*diagnostics*

MDPI

*Article*

# DeepSmile: Anomaly Detection Software for Facial Movement Assessment

**Eder A. Rodríguez Martínez [1,2,*,†], Olga Polezhaeva [1,3,†], Félix Marcellin [1,2], Émilien Colin [1,2,4], Lisa Boyaval [1,3], François-Régis Sarhan [1,2,5] and Stéphanie Dakpé [1,2,4,*]**

1 UR 7516 Laboratory CHIMERE, University of Picardie Jules Verne, 80039 Amiens, France
2 Institut Faire Faces, 80000 Amiens, France
3 Faculty of Odontology, University of Reims Champagne-Ardenne, 51097 Reims, France
4 Maxillofacial Surgery, CHU Amiens-Picardie, 80000 Amiens, France
5 Physiotherapy School, CHU Amiens-Picardie, 80000 Amiens, France
* Correspondence: eder.rodriguez@u-picardie.fr (E.A.R.M.); Dakpe.Stephanie@chu-amiens.fr (S.D.); Tel.: +33-(0)-22-08-90-48 (E.A.R.M.)
† These authors contributed equally to this work.

**Abstract:** Facial movements are crucial for human interaction because they provide relevant information on verbal and non-verbal communication and social interactions. From a clinical point of view, the analysis of facial movements is important for diagnosis, follow-up, drug therapy, and surgical treatment. Current methods of assessing facial palsy are either (i) objective but inaccurate, (ii) subjective and, thus, depending on the clinician's level of experience, or (iii) based on static data. To address the aforementioned problems, we implemented a deep learning algorithm to assess facial movements during smiling. Such a model was trained on a dataset that contains healthy smiles only following an anomaly detection strategy. Generally speaking, the degree of anomaly is computed by comparing the model's suggested healthy smile with the person's actual smile. The experimentation showed that the model successfully computed a high degree of anomaly when assessing the patients' smiles. Furthermore, a graphical user interface was developed to test its practical usage in a clinical routine. In conclusion, we present a deep learning model, implemented on open-source software, designed to help clinicians to assess facial movements.

**Keywords:** anomaly detection; deep learning; long-short term memory; facial paralysis

check for updates

**Citation:** Rodríguez Martínez, E.A.; Polezhaeva, O.; Marcellin, F.; Colin, É.; Boyaval, L.; Sarhan, F.-R.; Dakpé, S. DeepSmile: Anomaly Detection Software for Facial Movement Assessment. *Diagnostics* **2023**, *13*, 254. https://doi.org/10.3390/diagnostics13020254

Academic Editor: Sameer Antani

Received: 13 December 2022
Revised: 3 January 2023
Accepted: 5 January 2023
Published: 10 January 2023

**Copyright:** © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

## 1. Introduction

According to Jones et al. [1], the human face is an important social stimulus since it provides relevant information about the observed person's age [2] and sex [2,3]. Moreover, facial expressions are responsible for conveying emotional messages, enhancing communication, and establishing links between individuals [4].

From a clinical point of view, the analysis of facial movements is relevant for diagnosis, care, and follow-up. Primarily, this analysis provides quantitative criteria that ensure an efficient follow-up for patients with facial pathology, e.g., facial paralysis [5]. Several techniques for assessing facial movement have been developed [6], with a view to quantifying the extent of facial paralysis and facilitating diagnosis and therapy, e.g., plastic or reconstructive surgery. Generally speaking, techniques for assessing facial movement can be categorized as either subjective or objective [7].

Subjective assessment techniques are based on the observation made by experienced clinicians; examples include the House–Brackmann facial nerve grading system [8], the Yanagihara facial nerve grading system [9] and the Sunnybrook facial grading system [10]. These methods rely on the graded observation of specific movements. Hence, the method's level of repeatability can be criticized [11–14].

Objective techniques are based on the use of sensors to quantify and assess facial movements. Most of these techniques can be automated to some degree, to make them quicker to administer and to reduce variations. However, the signals generated by sensors can be difficult to interpret. Objective assessment techniques can be subdivided into four groups: electromyography (EMG), computer vision, three-dimensional (3D) imaging, and optical motion capture.

The EMG technique consists of measuring the muscle electric response to nerve stimulation. This technique requires one to puncture the small needles of the sensor into a specific facial muscle [15]. Once the sensor is placed, the patient is asked to exercise the muscle to read the electric signals. A non-invasive alternative to EMG is the surface EMG [16], which uses patches of electrodes instead of needles. However, uniformly placing the patches is a hard practice, and the signal is sensitive to external interference [7]. Electroneuronography is another alternative to EMG that consists of comparing distal facial muscle response to electrical impulses. By applying electrical stimulation to the facial nerve trunk, this method records compound muscle action potential as electric signals [17]. Nevertheless, some studies have invalidated distal muscle comparison when assessing facial palsy [18].

Computer vision techniques can be further divided into sparse and dense techniques. The former one leverages on face recognition to automatically place virtual landmarks on the image [19]. Then, some machine learning techniques, such as an ensemble of regression trees [20], or support vector machines [21], are used to classify different levels of facial paralysis based on asymmetry features. These techniques are capable to assess facial palsy [22,23] and social perception [1]. When dealing with the sequence of images, dense techniques are based on optical flow, which describes the face movement in the image space [24–26]. Otherwise, when dealing with single images, the assessment can be defined as a classification task, performed by a Convolutional Neural Network (CNN) [27], where asymmetry is used as a feature [28]. However, these techniques use each pixel in the image to predict or classify variants of facial palsy. Although computer vision techniques are faster than the other objective techniques, they are inaccurate since they rely on metric estimations defined on the image space [26].

In practice, the analysis through 3D scans [6] can be used for planning future maxillo-facial surgery [29], soft tissues changes quantification [30] and facial mimic variations of patients before and after treatment [31]. This class can be further subdivided depending on the sensor: laser-based scanning, stereophotogrammetry, structured-light scanning, or RGB-D (red, green, blue-depth) sensors [32]. Depending on the sensor, this technique can provide dense information (RGB-D) or sparse information (stereophotogrammetry) if landmarks are placed on the face [33]. Although the stereophotogrammetric technique is the most accurate and reliable, its cost, size, and complexity, are often unsuitable for incorporation into clinical environments with limited availability of resources. One alternative solution to these disadvantages is the use of RGB-D sensors since they can collect accurate static and dynamic 3D facial scans; however, further improvements prior to their implementation are required [32]. In conclusion, the main drawbacks of 3D scan techniques are that most of them do not measure the motion of the face but rather a single 3D model [34], and some of their evaluations rely on subjective analysis.

Optical motion capture techniques use photogrammetry to track the movement of markers in 3D over time. The markers are placed in relevant zones in the face to measure the movement of the skin [35,36].

In [37], a statistical analysis is carried out to assess the presence of unilateral facial palsy before and after surgery. The analysis consists of comparing the trajectories of each pair of distal markers to measure their symmetry. One of the main advantages of motion capture systems is their precision which depends on the camera's configuration. However, current methods to assess facial movement rely on models that do not fully exploit sequential data [38,39].

Here, we present a deep learning model that assesses facial movement by exploiting optical motion capture data. Compared with the study in [39], we decided to use a deep learning model to represent sequential data collected from healthy movement rather than using a statistical model. Furthermore, we implemented the model through a graphical user interface (GUI), available at https://github.com/PolezhaevaOlga/Face_Motion_Capture (accessed on 12 December 2022), to test its practical usage in clinical routines. Generally speaking, the software evaluates the movement of the patient's lower third of the face when smiling and provides relevant information on possible diagnoses. Moreover, the software provides a global degree of anomaly that further complements the clinician's diagnostic. Specifically, the evaluation is carried out with a long short-term memory (LSTM) model [40] defined as an anomaly detector; thus, it compares the patient's movement with its own prediction. When evaluating a smile, the model is able to predict a healthy estimation of it because it was trained on a dataset that contained solely smiles from healthy volunteers following a one-class anomaly detection strategy [41]. The main objective of this study is to present a different approach, through a deep learning model, to objectively assess facial movement. The second objective is to provide the open-source software, containing the trained model, that served as the proof of concept of our approach.

Below, Section 2 covers the data acquisition and preprocessing as well as the mathematical definitions of the baseline model and our proposed deep learning model. Then, Section 3 details the training and evaluation of the proposed model, the comparison between the latter model and the baseline, and the GUI built to assist clinicians. Lastly, the model's advantages and limitations are discussed in Section 4, and the paper provides a brief conclusion in Section 5.

## 2. Materials and Methods

This Section focuses on depicting the processes of data acquisition and preprocessing, as well as the models to be compared. The pipeline's various steps are depicted in Figure 1. Firstly, the data are acquired from motion capture sessions. Secondly, five preprocessing steps are applied to the data. Thirdly, the preprocessed data are used to generate the dataset. Lastly, the dataset is used to compute, train, and evaluate the models.

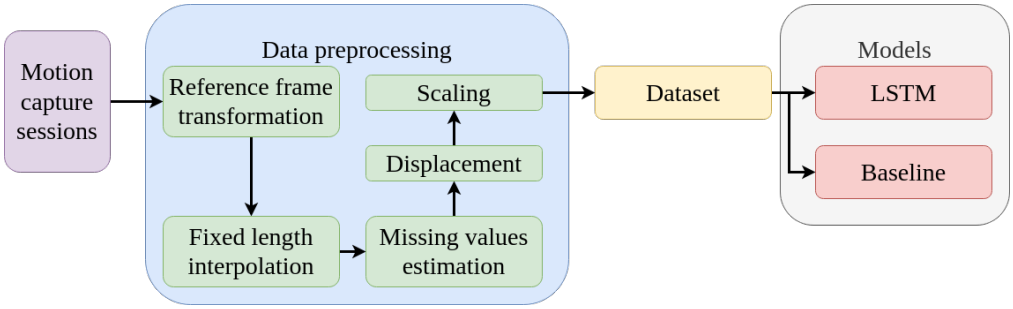



**Figure 1.** The pipeline's workflow.

### *2.1. Motion Capture Sessions*

The data used in this paper were acquired in several motion capture sessions. In each session, we recorded a neutral expression and five facial movements: gentle closure of the eyelids, forced closure of the eyelids, pronunciation of the [o] sound, pronunciation of the [p$\mu$] sound and broad smiling. The movements were recorded by tracking 105 reflective markers with an optical-passive motion capture system (Vicon Ltd., Oxford, UK). Prior to the sessions, a group of volunteers and patients were recruited. On the one hand, all volunteers were Caucasian men and women, between 18 and 30 years old, with no facial pathology known. On the other hand, patients were Caucasian men and women with facial pathology. For each volunteer and each patient, a 3D model of the face was generated using a stereo photogrammetry technique (Vectra M3 Imaging System, Canfield Scientific, Parsippany, NJ, USA). Later, a perforated mask was 3D printed (Form 2, Formlabs,

Somerville, MA, USA) so that the markers could be placed precisely on the face, using hypoallergenic glue. Moreover, rigid dental support made by a professional prosthetic defined the head's reference frame to disregard the head movements. During the session, each volunteer and each patient were asked to perform the 5 facial movements. Lastly, the movements were exported in a comma-separated value (csv) file format. The protocol used in the current study was approved by the Local Independent Ethics Committee (CPP Nord-Ouest II, Amiens, France) under references ID-RCB 2011-A00532-39/2011-23 and ID-RCB 2016-A00716-45/2016-55, registered at ClinicalTrials.gov (NCT02002572 and NCT03115203), and performed in accordance with the ethical standards of the 1964 Helsinki Declaration and its subsequent revisions. All participants provided written informed consent for study participation. For further details of the data acquisition process, please refer to [42].

Although the 5 movements previously described were recorded for each participant, this study focuses on the broad smile movement, as its production leads to large muscle displacements. For this purpose, 52 markers of the lower third of the face were chosen (highlighted in Figure 2) to be analyzed further. This consideration was implemented to reduce the complexity of the anomaly detection task, given the small number of csv samples on the dataset.

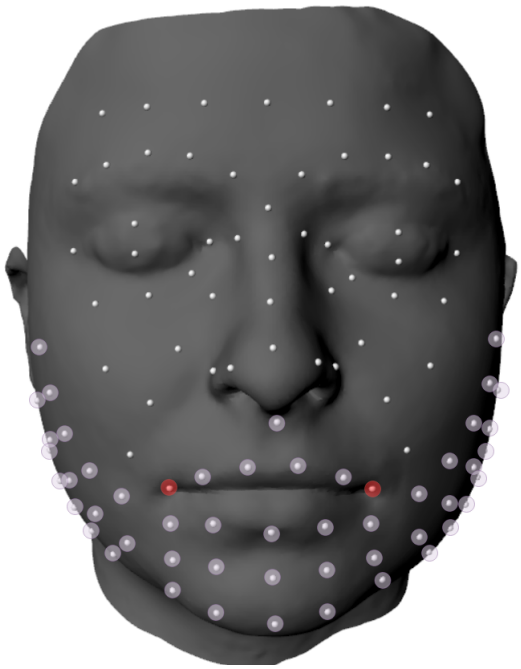


**Figure 2.** Set of markers virtually placed on a 3D model of the face. The selected markers appear highlighted in violet and red (commissure markers).

### *2.2. Data Preprocessing*

The motion capture system tracked the 3D ${}^{\mathcal{W}}\mathbf{P}_j(t) \in \mathbb{R}^3$ of the $j$-th marker, being $j = 1, 2, \ldots, 52$, over time in the world coordinate frame $\mathcal{W}$. The dental support is tracked to define the position and orientation of the head reference frame $\mathcal{H}$ in the Special Orthogonal group $SO(3)$. The start and the end of each smile are selected manually, and ${}^{\mathcal{W}}\mathbf{P}(t)$ was linearly interpolated, so the number of timesteps $|t| = 400$ remained constant.

A homogeneous transformation

$$ {}^{\mathcal{H}}\mathbf{P}' = {}^{\mathcal{H}}\mathbf{M}_{\mathcal{W}} {}^{\mathcal{W}}\mathbf{P}', \tag{1} $$

where $\mathbf{P}'$ is the homogeneous coordinate of $\mathbf{P}$ and ${}^{\mathcal{H}}\mathbf{M}_{\mathcal{W}}$ is the transformation matrix, is applied to express $\mathbf{P}$ in $\mathcal{H}$, i.e., ${}^{\mathcal{H}}\mathbf{P}$. This transformation disregards the head's translation and rotation, so the face's movements are precisely tracked.

Much as in [38,39,43], the markers' displacement was chosen as the feature vector that describes the smile. The markers' displacement, from their initial position to the current one, was defined by

$$D(\mathbf{P}_j(t)) = \|\mathbf{P}_j(0) - \mathbf{P}_j(t)\|, \tag{2}$$

where $\mathbf{P}_j(0)$ and $\mathbf{P}_j(t)$ are, respectively, the initial and current position of the $j$-th marker and $D \in \mathbb{R}$.

Once $D$ had been obtained, the *missing values* were estimated using linear regression [44]. However, this regression can be applied only if a small number of values are missing. In other cases, the csv file was not considered. Lastly, a scale transformation (also known as the min-max transformation) was applied to Equation (2) to normalize the feature vector:

$$S(D(\mathbf{P}_j)) = D(\mathbf{P_j})/\max(D(\mathbf{P}), D(\mathbf{P}_j)). \tag{3}$$

### 2.3. Datasets

Firstly, the training and validation datasets, which represent 70% of the smiles, were generated from healthy smiles only ($n = 25$ and $n = 7$, respectively). These datasets were used to compute the baseline and train the deep learning model. Then, the test dataset was generated from 4 healthy smiles (H-test) and 9 patients' smiles (P-test). The latter smiles were produced by 3 facial palsy patients and 1 facial transplantation patients. This dataset, which represents the 30% of the smiles, is used to evaluate both models. It is important to notice that the latter smiles were produced by volunteers that suffer from facial palsy. Specifically, all the healthy smiles samples were collected from 2014 to 2017 as previously described in [35,36,43,45–47].

### 2.4. Models

Two models are considered to evaluate the facial movement of healthy and pathological smiles: the baseline and the LSTM model. In brief, we compare the traditional approach [39], based on a statistical model, with a more complex model. Specifically, our LSTM model was a multivariate time series forecaster that follows a *seq-to-vector* [48] architecture. Although other deep learning models, such as multi-layer perceptron (MLP) [49], CNN, recurrent neural network (RNN) [50], and LSTM-CNN [51], were evaluated for this task, we decided only to include LSTM because its prediction better fitted the healthy smiles on the dataset. Similarly to [52], LSTM showed to be the best anomaly detector using sequential data by computing lower errors than other deep learning models.

#### 2.4.1. Baseline

The baseline model computes a single smile as the average of the markers' scaled displacement

$$B(\mathbf{P}) = \frac{1}{n}\sum S(D(\mathbf{P})), \tag{4}$$

where $n$ is the number of smiles ($n = 32$, see Section 2.3). In other words, the baseline can be roughly interpreted as the average smile of the training and validation datasets. Figure 3 shows two examples of the markers' average scaled displacement.

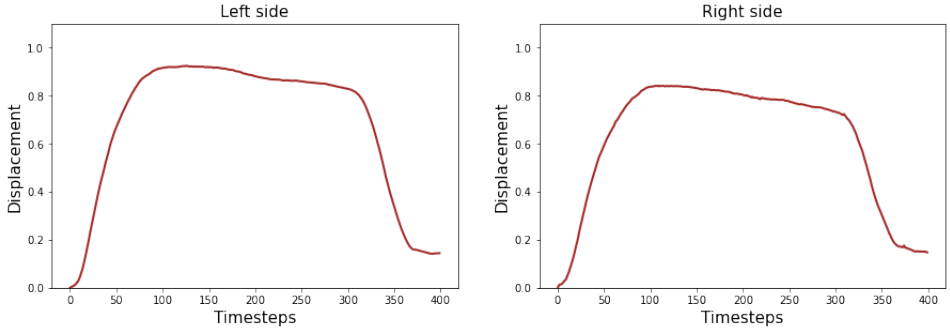



**Figure 3.** Average of the scaled displacement for the commissure markers.

#### 2.4.2. Long-Short Term Memory

An LSTM model was chosen to predict the smile because it had given a good level of performance for expression recognition by leveraging on sequential data [53]. Moreover, this model has outperformed other deep learning models as a dynamic and time-variant anomaly detector [52].

The LSTM cell [54], displayed in Figure 4, is defined as follows:

$$\begin{aligned}
\mathbf{i}_{(t)} &= \sigma(\mathbf{W}_{xi}^{\top}\mathbf{x}_{(t)} + \mathbf{W}_{hi}^{\top}\mathbf{h}_{(t-1)} + \mathbf{b}_i) \\
\mathbf{f}_{(t)} &= \sigma(\mathbf{W}_{xf}^{\top}\mathbf{x}_{(t)} + \mathbf{W}_{hf}^{\top}\mathbf{h}_{(t-1)} + \mathbf{b}_f) \\
\mathbf{o}_{(t)} &= \sigma(\mathbf{W}_{xo}^{\top}\mathbf{x}_{(t)} + \mathbf{W}_{ho}^{\top}\mathbf{h}_{(t-1)} + \mathbf{b}_o) \\
\mathbf{g}_{(t)} &= \tanh(\mathbf{W}_{xg}^{\top}\mathbf{x}_{(t)} + \mathbf{W}_{hg}^{\top}\mathbf{h}_{(t-1)} + \mathbf{b}_g) \\
\mathbf{c}_{(t)} &= \mathbf{f}_{(t)} \otimes \mathbf{c}_{(t-1)} + \mathbf{i}_{(t)} \otimes \mathbf{g}_{(t)} \\
\mathbf{y}_{(t)} &= \mathbf{h}_{(t)} = \mathbf{o}_{(t)} \otimes \tanh(\mathbf{c}_{(t)})
\end{aligned} \tag{5}$$

where $\mathbf{g}$ is the gate, $\mathbf{f}$, $\mathbf{i}$, and $\mathbf{o}$ are the controller of the forget, input, and output gates, respectively, $\mathbf{h}$ is the hidden state, $\mathbf{c}$ is the cell, $\mathbf{x}$ is the feature vector and

$$\begin{aligned}
\sigma(z) &= (1 + e^{-z})^{-1} \\
\tanh(z) &= \frac{e^{z} - e^{-z}}{e^{z} + e^{-z}}
\end{aligned} \tag{6}$$

are the logistic sigmoid and hyperbolic tangent functions, respectively, with $z \in \mathbb{R}$.

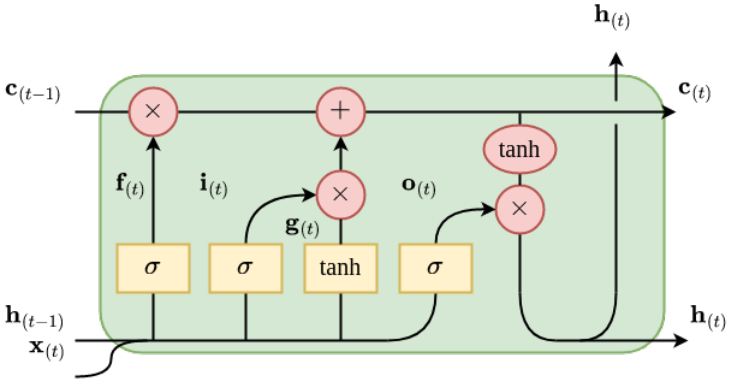



**Figure 4.** Representation of an LSTM cell.

In the view of the time series forecaster architecture, the mean square error (MSE):

$$\text{MSE} = \sum \frac{(\hat{\mathbf{y}} - \mathbf{y})^2}{n}, \tag{7}$$

where $\mathbf{y}$ and $\hat{\mathbf{y}}$ are the real and the predicted outputs, respectively, were chosen as the *loss function*. Similarly to [55], a windowing process was applied to $\mathbf{x}$. This process consists of generating a batch of *inputs* and *targets* by sliding a window through a vector. The size of the window, defined by $\texttt{window\_size} = \texttt{input\_size} + \texttt{target\_size}$, was experimentally set to $\texttt{input\_size} = 10$ and $\texttt{target\_size} = 1$, which results in $\texttt{window\_size} = 11$. Thus, resulting on an input size $n_x = (|j| \times \texttt{window\_size} \times |t|) = (52 \times 11 \times 400)$.

During training, the model's parameters were randomly initialized and then updated using the Adam algorithm [56] and the *stochastic gradient descent* method. Several learning rates, batch sizes, number of hidden units and number of hidden layers were experimentally tuned using the Keras Tuner library and the RandomSearch Tuner class for a maximum of 100 epochs with an early stop on the validation loss.

As an anomaly detector, the overall objective of the model was to minimize the degree of anomaly (cost) when evaluating the healthy smiles during training. To this end, the root mean square error (RMSE)

$$\text{RMSE} = \sqrt{\sum \frac{(\hat{\mathbf{y}} - \mathbf{y})^2}{n}}, \tag{8}$$

was selected as the cost function.

## 3. Results

This section presents the results of three experiments: (i) LSTM's training and evaluation, (ii) facial movement assessments comparison between LSTM and baseline, and (iii) LSTM model deployment on clinician diagnosis through a GUI. All the processes carried out in this study were executed on the same computer whose technical specifications are detailed in Table 1.

**Table 1.** System specifications.

| Hardware or Software | Settings |
|---|---|
| Model | Asus Strix G15 |
| Operative System | Windows home |
| GPU | NVIDIA GeForce RTX 3060 |
| Memory (RAM) | 16 GB |
| Processor | AMD Ryzen 7 5800H with Radeon Graphics 3.2 GHz |
| Memory storage capacity | 512 GB SSD |
| Programming languages | Python 3.10 |
| IDE | Jupyter notebook, Spyder |
| Libraries | Tensorflow, Pandas, Numpy, Tkinter, Scikit learn |

The goal of the training is to find the parameters that minimize the cost (or degree of anomaly) on the training dataset. The training's performance of the LSTM model is shown in Figure 5. The model achieved a performance of 0.0268, 0.0469, 0.0372, 0.0685, for the training, validation, H-test, and P-test datasets, respectively. Some LSTM model's predictions on the right and left commissure markers, for the H-test and P-Test, are presented in Figure 6.

The LSTM model's standard deviations were: 0.0119, 0.0274, 0.0174, and 0.0362 for the training, validation, H-Test, and P-Test datasets, respectively.

Next, the baseline and the LSTM models evaluate the test datasets (Table 2) by computing the degree of anomaly (Equation (8)). Figure 7 illustrates the evaluation, carried out by both models, on the left oral commissure marker of a smile that belongs to the H-Test.

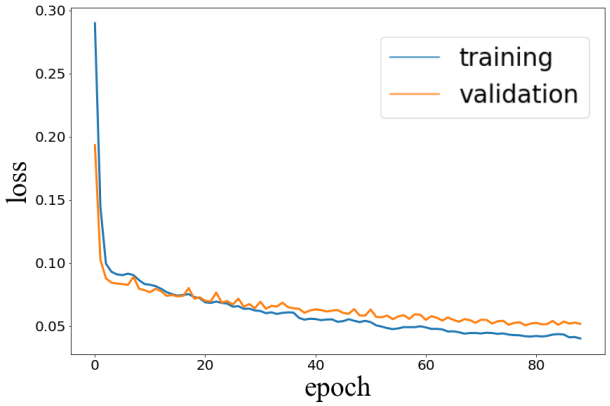
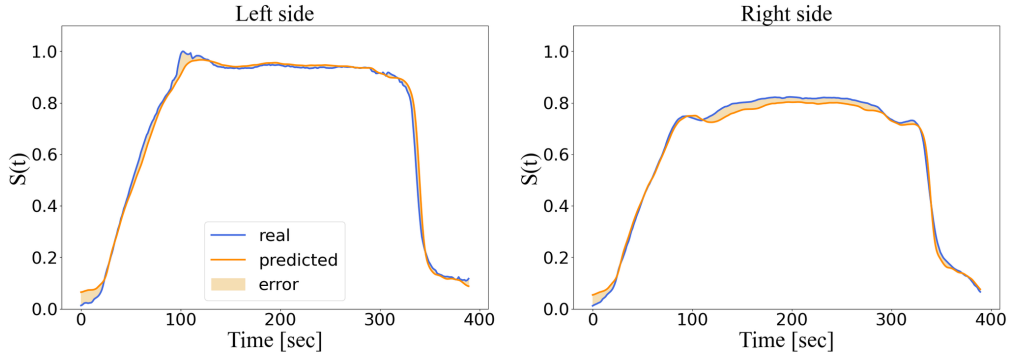
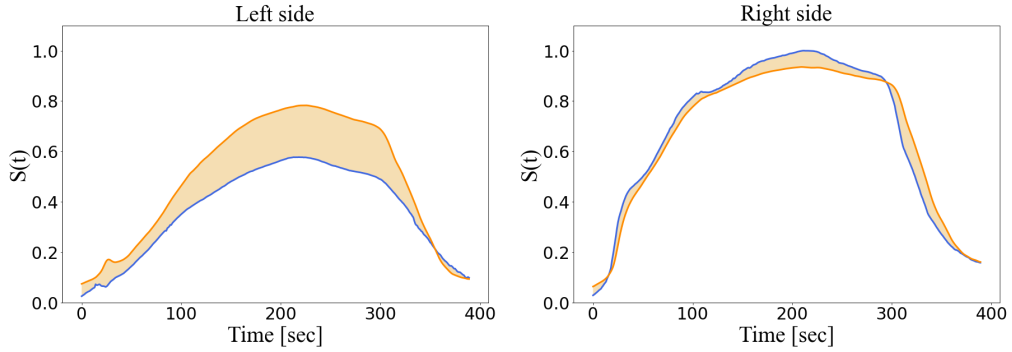


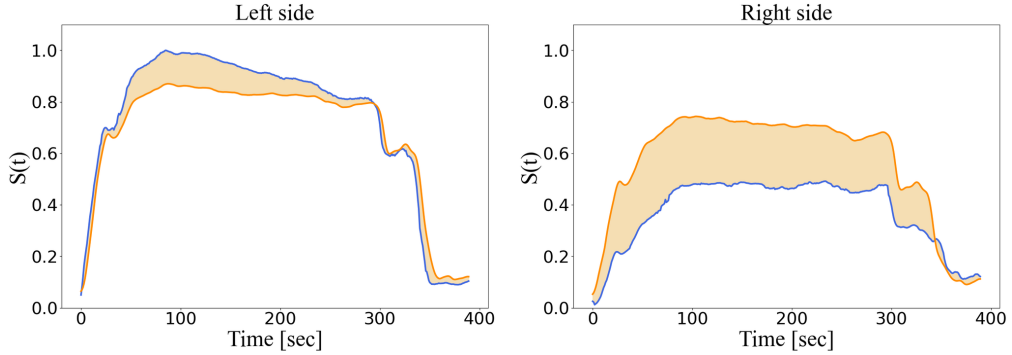

**Figure 5.** Performance of LSTM during training.




(**a**) Healthy control #1.




(**b**) Patient #1.

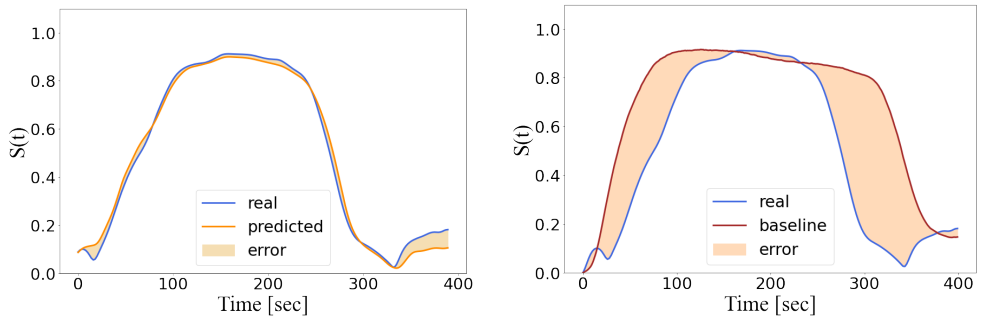



(**c**) Patient #2.

**Figure 6.** LSTM assessment of a relevant pair of distal markers on the test dataset.

**Table 2.** RMSE for the test dataset.

| Smile | H-1 | H-2 | H-3 | H-4 | P-1 | P-2 | P-3 |
|---|---|---|---|---|---|---|---|
| LSTM | 0.0454 | 0.0398 | 0.0527 | 0.0338 | 0.0895 | 0.0976 | 0.0625 |
| Baseline | 0.163 | 0.103 | 0.092 | 0.11 | 0.168 | 0.202 | 0.142 |




(**a**) LSTM-3 prediction. (**b**) Baseline: Average smile.

**Figure 7.** Predictions of the left oral commissure marker on a healthy smile.

In the view of the results presented here, we built a GUI for facial movement assessment to evaluate its deployment in clinical practice. Therefore, we created DeepSmile (Figure 8), which is an open-source software that assesses facial movement during a smile by running the trained LSTM model as an executable file. DeepSmile uses the csv file as the input to provide a report of its facial movement assessment with the following information: the patient's data; the performance of each marker (much as in Figure 6); a normalized and metric (in mm) degree of anomaly; a discrete indicator of relevant anomaly based on the normalized degree of anomaly relative to a predefined threshold.

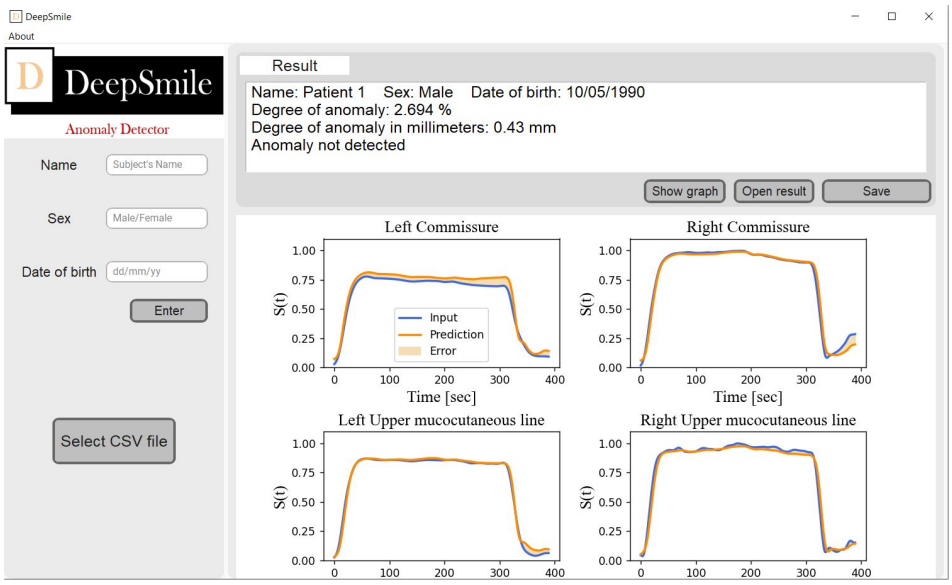



**Figure 8.** DeepSmile's graphical user interface.

To sum up, the DeepSmile's facial movement assessment can be carried out as follows. Firstly, the markers are placed on the patient's face (as in [42]) during the motion capture session. Secondly, the patient is asked to smile while the markers' locations are recorded. Thirdly, the markers are labeled (again as in [42]) and exported as a csv file. Lastly, the csv file is loaded into DeepSmile through a GUI and a report (based on the LSTM model's evaluation) is generated.

## 4. Discussion

In short, we have addressed a facial movement assessment problem as a non-linear optimization task rather than a classification task [57] to avoid a certain level of subjectivity, ex. the classes being defined by the grades of the House–Brackmann system. Furthermore, an anomaly detector is able to precisely track the degree of anomaly of a patient during follow-up which is expected to decrease over time if the rehabilition is successful. As stated in [52], LSTM models outperform other machine learning algorithms when sequential data is involved in anomaly detection tasks. On the one hand, the LSTM model is trained to minimize the degree of anomaly when evaluating the healthy smiles on the training dataset. Thus, its predictions are adapted to fit healthy smiles (Figure 6) and flag up a higher degree of anomalies for the patients' smiles. On the other hand, the baseline computes a single smile as the average of the training dataset. Hence, its assessment relies on a reference smile which does not fit other healthy smiles. Consequently, we consider that the LSTM's evaluations outperform those of the statistical model commonly used in the literature. Figure 7 illustrates an example where the LSTM's prediction is well adapted to evaluate healthy smiles whereas the baseline is not. Nonetheless, we consider that further investigation on LSTM variants, such as LSTM auto-encoders [58], should be carried out to exploit sequential data collected by motion capture system which might contain some null values.

Similarly to clinical diagnoses, the variation of the LSTM model's assessment is lower with healthy smiles than with patients' smiles. When evaluating the H-test, the RMSE was lower for the LSTM than for the baseline. Conversely, when evaluating the P-test, the RMSE was higher for the LSTM model than for the baseline (Table 2). This implies that the LSTM model is more robust than the baseline to differentiate a healthy smile from a pathological one. Furthermore, the degree of anomaly computed by the baseline on H-1 is higher than the corresponding one computed on P-3 (Table 2). In contrast, all the anomalies computed by the LSTM on the H-test are lower than those computed on the P-test. We, therefore, infer that a model that conveys temporal information is more suitable than a reference average healthy smile for assessing facial movements.

Compared with other deep learning models that leverage on qualitative grading system to evaluate facial pathologies ([59,60]), our LSTM model provides an objective assessment that exploits motion capture data. Furthermore, marker positions are more accurate than landmark positions because metric measurements are directly recorded rather than being estimated from the image space ([22]). Nevertheless, the cameras used by computer vision techniques are cheaper, and their installation requires less space than optical motion capture systems. Although the symmetry was not defined as a characteristic feature of healthy facial movement, ex. [22,61], we observed that our model outputs a low degree of anomaly when evaluating symmetric movement. Indeed, *Miller et al.* report that Emotrics computed more asymmetry in facial landmark positions than Auto E-Face when evaluating healthy volunteers [22] whereas our model did not present this phenomenon. It is interesting to note that the model predicted a displacement of similar magnitude when evaluating a patient with right-sided paralysis (Figure 6c); thus, the model computed a significant degree of anomaly on the paralyzed side. However, the model also computed a smaller degree of anomaly on the healthy side of the face. This reflects the phenomenon of compensation of the non-paralyzed side that we observe in clinical practice, but underlines the fact that, depending on the case, this side cannot really be considered as healthy, at least from the point of view of movement ([62,63]). This data objectively underlines the fact that the management of patients with facial paralysis concerns the whole face. One example is the use of botulinum toxin on the non-paralyzed side to induce a more symmetric mimicry ([64]).

In this study, one of the challenges was to train the model on a small number of healthy smiles. To address it, we opted to reduce the complexity of the anomaly detection task by experimentally selecting a small subset of markers. Indeed, the weights related to the selected markers were higher than the rest when training the deep learning model.

Another challenge related to our dataset is that, so far, our deep learning model is trained on healthy smiles produced by Caucasian volunteers only. Additionally, we observed in some patients a passive displacement of markers placed on the paralyzed side. Indeed, the markers were attracted in the direction of the non-paralyzed side by the soft tissue traction phenomenon. Therefore, we concluded that the displacement feature might not fully model abnormal movement, and another feature that includes the direction of movement should be considered.

The GUI we have developed calculates the level of anomaly in millimeters as well as in percentage for all the markers selected. It is, therefore, a global indicator of facial mobility, which can be used for the longitudinal follow-up of patients to quantify the evolution of paralysis. On the one hand, the interest of the GUI is to help clinicians to interpret facial movement, measured by a motion capture system, by displaying a global degree of anomaly. On the other hand, this principle leads to data simplification, whereas we could provide an enhanced diagnosis by fully exploiting another feature. In the future, it would, therefore, be interesting to diversify the algorithm so that it produces a group of scores, as in the Sunnybrook score [10], related to defined anatomical areas or particular functions. Moreover, this group of scores could be related to the relevant zones for each of the 5 movements of our complete protocol. Therefore, we must further curate our data, train more deep learning models on the other 4 movements of our motion capture protocol and explore other loss functions [65], meta-heuristic optimization algorithms [66] and features such as the markers' positions over time. Lastly, other deep learning architectures, such as auto-encoders [67], LSTM auto-encoders, graph neural networks (GNN) [68] or MLP, could be considered for facial movement assessment using motion capture data.

## 5. Conclusions

In this paper, we present an end-to-end deep learning framework to assess facial movement using optical motion capture data. Our deep learning model is able to detect abnormal movements because it was trained to predict healthy smiles via a one-class anomaly detection strategy. Compared with clinician-graded facial palsy evaluations, our novel technique is repeatable, reliable, objective, and not subjected to observer bias or human error; thus, it can further complement the clinician diagnosis. Furthermore, our training was deployed in a clinical environment, through a GUI, and it demonstrated its potential use, thus, validating the proof of concept. Although, further development is required.

**Author Contributions:** Conceptualization, E.A.R.M., F.-R.S. and S.D.; methodology, E.A.R.M.; software, O.P.; validation, E.A.R.M., O.P. and S.D.; formal analysis, E.A.R.M. and O.P.; investigation, E.A.R.M., F.-R.S., É.C. and S.D.; resources, F.-R.S., É.C. and S.D.; data curation, O.P., F.M. and L.B.; writing—original draft preparation, E.A.R.M.; writing—review and editing, E.A.R.M., O.P., É.C., F.-R.S. and S.D.; visualization, E.A.R.M.; supervision, E.A.R.M. and S.D.; project administration, F.-R.S., É.C. and S.D.; funding acquisition, É.C. and S.D. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded (as part of the FaceMoCap project) by the European Regional Development Fund (ERDF) N. 21004127, the Picardie Council, the Fondation. des Gueules Cassées N. 55-2022, and the FiGuRES EquipEx program N. ANR-10-EOPX-01-01.

**Institutional Review Board Statement:** The study protocols (healthy volunteers and patients) were approved by the Local Independent Ethics Committee (CPP Nord-Ouest II, Amiens, France) under references ID-RCB 2011-A00532-39/2011-23 and ID-RCB 2016-A00716-45/2016-55, registered at ClinicalTrials.gov (NCT02002572 and NCT03115203), and performed in accordance with the ethical standards of the 1964 Helsinki Declaration and its subsequent revisions.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Code available at https://github.com/PolezhaevaOlga/Face_Motion_Capture (accessed on 12 December 2022).

**Acknowledgments:** We thank Farouk Achakir for helping us during the reviewing process.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

## References


1. Jones, A.L.; Schild, C.; Jones, B.C. Facial metrics generated from manually and automatically placed image landmarks are highly correlated. *Evol. Hum. Behav.* **2021**, *42*, 186–193. [CrossRef]
2. Imai, T.; Okami, K. Facial cues to age perception using three-dimensional analysis. *PLoS ONE* **2019**, *14*, e0209639. [CrossRef] [PubMed]
3. Burton, A.M.; Bruce, V.; Dench, N. What's the difference between men and women? Evidence from facial measurement. *Perception* **1993**, *22*, 153–176. [CrossRef]
4. Bargiela-Chiappini, F.; Haugh, M. Face, communication and social interaction. *J. Politeness Res. Lang. Behav. Cult.* **2011**, *7*. [CrossRef]
5. Edward, L.; Dakpe, S.; Feissel, P.; Devauchelle, B.; Marin, F. Quantification of facial movements by motion capture. *Comput. Methods Biomech. Biomed. Eng.* **2012**, *15*, 259–260. [CrossRef] [PubMed]
6. Steinbacher, J.; Metz, A.A.; Tzou, C.H.J. 3D, 4D, Mobile APP, VR, AR, and MR Systems in Facial Palsy. In *Facial Palsy*; Springer: rBerlin/Heidelberg, Germany, 2021; pp. 405–425.
7. Jiang, C.; Wu, J.; Zhong, W.; Wei, M.; Tong, J.; Yu, H.; Wang, L. Automatic facial paralysis assessment via computational image analysis. *J. Healthc. Eng.* **2020**, *2020*, 2398542. [CrossRef]
8. House, J.W. Facial nerve grading systems. *Laryngoscope* **1983**, *93*, 1056–1069. [CrossRef]
9. Hato, N.; Fujiwara, T.; Gyo, K.; Yanagihara, N. Yanagihara facial nerve grading system as a prognostic tool in Bell's palsy. *Otol. Neurotol.* **2014**, *35*, 1669–1672. [CrossRef]
10. Neely, J.G.; Cherian, N.G.; Dickerson, C.B.; Nedzelski, J.M. Sunnybrook facial grading system: Reliability and criteria for grading. *Laryngoscope* **2010**, *120*, 1038–1045. [CrossRef]
11. Fattah, A.Y.; Gurusinghe, A.D.; Gavilan, J.; Hadlock, T.A.; Marcus, J.R.; Marres, H.; Nduka, C.C.; Slattery, W.H.; Snyder-Warwick, A.K. Facial nerve grading instruments: Systematic review of the literature and suggestion for uniformity. *Plast. Reconstr. Surg.* **2015**, *135*, 569–579. [CrossRef]
12. Revenaugh, P.C.; Smith, R.M.; Plitt, M.A.; Ishii, L.; Boahene, K.; Byrne, P.J. Use of objective metrics in dynamic facial reanimation: A systematic review. *JAMA Facial Plast. Surg.* **2018**, *20*, 501–508. [CrossRef] [PubMed]
13. Popat, H.; Richmond, S.; Benedikt, L.; Marshall, D.; Rosin, P.L. Quantitative analysis of facial movement—A review of three-dimensional imaging techniques. *Comput. Med. Imaging Graph.* **2009**, *33*, 377–383. [CrossRef] [PubMed]
14. Gaudin, R.A.; Robinson, M.; Banks, C.A.; Baiungo, J.; Jowett, N.; Hadlock, T.A. Emerging vs time-tested methods of facial grading among patients with facial paralysis. *JAMA Facial Plast. Surg.* **2016**, *18*, 251–257. [CrossRef] [PubMed]
15. Dalla Toffola, E.; Bossi, D.; Buonocore, M.; Montomoli, C.; Petrucci, L.; Alfonsi, E. Usefulness of BFB/EMG in facial palsy rehabilitation. *Disabil. Rehabil.* **2005**, *27*, 809–815. [CrossRef] [PubMed]
16. Kartush, J.M.; Lilly, D.J.; Kemink, J.L. Facial electroneurography: Clinical and experimental investigations. *Otolaryngol.—Head Neck Surg.* **1985**, *93*, 516–523. [CrossRef]
17. Lee, D.H. Clinical efficacy of electroneurography in acute facial paralysis. *J. Audiol. Otol.* **2016**, *20*, 8. [CrossRef] [PubMed]
18. Valls-Solé, J.; Montero, J. Movement disorders in patients with peripheral facial palsy. *Mov. Disord. Off. J. Mov. Disord.* **2003**, *18*, 1424–1435. [CrossRef]
19. King, D.E. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.* **2009**, *10*, 1755–1758.
20. Barbosa, J.; Seo, W.K.; Kang, J. paraFaceTest: An ensemble of regression tree-based facial features extraction for efficient facial paralysis classification. *BMC Med. Imaging* **2019**, *19*, 1–14. [CrossRef]
21. Wang, T.; Zhang, S.; Dong, J.; Liu, L.; Yu, H. Automatic evaluation of the degree of facial nerve paralysis. *Multimed. Tools Appl.* **2016**, *75*, 11893–11908. [CrossRef]
22. Miller, M.Q.; Hadlock, T.A.; Fortier, E.; Guarin, D.L. The Auto-eFACE: Machine learning–enhanced program yields automated facial palsy assessment tool. *Plast. Reconstr. Surg.* **2021**, *147*, 467–474. [CrossRef] [PubMed]
23. Guo, Z.; Dan, G.; Xiang, J.; Wang, J.; Yang, W.; Ding, H.; Deussen, O.; Zhou, Y. An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis. *IEEE J. Biomed. Health Inform.* **2017**, *22*, 835–841. [CrossRef] [PubMed]
24. Manohar, V.; Goldgof, D.; Sarkar, S.; Zhang, Y. Facial strain pattern as a soft forensic evidence. In Proceedings of the 2007 IEEE Workshop on Applications of Computer Vision (WACV'07), Austin, TX, USA, 21–22 February 2007; p. 42.
25. Manohar, V.; Shreve, M.; Goldgof, D.; Sarkar, S. Modeling facial skin motion properties in video and its application to matching faces across expressions. In Proceedings of the 2010 20th International Conference on Pattern Recognition, Istanbul, Turkey, 23–26 August 2010; pp. 2122–2125.
26. Guo, Z.; Shen, M.; Duan, L.; Zhou, Y.; Xiang, J.; Ding, H.; Chen, S.; Deussen, O.; Dan, G. Deep assessment process: Objective assessment process for unilateral peripheral facial paralysis via deep convolutional neural network. In Proceedings of the 2017 IEEE 14th international symposium on biomedical imaging (ISBI 2017), Melbourne, Australia, 18–21 April 2017; pp. 135–138.

27. LeCun, Y.; Bottou, L.; Bengio, Y.; Haffner, P. Gradient-based learning applied to document recognition. *Proc. IEEE* **1998**, *86*, 2278–2324. [CrossRef]
28. Song, A.; Wu, Z.; Ding, X.; Hu, Q.; Di, X. Neurologist standard classification of facial nerve paralysis with deep neural networks. *Future Internet* **2018**, *10*, 111. [CrossRef]
29. Adolphs, N.; Haberl, E.J.; Liu, W.; Keeve, E.; Menneking, H.; Hoffmeister, B. Virtual planning for craniomaxillofacial surgery–7 years of experience. *J. Cranio-Maxillofac. Surg.* **2014**, *42*, e289–e295. [CrossRef]
30. Verzé, L.; Bianchi, F.A.; Schellino, E.; Ramieri, G. Soft tissue changes after orthodontic surgical correction of jaws asymmetry evaluated by three-dimensional surface laser scanner. *J. Craniofacial Surg.* **2012**, *23*, 1448–1452. [CrossRef] [PubMed]
31. Bianchi, F.A.; Verze, L.; Ramieri, G. Facial mobility after bimaxillary surgery in class III patients: A three-dimensional study. In Proceedings of the XXI Congress f the European Association for Cranio-Maxillo-Facial Surgery. EACMFS2012, Dubrovnik, Croatia, 11–15 September 2012; p. 304.
32. Petrides, G.; Clark, J.R.; Low, H.; Lovell, N.; Eviston, T.J. Three-dimensional scanners for soft-tissue facial assessment in clinical practice. *J. Plast. Reconstr. Aesthetic Surg.* **2021**, *74*, 605–614. [CrossRef]
33. Sjögreen, L.; Lohmander, A.; Kiliaridis, S. Exploring quantitative methods for evaluation of lip function. *J. Oral Rehabil.* **2011**, *38*, 410–422. [CrossRef]
34. Ju, X.; Khambay, B.; O'Leary, E.; Al-Anezi, T.; Ayoub, A. Evaluation of the reproducibility of non-verbal facial animations. In Proceedings of the International Conference on Articulated Motion and Deformable Objects, Mallorca, Spain, 11–13 July 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 184–193.
35. Sarhan, F.R.; Mansour, K.; Neiva, C.; Godard, C.; Devauchelle, B.; Marin, F.; Dakpe, S. Apports d'une plateforme d'analyse du mouvement dans l'évaluation et la rééducation des atteintes de la mimique faciale. *Kinésithérapie Rev.* **2015**, *15*, 30–31. [CrossRef]
36. Sarhan, F.R.; Mansour, K.B.; Godard, C.; Neiva, C.; Devauchelle, B.; Marin, F.; Dakpe, S. Validation d'un protocole d'analyse quantifiée des mouvements de la mimique faciale. *Neurophysiol. Clin./Clin. Neurophysiol.* **2016**, *46*, 280. [CrossRef]
37. Sforza, C.; Frigerio, A.; Mapelli, A.; Mandelli, F.; Sidequersky, F.V.; Colombo, V.; Ferrario, V.F.; Biglioli, F. Facial movement before and after masseteric-facial nerves anastomosis: A three-dimensional optoelectronic pilot study. *J. Cranio-Maxillofac. Surg.* **2012**, *40*, 473–479. [CrossRef] [PubMed]
38. Sforza, C.; Frigerio, A.; Mapelli, A.; Tarabbia, F.; Annoni, I.; Colombo, V.; Latiff, M.; Ferreira, C.L.P.; Rabbiosi, D.; Sidequersky, F.V.; et al. Double-powered free gracilis muscle transfer for smile reanimation: A longitudinal optoelectronic study. *J. Plast. Reconstr. Aesthetic Surg.* **2015**, *68*, 930–939. [CrossRef]
39. Trotman, C.A.; Faraway, J.; Hadlock, T.; Banks, C.; Jowett, N.; Jung, H.J. Facial soft-tissue mobility: Baseline dynamics of patients with unilateral facial paralysis. *Plast. Reconstr. Surg. Glob. Open* **2018**, *6*, e1955. [CrossRef]
40. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef]
41. Chalapathy, R.; Menon, A.; Chawla, S. Anomaly Detection using One-Class Neural Networks. *arXiv* **2018**, arXiv:1704.06743v.
42. Sarhan, F.R. Quantification des mouvements de la mimique faciale par motion capture sur une population de volontaires sains. Ph.D. Thesis, University of Technology of Compiègne, Compiègne, France, 2017. Available online https://theses.fr/2017COMP2370 (accessed on 6 January 2023).
43. Dagnes, N.; Marcolin, F.; Vezzetti, E.; Sarhan, F.R.; Dakpé, S.; Marin, F.; Nonis, F.; Mansour, K.B. Optimal marker set assessment for motion capture of 3D mimic facial movements. *J. Biomech.* **2019**, *93*, 86–93. [CrossRef] [PubMed]
44. Sainani, K.L. Dealing with missing data. *PM&R* **2015**, *7*, 990–994.
45. Sarhan, F.R.; Olivetto, M.; Ben Mansour, K.; Neiva, C.; Colin, E.; Choteau, B.; Marie, J.P.; Testelin, S.; Marin, F.; Dakpé, S. Quantified analysis of facial movement, a reference for clinical applications. *J. Clin. Anat.* **2023**, *in press*.
46. Olivetto, M.; Sarhan, F.-R.; Mansour, K.B.; Marie, J.-P.; Marin, F.; Dakpe, S. Quantitative Analysis of Facial Palsy Based on 3D Motion Capture (SiMoVi-FaceMoCap Project). *Arch. Phys. Med. Rehabil.* **2019**, *100*, e112. [CrossRef]
47. Mansour, K.B.; Sarhan, F.; Neiva, C.; Godard, C.; Devauchelle, B.; Marin, F.; Dakpé, S. Analysis of mimic facial movements based on motion capture. *Comput. Methods Biomech. Biomed. Engin* **2014**, *17*, 78–79. [CrossRef]
48. Géron, A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2019.
49. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. *Learning Internal Representations by Error Propagation*; Parallel Distributed Processing: Explorations in the Microstructure of Cognition: Foundations; MIT Press: Cambridge, MA, USA, 1987; pp. 318–362.
50. Elman, J.L. Finding structure in time. *Cogn. Sci.* **1990**, *14*, 179–211. [CrossRef]
51. Canizo, M.; Triguero, I.; Conde, A.; Onieva, E. Multi-head CNN–RNN for multi-time series anomaly detection: An industrial case study. *Neurocomputing* **2019**, *363*, 246–260. [CrossRef]
52. Lindemann, B.; Maschler, B.; Sahlab, N.; Weyrich, M. A survey on anomaly detection for technical systems using LSTM networks. *Comput. Ind.* **2021**, *131*, 103498. [CrossRef]
53. Yu, Z.; Liu, G.; Liu, Q.; Deng, J. Spatio-temporal convolutional features with nested LSTM for facial expression recognition. *Neurocomputing* **2018**, *317*, 50–57. [CrossRef]
54. Yu, Y.; Si, X.; Hu, C.; Zhang, J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput.* **2019**, *31*, 1235–1270. [CrossRef] [PubMed]
55. Graves, A.; Schmidhuber, J. Framewise phoneme classification with bidirectional LSTM and other neural network architectures. *Neural Netw.* **2005**, *18*, 602–610. [CrossRef]

56. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
57. Gaber, A.; Taher, M.F.; Wahed, M.A.; Shalaby, N.M.; Gaber, S. Classification of facial paralysis based on machine learning techniques. *BioMed. Eng. Online* **2022**, *21*, 1–20. [CrossRef]
58. Nguyen, H.; Tran, K.P.; Thomassey, S.; Hamad, M. Forecasting and Anomaly Detection approaches using LSTM and LSTM Autoencoder techniques with the applications in supply chain management. *Int. J. Inf. Manag.* **2021**, *57*, 102282. [CrossRef]
59. Knoedler, L.; Baecher, H.; Kauke-Navarro, M.; Prantl, L.; Machens, H.G.; Scheuermann, P.; Palm, C.; Baumann, R.; Kehrer, A.; Panayi, A.C.; et al. Towards a Reliable and Rapid Automated Grading System in Facial Palsy Patients: Facial Palsy Surgery Meets Computer Science. *J. Clin. Med.* **2022**, *11*, 4998. [CrossRef]
60. Knoedler, L.; Miragall, M.; Kauke-Navarro, M.; Obed, D.; Bauer, M.; Tißler, P.; Prantl, L.; Machens, H.G.; Broer, P.N.; Baecher, H.; et al. A Ready-to-Use Grading Tool for Facial Palsy Examiners—Automated Grading System in Facial Palsy Patients Made Easy. *J. Pers. Med.* **2022**, *12*, 1739. [CrossRef]
61. 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. [CrossRef]
62. Jowett, N.; Malka, R.; Hadlock, T.A. Effect of weakening of ipsilateral depressor anguli oris on smile symmetry in postparalysis facial palsy. *JAMA Facial Plast. Surg.* **2017**, *19*, 29–33. [CrossRef] [PubMed]
63. Sahin, S.; Yaman, M.; Mungan, S.O.; Kiziltan, M.E. What happens in the other eye? Blink reflex alterations in contralateral side after facial palsy. *J. Clin. Neurophysiol.* **2009**, *26*, 454–457. [CrossRef] [PubMed]
64. de Sanctis Pecora, C.; Shitara, D. Botulinum toxin type a to improve facial symmetry in facial palsy: A practical guideline and clinical experience. *Toxins* **2021**, *13*, 159. [CrossRef] [PubMed]
65. Wang, Q.; Ma, Y.; Zhao, K.; Tian, Y. A comprehensive survey of loss functions in machine learning. *Ann. Data Sci.* **2022**, *9*, 187–212. [CrossRef]
66. Le-Duc, T.; Nguyen, Q.H.; Lee, J.; Nguyen-Xuan, H. Strengthening Gradient Descent by Sequential Motion Optimization for Deep Neural Networks. *IEEE Trans. Evol. Comput.* **2022**. [CrossRef]
67. Bengio, Y.; Lamblin, P.; Popovici, D.; Larochelle, H. Greedy layer-wise training of deep networks. In *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*; MIT Press: Cambridge, MA, USA, 2006.
68. Zheng, L.; Li, Z.; Li, J.; Li, Z.; Gao, J. AddGraph: Anomaly Detection in Dynamic Graph Using Attention-based Temporal GCN. In Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-1-19), Macao, China, 10–16 August 2019; pp. 4419–4425.


**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.