Pattern Recognition 144 (2023) 109866

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Deep learning-driven diagnosis: A multi-task approach for segmenting stroke and Bell's palsy

Sabina Umirzakova [a], Shabir Ahmad [a], Sevara Mardieva [a], Shakhnoza Muksimova [a], Taeg Keun Whangbo [b,*]

[a] Department of IT Convergence Engineering, Gachon University, Seongnam, South Korea
[b] Department of Computer Science, Gachon University, Seongnam, South Korea

ARTICLE INFO

Keywords:
Segmentation
Face parsing
Early stroke detection
Bell's palsy detection

ABSTRACT

Strong efforts have been undertaken to enhance the diagnosis and identification of diseases that cause facial paralysis, such as Bell's palsy and stroke, because of their detrimental social effects. Stroke is one of the most serious and potentially fatal conditions among the major cardiovascular disorders. We are introducing a deep-learning-based method for early diagnosis of facial paralysis diseases such as stroke and Bell's palsy. Recognizing the costs associated with traditional diagnostic techniques like magnetic resonance tomography (MRI) and computed tomography (CT) scan images, our model employs a multi-task network, integrating face parsing, facial asymmetry parsing, and category enhancement. Spatial inconsistencies are addressed via a depth-map estimation module that leverages an instance-specific kernel approach. To clarify the boundaries of facial components, we use category edge detection with a foreground attention module, generating generic geometric structures and detailed semantic cues. Our model is trained on two datasets, comprising individuals with regular smiles and those with one-sided facial weakness. This cost-effective, easily accessible solution can streamline the diagnostic process, minimizing data gaps, and reducing needless rescreening and intervention costs.

1. Introduction

Facial paralysis, characterized by a loss of voluntary muscle movement in an individual's face, often manifests in an asymmetrical form, affecting one side of the face. This facial asymmetry is a common symptom in various conditions, most notably strokes and Bell's palsy. As a medical emergency that requires immediate attention, a stroke is one of the leading causes of facial paralysis. A stroke occurs when the blood supply to a part of the brain is disrupted, causing brain cells to die from a lack of oxygen and nutrients. The result of this event can lead to various neurological impairments, including facial paralysis, as the nerves controlling facial muscles get damaged. According to the World Stroke Organization ([1], more than 12.2 million cases of new strokes are reported each year. One stroke occurs in the lifespan of one in four people over the age of 25 worldwide. Over 16% of all strokes happen to people between the ages of 15 and 49 each year, over 62% happen to people under the age of 70, 47% happen to males, and 53% happen to women. Approximately 6% of all stroke deaths occur in people aged 15 to 49; 34% of all stroke deaths occur in those under the age of 70, and men account for 51% of all stroke deaths compared to 49% for women. Each year, there are six and a half million stroke deaths worldwide. One factor contributing to the exceptional increase in stroke-related deaths is the lack of awareness of stroke symptoms [2]. Stroke is a condition that is frequently ignored. Because many patients are unaware of its warning signs and symptoms, the disease is not treated as quickly as it should be [3]. Even when employing established prehospital stroke detection screening measures, such as the Cincinnati prehospital stroke scale (CPSS), emergency medical service professionals may fail to diagnose stroke in more than half of the cases [4]. The CPSS and different screening instruments depend strongly on the provider's background and training to correctly identify neurological issues, which presents the greatest challenge to non-neurologists. Facial weakness showed the second-worst agreement (0.23 of all scale components in a survey of over 11 stroke professionals using the National Institute of Health stroke scale) [5]. The wide range of CPSS sensitivity and specificity observed in the prehospital setting is partly due to weak inter-operator variability. Consequently, more stroke cases are either misdiagnosed or improperly triaged, delaying or eliminating the possibility of prompt acute stroke

* Corresponding author.
E-mail address: tkwhangbo@gachon.ac.kr (T.K. Whangbo).

https://doi.org/10.1016/j.patcog.2023.109866
Received 6 March 2023; Received in revised form 5 July 2023; Accepted 31 July 2023
Available online 1 August 2023

treatment with thrombolysis. Many researchers have created diagnostics systems based on deep neural networks owing to the promising success of deep learning in computer vision tasks, including segmentation [6], detection, and classification [7]. Recent advances in the field of pattern recognition, particularly deep-learning algorithms, have enabled the distinction between nuanced facial expressions, such as planned and unplanned grins, by examining different patterns of facial muscle activation. Such algorithms are even capable of detecting minor asymmetrical facial expressions that may suggest a negative emotional valence. Facial weakness, one of the most prevalent stroke symptoms, has been significantly explored in recent studies on stroke detection techniques. Acute stroke-related facial paralysis often manifests as a drooping effect on one side of the face, leading to an uneven or lopsided smile and asymmetrical face [3]. The ultimate objective of this study is to develop a system capable of automatically segregating abnormal facial weakness from a patient's input image, without the need for specialized equipment. Timely and automated recognition of facial weakness serves as a critical initial step in stroke segmentation, enhancing the potential for early intervention and treatment [8]. By incorporating recent advancements in pattern recognition and deep-learning algorithms [41], this study aims to contribute significantly to the current body of knowledge on automated stroke detection.

Acute stroke therapy can be administered earlier and more frequently as a result of improved prehospital stroke identification, which can reduce disability in thousands of stroke patients worldwide. By creating this Artificial Intelligence (AI) model, we aim to enhance the prehospital detection of stroke, allowing for earlier and more effective intervention. This is particularly crucial given that swift treatment following a stroke can significantly reduce the risk of severe disability and improve patient outcomes. Thus, our model has the potential to contribute significantly to stroke treatment and patient care. Identifying stroke symptoms can be challenging, even for trained healthcare professionals. The proposed AI model could reduce this burden by automating the detection process, leaving less room for human error and ensuring patients receive timely care.

The key contributions of this study can be summarized as follows:

- We propose a light deep-learning-based model that can analyze a patient's facial asymmetry and parse six categories, namely, skin, eyebrow, eye, mouth, and stroke.
- The pilot study proposes the segmentation of facial weakness diseases, such as stroke and Bell's palsy, based on geometric structure dependencies among the facial categories.
- Instead of focusing on the entire feature map, we propose a new foreground attention block that extracts localized features by restricting cross attention within the foreground region of the anticipated mask for each query.
- To obtain more accurate segmentation results, we have added a depth-map estimator that integrates the depth-estimation pipelines and estimates the depth for each category using the same instance-specific kernel approach.

Instead of using a general deep learning model, we have developed a novel, lightweight deep learning model specifically designed for analyzing facial asymmetry. This model parses six categories related to facial structures, ensuring that it's more tailored and effective for this specific task. This study introduces a pioneering approach to segment facial weakness diseases like stroke and Bell's palsy. By utilizing geometric structure dependencies among the facial categories, it adds an extra layer of specificity to the process, which previous studies have not explored. We introduce a new method for extracting localized features. The proposed foreground attention block, rather than focusing on the entire feature map, restricts cross attention within the foreground region of the anticipated mask for each query. This technique brings a new way of focusing on the essential details within the foreground and ignoring the irrelevant background noise. Another major contribution is the inclusion of a depth-map estimator within the model, which estimates depth for each category using the same instance-specific kernel approach. This component allows the model to perceive the depth dimension, enhancing its understanding of the facial structure and hence the accuracy of the segmentation results. The proposed model can analyze patient images without the use of specialized tools, making it potentially more accessible to people worldwide. This could lead to more widespread and democratized stroke detection, particularly beneficial in under-resourced settings. Through this research, we hope to bridge the gap between the current state of stroke detection and the ideal state, in which stroke detection is quick, accurate, and accessible for all potential patients. This could revolutionize the field of stroke detection and contribute greatly to global health outcome.

2. Related work

Facial weakness is one of the main elements of the most commonly used prehospital screening tools. Facial weakness is one of the detected general neurological signs for a variety of neurological diseases, such as Bell's palsy and stroke. The most obvious symptom of stroke and Bell's palsy is facial paralysis, which manifests as drooping of one side of the face.

Dangerous diseases such as stroke are detected using CT and magnetic resonance tomography, which are not always available everywhere. Therefore, the main goal of our study was to present a more suitable method for detecting urgent diseases based on facial features. In the literature review, we mainly focus on recent research on facial weakness diseases, such as stroke and Bell's palsy, which are detected or analyzed based on facial features, brain signal analysis, and body motion.

2.1. Shape features approach

The most common method of detecting facial weaknesses is to examine facial asymmetry using geometric clues [1]. To perform classification on a single image or series of images, researchers frequently identify facial landmarks and directly evaluate the geometric properties of the face, such as the distances and angles between landmarks. Anuja et al. [9] focused on detecting face asymmetry based on feature coordinate differences between the left and right faces. Gemma et al. [10] attempted to measure facial abnormalities using a facial measure module to quantify the degree of facial asymmetry using facial landmarks, wherein a binary classifier using the multilayer perceptron approach delivered the output label. To create shape-based features for categorizing facial weaknesses, Zhexiao et al. [11] estimated the location and displacement of facial landmarks.

2.2. Machine learning approach

Machine-learning-based computer vision analysis can improve the clinical diagnosis of diseases that may be visually observable. Three machine-learning classifiers were tested and assessed by Chourib et al. [12] using a variety of feature-selection strategies. They identified the most pertinent information and used well-known machine-learning models to predict and categorize individuals with stroke. Instead of using ineffective classifiers, Amira et al. [13] described an ensemble learning strategy for facial weakness classification based on an SVM used as an estimator to improve prediction outcomes. With accuracy and sensitivity comparable to those of qualified paramedics, Chad et al. [14] demonstrate how a machine-learning algorithm utilizing computer vision analysis can identify facial weaknesses from videos. To predict strokes, Soumyabrata et al. [15] presented a thorough analysis of patient characteristics using electronic medical data. They analyzed multiple characteristics, performed feature correlation analyses, and used principal component analysis for step-wise analysis and selection of the best collection of features. To categorize individual stroke risk levels,

Xuemeng et al. [16] created nine models. The models created for the proposed method were designed to enhance the current screening procedure to minimize the influence of unmeasured values, increase the effectiveness of interventions for those at a high risk of stroke, and lower the cost of treatment.

2.3. Deep learning approach

Deep-learning-based techniques can learn and automatically extract traits unique to people with palsy. Muhammad et al. [17] used a deep-learning method to automatically extract traits unique to people with palsy from asymmetric facial photographs, which were used to categorize five asymmetric face grades. To automatically evaluate the severity of facial paralysis, Pengfei et al. [18] presented a network model using a combination of a dual-path long short-term memory (LSTM) and deep differentiated network. Syed et al. [19] used three deep-learning models to focus on the early detection of facial paralysis, which is one of the most common stroke symptoms. To fine-tune a modified VGGNet for stress inference and emotion identification, Cuiting et al. [20] collected a dataset of the facial and emotional expressions of patients with facial asymmetry. Using a bi-directional LSTM network, Yan et al. [21] introduced a framework for facial weakness identification from a video that simulated the temporal dynamics of both the shapes and appearance-based aspects of each target frame. Electroencephalogram signals are used by Mandeep et al. [22] to provide a rapid strategy based on a noninvasive method that uses raw data to train the framework and then predict if the person can have a stroke soon or if the signals are normal. Eui-Sun et al. [23] created an early-stage stroke analysis software based on artificial intelligence that can detect the early stage of stroke by analyzing facial muscle anomalies. Using a video of the neurological examinations, Taeho et al. [24] proposed a deep-learning-based screening technique for patients with stroke. Experiments were conducted using the ResNet-18 architecture with an overlapping N-divided recurrence plot pretrained with protocol testing, as well as three additional methods.

3. Proposed method

We investigated our network by combining the convolution module and enhanced stroke category (ESC) blocks, as shown in Fig. 1. We built our network as follows: (1) Without losing any spatial information from the feature maps, we added an ESC block after each convolution block. (2) An atrous-spatial pyramid pooling module is added to the final block. (3) We added a depth-map block that enhanced the segmentation of each category. (4) Finally, to focus on a specific category we added a foreground attention module.

3.1. Enhance stroke category

To reduce the impact of background variations and increase the geometric structure dependencies of the facial categories, an ESC block is added Fig. 2, which increases the global relation for locating the stroke category in the dynamically fused feature and is responsible for separating contextual relationships from features. Our ESC block consists of a 1×1 convolution layer followed by batch normalization as the activation function applied to the ReLU, followed by the implementation of adaptive average pooling with a size of 1×1, three fully connected layers, and bilinear upsampling. The ESC block distributes more

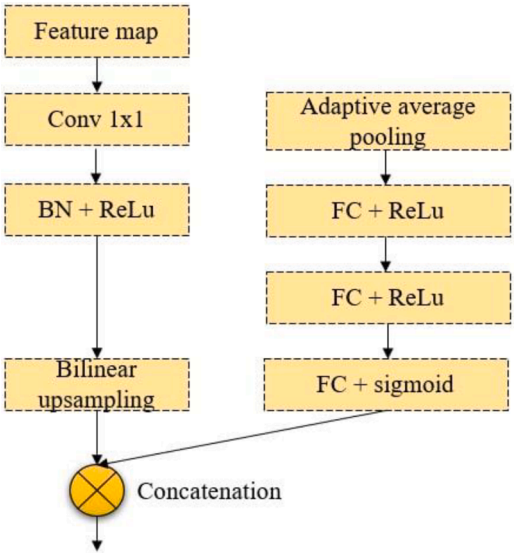

Fig. 2. Representation of the ESC block, which consists of a convolution layer (1×1), batch normalization (activation function ReLu), bilinear upsampling, adaptive average pooling, and three fully connected layers (two ReLu activation functions and one sigmoid).

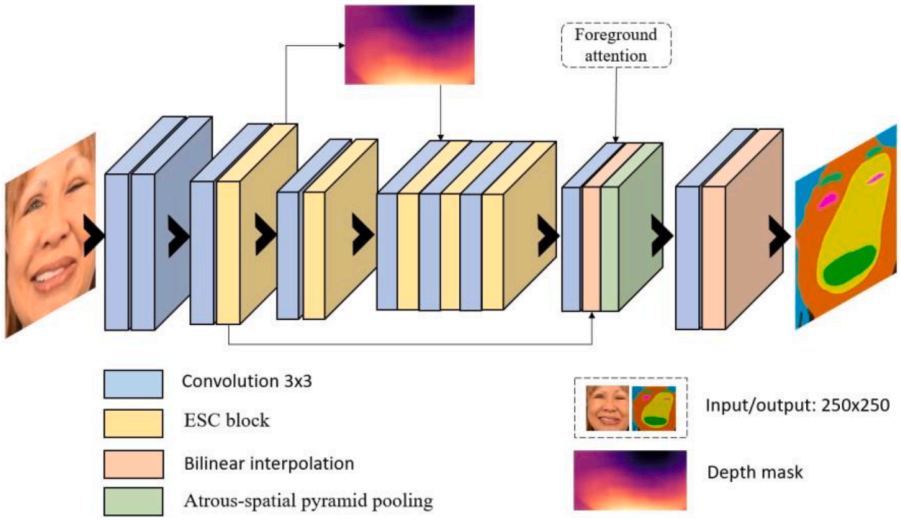

Fig. 1. Proposed architecture. The architecture uses a unique foreground attention block that emphasizes localized feature extraction, limiting cross attention to specific foreground regions, making it more efficient and effective at parsing relevant data. Furthermore, it incorporates a depth-map estimator that augments the depth-estimation pipelines, providing depth measurements for each category, and leading to more accurate segmentation results.

embedded high-context evidence throughout the wide neighborhoods of every feature map. Owing to the aggregated changes in adaptive average pooling, the ESC block is computationally more efficient and aids the network in learning higher-level contextual characteristics. Three fully connected layers were used in the excitation procedure of the self-gating system to extract geometric dependencies from the feature map. The role of the ESC block in our proposed network is to pass rich contextual information on the stroke category region for construction-dependent decision-making.

3.2. Depth of mask

Using the instance-specific kernel technique, which integrates pipelines for depth estimation and category segmentation, we predicted the depth for each instance. To produce the final complete depth map, as shown in Fig. 1, we first ran the depth kernels on the depth embedding to produce instance depth maps, which were then combined according to the findings of category segmentation.

Convolution and sigmoid activation were used to create the adjusted instance depth map M^*, which was then unadjusted to create depth map M.

$$M^* = Sig(D^m * L^m), \tag{1}$$

$$(M|M^*, m^r, m^s) = m_{max}*(m^r * M^* + m^s), \tag{2}$$

$$(M|M^*, m^r, m^s) = m_{max}*[(M^* - 0.5) * m^r + m^s] \tag{3}$$

Here, $L^m$is related to the depth kernels and depth embedding, as it's used in the calculation of the adjusted instance depth map M*, representing a set of learned weights or parameters associated with each instance m in the depth kernel, D^m. These are used in scaling the depth kernels for each instance in the scene, allowing the model to adjust its depth estimates for different instances. Specific depth kernels are defined as $D^m \in R^{L*s_1^m}$ and a common depth embedding $S^m \in R^{s_2^m * \frac{H}{2} * \frac{W}{2}}$ regulates the depth scale, where $m_{min} = 0$ and $m_{max} = 80$. Initially provided, $S^m$is described as a common depth embedding. In general, depth embeddings are representations of depth information that are learned by a machine learning model. The case that S^m is used in generating or adjusting D^m, the specific depth kernels, and in the procedures of the pipeline that aren't covered by these equations.

We predicted an adjusted depth map M' to simplify depth-estimation learning. It is created by combining two projected instance-degree depth parameters, depth range $m^r \in R^L$ and depth shift $m^s \in R^L$, with the instance depth map D. By simply setting $s_1^m = s_2^m + 2$, they may be calculated from high-level features in tandem with depth kernels, and thus, the depth bias and parameter of each instance are characterized. The normalized depth map M^* can be learned more quickly because it encodes the relative depth weights within each pattern.

We combined all category depth maps into an overall picture depth map using non-overlapping segmentation masks M. At the instance boundaries, this produces accurate depth values.

3.3. Foreground attention

The foreground attention section is an essential network component, which extracts localized features by restricting cross-attention within the foreground part of the expected mask for each category instead of attending to the entire feature map. In the foreground attention section, we applied optimization enhancements that improve the model efficiency without adding extra computing. We present foreground attention, a version of cross-attention that presents only the foreground area of the expected mask for each category. Contextual features are essential for image segmentation. However, according to recent research [25], the global context of the cross-attention layer is the cause of the slow convergence of models because cross-attention requires proper training epochs to learn to focus on specific object regions. We propose that local features are sufficient to renew request features and that self-attention can be used to obtain contextual information.

The proposed foreground attention module improves the attention matrix:

$$I_c = soft(F_{-1} + R_cK_c)*T_c + I_{-1}, \tag{4}$$

where c is the category index, $I_c \in R^{W*H}$ refers to W- and H-dimensional request features at the c^{th} category, and $R_c = f_R * I_{-c} \in R^{W*H}$. K_c, $T_c \in R^{W*H*N}$ are the category features. K_c is a transformation matrix to R_c, which changes the representation and highlights certain features of R_c.

Furthermore, the foreground attention F_{-c} at the x, y feature position is described as:

$$F_{-c}(x,y) = \begin{cases} 0 \ if \ F_{-c}(x,y) = 1 \\ -\infty \ else \end{cases}, \tag{5}$$

where $F_{-c} \in \{0,1\}^{L*H_c*W_c}$ is the binarized output with a threshold at 0.5 of the rescaled mask prediction of the preceding layer.

3.4. Loss function

In contrast to common semantic segmentation, facial segmentation features consist of small components. To maintain the pattern of tiny components and the geometrical dependencies of categories, we employ the hypergraph-induced semantic tuplet (HIST) loss [26]. The HIST loss consists of the summation of the $\mathscr{L}_{CE}$hypergraph node classification loss and $\mathscr{L}_D$ allocation loss:

$$\mathscr{L}_{hist} = \lambda_s * \mathscr{L}_{CE} + \mathscr{L}_D \tag{6}$$

where the scaling factor $\lambda s > 0$ equalizes the two loss values.

In addition to the above category classification optimization, we present a boundary-enhanced semantic loss that increases the segmentation loss of the boundary pixels according to the depth mask and foreground attention module boundary maps. Because depth maps are associated with segmentation maps, it is valuable to input two types of border cues into the segmentation module to refine the parsing accuracy of components with clear borders.

$$\mathscr{L}_{besl} = \frac{1}{NP} \sum_{i=1}^{N} \sum_{j=1}^{P} \frac{1}{p_{ij}} * w_{ij} \tag{7}$$

Here, N represents the total amount of samples in a batch, p_{ij} is the number of contour pixels of a defined category in label map $c_{ij} \in R^{H*W}$, w_{ij} represents the category-specific weighting to emphasize a particular category and is the cross-entropy between the expected segmenting output and ground truth. The total loss in our model is summarized as follows:

$$L_{loss} = \mathscr{L}_D + \lambda_s * \mathscr{L}_{CE} * \mathscr{L}_{besl} \tag{8}$$

4. Experimental results

Our solution specifically solves the major issues with stroke segmentation based on facial features in the following essential aspects: the correspondence between various facial components, which is critical in asymmetric facial analysis. However, current techniques such as face- and region-based segmentation lose sight of the links between various components. While exploring these correlations can undoubtedly improve the segmentation speed, face-based segmentation may overlook scale mismatches in various facial components, and region-based segmentation does not take advantage of between-region relationships. In addition, the structural variance of a face is substantially higher than that of an asymmetrical face, resulting in unreliable structural relations for efficient learning. The ESC is one of the best methods for modeling

such correlations for cases with asymmetric faces to address the aforementioned issue. To create and justify the component-wise relationship, we developed an enhancement block for asymmetric face representation. As borders are frequently confusing in real-world situations, segmenting pixels at the borders between components presents another issue. The border pixels cover most face photos more than they do for other segmentation tasks, making this problem more serious for face segmentation. Therefore, an efficient technique for improving the efficiency of face segmentation is to increase the segmentation accuracy of the border pixels. In this study, we accomplished precise border-accurate segmentation from two perspectives: (1) we added border attention to the depth-mask module, assigning border pixels more weight during feature aggregation; and (2) we designed a foreground attention module to reinforce each category of border pixels, which extracts localized features of the predicted segmenting map.

We conducted extensive tests on parsing face components using a publicly available face dataset for Bell's palsy [27] and a private Gil Hospital stroke dataset to assess the performance of the proposed method.

4.1. Datasets

In this study, we segmented the overall six categories as: skin, hair, eyebrow, eye, mouth, and stroke. However, there is no open-access dataset for stroke patients with category labels. We included the Facial Paralysis Possibility (FPP) category by obtaining asymmetric facial parts from the mouth and eyes, which were placed on the dropping side of the face.

In this study, two datasets were employed: (1) a public dataset for Bell's palsy patients [27] and (2) a private Gil Hospital dataset for stroke patients and healthy individuals. Because the dataset was not labeled, our research team labeled the dataset for several months according to expert guidelines.

4.1.1. YouTube-Facial-Palsy database

This is the first publicly accessible database established for the visual inspection of facial palsy symptoms. It includes 32 YouTube recordings of 22 patients with facial paralysis that are converted into images. When the deformation intensity was deemed sufficiently strong by experts, the local palsy zones were manually tagged [27].

For the training process, we applied 15 patient images of a total of 18,000 images. For the testing process, we used seven patient images of a total of 840 images.

4.1.2. Gil medical center database

Data on stroke patients were provided by Gil Hospital. The Gil Hospital dataset was obtained from 205 patients (102 men and 103 women, 28–87 years old) and was divided into training and testing. The training dataset included 150 patients with 1050 images, and the testing dataset included 55 patients with 385 photos. The inclusion criteria were as follows: patients aged 25–90 years; stroke confirmed based on CT or magnetic resonance tools; patients who could think reasonably 2 weeks after a stroke, and; informed consent signed by the patients. Before the trial, each volunteer signed an informed consent form.

4.2. Metrics

The segmentation results were evaluated using three standard metrics: F1 score, intersection over union (IoU), and pixel accuracy.

F1 score is calculated as:

$$F_1 = 2^* \frac{precision^* recall}{precision + recall} \tag{9}$$

IoU is calculated as:

$$IoU = \frac{1}{k+1} \sum_{i=0}^{k} \frac{p_{ii}}{\sum_{j=0}^{k} p_{ij} + \sum_{j=0}^{k} p_{ji} - p_{ii}}, \tag{10}$$

where IoU is determined using the average number of categories.

Pixel accuracy (PA) is calculated as:

$$PA = \frac{\sum_{i=0}^{k} p_{ii}}{\sum_{i=0}^{k} \sum_{j=0}^{k} p_{ij}} \tag{11}$$

4.3. Evaluation details

Workstations with specialized configuration capabilities are utilized to train the models (16 GB RAM, GPU—NVIDIA GeForce GTX 940, and Microsoft Windows 10 with a 64-bit platform). The proposed network was implemented in the PyTorch framework and trained with 300 epochs and a batch size of 32 by applying a stochastic gradient descent with a momentum of 0.9. During the training and testing stages, the input image size is consistently 250 × 250 pixels. The training data undergo augmentation processes, which involve random rotation, with the angle chosen anywhere between −10 and 10°, as well as random scaling where the factor varies between 0.75 and 1.25.

4.4. Comparison with state-of-the-art models (Segmentation)

We made use of several state-of-the-art instance segmentation models, including Mask R-CNN [31], DeepLabv3+ [28], and the models proposed by Zheng et al. [29], Hyungjoon et al. [30], Yiming et al. [32], Chang et al. [33], and Yang et al. [34]. Rather than using pre-trained models on widely-used datasets, such as the COCO keypoint, Cityscapes, and CelebAMask-HQ, Lapa datasets, we originally trained these models on the YouTube-Facial-Palsy [27] datasets. This means the models were tuned from the very start to perform optimally for our task and data. We followed standard procedures for training these models, with a specific learning rate, a certain number of epochs, and with selected batch size and other hyperparameters, as per the requirement of our data and task. After the original training, we then fine-tuned these models to further refine their performance. Fine-tuning involved continuing the training process, with a smaller learning rate to prevent significant alteration to the previously learned features. We ensured the original architecture of each model remained unchanged. The aim of fine-tuning was to strike a balance between adaptation to our task and preventing overfitting. The performance of each model was then evaluated based on their ability to correctly identify and segment different features in our images. These results were then compared to that of our proposed model.

The output results in Table 1 show that the proposed method greatly excels in face asymmetry cases when compared to other methods, with a mean F1 score of 96.93%. Specifically, for the YouTube-Facial-Palsy database [27] (Figs. 3 and 4), our model outperforms that proposed by Sabina et al. [6] by 9.43%. We conclude that our results present accurate parsing even over small details such as the categories of hair, eyebrows, and eyes. Our model is stable for strong geometrical changes in cases of facial asymmetry and yields highly accurate results.

According to the results presented in Table 2, and Fig. 4 the proposed approach presents an improvement of 9.69% on the Gil Medical Center database. The findings demonstrate that for the majority of facial components, the proposed model provides the best parsing performance for asymmetrical face instances.

Extensive testing reveal that the additional helper ESC block considerably improves the parsing quality of the network face components. In particular, the presented datasets improved the F1 scores by an average of 9% across all marker classes, including stroke.

According to the results presented in Table 3, the proposed model achieves a mean IoU of 6.28% and 5.91% of mean AP higher than another state - art models trained using the same parameters, where

Table 1
Comparison with other models based on the YouTube-Facial-Palsy database. In the calculation of eyebrow and eye, mean values of the paired category are presented.

Models	Skin	Hair	Eyebrow	Eye	Mouth	FPP	Mean F1 (%)
Mask R-CNN [31]	85.3	88.9	90.5	86.1	92.0	63.20	84.33
DeepLabv3+ [28]	92.3	85.3	96.5	88.6	91.6	62.90	86.2
Zheng et al. [29]	94.8	90.4	78.0	80.0	89.8	52.31	81.38
Hyungjoon et al. [30]	95.5	84.3	81.2	83.1	63.4	51.99	76.58
Yiming et al. [32]	**96.2**	94.9	85.7	88.5	95.0	61.3	86.93
Chang et al. [33]	93.8	93.9	83.2	86.5	63.4	51.6	78.73
Yang et al. [34]	86.9	91.3	84.5	82.0	62.9	52.9	76.75
Sabina et al. [6]	93.90	**95.90**	89.4	89.0	92.90	63.90	87.5
Proposed Method	96.0	94.5	**98.6**	**96.8**	**97.5**	**98.2**	**96.93**

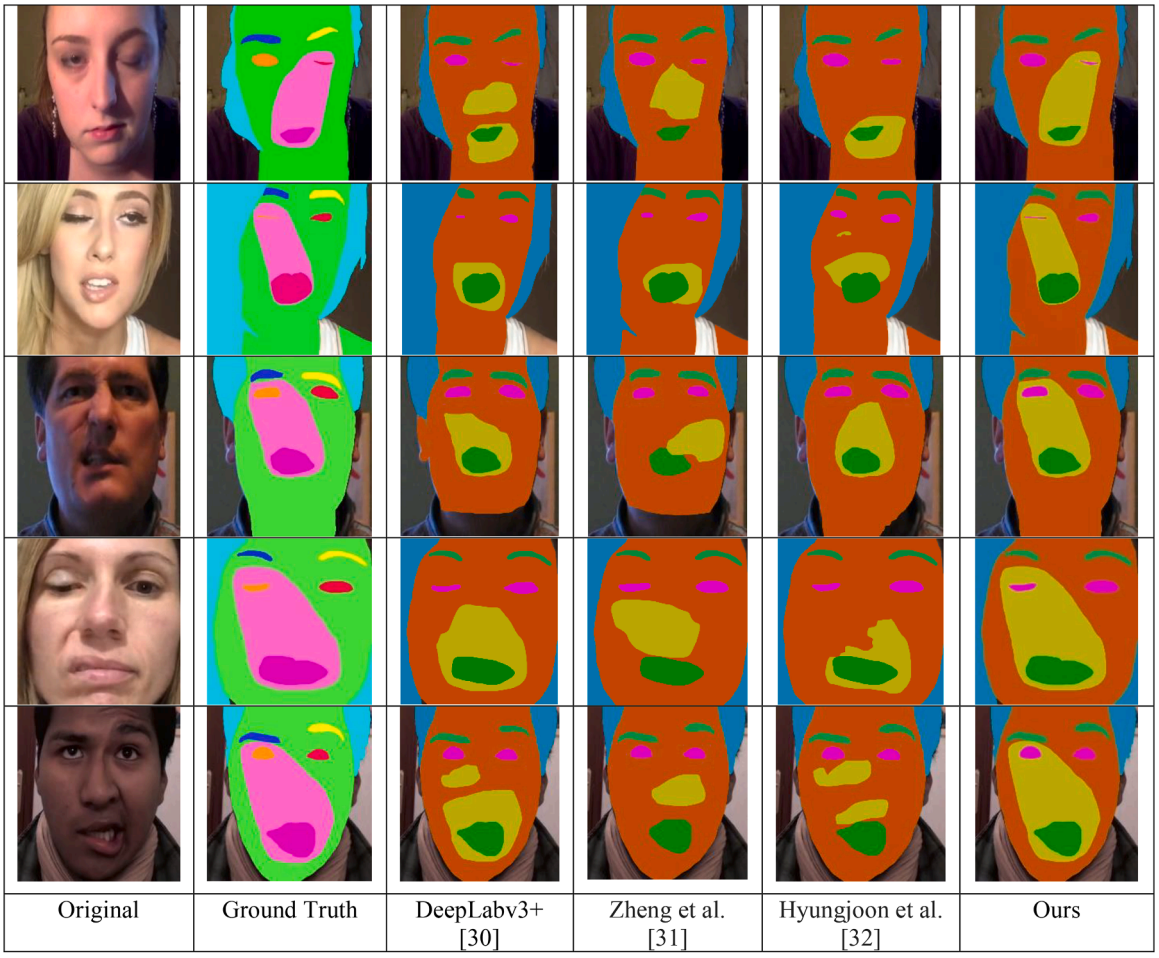

Fig. 3. The proposed model can receive complete facial components with clear boundaries in asymmetric facial cases. Visual samples in each column are created by the corresponding "YouTube-Facial-Palsy database" dataset. The input images, image mask, and results are presented in this figure. For this research, we used open access dataset, of which the creator's got permission from patients.

mean values of the matched categories were computed for the mouth and for stroke.

4.5. Comparison with state-of-the-art models (Facial paralysis models)

In this section, we focused on facial paralysis studies and demonstrated their accuracy using the F1 metric, which is suitable for detection and segmentation models. We compared state-of-the-art models and technologies with the current study by FPP level of face. As a calculation metrics were taken F1 score.

Chaoqun et al. [35] proposed an approach that leverages advanced image analysis algorithms to automatically evaluate the degree of facial paralysis. The system analyzes facial images or videos and extracts key facial landmarks, such as the position of the eyes, mouth, and other facial features. It then quantifies the asymmetry between the paralyzed and unaffected sides of the face, providing an objective measure of facial paralysis severity. The proposed method's effectiveness in accurately assessing more complex cases, such as partial paralysis or asymmetrical muscle involvement, should be evaluated. In this kind of situation presented method would struggle to provide accurate assessments in such scenarios, potentially leading to misinterpretations or inaccurate severity classifications. Xin et al. [36] presented CNN architecture that takes advantage of the hierarchical structure of the face, dividing it into multiple regions based on anatomical landmarks. Each region is processed independently by parallel CNN branches, allowing the network to capture region-specific features. By combining the outputs of these

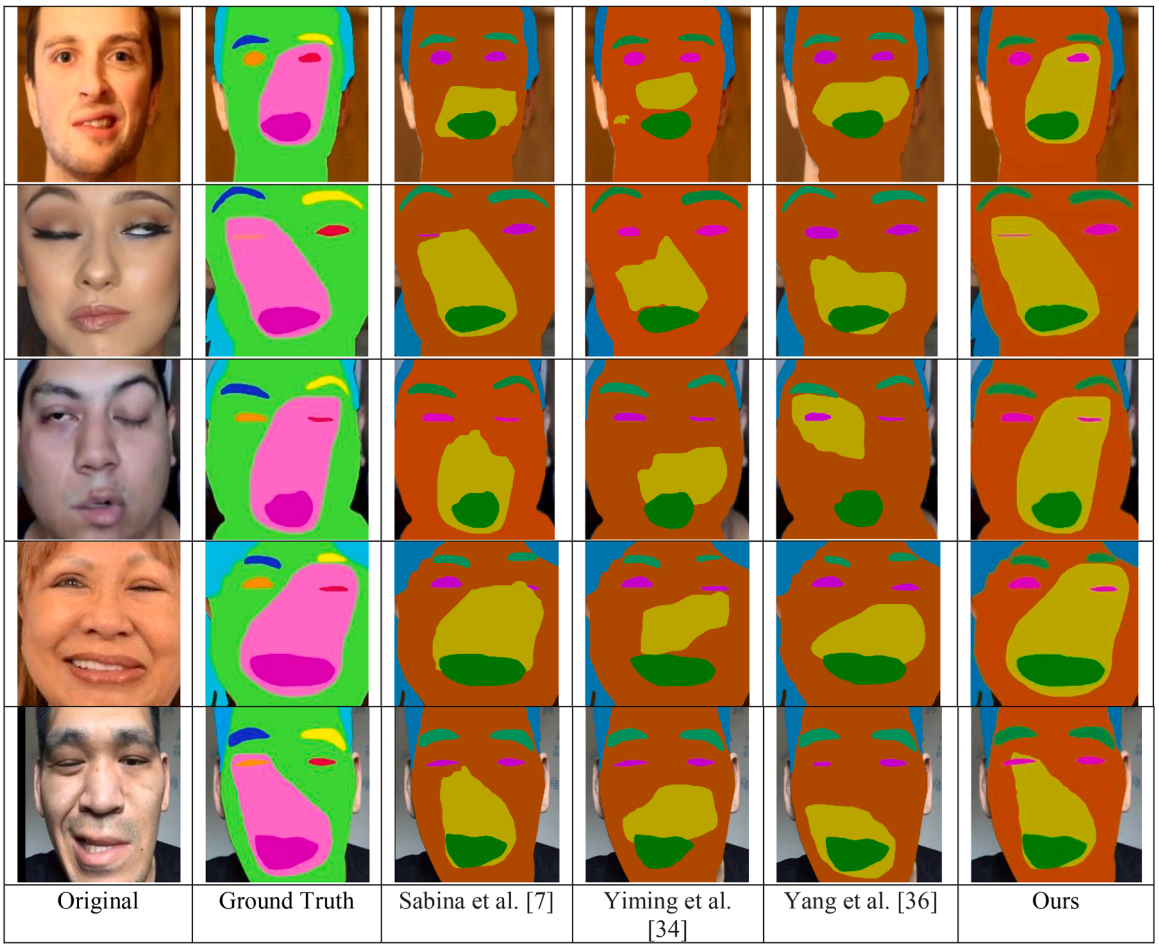

Fig. 4. Our method aims to provide a comprehensive comparison of its results with state-of-the-art models in order to showcase its effectiveness and superiority. In this comparison, we evaluated our method against three prominent models in the field: Sabina et al. [6], Yiming et al. [32] and Yang et al. [34].

Table 2
Comparison with other models based on the Gil Medical Center database.

Models	Skin	Hair	Eyebrow	Eye	Mouth	FPP	Mean F1 (%)
Mask R-CNN [31]	85.1	87.6	91.2	86.6	91.99	62.1	84.09
DeepLabv3+ [28]	90.0	82.9	94.6	90.1	90.8	61.5	84.98
Zheng et al. [29]	93.9	91.2	77.0	81.5	89.6	51.5	80.78
Hyungjoon et al. [30]	94.99	82.8	80.99	84.2	61.8	52.1	76.14
Yiming et al. [32]	**96.1**	95.2	84.3	87.8	94.99	60.1	86.41
Chang et al. [33]	94.1	93.4	82.9	83.99	65.8	56.8	79.49
Yang et al. [34]	86.4	90.1	81.7	82.3	60.7	51.4	75.43
Sabina et al. [6]	92.99	**95.6**	89.4	88.7	92.95	62.99	87.1
Proposed Method	95.99	94.1	**97.6**	**95.8**	**98.7**	**98.6**	**96.79**

Table 3
Comparison of the design of the models.

Models	Attention module based	Boundary enhancement module	Depth-map features	Disease included as category	Mean IoU (%)	Mean PA (%)
Mask R-CNN [31]			✓		83.63	80.64
DeepLabv3+ [28]			✓		88.29	86.92
Zheng et al. [29]	✓	✓			81.43	81.85
Hyungjoon et al. [30]	✓				78.18	75.93
Yiming et al. [32]	✓				83.62	83.37
Chang et al. [33]	✓				79.31	77.91
Yang et al. [34]	✓				76.89	75.63
Sabina et al. [6]		✓			87.52	86.99
Proposed Method	✓	✓	✓	✓	**93.8**	**92.9**

branches, the model generates a comprehensive evaluation of facial nerve paralysis. The effectiveness of the proposed method may be affected by the quality of input images. Factors such as lighting conditions, image resolution, and image artifacts can introduce noise and inconsistencies, potentially degrading the accuracy of facial paralysis evaluation. Preprocessing techniques or additional image quality enhancement methods are required to mitigate this limitation. Pengfei et al. [37] developed a deep learning-based framework that can automatically assess the degree of facial nerve paralysis. The proposed model consists of two parallel LSTM paths: one processes facial landmarks and the other focuses on facial textures. This dual-path architecture allows the model to capture both spatial and temporal information from facial images. The evaluation of facial nerve paralysis often relies on subjective visual assessments by medical professionals. These subjective assessments can introduce variability and may not always be accurate or consistent. Therefore, if the ground truth used for training the model is subjective, it can affect the reliability and validity of the automated evaluation. Ting et al. [38] the proposed method utilizes a deep learning approach to automatically assess the severity of facial paralysis based on facial images. The cascaded encoder network structure consists of multiple layers of encoders that progressively extract features from the input images. This allows for a hierarchical representation of the facial features, capturing both global and local details. Facial paralysis evaluation typically involves considering various factors, such as muscle tone, movement range, and overall facial expression. While the proposed method focuses on capturing facial asymmetry and distortion, the presented method does not fully capture the broader context required for a comprehensive evaluation of facial paralysis. Gee-Sern Jison et al. [39] offers an accurate and efficient deep-learning-based solution for facial palsy detection and diagnosis, utilizing component networks for face detection, landmark detection, and local palsy region detection. The presented method heavily relies on the accurate detection of facial landmarks. If the landmark detection network fails to locate the landmarks correctly, it affects the overall performance of the system. Inaccurate landmark detection can lead to the misidentification of palsy regions. Gemma et al. [40] presented method utilizes key point analysis based on CNN, which involves the identification and tracking of specific facial landmarks, such as the corners of the eyes, mouth, and nose. By comparing the positions and movements of these key points between the paralyzed and non-paralyzed sides of the face, the algorithm can detect and quantify facial asymmetry. In this section, we compared face asymmetry cases with the above methods in Table 4.

The findings presented in Table 4 demonstrate that our suggested approach significantly outperforms others in the area of facial paralysis, achieving an average F1 score of 98.2%. When utilizing the YouTube-Facial-Palsy dataset, our model surpasses the one introduced by Chaoqun and colleagues by 1.06%. Our results reveal that the model can accurately interpret data, even in the face of muscle variations and other types of interference, including lighting conditions, image quality, and visual artifacts. Our model maintains stability amid major geometric shifts, such as facial asymmetry, and consistently delivers highly accurate results.

Table 4
Comparison of the facial paralysis models.

Models	Mean F1 (%)	Method
Chaoqun et al. [35]	97.14	Facial Blood Flow Measure, facial landmarks, region segmentation, 3D model fitting
Xin et al. [36]	94.2	Face landmarks, region of interest, classification
Pengfei et al. [37]	72.0	Face landmarks, symmetry separation, classification
Ting et al. [38]	95.0	Facial paralysis grading prediction, facial instance segmentation
Gee-Sern Jison et al. [39]	92.0	Face landmarks, paralysis region detection
Gemma et al. [40]	91.07	Face landmarks, Facial measure, multi – layer classification
Proposed model	98. 2	Foreground attention block, category mask depth, facial instance segmentation

4.6. Limitations

Satisfactory results were obtained for asymmetrical face parsing using the proposed method. However, there was a slight degradation in the efficiency of some of the sampled faces. This is because our trained dataset does not have sufficient samples with bearded faces. We believe that our model is a useful tool for building a trustworthy face-parsing model on a large dataset with a variety of samples.

5. Discussion

The purpose of this pilot investigation was to propose a new method for the early detection of disorders causing facial paralysis, such as Bell's palsy and stroke, using a multi-task learning network with asymmetric face parsing. The results demonstrated that our proposed model significantly outperformed other methods on all datasets for asymmetric face cases, a significant advancement in the field of medical imaging and diagnosis. The method uniquely incorporated a foreground attention block and a depth-map estimator into the network, deviating from conventional methodologies and thereby improving both the specificity and relevance of data extraction.

The foreground attention block allowed the extraction of localized features by emphasizing the region of the anticipated mask for each query, thereby improving the nuance and specificity of the analysis. This represents a departure from traditional models that prioritize the entire feature map, demonstrating our method's innovative approach to data extraction and relevance. The depth-map estimator further enhanced the model's capabilities, increasing segmentation accuracy using an instance-specific kernel approach. By estimating the depth for each category, we achieved higher precision in segmentation results Figs. 3, 4 and Table 3, making the model more accurate and useful for detecting facial weakness diseases. We validated our approach through comprehensive experiments on the YouTube-Facial-Palsy database and the private Gil Medical Center database. The performance across these databases underscores the robustness of the model and its potential for real-world application.

However, our study is not without limitations. Deep learning models, while powerful, often require extensive datasets for high accuracy. Given the privacy and ethical concerns related to facial data collection, this may pose challenges to the future development and application of the model. Furthermore, the diversity in human faces may challenge the model's generalizability, necessitating further refinement. Despite these limitations Fig. 5, this study lays the groundwork for future studies in the field of AI-driven facial asymmetry analysis. The incorporation of innovative methodologies like the foreground attention block and depth-map estimator are significant advancements that have the potential to redefine how we diagnose and manage facial weakness

Fig. 5. Illustration of limitations of the proposed model. During testing, we were faced with a challenge with the bearded face since in our used dataset, samples with the bearded face were limited. For this research, we used open access dataset, of which the creators got permission from patients.

diseases. Future research should focus on improving the model's handling of diverse datasets and navigating ethical issues related to facial data usage. The early detection of high-risk diseases such as stroke and Bell's palsy is of paramount importance, and our method provides a promising tool for achieving this goal. We hope this study sparks further research into asymmetric face parsing, pushing the boundaries of what is possible in the early detection and management of facial paralysis disorders.

6. Conclusion

This study presented a novel, deep-learning-based model for the early detection of facial paralysis disorders, such as Bell's palsy and stroke. By integrating a depth-map estimation module and a foreground attention block into a multi-task learning network, our model significantly improved upon previous diagnostic methodologies. The unique approach of focusing on the foreground region of the anticipated mask allowed us to extract localized, specific features, improving the data's relevance and our model's overall effectiveness. The model's effectiveness was validated using the YouTube-Facial-Palsy database and the private Gil Medical Center database. It notably outperformed other methods in cases of asymmetric faces, which demonstrates its potential clinical utility. This is a significant advancement as early and accurate diagnosis is critical in the management of facial paralysis disorders.

Despite the promising results, we acknowledge the limitations of our model. One such limitation is the model's slight degradation in efficiency with certain types of faces, notably those with beards. This limitation emerged due to the lack of sufficient bearded face samples in our training dataset. As with any deep learning model, ours requires a diverse and extensive dataset to perform optimally. Therefore, data privacy, ethics, and the need for a large dataset represent potential challenges that need to be addressed. Furthermore, we observed that the model's generalizability needs further improvement due to the diverse range of human faces. Future efforts should focus on addressing these limitations while further refining the model's performance.

Despite these limitations, this study lays a solid foundation for future research. Our model, with its unique focus on foreground attention and depth-map estimation, introduces fresh insights into AI-driven diagnostics. It provides a more efficient, accessible, and cost-effective solution for early detection, thus inspiring further developments in this field. Our findings highlight the importance of incorporating diverse samples into training datasets for improved model efficiency. We believe that our model can serve as a reliable tool for building a comprehensive face-parsing model with a large dataset, ensuring a wide variety of samples.

Although our study offers a significant advancement in the early detection of facial paralysis disorders, there is still room for improvement and adaptation. Future studies should aim to address the identified limitations and expand the model's capability to handle a broader range of facial types and conditions. This research signifies a step forward in the field of AI-driven diagnostics, opening up new opportunities for tackling high-risk diseases such as stroke.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Taegkeun Whangbo reports administrative support, article publishing charges, statistical analysis, and writing assistance were provided by Gachon University. Taegkeun Whangbo reports a relationship with Gachon University that includes: consulting or advisory, employment, funding grants, and paid expert testimony. Taegkeun Whangbo has patent The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgement

This work was supported by the GRRC program of Gyeonggi province. [GRRC-Gachon2023(B02), Development of AI-based medical service technology].

References

[1] WSO global stroke fact sheet 2022, https://www.world-stroke.org/news-and-blog/news/wso-global-stroke-fact-sheet-2022, WSO, 2022.
[2] Y. Ohya, N.Sato R.Matsuo, F. Irie, Y. Wakisaka, T. Ago, M. Kamouchi, T. Kitazono, Modification of the effects of age on clinical outcomes through management of lifestyle-related factors in patients with acute ischemic stroke, J. Neurol. Sci. 446 (2023), https://doi.org/10.1016/j.jns.2023.120589.
[3] J.S. Kim, S. Umirzakova, T.K. Whangbo, Machine learning based analysis model for early stroke detection, J. Next-Gen. Converg. Inf. Serv. Technol. 8 (2019) 59–70.
[4] S. Elizabeth. R G. Eskes, A. Brodtmann, Executive function poststroke: concepts, recovery, and interventions, Stroke 54 (2022) 20–29, https://doi.org/10.1161/STROKEAHA.122.037946.
[5] O.M. Sun, K.H. Yu, J.H. Lee, S. Jung, I.S. Ko, J.H. Shin, S.J. Cho, H.C. Choi, H. H. Kim, B.C. Lee, Validity and reliability of a korean version of the national institutes of health stroke scale, Epub (2012) 177–183, https://doi.org/10.3988/jcn.2012.8.3.177.
[6] S. Umirzakova, T.K. Whangbo, Detailed feature extraction network-based fine-grained face segmentation, in: Knowledge-Based Systems, 250, Elsevier, 2022, https://doi.org/10.1016/j.knosys.2022.109036.
[7] Sh. Muksimova, U. Sabina, M. Sevara, Y.I. Cho, Novel video surveillance-based fire and smoke classification using attentional feature map in capsule networks, Sensors (2021), https://doi.org/10.3390/s22010098.
[8] F. Yiming, H. Wang, X. Zhu, X. Cao, C. Yi, Y. Chen, J. Jia, X. Lu, FER-PCVT: facial expression recognition with patch-convolutional vision transformer for stroke patients, Brain Sci. (2022), https://doi.org/10.3390/brainsci12121626.
[9] A. Anuja, A. Sinha, K. Bhansali, R. Goel, I. Sharma, A. Jayal, S.V.M. and, Logistic regression for facial palsy detection utilizing, in: Proceedings of the Fourteenth International Conference on Contemporary Computing, 2022, pp. 43–48, https://doi.org/10.1145/3549206.3549216.
[10] P. Dominguez, G.S. Raul, E.S. Yanez, C.H. Garcia-Capulin, Facial paralysis detection on images using key point analysis, Appl. Sci. (2021), https://doi.org/10.3390/app11052435.
[11] G. Zhexiao, G. Dan, J. Xiang, J. Wang, W. Yang, H. Ding, O. Deussen, Y. Zhou, An unobtrusive computerized assessment framework for unilateral peripheral facial paralysis, IEEE J. Biomed. Health Inform. 22 (2018), https://doi.org/10.1109/JBHI.2017.2707588.
[12] I. Chourib, G. Guillard, I.R. Farah, B. Solaiman, Stroke treatment prediction using features selection methods and machine learning classifiers, IRBM (2022) 678–686, https://doi.org/10.1016/j.irbm.2022.02.002.
[13] G. Amira, M.F. Taher, M.A. Wahed, N.M. Shalaby, S. Gaber, Classification of facial paralysis based on machine learning techniques, Biomed. Eng. Online (2022), https://doi.org/10.1186/s12938-022-01036-0.
[14] C.M. Aldridge, M.M. McDonald, M. Wruble, Y. Zhuang, O. Uribe, T.L. McMurry, I. Lin, H. Pitchford, B.J. Schneider, W.A. Dalrymple, J.F. Carrera, Human vs. machine learning based detection of facial weakness using video analysis, Front. Neurol. (2022), https://doi.org/10.3389/fneur.2022.878282.
[15] D. Soumyabrata, H. Wang, C.S. Nwosu, N. Jain, B. Veeravalli, D. John, A predictive analytics approach for stroke prediction using machine learning and neural networks, Healthc. Anal. (2022), https://doi.org/10.1016/j.health.2022.100032.
[16] L. Xuemeng, D. Bian, J. Yu, M. Li, D. Zhao, Using machine learning models to improve stroke risk level classification methods of China national stroke screening, BMC Med. Inform. Decis. Mak. (2019), https://doi.org/10.1186/s12911-019-0998-2.
[17] S. Muhammad, T. Shafique, M. Jabbar A. Baig, I. Riaz, S. Amin, S. Manzoor, Automatic grading of palsy using asymmetrical facial features: a study complemented by new solutions, Symmetry (Basel) (2018), https://doi.org/10.3390/sym10070242.
[18] X. Pengfei, F. Xie, T. Su, Z. Wan, Z. Zhou, X. Xin, Z. Guan, Automatic evaluation of facial nerve paralysis by dual-path LSTM with deep differentiated network, Neurocomputing (2020), https://doi.org/10.1016/j.neucom.2020.01.014.
[19] H.S. Maaher, Z. Jamal, A.A. Noshin, M.M. Khan, Comparative study of deep learning algorithms for the detection of facial paralysis, in: Proceedings of the IEEE 13th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON), 2022, https://doi.org/10.1109/IEMCON56893.2022.9946491.
[20] X. Cuiting, C. Yan, M. Jiang, F. Alenezi, A. Alhudhaif, N. Alnaim, K. Polat, W. Wu, A novel facial emotion recognition method for stress inference of facial nerve paralysis patients, Expert Syst. Appl. 197 (2022), 116705, https://doi.org/10.1016/j.eswa.2022.116705.

[21] Z. Yan, M.M. McDonald, C.M. Aldridge, M.A. Hassan, O. Uribe, D. Arteaga, A. M. Southerland, G.K. Rohde, Video-based facial weakness analysis, IEEE Trans. Biomed. Eng. 68.9 (2021) 2698–2705, https://doi.org/10.1109/TBME.2021.3049739.
[22] K. Mandeep, S.R. Sakhare, K. Wanjale, F. Akter, Early stroke prediction methods for prevention of strokes, Hindawi, Behav. Neurol. (2022), https://doi.org/10.1155/2022/7725597.
[23] K.E. Sun, J.M. Heo, S.J. Eun, J.Y. Lee, Development of early-stage stroke diagnosis system for the elderly neurogenic bladder prevention, Int. Neurourol. J. (2022), https://doi.org/10.5213/inj.2244030.015.
[24] L. Taeho, E.T. Jeon, J.M. Jung, M. Lee, Deep-learning-based stroke screening using skeleton data from neurological examination videos, Per. Med. 12.10 (2022) 1691, https://doi.org/10.3390/jpm12101691.
[25] G. Peng, M. Zheng, X. Wang, J. Dai, H. Li, Fast convergence of DETR with spatially modulated co-attention, Comput. Vision Pat. Recogn. (2021).
[26] L. Jongin, S. Yun, S. Park, J.Y. Choi, Hypergraph-induced semantic tuplet loss for deep metric learning, in: , 2022, pp. 212–222.
[27] H.G.S. Jison, J.H. Kang, W.F. Huang, Deep hierarchical network with line segment learning for quantitative analysis of facial palsy, IEEE Access 7 (2018) 4833–4842, https://doi.org/10.1109/ACCESS.2018.2884969.
[28] C.L. Chieh, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, Encoder-decoder with atrous separable convolution for semantic image segmentation, in: Proceedings of the European Conference on Computer Vision (ECCV), 2018, pp. 801–818.
[29] Z. Qingping, J. Deng, Z. Zhu, Y. Li, S. Zafeiriou, Decoupled multi-task learning with cyclical self-regulation for face parsing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 4156–4165.
[30] K. Hyungjoon, H. Kim, S. Cho, E. Hwang, An end-to-end face parsing model using channel and spatial attentions, Measurement 191 (2022), 110807, https://doi.org/10.1016/j.measurement.2022.110807.
[31] H. Kaiming, G. Gkioxari, P. Dollár, R. Girshick, Mask R-CNN, in: Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2961–2969.
[32] L. Yiming, J. Shen, Y. Wang, M. Pantic, FP-age: leveraging face parsing attention for facial age estimation in the wild, EEE Trans. Image Process. (2022), https://doi.org/10.1109/TIP.2022.3155944.
[33] Y. Chang, X. Zhu, X. Zhang, Z. Wang, Z. Zhang, Z. Lei, HP-capsule: unsupervised face part discovery by hierarchical parsing, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022, pp. 4032–4041.
[34] Z. Yang, X. Yu, X. Lu, P. Liu, Pro-UIGAN: progressive face hallucination from occluded thumbnails, IEEE Trans. Image Process. (2022) 3236–3250, https://doi.org/10.1109/TIP.2022.3167280.
[35] J. Chaoqun, J. Wu, W. Zhong, M. Wei, J. Tong, H. Yu, L. Wang, Automatic facial paralysis assessment via computational image analysis, J. Healthc. Eng. 2020 (2020), https://doi.org/10.1155/2020/2398542.
[36] L. Xin, Y. Xia, H. Yu, J. Dong, M. Jian, T.D. Pham, Region based parallel hierarchy convolutional neural network for automatic facial nerve paralysis evaluation, IEEE Trans. Neural Syst. Rehabil. Eng. 28 (2020) 2325–2332, https://doi.org/10.1109/TNSRE.2020.3021410.
[37] X. Pengfei, F. Xie, T. Su, Z. Wan, Z. Zhou, X. Xin, Z. Guan, Automatic evaluation of facial nerve paralysis by dual-path LSTM with deep differentiated network, Neurocomputing 388 (2020), https://doi.org/10.1016/j.neucom.2020.01.014.
[38] W. Ting, S. Zhang, L. Liu, G. Wu, J. Dong, Automatic facial paralysis evaluation augmented by a cascaded encoder network structure, IEEE Access 7 (2019) 135621–135631, https://doi.org/10.1109/ACCESS.2019.2942143.
[39] H.G.S. Jison, W.F. Huang, J.H. Kang, Hierarchical network for facial palsy detection, in: Procedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2018, pp. 580–586.
[40] P.D. Gemma, R.E.S. Yanez, C.H.G. Capulin, Facial paralysis detection on images using key point analysis, Appl. Sci. (2021) 2435, https://doi.org/10.3390/app11052435.
[41] C. Hu, Y. Li, Z. Feng, X. Wu, Attention-guided evolutionary attack with elastic-net regularization on face recognition, Pattern Recognit. 143 (2023), https://doi.org/10.1016/j.patcog.2023.109760.