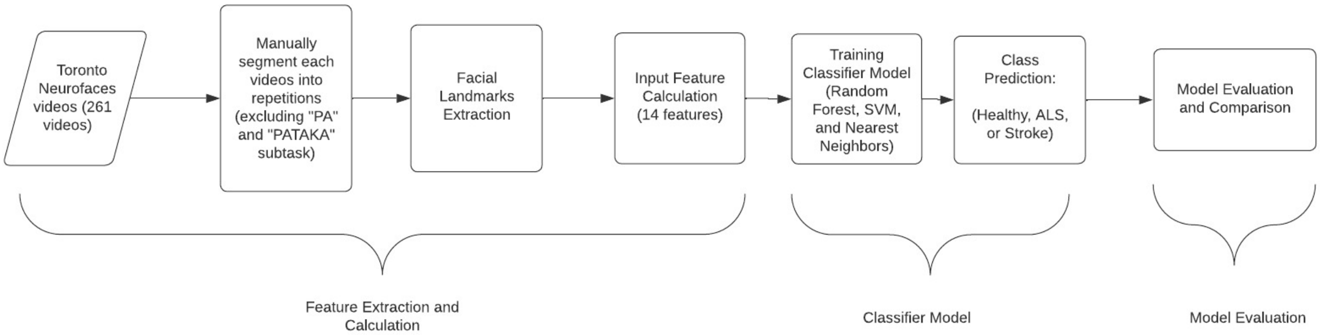

Figure 2: End-to-end Process Flow

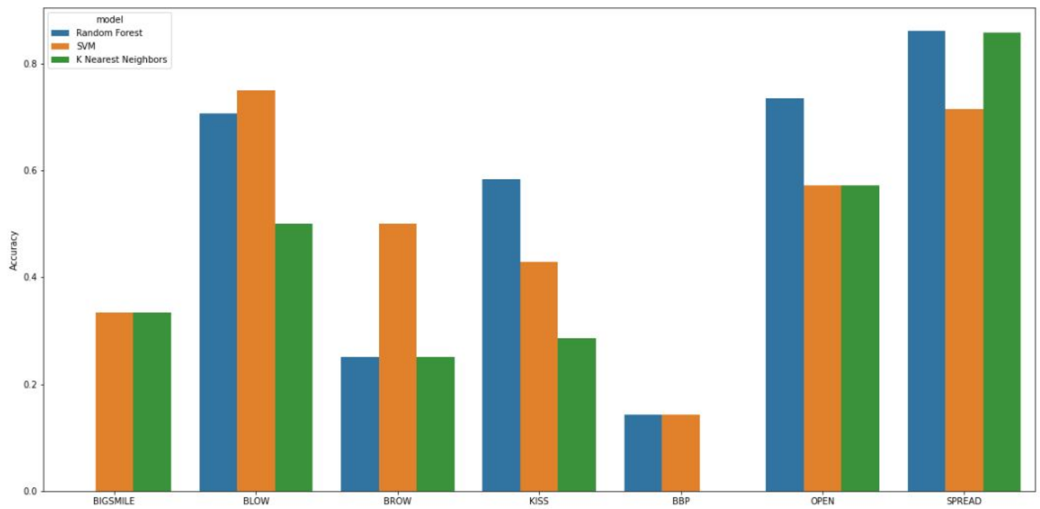

Figure 3: Model Accuracy per Subtask

Table 5: Model Performance per Subtask

Subtask	Accuracy	Precision	Recall	Best Performing Model
BBP	0.14	0.14	0.14	Random Forest / SVM
BLOW	0.75	0.88	0.75	SVM
KISS	0.57	0.42	0.57	Random Forest
OPEN	0.71	0.76	0.71	Random Forest
SPREAD	**0.86**	**0.89**	**0.86**	Random Forest
BIGSMILE	0.33	0.17	0.33	SVM / Nearest Neighbors
BROW	0.50	0.38	0.50	SVM

best with an accuracy of 86%. This result is aligned with the previous studies wherein the each of the binary classifiers for ALS and Post-stroke orofacial impairment both performed well, above 80% accuracy, within the "SPREAD" subtask [2, 3]. For the "BROW" subtask, we see that it does not perform well and this is possibly due to the lack of movement required from the muscles near the mouth area as compared to the other tasks.

Looking at the distribution of values of the certain features between the "KISS", "SPREAD", and "BROW" subtask, we see that the asymmetry in facial movements during the task (e.g., the r_{LCRC} feature) differs the most across the different classes during the "SPREAD" subtask as illustrated in Figure 4; whereas in the "KISS" subtask and "BROW" subtask, the differences are not as prominent. In the "SPREAD" subtask, we see that both the patients with ALS and Post-stroke Orofacial Impairment possibly manifest their facial weakness in the form of lowered asymmetry in facial movement when performing tasks vs healthy patients with post-stroke patients manifesting greater facial asymmetry. For the "KISS", and "BROW" subtask, it is possible that the weakness of facial muscles do not manifest as much in these tasks due to either the area of the muscles affected when performing the task (in the case of the "BROW" task being performed on the upper part of the face), or the level of effort required by the muscles in performing the tasks ("KISS" task does not require face to move as wide of a range compared to the "SPREAD" subtask).

As for the comparison of the model performances, we see that across most of the tasks, Random Forest yields the highest accuracy metric except for the tasks "BLOW", and "BROW" where SVM yields the best performance. As for KNN, it does not perform as well as Random Forest in most cases except for the "SPREAD" subtask, and the "BIGSMILE" subtask wherein all models perform poorly.

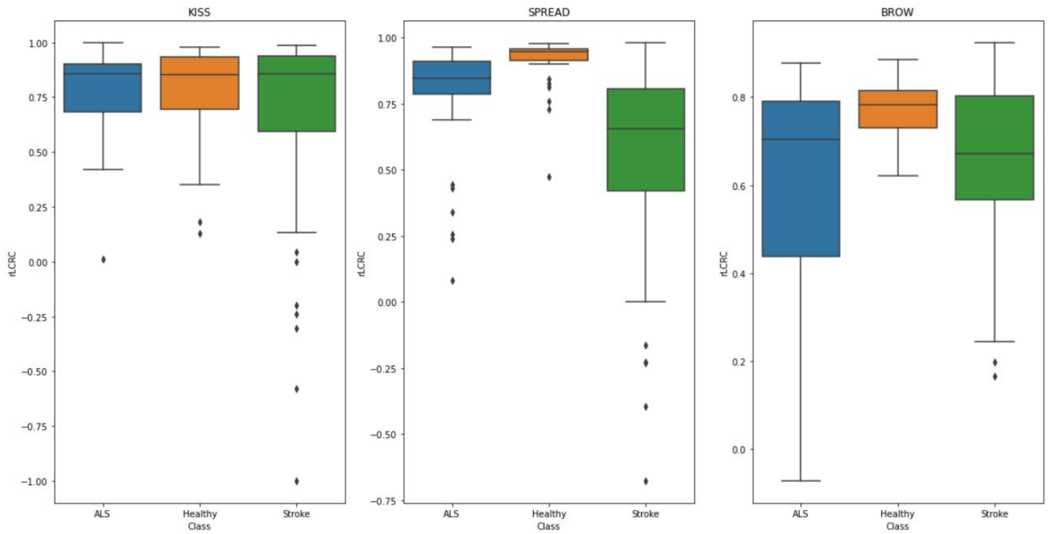

Figure 4: Comparison of r$_{LCRC}$ feature by Class per Subtask

For the most tasks, it is possible that Random Forest performs well due to the complex relationships and conditions present in the variables that require for distinguishing the facial movement and attributes between 3 distinct classes of neurological condition; something that tree-based classification model is able to determine based on its method for model training. As for the "BLOW" and "BROW" task, it is possible that these tasks present features which are more linearly separable, and possibly less complex in relationship, in terms of distinguishing between classes, which is why the the SVM model performs the best.

5 DISCUSSION AND CONCLUSION

From the results, we see that using different features derived from the face and its movements altogether, it is possible to create a single multi-class model that can detect whether an individual may have ALS, Orofacial Impairment in Post-Stroke, or is Healthy.

This allows identification of these neurological diseases with less need for a prior hypothesis on what neurological disease a patient may by have through the symptoms exhibited on the face, something which would normally be required should models for identifying be separate per type of disease.

From the results, we see that the best performing models were found when applied to the patients performing the SPREAD subtask. These findings show a multiclass accuracy of 86% on the test set with balanced class idenfitication as seen in the similar performance of the precision and recall at 89% and 86%, respectively.

This allows future users of this model to know which subtasks should be recommended for patients to perform on the video to allow the model to best distinguish, based on the highest accuracy show in the experiments, whether they have the disease and which disease it is.

In this work, we see that there is still room for improvement in the accuracy of the classification for the various subtasks, including the best performing models found in specific subtasks. Compared to the prior research which used separate binary classifier models for each neurological disorder, while the accuracy of best performing models in this experiment fall 1% lower in terms of accuracy, the added advantage of simplifying classification of both neurological disorders into one model becomes available to the possible users; reducing the need of having a prior assumption on which of the two neurological disorders, both of which manifest symptoms in the form of facial muscle weakness, the patient might have before deciding which model to use.

For future studies, other facial alignment models may be considered to explore whether improvement in the accuracy of the detection of the facial landmarks will result in a higher accuracy of the model due to having a more accurate representation of the features extracted. Testing other more complex classifiers (e.g., XGBoost) along with hyperparameter tuning of the models used in this study may also be tested for all the different subtasks to see whether some classifier models will yield better accuracy for different subtasks performed.

ACKNOWLEDGMENTS

(Portions of) the research in this paper uses the NeuroFace Database collected by Dr. Yana Yunusova and the Vocal Tract Visualization and Bulbar Function Lab teams at UHN-Toronto Rehabilitation Institute and Sunnybrook Research Institute respectively, financially supported by the Michael J. Fox Foundation, NIH-NIDCD, Natural Sciences and Engineering Research Council, Heart and Stroke Foundation Canadian Partnership for Stroke Recovery and AGE-WELL NCE.

We would like to thank the researchers who have carefully prepared this dataset for research purposes and for allowing us access to help further our research.

REFERENCES

[1] U. Rajendra Acharya, Kristen M. Meiburger, Oliver Faust, Joel En Wei Koh, Shu Lih Oh, Edward J. Ciaccio, Asit Subudhi, V. Jahmunah, and Sukanta Sabut.

2019. Automatic detection of ischemic stroke using higher order spectra features in brain MRI images. *Cognitive Systems Research* 58 (Dec. 2019), 134–142. https://doi.org/10.1016/j.cogsys.2019.05.005

[2] Andrea Bandini, Jordan Green, Brian Richburg, and Yana Yunusova. 2018. Automatic Detection of Orofacial Impairment in Stroke. In *Interspeech 2018*. ISCA. https://doi.org/10.21437/interspeech.2018-2475

[3] Andrea Bandini, Jordan R. Green, Babak Taati, Silvia Orlandi, Lorne Zinman, and Yana Yunusova. 2018. Automatic Detection of Amyotrophic Lateral Sclerosis (ALS) from Video-Based Analysis of Facial Movements: Speech and Non-Speech Tasks. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*. IEEE. https://doi.org/10.1109/fg.2018.00031

[4] Andrea Bandini, Jordan R. Green, Jun Wang, Thomas F. Campbell, Lorne Zinman, and Yana Yunusova. 2018. Kinematic Features of Jaw and Lips Distinguish Symptomatic From Presymptomatic Stages of Bulbar Decline in Amyotrophic Lateral Sclerosis. *Journal of Speech, Language, and Hearing Research* 61, 5 (May 2018), 1118–1129. https://doi.org/10.1044/2018_jslhr-s-17-0262

[5] Andrea Bandini, Sia Rezaei, Diego L. Guarín, Madhura Kulkarni, Derrick Lim, Mark I. Boulos, Lorne Zinman, Yana Yunusova, and Babak Taati. 2021. A New Dataset for Facial Motion Analysis in Individuals With Neurological Disorders. *IEEE Journal of Biomedical and Health Informatics* 25, 4 (2021), 1111–1119. https://doi.org/10.1109/JBHI.2020.3019242

[6] Robert H. Brown and Ammar Al-Chalabi. 2017. Amyotrophic Lateral Sclerosis. *New England Journal of Medicine* 377, 2 (July 2017), 162–172. https://doi.org/10.1056/nejmra1603471

[7] M. Chawla, S. Sharma, J. Sivaswamy, and L.T. Kishore. 2009. A method for automatic detection and classification of stroke from brain CT images. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE. https://doi.org/10.1109/iembs.2009.5335289

[8] Joseph R. Duffy. 2000. Motor Speech Disorders: Clues to Neurologic Diagnosis. In *Parkinson's Disease and Movement Disorders*. Humana Press, 35–53. https://doi.org/10.1007/978-1-59259-410-8_2

[9] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker. 2010. Multi-PIE. *Image and Vision Computing* 28, 5 (May 2010), 807–813. https://doi.org/10.1016/j.imavis.2009.08.002

[10] Diego L. Guarin, Andrea Bandini, Aidan Dempster, Henry Wang, Siavash Rezaei, Babak Taati, and Yana Yunusova. 2021. Improving Face Alignment Accuracy on Clinical Populations and its effect on the Video-based Detection of Neurological Diseases. (11 2021). https://doi.org/10.36227/techrxiv.12950279.v2

[11] Diego L. Guarin, Yana Yunusova, Babak Taati, Joseph R. Dusseldorp, Suresh Mohan, Joana Tavares, Martinus M. van Veen, Emily Fortier, Tessa A. Hadlock, and Nate Jowett. 2020. Toward an Automatic System for Computer-Aided Assessment in Facial Palsy. *Facial Plastic Surgery & Aesthetic Medicine* 22, 1 (Feb. 2020), 42–49. https://doi.org/10.1089/fpsam.2019.29000.gua

[12] Vahid Kazemi and Josephine Sullivan. 2014. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 1867–1874. https://doi.org/10.1109/CVPR.2014.241

[13] Davis E. King. 2009. Dlib-ml: A Machine Learning Toolkit. *Journal of Machine Learning Research* 10 (2009), 1755–1758.

[14] Jane Larkindale, Wenya Yang, Paul F. Hogan, Carol J. Simon, Yiduo Zhang, Anjali Jain, Elizabeth M. Habeeb-Louks, Annie Kennedy, and Valerie A. Cwik. 2014. Cost of illness for neuromuscular diseases in the United States. *Muscle & Nerve* 49, 3 (Jan. 2014), 431–438. https://doi.org/10.1002/mus.23942

[15] M.D. Larry E. Davis and Assistant P Sarah Pirio Richardson, M.D. 2015. *Fundamentals of Neurologic Disease*. Springer New York. https://doi.org/10.1007/978-1-4939-2359-5

[16] Raquel Norel, Mary Pietrowicz, Carla Agurto, Shay Rishoni, and Guillermo Cecchi. 2018. Detection of Amyotrophic Lateral Sclerosis (ALS) via Acoustic Analysis. In *Interspeech 2018*. ISCA. https://doi.org/10.21437/interspeech.2018-2389

[17] Gemma S. Parra-Dominguez, Raul E. Sanchez-Yanez, and Carlos H. Garcia-Capulin. 2021. Facial Paralysis Detection on Images Using Key Point Analysis. *Applied Sciences* 11, 5 (March 2021), 2435. https://doi.org/10.3390/app11052435

[18] Ralph L. Sacco, Scott E. Kasner, Joseph P. Broderick, Louis R. Caplan, J.J. (Buddy) Connors, Antonio Culebras, Mitchell S.V. Elkind, Mary G. George, Allen D. Hamdan, Randall T. Higashida, Brian L. Hoh, L. Scott Janis, Carlos S. Kase, Dawn O. Kleindorfer, Jin-Moo Lee, Michael E. Moseley, Eric D. Peterson, Tanya N. Turan, Amy L. Valderrama, and Harry V. Vinters. 2013. An Updated Definition of Stroke for the 21st Century. *Stroke* 44, 7 (July 2013), 2064–2089. https://doi.org/10.1161/str.0b013e318296aeca

[19] Christos Sagonas, Epameinondas Antonakos, Georgios Tzimiropoulos, Stefanos Zafeiriou, and Maja Pantic. 2016. 300 Faces In-The-Wild Challenge: database and results. *Image and Vision Computing* 47 (March 2016), 3–18. https://doi.org/10.1016/j.imavis.2016.01.002

[20] M. SCHIMMEL, B. LEEMANN, P. CHRISTOU, S. KILIARIDIS, F. R. HERRMANN, and F. MÜLLER. 2011. Quantitative assessment of facial muscle impairment in patients with hemispheric stroke. *Journal of Oral Rehabilitation* 38, 11 (March 2011), 800–809. https://doi.org/10.1111/j.1365-2842.2011.02219.x

[21] N.-C. TAN. 2002. Hemifacial spasm and involuntary facial movements. *QJM* 95, 8 (Aug. 2002), 493–500. https://doi.org/10.1093/qjmed/95.8.493

[22] Jerome Thevenot, Miguel Bordallo Lopez, and Abdenour Hadid. 2018. A Survey on Computer Vision for Assistive Medical Diagnosis From Faces. *IEEE Journal of Biomedical and Health Informatics* 22, 5 (Sept. 2018), 1497–1511. https://doi.org/10.1109/jbhi.2017.2754861

[23] Jun Wang, Prasanna V. Kothalkar, Beiming Cao, and Daragh Heitzman. 2016. Towards Automatic Detection of Amyotrophic Lateral Sclerosis from Speech Acoustic and Articulatory Samples. In *Interspeech 2016*. ISCA. https://doi.org/10.21437/interspeech.2016-1542

[24] Yana Yunusova, Jordan R. Green, Jun Wang, Gary Pattee, and Lorne Zinman. 2011. A Protocol for Comprehensive Assessment of Bulbar Dysfunction in Amyotrophic Lateral Sclerosis (ALS). *Journal of Visualized Experiments* 48 (Feb. 2011). https://doi.org/10.3791/2422