Facial Nerve Paralysis Severity Grading by Computer Vision and Machine Learning

Bingnan Huo

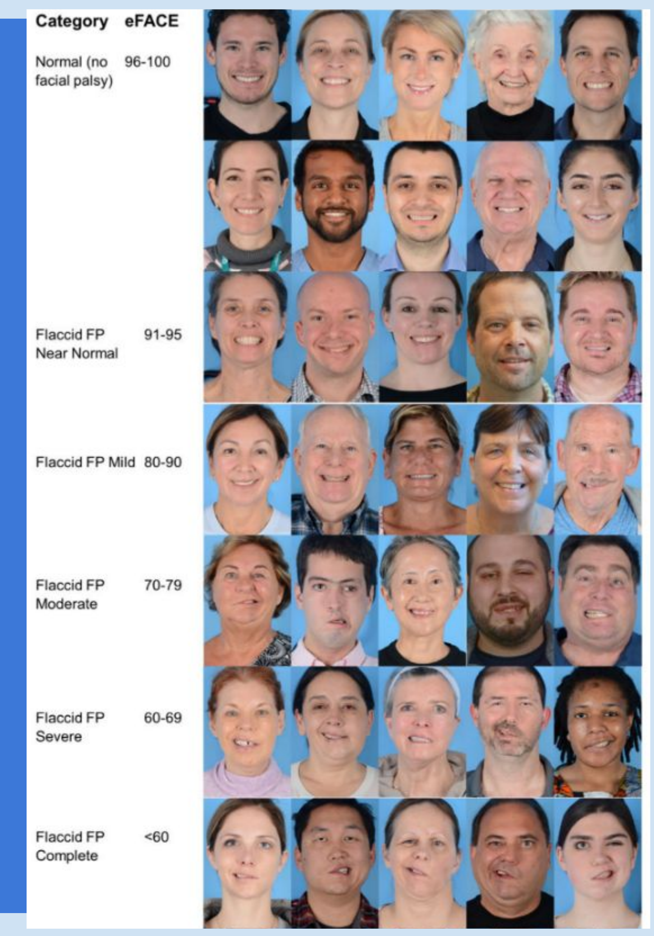
Advisors: Dr. Keith Buffinton[1], Dr. Joshua Stough[2], Dr. Arun Gadre[3]

[1]Department of Mechanical Engineering / [2]Department Computer Science, Bucknell University, Lewisburg, PA

[3]Geisinger Commonwealth Medical School, Geisinger Medical Center

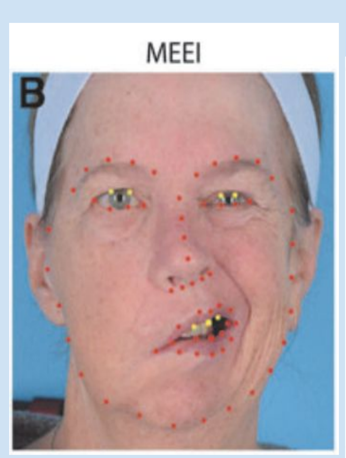
Background & Motivation

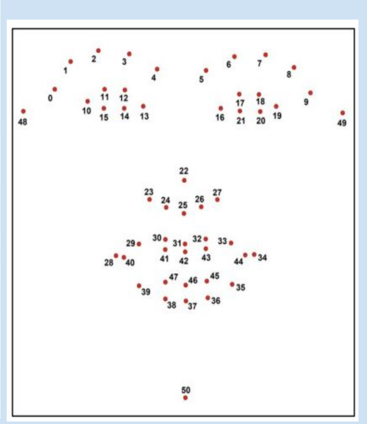
- **Goal**: Develop computer vision and machine learning software to grade **Facial Nerve Paralysis** (FNP) and to assist diagnosis and recovery tracking.
- **Observer bias** commonly arises when FNP patients are seen and diagnosed by clinicians, as reported in [1], which showed that a machine learning (ML) based approach found less facial asymmetry in severe FNP patients and more asymmetry in healthy faces than clinicians.
- We hope to train an ML model that grades patient facial palsy severity on the House-Brackmann scale.

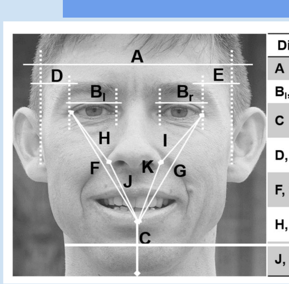


Approach

Facial Landmarks Extraction: Landmark prediction & head tilt correction → **(Optional) Facial Feature Computation**: Distance calculation & feature computation → **Machine Learning**: Severity grading







Distance	Description
A	Distance between the first and last jaw points
B_l, B_r	Width of the eye (left and right)
C	Distance from the middle mouth point to the middle of the jaw point
D, E	Distance from the outer eye corner to the head side
F, G	Distance from the outer eye corner to the middle mouth point
H, I	Distance from the outer eye corner to the nose edge point
J, K	Distance from the nose edge point to the middle mouth point

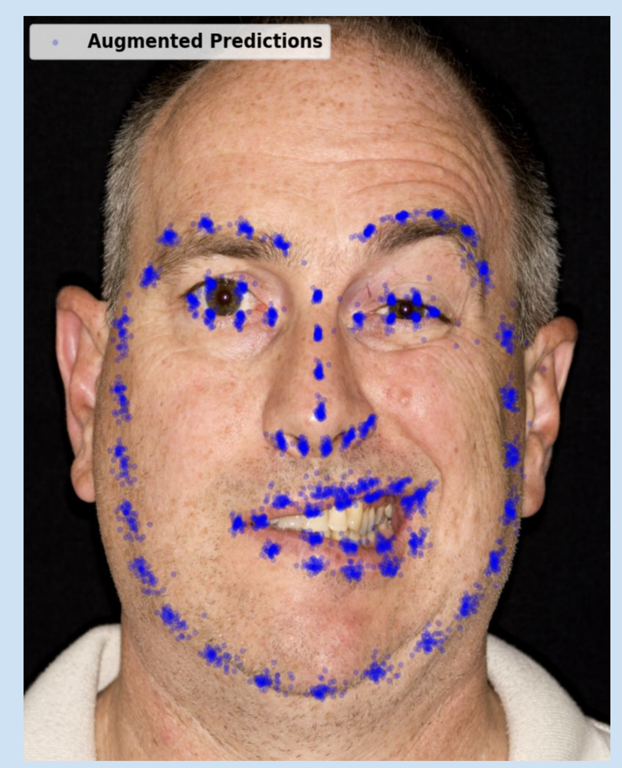
Feature	Face Element	Description	Formula
$f0$	Eyebrows	Angle	$\lvert\angle(P0, P9)\rvert$
$f1$	Eyebrows	Angle	$\lvert\angle(P2, P7)\rvert$
$f2$	Eyebrows	Angle	$\lvert\angle(P4, P5)\rvert$
$f3$	Eyebrows	Maximum value	$\max(L/M, M/L)$
$f4$	Eyebrows	Slope	$m(P0, P9)$
$f5$	Eyebrows	Slope	$m(P2, P7)$
$f6$	Eyebrows	Slope	$m(P4, P5)$
$f7$	Eyes	Angle	$\lvert\angle(P10, P19)\rvert$
$f8$	Eyes	Maximum value	$\max(B_l/B_r, B_r/B_l)$
$f9$	Eyes	Maximum value	$\max(D/E, E/D)$
$f10$	Eyes	Maximum value	$\max(H/I, I/H)$

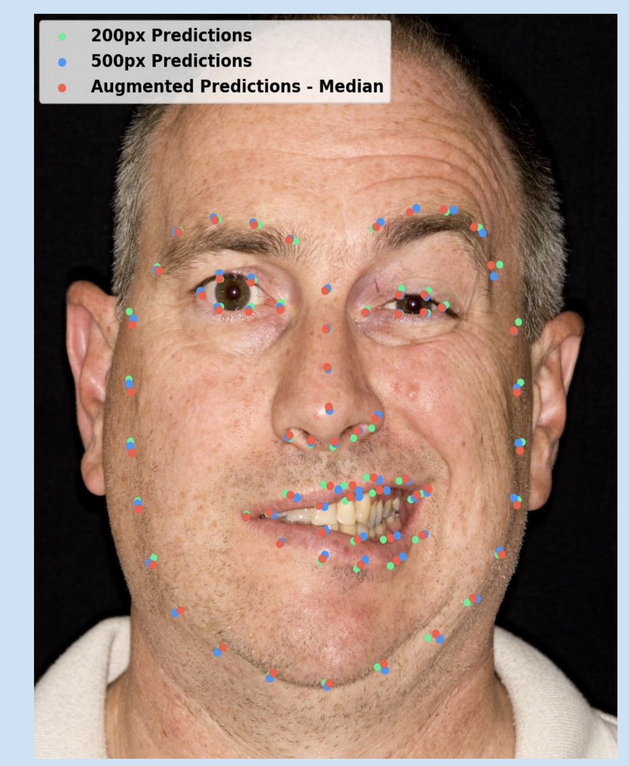
Grade	Appearance
Grade I - normal	
Grade II - slight dysfunction	Motion: Forehead - moderate to good function Eye - complete closure with minimum effort Mouth - slight asymmetry
Grade III - moderate dysfunction	Motion: Forehead - slight to moderate movement Eye - complete closure with effort; Mouth - slightly weak with maximum effort
Grade IV - moderate severe dysfunction	Motion: Forehead - none; eye - incomplete closure Mouth - asymmetric with maximum effort
Grade V - severe dysfunction	Motion: Forehead - none; eye - incomplete closure Mouth - slight movement mouth - slight movement
Grade VI - total paralysis	No movement

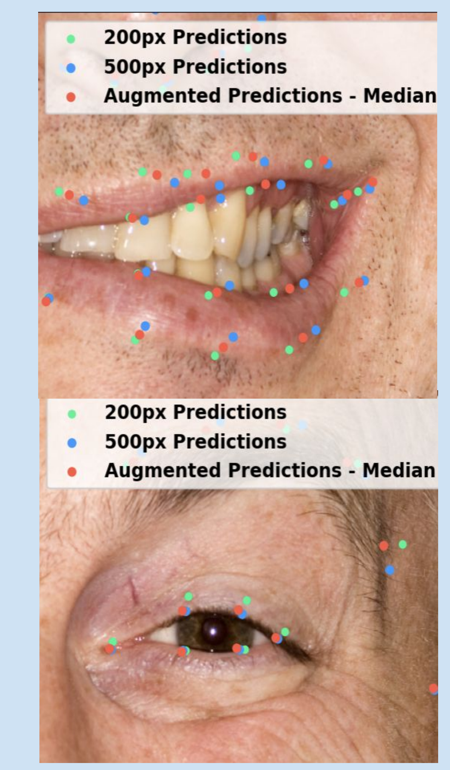
- Work in [2] demonstrated successful application of ML in determining whether a face is healthy or ill with FNP by using the ML-based software *Emotrics* [3] to automatically predict key facial landmarks from facial images and then computing facial features to determine health or illness.
- ***Emotrics*** [3] is used as a landmark detector after training on a dataset [4] consisting of 60 patients with a spectrum of types and severities of FNP.
- A landmark detector trained on the **patient population** minimizes potential **model bias** [5].

Validating Landmark Detection

- Emotrics [2] claims to have accurate results under various conditions; however, small image changes, i.e., rotations or adding noise, lead to less accurate results.
- Applying **test-time augmentation** leads to better results by intentionally introducing random small rotations and/or noise and using the resulting median of the predicted landmarks.



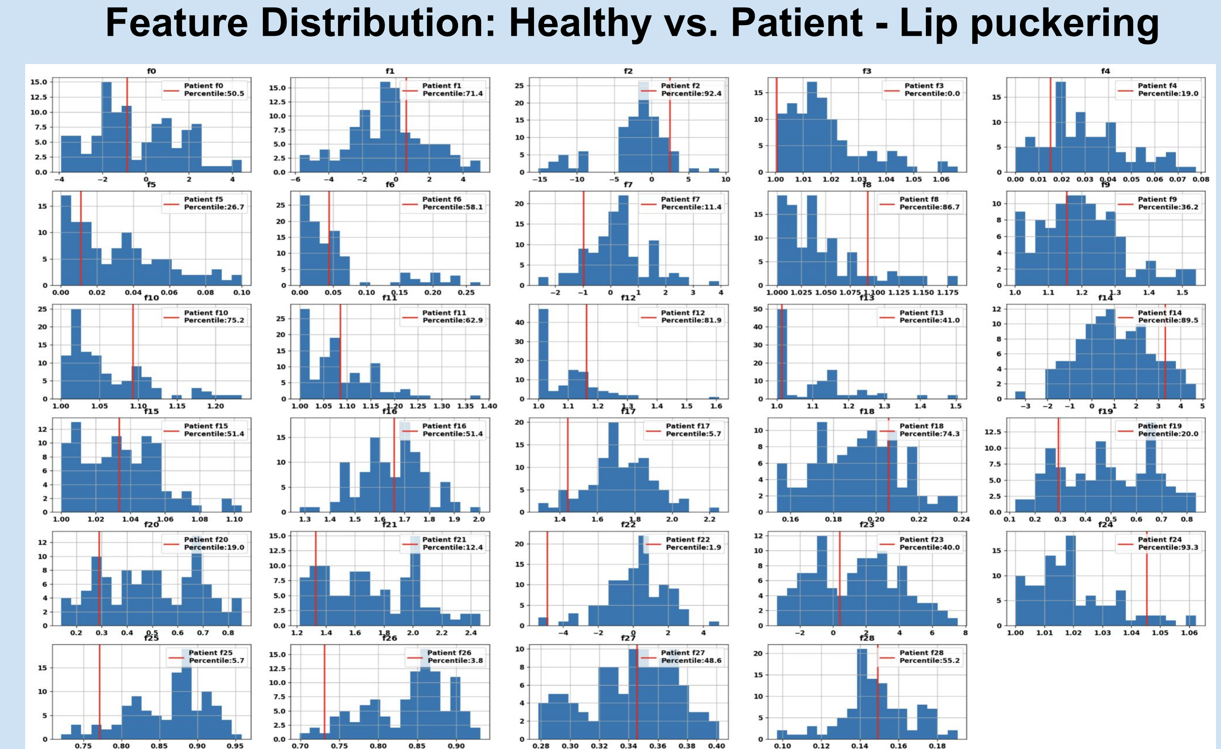


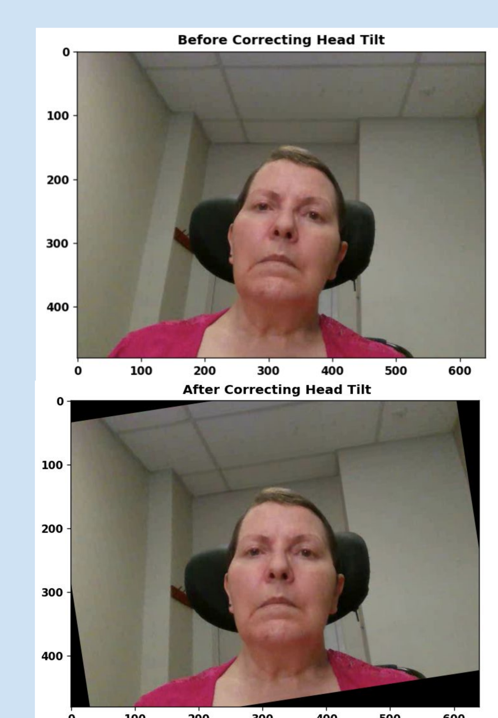


Feature Computation

- While feature computation can be used to train the ML algorithm to grade the severity of FNP, landmarks alone can be used in the training and may reduce structural bias.
- Features also serve informative purposes by highlighting certain characteristics (i.e., asymmetry in certain regions of the face) that are abnormal relative to healthy patients.

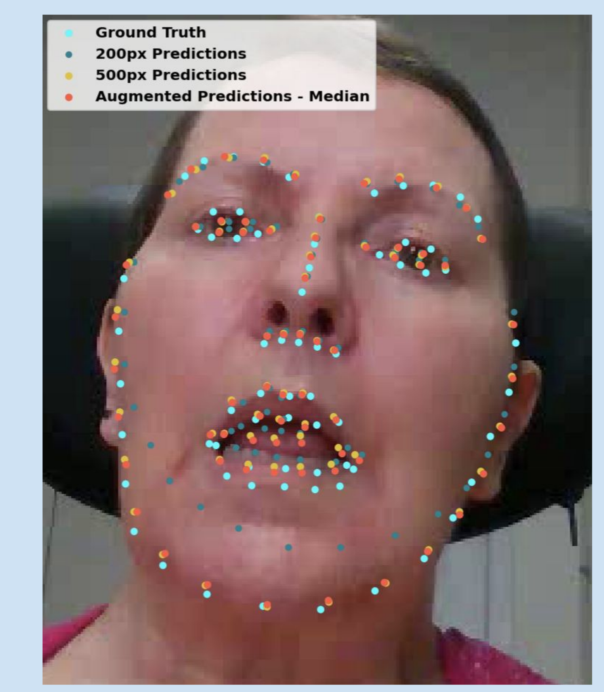
Feature Distribution: Healthy vs. Patient - Lip puckering

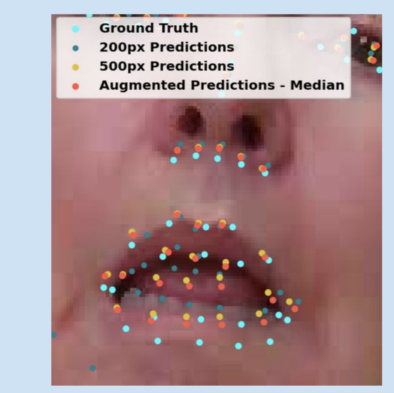


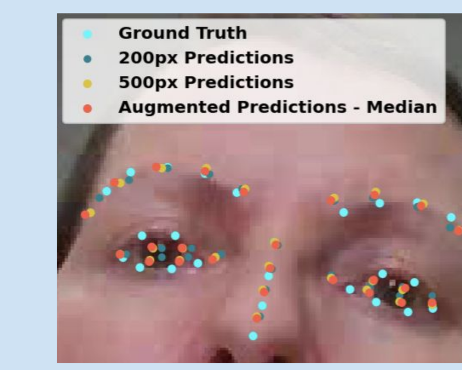


Preliminary Results

- Difference between predicted landmarks and Ground Truth landmarks using the Toronto NeuroFace (TNF) Dataset [5].
- Calculation of **RMSE** to quantitatively compare landmark detection accuracy using various algorithm on one image.







Predictor	RMSE
Emotrics – 200px	7.986
Emotrics – 500px	4.364
Augmented Median	3.865

Conclusions

- Assessed the performance of a landmark predictor and applied test-time augmentation to improve accuracy.
- Using the NeuroFace dataset, preliminarily showed that the median of augmented points is more accurate.
- After acquiring more patient data, will train the model to provide FNP grading with the landmark points as input.

References

- [1] Miller, Matthew Q., *et al.* "The Auto-eFACE: Machine Learning-Enhanced Program Yields Automated Facial Palsy Assessment Tool." *Plastic and reconstructive surgery* vol. 147, 2 (2021): 467-474. doi:10.1097/PRS.0000000000007572
- [2] Parra-Dominguez, *et al.* "Facial Paralysis Detection on Images Using Key Point Analysis." *Appl. Sci.* 2021, 11, 2435. https://doi.org/10.3390/app11052435
- [3] Guarin, Diego L., *et al.* "Toward an Automatic System for Computer-Aided Assessment in Facial Palsy." *Facial plastic surgery & aesthetic medicine* vol. 22, 1 (2020): 42-49. doi:10.1089/fpsam.2019.29000.gua
- [4] Greene, Jacqueline J., *et al.* "The spectrum of facial palsy: The MEEI facial palsy photo and video standard set." *The Laryngoscope* vol. 130, 1 (2020): 32-37. doi:10.1002/lary.27986
- [5] Bandini, Andrea, *et al.* "A New Dataset for Facial Motion Analysis in Individuals With Neurological Disorders." *IEEE journal of biomedical and health informatics* vol. 25, 4 (2021): 1111-1119. doi:10.1109/JBHI.2020.3019242