CZECH TECHNICAL
UNIVERSITY
IN PRAGUE

FACULTY
OF ELECTRICAL
ENGINEERING

DIPLOMA THESIS

2023

LOUIS
KÄLBLE

Diploma Thesis Louis Kälble 2023

Czech Technical University in Prague

Faculty of Electrical Engineering

Department of Circuit Theory

Diploma thesis

Automatic Assessment of Facial Movement in Multiple Sclerosis Patients

Author: Louis Kälble

Supervisor: Ph.D. Ing. Michal Novotny

Study program: Medical Electronics and Bioinformatics

Prague 2023

ZADÁNÍ DIPLOMOVÉ PRÁCE

I. OSOBNÍ A STUDIJNÍ ÚDAJE

Příjmení: **Kälble** Jméno: **Louis** Osobní číslo: **505814**

Fakulta/ústav: **Fakulta elektrotechnická**

Zadávající katedra/ústav: **Katedra teorie obvodů**

Studijní program: **Medical Electronics and Bioinformatics**

Specializace: **Signal Processing**

II. ÚDAJE K DIPLOMOVÉ PRÁCI

Název diplomové práce:

Automatické hodnocení obličejových pohybů u pacientů trpících roztroušenou sklerózou

Název diplomové práce anglicky:

Automatic Assessment of Facial Movement in Multiple Sclerosis Patients

Pokyny pro vypracování:

Seznam doporučené literatury:

[1] Dobson, R., & Giovannoni, G. (2019). Multiple sclerosis–a review. European journal of neurology, 26(1), 27-40.
[2] Fukazawa, T., Moriwaka, F., Hamada, K., Hamada, T., & Tashiro, K. (1997). Facial palsy in multiple sclerosis. Journal of neurology, 244, 631-633.
[3] Ghosh, R., Roy, D., Dubey, S., Das, S., & Benito-León, J. (2022). Movement disorders in multiple sclerosis: an update. Tremor and Other Hyperkinetic Movements, 12.
[4] Mehanna, R., & Jankovic, J. (2013). Movement disorders in multiple sclerosis and other demyelinating diseases. Journal of the neurological sciences, 328(1-2), 1-8.
[5] Di Stadio, A., Dipietro, L., Ralli, M., Greco, A., Ricci, G., Messineo, D., & Bernitsas, E. (2020). Clinical and radiological findings of facial paralysis in multiple sclerosis. Multiple Sclerosis and Related Disorders, 37, 101456.

Jméno a pracoviště vedoucí(ho) diplomové práce:

Ing. Michal Novotný, Ph.D. katedra teorie obvodů FEL

Jméno a pracoviště druhé(ho) vedoucí(ho) nebo konzultanta(ky) diplomové práce:

Datum zadání diplomové práce: **16.02.2023** Termín odevzdání diplomové práce: **26.05.2023**

Platnost zadání diplomové práce: **22.09.2024**

Ing. Michal Novotný, Ph.D.	doc. Ing. Radoslav Bortel, Ph.D.	prof. Mgr. Petr Páta, Ph.D.
podpis vedoucí(ho) práce	podpis vedoucí(ho) ústavu/katedry	podpis děkana(ky)

III. PŘEVZETÍ ZADÁNÍ

Diplomant bere na vědomí, že je povinen vypracovat diplomovou práci samostatně, bez cizí pomoci, s výjimkou poskytnutých konzultací. Seznam použité literatury, jiných pramenů a jmen konzultantů je třeba uvést v diplomové práci.

. | Datum převzetí zadání

Podpis studenta

Declaration

I declare that the presented work was developed independently and that I have listed all sources of information used within it in accordance with the methodical instructions for observing the ethical principles in the preparation of university theses.

Prague, date ....

Author's signature

Acknowledgement

I would like to express my biggest appreciation and deepest gratitude to my thesis supervisor. Without his ongoing feedback and expertise, this endeavor would not have been possible. Additionally, big thanks also go to all university staff and study participants as well as all hospital staff involved in the collection of data.

Moreover, I cannot miss to mention the never-ending support from my girlfriend, keeping my motivation high and believing in me until the very end as well as always offering support. Lastly, I would like to also thank my best friend and family for all the moral support and long feedback-sessions and my dog for always happily accompanying me on life-saving walks in nature.

Abstract

Multiple Sclerosis (MS) is the most common non-traumatic, neurodegenerative disease affecting young adults worldwide with a total of 2.8 million people diagnosed. Main characteristics of the disease are the appearance of demyelinating plaque on neuronal structures. When demyelination occurs in the facial pathway, motor manifestations affecting the facial expression may arise which can include facial myokymia, hemifacial spasm or facial palsy. These all belong to the most common presentations of movement disorders in MS. Currently, there is no existing easy-to-use, objective and fully automatic tool enabling fast and reliable assessment of facial movement disruption for MS. The goal of this presented thesis is to develop a tool for the computerized, video-based assessment of facial disruption to facilitate an accurate, objective, easily applicable, and cost-effective method to evaluate facial movement in patients with MS.

Forty native Czech-speakers have been recorded during a facial expressivity examination and one-minute-long video-recordings of their freely spoken monologue were used in the subsequent analysis. To analyze the disruption of facial movement, an end-to-end neural network based facial landmark detection algorithm "Face Mesh" was applied and a total of six facial movement markers have been proposed to parametrize asymmetric movement of the face.

Significant differences were found in the face symmetry between MS patients with facial palsy and a healthy control group. Subsequently, a classification algorithm was trained using multinomial logistic regression that reached an AUC of 0.71.

The results of this work confirmed the utility of an automated objective tool for facial disruptions in MS and present the disruption of facial movement as a possible disease biomarker. Moreover, the classification experiment emphasized the need of regional assessment in the evaluation of facial manifestations in MS.

Keywords: Multiple Sclerosis, Facial Palsy, Facial Landmark Detection, Diagnostic Markers

List of abbreviations

AAM	Active Appearance Model
ADS	Arianna Disease Scale
AR	Augmented Reality
AUC	Area under the receiver operating curve
BDI	Beck Depression Inventory
BP	Bell's Palsy
cFP	Central Facial Palsy
CI	Confidence Interval
CIS	Clinically Isolated Symptom
CLM	Constrained Local Model
CNN	Convolutional Neural Networks
CNS	Central nervous system
EBV	Epstein-Barr-Virus
EDSS	Expanded Disability Status Scale
ENT	Ear, nose and throat
ES	Cohen's Effect Size
FM	Facial Myokymia
FNGS	Facial Nerve Grading Systems
FP	Facial Palsy
HBS	House-Brackmann-Scale
HC	Healthy Control
HS	Hemifacial Spasm
IQR	Interquartile Range
MRI	Magnetic Resonance Imaging
MS	Multiple Sclerosis
NEDA	No Evident Disease Activity
pFP	Peripheral Facial Palsy
PPMS	Primary Progressive Multiple Sclerosis
RIS	Radiologically Isolated Symptom
RRMS	Relapse Remitting Multiple Sclerosis
SD	Standard deviation
SPMS	Secondary Progressive Multiple Sclerosis

List of figures

Fig. 1 The Facial Pathway from the Periphery to the Brain Motor Cortex Area M1 taken from (Di Stadio and Bernitsas, 2018) 14

Fig. 2 Major categories of facial Landmark detection algorithms 18

Fig. 3 468 Facial Landmarks of Google's Mediapipe Face Mesh Algorithm with highlighted areas of interest 22

Fig. 4 Inference and Tracking Pipeline of Google's Mediapipe Face Mesh (Kartynnik *et al.*, 2019) 23

Fig. 5 Comparative figures for the mouth horizontal movement markers 25

Fig. 6 Between-group differences for each facial marker 30

Fig. 7 Receiver operating characteristic curve for the multinomial logistic regression using the mouth horizontal movement, eyebrow curvature and palpebral fissure marker 32

List of tables

Table 1 Demographical and clinical information about the MS patients 21

Table 2 Description of analyzed facial areas and depiction of their estimation 28

Table 3 Overview of Results 29

Table 4 Results for Spearman's correlation coefficient for each marker 31

Table 5 Results of diagnostic sensitivity analysis for facial palsy listing AUC, sensitivity, specificity and accuracy for single facial paralysis markers 32

Table 6 Multinomial logistic regression model coefficient statistics 33

List of supplementary material

Supplementary Material 1 Results of diagnostic sensitivity analysis showing AUC, sensitivity, specificity and accuracy for the four best predictor combinations 44

Supplementary Material 2 Close up view the eyebrow landmarks of the Face Mesh landmark detection algorithm 45

Supplementary Material 3 Close up view of the eye landmarks of the Face Mesh landmark detection algorithm 45

Supplementary Material 4 Close up view of the mouth landmarks of the Face Mesh landmark detection algorithm 45

Contents

1. Introduction – Multiple Sclerosis 10

1.1. Multiple Sclerosis - Etiology and Clinical Features 11

1.2. Treatment of MS 12

1.3. Facial Manifestations in MS 14

1.4. Clinical Assessment of Facial Palsy 16

1.5. Automatic Assessment of Facial Movement 17

1.6. Objectives of this thesis 19

2. Data and Methods 20

2.1. Subjects 20

2.2. Facial movement examination 20

2.3. Perceptual analysis 21

2.4. Computerized assessment of facial movement 21

2.4.1. Facial landmark detection 21

2.4.2. Estimation of makers for facial paralysis 24

2.5. Statistical Analysis 27

3. Results 29

3.1. Between-group differences 29

3.2. Comparison with perceptual assessment 31

3.3. Classification 31

4. Discussion 34

4.1. Objectives of this thesis – completion 37

5. Conclusion 38

References 39

Supplementary Material 44

1. Introduction – Multiple Sclerosis

Multiple Sclerosis (MS) is the most common non-traumatic, neurodegenerative disease affecting young adults worldwide (Calabresi, 2004; Dobson and Giovannoni, 2019) with a total of 2.8 million people diagnosed (MSIF, 2020) . While its prevalence is rising similarly in Africa, Australia, Asia and Europe, actual numbers vary widely around the globe with Europe showing approximately half of all MS cases (Evans *et al.*, 2013; Kingwell *et al.*, 2013; Makhani *et al.*, 2014). After a usually silent onset, the disease eventually leads to full disability of its sufferer, gradually accruing irreversible axonal damage. While many environmental and genetic factors have been described that increase one's individual risk of developing MS, its direct cause remains unknown. However, recent studies report a 32-fold risk increase of MS development after infection with Epstein-Barr-Virus (EBV), suggesting it to be a leading cause in the development of MS (Bjornevik *et al.*, 2022) . Smoking increases the risk by about 50% (Palacios *et al.*, 2011) and an increase in geographical latitude which is linked to decreasing UV-B exposure and lower vitamin D levels, correlates with increased MS prevalence in the population (Sintzel, Rametta and Reder, 2018; Dobson and Giovannoni, 2019). Furthermore, even though women seem to be twice as likely to be affected by MS than men, their prognosis is often better and the course of the disease less aggressive (Di Stadio and Bernitsas, 2018).

When symptomatic, MS can manifest as both motor and sensory disability. Most common symptoms include sensory or cerebellar ataxia, cognitive and progressive visual failure as well as spastic paraparesis. Additionally more than 50% of patients with MS suffer from depression (Calabresi, 2004). It has been estimated that the disease prodromal phase lasts between 5 to 10 years and is characterized by the presence of subtle nonspecific symptoms such as cognitive decline, psychiatric morbidity, fatigue, sleep disorders, pain, fibromyalgia, bowel, bladder and dermatological issues (Tremlett, Munger and Makhani, 2022). MS' subtle and nonspecific onset makes an early diagnosis especially challenging and hinders the success of upcoming promising immune reconstitution therapies. For people with MS these therapies offer a first significant chance to alter the disease course to the point of no progression or even a potential cure. Early diagnosis is however critical in order to reduce disease progression in these treatment options (Dobson and Giovannoni, 2019).

1.1. Multiple Sclerosis - Etiology and Clinical Features

By clinical definition, MS describes the phenomenon of plaque on structures of the nervous system, primarily on axons and their myelin sheath (Di Stadio *et al.*, 2020). The demyelinating plaque is a result of perivenular inflammatory lesions. Active lesions contain lymphocytes such as T-cells and B-cells that cause not only demyelination but also oligo dendrite damage (Dobson and Giovannoni, 2019). While axons often fully recover after remission in early stages of the disease, irreversible damage accumulates over the course of MS leading to dysfunction of the affected axon. Inflammation can appear wherever the immune cells are located within the nervous system but in most cases it primarily appears in the central nervous system (CNS) as opposed to the periphery (Di Stadio *et al.*, 2020). When inactive, lesions are surrounded by plaque and the inflammation core by activated macrophages and micro philia (Prineas *et al.*, 2001). Symptomatology of MS depends on the location of an active or inactive lesions. When a motor area is affected, the patient might experience stiffness or numbness at the innervated location and may develop a movement disorder. Analogously, with inflammation of sensory nervous tissue, the patient's sensorial perception will be affected (Di Stadio and Bernitsas, 2018). Over time, MS leads to disability, brain atrophy and end-organ damage. As clinically apparent symptoms differ with location and size of a lesion it is not possible to predict individual symptomatology and its severity. A post-mortem pathological evidence found 25% of MS cases that were never clinically diagnosed during their lifetime (Engell, 1989).

Clinical MS is divided into two stages; the inflammation dominant stage with profound active lesions called relapse remitting MS (RRMS) and the often-subsequent neurodegeneration dominant phase known as primary or secondary progressive MS (PPMS or SPMS). While traditionally described as separate diseases, research suggests that those phases exist on a continuous spectrum. A patient usually develops SPMS 10 to 15 years after being clinically diagnosed with RRMS (Dobson and Giovannoni, 2019). Only around 5-15% of people with MS are diagnosed with PPMS when the disease becomes clinically apparent, indicating that it most often gradually progresses from relapses to non-relapsing progression (Dobson and Giovannoni, 2019). Inflammatory rates are commonly highest in RRMS. As a new relapse develops, it reaches its peak within hours to days and stays highly symptomatic for any time period from weeks up to months. It then remits over time after which the patient recovers seemingly to full ability. However, each discrete relapse produces permanent neuronal damage even if in the form of microscopic lesions. It has been shown that up to 10 asymptomatic lesions can appear during one relapse and silently promote the disease evolution (Dobson and Giovannoni, 2019).

Moreover, pathological studies show that the preclinical phase of MS could be between 5 to 10 years (Tremlett, Munger and Makhani, 2022); neuroinflammation and neurodegeneration

usually occur before the first clinical attack while end-organ damage is proven to progress from the onset of MS. Because of this, inactive lesions may already arise in the prodromal phase likely due to asymptomatic relapses. Patients found with inactive lesions will be diagnosed with clinically isolated symptom (CIS) which describes a first clinical appearance of MS usually supported with explanatory lesions on an MRI scan. Further prodromal diagnoses may include asymptomatic MS or radiologically isolated symptom (RIS). RIS is diagnosed upon isolated findings of white matter lesions that are typically discovered on unrelated MRIs without ever having been clinically apparent. In young people with CIS, brain volume loss has been shown (Aubert-Broche *et al.*, 2014) and in 25% of people diagnosed with RIS significant cognitive impairment has been found similar to patients with established MS diagnoses (Amato *et al.*, 2012). This strongly highlights the importance of enabling pre-symptomatic diagnosis. Tremlett et al. (2022) argues that biomarkers may be the key to detecting MS in the prodromal phase arguing that they may increase specificity and accuracy of the diagnosis. However, this is in need of further formal validation (Tremlett, Munger and Makhani, 2022). In addition, the onset of clinical MS may be sudden, and the naturally resolving nature of relapses may increase difficulty for the affected patient to first understand that the underlying cause is a neurological disease.

As of now, the clinical diagnosis of MS requires dissemination of MRI lesions in time and space over a period of three months, without any other causal explanation. This exclusion of asymptomatic and even prodromal MS is being criticized since it withholds crucial treatment that might prevent the clinical appearance overall. Patients diagnosed with CIS or RIS are ineligible for MS treatment (Dobson and Giovannoni, 2019).

1.2. Treatment of MS

Treatment of MS can be divided into two categories: disease-modifying therapies and symptomatic therapies. Symptomatic therapies are commonly non-MS specific treatments that are used in any neurological disorder. They can include pharmaceutical and physical therapies that target specific symptoms e.g., urinal bladder disfunction or even certain motor disorders. Disease-modifying therapies on the other hand are the closest opportunity for people with MS to a long-term remission of the disease. These treatments are MS-specific and are now focusing on early treatment admission after years of research progression (Dobson and Giovannoni, 2019). The three existing types of therapy, immunosuppressant, immunomodulatory and immune reconstitution therapy, differ in treatment duration and effect. Previously, immunosuppressant and -modulatory treatments have been used to prevent further inflammation and MS progression, immune reconstitution therapy however brings the added benefit of having a finite treatment length while allowing long term results. These are said to be the closest to a potential cure of MS to date (Dobson and Giovannoni, 2019). Dobson et al even raises the question if this is the proof

that early or pre-symptomatic treatment could be preventing the clinical appearance of the disease overall (Dobson and Giovannoni, 2019). Subsequently to improve categorization of nonclinical MS, the concept of no evident disease activity (NEDA) was introduced. Five levels of NEDA describe disease activity ranging from absence of relapses and clinical disease progression (NEDA-1) over inflammatory MRI activity to MRI atrophy and levels of biomarkers (NEDA-5).

The success of early treatment in MS and the possibility of precise identification of the at-risk population strongly suggests that a pre-symptomatic screening of said population could further improve treatment and significantly reduce clinically apparent MS overall and even set the stage to start population screening in other neurodegenerative diseases (Dobson and Giovannoni, 2019). Until then, research stresses early treatment which inevitably follows early, fast, and definite detection of signs and symptoms of the disease at any stage, be it asymptomatic, prodromal, or clinical. To attain to this target, this thesis handles a specific subset of symptoms, facial manifestations, and their reliable detection.

1.3. Facial Manifestations in MS

Any damage to the facial pathway in MS may lead to motor manifestations that affect the facial expression. The facial pathway describes the 7th cranial nerve, the facial nerve, that leads from its terminal branch on one side of the face through the pons up to the contralateral motor cortex area M1 (see Fig. 1) (Campbell and DeJong, 2005). Movement disorders that develop upon impairment of the facial pathway may be (continuous) facial myokymia (FM), hemifacial spasm (HS) or facial palsy (FP). FM, HS and FP belong to the most common presentations of movement disorders in MS (Ghosh *et al.*, 2022). Facial myokymia is defined as involuntary wave-like movements across all facial muscles, often related to lesions in the brainstem. Hemifacial spasm presents itself as synchronous spasms on one side of the face for which lesions attributed to the ipsilateral facial nucleus are the cause (Mehanna and Jankovic, 2013; Etemadifar *et al.*, 2022). And lastly, facial palsy, the most common presenting facial disorder of the group, describes the inability of a patient to voluntarily move one side of their face (Etemadifar *et al.*, 2022).

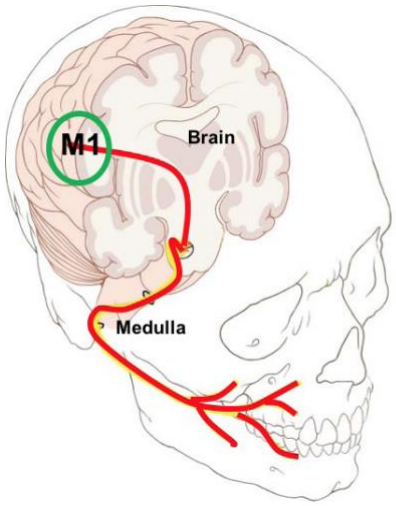


Fig. 1 The Facial Pathway from the Periphery to the Brain Motor Cortex Area M1 taken from (Di Stadio and Bernitsas, 2018)

The actual incidence of facial manifestations in literature is ambiguous. According to Ghosh et al., FM is frequently reported as a presenting feature of MS whereas in another recent study of 2260 patients, FP had the highest prevalence at 3.27% followed by FM with 1.28% and HS with 0.84% (Etemadifar *et al.*, 2022; Ghosh *et al.*, 2022). Moreover, the prevalence of FP alone varies throughout research between 3% and 20% (Fukazawa *et al.*, 1997; Danesh-Sani *et al.*, 2013; Lassemi *et al.*, 2014; Di Stadio and Bernitsas, 2018) which suggests difficulties and methodological differences in the diagnosis of facial manifestations. Contributing to the issue is that every author has their own way of documenting symptoms, and it is difficult to ascertain if

the symptom was not only verbally reported by the patient to the author. Diagnosis is often carried out by trained neurologists or ear, nose and throat (ENT) specialists using classical neurological examinations and grading the results based on a set scale like the House-Brackmann scale (HBS) (Fukazawa *et al.*, 1997; Di Stadio, 2015; Etemadifar *et al.*, 2022).

Contributing to the discussion of FP diagnosis is its clinical separation into two types: central facial palsy (cFP) or peripheral facial palsy (pFP). cFP denotes the paresis or reduced motor function of facial muscles of only the lower two thirds of the face while pFP is defined to affect all facial muscles on one side (Di Stadio, 2015). This distinction originates in the bilateral innervation of the frontalis muscle. When a central lesion in one hemisphere is present, the frontalis muscle will be unaffected due to its secondary innervation by the contralateral hemisphere, thus classifying this type as central FP.

In a group of 107 patients diagnosed with MS, 4.7% developed FP as an early onset symptom (Fukazawa *et al.*, 1997). Another study of a subset of MS patients that were clinically evaluated to show FP, found about 25.8% developed FP as an early onset symptom (Di Stadio *et al.*, 2020). The general rate of FP manifestation during MS is more variable ranging from 2.6% (Kurtzke *et al.*, 1977) to 31% (Fukazawa *et al.*, 1992). On top of that, there has been evidence suggesting that facial nerve damage is much more profound than actually reported; an autopsy series found facial nerve damage in over 50% of the examined cases (Fukazawa *et al.*, 1997). Di Stadio et al points out the difficulties in correctly diagnosing the origin of FP and warns of the dangers of misdiagnosing the underlying cause as Bell's Palsy (BP) instead of FP. In the mentioned study 21.7% of a group that presented with FP were initially misdiagnosed as Bell's Palsy or stroke (Di Stadio *et al.*, 2020). BP is defined by damage of the peripheral facial nerve (thus is symptomatic as pFP) and is often treated with corticosteroids after examination without further investigation, possibly missing MS as an underlying cause (Saleh *et al.*, 2016; Di Stadio *et al.*, 2020). To summarize, the prevalence of FP in MS reportedly has a wide variability, suggesting possible difficulties in the diagnosis of FP itself as well as in the correct assignment of the underlying cause of FP as MS. This may be due to the nature of the clinical assessment used to evaluate a patient that shows signs of FP. The next subchapter will further investigate methods used to evaluate FP.

1.4. Clinical Assessment of Facial Palsy

The assessment of facial manifestations is traditionally carried out by experts of various fields by visually evaluating the patient's face in static, dynamic and synkinetic state. To allow comparison, experts adhere to facial nerve grading systems (FNGS) when assessing severity. The most used FNGS internationally remains the House-Brackmann scale that was presented in 1985 likely because of its endorsement by the Facial Nerve Disorders Committee (House and Brackmann, 1985). Over the years, there have been numerous proposals for further FNGS and even algorithms for a computerized comparison of different FNGS (Bansal *et al.*, 2020). However, the simplicity of the HBS leads to its ongoing preference (Di Stadio, 2015). Criticism of the HBS addresses the importance it is putting on the ocular area and lack of importance of other facial muscles even when considering the revised version HBS-2 which was designed to be more regional (Vrabec *et al.*, 2009). Di Stadio et al. proposed the use of the Arianna Disease Scale (ADS) which parts the face in three sub areas where further grouping focuses on the evaluation of smaller muscle groups that are all weighted equally. Other proposed scales over the years include the Sunnybrook and Yanagihara Scale (Ross, Fradet and Nedzelski, 1996; Yanagihara and Hato, 2003) that are similar to the HBS but require more time for the assessment. The frequent introduction of new grading scales itself argues for an underlying issue with the diagnosis (Bansal *et al.*, 2020). Moreover, the proposed systems do not manage to reduce the clinicians' subjectivity which can pose as the biggest problem especially when used by clinicians who do not frequently work with the applied scales (Lee *et al.*, 2013). Already in 1983, House observed that the subjective interpretation in clinician-based FGNS should be reduced as much as possible (House, 1983).

A recent proposal aims to eliminate the subjective part of the evaluation by developing a simple, objective and mathematical model for the quantification of the severity of FP (Bansal *et al.*, 2020). It requires physicians to physically take measurements of defined distances in the face of the patient in various static states. While this approach may reduce subjectivity in evaluation, reasonable doubt can be expressed about the accuracy and reproducibility of measurements taken by clinicians with a Vernier Caliper and about the safety of patients since the measurements are in close vicinity to extremely sensitive structures e.g., the eyes.

The choice of an automated, computerized method may be the most feasible option to reduce the subjective component in assessing a patient's face whilst achieving reproducible and fast results for an evaluation. The next subchapter presents options for implementing automated facial evaluation methods.

1.5. Automatic Assessment of Facial Movement

Automatically assessing and evaluating facial motor dysfunctions requires the capturing of facial movement over an extended period of time. One possible implementation of this is the application of automatic facial landmark detection algorithms on video-recordings of freely spoken monologues.

Facial landmark detection describes the localization of key points (landmarks) around facial components and facial contours that can further be used for facial analysis tasks. Algorithms that perform this task commonly use RGB or grayscale images and return a vector of 2D or 3D coordinates of the landmarks. In these algorithms, face detection is often assumed to have already been carried out. Facial landmark detection has been one of the heaviest researched topics in computer vision over the last decade (Bulat and Tzimiropoulos, 2017). This is likely due to the opportunities its application provides in a range of different fields such as security, medical applications and the field of augmented reality applications (Wu and Ji, 2019). The localization of fiducial facial landmark points however is not a trivial task especially as many applications rely on the accurate detection of the landmarks. Challenges in the automated detection may arise for several reasons such as changing environmental factors like illumination, differing facial expressions leading to a change of facial appearance and lastly possible facial occlusion by third objects or even self-occlusion as a result of extreme head poses that may lead to incomplete facial appearance information (Wu and Ji, 2019). Several different algorithms have been developed over the past decades to best tackle the task of landmark detection. According to Wu and Ji (2019) the developed algorithms can be classified into three categories, the holistic methods, the Constrained Local Model (CLM) methods or the regression-based methods (see Fig. 2) (Wu and Ji, 2019). Between them, they differ mainly on how they model facial appearance and facial shape patterns. Holistic methods describe algorithms that explicitly model the holistic facial appearance and global facial shape patterns while CLMs rely on explicit local facial appearance and explicit global facial shape patterns. Whilst these two methods share modelling of the global facial shape, the CLMs have an advantage in using the local appearance around landmarks instead of a holistic one, making it more robust towards illumination changes and facial occlusion (Wu and Ji, 2019). Both methods can be further subdivided into different approaches to their general solutions such as the Active Appearance Model (AAM) as the classic holistic model (Cootes, Edwards and Taylor, 1998) or local appearance models as an example for CLMs (Cristinacce and Cootes, 2006). Regression-based methods differ in the sense that they directly learn the mapping from image appearance to the landmark locations i.e., they do not build a global face shape model (Wu and Ji, 2019). Within the regression-based methods direct regression, cascaded regression or deep-learning based regression methods can be further classified. Altogether, regression-based

models are viewed as the most promising in successfully accomplishing the task of facial landmark detection. They directly predict the landmarks as opposed to predicting model coefficients as in the holistic methods and CLMs which can turn a small error in the model coefficients into a large landmark error (Wu and Ji, 2019).

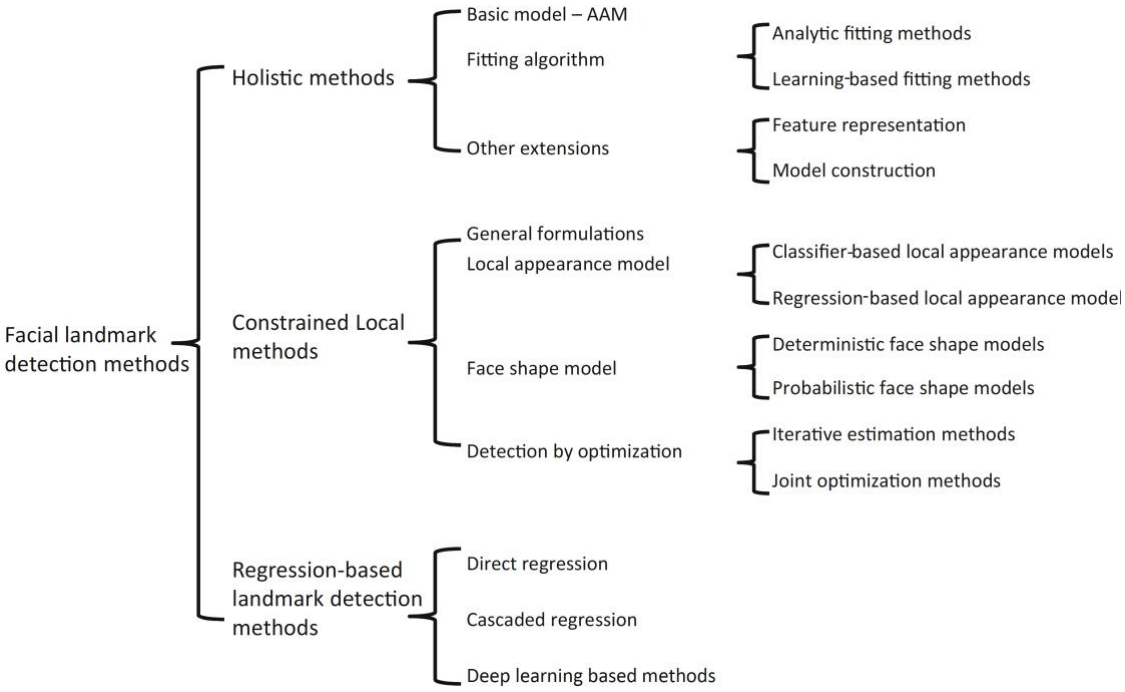


Fig. 2 Major categories of facial Landmark detection algorithms
Taken from (Wu and Ji, 2019)

Over the past years, deep learning based regression methods have shown incomparable advances in accuracy and efficiency when applied for the localization task (Bulat and Tzimiropoulos, 2017). Especially the use of Convolutional Neural Networks (CNN) has become a popular choice for deep learning models and most of them follow the direct regression or cascaded regression framework (Wu and Ji, 2019). One of the first developed algorithms using CNNs has been presented by Sun et al. (2013) which uses four convolutional layers to predict five facial landmarks in a cascaded manner (Sun, Wang and Tang, 2013). Their proposed solution has been improved over years in performance and in cascaded procedure to increase the number of predicted landmarks e.g., from 5 landmarks to 68 (Zhou *et al.*, 2013).

The application of these methods has been previously done in the field of neuropsychiatric and -degenerative disorders for example with the aim to draw conclusions about facial expressions, which are then used in multimodal fusion for emotion recognition as in Schoneveld et al (2021).

Here, audio and visual deep convolutional neural networks were applied to extract voice and facial features (Schoneveld, Othmani and Abdelkawy, 2021). Another similar application uses deep convolutional neural networks that extract audiovisual features to develop objective automatic depression estimation systems (He *et al.*, 2022). In the field of neurodegenerative pathologies, Polet et al (2022) used an eye-tracker to evaluate the eye gaze of patients with Alzheimer disease and Parkinson disease to determine whether facial emotion recognition impairment is linked with inappropriate eye-gaze strategies (Polet *et al.*, 2022). In the field of Parkinson disease, Novotny et al (2022) employed a convolutional neural network that localizes 68 facial landmarks (Gross *et al.*, 2010) in each frame and used a selected combination of the landmarks to describe the dynamics of different facial regions for an automatic assessment of hypomimia (Novotny *et al.*, 2022). However, no method for computerized assessment of facial motor impairments has been previously applied in the field of MS at the time of this paper.

1.6. Objectives of this thesis

With the goal of eliminating subjective errors in the clinical assessment of facial movement disorders, this paper aims to develop a tool for their computerized assessment to facilitate an objective, simple, economical, and reproducible method to evaluate facial movement without bias. The detailed objectives of this thesis are:

(1) The application of a deep learning based landmark detection algorithm to extract key points of facial expressivity of the subjects for further analysis
(2) The analysis of disruption of facial movement in patients with multiple sclerosis
(3) The proposition of an automatic approach for facial movement parametrization in patients with multiple sclerosis that exhibit signs of facial paralysis

2. Data and Methods

2.1. Subjects

Between 2016 and 2017, a total of 40 native Czech-speakers have been subjected to the facial expressivity examination for this study including 20 people (10 women) diagnosed with MS and 20 healthy controls (HC) (10 women). Video data of the MS participants were collected within the scope of an original study (Rusz *et al.*, 2018) investigating the relationship between the severity of speech disorders and neurological involvement in MS. The MS patients that were clinically confirmed with a diagnosis of definite MS based on the revised McDonald Criteria 2010 (Polman *et al.*, 2011). Within the MS group, 15 patients were diagnosed with relapse-remitting MS, 3 with secondary progressive MS, 1 with primary progressive MS and 1 with clinically isolated symptom. All MS patients have not had a relapse for at least 30 days prior to the data collection. Additionally, each patient has been ranked by a board certified neurologist according to the Expanded Disability Status Scale (EDSS) (Kurtzke, 1983) and received scores evaluating their severity of depression based on the Beck Depression Inventory (BDI) (Beck, Steer and Brown, 1987).

The HC group consisted of 20 age- and sex-matched individuals with no previous history of neurological or facial disorders which would interfere with the analysis of facial movements.

Each participant provided written and informed consent for their inclusion. The study received approval from the ethics committee of General University Hospital in Prague, Czechia, and has been performed in accordance with the ethical principles laid down by the Declaration of Helsinki.

2.2. Facial movement examination

The facial movement examination was conducted in a room with standard indoor lighting using a digital camera (Panasonic Handycam HDR-CX410, Osaka, Japan) that was fixed in position with regards to artificial and natural sources of light. It was set up at a distance of approximately one meter to the subjects' faces. The recordings were captured at a resolution of 1440 x 1080 pixels (HD) and a frame rate of 25 RGB frames (24-bits) per second. The recordings contain one minute of freely spoken dialogue from each patient on a topic given by a speech specialist conducting the examination.

2.3. Perceptual analysis

A speech-language specialist trained in facial expressivity examination provided an assessment of facial paralysis from the recordings. Each patient was assigned a score based on a five-point scale with 0 representing normal facial movement, 1 representing slight paralysis, 2 mild, 3 moderate and 4 severe paralysis. A summary of the scores is shown in Table 1.

Table 1 Demographical and clinical information about the MS patients

Multiple sclerosis participants (n = 20, 10 women)	Mean(Standard deviation, range)
Age (years)	49.6 (8.59, 33-70)
Disease duration (years)	14.75 (8.43, 2-31)
EDSS score	1.85 (0.37, 1-2)
BDI score	10.5 (6.75, 1-27)
Perceptual analysis score (0-5)	1.8 (0.62, 1-3)
Healthy control group (n = 20, 10 women)	
Age (years)	48.8 (8.18, 40-77)
Perceptual analysis score (0-5)	0 (0, 0)

EDSS – Expanded Disability Status Scale

BDI – Beck Depression Inventory

2.4. Computerized assessment of facial movement

2.4.1. Facial landmark detection

To detect the facial landmarks, the end-to-end neural network based facial landmark detection algorithm "Face Mesh" described in Kartynnik et al (2019) was applied. It creates 3D facial landmark estimations forming a mesh of 468 points (see Fig. 3) with only single RGB video frames as its input. The algorithm worked with a machine learning pipeline of two real-time, deep learning neural networks: it first applied a face detector that creates a rectangle around the detected face(s) as well as a number of basic landmarks like the center of the eyes or the nose tip. The original frame was then cropped and fed into the second 3D mesh prediction neural network that predicted the approximate 3D surface via regression and outputs a vector of the desired 3D landmark coordinates (see Fig. 4) (Kartynnik *et al.*, 2019). Along with the 3D landmark estimations, the mesh prediction network returned the probabilities of a face being contained

within the cropped frame and reasonably aligned. The Face Mesh algorithm has been primarily developed for real-time augmented-reality (AR) applications and has been pre-trained on 30K in-the-wild mobile camera photos taken from a wide range of sensors and under different lighting conditions. Considering the standardized recording conditions during the clinical assessment the 30K in-the-wild provided sufficient sensitivity and robustness for the subsequent feature extraction.(Kartynnik *et al.*, 2019).

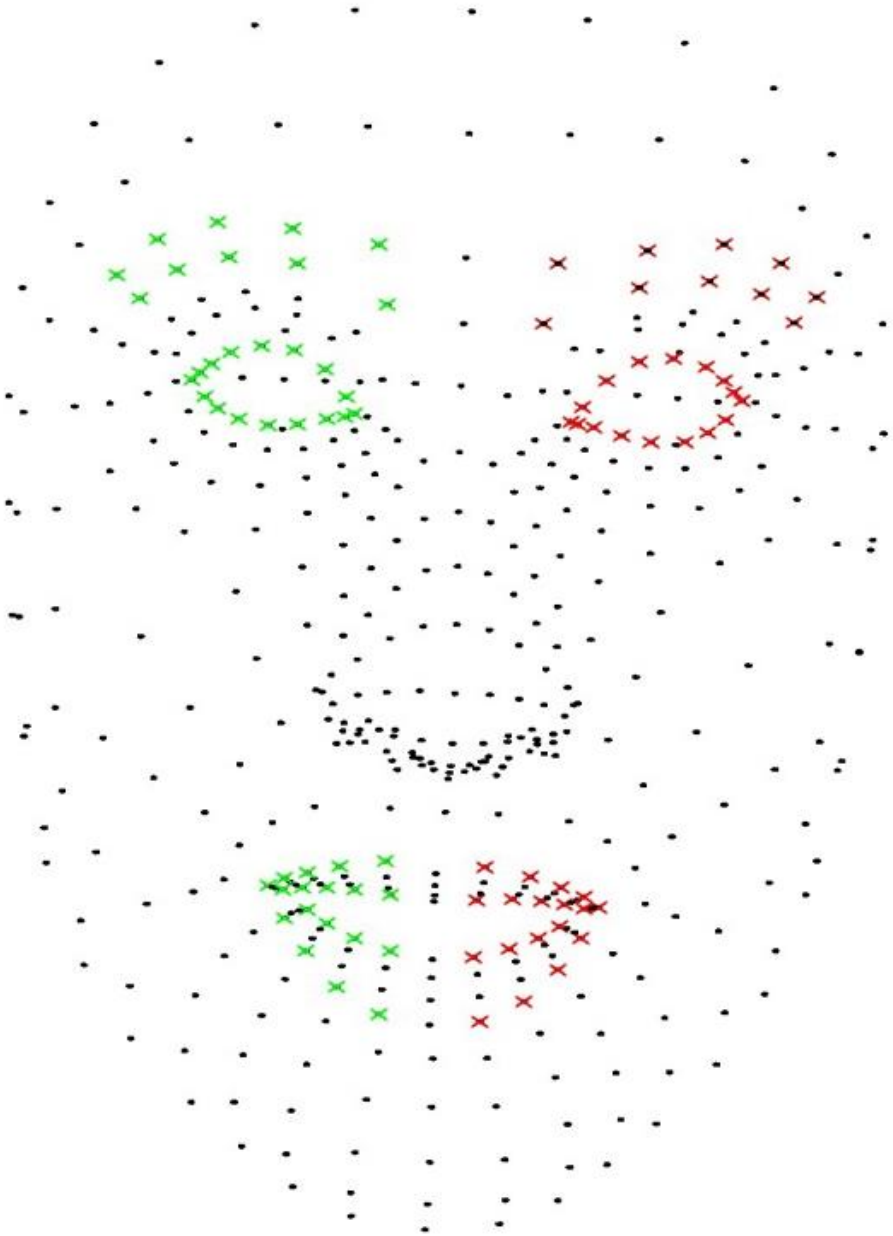


Fig. 3 468 Facial Landmarks of Google's Mediapipe Face Mesh Algorithm with highlighted areas of interest

In addition to the 3D face landmark detection models, another so-called Attention Mesh model was available that further refined landmarks of higher interest in areas such as the lips and eyes and adds further ten landmarks describing both irises, bringing the total number of estimated landmarks to 478. At a cost of higher computing, a more precise measurement around these areas can be achieved.

The landmark detection algorithm was executed in the PyCharm Professional environment with Python 3.8 (JetBrains s.r.o.). The described Face Mesh solution provides the option to define a maximum number of faces to track in an image, *max_num_faces*, which has been set to one. The *refine_landmarks* option has been activated, leading to 478 more robust landmarks that are refined around the mouth- and eye-areas. The script was modified to create a mat-file for each frame containing all 478 landmarks as 3D coordinates, normalized to the image size, that would subsequently be used for the estimation of facial paralysis markers.

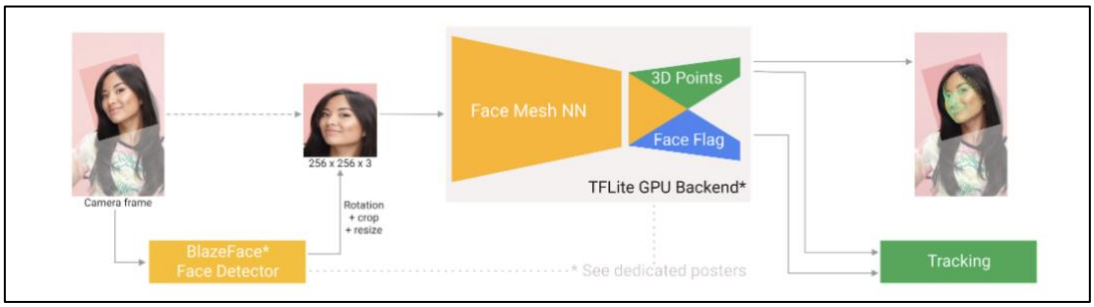


Fig. 4 Inference and Tracking Pipeline of Google's Mediapipe Face Mesh (Kartynnik *et al.*, 2019)

2.4.2. Estimation of makers for facial paralysis

The marker assessment was performed in the MATLAB environment (Mathworks, Natick, MA, USA). Using the detected landmarks, differences between the left and right variants of the following six facial measures were estimated:

(i) Eyebrow elevation / depression
(ii) Eyebrow curvature angle
(iii) Palpebral fissure dynamic behavior
(iv) Mouth corner elevation / depression
(v) Mouth corner horizontal movement
(vi) Mouth asymmetry

To assess differences between left and right variants in the case of all markers aside from (ii) and (vi), the areas under the curves describing the distances over all frames have been calculated and subsequently subtracted from each other. The result was then normalized with regards to the total number of frames. This measure has been chosen after the distance measures revealed high movement artefacts that introduce significant bias into any measure of deviation, overshadowing small differences between each side. To illustrate this, an additional video of the author of this paper was fed into the landmark detection algorithm where the head was kept still and oriented frontal to the camera. Fig. 5 shows the comparison between the reference video and one patient's measurement of the mouth horizontal marker.

By subtracting the total areas, the movement artefacts are reduced while differences between each side remain. In case of the mouth asymmetry marker, the median has been selected in order to describe median total mouth asymmetry for one subject. Lastly, for the eyebrow curvature marker, the difference of the deviation measure of each side was calculated to assess their differences. All markers are described in greater detail in the following paragraphs.

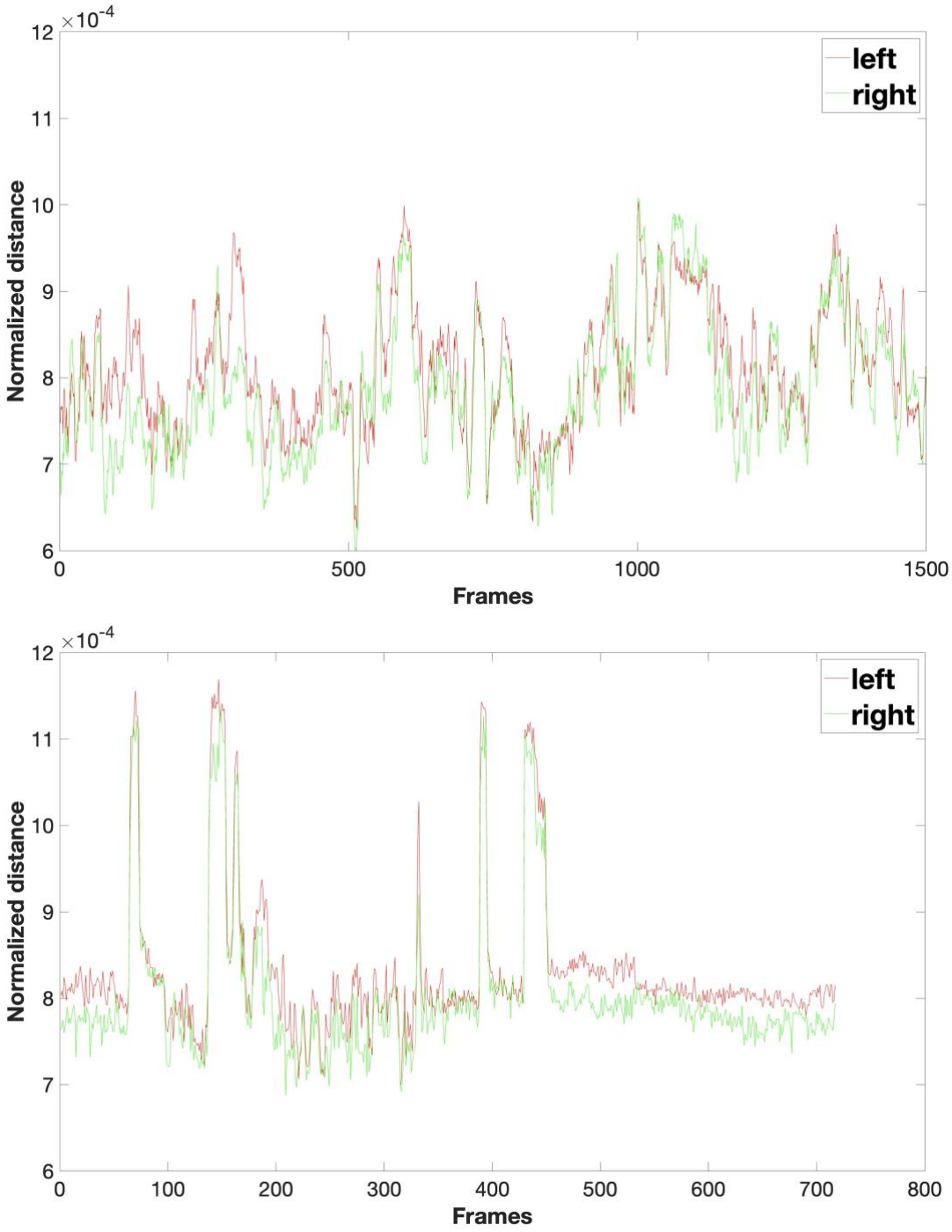


Fig. 5 Comparative figures for the mouth horizontal movement markers

The upper figure shows the mouth horizontal movement marker for one subject in one recording.

The lower figure shows the mouth horizontal movement marker from a monologue recorded from the author where the head was kept frontal to the camera and not moved. Horizontal movements of the lips that denote a smile are clearly visible.

The *eyebrow elevation/depression* marker was estimated as a measure of the normalized distance between the nose tip and a centroid created from all ten eyebrow landmarks per side. Dissimilarity between sides was calculated as the difference between areas under their distance curves. This marker describes differing dynamic movement of the eyebrows where a higher value indicates greater digression from symmetric movement.

The *eyebrow curvature* marker describes the differences in dynamic change of the angle in the middle of the eyebrow. To estimate the angle two lines leading up to the middle of each brow had to be fitted to the landmarks. Single value decomposition was used to fit the lines and determine their directional vector as the smallest singular value. The angle of the eyebrow could then be determined from the dot-product of the resulting two directional vectors per eyebrow. Following, the absolute difference of the median absolute deviation (MAD) was calculated between both sides. This marker describes the dynamic changes in eyebrow shape and shows greater difference in symmetry at a higher value.

To describe the *palpebral fissure's difference in dynamic behavior*, the area encompassed by the eyelids of a subject was determined by triangulating the area that the sixteen eye landmarks encompass and subsequently calculating each triangles' area. The triangulation was performed using the function *delaunayTriangulation()*. The determined area was then normalized by the square of all summed distances to the nose tip. To determine the difference between sides, the areas under the curves were calculated, subtracted and normalized.

The *mouth corner elevation/depression* marker was calculated as the difference of area under the curves of the normalized distance between a mouth corner centroid and the nose tip. The mouth corner centroid is composed of four landmarks per side (see Table 2). Analogously, the *mouth corner horizontal movement* marker was determined using the normalized distance to a centroid in the middle of the mouth, created from eight landmarks.

Lastly, the *mouth asymmetry* marker was calculated by mirroring the left side of the mouth onto the right side of the mouth using a sagittal plane fitted through the middle mouth landmarks as the reflector (see Table 2). The plane was created using eigenvalue decomposition of the landmarks in order to find the normal vector to the plane and construct the plane equation. Next, normalized distances between each corresponding mouth landmark were calculated and summed to create one asymmetry metric per frame. For the total asymmetry, the median of all values was calculated.

Prior to the decision of the descriptive measure of the markers, all measurement distributions i.e., the distribution of measurements over all frames per marker, were tested for normality using the Kolgomorov-Smirnov test. The result showed a non-normal distribution for all markers. Close inspection revealed that the majority of distributions exhibit negative kurtosis and therefore, non-

parametric measures such as the median have been chosen for the estimation of the expected value or the median absolute deviation for the measure of deviation.

A more detailed illustration of the landmarks including their indices from all areas of interest (mouth, eyes and eyebrows) can be found in Supplementary Material 2-4.

2.5. Statistical Analysis

Because the one-sample Kolmogorov-Smirnov test revealed non-normal distributions of the proposed markers, between group differences were assessed using the Kruskal-Wallis test with the possible presence of outliers. Furthermore, the strength of between group differences was calculated using Cohen's effect size (ES). The level of significance for testing between group differences was set to $\alpha = 0.05$.

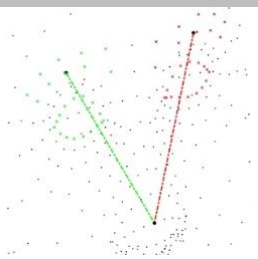
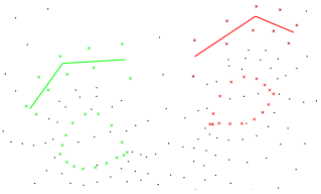
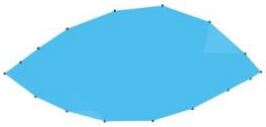
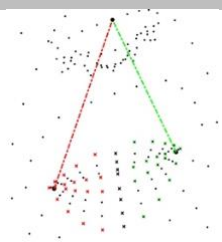
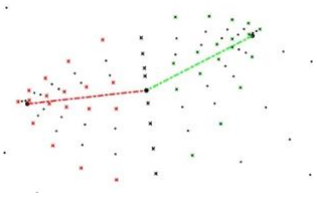
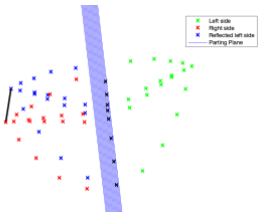
As a test for the relationship between perceptual analysis and the automated video-based analysis, spearman's correlation coefficient was used for the merged MS- and HC-group.

To estimate the ability to distinguish between MS patients and healthy control using the calculated markers, multinomial logistic regression was applied in combination with leave-one-out cross-validation. In order to find the optimal combination of markers, multiple combinations have been considered with the aim to maximize the area under the receiver operating characteristic curve (AUC) using the grid-search approach. The regression model was then fitted using the determined optimal markers and the grouping categories "MS" or "HC" were chosen as the categorical outcomes for the algorithm with "MS" as the reference category. The algorithm fits the predictors to the following equation, that determines the log-odds of a sample belonging to the secondary category as opposed to the reference category.

$$ln\left(\frac{\pi_{HC}}{\pi_{MS}}\right) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 \beta X_3 \quad (1)$$

The diagnostic accuracy of the classifier was then obtained from the receiver operating characteristic curve and reported.

Table 2 Description of analyzed facial areas and depiction of their estimation

Marker	Definition	Illustration
Eyebrows		
Eyebrow elevation / depression	The normalized difference between the area under the curve of left and right variants of the normalized distance of one eyebrow centroid to the nose tip	
Eyebrow Curvature	The absolute difference between left and right median absolute deviations of the eyebrow curvature angle between the directional vector of two lines fitted through half of each eyebrow	
Eyes		
Palpebral fissure dynamic behavior	The normalized difference between the area under the curve of left and right variants of the total area encompassed by the eyelids	
Mouth		
Mouth corner elevation / depression	The normalized difference between the area under the curve of left and right variants of the normalized distances between a centroid of the mouth corners taken from four landmarks and the nose tip	
Mouth corner horizontal movement	The normalized difference between the area under the curve of left and right variants of the normalized distances between a centroid of the mouth corner and a centroid representing the middle of the mouth in each frame for each side	
Mouth asymmetry	The median of the summed normalized distances between corresponding mouth landmarks from both sides when one side is mirrored onto the other side by the vertical mouth parting-plane (for simplicity, only one distance is marked)	

3. Results

3.1. Between-group differences

All results for each marker per group including their median and interquartile range (IQR) as well as effect size are summarized in Table 3. There was a statistically significant relationship for the mouth horizontal movement marker between the MS- and HC-group ($\chi^2 = 5.41, p = 0.02$) with a median of 5.01e-5 for the MS-group and 2.37e-5 for the HC. The remaining markers did not reach the level of statistical significance, yet the mouth asymmetry marker ($\chi^2 = 3.28, p = 0.0699$) and the mouth elevation depression marker ($\chi^2 = 3.09, p = 0.0787$) suggest a possible trend, both showing also higher medians for the MS-group than for the HC. Moreover, the horizontal mouth marker reached the largest effect size ($ES = 0.818, CI = [0.179\ 1.45]$), followed by the mouth asymmetry marker reaching medium effect size ($ES = 0.5046, CI = [-0.117\ 1.12]$) and the mouth elevation/depression marker reaching small effect size ($ES = 0.461, CI = [-0.158\ 1.09]$). A visual illustration of the marker comparison is displayed in Fig. 6.

Table 3 Overview of Results

#	Feature	HC		MS		Effect
		Median	IQR	Median	IQR	Size
Eyebrows						
(i)	Elevation / Depression	0.971e-4	1.78e-4	1.34e-4	1.33e-4	0.00125
(ii)	Curvature	0.182	0.275	0.2321	0.364	0.234
Eyes						
(iii)	Palpebral Fissure	0.922e-7	1.30e-7	1.26e-7	1.14e-7	0.366
Mouth						
(iv)	Elevation / Depression	3.684e-5	6.15e-5	7.51e-5	8.49e-5	0.461
(v)	Horizontal	2.366e-5	2.12e-5	5.01e-5	6.48e-5	0.818*
(vi)	Asymmetry	1.98e-3	0.0018	2.43e-3	0.0011	0.505

Asterisks denote significance of values: *$p < 0.05$

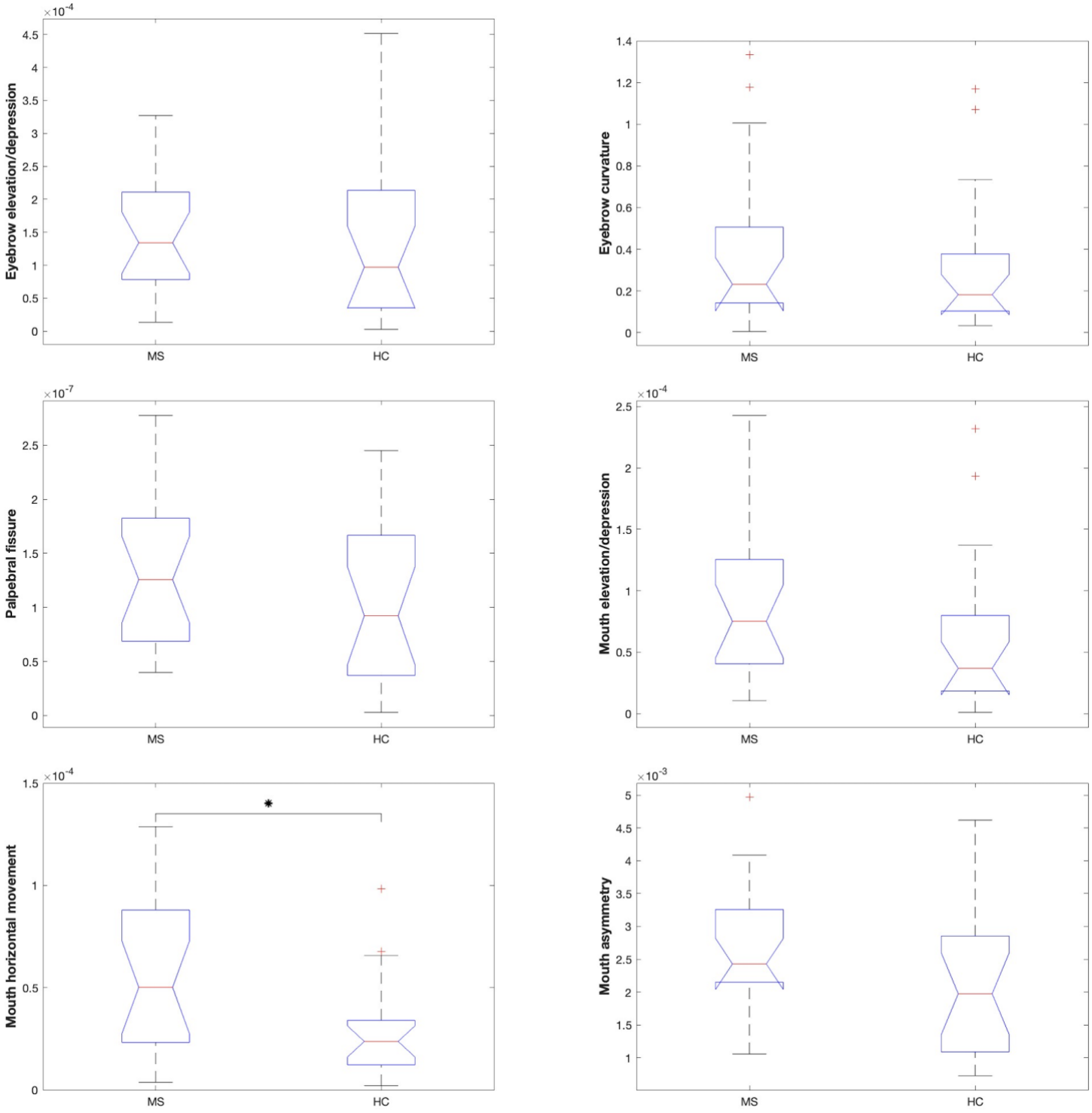


Fig. 6 Between-group differences for each facial marker

The line in the boxplots represents the median and the boxes denote the 25th and 75th percentiles. Vertical lines denote the standard deviation. The Kruskal-Wallis test has been used to test for group differences: *p < 0.05. MS – Multiple Sclerosis group, HC – Healthy Control group

3.2. Comparison with perceptual assessment

Spearman's correlation coefficient revealed a significant positive correlation in the case of the mouth horizontal movement marker ($r = 0.39, p = 0.014$). Other facial markers also showed a positive correlation coefficient, but not significant. Table 4 shows the summary of the correlation analysis.

Table 4 Results for Spearman's correlation coefficient for each marker

#	Feature	Spearman's correlation	
		r	**p**
Eyebrows			
(i)	Elevation / Depression	0.0935	0.566
(ii)	Curvature	0.0565	0.729
Eyes			
(iii)	Palpebral Fissure	0.231	0.151
Mouth			
(iv)	Elevation / Depression	0.181	0.263
(v)	Horizontal	0.387	0.0137
(vi)	Asymmetry	0.245	0.128

Asterisks denote significance of values: $*p < 0.05$

3.3. Classification

In the classification experiment, the best combination of markers to fit a multinomial logistic regression was determined using grid-search, followed up by leave-one-out cross-validation with AUC as the measure of performance. When considered separately, the mouth horizontal movement marker performed best ($AUC = 0.64$), followed by the mouth asymmetry marker ($AUC = 0.55$), the mouth elevation/depression marker ($AUC = 0.53$), the palpebral fissure marker ($AUC = 0.49$), the eyebrow curvature marker ($AUC = 0.31$) and lastly the eyebrow nose-distance marker ($AUC \approx 0$) (see Table 5). The most performant combination of markers consisted of the mouth horizontal movement marker, the palpebral fissure marker and the eyebrow curvature marker. Together, they achieved the highest overall AUC (0.71) with an accuracy of 63% (sensitivity of 65% and specificity of 60%). The receiver operating characteristic curve for this combination of predictors is displayed in Fig. 7. Performance metrics for the four best combinations of predictors are shown in Supplementary Material 1.

Table 5 Results of diagnostic sensitivity analysis for facial palsy listing AUC, sensitivity, specificity and accuracy for single facial paralysis markers

#	Feature	AUC	Sensitivity	Specificity	Accuracy
Eyebrows					
(i)	Elevation / Depression	0	0	0	0
(ii)	Curvature	0.31	0.60	0.30	0.45
Eyes					
(iii)	Palpebral Fissure	0.49	0.60	0.50	0.55
Mouth					
(iv)	Elevation / Depression	0.53	0.60	0.55	0.58
(v)	Horizontal	0.64	0.80	0.60	0.70
(vi)	Asymmetry	0.55	0.75	0.50	0.63

AUC = Area under the curve

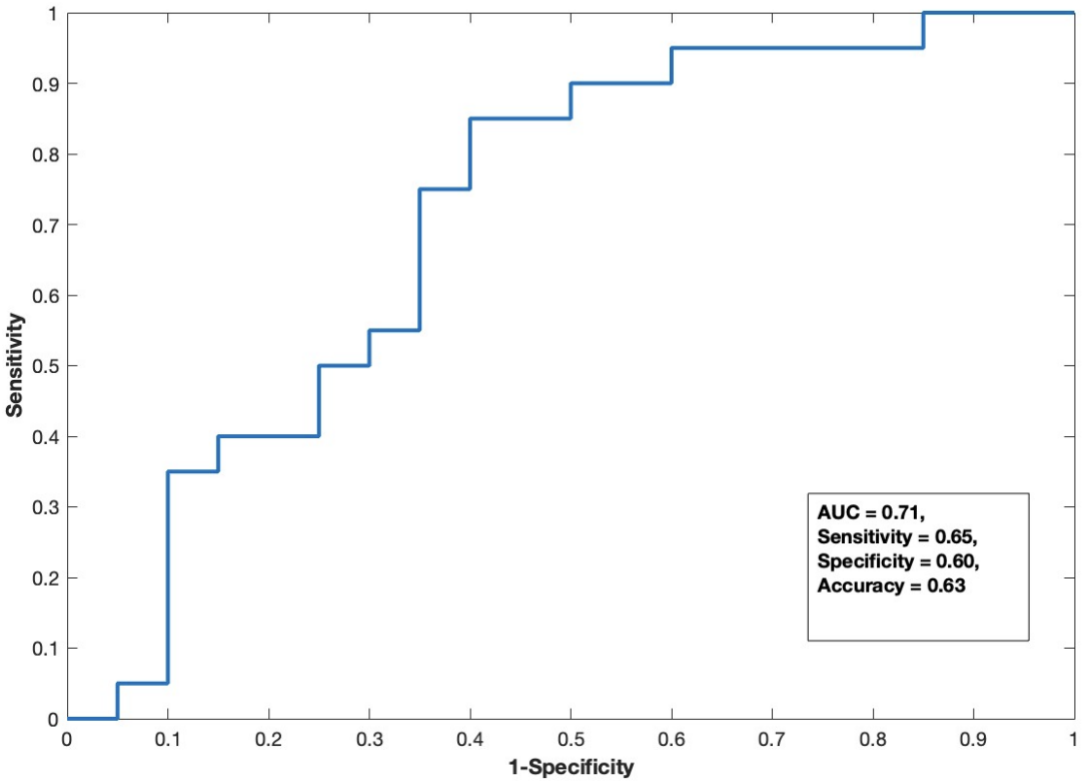


Fig. 7 Receiver operating characteristic curve for the multinomial logistic regression using the mouth horizontal movement, eyebrow curvature and palpebral fissure marker

A summary of mean coefficients and deviations of the optimal multinomial logistic regression model with the three determined markers as predictors: mouth horizontal, palpebral fissure and eyebrow curvature can be found in Table 6. For the palpebral fissure predictor, the mean coefficient was equal to $\beta_1 = -1.05 \cdot 10^7$, indicating that the relative risk of a sample belonging to the group HC as opposed to MS decreases by $e^{1.05 \cdot 10^7}$ with one unit increase of the predictor while all other predictors remain constant. Analogously, the risk would decrease for the mouth horizontal predictor with $\beta_2 = -2.95 \cdot 10^4$ and for the eyebrow curvature predictor with $\beta_3 = -1.82$. The mean intercept was $\beta_0 = 3.10$.

Table 6 Multinomial logistic regression model coefficient statistics

Predictors		μ	σ	**95% CI**
Intercept	β_0	3.10	0.207	[0.573 5.62]
Palpebral fissure	β_1	$-1.05e7$	$9.95e5$	$[-2.24e7$ $1.41e6]$
Mouth horizontal movement	β_2	$-2.95e4$	$2.64e3$	$[-5.42e4$ $4.89e3]$
Eyebrow curvature	β_3	-1.82	0.236	$[-4.15$ $0.516]$
Deviance of the fit		43.2	1.11	

95% CI – 95% confidence interval

σ – standard deviation

μ – mean

The deviance of the fit denotes the sum of deviance residuals

4. Discussion

In this thesis, a novel approach to identifying facial manifestations in MS has been presented. The Face Mesh algorithm was successfully used to extract 478 face landmarks that have then been used for the design of 6 individual facial movement markers, describing the difference in symmetric movement from 3 distinct face regions: the eyebrows, the eyes and the mouth. Using the designed markers, facial movement of 20 MS patients and 20 healthy controls has been analyzed with an automatic facial movement parametrization algorithm. Subsequently, multinomial logistic regression has been applied to create a classifier between MS patients that exhibit signs of facial palsy and healthy controls. The overall objectives of this paper, to do the first employment of a deep learning based landmark detection algorithm in the field of MS, to analyze the disruption of facial movement in patients with MS and to propose an automatic and objective approach for facial movement parametrization in patients with multiple sclerosis have been achieved. The strength of this algorithm lies in its ability to categorize a possible disruption of facial movement from freely spoken monologue that is the most naturally available form of facial movements. A computerized method of assessing facial manifestations in MS has great potential as it is non-invasive, easily interpretable and inexpensive. Specifically, the employed facial landmark detection algorithm was designed for a robust performance on mobile phones, which provides great scalability for possible future application in a wider population and may enable easy at-risk population-screening as it only requires e.g., a video-call with any subject. As opposed to previous proposed facial nerve grading systems that all require a physician to execute the assessment, the computerized aspect of the presented approach is able to eliminate any subjective component from the assessment and adds reproducibility to the measurement; it marks the first approach of an automatic, objective and video-based assessment for facial manifestations in MS. The proposed classification algorithm was able to differentiate between the MS group and the healthy control group with only natural speech recordings as its input with a moderately high accuracy of 62.5% and an area under the receiver operating curve of 0.71.

The three mouth markers were reported to have the most pronounced difference between the HC- and MS-group, with the mouth asymmetry marker reaching the level of significance ($p < 0.05$). This shows that facial asymmetry was by far best captured by analyzing the behavior of the opposing sides of the mouth. However, the reason of their above average performance may partly lie in the nature of the recordings, since in natural speech, the mouth shows the most voluntary movement, which inevitably may lead to a higher difference between groups because of more overall movement (McGettigan and Scott, 2014). Markers other than the mouth horizontal marker showed the expected trends towards decreased symmetry of facial movement

(see Fig. 6) but did not reach significance. The difference in mean ranks agrees with the design of the markers, where higher values indicate larger deviations from facial symmetry.

In the comparison of the computerized analysis and the perceptual assessment of the merged group of MS and HC, every marker agreed with the perceptual assessment score, showing a positive correlation. The mouth horizontal marker here too showed the highest significant correlation which supports that the markers indeed work analogously to a perceptual assessment and can be used in the future to create a similar grading system as it exists in the perceptual assessment beyond a binary classification. The reason for the moderate levels of correlation may lie in the crude nature of the perceptual assessment scale and the relatively limited sample size.

When considering highest performance of markers for the machine learning algorithm using a single predictor, the three markers describing movement of the mouth again performed best in the same order i.e., the mouth horizontal movement marker as best, then the mouth asymmetry marker and the mouth elevation/depression marker. On the contrary, the marker combination that led to the best overall AUC consisted of three markers, one of each examined separate facial region i.e., the eyebrow curvature marker, the palpebral fissure marker and the mouth horizontal marker. This may support the importance of regional assessment for classification in FNGS, as indicated by the many criticisms of the House-Brackman scale (Yanagihara and Hato, 2003; Di Stadio, 2015; Bansal *et al.*, 2020). Remarkable is that the algorithm was able to distinguish moderately well between the two groups even though the patients were not asked to perform specific unnatural movements that are commonly used in FNGS to assess the performance of specific muscles such as raising the forehead or closing of the eyes (House and Brackmann, 1985; Di Stadio, 2015).

A challenge to the performance of the landmark detection algorithm has been the natural head movements of the patients during the recording. As the head positions of the subjects were not fixed in order to not restrain the patient too much, subjects where free to move. Any head movement may have affected the sensitivity of the algorithm by introducing projection artefacts as noise to the measurement that can overshadow actual differences in movement between sides. Even though several steps have been made to minimize its effect, it may explain the lower between group differences of longer distance measure markers like the eyebrow elevation/depression marker. The low-frequency drift that is introduced by head movements could have eliminated small diversions in the relatively long distance between eyebrow and nose between the face sides. Nevertheless, distance markers concerning the mouth have performed much better, likely because the whole mouth is similarly affected by any head movement on both sides and the measured distances overall are smaller. For future applications, this challenge could be resolved or its effect further minimized either by determining the rotation and translation of each face by solving the pose computation problem and rotating the landmarks into a neutral

position (Marchand, Uchiyama and Spindler, 2016), or possibly by filtering the acquired measurements of both sides in order to eliminate the drift. It can be assumed that minimizing the issue of head movement will considerably improve the marker performance. Further limitations of this study include that the subjects were only European and, more specifically, Czech native speakers. This may raise uncertainty about the algorithm's performance for different ethnic groups. The markers however are mostly based on normalized distances, reducing the impact of differing facial features. Plus, the Face Mesh algorithm has been trained on 30k in-the-wild photos that include all ethnic groups. Lastly, the size of the database presents a limiting factor for this work that reduces the power of statistical inference. Because of this, no Bonferroni correction for multiple comparison has been applied as it would be too strict at $\alpha = \frac{0.05}{6} = 0.008\overline{3}$. With a bigger database, features that showed to be significant or close to it e.g., the mouth markers, may reach higher significance that will be important for possible future applications such as clinical assessment.

For future works, this analysis may be rerun on an enlarged database and may be enhanced with regards to the noises introduced from the nature of the landmark detection. The selection of proposed markers can also be extended e.g., markers for the cheek movement or forehead movement can be introduced to explore further areas to possibly observe facial asymmetry and ensure a more robust classification.

On top of that, the proposed solution shows great possibilities for future applications. On one side, it has high potential to be expanded into a variety of use-cases e.g., creating a mobile version of the algorithm to minimize the threshold for the at-risk population of attending an examination. Or, on the other side, the proposed algorithm may be expanded to assess other facial manifestations such as facial myokymia or hemifacial spasm. With revised markers, a regional score can be implemented in addition to an overall score, allowing a precise tracking of disease development or treatment impact.

4.1. Objectives of this thesis – completion

(1) The application of a deep learning based landmark detection algorithm to extract key points of facial expressivity of the subjects for further analysis

I have successfully applied Google's Mediapipe toolkit to extract facial landmarks. Based on these landmarks, six parameters describing the disruption of facial movement were extracted.

(2) The analysis of disruption of facial movement in patients with multiple sclerosis

I have found significant differences in the symmetry of mouth movement. The proposed marker shows positive correlation with perceptual assessment.

(3) The proposition of an automatic approach for facial movement parametrization in patients with multiple sclerosis that exhibit signs of facial paralysis

I have employed a fully automatic approach as a classification experiment reaching an AUC of 0.71, showing the utility of fully automatic video-based assessment of facial symmetry in MS.

5. Conclusion

In conclusion, this work presented a fully automatic, video-based approach to analyzing asymmetries in facial movements for patients with multiple sclerosis. It has successfully been the first employment of a deep learning based facial landmark detection algorithms in the analysis of facial manifestations in MS. A total of six facial movement markers have been proposed to parametrize asymmetric movement of the face. Significant differences were found in the face symmetry between MS patients with facial palsy and a healthy control group. Subsequently, a machine learning categorization algorithm, multinomial logistic regression, showed the applicability of the presented approach to automatically detect signs of facial paralysis and differ between the two groups. Results of this work confirmed and emphasized the need of regional assessment in the evaluation of facial movement disruption, the utility of an automated objective tool for facial disruptions in MS and presents disruption of facial movement as a possible disease biomarker.

References

Amato, M.P. *et al.* (2012) 'Association of MRI metrics and cognitive impairment in radiologically isolated syndromes', *Neurology*, 78(5), pp. 309–314. Available at: https://doi.org/10.1212/WNL.0b013e31824528c9.

Aubert-Broche, B. *et al.* (2014) 'Onset of multiple sclerosis before adulthood leads to failure of age-expected brain growth', *Neurology*, 83(23), pp. 2140–2146. Available at: https://doi.org/10.1212/WNL.0000000000001045.

Bansal, M. *et al.* (2020) 'A Simple, Objective, and Mathematical Grading Scale for the Assessment of Facial Nerve Palsy', *Otology & Neurotology*, 41(1), pp. 105–114. Available at: https://doi.org/10.1097/MAO.0000000000002450.

Beck, A.T., Steer, R.A. and Brown, G.K. (1987) *Beck depression inventory*. Harcourt Brace Jovanovich New York:

Bjornevik, K. *et al.* (2022) 'Longitudinal analysis reveals high prevalence of Epstein-Barr virus associated with multiple sclerosis', *Science*, 375(6578), pp. 296–301.

Bulat, A. and Tzimiropoulos, G. (2017) 'How Far Are We From Solving the 2D & 3D Face Alignment Problem? (And a Dataset of 230,000 3D Facial Landmarks)', in. *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1021–1030. Available at: https://openaccess.thecvf.com/content_iccv_2017/html/Bulat_How_Far_Are_ICCV_2017_paper.html (Accessed: 27 January 2023).

Calabresi, P.A. (2004) 'Diagnosis and management of multiple sclerosis', *American Family Physician*, 70(10), pp. 1935–1944.

Campbell, W.W. and DeJong, R.N. (2005) *DeJong's the Neurologic Examination*. Lippincott Williams & Wilkins.

Cootes, T.F., Edwards, G.J. and Taylor, C.J. (1998) 'Active appearance models', in H. Burkhardt and B. Neumann (eds) *Computer Vision — ECCV'98*. Berlin, Heidelberg: Springer (Lecture Notes in Computer Science), pp. 484–498. Available at: https://doi.org/10.1007/BFb0054760.

Cristinacce, D. and Cootes, T.F. (2006) 'Feature detection and tracking with constrained local models.', in *Bmvc*. Citeseer, p. 3.

Danesh-Sani, S.A. *et al.* (2013) 'Clinical Assessment of Orofacial Manifestations in 500 Patients With Multiple Sclerosis', *Journal of Oral and Maxillofacial Surgery*, 71(2), pp. 290–

294. Available at: https://doi.org/10.1016/j.joms.2012.05.008.

Di Stadio, A. (2015) 'Another Scale for the Assessment of Facial Paralysis? ADS Scale: Our Proposition, How to Use It', *JOURNAL OF CLINICAL AND DIAGNOSTIC RESEARCH* [Preprint]. Available at: https://doi.org/10.7860/JCDR/2015/15366.6953.

Di Stadio, A. *et al.* (2020) 'Clinical and radiological findings of facial paralysis in multiple sclerosis', *Multiple Sclerosis and Related Disorders*, 37. Available at: https://doi.org/10.1016/j.msard.2019.101456.

Di Stadio, A. and Bernitsas, E. (2018) 'The Origin of Facial Palsy in Multiple Sclerosis', *Otolaryngology - Open Journal*, 4. Available at: https://doi.org/10.17140/OTLOJ-4-e006.

Dobson, R. and Giovannoni, G. (2019) 'Multiple sclerosis – a review', *European Journal of Neurology*, 26(1), pp. 27–40. Available at: https://doi.org/10.1111/ene.13819.

Engell, T. (1989) 'A clinical patho-anatomical study of clinically silent multiple sclerosis', *Acta Neurologica Scandinavica*, 79(5), pp. 428–430. Available at: https://doi.org/10.1111/j.1600-0404.1989.tb03811.x.

Etemadifar, M. *et al.* (2022) 'Facial involvement in multiple sclerosis', *Multiple Sclerosis and Related Disorders*, 67, p. 104110. Available at: https://doi.org/10.1016/j.msard.2022.104110.

Evans, C. *et al.* (2013) 'Incidence and prevalence of multiple sclerosis in the Americas: a systematic review', *Neuroepidemiology*, 40(3), pp. 195–210. Available at: https://doi.org/10.1159/000342779.

Fukazawa, T. *et al.* (1992) 'Multiple sclerosis in Hokkaido, the northernmost island of Japan: prospective analyses of clinical features', *Internal Medicine*, 31(3), pp. 349–352.

Fukazawa, T. *et al.* (1997) 'Facial palsy in multiple sclerosis', *Journal of Neurology*, 244(10), pp. 631–633. Available at: https://doi.org/10.1007/s004150050158.

Ghosh, R. *et al.* (2022) 'Movement Disorders in Multiple Sclerosis: An Update', *Tremor and Other Hyperkinetic Movements*, 12, p. 14. Available at: https://doi.org/10.5334/tohm.671.

Gross, R. *et al.* (2010) 'Multi-pie', *Image and vision computing*, 28(5), pp. 807–813.

He, L. *et al.* (2022) 'Deep learning for depression recognition with audiovisual cues: A review', *Information Fusion*, 80, pp. 56–86.

House and Brackmann (1985) 'Facial nerve grading system', *Otolaryngol Head Neck Surg*, 93, pp. 184–193.

House, J.W. (1983) 'Facial nerve grading systems.', *The Laryngoscope*, 93(8), pp. 1056–1069.

Available at: https://doi.org/10.1288/00005537-198308000-00016.

Kartynnik, Y. *et al.* (2019) 'Real-time Facial Surface Geometry from Monocular Video on Mobile GPUs'. arXiv. Available at: https://doi.org/10.48550/arXiv.1907.06724.

Kingwell, E. *et al.* (2013) 'Incidence and prevalence of multiple sclerosis in Europe: a systematic review', *BMC Neurology*, 13(1), p. 128. Available at: https://doi.org/10.1186/1471-2377-13-128.

Kurtzke, J.F. *et al.* (1977) 'Studies on the natural history of multiple sclerosis—8: Early prognostic features of the later course of the illness', *Journal of chronic diseases*, 30(12), pp. 819–830.

Kurtzke, J.F. (1983) 'Rating neurologic impairment in multiple sclerosis: an expanded disability status scale (EDSS)', *Neurology*, 33(11), pp. 1444–1444.

Lassemi, E. *et al.* (2014) 'Oral and facial manifestations of patients with multiple sclerosis', *Dentistry*, 4(2), p. 1.

Lee, L.N. *et al.* (2013) 'A Comparison of Facial Nerve Grading Systems', *Annals of Plastic Surgery*, 70(3), pp. 313–316. Available at: https://doi.org/10.1097/SAP.0b013e31826acb2c.

Makhani, N. *et al.* (2014) 'MS incidence and prevalence in Africa, Asia, Australia and New Zealand: A systematic review', *Multiple Sclerosis and Related Disorders*, 3(1), pp. 48–60. Available at: https://doi.org/10.1016/j.msard.2013.06.015.

Marchand, E., Uchiyama, H. and Spindler, F. (2016) 'Pose Estimation for Augmented Reality: A Hands-On Survey', *IEEE Transactions on Visualization and Computer Graphics*, 22(12), pp. 2633–2651. Available at: https://doi.org/10.1109/TVCG.2015.2513408.

McGettigan, C. and Scott, S.K. (2014) 'Voluntary and involuntary processes affect the production of verbal and non-verbal signals by the human voice', *Behavioral and Brain Sciences*, 37(6), pp. 564–565.

Mehanna, R. and Jankovic, J. (2013) 'Movement disorders in multiple sclerosis and other demyelinating diseases', *Journal of the Neurological Sciences*, 328(1), pp. 1–8. Available at: https://doi.org/10.1016/j.jns.2013.02.007.

MSIF (2020) *Number of people with MS | Atlas of MS*. Available at: https://www.atlasofms.org/map/global/epidemiology/number-of-people-with-ms (Accessed: 10 February 2023).

Novotny, M. *et al.* (2022) 'Automated video-based assessment of facial bradykinesia in de-novo Parkinson's disease', *npj Digital Medicine*, 5(1), pp. 1–8. Available at:

https://doi.org/10.1038/s41746-022-00642-5.

Palacios, N. *et al.* (2011) 'Smoking and Increased Risk of Multiple Sclerosis: Parallel Trends in the Sex Ratio Reinforce the Evidence', *Annals of Epidemiology*, 21(7), pp. 536–542. Available at: https://doi.org/10.1016/j.annepidem.2011.03.001.

Polet, K. *et al.* (2022) 'Eye-gaze Strategies During Facial Emotion Recognition in Neurodegenerative Diseases and Links With Neuropsychiatric Disorders', *Cognitive and Behavioral Neurology*, 35(1), p. 14. Available at: https://doi.org/10.1097/WNN.0000000000000288.

Polman, C.H. *et al.* (2011) 'Diagnostic criteria for multiple sclerosis: 2010 revisions to the McDonald criteria', *Annals of neurology*, 69(2), pp. 292–302.

Prineas, J.W. *et al.* (2001) 'Immunopathology of secondary-progressive multiple sclerosis', *Annals of Neurology*, 50(5), pp. 646–657. Available at: https://doi.org/10.1002/ana.1255.

Ross, B.G., Fradet, G. and Nedzelski, J.M. (1996) 'Development of a Sensitive Clinical Facial Grading System', *Otolaryngology–Head and Neck Surgery*, 114(3), pp. 380–386. Available at: https://doi.org/10.1016/S0194-59989670206-1.

Rusz, J. *et al.* (2018) 'Characteristics of motor speech phenotypes in multiple sclerosis', *Multiple Sclerosis and Related Disorders*, 19, pp. 62–69. Available at: https://doi.org/10.1016/j.msard.2017.11.007.

Saleh, C. *et al.* (2016) 'Peripheral (Seventh) Nerve Palsy and Multiple Sclerosis: A Diagnostic Dilemma - A Case Report', *Case Reports in Neurology*, 8(1), pp. 27–33. Available at: https://doi.org/10.1159/000443681.

Schoneveld, L., Othmani, A. and Abdelkawy, H. (2021) 'Leveraging recent advances in deep learning for audio-visual emotion recognition', *Pattern Recognition Letters*, 146, pp. 1–7.

Sintzel, M.B., Rametta, M. and Reder, A.T. (2018) 'Vitamin D and Multiple Sclerosis: A Comprehensive Review', *Neurology and Therapy*, 7(1), pp. 59–85. Available at: https://doi.org/10.1007/s40120-017-0086-4.

Sun, Y., Wang, X. and Tang, X. (2013) 'Deep convolutional network cascade for facial point detection', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3476–3483.

Tremlett, H., Munger, K.L. and Makhani, N. (2022) 'The Multiple Sclerosis Prodrome: Evidence to Action', *Frontiers in Neurology*, 12. Available at: https://www.frontiersin.org/articles/10.3389/fneur.2021.761408 (Accessed: 25 November 2022).

Vrabec, J.T. *et al.* (2009) 'Facial Nerve Grading System 2.0', *Otolaryngology–Head and Neck Surgery*, 140(4), pp. 445–450. Available at: https://doi.org/10.1016/j.otohns.2008.12.031.

Wu, Y. and Ji, Q. (2019) 'Facial Landmark Detection: A Literature Survey', *International Journal of Computer Vision*, 127(2), pp. 115–142. Available at: https://doi.org/10.1007/s11263-018-1097-z.

Yanagihara, N. and Hato, N. (2003) 'Assessment of Facial Nerve Function Following Acoustic Neuroma Surgery: Facial Nerve Grading Systems', in J. Kanzaki et al. (eds) *Acoustic Neuroma*. Tokyo: Springer Japan (Keio University International Symposia for Life Sciences and Medicine), pp. 91–98. Available at: https://doi.org/10.1007/978-4-431-53942-1_16.

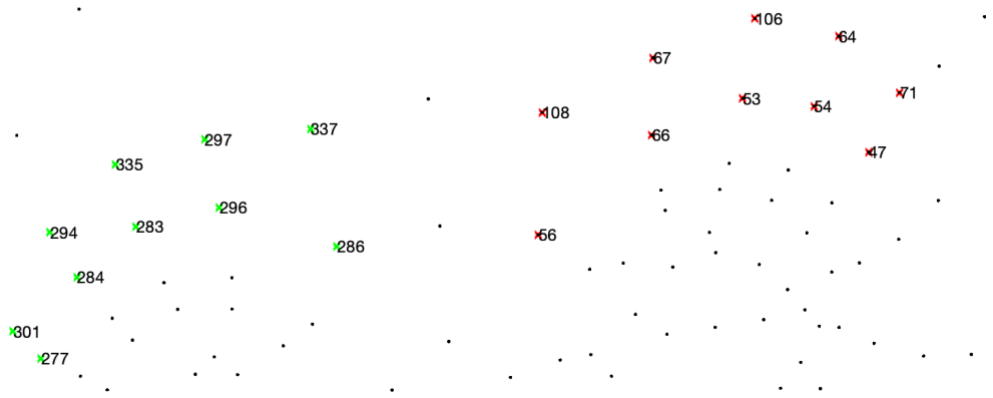
Zhou, E. *et al.* (2013) 'Extensive facial landmark localization with coarse-to-fine convolutional network cascade', in *Proceedings of the IEEE international conference on computer vision workshops*, pp. 386–391.

Supplementary Material

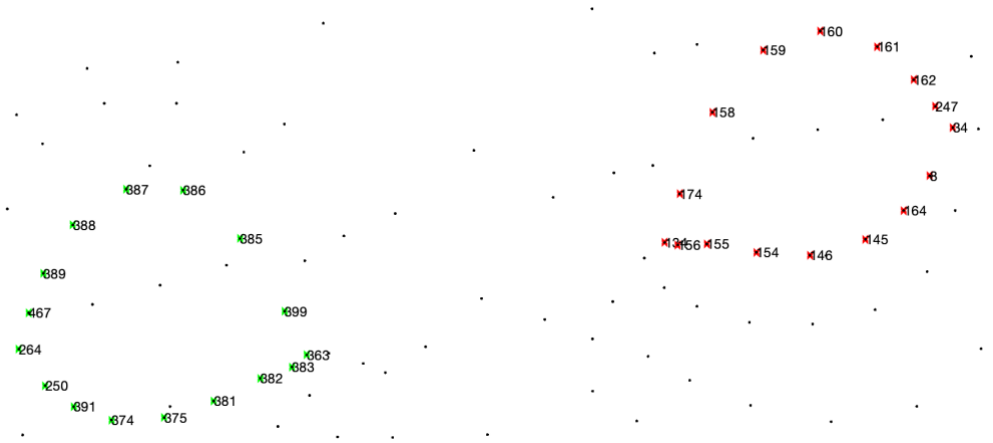
Supplementary Material 1 Results of diagnostic sensitivity analysis showing AUC, sensitivity, specificity and accuracy for the four best predictor combinations

Predictor combination	AUC	Sensitivity	Specificity	Accuracy
Best Performing				
(i) Eyebrow curvature *(ii) Palpebral Fissure* *(iii) Mouth horizontal*	0.713	0.65	0.6	0.63
2nd Best				
(i) Eyebrow curvature *(ii) Palpebral Fissure* *(iii) Mouth horizontal* *(iv) Mouth elevation/depression* *(v) Mouth asymmetry*	0.695	0.70	0.60	0.65
3rd Best				
(i) Palpebral Fissure *(ii) Mouth horizontal* *(iii) Mouth elevation/depression* *(iv) Mouth asymmetry*	0.693	0.65	0.55	0.60
4th Best				
(i) Eyebrow curvature *(ii) Palpebral Fissure* *(iii) Mouth horizontal* *(iv) Mouth asymmetry*	0.690	0.70	0.55	0.63

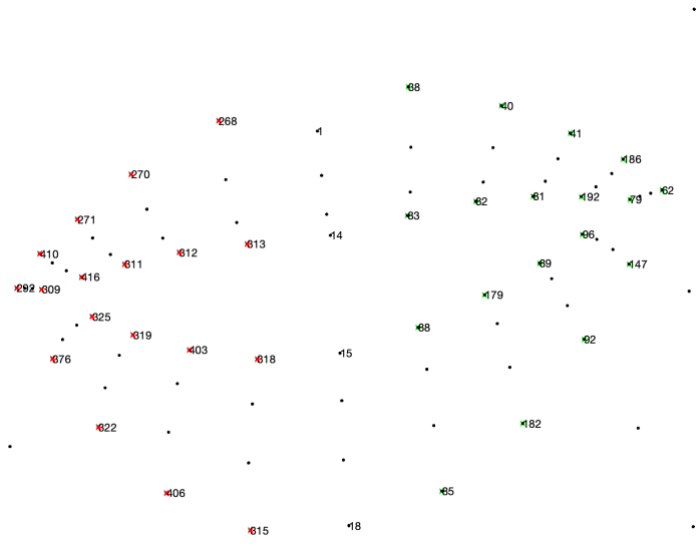
AUC = Area under the curve



Supplementary Material 2 Close up view the eyebrow landmarks of the Face Mesh landmark detection algorithm



Supplementary Material 3 Close up view of the eye landmarks of the Face Mesh landmark detection algorithm



Supplementary Material 4 Close up view of the mouth landmarks of the Face Mesh landmark detection algorithm