applied sciences

MDPI

Article

Advanced Analysis of 3D Kinect Data: Supervised Classification of Facial Nerve Function via Parallel Convolutional Neural Networks

Mohsen Shayestegan [1], Jan Kohout [2], Karel Štícha [2] and Jan Mareš [1,2,*]

[1] Faculty of Electrical Engineering and Informatics, University of Pardubice, Nam. Cs. Legii 565, 530 02 Pardubice, Czech Republic; mohsen.shayestegan@upce.cz
[2] Department of Computing and Control Engineering, University of Chemistry and Technology Prague, Technická 1905/5, 166 28 Praha 6, Czech Republic; jan.kohout@vscht.cz (J.K.); karel.sticha@vscht.cz (K.Š.)
* Correspondence: jan.mares@vscht.cz

Abstract: In this paper, we designed a methodology to classify facial nerve function after head and neck surgery. It is important to be able to observe the rehabilitation process objectively after a specific brain surgery, when patients are often affected by face palsy. The dataset that is used for classification problems in this study only contains 236 measurements of 127 patients of complex observations using the most commonly used House–Brackmann (HB) scale, which is based on the subjective opinion of the physician. Although there are several traditional evaluation methods for measuring facial paralysis, they still suffer from ignoring facial movement information. This plays an important role in the analysis of facial paralysis and limits the selection of useful facial features for the evaluation of facial paralysis. In this paper, we present a triple-path convolutional neural network (TPCNN) to evaluate the problem of mimetic muscle rehabilitation, which is observed by a Kinect stereovision camera. A system consisting of three modules for facial landmark measure computation and facial paralysis classification based on a parallel convolutional neural network structure is used to quantitatively assess the classification of facial nerve paralysis by considering facial features based on the region and the temporal variation of facial landmark sequences. The proposed deep network analyzes both the global and local facial movement features of a patient's face. These extracted high-level representations are then fused for the final evaluation of facial paralysis. The experimental results have verified the better performance of TPCNN compared to state-of-the-art deep learning networks.

Keywords: rehabilitation; House–Brackman scale; functional data analysis; multi class classification; deep learning; Kinect

Citation: Shayestegan, M.; Kohout, J.; Štícha, K.; Mareš, J. Advanced Analysis of 3D Kinect Data: Supervised Classification of Facial Nerve Function via Parallel Convolutional Neural Networks. *Appl. Sci.* **2022**, *12*, 5902. https://doi.org/10.3390/app12125902

Academic Editor: Baiba Vilne

Received: 22 April 2022
Accepted: 7 June 2022
Published: 9 June 2022

1. Introduction

Deep learning has emerged as a potential and promising tool in biomedical signal analysis. It is employed in early diagnosis, onco-surgery and rehabilitation. A number of systematic reviews can be found in the literature, starting with a review of the applications of deep learning in rehabilitation by [1] and ending with a review of assistive technologies for patients in [2]. A systematic review of deep learning methods in biomedicine can be found in [3]

We can also find references dealing with specific software tools (based on deep learning) that help in biomedical image analysis. For example, a deep learning approach has been used in breast tumor histopathological images analysis [4,5], where the authors introduce a convolutional neural network to detect tumor targets from pathological images. Deep learning algorithms have recently been shown to be reliable and time-efficient in segmenting pathological lungs and quantification of aeration of the chest [6]. Deep neural networks also play an important role in data augmentation for brain tumor detection in magnetic resonance imaging [7].

Surgical treatment of a specific brain tumor can result in damage to the facial nerve. This leads to either complete paresis or, at best, increased fatigue of the facial muscles. Consequently, the patient's daily life is affected by misinterpretation of their emotions through facial expressions [8]

1.1. Facial Nerve Function Analysis

Some evaluation methods are based on facial regions, and the facial nerve function is evaluated by comparing the difference of the texture and shape features between the two symmetric regions. For example, [9] proposed a classification approach that is based on Gabor feature and SVM. Ref. [10] quantitatively evaluated facial nerve function using Gabor and LBP features. In addition, several evaluation approaches are based on dynamic and 3D extracted features to solve the problems in these. For example, approaches based on the facial image only consider the facial asymmetry information and ignore the features of the movements of the facial muscle.

Ref. [11] employed a 3D model to quantitatively study facial movements by directing patients to perform facial activities. Furthermore, Ref. [12] used the extracted facial temperature distribution features in the particular facial regions to evaluate facial nerve function. From previous related research, we have found a few researchers who pay attention to the deep features of facial muscle movements, where conventional methods have mainly been applied to extract the artificial features from static facial images by considering the facial asymmetry.

Recently, deep learning approaches have had many successful applications in facial expression and recognition [13]. Several works are presented to evaluate facial nerve function based on deep learning methods. Automatic facial paralysis feature point detection methods based on deep convolutional networks have been proposed [14]. In [15], the authors employed the GoogleNet model [16] to drive transfer learning and reported good results. However, these approaches usually apply a classical network for facial paralysis analysis with few modifications to the original network. Therefore, the evaluation approach for facial nerve function based on deep learning has a more expansive opportunity for development.

1.2. Main Aims

Our work concentrates on a very special type of tumor in the inner ear (so called Vestibular Schwannoma) and the reconstruction and analysis of muscle rehabilitation after the surgical treatment. We focus on both: (i) data acquisition in a clinical environment; and (ii) advanced data analysis.

We previously introduced a methodology for the reconstruction and analysis of mimetic data [17], and to compare basic classification methods [18]. In these previous works, feature engineering and feature selection were applied to train by traditional algorithms. It has shown that the most algorithms have reached less than 50 percent accuracy.

In this work, we were inspired by the neural network with convolutional layers (CNN) architecture, and our goal was to do whole process necessary to achieve higher accuracy than previous models in our system. So, the whole process included the data processing with our raw dataset, augmentation, and design of a suitable deep learning algorithm for classification. Our goal was to make it more applicable in our system by improving the accuracy. We will now focus on the deep learning methodology to make full use of its capacity.

Compared with the existing evaluation approaches, the main contribution of TPCNN is that it evaluates the global and local movement features of the patient's face from their facial activities, which are fused for the evaluation of facial nerve function. Another contribution of this paper is that we have applied data augmentation methods for facial nerve function classification to increase our dataset size and to improve the accuracy of classification. No previous study has applied data augmentation methods for time series facial nerve function classification. For instance, Ref. [19] only compares jittering, permutation, scaling, and

time-warping using a Fully Convolutional Network (FCN) and a Residual Neural Network (ResNet).

2. Materials and Methods

During diagnosis, clinicians evaluate the severity of facial paralysis when the patients perform several standard facial actions. They also focus on the asymmetry that occurs around specific facial regions. Based on the patient's facial action, the asymmetry of the texture changes and the facial shape would be taken from the whole face, and the largest asymmetry of the changes appears around a particular facial region. For example, the largest asymmetry of changes occurs around the mouth when the patient purses their lips or bares their teeth.

Based on these facts, to evaluate facial paralysis, we proposed a triple-path CNN (TPCNN) to evaluate the high-level movement features of facial muscles from the facial diagnostic actions. The TPCNN employs one CNN sub-network to evaluate the global movement features from whole faces and another two CNN sub-networks to evaluate the local movement features from the relevant facial regions. The extracted features by TPCNN are fused for final evaluation.

2.1. Measurement Scheme

A typical patient is shown in Figure 1. Facial data in the form of facemarks (facial landmarks) were obtained during the patient's rehabilitation of mimetic muscles.

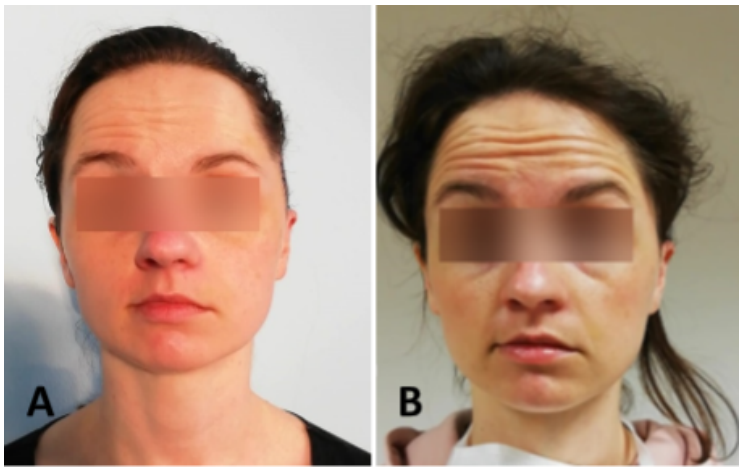

Figure 1. Typical patient (**A**)—HB6, (**B**)—HB3. Reader may notice movement improvement in the forehead part of the face—more pronounced wrinkles in (**B**).

The rehabilitation has been standardized by Prof. Janda [20] and consists of the mimetic exercises that are presented in Table 1.

Table 1. Hospital control rehabilitation exercises using Prof. Janda's methodology.

Order	Exercise	Order	Exercise
1	Raising eyebrows	6	Pursing the lips
2	Frowning	7	Blowing out the cheeks
3	Closing eyes tightly	8	Closing eyes tightly and baring the teeth
4	Smiling	9	Raising eyebrows and pursing the lips
5	Baring the teeth		

The physician evaluates each rehabilitation (set of nine exercises in Table 1) using the HB classification Table 2 [21].

Table 2. House-Brackmann (HB) scale for facial movements.

Grade	Characteristic
I	normal facial function in all areas
II	slight weakness on close inspection
III	obvious but not disfiguring
IV	obvious weakness and disfiguring asymmetry
V	only barely perceptible motion
VI	no movement

The patient is scanned at the same time with the Kinect v2 camera during the exercises. The camera computes the facemarks online from the acquired depth frames using an Active Appearance Model (AAM) based algorithm [22].

The space coordinates of the facemarks, the time, and name of the exercise are stored in data files for further analysis. These facemarks are 21 tracked facial points mapped as described in Table 3 and shown in Figure 2.

Table 3. Indices of points of interest (facemarks), *p* is an internal index number.

p	Kinect	Position	*p*	Kinect	Position
0	1104	left eye, bottom	11	849	left eyebrow, centre
1	241	left eye, top	12	18	nose tip
2	210	left eye, inner corner	13	8	mouth lower lip, centre-bottom
3	469	left eye, outer corner	14	91	mouth, left corner
4	346	left eyebrow, inner	15	687	mouth, right corner
5	222	left eyebrow, centre	16	19	mouth upper lip, centre-top
6	1090	right eye, bottom	17	4	chin, centre
7	731	right eye, top	18	28	forehead, centre
8	843	right eye, top	19	412	left cheek, centre
9	1117	right eye, outer	20	933	right cheek, centre
10	803	right eyebrow, inner			

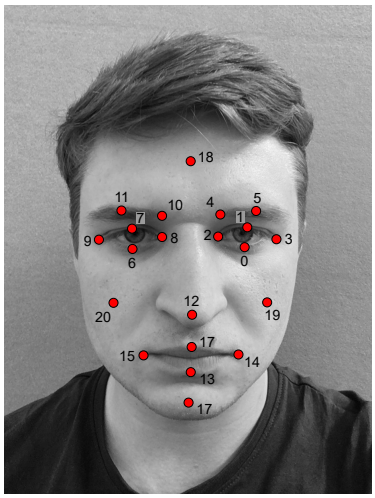

Figure 2. Points of interest illustration.

2.2. Dataset

The research was approved by the Ethics Committee of the University Hospital Královské Vinohrady Prague (EK-VP/4310120), where measurements were made. Each patient signed their informed consent to the research conditions.

Up to February 2022, the dataset contained 236 successful rehabilitations of 127 patients. More detailed information on the study cohort can be found in Table 4.

Table 4. Dataset overview.

Start Date	22 January 2019
End Date	20 January 2022
Number of Patients	127
Number of Sessions	236
Male	71
Female	56
Average Age (years)	58.4

The patients had been performing nine exercises using the mimetic muscles during the examination. The samples in Figure 3 are the amount of data that totalled around 2214 samples (236 measurements where each measurement included nine exercises). Time steps are expressed as the number of frames (sequences) recorded by the camera. Facial points are the number of features in every frame (time step) recorded from measurements that are the input of our network instead of images.

Figure 3 shows both the grade score distribution (classes) and the time steps distribution of our dataset. Data size is a very important factor in training deep neural networks. Larger datasets can help deep networks learn the model parameters better. A deep neural network trained with small datasets generally exhibits a poorer performance than that of conventional machine learning methods [18].

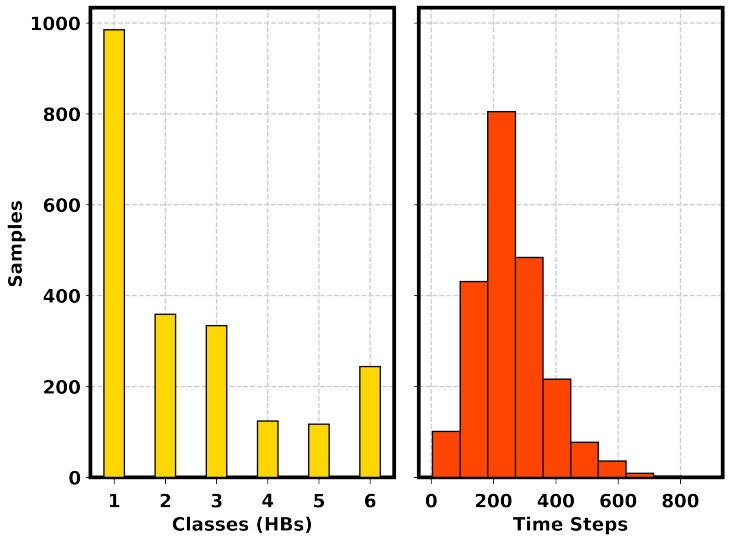

Figure 3. Classes and time step distribution in our dataset.

2.3. Data Preprocessing

A more detailed description of the registration technique that we used can be found in [17]. As can be seen in Figure 3, the class distribution is imbalanced and the time steps for each subject in our dataset vary widely. There are many more patients with class 1 (HB1) than patients with other classes, and classes 4 (HB4) and 5 (HB5) are worse with a minimum number of samples. To tackle unbalanced datasets, there are various techniques: under-sampling is not applicable in our case because of a small dataset; over-sampling can work, but there is a risk of model over-fitting, especially in our dataset, which is very small. To overcome these problems, we applied some techniques such as data augmentation and optimum weights for each class in the training process.

Furthermore, due to variation in time steps, all sequences are padded with zeros at the beginning according to the length of the longest sequence or a length chosen longer than the longest length, as shown in Figure 4.

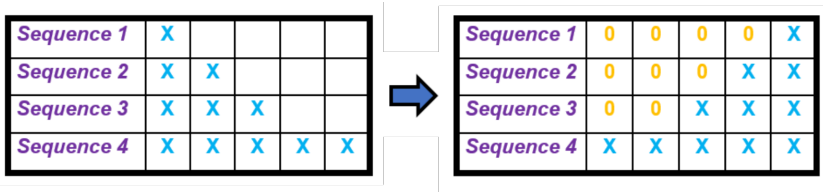

Figure 4. Pre-Sequence Padding.

Pre- and post-padding do not matter much to CNN because it tries to find the pattern in the given data, but, generally, pre-padding would be more useful when multiple types of neural networks are fused to execute a task [23]. In addition, as the accuracy is not consistent for different iterations in the unbalanced dataset, we applied the F1 score metric, which is the harmonic mean of precision and recall and is appropriate for an unbalanced data set.

Data Augmentation Strategies

A deep neural network has millions of parameters to learn, which means that it requires many iterations before it discovers the optimum values. If there are small volumes of data, the execution of many iterations can result in overfitting. A large dataset helps the network to avoid overfitting with a better performance. Data augmentation is a useful method for dealing with a small dataset without overfitting. Furthermore, data augmentation can boost the generalization ability of trained models by expanding the decision boundary of the model and decreasing overfitting [24]. The need for generalization is necessary for real-world data and can help designed networks, especially deep learning networks, to overcome datasets with imbalanced classes [25] or small datasets [26]. Most data augmentation strategies for time series classification are based on random transformations of the data, such as adding jittering [27], slicing [28], scaling, magnitude warping, and time dimension warping [29].

Jittering is one of the effective data augmentation methods [30], which can be defined as:

$$x^{'} = x_1 + \epsilon_1, \cdots + x_t + \epsilon_t, \cdots + x_T + \epsilon_T, \quad (1)$$

where ϵ is the Gaussian noise added to each time step.

In [30], the authors report that **rotation** data augmentation can increase accuracy when combined with other augmentation methods, where rotation is defined as:

$$x^{'} = Rx_1, \ldots, Rx_t, \ldots, Rx_T, \quad (2)$$

where R is a rotation matrix for flipping for univariate time series and angle for multivariate time series.

By a random scalar value, **scaling** can change the global intensity of a time series [30], where the scaling is defined as:

$$x^{'} = \alpha x_1, \ldots, \alpha x_t, \ldots, \alpha x_T, \quad (3)$$

where the scaling parameter α can be determined from a random value from a predefined set.

Magnitude warping is a technique that warps the magnitude of a signal by a smoothed curve [30], which is defined as:

$$x^{'} = \alpha_1 x_1, \ldots, \alpha_t x_t, \ldots, \alpha_T x_T, \tag{4}$$

where $\alpha_1, \ldots, \alpha_t, \ldots, \alpha_T$ is a sequence built by inserting a cubic spline $S(u)$ with knots $u = u_1, \ldots, u_i, \ldots, u_I$.

Slicing augments the data by slicing time steps off the ends of the pattern [30], where slicing is defined as:

$$x^{'} = x_\varphi, \ldots, x_t, \ldots, x_{W+\varphi}, \tag{5}$$

where W is the size of a window and φ is a random integer.

Time warping is the perturbation of a pattern in the temporal dimension using a smooth warping path [30], which is defined as:

$$x^{'} = x_{\tau(1)}, \ldots, x_{\tau(t)}, \ldots, x_{\tau(T)}, \tag{6}$$

where $\tau()$ is a warping function that warps the time steps based on a smooth curve.

Only a few works have applied data augmentation methods for time-series classification data. In this paper, the data augmentation strategies that we used include jittering, scaling, magnitude warping, time warping, slicing, and rotation as shown in Figure 5.

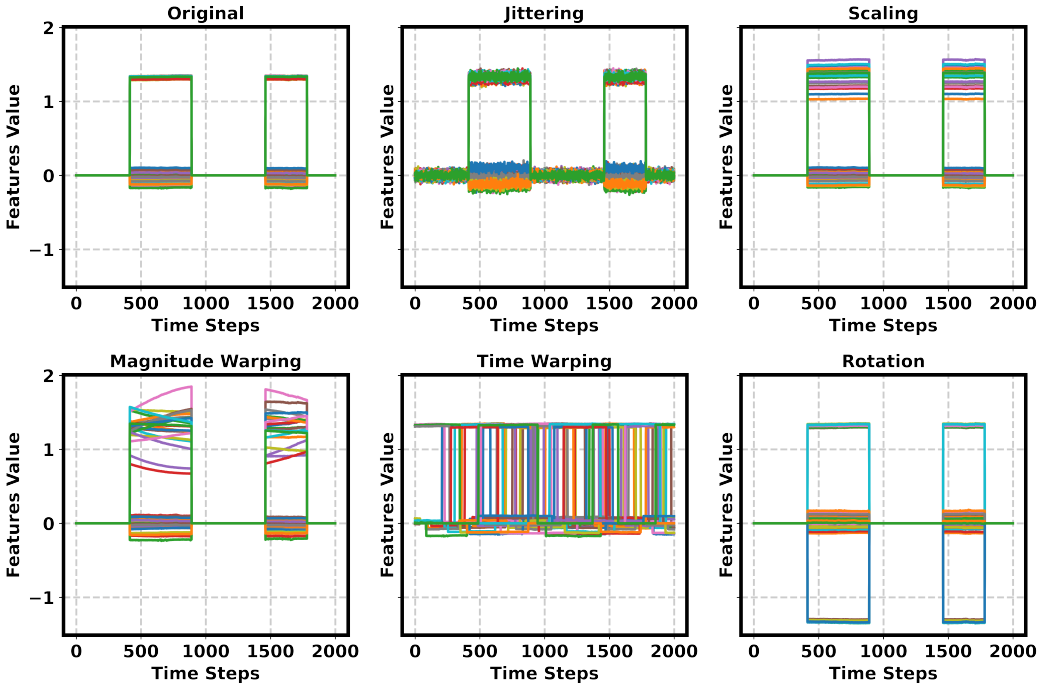

Figure 5. A sample of data augmentation strategies in our dataset.

2.4. Classification Methodology

Facial landmarks are an important part of facial expression analysis and facial recognition [31]. A patient who has a sign of facial nerve palsy will probably also have signs of deformation in important regions, such as the inability to close the eye, bare teeth, or the inability to purse the lips [32,33]. Facial nerve function can be recognized by extracting the position and distance between the salient points. Reference [34] demonstrated that the features of the mouth region had a direct correlation with the HB scores. Meanwhile, Ref. [35] used the boundaries of the eye region as a region-based feature for the classification of facial nerve function. In this paper, the facial landmarks of these two regions of interest, including the mouth region and the eye region, are selected for different subnetworks. The proposed parallel networks can learn the spatial features of the regional nerve function

from sequences of different facial expressions. We introduce two types of subnet networks for global and regional areas: a subnetwork for the whole face, and two subnetworks for the regions of the eyes and mouth. Figure 6 illustrates the framework of the proposed TPCNN.

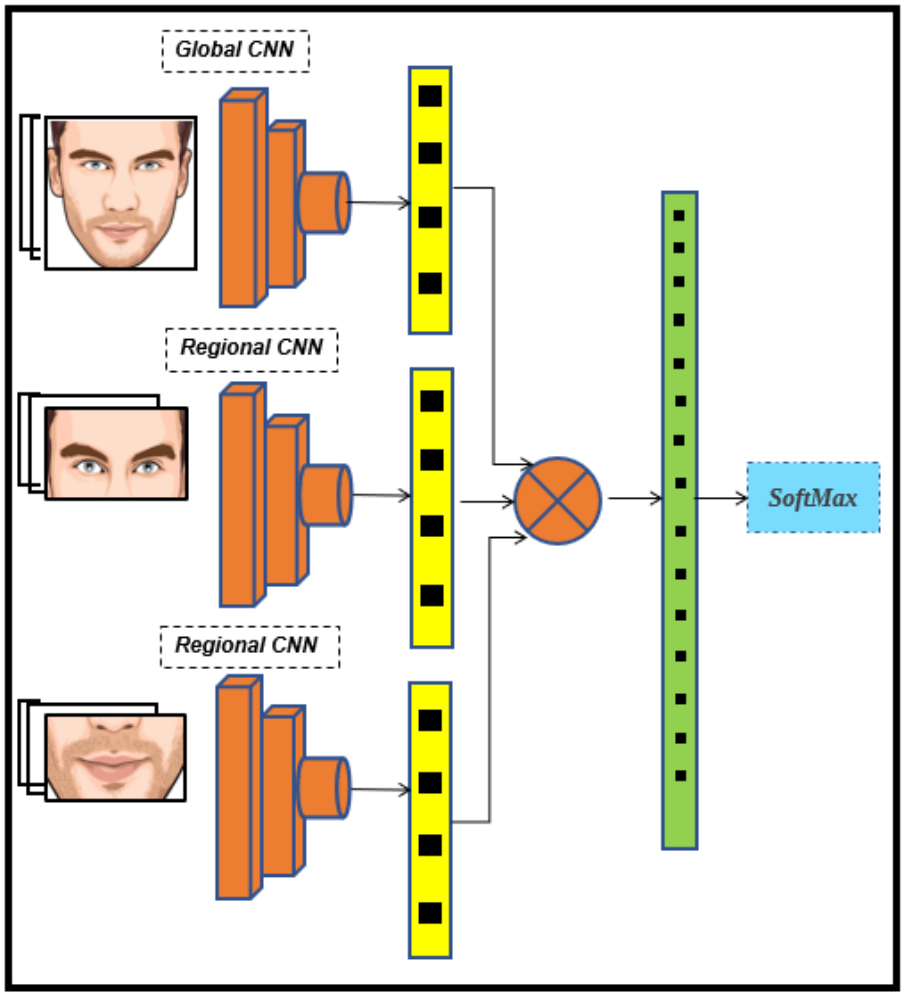

Figure 6. The overall framework of TPCNN.

For the region of the whole face, there are 21 key points in each sequence and each key point has three axes of data, this means that there is a total of 63 variables for each time step. Furthermore, each series of data has been partitioned into 891 time steps (891 are sequences), and there are also nine exercises for each labeled case. This means that the total time steps is around 891 $\times$ 9, or 8019 steps. Therefore, a row of data has (801 $\times$ 63 as is shown in Figure 7, the input of the subnetwork on the right side) or 505,197 elements. This is exactly how we loaded the data, where one sample is one window of the time series data, each window has 8019 time steps, and a time step has 63 features. The output of the model will be a six-element vector that contains a given window that belongs to each of the six classes. For the mouth and eye regions, the process is similar but with the number of features of their regions.

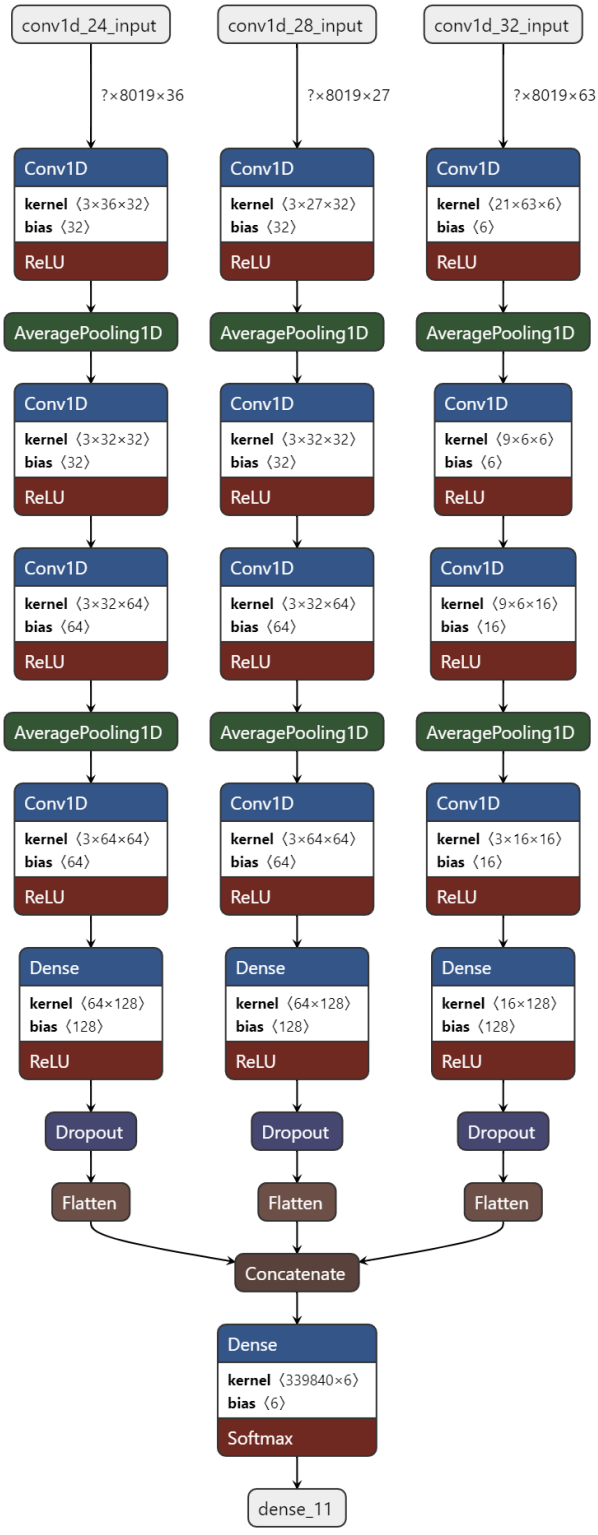

Figure 7. Graphical representation of TPCNN.

3. Results

To evaluate facial nerve function, we extracted and analyzed temporal features in changes in facial texture. We used a convolutional neural network to learn the dynamic features of facial movements to extract sequential features. In addition, we not only focused on the features of the movement on the overall faces, but also on the detailed facial regions corresponding to facial activities. Hence, a new network structure that is included with

the triple-path CNN was designed to extract the temporal features of the facial feature movements of the whole face and the detailed facial regions to evaluate facial nerve function. For the one path of the TPCNN, a recorded sequence of facial movements was used as input data.

As shown in Figure 7, there are three network subnets for each region of the face region. The network consists of single convolutional layers, average pooling layers, and a fully connected layer (FC). These piles of convolution layers are used to learn the asymmetry from low-level features to semantic-level features of the different sequences for each subject. These extracted features are then flattened to feature vectors in regional networks and in a global network. The ReLU function is used for the activation function of each subnet. In addition, the dropout is set to 0.2 to avoid the overfitting problem. Finally, the global temporal features and the local temporal features are concatenated to form a new vector of features for the final evaluation of facial nerve function.

The full features of the mouth, eye regions, and whole face sequences are used to analyze spatial position changes and temporal features among different facial movements. As shown in Figure 7, the feature vector for evaluation consists of global and regional features. In the last layer, weighted cross-entropy and soft-max are the loss and activation functions, respectively. The evaluation score according to the HB provided by a clinician is used as a basis for training, and the fused feature vectors of TPCNN are used to evaluate the severity of facial paralysis. As can be seen from Figure 3, the results of the analysis of the dataset show that the distribution of patients with different HB scores (classes) is quite unbalanced. Due to this problem, we introduce a class weight coefficient k_i, and the optimal weights w_i for each class in the training process are calculated using the following equations:

$$k_i = \sum_{i=1}^{n_c} \frac{1}{n_c n_{s_i}} \sum_{j=1}^{n_c} n_{s_j} \tag{7}$$

$$w_i = \sum_{i=1}^{n_c} k_i^2 . min(e^{k_i}, max(1.5k_i)), \tag{8}$$

where n_c is the number of classes, n_{s_i} or n_{s_j} are the number of samples in a particular class.

3.1. Experiment Set-Up

To test the models for the multiclassification tasks, the K-fold cross-validation process is applied in our experiment, where the value of k is determined as 5. This allows us to test on unseen facial feature samples while, accordingly, decreasing the possibility of overfitting to previously seen samples. The five-fold cross-validation technique divides the dataset into five subsets. Each subset is included as validation data and the other four subsets are used as training data. This can guarantee that the test data are not touched on in each evaluation. This procedure is repeated five times, and each class has the same probability for validation.

We modified the existing training algorithm to consider the skew distribution of classes. This required different weights to be assigned to the minority and the majority of samples. In the training stage, the disparity in weights will affect the classification of the class. The goal here is to compensate for the misclassification of classes with a few samples by assigning higher weights and lessening the weights for classes with higher samples.

In the training process, we assign a higher weight to the minority class in the training method. Therefore, training can focus on decreasing the error of the minority class. Furthermore, given that the F1 score is the key to measuring classification imbalance, the F1 score was used as an evaluation metric rather than accuracy because the F1 score is just a harmonic mean of recall and precision.

A batch-based ADAM algorithm is used to optimize the model. Adam optimization is a replacement method for stochastic gradient descent (SGD) for training deep learning models. For the training cycle, the batch size is set to 64 for each session. Finally, the best

model is chosen after 50 epochs to test the performance of the network in classifying the classes of facial nerve function sequences.

3.2. Evaluating the Experiment

The proposed method can solve the classification of facial nerve function with 21 landmark key points in the whole face that are detected by facial landmark detection. We also separate the eye region and the mouth region following the detection results for other networks. These features of the sequences are then fed into the TPCNN for classification purposes. The 21 point landmarks include 12 marks for the top region (e.g., four marks for the eyebrow and eight marks for two eyes) and nine marks for the down region (e.g., three marks for the nose and six marks for the lip and chin). In the model training procedure, facial, eye, and mouth features are used as inputs to the proposed parallel subnets. Each parallel subnet in TPCNN is a CNN structure with a low number of parameters and hidden layers, which can speed up the model's training procedure.

The first and second subnets are CNNs that focus on region-based, while the third CNN only focuses on the whole face. Performance metrics, such as the F1 score, are used to evaluate the performance of the proposed methodology, which are defined as follows:

$$\mathrm{F1} = \frac{2\mathrm{PR}}{\mathrm{P} + \mathrm{R}}, \tag{9}$$

where P stands for precision and R stands for recall. The precision represents the fraction of correctly predicted positive samples from the positive predicted samples, and the recall represents the ratio of real positive samples that are truly found by the model. The F1 score is the harmonic mean of precision and recall, which is a meaningful criterion (specifically in unbalanced data).

To demonstrate the prediction ability of the proposed method with parallel inputs, it has been compared with CNN-LSTM [36], ResNet [37], and FCN [37]. Table 5. As can be seen, most F1 scores are greater than 80% with the proposed method (see the Supplementary Material for more information). The classification average F1 score compared to different network structures in a five-fold cross-validation experiment is shown in Figure 8. The F1 score of the proposed model shows the percentage of correctly predicted samples. The proposed model is found to have achieved an excellent performance in the classification of facial nerve functions. The average rates in the five-fold cross-validation experiment with TPCNN are 88.31%, 81.18%, 81.76%, 83.73%, 85.02%, and 76.34%, as shown in Table 5 and Figure 8.

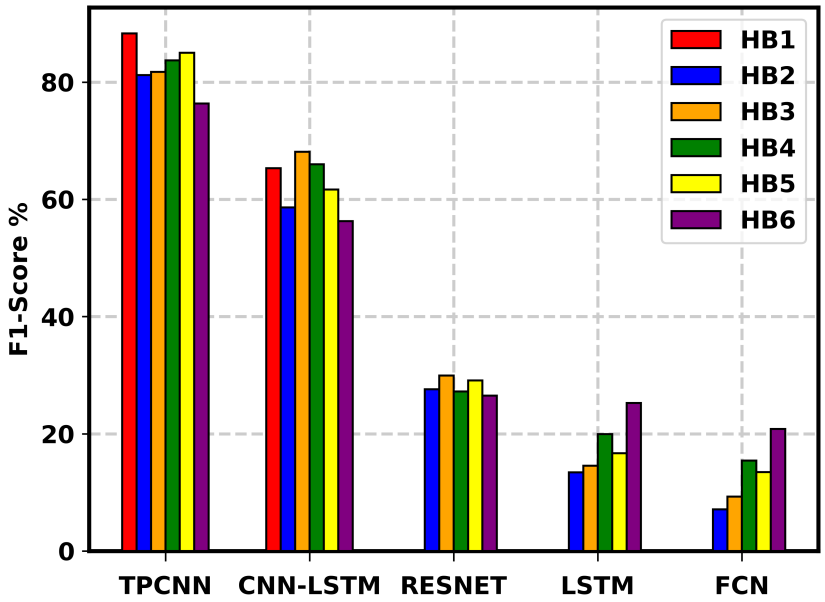

Figure 8. Average F1-score comparison.

Table 5. Average F1-Score comparison.

HB by a Clinician	TPCNN [%]	CNN-LSTM [%]	LSTM [%]	FCN [%]	ResNet [%]
1	88.31	65.32	0	0	0
2	81.18	58.63	13.42	7.1	27.58
3	81.76	68.10	14.56	9.26	29.93
4	83.73	65.96	19.93	15.44	27.22
5	85.02	61.69	16.67	13.46	29.11
6	76.34	56.30	25.24	20.82	26.51

4. Discussion

We also compare the performance of deep learning models with the test dataset, which is shown in Tables 6–10. The precision, recall, and F1 score are considered in comparison to the models by HB grades (classes). It is evident from the overall F1 score of the classification that the TPCNN classification method also produces the best results of all the selected deep learning methods on the test dataset.

When analyzing the confusion matrix, it is obvious that the proposed approach can predict the palsy classes (HBs) satisfactorily. The highest classification F1 score is 93%, which is attributed to face palsy class 1 (HB1), while the lowest classification F1 score is 67%, which is attributed to classes 4 (HB4) and 5 (HB5). The precision, recall, and F1 score metrics for our proposed method are shown in Table 6.

Precision and recall metrics describe the cases that correctly predicted overall positive predictions and observations, respectively. The precision and recall values of the proposed approach for face palsy class 1 (HB1) are 92% and 94%, compared to face palsy class 4 (HB4) with 100% and 50%, respectively. It can be seen from Tables 6–10 that the prediction F1 score of our classification model is better than that of the deep networks presented.

The proposed grading classification is robust and more effective because TPCNN has been used with automatic face features representation to achieve excellent classification accuracy across a range of facial palsy severities. By data augmentation, our model demonstrated that the classification model is appropriate for learning the most discriminative characteristics of the expected task. Furthermore, data augmentation makes our model capable of learning varying palsy severities, which can substitute the repeated measurements, and hence is useful in the repeatability of the classification method.

As can be seen in Figure 9, the proposed TPCNN also has the highest classification accuracy in the test set experiment, which is 89% compared to 69% and 13% for CNN-LSTM and LSTM, respectively. Other deep learning models have shown very poor performance in the test set. This is due to the proper design of TPCNN with merged global and local features for facial nerve function evaluation, rather than only operating on the extracted facial motion features of the entire face. These selected features of the facial nerve function characteristics for evaluation are compatible with the subjective evaluation of a clinician.

Table 6. Best TPCNN model result on the test set.

HB by a Clinician	Precision [%]	Recall [%]	F1-Score [%]
1	92	94	93
2	85	100	92
3	92	85	88
4	100	50	67
5	75	60	67
6	75	75	75

Table 7. Best CNN-LSTM model result on the test set.

HB by a Clinician	Precision [%]	Recall [%]	F1-Score [%]
1	100	67	80
2	64	82	72
3	69	69	69
4	33	50	40
5	30	60	40
6	43	75	55

Table 8. Best ResNet model result on the test set.

HB by a Clinician	Precision [%]	Recall [%]	F1-Score [%]
1	0	0	0
2	0	0	0
3	0	0	0
4	3	50	6
5	4	20	7
6	6	25	10

Table 9. Best LSTM model result on the test set.

HB by a Clinician	Precision [%]	Recall [%]	F1-Score [%]
1	0	0	0
2	0	0	0
3	67	31	42
4	6	50	11
5	13	80	22
6	0	0	0

Table 10. Best FCN model result on the test set.

HB by a Clinician	Precision [%]	Recall [%]	F1-Score [%]
1	0	0	0
2	0	0	0
3	5	8	6
4	0	0	0
5	0	0	0
6	50	25	33

Hypothetically, TPCNN (inspired by CNN architecture) is better at capturing local feature and neighborhood information more robustly, and it also considers features in the feature vector sequentially. In addition, the hyperparameters of the comparison methods are not tuned in this work and may be one reason for their failure cases. Zero padding should have a negative effect on its performance, where the TPCNN will identify patterns locally around the kernel. Therefore, the long sequences with zero padding should not have a negative effect on TPCNN. This will be added in our future work to investigate different deep learning methods in the imbalance dataset for failure cases.

We also provide a brief evaluation to illustrate the effect of architecture hyperparameters on the floating point operations (FLOPs) consumption to compare with the presented

methods. The performance of deep neural networks is highly dependent on the complexity of the model, which is measured by the size of a model's parameters multiplied by accumulating operations (MAC) or the number of floating-point operations. FLOPs are often used to illustrate how many operations are required to run a single instance of a given model. MACs contain multiplication and addition, so they can be counted as two separate floating-point operations. In other words, the Macs are approximately half of the FLOPs by ignoring bias terms, as the number of bias addition operations is much fewer than that of the MACs. A general trend is that the larger the model, the higher the accuracy it can achieve in a given task. Table 11 shows the comparison of the model parameters, MACs, and FLOPs of the methods presented. For the proposed model, the total number of FLOPs is given as 8480 K. As shown in the table, the LSTM variant requires the least number of model parameters, FLOPs, and MACs, but it is among low-accuracy methods. The proposed model has shown a larger number of parameter sizes, FLOPs and MACs, but, significantly, the highest accuracy among the other methods.

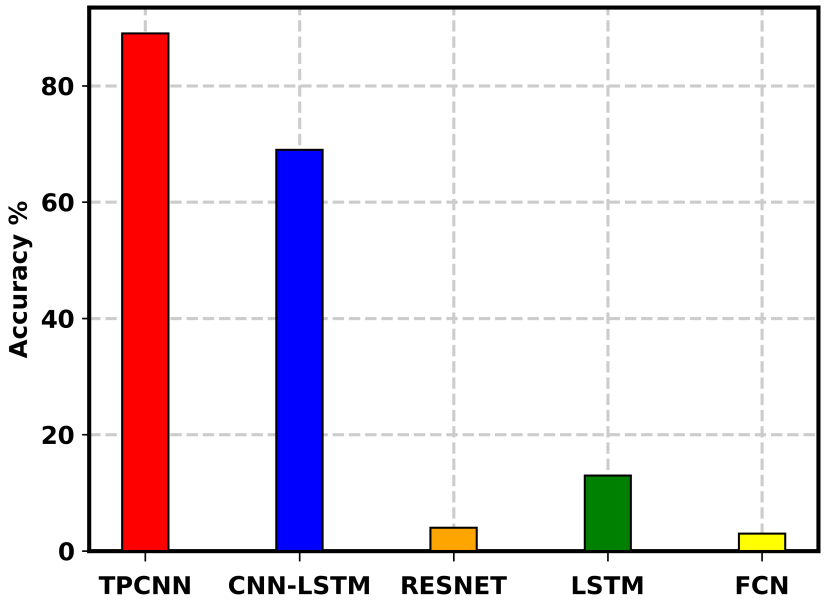

Figure 9. Accuracy comparison.

Table 11. Metrics of model efficiency.

Models	Parameters [K]	MACs [K]	FLOPs [K]	Accuracy [%]
TPCNN	2117	4240	8480	89
CNN-LSTM	1217	2470	4940	69
FCN	330	657.37	1314.74	3
ResNet	151	298.61	597.22	4
LSTM	42	115.71	231.42	13

5. Conclusions

In this paper, a region-based parallel network model is introduced to classify the facial nerve function of sequences based on facial regions. Due to an imbalanced data sample, different time step length, and lack of enough data samples for training, we applied techniques such as prepadding, data augmentation, F1-score metric evaluation, and optimum weights for each class in the training process. Based on the TPCNN network, this method automatically learns region features to distinguish the difference between normal faces and a face with facial nerve palsy. This shows an improvement in robustness and consistency as a result of the classification of facial nerve functions.

We compare our algorithm with other deep networks in terms of accuracy and evaluation of the F1 score. The other methods could not achieve good performances in evaluating facial nerve function. This may happen because the design of their architecture is just not suitable for data augmentation, where our TPCNN method considers the features of global and local facial movements and could therefore improve accuracy, precision, recall, and F1 score over other methods.

Future Work

The current study has some limitations. Most importantly, the lack of various facial expressions and the varying distributions of the patients' facial sequence lengths with different HB grades have limited the optimization of the deep network and the learning of the dynamic features of muscle movement.

Our results are promising, and there are many areas in which this research can be pursued. First, if we have enough data, then there is the possibility to analyze the effect of different sequence lengths on the model's accuracy. Furthermore, because our method uses a supervised deep network and is dependent on labels prepared by the subjective opinion of the clinician, it will be worth testing the unsupervised method where the data samples will be processed independently of the clinical procedure. Finally, the effects of each data augmentation strategy on the accuracy of the model can be considered as future research problems. Because time-series data augmentation is not used as much as image data augmentation, there is good potential for time-series data augmentation to grow, especially in the domain of facial nerve function. There are other advanced data augmentation methods (e.g., metalearning, filters, and style transfer) that have been applied in the image domain but are still not used by time series. Therefore, there is a good opportunity for new research to work on the augmentation of time-series data.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/app12125902/s1, Table S1. F1-Score of TPCNN. Table S2. F1-Score of CNN-LSTM. Table S3. F1-Score of ResNet. Table S4. F1-Score of LSTM. Table S5. F1-Score of FCN.

Author Contributions: All authors contribute equally, including manuscript preparation, namely: M.S., neural network; J.K., J.M. and K.Š., medical background and data acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: The work of J.K., K.Š. and J.M. was funded by the Ministry of Education, Youth and Sports by grant 'Development 406 of Advanced Computational Algorithms for evaluating post-surgery rehabilitation' number LTAIN19007. This support is gratefully acknowledged.

Institutional Review Board Statement: This study was conducted according to the guidelines of the Declaration of Helsinki, and it was approved by the Institutional Ethics Committee of Charles University Prague, University Hospital Kralovske Vinohrady, EK-VP/431012020, approved 22 June 2020.

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

Data Availability Statement: The data presented in this study are available on request from the corresponding author. The data are not publicly available due to ethical restrictions.

Acknowledgments: We thank the Ministry of Education, Youth and Sports; Faculty Hospital Královské Vinohrady; University of Chemistry and Technology Prague.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

Adam	Adaptive Moment Estimation
CNN	Neural Network with Convolutional layers
FC	Fully connected
FCN	Fully Convolutional Network
FLOPs	Floating-point operations
HB	House–Brackmann facial nerve grading system
LSTM	Long short-term memory
MACs	Multiply-accumulate operations
ReLU	Rectified Linear Unit
ResNet	Residual Network
SGD	Stochastic gradient descent
TPCNN	Triple-path convolutional neural network

References

1. Nussbaum, R.; Kelly, C.; Quinby, E.; Mac, A.; Parmanto, B.; Dicianno, B.E. Systematic Review of Mobile Health Applications in Rehabilitation. *Arch. Phys. Med. Rehabil.* **2019**, *100*, 115–127. [CrossRef]
2. Geman, O.; Postolache, O.; Chiuchisan, I. *Mathematical Models Used in Intelligent Assistive Technologies: Response Surface Methodology in Software Tools Optimization for Medical Rehabilitation*; *Recent Advances in Intelligent Assistive Technologies: Paradigms and Applications*; Intelligent Systems Reference Library; Springer: Cham, Switzerland, 2020; Volume 170; pp. 83–110.
3. Chen, H.; Gao, J.; Zhao, D.; Song, H.; Su, Q. Review of the Research Progress in Deep Learning and Biomedical Image Analysis Till 2020. *J. Image Graph.* **2021**, *26*, 101874.
4. Ameh Joseph, A.; Abdullahi, M.; Junaidu, S.B.; Hassan Ibrahim, H.; Chiroma, H. Improved Multi-classification of Breast Cancer Histopathological Images Using Handcrafted Features and Deep Neural Network (Dense Layer). *Intell. Syst. Appl.* **2022**, *14*, 200066. [CrossRef]
5. Hirra, I.; Ahmad, M.; Hussain, A.; Ashraf, M.U.; Saeed, I.A.; Qadri, S.F.; Alghamdi, A.M.; Alfakeeh, A.S. Breast Cancer Classification from Histopathological Images Using Patch-Based Deep Learning Modeling. *IEEE Access* **2021**, *9*, 24273–24287. [CrossRef]
6. Maiello, L.; Ball, L.; Micali, M.; Iannuzzi, F.; Scherf, N.; Hoffmann, R..; Gama de Abreu, M.; Pelosi, P.; Huhle, R. Automatic Lung Segmentation and Quantification of Aeration in Computed Tomography of the Chest Using 3D Transfer Learning. *Front. Physiol.* **2022**, *12*, 2508. [CrossRef]
7. Anaya-Isaza, A.; Mera-Jimenez, L. Data Augmentation and Transfer Learning for Brain Tumor Detection in Magnetic Resonance Imaging. *IEEE Access* **2022**, *10*, 23217–23233. [CrossRef]
8. Rosahl, S.; Bohr, C.; Lell, M.; Hamm, K.; Iro, H. Diagnostics and Therapy of Vestibular Schwannomas: An Interdisciplinary Challenge. *GMS Curr. Top. Otorhinolaryngol. Head Neck Surg.* **2017**, *16*, Doc03. doi: [CrossRef]
9. Wachtman, G.S.; Cohn, J.F.; VanSwearingen, J.M.; Manders, E.K. Automated Tracking of Facial Features in Patients with Facial Neuromuscular Dysfunction. *Plast. Reconstr. Surg.* **2001**, *107*, 1124–1133. [CrossRef]
10. Ngo, T.H.; Seo, M.; Chen, Y.W.; Matsushiro, N. Quantitative Assessment of Facial Paralysis Using Local Binary Patterns and Gabor Filters. In Proceedings of the Fifth Symposium on Information and Communication Technology, Hanoi, Vietnam, 4–5 December 2014; pp. 155–161.
11. Hontanilla, B.; Aubá, C. Automatic Three-dimensional Quantitative Analysis for Evaluation of Facial Movement. *J. Plast. Reconstr. Aesthetic Surg.* **2008**, *61*, 18–30. [CrossRef]
12. Liu, X.; Dong, S.; An, M.; Bai, L.; Luan, J. Quantitative Assessment of Facial Paralysis Using Infrared Thermal Imaging. In Proceedings of the 2015 8th International Conference on Biomedical Engineering and Informatics (BMEI), Shenyang, China, 14–16 October 2015; pp. 106–110.
13. Ben, X.; Zhang, P.; Yan, R.; Yang, M.; Ge, G. Gait Recognition and Micro-expression Recognition Based on Maximum Margin Projection with Tensor Representation. *Neural Comput. Appl.* **2016**, *27*, 2629–2646. [CrossRef]
14. Yoshihara, H.; Seo, M.; Ngo, T.H.; Matsushiro, N.; Chen, Y.W. Automatic Feature Point Detection Using Deep Convolutional Networks for Quantitative Evaluation of Facial Paralysis. In Proceedings of the 2016 9th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Datong, China, 15–17 October 2016; pp. 811–814.
15. 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.
16. Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going Deeper with Convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9.

17. Kohout, J.; Verešpejová, L.; Kříž, P.; Červená, L.; Štícha, K.; Crha, J.; Trnková, K.; Chovanec, M.; Mareš, J. Advanced Statistical Analysis of 3D Kinect Data: Mimetic Muscle Rehabilitation Following Head and Neck Surgeries Causing Facial Paresis. *Sensors* **2021**, *21*, 103. [CrossRef] [PubMed]
18. Červená, L.; Kříž, P.; Kohout, J.; Vejvar, M.; Verešpejová, L.; Štícha, K.; Crha, J.; Trnková, K.; Chovanec, M.; Mareš, J. Advanced Statistical Analysis of 3D Kinect Data: A Comparison of the Classification Methods. *Appl. Sci.* **2021**, *11*, 4572. [CrossRef]
19. Liu, B.; Zhang, Z.; Cui, R. Efficient Time Series Augmentation Methods. In Proceedings of the 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Chengdu, China, 17–19 October 2020; pp. 1004–1009.
20. Zemanová, M.; Janda, V.; Ondráčková, Z. *Rehabilitace po obrně lícního nervu*; Technical Report; Státní zdravotní ústav: Praha, Czech Republic, 1998.
21. House, W. Facial Nerve Grading System. *Otolaryngol. Head Neck Surg.* **1985**, *93*, 184–193. [CrossRef] [PubMed]
22. Han, J.; Shao, L.; Xu, D.; Shotton, J. Enhanced Computer Vision With Microsoft Kinect Sensor: A Review. *IEEE Trans. Cybern.* **2013**, *43*, 1318–1334. doi: [CrossRef]
23. Dwarampudi, M.; Reddy, N. Effects of Padding on LSTMs and CNNs. *arXiv* **2019**, arXiv:1903.07288.
24. Shorten, C.; Khoshgoftaar, T.M. A Survey on Image Data Augmentation for Deep Learning. *J. Big Data* **2019**, *6*, 60. [CrossRef]
25. Hasibi, R.; Shokri, M.; Dehghan, M. Augmentation Scheme for Dealing with Imbalanced Network Traffic Classification Using Deep Learning. *arXiv* **2019**, arXiv:1901.00204.
26. Olson, M.; Wyner, A.; Berk, R. Modern Neural Networks Generalize on Small Data Sets. In Proceedings of the Advances in Neural Information Processing Systems 31, Montreal, QC, Canada, 2–8 December 2018.
27. Fields, T.; Hsieh, G.; Chenou, J. Mitigating Drift in Time Series Data with Noise Augmentation. In Proceedings of the 2019 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, 5–7 December 2019; pp. 227–230.
28. Le Guennec, A.; Malinowski, S.; Tavenard, R. Data Augmentation for Time Series Classification Using Convolutional Neural Networks. In Proceedings of the ECML/PKDD Workshop on Advanced Analytics and Learning on Temporal Data, Riva Del Garda, Italy, 19–23 September 2016.
29. Um, T.T.; Pfister, F.M.; Pichler, D.; Endo, S.; Lang, M.; Hirche, S.; Fietzek, U.; Kulić, D. Data Augmentation of Wearable Sensor Data for Parkinson's Disease Monitoring Using Convolutional Neural Networks. In Proceedings of the 19th ACM International Conference on Multimodal Interaction, Glasgow, UK, 11–13 December 2017; pp. 216–220.
30. Iwana, B.K.; Uchida, S. An Empirical Survey of Data Augmentation for Time Series Classification with Neural Networks. *PLoS ONE* **2021**, *16*, e0254841.
31. Lou, J.; Cai, X.; Wang, Y.; Yu, H.; Canavan, S. Multi-subspace Supervised Descent Method for Robust Face Alignment. *Multimed. Tools Appl.* **2019**, *78*, 35455–35469. [CrossRef]
32. Samsudin, W.S.W.; Samad, R.; Ahmad, M.Z.; Sundaraj, K. Forehead Lesion Score for Facial Nerve Paralysis Evaluation. In Proceedings of the 2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS), Selangor, Malaysia, 29 June 2019; pp. 102–107.
33. Lafer, M.P.; O, T.M. Management of Long-standing Flaccid Facial Palsy: Static Approaches to the Brow, Midface, and Lower Lip. *Otolaryngol. Clin. N. Am.* **2018**, *51*, 1141–1150. [CrossRef]
34. 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]
35. 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*, 30. [CrossRef] [PubMed]
36. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.c. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In Proceedings of the Advances in Neural Information Pprocessing Systems 28, Montreal, QC, Canada, 8–9 December 2015.
37. Wang, Z.; Yan, W.; Oates, T. Time Series Classification from Scratch with Deep Neural Networks: A Strong Baseline. In Proceedings of the 2017 International Joint Conference on Neural Networks (IJCNN), Anchorage, AK, USA, 14–19 May 2017; pp. 1578–1585.