Asunto: Notificación de Dictamen de Modalidad de Titulación.

GEMMA SILVANA PARRA DOMINGUEZ ESTUDIANTE DEL DOCTORADO EN INGENIERÍA ELÉCTRICA

P R E S E N T E

Una vez concluido el análisis y discusión sobre la propuesta del proyecto titulado: ***"Análisis de Imágenes Usando Propiedades de Simetría Facial para las Tareas de Detección y Medición del Grado de Parálisis Facial y el Reconocimiento de Gestos en Pacientes"*** , bajo la revisión del Asesor(es): *Dr. Raúl Enrique Sánchez Yáñez* y *Dr. Carlos Hugo García Capulín*, el Comité Académico del Doctorado en Ingeniería Eléctrica, acordó la siguiente resolución:

Con fundamento en las fracciones primera a cuarta del Artículo 67, el Artículo 68 del Estatuto Académico del Reglamento Académico, el comité acordó **POR UNANIMIDAD DE VOTOS AUTORIZAR SU MODALIDAD DE TITULACIÓN POR TRABAJO DE TESIS** designando el siguiente jurado:

Presidente: DR. HORACIO ROSTRO GONZÁLEZ

Secretario: DR. RAÚL ENRIQUE SÁNCHEZ YÁÑEZ

Vocal1: DR. JOSÉ RUÍZ PINALES

Vocal 2: DR. IVÁN CRUZ ACEVES

Vocal 3: DR. JOSÉ JOEL GONZÁLEZ BARBOSA

Se extiende la presente notificación de la resolución el día 22 de junio del 2022.

A T E N T A M E N T E
"LA VERDAD OS HARÁ LIBRES"

DR. JOSÉ AMPARO ANDRADE LUCIO
DIRECTOR DE LA DIVISIÓN DE INGENIERÍAS DEL CIS UNIVERSIDAD DE GUANAJUATO

C.c.p. **Archivo**

M. en I. HERIBERTO GUTIÉRREZ MARTIN
COORDINADOR DE ASUNTOS ESCOLARES
P R E S E N T E.-

Por medio de la presente, se otorga autorización para proceder a los trámites de impresión, empastado de tesis y titulación al alumno(a) _Gemma Silvana Parra Domínguez_ del *Programa de Doctorado en_ Ingeniería Eléctrica* y cuyo número de *NUA* es: _144045_ del cual soy director. El título de la tesis es: _Análisis de Imágenes Usando Propiedades de Simetría Facial para las Tareas de Detección y Medición del Grado de Parálisis Facial y el Reconocimiento de Gestos en Pacientes_

Hago constar que he revisado dicho trabajo y he tenido comunicación con los sinodales asignados para la revisión de la tesis, por lo que no hay impedimento alguno para fijar la fecha de examen de titulación.

A T E N T A M E N T E

Dr. Raúl Enrique Sánchez Yáñez
NOMBRE Y FIRMA
DIRECTOR DE TESIS
SECRETARIO

Dr. Carlos Hugo García Capulín
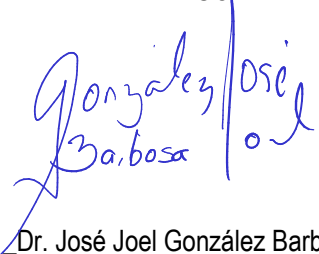NOMBRE Y FIRMA
DIRECTOR DE TESIS

Dr. Horacio Rostro González
NOMBRE Y FIRMA
PRESIDENTE

Dr. José Ruíz Pinales
NOMBRE Y FIRMA
VOCAL

Dr. Iván Cruz Aceves
NOMBRE Y FIRMA
VOCAL

Dr. José Joel González Barbosa
NOMBRE Y FIRMA
VOCAL
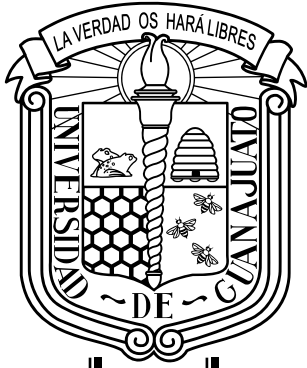
UNIVERSIDAD DE GUANAJUATO

CAMPUS IRAPUATO-SALAMANCA
DIVISIÓN DE INGENIERÍAS

Análisis de Imágenes Usando Propiedades de Simetría Facial para las Tareas de Detección y Medición del Grado de Parálisis Facial y el Reconocimiento de Gestos en Pacientes

TESIS

QUE PARA OBTENER EL GRADO DE:

DOCTORA EN INGENIERÍA ELÉCTRICA

PRESENTA:

M. en I. Gemma Silvana Parra Domínguez

DIRECTOR DE TESIS:

Dr. Raúl E. Sánchez Yáñez

CODIRECTOR DE TESIS:

Dr. Carlos H. García Capulín

SALAMANCA, GTO.   18 de Agosto de 2022
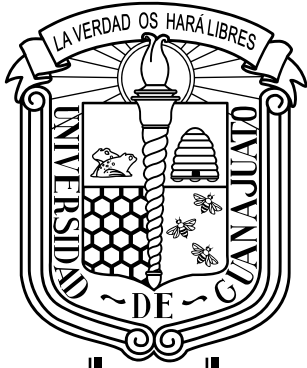
UNIVERSITY OF GUANAJUATO

CAMPUS IRAPUATO-SALAMANCA
ENGINEERING DIVISION

Image Analysis Using Facial Symmetry Features for the Tasks of Detection and Assessment of Facial Palsy and Gesture Recognition in Patients

THESIS

TO OBTAIN THE DEGREE OF:

DOCTOR IN ELECTRICAL ENGINEERING

PRESENTS:

M.Eng. Gemma Silvana Parra Domínguez

DOCTORAL SUPERVISORS:

Dr. Raúl E. Sánchez Yáñez

DOCTORAL COSUPERVISORS:

Dr. Carlos H. García Capulín

SALAMANCA, GTO.   August 18, 2020

ABSTRACT

Facial paralysis is a physical condition that negatively affects the facial structure. It is characterized by the incapacity to move some muscles of the face. Different factors can develop facial paralysis, for example, a congenital condition, trauma, disease like stroke, brain tumor, or Bell's palsy. Since paralysis can affect one or both sides of the face, there is a remarkable drooping of facial movements. Therefore, the patient exhibits difficulty performing daily life activities such as speaking, blinking, swallowing saliva, eating, or communicating through natural facial expressions.

Recently, computer-based systems have been designed to automatically diagnose facial paralysis through images. Those systems are essential in developing standardized tools for medical assessment, treatment, and monitoring; additionally, they are expected to provide user-friendly tools for patient monitoring at home.

Simultaneously, computed-based systems are also being developed to recognize facial expressions in a wide variety of situations. Facial expressions are crucial in designing security, healthcare, entertainment, advertisement, education, and robotics applications. Performing these recognition tasks in palsy patients could allow their integration in developing and using such applications.

This research seeks to diagnose facial palsy in a photograph using computer vision techniques and machine learning algorithms. The analysis of faces is performed in terms of symmetry. Knowing that almost all humans exhibit a level of asymmetry between the left and right sides of the face, it is expected to find a boundary capable of distinguishing between healthy and palsy faces. In other words, this work seeks to characterize human faces by assuming that a healthy face is pretty symmetrical and a palsy face is not.

This project was split into three sub-projects to facilitate its reading: (1) facial palsy detection, (2) grading of facial palsy, and (3) gesture recognition in palsy patients. Facial palsy detection is a binary classification that aims to discriminate between healthy subjects and palsy patients. The diagnosis of facial paralysis involves at least the task of palsy detection. Assessing facial palsy is the process of grading the level of palsy severity, and it is addressed as a multi-class classification. Gesture recognition in palsy patients is also

interpreted as a multi-class classification task.

The proposed methodology to analyze facial paralysis consists of a three modules system performing (1) facial landmark prediction, (2) facial feature extraction, and (3) classification tasks. This framework predicts facial landmarks from a face image to later compute measures before extracting facial features. The proposed features seek to characterize the face's asymmetry, and this is a core trait in the methodology. Predictive models that perform different classification tasks are built using those facial features.

The proposed methodology successfully explores the use of regional information, meaning that only certain areas of the face are of interest, and those are the eyes and the mouth region. This perspective leads us to believe that the analysis of facial paralysis is possible with partial occlusions if face detection is accomplished and facial features are obtained adequately.
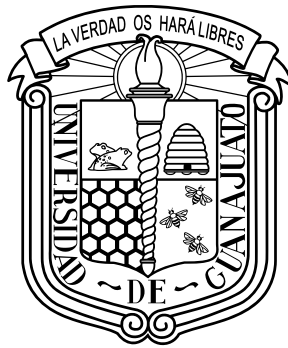
Tests on four publicly available databases reveal outstanding classification results for detecting and grading facial palsy. The achieved performance also shows that the methodology is suited to operate with other databases while attaining high performance, even though the image characteristics are different and the participants do not perform equivalent facial expressions.

Detecting facial expressions via gesture recognition is a complex and challenging problem, especially in persons who suffer face impairments, such as patients with facial paralysis. The proposed gesture recognition system is the first attempt to perform this task in the medical field. It was evaluated on a publicly available database and achieved good results.

Recognizing facial gestures in patients with facial paralysis should consider the lesion's severity. It is found that the paralyzed facial elements exhibit different behavior than the healthy ones. Any recognition system must be capable of discerning these behaviors to accomplish this task.

At the end of this project, it is found that analyzing facial paralysis from images requires new data and models to overcome the bias produced using healthy people photographs in the design, training, and testing of algorithms. Improvements to almost any methodology would require more samples from palsy patients with annotated data, including information about the lesion's severity, exhibited facial expression, and felt emotion.

INSTITUTIONAL ACKNOWLEDGEMENTS

national scholarship program Becas Nacional (Tradicional) 2018 - 1.

CONTENTS

LIST OF FIGURES

Chapter List of Figures

LIST OF TABLES

CHAPTER 1

PROJECT DESCRIPTION

This research project is related to the abilities that most humans display through facial expressions in our daily activities. It addresses solutions based on computer vision techniques and machine learning algorithms to a problem observed when those abilities are not adequately performed due to facial paralysis.

1.1 Introduction

It is widely known that our face is perhaps our most important channel of nonverbal communication; it encodes messages through our facial expressions. We might be able to understand what the others are feeling by seeing their faces, and they can do the same by seeing ours. It is said that our identity, intent, and emotion can be extracted from this visual nonverbal communication [51]. Moreover, the human ability to effectively communicate emotion is essential to performing daily activities and is required for personal and social sufficiency. It has been shown that 93% of our communication is non-verbal; thus, facial expressions and body gestures play an essential role in it [31].

Facial paralysis is a physical condition that negatively affects the structure of the face. This paralysis is characterized by the inability to move some facial muscles. It can be produced by a congenital condition, trauma, diseases like stroke, brain tumor, or Bell's palsy. Since the problem can affect one or both sides of the face, there is a noticeable drooping on the facial movements and, therefore, difficulty performing daily life activities such as speaking, blinking, swallowing saliva, eating, or communicating through natural facial expressions.

In the literature, facial paralysis is also referred to as facial palsy; both are correct and used interchangeably in this thesis. Also, in this research, we refer to people with facial paralysis when talking about palsy patients. And by palsy face, we mean a person's face with facial paralysis.

Figure 1.1: Example of face images from palsy patients obtained from [19].

When diagnosing this physical condition, the visual signs of paralysis provide valuable information to the healthcare professional about the state of the patient [45]. The diagnosis begins with visually inspecting the patient's face to evaluate its symmetry. A second inspection could occur by asking the patient to perform certain gestures, for example, smile, close the eyes, lift the eyebrows, wrinkle the nose, or remain with a neutral expression. After it, the practitioner can determine how affected are the facial muscles, which region is compromised, and the type of palsy the patient suffers. Figure 1.1 shows a set of images displaying different levels of paralysis.

The visual inspection of facial paralysis is a subjective evaluation that may change from practitioner to practitioner. Looking to perform an objective assessment, many approaches seek to develop automated computer-based systems to provide standardized tools for detecting and evaluating facial paralysis. Such standardized methods are also expected to help in the treatment and monitoring of the patient by providing user-friendly tools and reducing healthcare costs with the inclusion of automatic processes [45].

The analysis of facial signs has motivated a lot of studies in the field of computer vision, which aim to automatically assess the facial nerve function from a biomedical visual capture of the face. Visual capture of the face goes from traditional color images and video to infrared (thermal imaging) and depth images (stereophotogrammetry). Although various automated solutions have been proposed, their outcomes fall into two main categories, as described in [27]. The first provides, as a result, a non-semantic numerical value that quantifies static, dynamic, and synkinetic facial features. The second produces a semantic grade of the facial nerve function according to a scale designed by a clinician.

Many existing solutions belong to the first category. However, most of them stay in the discussion phase, and only a few are implemented into prototypes. A typical system of its kind consists of a user-friendly graphical interface to acquire and process the facial

data, with a measuring algorithm embedded to evaluate the functionality of the face. The outcome shows the measures with plots and tables in a graphical manner known as facegram. These solutions provide detailed insights and quantifications about abnormal conditions, but they still need clinicians to judge the severity of facial nerve dysfunction. An example is the *Emotrics* software developed by Guarin *et al.* and introduced in [14], which can easily be downloaded from the web.

Solutions in the second category aim to quantify the facial nerve function. In other words, to determine how much facial paralysis the patient suffers according to a grading scale. Those scales are clinically designed to split the damage's severity into discrete levels, using some rigorously-validated measures. Those measures are related to the symmetry of the face while displaying a neutral expression or showing a set of voluntary facial muscle movements; they are also associated with other dysfunctions such as synkinesis, as mentioned by [45]. The grading scales include House-Brackmann, Sunnybrook, Yanagihara, FNGS 2.0, and eFACE.

Solutions aiming to quantify the facial nerve function mainly apply machine learning algorithms to build a predictive model. Notice that labelled data of palsy patients and healthy subjects are required to train and test the model. The output is a grade representing the facial nerve function, but it could also be a binary value indicating whether the person has facial palsy or not.

A system that quantifies facial nerve function consists of three modules: pre-processing, processing, and classification. In the first, the face detection and landmark extraction process occur independently if the image has been previously enhanced or not. Depending on the methodology, handcrafted features are extracted in the second module, although features can also be computed via deep learning approaches. And in the third, a classifier is trained to map the features to a pre-defined grade of the facial nerve function. Classifiers for these applications are based on the methods of support vector machine (SVM), artificial neural network (ANN), k-nearest-neighbor (KNN), and multinomial logistic regression (MNLR), or hybrid classifier [27].

Landmarks are preferred interest points in an image for various computer vision tasks such as image measurement, registration, camera calibration, motion analysis, 3D scene reconstruction, and object recognition. They are also known as reference points because landmarks (should) remain present in the image after (a certain degree of) translation, rotation, change of point of view, different illumination, etc.

Facial landmarks are key reference points from a human face extracted according to a specific face model. This model was previously trained on thousands of images, particularly healthy white people imagery. There are some publicly available models, also known as shape predictors. For example, the 68-point model depicted in Figure 1.2 is widely known and employed in facial analysis.

It has been observed that predicting key points in palsy patients is somehow inaccurate because the model was trained using imagery from healthy persons [14, 20, 46]. The extraction of facial landmarks is crucial because these key points are used to compute other measures before extracting the facial features. Recently, some authors have been
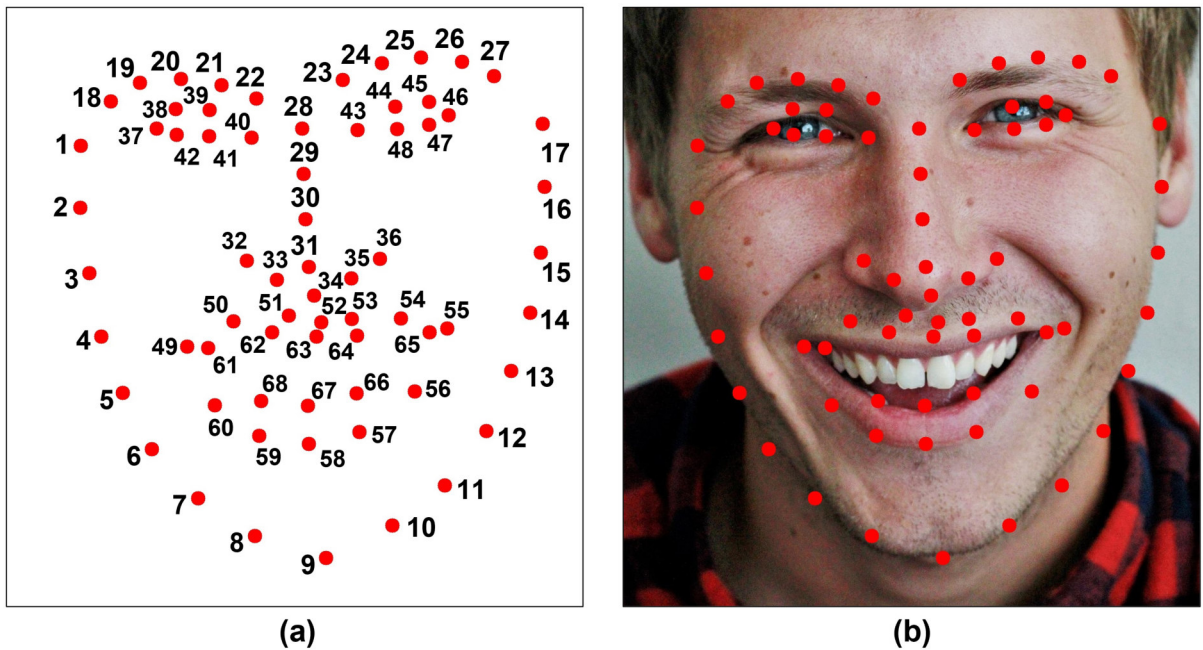
Figure 1.2: Example of the 68-point model proposed by Matthews and Baker [30]: (a) the predicted facial landmarks and (b) the facial landmarks overlaid on the corresponding color image.

working on improving the available shape predictors by using palsy patients' images. For example, the model developed and released by Guarin *et al.* in [15] is of great interest to our research.

This project is based on predicting facial landmarks from a face image to later compute measures before extracting facial features. Our handcrafted features seek to characterize the face's asymmetry, and this is a core trait in our methodology. Predictive models that perform different classification tasks are built using those facial features.

This project was split into three sub-projects to facilitate its reading: (1) facial palsy detection, (2) grading of facial palsy, and (3) gesture recognition in palsy patients. Facial palsy detection is a binary classification that aims to discriminate between healthy subjects and palsy patients. The diagnosis of facial paralysis involves at least the task of palsy detection. Assessing facial palsy is the process of grading the level of palsy severity, and it is addressed as multi-class classification. Gesture recognition in palsy patients is also interpreted as a multi-class classification task.

In the literature, the works in [5, 6, 16, 20, 22, 46] evaluate facial palsy following a methodology based on features extracted from facial landmarks. It is essential to mention that most approaches operate on specific facial expressions and others require a set of movements to output a result; our method does not depend on a particular facial expression or a collection of images to produce an outcome. Most of those methodologies operate on proprietary databases. To our knowledge, those are unavailable to the research community;

thus, a direct comparison of the performance is unfeasible. Therefore, we searched publicly available databases to develop our facial palsy analysis system.

1.1.1 Diagnosis of facial paralysis

Next, some strategies that detect facial paralysis as a binary classification problem are mentioned. Kim *et al.* in [22] proposed a smartphone-based automatic diagnosis system, which consists of three modules and analyzes three specific facial movements. The authors reported their highest classification accuracy at 88.9%. Barbosa *et al.* proposed in [6] a method to diagnose facial paralysis in two stages; first, facial palsy is detected, and then a grading among the unhealthy subjects is performed. Also, four specific facial expressions are used to measure the subject's facial symmetry. The authors reported a sensitivity of 98.12% and specificity of 98.34% in the first task; 97.48% and 94.91%, respectively, for the second task.

Next, some approaches that seek to evaluate the level of palsy severity as a multi-class classification problem are mentioned. Wang *et al.* in [46] proposed an evaluation method considering both static facial asymmetry and dynamic transformation factors. Their analysis uses five specific facial expressions; the best-reported accuracy is 97.56%. Barbosa *et al.* in [5] introduced an approach based on iris segmentation to classify facial paralysis quantitatively. Their methodology includes three stages and operates on five specific facial movements. The authors reported an overall accuracy of 93.7%. Guo *et al.* [16] proposed 49 features to quantify facial symmetry in particular areas of the face. Using five specific facial expressions, the authors reported an accuracy rate of 90.01%. Jian *et al.* in [20] introduced the hypothesis that facial skin perfusion can measure the severity of the lesion. Then, their methodology mainly calculates characteristics related to the blood flow distribution on the face. The authors reported a classification accuracy of up to 97.14%.

Neural networks play an essential role in solving classification problems in the medical field. A few methodologies are working with deep neural networks to diagnose facial palsy (*e.g.,* [17, 18, 40, 41, 26, 1]). Deep learning methods can improve the facial palsy detection rate, but insufficient data limit their efficiency and class imbalance, according to [1]. These methods automatically learn discriminative features, meaning they do not compute handcrafted features but perform pre-processing steps before training and evaluating the network. Due to the enormous amount of required data, geometric and color transformations, image cropping, and resizing are needed to increase the number of samples.

Two of those deep learning approaches are of our interest because of the databases they employed, the work of Hsu *et al.* in [18] and Liu *et al.* in [26]. Hsu *et al.* formulated facial palsy identification as an object detection problem and considered the deformation regions on the patient's face as the target objects. The authors reported a precision of 93% at recall 88% on a database of 21 patients. Liu *et al.* proposed segmentation of the facial area into two palsy regions to improve the detection rate. The authors reported an average accuracy of 94.81% on their grading scale with three severity levels (normal, mild-to-moderate, and severe), using a database with 26 patients.

1.1.2 Gesture recognition in facial paralysis

Facial expressions are among the most flexible ways humans communicate emotion. Their universality allows us to rapidly convey information to people of different ages, cultures, and languages [24]. This is why gesture and facial expression recognition has recently become a popular topic with applications in various fields, such as security, healthcare, entertainment, advertisement, education, and robotics [54].

Computer vision has reached high accuracy in the automatic recognition of facial expressions in healthy people; extensive research is found in [33, 54, 12, 47]. Facial Expression Recognition (FER) systems are computer-based technology that uses mathematical algorithms to analyze faces in images or video to perform recognition tasks [35, 11].

One or more gestures can represent a facial expression; we assume that identifying those could help discriminate one expression from another. Analyzing facial expressions allows us to classify them into emotion categories; in other words, giving meaning to a facial expression lets us recognize and label it as anger, disgust, fear, happiness, sadness, surprise, contempt, etc. Usually, an algorithm designed to recognize emotions also performs facial expression recognition tasks. Thus, it is said that the recognition of emotional states is based on facial expressions.

Detecting facial expressions continues to be a complex and challenging problem [33]. Most FER systems operate in three basic stages: face detection, feature extraction, and classification [42]. They can be divided into appearance or geometric-based methods. Appearance-based methods extract features from textural information about the face. In contrast, geometric-based methods rely on features computed from the shape change information of the face during expressions [53]. Several machine-learning algorithms accurately classify facial expressions and emotions. Some popular classifiers are the linear discriminant classifiers, k-nearest neighbor, Gaussian mixture model, support vector machines, artificial neural networks, deep neural networks, decision tree algorithms, and hidden Markov models [12].

Emotion detection based on facial expressions is currently a popular research area in computer science. However, the face analysis for medical and health applications is still in an embryonic state [24]. Applications in the medical field are designed to reduce the patient's stress, depression, and anxiety. In other words, automated systems are being developed to recognize their emotions and then provide appropriate therapy to manage any adverse reaction in the patient [12].

According to Leo *et al.* in [24], some healthcare frameworks include an emotion or expression recognition module. They have been introduced to provide suitable solutions for the following: (1) long-term care for person-centered and integrated healthcare systems; (2) diagnosis or assessment of cognitive impairments (*e.g.,* autism, schizophrenia, and profound intellectual and multiple disabilities [10]) and automated pain detection; (3) the use of technological rehabilitation frameworks [44, 43], and (4) the design of smart environments that react in a friendly manner according to the patient's necessities.

Yolcu *et al.* in [53] argued that some neurological disorders show well-known impaired

facial expressions. Therefore, it is necessary to design automated systems to detect emotions in–for example–patients with Parkinson's, stroke, and facial palsy. In the literature, and to the best of our knowledge, only two recent studies aim to evaluate the emotions from facial expressions in patients with facial palsy. Those are the work of Xu *et al.* [52] and the analysis of Kuttenreich *et al.* [23]. The latter is a remarkable study on performing emotion recognition in palsy patients, but it does not propose an automated system but the result of a self-assessment evaluation.

Our research seeks to classify facial gestures in palsy patients as a first step in recognizing emotions from facial expressions. Xu *et al.* in [52] stated that the automatic recognition of emotions could be a solution to help understand facial palsy patients, acknowledge their stress in advance, and assist their treatment. The authors also stated that their facial expressions are different from healthy subjects due to the inability of the facial muscles, making the existing data and models from healthy people invalid [52].

1.2 Motivation

In this thesis, we wish to create a multidisciplinary project involving image processing and artificial intelligence and their applications in the biomedical field. We have identified the need for specific analysis, data, and models to be extracted from images of facial palsy patients, mainly because the actual work was developed and executed on healthy people imagery. We assume that by having such specific information and algorithms, the detection rate of facial paralysis from an image can be improved.

From a different perspective, living in a digital area with many smart applications trying to control our environment using our body, hand, and facial gestures, we found it crucial to develop systems capable of recognizing facial gestures and expressions in persons with facial paralysis. We assume this would facilitate their integration in using that security, healthcare, and entertainment-related applications, where facial expressions play an important role.

1.3 Objectives

This thesis work's primary objective has been to analyze facial paralysis based on computer vision techniques and machine learning algorithms. Our objectives focus on three principal areas: (1) the detection of facial palsy on an image, (2) the assessment of the lesion's severity, and (3) the recognition of facial gestures in palsy patients. The design of our methodology included partial goals such as a literature review, obtaining a publicly available image database, and submitting our findings to a peer-review journal.

1.4 Contributions

The main contribution of this research is an image analysis for detecting facial paralysis using computer vision techniques and machine learning algorithms. It is such a challenging project that it was split into three sub-projects to cover more areas of that facial palsy analysis. Those sub-projects are deeply discussed in the following chapters. Still, here we mention a list of our contributions:

- An analysis of three publicly available image databases of palsy patients to effectively extract facial landmarks from a face image.

- A set of facial symmetry features is easily computed from measures extracted from the predicted facial landmarks.

- A methodology to design classifiers based on facial features and a single face image to output a result.

- A framework to detect and grade facial palsy using information extracted from the entire face or specific facial regions.

- Performance analysis on detecting and grading facial palsy using different classifiers.

- A set of palsy data and classification models to improve the detection and assessment of facial paralysis.

- A methodology based on facial features and classifiers to recognize facial gestures in palsy patients.

- An analysis of facial gestures in persons with facial palsy as a first step in achieving emotion recognition.

1.5 Thesis outline

This section describes the organization of this thesis. In Chapter 2, our proposed facial features are introduced. Later, Chapter 3 describes our methodology to design classifiers for detecting and grading facial paralysis. Similarly, Chapter 4 describes our method to create classifiers to recognize facial gestures in palsy patients. Finally, our conclusions and aims for future work are provided.

1.5.1 Analysis of facial paralysis

This chapter introduces face analysis in terms of symmetry. Here, it is assumed that a healthy person will present a pretty symmetric face, and a palsy patient will not. Our

proposed facial features seek to characterize human faces in terms of symmetry. Those features are composed using facial measures computed from previously predicted landmarks. Landmark extraction is crucial to calculate our features accurately; it is the core of our proposed method.

1.5.2 Detection of facial paralysis

Chapter three introduces our methodology to design classifiers based on facial features and machine learning algorithms for two specific tasks. First, we seek to detect palsy patients in a set of photographs through a binary classification (*i.e.,* to discriminate between healthy and palsy faces). Second, we address the challenge of grading the level of paralysis found in the patient as a multi-class classification problem. We perform different evaluations aiming to discriminate between slight and strong palsy patients; between healthy, slight, and strong palsy patients. The work of this chapter produced one article that was published in the *Applied Sciences* journal (refer to [38]) and another that was published in the *Diagnostics* journal (refer to [37]).

1.5.3 Gesture recognition in palsy patients

Chapter four introduces an initial methodology to recognize facial gestures in palsy patients. It attempts to perform a facial analysis with an intention other than diagnosing the paralysis. Here, it is established that a set of facial muscle movements compose a gesture, one or more gestures compose a facial expression, and facial expressions play an essential role in human emotions. Our work argues that detecting gestures is crucial before achieving facial expression recognition. This method is also based on our proposed facial features and a multi-class classifier. This chapter's work produced one article that was accepted and published in the *Healthcare* journal (refer to[39]).

CHAPTER 2

ANALYSIS OF FACIAL PARALYSIS

The analysis of the human face in terms of symmetry/asymmetry is not a new topic, it was mentioned by Borod and van Gelder in [8], and there is plenty of literature about it. We talk about the symmetry of a person's face knowing that we all are asymmetric. In other words, facial asymmetry responds to the fact that the left and right sides of our face showing no movement or while performing an expression are not identical. This asymmetry can result from various factors, including anatomical, neurological, physiological, pathological, psychological, and socio-cultural variables [8]. The growth and development of bones, nerves, and muscles should not produce additional asymmetries. Therefore, we assume that healthy faces tend to have identical sides at a neutral expression or while showing specific gestures, while palsy faces do not. This allows us to characterize the face of a palsy patient using machine-learning algorithms and grading scales.

Our facial palsy analysis relies on the extraction of handcrafted features. Handcrafted features generally refer to properties derived from the information in the image itself, using various algorithms [34]. The most common example of simple attributes extracted from images is edges and corners. Features are manually designed to overcome specific issues like occlusions and variations in scale and illumination. The design of handcrafted features often involves finding the suitable trade-off between accuracy and computational efficiency.

The computation of handcrafted features is usually a two-step procedure once the face is detected. First, a key point detector locates distinct regions of an image; in our case, this detector is represented by the shape predictor that extracts facial landmarks. Second, a descriptor is built using cues that depend on the features. Our descriptor comprises 29 facial features extracted from arithmetic operations and landmarks. This descriptor can be defined as a vector of measurements set to train a classifier.

Handcrafted features have been implemented for decades and still serve as a powerful tool when combined with machine learning classifiers [25]. Supervised learning algorithms, which map input data to output labels from a set of training examples, are widely used in medical image analysis. Deep-learning approaches are becoming increasingly popular

10
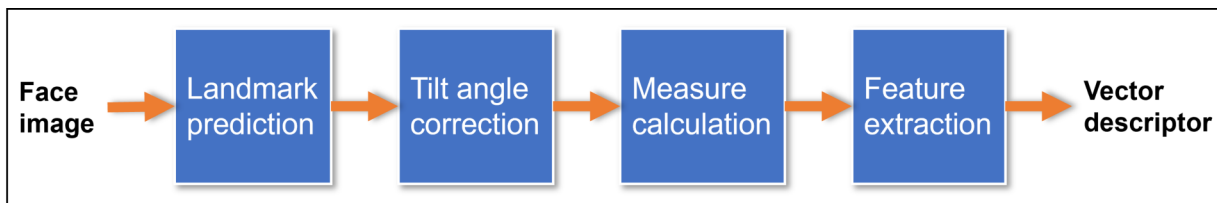
Figure 2.1: Our process to extract facial features.

because of their impressive classification performance. They have the particularity of automatically learning image features from the inputs, requiring a large number of training examples. Obtaining such information can be challenging in the medical field due to confidentiality constraints, financial limitations, and the time required for expert annotations [25]. Thus, traditional methods using handcrafted features remain helpful in dealing with limited samples of medical imaging data.

2.1 Our facial palsy analysis

When working with images of facial palsy patients, the asymmetry of the face must be considered. Usually, the main difference is observed between the left and right sides of the face, but others can be observed depending on the disease or the affected nerves. Our proposed analysis mainly compares and quantifies the differences between the left and right sides by analyzing the location and position of the face organs (eyebrows, eyes, nose, cheeks, and mouth). Initial tests concluded that the regions of the eyebrows, eyes, and mouth provide the most meaningful information for this challenge. This conclusion is similar to the findings of other authors [5, 16, 18, 26].

Notice that some approaches require a set of images from the same subject performing specific gestures to operate (*e.g.,* [5, 6]). Although we firmly believe that the facial palsy assessment should not be conditioned to specific facial movements. Similar to [41], when a face image is loaded into our system, the facial gesture performed by the person does not need to be identified; the system's output is a label obtained after an objective evaluation.

2.1.1 Facial symmetry features

The methodology to extract our facial features is shown in Figure 2.1. In there, four main steps are followed once an image is available and the face is detected: (1) a landmark prediction using a face model trained with palsy patients' imagery, (2) a correction of the head's tilt angle using a transformation matrix, (3) using the transformed key points a calculation of measures (values of distances and averages), and (4) the extraction of facial features to create a vector descriptor.
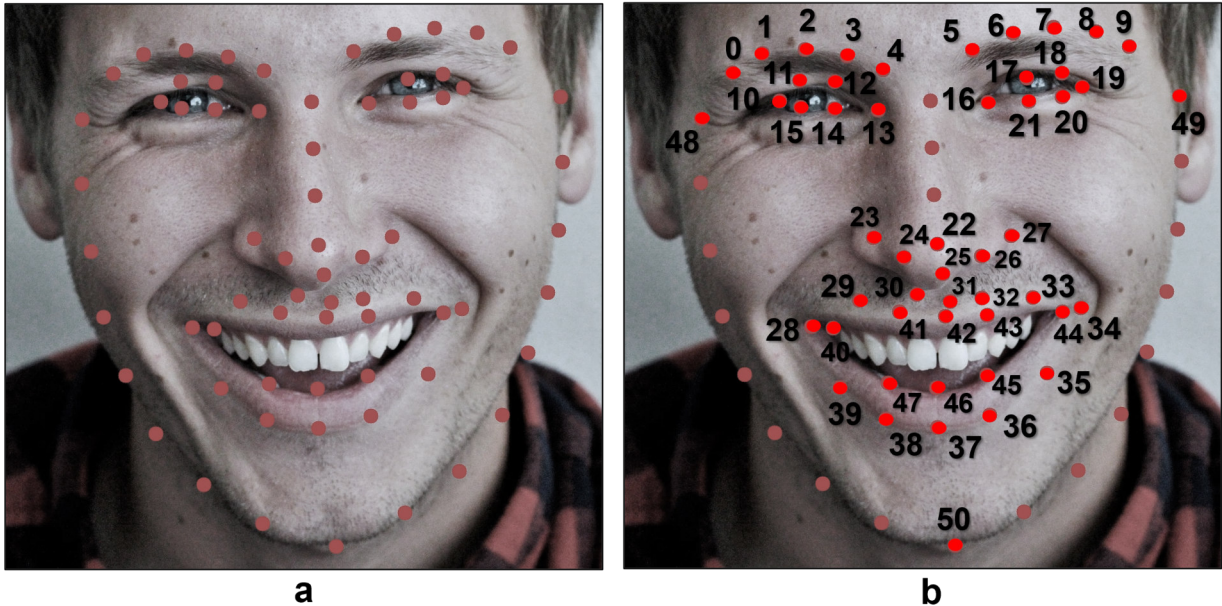
Figure 2.2: (**a**) The predicted 68 key points and (**b**) The 51 renumbered key points of our interest inspired by the model proposed by Matthews and Baker [30].

Landmark prediction

Our methodology uses the widely-known face detector implemented in the open-source *dlib C++ Library*. Once the face is detected within the image, the facial landmarks are extracted using the shape predictor proposed by Guarin *et al.* and introduced in [15]. The complete extraction process, as suggested by the authors, goes as follows:

1. Transform the face image to gray levels.

2. (Optional) Resize the transformed image according to a scale factor (sf) of $sf = \frac{W}{nW}$, having $nW = 200$ and $nH = \frac{H}{sf}$. Where W and H refer to the width and height of the face image.

3. Detect the face rectangle on the smaller image.

4. (Optional) Re-scale the detected face rectangle to its original size using sf.

5. Extract the facial landmarks on the transformed image.

6. Store the predicted information for future processing.

Note that the shape predictor is a 68-point model, but only 51 points are of interest in this work. As sketched in Figure 2.2, the 51 points are renumbered to ease the further calculation of attributes.

Tilt angle correction

It is known that the head's tilt angle can influence the accuracy of the facial symmetry quantification [16]. Thus, a tilt correction is performed using known points and a transformation matrix before computing any symmetry measure. The process uses the eyes' landmarks, particularly the left corner of the left eye and the right corner of the right eye (points 10 and 19 as seen in Figure 2.2b) are set as known points. The complete process goes as follows:

1. Set as input data the eye-corner points: **P10** for the left eye's left corner and **P19** for the right eye's right corner.

2. Set the destination data on such points, ensuring they are horizontally aligned.

3. Calculate a transformation matrix $\mathbf{T_f}$ using the eye-corner points and the similarity transform approach.

4. (Optional) Transform the input image using the matrix $\mathbf{T_f}$.

5. Rotate the predicted landmark $\mathbf{P_i}$ using $\mathbf{T_f}$ and a multiplication of matrices to obtain $\mathbf{P'_i}$, for example:

$$\begin{pmatrix} P'_x \\ P'_y \end{pmatrix} = \begin{bmatrix} T_{f(1,1)} & T_{f(1,2)} & T_{f(1,3)} \\ T_{f(2,1)} & T_{f(2,2)} & T_{f(2,3)} \end{bmatrix} \begin{bmatrix} P_x \\ P_y \\ 1 \end{bmatrix} \tag{2.1}$$

where $i = \{0, \ldots, 50\}$, meaning that $\mathbf{P_i}$ represent each of the 51 key points.

6. Store the rotated data.

Measure calculation

A total of 32 measures (28 distances and four average values) are calculated using the key points, and they are described in Figure 2.3. Ostrofsky *et al.* in [36] evaluated a set of objective measures from face photographs to characterize the human face, their work inspired our distances from A to K. Notice that the authors did not target the analysis of facial palsy, but their measures appeared to be an excellent reference. In our study, distances A to Q were proposed to detect any level of asymmetry among face sides (left vs. right). At the same time, measures P to X were designed to characterize the facial movement executed by the patient. Providing robustness against facial expressions allows us to obtain a result from a single image rather than a set of facial gestures.

The widely-known Euclidean distance was employed to measure the distance between two points, according to Equation (2.2). Most of these distances are computed in pairs; the left and right sides will be compared to extract a specific feature. The other lengths are used to normalize the range of the measures between 0 and 1. The average value refers to the arithmetic mean of a group of values, computed according to Equation (2.3).

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

(a)

Distance	Description
L, M	Height average of the 5 points within each eyebrow
N_l, N_r, O_l, O_r	Distance between the upper and lower eyelids (left and right) of both eyes
N	Average height of the eye $(N_l + N_r)/2$
O	Average height of the eye $(O_l + O_r)/2$
P_l, P_u, Q_l, Q_u	Distance between the upper and lower lip of both sides of the mouth

(b)

Distance	Description
R, S	Distance from the fourth eyebrow point to the middle of the jaw point (left and right)
T, U	Distance from the third eyebrow point to the middle of the jaw point (left and right)
V_l, V_r	Distance from the mouth corner to the outer middle point of the mouth (left and right)
W	Distance from the mouth corners.
X	Distance from the middle nose tip to the outer middle point of the mouth

(c)

Figure 2.3: Facial measures to compute spatial relations between facial landmarks. (**a**) Measures A to K; (**b**) measures L to Q; and (**c**) measures R to X.

$$d(\mathbf{Pa}, \mathbf{Pb}) = \sqrt{(Pa_x - Pb_x)^2 + (Pa_y - Pb_y)^2} \tag{2.2}$$

$$avrg = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.3}$$

Here are some additional remarks regarding our facial measures:

- Distances A, C, W, and X are normalization parameters.

- Values L and M look for unevenness in the height of the eyebrows.

- Distances R, S, T, and U search for the larger side (if any) of the face from the eyebrow to the mouth.

- Distances B_l, B_r, D, E, N_l, N_r, O_l and O_r inspect for any deformity in the eyes.

- Distances F, G, H, and I search for the larger side (if any) of the face from the eye to the mouth.

- Distances J and K look for the larger side (if any) of the face from the nose to the mouth.

- Distances P_l, P_u, Q_l, Q_u, V_l and V_r inspect for any deformity in the mouth.

Feature extraction

Our facial symmetry features were computed using the 32 measures mentioned above. In the beginning, 33 features were proposed to analyze facial palsy, but looking to reduce our feature space, several experiments were conducted to evaluate the worth of each attribute. After many tests, only 29 were found helpful in detecting a level of asymmetry within the face and performing gesture recognition tasks. Our facial symmetry features are extracted as follows: using Equation (2.4) to obtain the angle between two points, Equation (2.5) to calculate the slope between points, Equation (2.2) to compute the Euclidean distance, and Equation (2.6) to calculate the perimeter of a closed shape.

$$\angle(\mathbf{Pa}, \mathbf{Pb}) = \arctan 2(\triangle x, \triangle y) \times 180/\pi \tag{2.4}$$

where $\triangle x = Pa_x - Pb_x$ and $\triangle y = Pa_y - Pb_y$.

$$m(\mathbf{Pa}, \mathbf{Pb}) = \left| \frac{Pa_y - Pb_y}{Pa_x - Pb_x} \right| \tag{2.5}$$

$$\bar{\mathrm{S}}(\mathbf{P_s}, \ldots, \mathbf{P_l}) = \sum_{x=s}^{l-1} d(\mathbf{P_x}, \mathbf{P_{x+1}}) + d(\mathbf{P_s}, \mathbf{P_l}) \tag{2.6}$$

where $\overline{\text{S}}$ is a closed shape, $\mathbf{P_s}$ is the start point, and $\mathbf{P_l}$ is the last one and represents the endpoint within the shape.

The 29 facial symmetry features are described in Table 2.1, refer to Figure 2.3 if further detail is required. They are numbered from $f0$ to $f28$ and grouped by facial regions: eyebrows, eyes, mouth, nose and combined. The formula to compute each is also depicted. It is easy to observe that most of the distances provide ratios in the $[0, 1]$ range, where 0 means somewhat asymmetric (palsy face), and 1 means closer to symmetric (healthy face).

- Features $f0$ to $f2$, $f4$ to $f6$, $f7$, $f14$, $f22$, and $f23$: they are designed to measure the inclination between two key points. It is expected for a healthy face to have a slope or angle close to 0 (to be in a horizontal position), except for feature $f23$, which is expected to be vertical on a healthy subject.

- Features $f3$, $f8$ to $f13$, $f15$ to $f21$, and $f24$ to $f26$: they are analogous ratios between the left and right sides. Those ratios are compared, and the maximum value is selected as an attribute.

- Features $f27$ and $f28$: they reflect the asymmetry of the face. Smaller values relate to a healthy subject, and bigger values to a palsy patient.

An additional description is provided on Table 2.1. Notice that in *f21*, W is the distance shown in Figure 2.3c, and the perimeter values W_l and W_r are computed as $W_l = \overline{\text{S}}(\mathbf{P28}, \mathbf{P29}, \mathbf{P30}, \mathbf{P31}, \mathbf{P37}, \mathbf{P38}, \mathbf{P39})$ and $W_r = \overline{\text{S}}(\mathbf{P31}, \mathbf{P32}, \mathbf{P33}, \mathbf{P34}, \mathbf{P35}, \mathbf{P36}, \mathbf{P37})$.

2.1.2 Machine learning algorithms

Machine learning (ML) is helpful in complex problems where humans cannot explain a solution, or there is no expertise on how to solve a specific problem [2]. ML is a set of techniques for automatically extracting knowledge from data [7]. In other words, it creates a model using exemplary data from the problem. The model is iteratively refined using the sample data to tune its parameters. This iterative process is called the learning process.

Machine learning algorithms learn from sample data. The quality and the quantity of the samples have a significant impact during the learning process. There are two types of learning algorithms: (1) supervised learning, where the labels (output) of the examples (input) are known, and (2) unsupervised learning, where the labels of the data are unknown. The first aims to create a model that associates an input with its correct label, and the second seeks associations among the data to find a model representing these relationships.

In this project, we propose to create models based on supervised machine learning algorithms and using data from palsy patients; restrictions due to the limited amount of data will be discussed later. Our proposed classifiers were designed using the Waikato Environment for Knowledge Analysis (Weka) suite. Weka is a platform for general-purpose machine learning tasks, such as classification, prediction, detection, etc. Weka is a graphical user interface that contains the full pipeline for managing data sets, training, and

Table 2.1: Our proposed facial symmetry features introduced in [38].

No.	Facial region	Formula	Description
f0	Eyebrows	$\lvert\angle(P0, P9)\rvert$	Vertical position of the left and right sides.
f1	Eyebrows	$\lvert\angle(P2, P7)\rvert$	Vertical position of the left and right sides.
f2	Eyebrows	$\lvert\angle(P4, P5)\rvert$	Vertical position of the left and right sides.
f3	Eyebrows	$\max(L/M, M/L)$	Ratio of the eyebrows' elevation.
f4	Eyebrows	$m(P0, P9)$	Vertical position of the eyebrows.
f5	Eyebrows	$m(P2, P7)$	Vertical position of the eyebrows.
f6	Eyebrows	$m(P4, P5)$	Vertical position of the eyebrows.
f7	Eyes	$\lvert\angle(P10, P19)\rvert$	Vertical position of the eyes' corner.
f8	Eyes	$\max(B_l/B_r, B_r/B_l)$	Ratio of the eyes' width.
f9	Eyes	$\max(D/E, E/D)$	Ratio of the eye to head-side distance.
f10	Eyes	$\max(H/I, I/H)$	Ratio of the eye to nose distance.
f11	Eyes	$\max(N/O, O/N)$	Ratio of the eyes' average opening.
f12	Eyes	$\max(N_l/O_r, O_r/N_l)$	Ratio of the eyes' outer opening.
f13	Eyes	$\max(N_r/O_l, O_l/N_r)$	Ratio of the eyes' inner opening.
f14	Mouth	$\lvert\angle(P28, P34)\rvert$	Vertical position of the corners.
f15	Mouth	$\max(F/G, G/F)$	Ratio of the eye to mouth distance.
f16	Mouth	$\max(P_l/Q_l, Q_l/P_l)$	Ratio of the mouth's outer opening.
f17	Mouth	$\max(P_u/Q_u, Q_u/P_u)$	Ratio of the mouth's inner opening.
f18	Mouth	$\max(V_l/A, V_r/A)$	Ratio of the mouth's half-width divided by the head's width.
f19	Mouth	$\max(P_l/W, Q_l/W)$	Ratio of the mouth's outer height divided by its width.
f20	Mouth	$\max(P_u/W, Q_u/W)$	Ratio of the mouth's inner height divided by its width.
f21	Mouth	$\max(W_l/W, W_r/W)$	Ratio of the mouth's half perimeter divided by its width.
f22	Nose	$\lvert\angle(P23, P27)\rvert$	Vertical position of the nose.
f23	Combined	$\lvert\angle(P22, P37)\rvert$	Vertical position of the nose and mouth.
f24	Combined	$\max(J/K, K/J)$	Ratio of the nose-to-mouth distance.
f25	Combined	$\max(T/A, U/A)$	Ratio of the eyebrows-to-mouth distance divided by the head's width.
f26	Combined	$\max(R/A, S/A)$	Ratio of the eyebrows-to-mouth distance divided by the head's width.
f27	Combined	C/A	Ratio of the mouth-to-chin distance divided by the head's width.
f28	Combined	X/A	Ratio of the nose-to-mouth distance divided by the head's width.

In *f21*, $W_l = \overline{\mathrm{S}}(P28, P29, P30, P31, P37, P38, P39)$ and $W_r = \overline{\mathrm{S}}(P31, P32, P33, P34, P35, P36, P37)$.

testing according to multiple machine learning models; therefore, it was appropriate for the analysis performed in this research. Weka operates using Attribute-Relation File Format (arff) files, which are text files that describe a list of samples sharing a set of features and labels. Hints on creating arff files can be found at [3]; more information concerning the Weka suite is available at [50].

Initially, we wanted to show that handcrafted features and simple neural networks are still effective in dealing with the automated analysis of facial paralysis. Then, we designed a classifier based on the multi-layer perceptron strategy and used the Weka function known as the Multilayer Perceptron (MLP). This MLP model requires a few parameters to learn, as described in Table 2.2. Some of those can be easily set up in the Weka environment, for example, the learning rate (L), momentum (M), training time (N), number of neurons in the hidden layers (H), and seed (S). During the training phase of the MLP strategy, the learning rate and the momentum are required to update the weights of the connections between network nodes. As indicated by its name, the training time refers to the number of iterations (epochs) to train through, and the value in H refers to the number of neurons in each layer within the network. Finally, the value of S refers to a number employed to randomly initialize the network's weights. The MLP strategy uses backpropagation to learn and classify instances, and the network nodes are all sigmoid.

All the proposed MLP models are 3-layer networks in this project. The input layer consists of 29 neurons (one for each feature) and some neurons in the output layer that correspond to the number of the required labels (2 for binary classification and 3 for a multi-class classification). The number of neurons in the hidden layer will be specified for each experiment using the parameter H.

Lastly, we wished to show that our methodology to design classifiers could be applied to other models. Then, three additional classifiers were configured based on the support vector machine (SVM), k-nearest-neighbor (KNN), and multinomial logistic regression (MNLR) methods. Depending on the implementation, each classifier requires specific parameters to operate, and the parameters are optimized to reach the best performance for the particular task. Again in Table 2.2, the function known as the LibSVM is an implementation of an SVM. The radial basis function (RBF) kernel was used in all experiments, and the cost and gamma parameters were optimized. The function IBk implements the KNN strategy using one neighbor; the distance function was optimized for each experiment. The function Logistic is an implementation of the MNLR method; in our design, only the Ridge parameter was refined.

The process of designing our classifiers consisted of having a vector descriptor (input), computed according to Section 2.1.1, and its associated label (output) to perform a learning process for each algorithm. Many arff files were created containing such input and output data to facilitate our learning environment. After the files are loaded into Weka and the classifiers are adequately configured, the training and evaluation process occur. Several tests were conducted to reach the highest performance for each task.

Table 2.2: Required parameters to operate the classifier according to Weka [50].

Method	Parameters	Weka function
MLP	Learning rate (L), momentum (M), training time (N), number of neurons in the hidden layers (H), and seed (S)	MultilayerPerceptron
SVM	Cost (C), gamma (G), kernel type	LibSVM
KNN	Number of neighbors (KNN) and distance function (A)	IBk
MNLR	Ridge (R)	Logistic

2.2 Available image databases

We already mentioned that some methodologies aim to detect facial paralysis in a photograph. However, collaboration among the research community has been complex due to the unavailability of public datasets, mainly because of patient privacy (patients prefer not to share their biometrical data). This situation motivated us to perform our analysis in three databases containing images from palsy patients that recently became publicly available. Some remarks on those databases are given here.

MEEI database

The Massachusetts Eye and Ear Infirmary (MEEI) database is an open-source set of facial photographs and videos representing the entire spectrum of flaccid and non-flaccid facial palsy collected by Greene *et al.* and introduced in [13]. The database was launched to serve as a resource for facial paralysis research and education; some works that have employed the MEEI database are [15, 29, 32, 38, 39].

The MEEI dataset consists of 480 high-resolution images from 60 subjects: ten healthy participants, 25 patients with flaccid palsy, and 25 patients with non-flaccid paralysis. Each subject performed eight facial movements: (1) at rest, (2) eyebrow elevation, (3) light effort eye closure, (4) full effort closure, (5) light effort smile, (6) full effort smile, (7) pucker, and (8) lip depression. An example of these photographs is shown in Figure 2.4. The MEEI database is also annotated according to the eFace, Sunnybrook, and HB grading scores, meaning that six levels of paralysis are provided accordingly: normal, near-normal, mild, moderate, severe, and complete.

In previous sections, we mentioned that three sub-projects were developed in this research; the MEEI database was employed in two: to detect facial palsy and recognize facial gestures. The first was achieved using the annotated data provided by the owners; for the second task, we extracted the gesture information from the eight facial expressions performed by each participant.
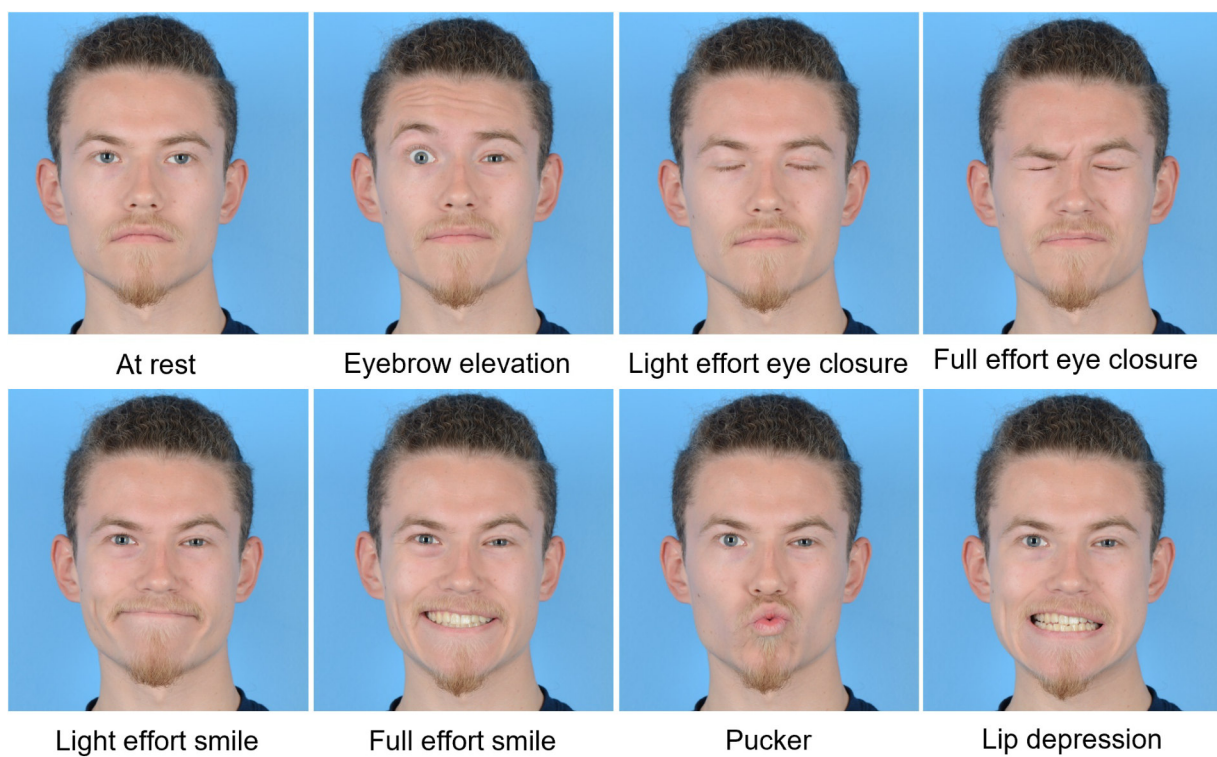
Figure 2.4: Example of a set of photographs extracted from the MEEI dataset [13].

Toronto NeuroFace

The Toronto NeuroFace (TNF) is a publicly available database that aims to assess neurological disorders. It was collected by Bandini *et al.* and introduced in [4]. Similarly to the MEEI dataset, the Toronto NeuroFace dataset was released to foster the development of novel and robust approaches for face alignment and oro-facial assessments that can be used to track and analyze facial movements in clinical populations suffering from amyotrophic lateral sclerosis and other neurological diseases. Authors in [4] analyzed the importance of using algorithms trained with data from the target population to improve facial landmarks' localization and accuracy in face alignment. The TNF dataset consists of 261 videos, clinical scores per video, and more than 3300 annotated frames of faces from individuals performing oro-facial tasks typical of the clinical assessment. Due to privacy constraints, it is not possible to show an image from this database.

YFP database

The YouTube Facial Palsy (YFP) database is an image collection provided by Hsu *et al.* and introduced in [19]. The YFP database is a compilation of 32 videos from 21 patients obtained from YouTube; 10 additional patients were included in a second release. The patient talks to the camera in each video, and the facial expression variation across time is recorded. Each video was converted into an image sequence at 6 FPS, yielding almost 3,000 images. Independent clinicians labeled the palsy regions in each frame; the junction of the cropped areas is considered the ground truth. The authors also provided additional labels to classify the intensity exhibited in each palsy region. We understand that the authors in [19] determined these intensity labels; they did not declare that the intensity label was approved by a clinician, which could lead to discrepancies in the classification results among methodologies. An example of these images is shown in Figure 1.1.

CK+

The Extended Cohn-Kanade (CK+) database distribution is a well-known dataset in the research community to prototype and benchmark systems for the automated detection of facial expression [28]. The CK+ database collects 593 sequences across 123 subjects, close to 10,800 images, and all sequences go from a neutral face to a peak expression. An example of some pictures is shown in Figure 2.5. The CK+ database is included in this project to make our methodology robust against expression variation, as suggested in the literature [19, 26]. In other words, as explained later, the palsy samples came from the YFP database, and the healthy subjects were from the CK+.

Figure 2.5: Example of some images extracted from the CK+ dataset [28], ©Jeffrey Cohn.

2.2.1 Additional remarks

To the best of our knowledge, neither the MEEI database nor the TNF dataset was employed to detect facial paralysis in frontal face photography before our work in [38]; or to recognize facial gestures before our analysis in [39]. On the other hand, the YFP database has been used to diagnose facial palsy in [19, 26]. It is worth noticing that those databases intend to facilitate information for the development of clinical applications; however, they are not equivalent in terms of image quality, lighting, pose conditions, and the tasks performed by the participants. In other words, the three datasets are not directly comparable to our classification problem, but they were helpful in the design process.

To conclude this subsection, the databases above were employed in our project as follows: (1) the MEEI, TNF, YFP, and CK+ databases were used to detect facial palsy; (2) the YFP and CK+ databases were used to grade the lesion's severity; and (3) the MEEI database was used to recognize facial gestures in palsy patients. Such tasks will be described in detail in the following chapters.

CHAPTER 3

DETECTION OF FACIAL PARALYSIS

Detecting facial paralysis in this research is achieved by observing the symmetry of the subject's face through a photograph. It has already been stated that a healthy face (*i.e.,* not affected by any paralysis) has quite an identical left and right sides when showing a neutral facial expression. On the other hand, the face of a palsy patient is expected to be more asymmetrical, which allows us to characterize the faces of healthy and palsy subjects. Such characterization refers to obtaining a threshold value using machine learning algorithms to determine to what extent the facial asymmetry is expected and where it is considered an unhealthy condition.

In this palsy detection task, we aim to detect facial paralysis as a binary classification problem. That is, to identify if a person is healthy or a palsy patient. Our methodology evaluates the input image independently of the facial movement executed by the subject. Other approaches use a set of facial gestures to compute symmetry variations between expressions and detect facial paralysis. Our system measures predicted facial landmarks using basic mathematical operations that keep the implementation uncomplicated but effective. Our facial measures aim to see the asymmetry between the two sides of the face and characterize facial gestures so that the learning algorithm can relate each asymmetry level with the expression found in the image. Our measures extract information from the eyebrows, eyes, nose, and mouth. Our classifiers provide a label as an output.

In the grading of facial paralysis, we define the task as a multi-class classification problem. Mainly, a three-level classifier was designed using different machine learning algorithms. We explore the use of regional information to achieve such classification tasks. In the end, we show that the assessment of facial paralysis is possible with partial occlusions of the face if the analysis is executed on some areas of the face. Some of those areas are the eyebrows, the eyes, the nose, and the mouth, as shown in Figure 3.1. We refer to the features computed from those facial areas as regional information. Generally speaking, our experiments are carried out on the entire face, but additional evaluations were performed on the eyes and the mouth region. By the entire face, we refer to the use of the 29 proposed
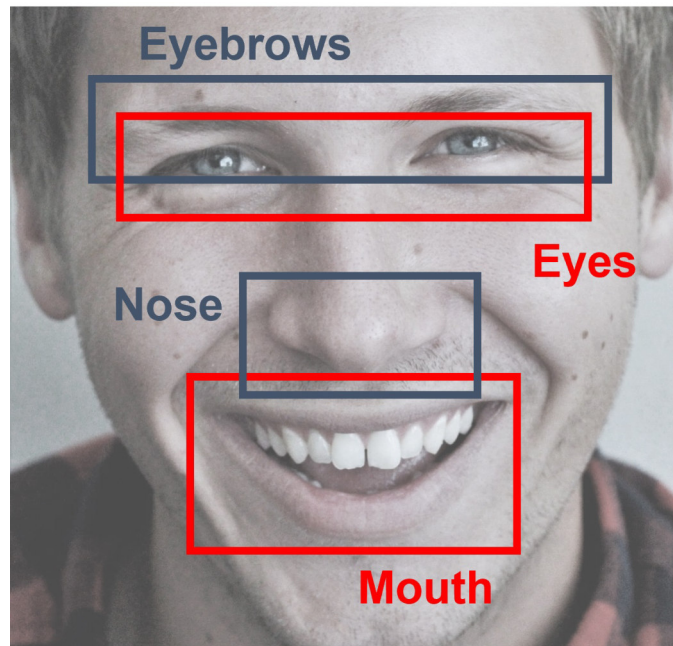
Figure 3.1: Example of a face image divided into four facial regions.

symmetry features. In the case of evaluating using the eyes, only 19 features are considered, and 15 features for the mouth region. Further details on the performed experiments to evaluate our framework will be provided.

3.1 Training of our classifier

The framework of our facial palsy detection was introduced in Section 2.1, and it is shown in Figure 3.2. Firstly, the face must be detected in the input image, followed by facial landmarks extraction using a shape predictor. Our proposed 29 facial symmetry features are subsequently computed using these key points, as explained in Section 2.1.1. Finally, such features are fed to a classifier that will provide an output depending on the system's goal.

Section 2.1.2 mentioned that the Weka platform was employed to design a set of classifiers. In our first attempt to detect facial palsy, we trained and evaluated a classifier based on the multi-layer perceptron. Later, three additional methods were considered for comparison purposes and to validate our features using different strategies. We refer to the support vector machine (SVM), k-nearest-neighbor (KNN), and multinomial logistic regression (MNLR) methods. Some parameters needed to operate those classifiers using the Weka platform were introduced in Table 2.2. We have refined some of those parameter values to reach the highest performance; they are mentioned in Table 3.1. The optimization process was also performed using the Weka suite and its options, for example:
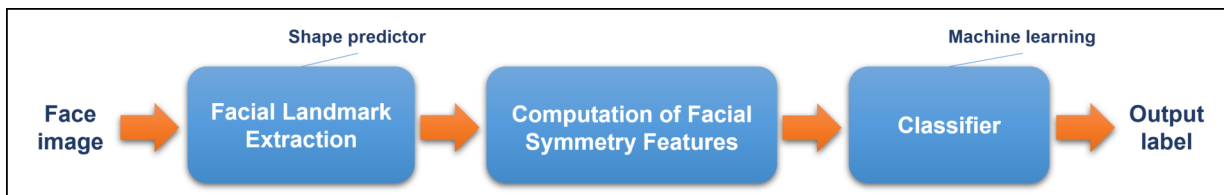
Figure 3.2: Framework of the proposed facial palsy detection system.

- CVParameterSelection: it is a function that performs a parameter selection for any classifier using cross-validation. We use it to choose the best L and M values in a range from 0.05 to 0.95 with increments of 0.05; the best H value in a range between 40 to 70 with increments of 1; and the best S value in a range from 0 to 99 with increments of 1. More information is available at [48].

- Gridsearch: it is a function that performs an exhaustive search of parameter pairs for a classifier and chooses the best pair found for the actual prediction. We use it to find the best C and G using accuracy as evaluation criteria. More information is available at [49].

- Weka experimenter: it is an application to create experiments using different datasets or algorithms simultaneously. We used it to locate which distance measure to use; we tested with the Euclidean, Manhattan, and Chebyshev distances. We also use it to repeat each experiment ten times because it easily stores the results and calculates performance metrics. The complete procedure is explained in [9].

Once the parameters were refined, several evaluations were executed.

Widely known evaluation metrics were computed to measure the performance of our classifiers. These are accuracy, recall (or true positive rate), specificity, precision, F1 score, true negative rate, false negative rate, and false positive rate. They are calculated according to Equations (3.1)–(3.8), respectively.

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.1}$$

$$Rec = \frac{TP}{TP + FN} \tag{3.2}$$

$$Spec = \frac{TN}{TN + FP} \tag{3.3}$$

$$Prec = \frac{TP}{TP + FP} \tag{3.4}$$

$$F1s = \frac{2 \times TP}{2 \times TP + FP + FN} \tag{3.5}$$

$$TNR = \frac{TN}{TN + FP} \tag{3.6}$$

$$FNR = \frac{FN}{FN + TP} \tag{3.7}$$

$$FPR = \frac{FP}{FP + TN} \tag{3.8}$$

where TP is for true positives, TN for true negatives, FP for false positives, and FN for false negatives.

Table 3.1: Optimized parameter to reach the highest performance.

Method	Parameters	Weka option
MLP	L, M, H and S	CVParameterSelection
SVM	C and G	GridSearch
KNN	A	Weka experimenter

3.2 Classification of facial palsy

Two main experiments were performed to show the capacity of our methodology to detect facial paralysis in an image. They are mentioned as (1) facial palsy detection and (2) palsy region detection in the following subsections. Only the MLP strategy is discussed in the first evaluation, and the other classifying approaches are included in the second. In the experiments, the employed databases significantly differ in the characteristics of the images (*e.g.,* the lighting conditions and pose). However, the methodology to train and evaluate is the same as described in Figure 3.2, and excellent results were achieved.

3.2.1 Facial palsy detection

The first evaluation was performed to show that our methodology can detect facial palsy in a photograph. Two binary classifiers based on the MLP strategy were trained using images from the MEEI and TNF databases; both were introduced in Section 2.2. The parameter's configuration was set as $L = 0.2045$, $M = 0.1909$, $N = 500$, $S = 0$, and $H = 56$. The 10-fold cross-validation technique, a statistical method used to estimate the model's performance on new data, was employed to train and evaluate.

While testing using the MEEI database, it was observed that the MLP function in Weka does not perform sequential learning, resulting in a negative impact on the performance because the available dataset is unbalanced. There are fewer healthy samples than unhealthy ones (80 vs. 400). The healthy set was replicated three times with a sample augmentation process to overcome this situation. Similar to the approach suggested in [16, 22], the

Table 3.2: Cross-validation results for tests on the MEEI database.

n-fold	TN	TP	FN	FP	Accuracy [%]
1	22	32	8	2	84.37
2	24	37	3	0	95.31
3	24	33	7	0	89.06
4	24	38	2	0	96.87
5	24	40	0	0	100
6	23	38	2	1	95.31
7	24	38	2	0	96.87
8	22	38	2	2	93.75
9	24	35	5	0	92.19
10	24	38	2	0	96.87
Average	-	-	-	-	**94.06**

Table 3.3: Resulting confusion matrix of the test on the original MEEI database.

Healthy = 0	Palsy = 1	Correct label
80	0	healthy
1	399	palsy

healthy images were rotated in two opposite directions, increasing the available data and verifying that our algorithm is invariant to rotation. In previous experiments, it was observed that increasing the number of healthy instances by three times is enough to learn and discriminate this class without over-fitting. In the end, the training of the MLP classifier was executed with 640 samples (240 healthy and 400 unhealthy instances).

In Table 3.2, the model's performance for the testing part of the data for each fold is given. There, TN stands for true negative (*i.e.,* healthy samples), TP stands for true positive (*i.e.,* palsy samples), and false negative (FN) and false positive (FP) are incorrectly classified samples. Here, the worst performance is 84.37% for the 1st fold, and the best is 100% for the 5th fold. The average performance of this methodology is 94.06%, as shown in Table 3.2. After evaluating our MLP classifier with the original 480 samples, there is a 99.79% correct classification of the MEEI database, yielding a sensitivity of 99.75% and a specificity of 100%. The confusion matrix of the evaluation using the original MEEI database is given in Table 3.3.

It was already mentioned that the Toronto NeuroFace (TNF) is a database collected for clinical assessment. To the best of our understanding and the information provided by the authors, the subset called *Stroke* can be used to evaluate our methodology. There are eight subjects with facial asymmetry and 3 participants without it; in total, 817 asymmetrical samples and 219 symmetrical ones. It is essential to notice that the TNF images refer to frames extracted from the subjects' videos performing oro-facial tasks. Following the proposed methodology, a 10-fold cross-validation technique was employed to train another

Table 3.4: Cross-validation results for tests on the TNF database.

n-fold	TN	TP	FN	FP	Accuracy [%]
1	66	78	4	0	97.30
2	65	78	4	1	96.62
3	66	80	2	0	98.65
4	65	80	2	1	97.97
5	64	75	6	2	94.56
6	65	79	2	1	97.96
7	66	79	2	0	98.64
8	65	77	5	0	96.60
9	65	78	4	0	97.28
10	64	78	4	1	96.60
Average	-	-	-	-	**97.22**

Table 3.5: Resulting confusion matrix of the test on the original TNF database.

Healthy = 0	Palsy = 1	Correct label
218	1	healthy
14	803	palsy

MLP classifier using data computed from the TNF database.

In Table 3.4, the model's performance on the testing part of the data for each fold is given. Here, the worst performance was 94.56% for the 5th fold, and the best was 98.65% for the 3rd fold. The average performance of this methodology for the second dataset was 97.22%, as shown in Table 3.4. A 98.55% correct classification was found after evaluating the MLP classifier with the original 1036 samples. The confusion matrix of the test on the original TNF database is described in Table 3.5, which yields a sensitivity of 98.29% and a specificity of 99.54%, proving that our methodology to design a binary classifier to detect facial paralysis can be extended to other databases of face photographs with an outstanding performance.

3.2.2 Palsy region detection

In the literature, it is said that deformed regions, or simply palsy regions, are caused by facial paralysis on a patient's face; and those are the target regions to locate. Here, we called it palsy region detection because our algorithms analyze two specific facial regions: the eyes and the mouth. The features related to the eyes' regional information are the features named Eyebrows, Eyes, Nose, and Combined (only *f23* and *f25–f27*) in Table 2.1. Similarly, the mouth's regional information are all features, except those named as Eyebrows and Eyes features.

The following experiments evaluate our methodology using the above regional informa-

Table 3.6: Detection of palsy regions: classifiers' configuration.

Method	Parameters
MLP	$L = 0.2045$, $M = 0.1909$, $H = 59$, $N = 5000$, $S = 0$
SVM	$C = 1000$, $G = \{0.1, 0.001, 0.01\}$ *, kernel type = RBF
KNN	KNN $= 1$, $A =$ Euclidean distance
MNLR	$R = 1 \times 10^{-8}$

* Refers to the gamma value for the face, eyes, and mouth evaluation, respectively.

Table 3.7: Results of the detection of palsy regions on the criteria of accuracy.

Classifier	Face	Eyes	Mouth
MLP	$95.03 \pm 1.69\%$	$92.86 \pm 1.99\%$	$92.98 \pm 2.13\%$
SVM	$95.61 \pm 1.40\%$	$93.42 \pm 1.84\%$	$90.93 \pm 2.49\%$
KNN	$92.34 \pm 1.80\%$	$89.33 \pm 2.24\%$	$91.28 \pm 2.21\%$
MNLR	$94.24 \pm 1.60\%$	$92.16 \pm 1.99\%$	$91.07 \pm 2.35\%$

tion, but our initial proposal is to inspect the whole face using our symmetry features. Three tests were performed: (1) the detection of palsy using 29 features, (2) the detection of palsy using the eyes information, and (3) the detection of palsy using the mouth information. Four classifiers for each test were evaluated to compare their performance in this classification task. The 5-fold cross-validation technique was used during the training process. Each validation was repeated ten times to observe the behavior of the classifiers and provide information regarding their repeatability. The average performance will be given for each test.

The data set comprises 19 palsy patients and 19 healthy subjects extracted from the YFP database and the CK+ collection. Both datasets were introduced in Section 2.2. The palsy patients are subjects 1, 5, 6, 7, 11, 12, 13, 14, 15, 19, 20, 21, 23, 24, 25, 28, 29, 30, and 31 from the YFP database. Patients with less than 20 images and patients with facial occlusions were excluded. The healthy subjects belong to the S022, S026, S028, S034, S042, S046, S050, S054, S057, S102, S105, S124, S130, S131, S132, S133, S134, S135, and S136 folders in the CK+ collection.

The data set for this evaluation comprises 20 images from each of the 38 participants (760 photos in total); this arrangement is expected to have the same amount of healthy and palsy samples, making it a balanced data set for the experiment. Notice that the healthy subjects are labeled as class 0 and the palsy subjects as class 1. For this experiment, the classifiers were configured as described in Table 3.6. The average performance is shown in Table 3.7 for each of the three tests performed in this palsy region detection task.

Great results are obtained for the MLP and SVM classifiers, 95.03% and 95.61%, respectively. Few samples were required during the training phase compared to other approaches that required thousands of images to output a label. Good results are also reached using only regional features, in the eye and the mouth, 93.42% and 92.98%, respectively.
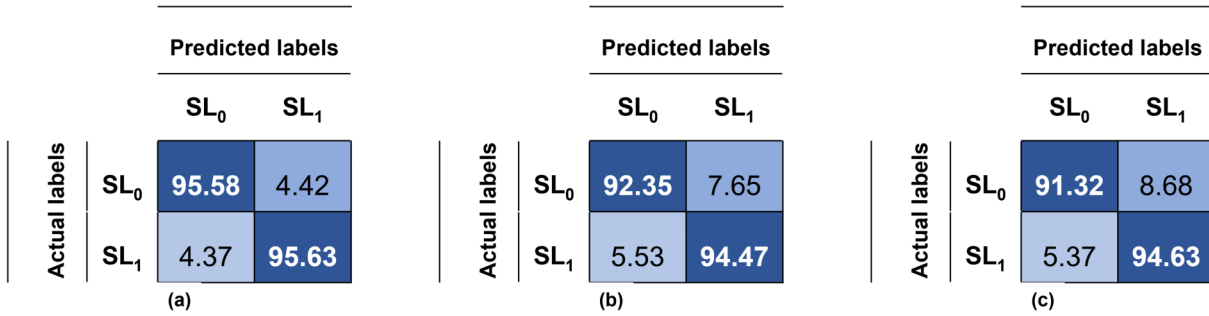
Figure 3.3: Confusion matrix for palsy detection using: (**a**) the entire face, (**b**) the eyes' information and (**c**) the mouth's information.

Table 3.8: Performance results for the detection of palsy regions.

Region	Classifier	TNR	FNR	TPR	FPR
Face	SVM	$95.59 \pm 2.28\%$	$4.37 \pm 2.05\%$	$95.63 \pm 2.05\%$	$4.41 \pm 2.28\%$
Eyes	SVM	$92.35 \pm 3.11\%$	$5.53 \pm 2.81\%$	$94.47 \pm 2.81\%$	$7.65 \pm 3.11\%$
Mouth	MLP	$91.32 \pm 3.51\%$	$5.37 \pm 2.85\%$	$94.63 \pm 2.85\%$	$8.68 \pm 3.51\%$

Still, the entire face analysis is better than focusing on a single region when discriminating between healthy and palsy patients.

The confusion matrix of this experiment using the entire face and SVM is depicted in Figure 3.3a, the system's average accuracy is 95.61%, recall 95.63%, precision 95.61% and F1-score 95.62%. The confusion matrix for the test using the eyes' information with SVM is depicted in Figure 3.3b, the system's average accuracy is 93.42%, recall 94.47%, precision 92.55% and F1-score 93.50%. Finally, the confusion matrix for the experiment using the mouth's information with MLP is depicted in Figure 3.3c, the system's average accuracy is 92.98%, recall 94.63%, precision 91.64% and F1-score 93.11%.

Additional performance results for the three tests are provided in Table 3.8. For our methodology, the true negative rate (TNR) reflects the number of healthy subjects detected as regular participants, while the true positive rate (TPR) shows the number of palsy patients predicted as unhealthy subjects. Great results are achieved using the SVM classifier: 95.59% and 95.63% for the face, 92.35%, and 94.47% for the eyes. On the other hand, the false negative and false positive rates are expected to be as low as possible because both represent a wrong diagnosis. Good results are obtained for this metric: 4.37% and 4.41% for the face, 5.53% and 7.65% for the eyes. Similarly, good results are reached for the mouth using the MLP classifier, 91.32% and 94.63% of true detection rates, and 5.37% and 8.68% of false detection rates.

Table 3.9: Performance results of different facial palsy classification approaches.

Method	Database	Rec	Spec	Prec	F1s	Acc
Kim *et al.* [22]	*	80.00%	–	92.30%	–	88.90%
Hsu *et al.* [18]	YFP	88.00%	–	93.00%	–	–
Barbosa *et al.* [6]	*	98.12%	94.06%	–	–	–
Abayomi *et al.* [1]	YFP	91.85%	–	81.06%	85.91%	78.62%
Our methodology	MEEI	91.75%	97.92%	98.66%	95.08%	94.06%
Our methodology	TNF	95.72%	99.09%	99.24%	97.44%	97.22%
Our methodology	YFP	95.63%	95.59%	95.61%	95.62%	95.61%

*Tested on their private dataset, not available due to patient privacy. Note that test cases cannot be compared straightforwardly.

3.2.3 Findings and discussion

Our methodology detects facial asymmetry levels within an image independently of the gesture performed by the subject. At the same time, most methods in the literature compute asymmetry measures from a set of different facial gestures from the same subject. We can summarize three methodologies if we focus on the detection systems that compute facial landmarks at some point in their process. Those were introduced in Section 1.1.1. Such relation is shown in Table 3.9. We observe that a direct comparison of methods in this table is not feasible because of discrepancies in the goals, training configuration, and metrics. Nonetheless, we include the results of the works closer to ours. We also include the work of Abayomi *et al.* in [1] only because the YFP database was employed.

Table 3.9 shows the performance results of our methodology in three different databases. Great accuracy values are reached, more than 94%, showing that our algorithm can be extended to other databases with varying image characteristics. As explained earlier, a direct comparison with other methods is unfeasible; however, the performance suggests that our strategy to detect facial palsy exhibits a better performance or, at least, one similar to that of other approaches.

Although out of the scope of this work, those scores using either the eye or mouth information lead us to believe that we can use these features to detect facial palsy on images with partial occlusions. If face detection is achieved and landmarks are predicted adequately, an analysis to detect facial palsy might be possible. In other words, our analysis allows us to determine to what extent regional information is needed to diagnose the severity of the lesion with satisfactory results.

3.3 Grading of facial palsy

Two experiments were performed to assess the facial palsy severity in the YFP database. The palsy information was provided by Hsu *et al.* in [19] to indicate the grade of paralysis

Table 3.10: Data distribution for the prediction of two palsy levels.

Test	Total of Images	Data Distribution
Eyes region	680	Original data: 208 low-intensity and 472 high-intensity samples
	2040	Augmented data: 624 low-intensity and 1416 high-intensity samples
Mouth region	680	Original data: 141 low-intensity and 539 high-intensity samples
	2040	Augmented data: 423 low-intensity and 1617 high-intensity samples

found in that specific palsy region. The authors analyzed only two palsy regions: the eyes and the mouth. Notice that the label provided by Hsu *et al.* does not reflect the lesion's severity on the face; it describes the severity of the eyes or the mouth. There might be a case where the intensity is not the same for both, then a separate analysis is required.

We also explore the use of regional information in the following experiments. As explained earlier, we refer to regional information to the use of specific features, depending on the region under analysis. Then, the eyes' information corresponds to features named Eyebrows, Eyes, Nose, and Combined (only *f23* and *f25–f27*) in Table 2.1, 19 attributes in total. And the mouth's information corresponds to all features, except those named Eyebrows and Eyes features in Table 2.1, a total of 15 attributes.

A 5-fold cross-validation technique was employed in these evaluations to train and evaluate the algorithms. Each validation was repeated ten times to observe the behavior of the classifiers and provide information regarding their repeatability. Therefore, the average performance of those repetitions is given for each test. More details of the experiments and results are now described.

3.3.1 Prediction of two palsy levels

In the first experiment, the data set is composed of 19 patients from the YFP database and is now divided into class SL_1 (low-intensity) and class SL_2 (high-intensity). After preliminary tests, it was found that 40 images per patient were adequate to train the classifier, but for those who did not have enough photos, only 20 were employed. To improve the learning process, a data augmentation was performed as suggested in [16, 22]. This process consisted of rotating in two opposite directions the palsy images to increase the amount of available data. This augmentation also allows us to verify that the algorithm is invariant to rotation, as previously stated. This experiment is divided into two tests: (a) using the entire face information (29 features) and (b) using regional information (19 features for the eyes and 15 features for the mouth region).

The data distribution for this experiment is described in Table 3.10. To evaluate the eyes' level of paralysis, 208 low-intensity and 472 high-intensity images are included. After

Table 3.11: Prediction of two palsy levels: classifiers' configuration using the entire face information.

Method	Parameters
MLP	$L = 0.2045$, $M = 0.1909$, $H = 59$, $N = \{500, 2000\}$ *, $S = \{2, 37\}$ *
SVM	$C = 1000$, $G = 1.0$, kernel type = RBF
KNN	KNN = 1, A =Manhattan distance
MNLR	$R = 1 \times 10^{-8}$

* Refers to the values for the eyes and mouth evaluation, respectively.

Table 3.12: Results on the criteria of accuracy using the entire face information.

Classifier	Eyes	Mouth
MLP	$92.39 \pm 1.23\%$	$90.20 \pm 1.27\%$
SVM	$95.05 \pm 1.14\%$	$92.69 \pm 1.01\%$
KNN	$93.54 \pm 1.24\%$	$92.14 \pm 1.16\%$
MNLR	$82.96 \pm 1.83\%$	$81.09 \pm 1.47\%$

data augmentation, the data set is formed by 624 low-intensity and 1416 high-intensity samples (2040 images in total). Similarly, to evaluate the mouth's level of paralysis, 141 low-intensity and 539 high-intensity images are included. After data augmentation, the data set is formed by 423 low-intensity and 1617 high-intensity samples (2040 images in total).

For the first test, using the entire face information, the configuration of the classifiers is described in Table 3.11, and the average performance for both regions is shown in Table 3.12. Using SVM, excellent results are obtained assessing the eyes region, up to 95.05%. Similarly, good results are reached for the mouth area, 92.69%. The KNN classifier achieved better results in this task than the MLP for both cases. Still, the proposed MLP yielded good results using few samples compared to other published deep learning approaches that require thousands of images and complex neural network structures.

For the second test, the configuration of the classifiers is described in Table 3.13, and the average performance for both regions using fewer features is shown in Table 3.14. Lower performance was expected because the information fed to the classifiers was decreased; still, good performance (more than 90%) was achieved using SVM. A slight increase in performance is observed when using the information from the mouth region. This evaluation leads us to believe that a classification of the palsy intensity is possible to a certain degree in partial occlusions of the face.

3.3.2 Prediction of three palsy levels

The goal of assessing the severity of the lesion is to determine how diminished is the facial nerve function. It can be evaluated once the palsy has been detected or assessed simultaneously. For the last experiment, the data set is divided into class SL_0 (healthy),

Table 3.13: Prediction of two palsy levels: classifiers' configuration using regional information.

Method	Parameters
MLP	$L = 0.2045$, $M = 0.1909$, $H = 59$, $N = \{1000, 500\}$ *, $S = \{87, 7\}$ *
SVM	$C = 10$, $G = 1.0$, kernel type $=$ RBF
KNN	KNN $= 1$, $A =$ Manhattan distance
MNLR	$R = 1 \times 10^{-8}$

* Refers to the values for the eyes and mouth evaluation, respectively.

Table 3.14: Results on the criteria of accuracy using regional information.

Classifier	Eyes	Mouth
MLP	$88.75 \pm 1.87\%$	$89.60 \pm 1.27\%$
SVM	$91.12 \pm 1.57\%$	$93.29 \pm 1.22\%$
KNN	$89.99 \pm 1.54\%$	$91.70 \pm 0.94\%$
MNLR	$79.97 \pm 1.97\%$	$80.15 \pm 1.13\%$

class SL_1 (slight or low-intensity palsy), and class SL_2 (strong or high-intensity palsy). A sample of healthy subjects and patients with facial palsy was introduced in Figure 1.2. The images show the two palsy regions of interest in this work: the eyes and mouth.

This experiment is also divided into two tests (a) using the entire face information (29 features) and (b) using regional information (19 features for the eyes and 15 features for the mouth). The data set consists of 19 patients from the YFP database and 19 participants from the CK+ database. In this case, 40 images per subject were used; for those who did not have enough photos, only 20 were employed. Once more, the same data augmentation process was performed to provide enough information to the classifiers, increasing the amount of data and verifying that the methodology is rotation invariant.

The data distribution is described in Table 3.15. For the eyes region, 740 healthy, 208 low-intensity, and 472 high-intensity samples are included; it is easy to observe that the classes are unbalanced and that there are few examples for the class SL_1. After sample

Table 3.15: Data distribution for the prediction of three palsy levels.

Test	Total of Images	Data Distribution
Eyes region	1420	Original data: 740 healthy, 208 low-intensity and 472 high-intensity samples
	4260	Augmented data: 2220 healthy, 624 low-intensity and 1416 high-intensity samples
Mouth region	1420	Original data: 740 healthy, 141 low-intensity and 539 high-intensity samples
	4260	Augmented data: 2220 healthy, 423 low-intensity and 1617 high-intensity samples

Table 3.16: Prediction of three palsy levels: classifiers' configuration using the entire face information.

Method	Parameters
MLP	$L = 0.2045$, $M = 0.1909$, $H = 59$, $N = \{1000, 7000\}$ *, $S = \{18, 25\}$ *
SVM	$C = \{10, 1000\}$ *, $G = 1.0$, kernel type $=$ RBF
KNN	$KNN = 1$, $A =$Manhattan distance
MNLR	$R = 1 \times 10^{-8}$

* Refers to the values for the eyes and mouth evaluation, respectively.

Table 3.17: Results on the criteria of accuracy using the entire face information.

Classifier	Eyes	Mouth
MLP	$93.67 \pm 0.94\%$	$92.77 \pm 0.87\%$
SVM	$95.58 \pm 0.71\%$	$94.44 \pm 0.63\%$
KNN	$93.21 \pm 0.80\%$	$92.94 \pm 0.80\%$
MNLR	$86.48 \pm 0.82\%$	$85.95 \pm 0.77\%$

augmentation, 4260 images compose the training set with 2220 healthy, 624 low-intensity and 1416 high-intensity samples. For the mouth region, 740 healthy samples, 141 low-intensity, and 539 high-intensity samples are included; again, there are fewer examples for the class SL_1. After sample augmentation, 4260 images compose the training set with 2220 healthy, 423 low-intensity and 1617 high-intensity samples. In both tests, the classes remained unbalanced, but after several evaluations, this data distribution provided the best results.

The configuration of the classifiers is described in Table 3.16, and the average performance for both regions using the entire face information is shown in Table 3.17. For the SVM strategy, excellent results are obtained in the eyes regions, up to 95.58%. Similarly, good results are reached for the mouth area, 94.44%. In both cases, the proposed MLP yielded similar results using few samples compared to other deep learning strategies that require thousands of samples and complex neural network structures.

For the second test, using regional information, the configuration of the classifiers is described in Table 3.18, and the average performance for both regions using fewer features is shown in Table 3.19. It was expected a lower performance because the information fed to the classifiers was decreased; still, good performance of 92.08% and 93.95% was achieved using SVM. Once more, this evaluation leads us to believe that palsy intensity classification is possible to a certain degree in partial occlusions of the face.

3.3.3 Findings and discussion

To the best of our knowledge, only the work of Liu *et al.* in [26] introduces a three-level classification system to distinguish between healthy and palsy faces. The authors argue that a clinician manually labeled all palsy subjects according to the HB grading scale.

Table 3.18: Prediction of three palsy levels: classifiers' configuration using regional information.

Method	Parameters
MLP	$L = 0.2045$, $M = 0.1909$, $H = 59$, $N = \{4000, 7000\}$ *, $S = \{18, 25\}$ *
SVM	$C = 10$, $G = 1.0$, kernel type = RBF
KNN	KNN $= 1$, $A = \{$Manhattan, Euclidean distance$\}$
MNLR	$R = 1 \times 10^{-8}$

* Refers to the values for the eyes and mouth evaluation, respectively.

Table 3.19: Results on the criteria of accuracy using regional information.

Classifier	Eyes	Mouth
MLP	$89.63 \pm 1.04\%$	$91.91 \pm 0.94\%$
SVM	$92.08 \pm 0.79\%$	$93.95 \pm 0.62\%$
KNN	$89.24 \pm 0.92\%$	$92.08 \pm 0.83\%$
MNLR	$83.07 \pm 1.07\%$	$84.30 \pm 0.76\%$

Based on those labels, they grouped the lesion's severity into three levels: (I) grade 1, (II) grade 2–4, and (III) grade 5-6. The information regarding which palsy participant of the YFP database belongs to each group is not publicly available after we requested it from the authors. Then, a direct comparison of the methodologies is not feasible because the data might be grouped into different classes. The performance is evaluated because the same databases are employed, and the same classification goal is aimed.

The authors in [26] reported accuracy values of 94.81% and recall of 94.8%. We must mention that their methodology requires a sequence of images from the same subject to output a result, while our algorithm only needs one image. It is unclear how they measured their results; we reported them separately: accuracy values of 93.67% and recall of 97.86% for the eyes, accuracy values of 92.76% and recall of 97.89% for the mouth, using the MLP classifier in both cases. It seems that our classifiers based on neural networks reached a lower performance than those reported by Liu *et al.* in terms of accuracy. However, we wish to emphasize that our classification system requires fewer images (3,408 against 21,612) in the training phase and a simpler architecture to perform the analysis and output a label. Then, we firmly believe our results are competitive with those in [26]. Finally, most of our best results were obtained using the SVM classifier (95.58% for the eyes and 94.44% for the mouth).

CHAPTER 4

GESTURE RECOGNITION IN PALSY PATIENTS

For this task, we seek to classify facial gestures in palsy patients as a first step to recognizing emotional states based on facial expressions. We assume that a facial expression is composed of one or more facial gestures, and a gesture is produced by activating one or more muscles. In computer vision, such muscle activation is described through action units. An action unit is a code that names the muscle activity, making its identification easier and providing a common ground in this field.

If analyzing faces, the related action units (AUs) are encoded in the Facial Action Coding System (FACS) according to [21, 28]. The FACS consists of many action units; some are symmetrical or not, and some vary in intensity. The action units may occur in combinations or show serial dependence. For example, for raising the eyebrows, the inner brow and outer brow raisers must be activated; for smiling with a closed mouth, the lip corner puller and lip stretcher must be activated. The seven basic emotions (anger, contempt, disgust, fear, happiness, sadness, and surprise) are also described, among other characteristics, in terms of facial action units [28].

We argue that identifying specific FACS leads us to classify specific facial gestures and, later, to recognize facial expressions [54]. As represented in Figure 4.1, a common recognition system consists of a machine learning algorithm fed with specific facial gestures to identify a particular facial expression. Again, we think of this as a first step in recognizing emotions from facial expressions.

As stated in the previous chapters, specific analysis, data, and models are necessary to deal with palsy patients' images, other than the actual work developed and performed on healthy people (*i.e.,* not affected by paralysis). A similar necessity is found for detecting facial gestures [53, 13, 52]. The main contributions of this sub-project are (1) an analysis of facial gestures in persons with facial palsy and (2) a system for recognizing gestures on photographs of patients exhibiting facial paralysis. Such a system consists of a classifier based on neural networks using our handcrafted features extracted from facial landmarks.
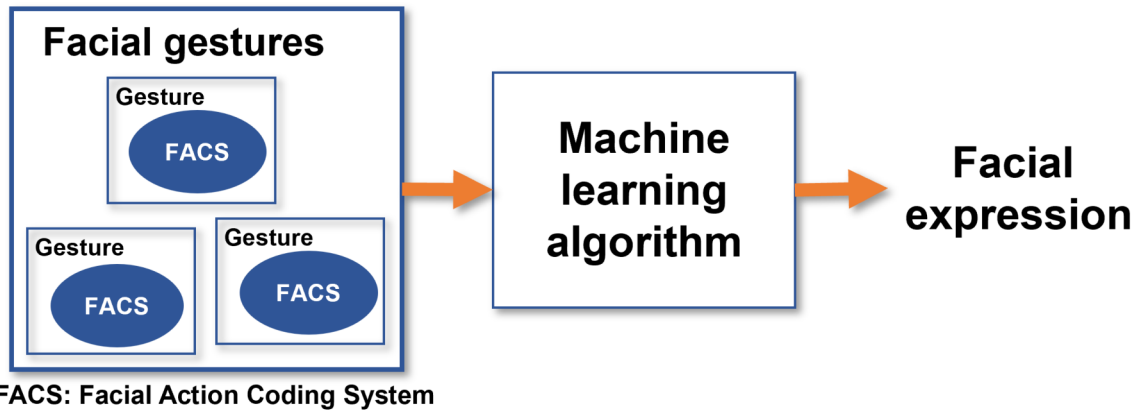
Figure 4.1: A common facial expression recognition system based on analyzing face images and machine-learning algorithms to output a facial expression.
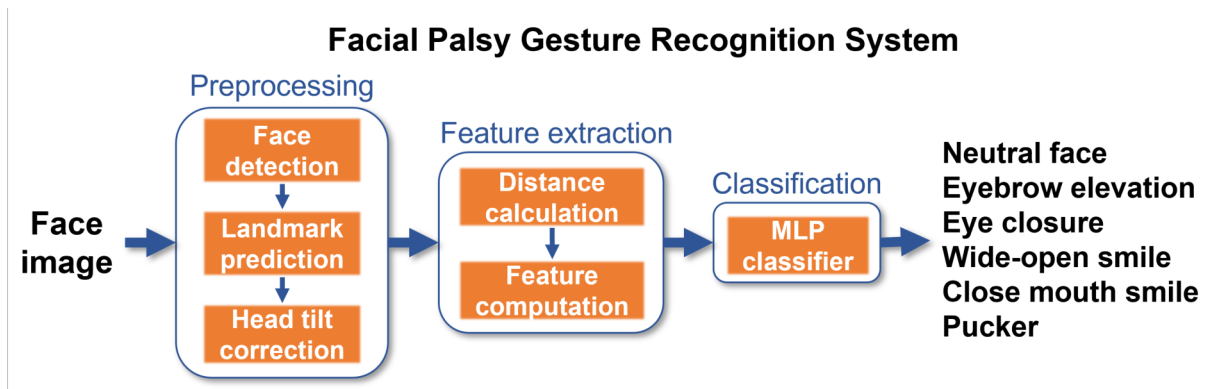
Figure 4.2: Framework of the proposed facial palsy gesture recognition system.

4.1 Framework

Our methodology is shown in Figure 4.2. It is easy to observe that it inspects a single image aiming to recognize its facial gesture. Our approach performs face detection and later predicts facial landmarks, as explained in Section 2.1.1. Facial symmetry features are computed from those landmarks in mainly three regions: the eyebrows, eyes, and the mouth, which is similar to other approaches [24]. A multi-layer perceptron classifier is trained, as described in Section 3.1, to detect six gestures using the proposed symmetry features. Figure 4.3 and Table 4.1 introduce those facial gestures: rest (neutral), eyebrow elevation, eye closure, wide-open smile, closed mouth smile, and pucker face.

The MEEI database was employed in this classification task. As mentioned in Section 2.2, the participants performed eight facial movements that are of interest: (1) at rest, (2) eyebrow elevation, (3) light effort eye closure, (4) full effort closure, (5) light effort smile, (6) full effort smile, (7) pucker, and (8) lip depression. A set of those facial movements is shown

Six classes of facial gestures

	Class 0 Neutral face	Class 1 Eyebrow elevation	Class 2 Eye closure	Class 3 Wide-open smile	Class 4 Close mouth smile	Class 5 Pucker face
Eyebrows						
Eyes						
Mouth						
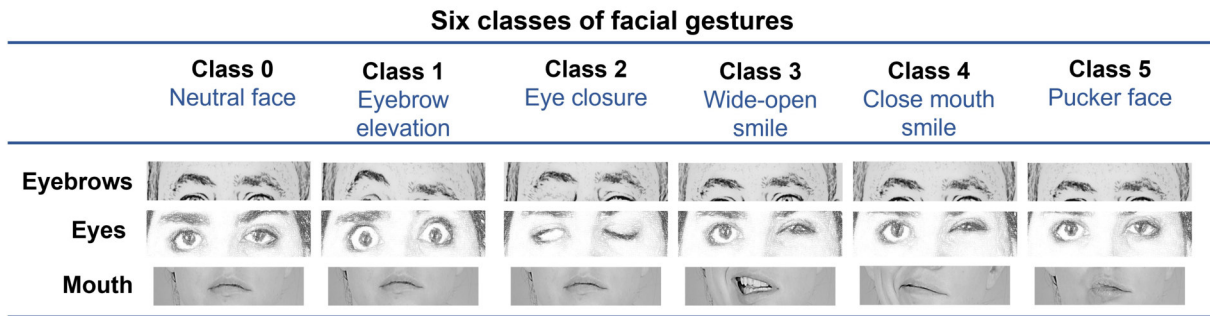
Figure 4.3: The proposed facial palsy gestures are grouped into six classes.

in Figure 2.4. We wish to emphasize that the MEEI database is annotated according to the level of paralysis. This research aims to recognize facial gestures using the information we extracted according to the AUs described in [21, 28].

The eight aforementioned facial movements are grouped into six classes following these considerations: (1) closing the eyes is considered a single gesture since it is not our intention to measure the intensity of the movement; (2) smiling is split into smiles showing the teeth (wide-open smile) or just enlarging the lips (closed mouth smile), measuring the intensity of the gesture is out of our scope; (3) during preliminary tests, we observed that for a palsy patient, performing a lip depression—which consists of showing the lower teeth—is very similar to a wide-open smile; thus, it was grouped as a single gesture. Figure 4.3 depicts the above six classes of facial gestures.

A description of the analyzed facial gestures is given in Table 4.1, divided into three facial regions and the more relevant AUs involved in each. When the patient is asked to raise the eyebrows, we observed that the movement comes with a whole eye-opening; the mouth is irrelevant in this case. Smiling shows the most considerable asymmetry within the lips, vertically in the case of a wide-open smile and horizontally in the other case. Here, it is also observed that patients tend to show a slight asymmetry in the eye-opening when smiling or performing a pucker gesture. In some cases, not all the AUs are meaningful to recognize the gesture; they are marked as irrelevant to the methodology. It does not mean they are not relevant in a different context.

Our facial symmetry features can discriminate between different positions and shapes of the face organs involved in the analyzed facial gestures. Additional information showing the behavior of the face elements is given in Figure 4.3. For example, the eyebrows exhibit the most significant asymmetry in the eyebrow elevation movement. Depending on the palsy severity, such asymmetry can also be observed in other expressions, such as the neutral face. For the eyes, the most asymmetrical level is found for the eye closure; however, a significant change is also observed when the subject elevates the eyebrows. Depending on the palsy severity, this asymmetrical eye-opening could be observed when resting the face. Notice that the photographs represent posed expressions; a different behavior could be observed in in-the-wild images.

Table 4.1: General description of the facial gestures following [28].

Facial Gesture	Face Region		
	Eyebrows	Eyes	Mouth
Neutral face	No movement	No movement	No movement
Eyebrow elevation	Inner brow raiser (AU1) and outer brow raiser (AU2)	Irrelevant	Irrelevant
Eye closure	Irrelevant	Eyes closed (AU43)	Irrelevant
Wide-open smile	Irrelevant	Irrelevant	Lip corner puller (AU12), lips part (AU25) and jaw drop (AU26)
Closed mouth smile	Irrelevant	Irrelevant	Lip corner puller (AU12) and lip stretcher (AU20)
Pucker face	Irrelevant	Irrelevant	Lip puckerer (AU18)

4.2 Experiments and results

The 5-fold cross-validation protocol is adopted during the training process to evaluate the model's performance on new data. Accuracy, recall, precision, and F1 score values are computed as evaluation metrics according to Equation (3.1), (3.2), (3.4), and (3.5), respectively. Due to the limited amount of data, 80 healthy and 400 palsy samples (200 flaccid and 200 non-flaccid), it was necessary to augment the data set. Following the strategy suggested in [16, 22], the samples were rotated in opposite directions, increasing the amount of data and verifying that the proposed methodology is invariant to rotation. In preliminary tests, increasing the amount of data to 560 healthy and 2800 palsy samples, meaning three rotations in opposite directions and the original data, was shown to be sufficient to obtain good results without over-fitting.

For this facial gesture recognition task, an MLP classifier was trained and configured using $L = 0.2045$ and $M = 0.1909$ because it yielded a good result in previous classification challenges. The rest of the parameters were defined after several trials and an optimization process using the Weka platform, as explained in Section 3.1. We found that setting the parameters to $H = 95$, $S = 0$, and $N = 5000$ reached the best results.

The following experiment was carried out to illustrate the necessity of having specific models designed with palsy data. A classifier was trained and tested using only the healthy data from the MEEI database. The 560 samples were distributed as follows: 70 samples for class 0, 70 samples for class 1, 140 samples for class 2, 161 samples for class 3, 49 samples for class 4, and 70 samples for class 5. The overall accuracy of 90.71% was reached

Table 4.2: The results per fold of the facial gesture recognition system.

n-Fold	Correct	Errors	Accuracy
1-fold	509	51	90.89%
2-fold	506	54	90.36%
3-fold	501	59	89.46%
4-fold	507	53	90.54%
5-fold	504	56	90.00%
Overall	2527	273	90.25%

Table 4.3: The results per class of the facial gesture recognition system.

Class	Recall	Precision	F1-Score
Class 0	83.71%	78.55%	81.05%
Class 1	86.57%	89.38%	87.95%
Class 2	89.43%	91.92%	90.66%
Class 3	97.12%	92.84%	94.93%
Class 4	83.52%	86.86%	85.15%
Class 5	88.57%	95.68%	91.99%

Class 0 refers to neutral face, class 1 to eyebrow elevation, class 2 to eye closure, class 3 to wide-open smile, class 4 to closed mouth smile, and class 5 to pucker face.

in recognizing facial gestures using the 5-fold cross-validation strategy. The palsy data was also evaluated with this model trained with healthy samples, and the performance decreased to 58.96%. This performance decrement shows that a model trained with palsy information is required to reach acceptable results, as discussed earlier.

Another classifier was trained using the 2800 palsy samples, which are distributed as follows: 350 samples for class 0, 350 samples for class 1, 700 samples for class 2, 868 samples for class 3, 182 samples for class 4, and 350 samples for class 5. The results are provided in Table 4.2. The values correspond to the performance of the testing part of the data for each fold. The system accuracy is 90.25%, with the best result of 90.86% and the worst of 89.46%. The results provided in Table 4.3 reflect the capacity of the system to recognize each class. It is easy to observe that the best-recognized gesture is class 3 (wide-open smile), while class 4 (close-mouth smile) is the worst in recall values, 97.12% against 83.52%, likely because of the data distribution previously described. A deeper analysis of the capacity of the system to recognize each class is provided in Section 4.3.

The confusion matrix with actual and predicted labels is given in Figure 4.4, showing the accuracy of the proposed methodology. By looking into it, we can observe that the classification system can recognize different facial gestures. The best result is 97.12% for class 3, and the worst is 83.52% for class 4. As mentioned before, the system's overall accuracy is 90.25%, representing a good result if considering this is a six-class classification system. Furthermore, considering that the baseline accuracy was set at 58.96%, a
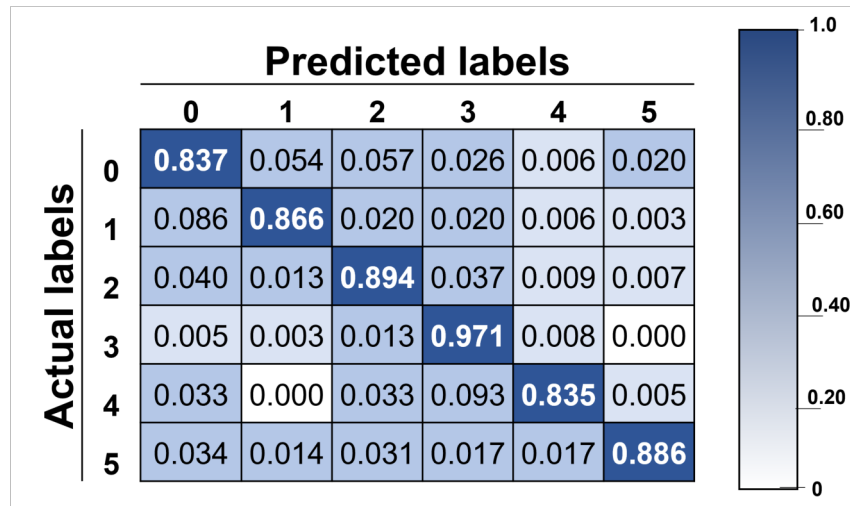
Figure 4.4: Confusion matrix with actual and predicted labels: class 0 for neutral face, class 1 for eyebrow elevation, class 2 for eye closure, class 3 for wide-open smile, class 4 for closed mouth smile, and class 5 for pucker face.

significant improvement was achieved.

Additional experiments were performed to observe the capacity of the recognition system based on the level of facial palsy. For this test, six classifiers were trained under the same parameters (*i.e.*, $L = 0.2045$ and $M = 0.1909$, $H = 95$, $S = 0$, and $N = 5000$); however, only the data corresponding to the same grade of paralysis was employed. The data distribution consists of 10 participants for each grade: normal, near-normal, mild, moderate, severe, and complete.

The recognition system results according to the facial palsy level are shown in Table 4.4. Good results are revealed, especially for the complete level of paralysis with an overall accuracy of 96.79%. The lowest performance was 87.14% for the mild grade; this palsy level was the hardest to characterize for our system. The overall accuracy of 90.71% seems acceptable for normal subjects; however, higher accuracy values are found in the literature. However, it is not our scope to perform facial gesture recognition on healthy people.

In Figure 4.5, the confusion matrix for each of the evaluations is depicted. In Figure 4.5a, it is possible to observe that a neutral face is mainly confused with closed eyes in a healthy face; this is expected behavior since a little change in the organs' shape would be observed because of the symmetry of the face. The same justification applies for eyebrow elevation being confused with a neutral face. In the case of near-normal palsy, shown in Figure 4.5b, the worst class is closed mouth smile, which is confused with eye closure. We argue that this is produced by the lip corner puller activation, making what looks like a smile. In the case of mild palsy, shown in Figure 4.5c, it is easy to observe that this level of paralysis is the most difficult to characterize.

Here, the neutral face is confused with four possible classes. In the case of moderate palsy, shown in Figure 4.5d, the worst class is puckered face, which is confused with all

Table 4.4: The results per class for each level of paralysis of the facial gesture recognition system.

Test	Class 0	Class 1	Class 2	Class 3	Class 4	Class 5	Overall
Normal	75.71%	87.14%	86.43%	98.76%	91.84%	97.14%	90.71%
Near-normal	87.14%	94.29%	85.71%	98.81%	80.95%	91.43%	91.25%
Mild	71.43%	80.0%	80.0%	98.41%	76.19%	97.14%	87.14%
Moderate	87.14%	95.71%	87.86%	97.80%	89.29%	81.43%	91.25%
Severe	88.57%	85.71%	86.43%	96.89%	93.88%	91.43%	90.89%
Complete	91.43%	95.71%	93.57%	100%	100%	100%	96.79%
Average	83.57%	89.76%	86.67%	98.44%	88.69%	93.09%	91.34%

Class 0 refers to neutral face, class 1 to eyebrow elevation, class 2 to eye closure, class 3 to wide-open smile, class 4 to closed mouth smile, and class 5 to pucker.

possible classes, likely because the asymmetry of the mouth is not sufficient to characterize it into specific gestures. In the case of severe palsy, shown in Figure 4.5e, the worst class is eyebrow elevation, and we argue that it is produced by the patient's incapacity to raise the eyebrows. Finally, in the case of complete palsy, shown in Figure 4.5e, the worst class is the neutral face, which is slightly confused with three of the five possible classes.

At the introduction of this thesis, we mentioned the work of Xu *et al.*, which aimed to evaluate emotions in patients with facial palsy through deep learning [52]. To start, the authors collected images representing facial expressions of six basic emotions from 45 patients with three levels of facial palsy. The authors reported a performance increase from 23.89% to 66.58% after training using the palsy data. Moreover, their emotion recognition accuracy rates were 72.22% for mild, 65.83% for moderate, and 61.67% for severe levels of paralysis. Their work is not a direct comparison to ours. They are related because both seek to provide solutions in analyzing images of facial paralysis, and both concluded that new models trained with palsy data are required.

4.3 Findings and discussion

With our methodology, the levels of asymmetry were found by measuring the change rate in the height and width of each eye and the mouth and by observing the difference in the vertical position between the eyes and eyebrows. The method mainly compares the left and right sides of the face, as depicted in Figure 4.6, thereby allowing good results in detecting facial gestures.

In a further analysis and referring to Figures 4.5 and 4.6, we found that:

- A neutral face is confused with closing the eyes because there is a slight difference between the eyebrows' position (see Figure 4.6a) and the mouth shape for both gestures (see Figure 4.6b); however, if there is enough paralysis where one eye remains opened and the other closed, the change is perceived by the classifier, and the movement is
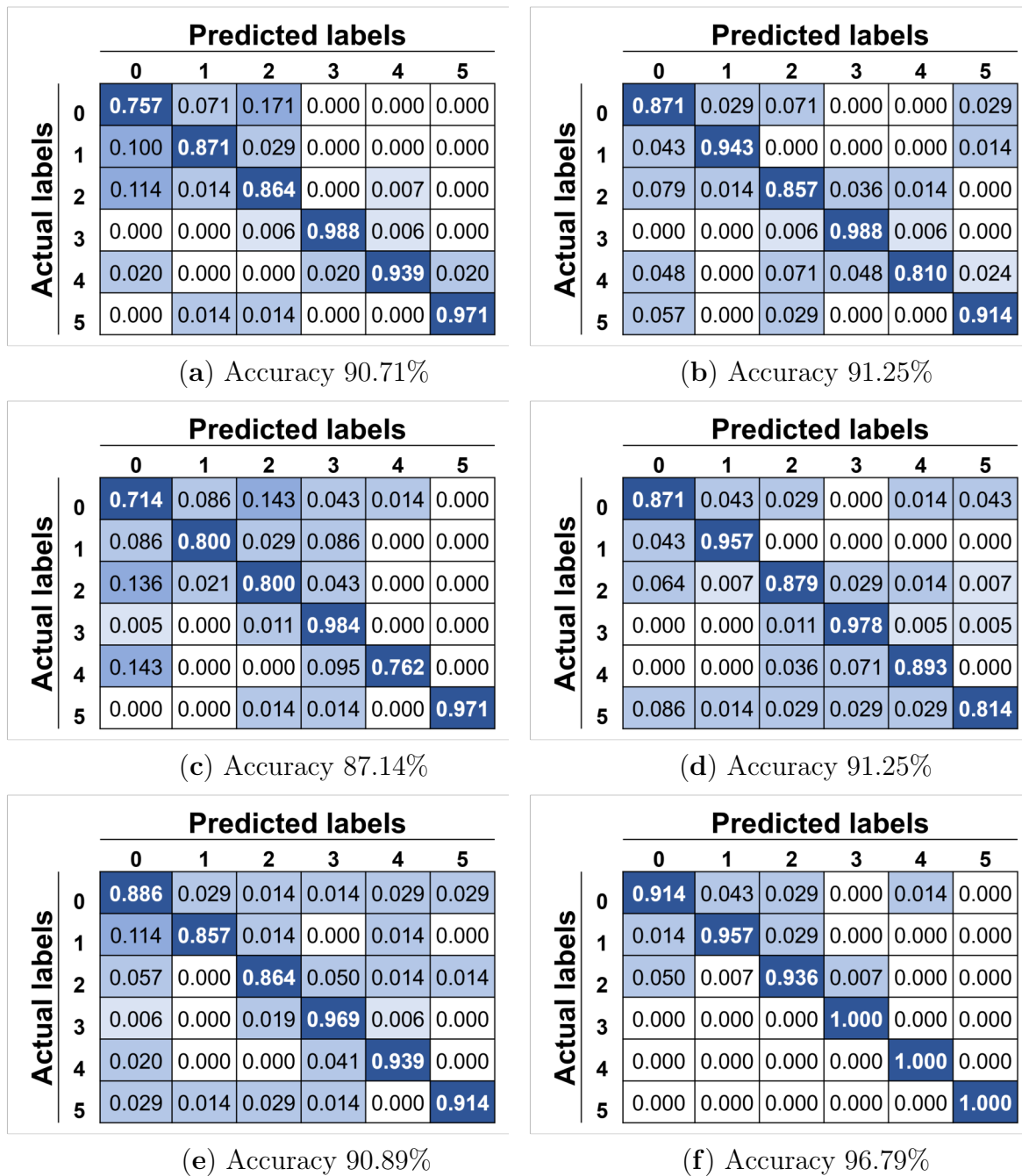
Predicted labels

	0	1	2	3	4	5
0	**0.757**	0.071	0.171	0.000	0.000	0.000
1	0.100	**0.871**	0.029	0.000	0.000	0.000
2	0.114	0.014	**0.864**	0.000	0.007	0.000
3	0.000	0.000	0.006	**0.988**	0.006	0.000
4	0.020	0.000	0.000	0.020	**0.939**	0.020
5	0.000	0.014	0.014	0.000	0.000	**0.971**

(**a**) Accuracy 90.71%

Predicted labels

	0	1	2	3	4	5
0	**0.871**	0.029	0.071	0.000	0.000	0.029
1	0.043	**0.943**	0.000	0.000	0.000	0.014
2	0.079	0.014	**0.857**	0.036	0.014	0.000
3	0.000	0.000	0.006	**0.988**	0.006	0.000
4	0.048	0.000	0.071	0.048	**0.810**	0.024
5	0.057	0.000	0.029	0.000	0.000	**0.914**

(**b**) Accuracy 91.25%

Predicted labels

	0	1	2	3	4	5
0	**0.714**	0.086	0.143	0.043	0.014	0.000
1	0.086	**0.800**	0.029	0.086	0.000	0.000
2	0.136	0.021	**0.800**	0.043	0.000	0.000
3	0.005	0.000	0.011	**0.984**	0.000	0.000
4	0.143	0.000	0.000	0.095	**0.762**	0.000
5	0.000	0.000	0.014	0.014	0.000	**0.971**

(**c**) Accuracy 87.14%

Predicted labels

	0	1	2	3	4	5
0	**0.871**	0.043	0.029	0.000	0.014	0.043
1	0.043	**0.957**	0.000	0.000	0.000	0.000
2	0.064	0.007	**0.879**	0.029	0.014	0.007
3	0.000	0.000	0.011	**0.978**	0.005	0.005
4	0.000	0.000	0.036	0.071	**0.893**	0.000
5	0.086	0.014	0.029	0.029	0.029	**0.814**

(**d**) Accuracy 91.25%

Predicted labels

	0	1	2	3	4	5
0	**0.886**	0.029	0.014	0.014	0.029	0.029
1	0.114	**0.857**	0.014	0.000	0.014	0.000
2	0.057	0.000	**0.864**	0.050	0.014	0.014
3	0.006	0.000	0.019	**0.969**	0.006	0.000
4	0.020	0.000	0.000	0.041	**0.939**	0.000
5	0.029	0.014	0.029	0.014	0.000	**0.914**

(**e**) Accuracy 90.89%

Predicted labels

	0	1	2	3	4	5
0	**0.914**	0.043	0.029	0.000	0.014	0.000
1	0.014	**0.957**	0.029	0.000	0.000	0.000
2	0.050	0.007	**0.936**	0.007	0.000	0.000
3	0.000	0.000	0.000	**1.000**	0.000	0.000
4	0.000	0.000	0.000	0.000	**1.000**	0.000
5	0.000	0.000	0.000	0.000	0.000	**1.000**

(**f**) Accuracy 96.79%

Figure 4.5: Confusion matrix with actual and predicted labels for test: (**a**) normal, (**b**) near-normal, (**c**) mild, (**d**) moderate, (**e**) severe, and (**f**) complete paralysis. Class 0 refers to neutral face, class 1 to eyebrow elevation, class 2 to eye closure, class 3 to wide-open smile, class 4 to closed mouth smile, and class 5 to pucker face.

classified correctly (see Figure 4.6c).

- A neutral face is confused with eyebrow elevation; some patients show similar asymmetry in the eyebrows' vertical position due to their level of paralysis.

- A closed mouth smile is confused with a wide-open smile mainly because a similar rate of change is produced in the eyes opening and in the mouth width or opening (see Figure 4.6d).

- A closed mouth smile is confused with closing the eyes because, in some patients, a full effort eye closure activates the lip corner puller, producing what looks like a smile (see Figure 4.6e).

- A closed mouth smile is confused with a neutral face because, in some patients, the asymmetry in the mouth shape is insufficient to classify it as a smile.

- A pucker face is confused with a neutral face because some patients cannot fully perform the gesture. Their mouth looks asymmetrical; however, the eyes and the eyebrows remain close to no movement (see Figure 4.6f).

- A pucker face is confused with closing the eyes because some patients cannot fully perform the gesture but show a change in the eyes' opening.

.

In conclusion, we emphasize that to achieve facial gesture recognition and later emotion classification in patients with facial paralysis, the level of damage must be considered in the recognition system design because the paralyzed facial organs exhibit different behavior compared with healthy organs. Regarding accuracy and face generalization, improvements to the methodology would require more data from facial palsy patients with labeled facial expressions and emotions.
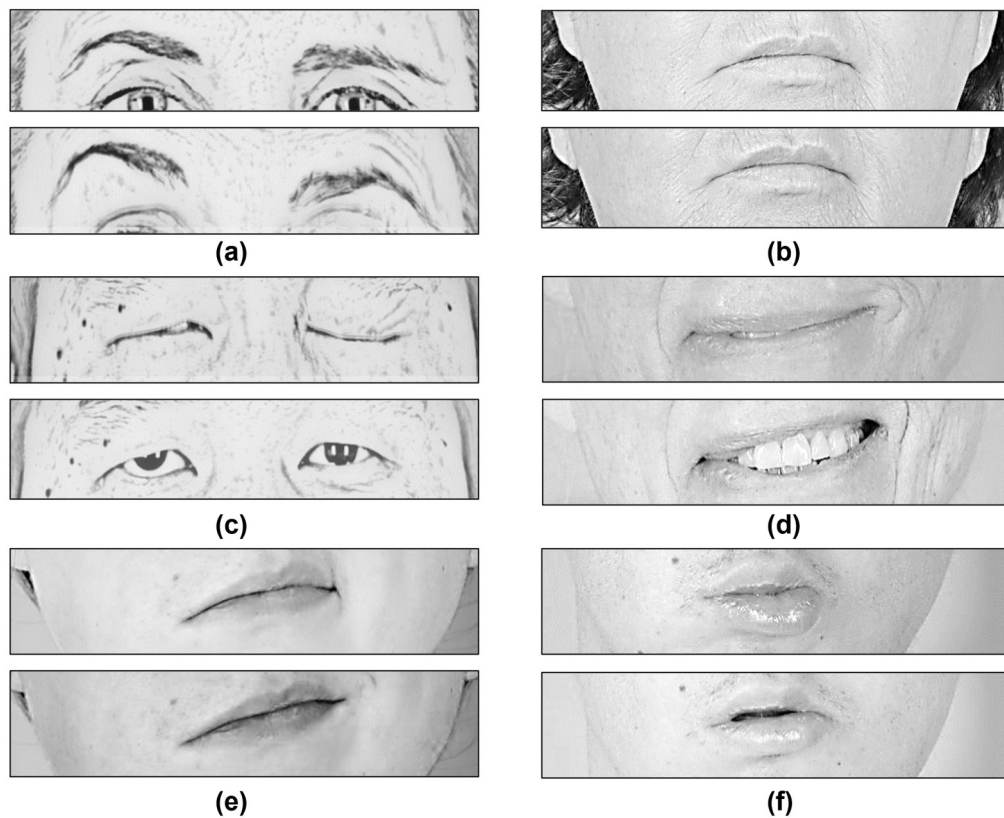
Figure 4.6: Analysis of facial gestures in patients who have facial palsy: (**a**) An small change in the eyebrows' height. (**b**) A similar mouth shape (width and height). (**c**) A slightly different change in the eyes' height. (**d**) A similar change in the mouth shape. (**e**) An activated lip corner puller. (**f**) A similar change in the mouth shape.

CONCLUSIONS AND FUTURE WORK

Conclusions

In this thesis, we introduced an analysis of facial paralysis using computer vision techniques and machine learning algorithms. We focused our facial analysis on terms of symmetry, knowing that almost all humans exhibit a level of asymmetry between the left and right sides of the face. We seek to find a boundary capable of distinguishing between healthy and palsy faces. In other words, we wished to characterize human faces by assuming that a healthy face is pretty symmetrical and a palsy face is not.

As a result of such facial palsy analysis, we have designed 29 facial symmetry features that measure the position between the face elements and their shape at some level. Our proposed features are extracted from facial landmarks, which are predicted after a face detection process on the image and using shape predictors trained on palsy patients' imagery. Landmark extraction is crucial to calculating our facial features.

Our proposed methodology is suitable for performing different classification tasks using our facial symmetry features. We have explored three objectives: (1) detection of facial paralysis, (2) assessment of the palsy lesion's severity, and (3) recognition of facial gestures in palsy patients. Experiments were performed on four (three palsy and one healthy) publicly available image databases to validate our methodology, and different classification approaches were tested.

Our first experiment aimed to detect facial paralysis using a binary classifier based on the multi-layer perceptron strategy, which provides a healthy or unhealthy label as an output. It was evaluated on the MEEI and TNF databases. In the first one, the approach attained a 94.06% correct classification; a performance of 97.22% was reached on the second database, showing that our methodology to design a binary classifier can be expanded to other databases with excellent results.

Four different classifiers were evaluated in the following experiments aiming to detect palsy regions, discriminate among two levels of the palsy, and assess the lesion's severity

among three palsy levels. Tests were performed on the YouTube Facial Palsy and the CK+ databases. The best results were achieved using SVM, but a similar performance with a slight decrease is obtained using the multi-layer perceptron method. After the evaluations, it was found that dividing the face into specific regions is convenient to detect and assess paralysis with fewer features. This feature reduction leads us to believe that the analysis of facial paralysis is possible with partial occlusions of the face if face detection is achieved and landmarks are predicted adequately.

In detecting palsy regions, the proposed system achieved the highest accuracy of 95.61%. In the discrimination task of two palsy levels (low-intensity vs. high-intensity), the system achieved the highest accuracy of 95.05% in the eyes and 93.29% in the mouth region. In assessing the lesion's severity (healthy, low-intensity, and high-intensity), the system achieved the highest accuracy of 95.58% for the eyes and 94.44% for the mouth region.

Our final experiment aimed to recognize six gestures in patients with facial paralysis. The gesture recognition system was evaluated using the patients from the MEEI database. The recognition system obtained an overall accuracy of 90.25%; the best result yielded 97.12% and the worst 83.52%. After the evaluations, we found that any recognition system must consider the level of paralysis exhibited by the patient because the healthy facial organs show different behavior compared to the paralyzed face elements. In other words, the recognition systems must be capable of discerning such behavior.

The results show that our methodology helps to solve various issues of interest in facial palsy. Other approaches have aimed to analyze faces in subjects with cognitive impairments and pain; however, this research is the first attempt to perform facial gesture recognition in patients with facial palsy.

The accomplished results show that our methodology to design classifiers can be adapted to other datasets with outstanding results for various classification tasks. It is a methodology easy to replicate compared to the other complex systems and achieves similar results for detecting and evaluating facial paralysis. Remarkably, the proposed classifiers based on neural networks require fewer samples in the training stage compared to other approaches based on deep neural networks.

After submitting our research to several peer-review journals, it was noted that evaluating facial paralysis using symmetry/asymmetry values is risky because the human face is not identical concerning its left and right sides. Then, to thoroughly verify the usefulness of our algorithm in clinical practice, a much larger sample of healthy controls with different degrees of facial asymmetry (not caused by facial palsy) is needed. Achieving this monumental task would require a specific database of healthy participants (*i.e.,* showing no facial palsy of any kind) and a multidisciplinary team of experts to design a grading scale of healthy asymmetry, to calculate how asymmetric is the subject's face according to it, and to label each participant's image manually. To our knowledge, no such dataset is available for the research community.

Future work

Analyzing facial paralysis from images requires new data and models to overcome the bias produced using healthy people photographs in the design, training, and testing of algorithms. Improvements to the methodology would require more publicly available samples from palsy patients with annotated data, including information about the lesion's severity, exhibited facial expression, felt emotion, and healthy controls to characterize the facial asymmetry in non-palsy subjects.

Developing a complete facial paralysis system was out of the scope of this research. Still, as future work, such a system should start with detecting facial palsy in a face photograph (or video). Then, a classifier trained on healthy people or palsy patients could be selected for the following tasks depending on the detection result. The next module should include the analysis of facial gestures to recognize facial expressions to detect emotion or any other application involving the face. To date, we proposed a system to detect facial palsy in photographs and an approach to classify facial gestures. Annotated data, including the patient's emotional state, is needed to go further.

The design of a mobile application to diagnose facial palsy at home is also desirable to improve the rate of early diagnosis. Furthermore, developing an application to monitor the treatment and improvement of the patient easily would represent a milestone in the healthcare sector, with many other benefits such as cost reductions and resource optimization.

SCIENTIFIC PRODUCTION

As a result of our research on diagnosing facial paralysis using computer vision techniques and machine learning algorithms, three articles were published in different scientific journals with a JCR impact factor. The publications are listed next:

- Parra-Dominguez, G. S.; Sanchez-Yanez, R. E.; Garcia-Capulin, C. H. Facial Paralysis Detection on Images Using Key Point Analysis. Appl. Sci. 2021, 11, 2435. https://doi.org/10.3390/app11052435

- Parra-Dominguez, G.S.; Sanchez-Yanez, R.E.; Garcia-Capulin, C.H. Towards Facial Gesture Recognition in Photographs of Patients with Facial Palsy. Healthcare 2022, 10, 659. https://doi.org/10.3390/healthcare10040659

- Parra-Dominguez, G.S.; Garcia-Capulin, C.H.; Sanchez-Yanez, R.E. Automatic Facial Palsy Diagnosis as a Classification Problem Using Regional Information Extracted from a Photograph. Diagnostics 2022, 12, 1528. https://doi.org/10.3390/diagnostics12071528

REFERENCES

[1] O. O. Abayomi-Alli, R. Damaševičius, R. Maskeliūnas, and S. Misra. Few-shot learning with a novel voronoi tessellation-based image augmentation method for facial palsy detection. *Electronics*, 10(8), 2021.

[2] E. Alpaydin. *Introduction to Machine Learning*. The MIT Press, 2 edition, 2010.

[3] A.-R. F. F. (ARFF). https://www.cs.waikato.ac.nz/ml/weka/arff.html, 4 2002. Accessed: 2021-03-22.

[4] A. Bandini, S. Rezaei, D. L. Guarin, M. Kulkarni, D. Lim, M. I. Boulos, L. Zinman, Y. Yunusova, and B. Taati. A new dataset for facial motion analysis in individuals with neurological disorders. *IEEE Journal of Biomedical and Health Informatics*, 25(4):1111–1119, 2020.

[5] J. Barbosa, K. Lee, S. Lee, B. Lodhi, J.-G. Cho, W.-K. Seo, and J. Kang. Efficient quantitative assessment of facial paralysis using iris segmentation and active contour-based key points detection with hybrid classifier. *BMC Med Imaging*, 16(23), 2016.

[6] J. Barbosa, W.-K. Seo, and J. Kang. paraFaceTest: an ensemble of regression tree-based facial features extraction for efficient facial paralysis classification. *BMC Med Imaging*, 19(30), 4 2019.

[7] C. M. Bishop. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.

[8] J. C. Borod and R. S. van Gelder. Introduction. volume 25, pages 135–139. 1990.

[9] J. Brownlee. How to tune a machine learning algorithm in weka, 2 2014. Accessed: 2022-01-18.

Chapter References

[10] C. Campomanes-Álvarez and B. R. Campomanes-Álvarez. Automatic facial expression recognition for the interaction of individuals with multiple disabilities. In *2021 International Conference on Applied Artificial Intelligence (ICAPAI)*, pages 1–6, Halden, Norway, 5 2021. IEEE.

[11] A. Carnevale. *Machine learning algorithms for facial gesture recognition: a first analysis based on event-driven sEMG acquisition*. PhD thesis, Politecnico di Torino, Torino, Italy, 3 2021.

[12] D. S. David and M. Samraj. A comprehensive survey of emotion recognition system in facial expression. *Artech Journal of Effective Research in Engineering and Technology*, 1:76–81, 2020.

[13] J. J. Greene, D. L. Guarin, J. Tavares, E. Fortier, M. Robinson, J. Dusseldorp, O. Quatela, N. Jowett, and T. Hadlock. The spectrum of facial palsy: the MEEI facial palsy photo and video standard set. *The Laryngoscope*, 130(1):32–37, 2020.

[14] D. L. Guarin, J. R. Dusseldorp, T. A. Hadlock, and N. Jowett. A machine learning approach for automated facial measurements in facial palsy. *JAMA Facial Plastic Surgery*, 20(4):335–337, 2018. PMID: 29543955.

[15] D. L. Guarin, Y. Yunusova, B. Taati, J. R. Dusseldorp, S. Mohan, J. Tavares, M. M. van Veen, E. Fortier, T. A. Hadlock, and N. Jowett. Toward an automatic system for computer-aided assessment in facial palsy. *Facial Plastic Surgery & Aesthetic Medicine*, 22(1):42–49, 2020. PMID: 32053425.

[16] Z. Guo, G. Dan, J. Xiang, J. Wang, W. Yang, H. Ding, O. Deussen, and Y. Zhou. An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis. *IEEE Journal of Biomedical and Health Informatics*, 22(3):835–841, 2018.

[17] Z. Guo, M. Shen, L. Duan, Y. Zhou, J. Xiang, H. Ding, S. Chen, O. Deussen, and G. Dan. Deep assessment process: Objective assessment process for unilateral peripheral facial paralysis via deep convolutional neural network. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 135–138, Melbourne, Australia, 4 2017.

[18] G.-S. J. Hsu, W.-F. Huang, and J.-H. Kang. Hierarchical network for facial palsy detection. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 693–699, Salt Lake City, UT, USA, 6 2018.

[19] G.-S. J. Hsu, J.-H. Kang, and W.-F. Huang. Deep hierarchical network with line segment learning for quantitative analysis of facial palsy. *IEEE Access*, 7:4833–4842, 2019.

[20] C. Jiang, J. Wu, W. Zhong, M. Wei, J. Tong, H. Yu, and L. Wang. Automatic facial paralysis assessment via computational image analysis. *Journal of Healthcare Engineering*, 2020, 2 2020.

[21] T. Kanade, J. F. Cohn, and Y. Tian. Comprehensive database for facial expression analysis. In *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, pages 46–53, Grenoble, France, 3 2000. IEEE.

[22] H. S. Kim, S. Y. Kim, Y. H. Kim, and K. S. Park. A smartphone-based automatic diagnosis system for facial nerve palsy. *Sensors*, 15(10):26756–26768, 2015.

[23] A.-M. Kuttenreich, G. F. Volk, O. Guntinas-Lichius, H. von Piekartz, and S. Heim. Facial emotion recognition in patients with post-paralytic facial synkinesis-a present competence. *Diagnostics*, 12(5), 2022.

[24] M. Leo, P. Carcagnì, P. L. Mazzeo, P. Spagnolo, D. Cazzato, and C. Distante. Analysis of facial information for healthcare applications: A survey on computer vision-based approaches. *Information*, 11(3), 2020.

[25] W. Lin, K. Hasenstab, G. Moura Cunha, and A. Schwartzman. Comparison of hand-crafted features and convolutional neural networks for liver mr image adequacy assessment. *Scientific Reports*, 10, 11 2020.

[26] X. Liu, Y. Xia, H. Yu, J. Dong, M. Jian, and T. D. Pham. Region based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(10):2325–2332, 2020.

[27] J. Lou, H. Yu, and F.-Y. Wang. A review on automated facial nerve function assessment from visual face capture. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(2):488–497, 2020.

[28] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. The extended cohn-kanade dataset (ck+): A complete dataset for action unit and emotion-specified expression. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*, pages 94–101, San Francisco, CA, USA, 6 2010.

[29] R. Malka, M. Miller, D. Guarin, Z. Fullerton, T. Hadlock, and C. Banks. Reliability between in-person and still photograph assessment of facial function in facial paralysis using the eface facial grading system. *Facial Plastic Surgery & Aesthetic Medicine*, 23(5):344–349, 9 2020. PMID: 33325774.

[30] I. Matthews and S. Baker. Active appearance models revisited. *International Journal of Computer Vision*, 60(2):135–164, 2004.

Chapter References

[31] A. Mehrabian. Communication without words. *Psychology Today*, 2(4):53–56, 1968.

[32] M. Miller, T. Hadlock, E. Fortier, and D. L. Guarin. The auto-eface: Machine learning–enhanced program yields automated facial palsy assessment tool. *Plastic and Reconstructive Surgery*, 147(2):467–474, 2021.

[33] S. Mishra, R. Gupta, and S. K. Mishra. Facial expression recognition system (fers): A survey. In D. Mishra, R. Buyya, P. Mohapatra, and S. Patnaik, editors, *Intelligent and Cloud Computing*, pages 43–50, Singapore, 2021. Springer Singapore.

[34] L. Nanni, S. Ghidoni, and S. Brahnam. Handcrafted vs. non-handcrafted features for computer vision classification. *Pattern Recognition*, 71:158–172, 2017.

[35] F. Nonis, N. Dagnes, F. Marcolin, and E. Vezzetti. 3d approaches and challenges in facial expression recognition algorithms—a literature review. *Applied Sciences*, 9(18), 2019.

[36] J. Ostrofsky, D. J. Cohen, and A. Kozbelt. Objective versus subjective measures of face-drawing accuracy and their relations with perceptual constancies. *Psychology of Aesthetics, Creativity, and the Arts*, 8(4):486—-497, 2014.

[37] G. S. Parra-Dominguez, C. H. Garcia-Capulin, and R. E. Sanchez-Yanez. Automatic facial palsy diagnosis as a classification problem using regional information extracted from a photograph. *Diagnostics*, 12(7), 2022.

[38] G. S. Parra-Dominguez, R. E. Sanchez-Yanez, and C. H. Garcia-Capulin. Facial paralysis detection on images using key point analysis. *Applied Sciences*, 11(5), 2021.

[39] G. S. Parra-Dominguez, R. E. Sanchez-Yanez, and C. H. Garcia-Capulin. Towards facial gesture recognition in photographs of patients with facial palsy. *Healthcare*, 10(4), 2022.

[40] M. Sajid, T. Shafique, M. J. A. Baig, I. Riaz, S. Amin, and S. Manzoor. Automatic grading of palsy using asymmetrical facial features: A study complemented by new solutions. *Symmetry*, 10(7):242–254, 6 2018.

[41] A. Song, Z. Wu, X. Ding, Q. Hu, and X. Di. Neurologist standard classification of facial nerve paralysis with deep neural networks. *Future Internet*, 10(11), 2018.

[42] P. Srinivas and P. Mishra. Prediction of emotional score of the multiple faces of a photo frame through facial emotion recognition using the deep convolutional neural network. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, 12(10):1178–1186, 2021.

[43] J. Taeger, S. Bischoff, R. Hagen, and K. Rak. Development of a smartphone app for neuromuscular facial training. *HNO*, 68(2):79–85, 2020.

Chapter References

[44] T. Tasneem, A. Shome, and S. K. A. Hossain. A gaming approach in physical therapy for facial nerve paralysis patient. In *16th Int'l Conf. Computer and Information Technology*, pages 345–349, Khulna, Bangladesh, 3 2014.

[45] J. Thevenot, M. B. López, and A. Hadid. A survey on computer vision for assistive medical diagnosis from faces. *IEEE Journal of Biomedical and Health Informatics*, 22(5):1497–1511, 2018.

[46] T. Wang, S. Zhang, H. Yu, J. Dong, and L.-A. Liu. Automatic evaluation of the degree of facial nerve paralysis. *Multimedia Tools and Applications*, 75:11893—-11908, 2016.

[47] H. Wei and Z. Zhang. A survey of facial expression recognition based on deep learning. In *2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA)*, pages 90–94, 11 2020.

[48] I. H. Witten. More data mining with weka. class 5 – lesson 1. simple neural networks, 5 2014. Accessed: 2020-09-14.

[49] I. H. Witten. Advanced data mining with weka. class 3 – lesson 1. libsvm and liblinear, 6 2016. Accessed: 2022-01-18.

[50] I. H. Witten, E. Frank, and M. A. Hall. *Data mining: practical machine learning tools and techniques*. Morgan Kaufmann, Burlington, MA, USA, 3 edition, 2011.

[51] Y. Wu and Q. Ji. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, 127:115—142, 2019.

[52] C. Xu, C. Yan, M. Jiang, F. Alenezi, A. Alhudhaif, N. Alnaim, K. Polat, and W. Wu. A novel facial emotion recognition method for stress inference of facial nerve paralysis patients. *Expert Systems with Applications*, 197:116705, 2022.

[53] G. Yolcu, I. Oztel, S. Kazan, C. Oz, K. Palaniappan, T. E. Lever, and F. Bunyak. Facial expression recognition for monitoring neurological disorders based on convolutional neural network. *Multimedia Tools and Applications*, 78(22):31581–31603, 2019.

[54] R. Zhi, M. Liu, and D. Zhang. A comprehensive survey on automatic facial action unit analysis. *The Visual Computer*, 36(5):1067–1093, 2020.