Leona Schild*, Jana Zang, Till Flügel, Deike Weiss, Alexander Schlaefer, and Sarah Latus

Automated Detection of Palatal Deformations Using Deep Learning on Endoscopic Images

https://doi.org/10.1515/cdbme-2024-1017

Abstract: A deformation of the hard palate can occur in spinal muscular atrophy and leads to problems with feeding and swallowing in early childhood. An objective analysis of the palatal changes is therefore desirable for early treatment initiation. In this study, we investigate a deep learning approach to automatically detect deformation in endoscopic images which were collected in a prospective in-vivo study on 33 infants. Ratings of five different experts were used to quantify the deformation and to train our network. We investigate different network architectures and data set splits and achieve classification performances of up to 0.85 ± 0.05 when distinguishing between normal and deformation using the EfficientNet architecture. This combination of endoscopic imaging and deep learning offers a first approach for the objective assessment of palatal changes.

Keywords: palatal deformations, endoscopic analysis, convolutional neural networks, spinal muscular atrophy

1 Introduction

Spinal muscular atrophy (SMA) is a genetic disease that affects around 12 in 100,000 children [1]. This disease is associated with a degeneration of motor neurons resulting in muscle atrophy and weakness [2] and often associated with a swallowing disorder [3]. Before the introduction of new disease-modifying therapies, the disease was classified into different types, reflecting the severity of the symptoms that range from severe hypotonia and survival rates of a maximum of two years in 90% of the patients (Type I) to minor muscular weakness correlated with scoliosis or osteoporosis in adulthood (Type III) [4]. In Type I, the frequent occurrence of a hyper-ogival palate was described [4] which was recently identified as a risk marker of sudden unexpected death in infancy [5]. So far, no effective medical treatment strategies have been found, and current research is focusing on the search for suitable therapeutic strategies [2]. In order to plan treatment, the stage of the disease must be determined, and identifying the onset of palatal deformity is a crucial step.

Various methods have been proposed to assess the geometry of the palate, e.g., by taking plaster casts [6], creating digitized maxillary dental casts [7], acquiring cone beam CT images [8], or analyzing trans oral images acquired with a mobile device [9]. However, none of these approaches are suitable for repeated use for close monitoring of infants or toddlers. For these patients, it is crucial to find minimally invasive methods that do not require anesthesia or a high level of patient cooperation. Therefore, we investigate the application of a thin flexible chip-tip Videoendoscope with 3.2mm diameter (PatCom Medical GmbH) to examine the palate of healthy infants and SMA patients.

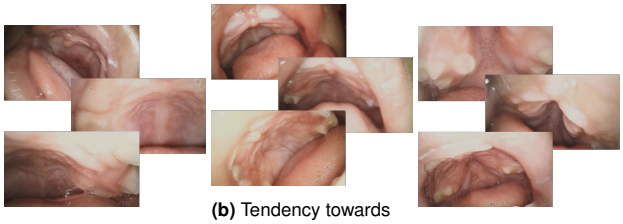
In several studies, the benefit of deep learning for classification, segmentation, or anomaly detection tasks has been shown, e.g., by improving the efficiency and accuracy of diagnosis [10]. Especially convolutional neural networks (CNNs) have been proven to be efficient in processing medical image data, including endoscopy [11]. In this work, we present a deep learning approach for an automated analysis of the endoscopic image data to assist early detection of palatal deformations. We use the assessment of five clinical experts as the basis for our training. In our experiments, we investigate different deep-learning architectures for deformation classification and the impact that data sets of individual patients have on the classification. In particular, we are examining whether it is possible to use deep learning to detect deformations even in data sets where experts could not find a consensus.

2 Material and Methods

2.1 Data set

As part of the study "OSMA (Objective capture of swallow-related outcomes in infants and young children with SMA)" carried out by the University Medical Center Hamburg-Eppendorf we acquired images of the palate using a flexible chip-tip Videoendoscope. Patients are stimulated to open their mouths and, while opening the mouth, image data of the palate is recorded from various directions. Depending on the coop-

***Corresponding author: Leona Schild,** Hamburg University of Technology, Institute of Medical Technology and Intelligent Systems, Am Schwarzenberg-Campus 3, Hamburg, Germany, e-mail: leona.schild@tuhh.de
Alexander Schlaefer, Sarah Latus, Hamburg University of Technology, Institute of Medical Technology and Intelligent Systems, Hamburg, Germany
Jana Zang, Till Flügel, Deike Weiss, University Medical Center Hamburg-Eppendorf, Martinistrasse 52, Hamburg, Germany



(a) No anomalies **(b)** Tendency towards deformation **(c)** Visible Deformation

Fig. 1: Example images for each medical classification

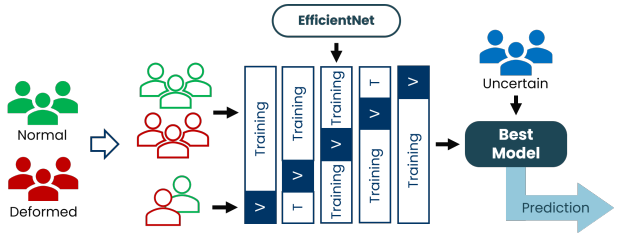


Fig. 2: For the five-fold cross-validation each patient group is used as validation data while data from the other groups are used in the training (ratio: 20% to 80%). The trained models based on the EfficientNet architecture, are evaluated using the test data to identify the optimal model. This best-performing model is then employed to classify the uncertain data.

erativity of the patients, image sequences could be recorded, varying in length between 2 and 19 seconds. Our data set consists of a total of 44 video sequences of 11 healthy subjects and 22 patients suffering from SMA, that have been converted to 8304 images. The SMA patients show different stages of palate deformation. To establish ground truth for the classification, we conducted a blinded grading study in which one set of images per subject was evaluated by five clinical experts, i.e., three otorhinolaryngologists and two speech therapists. These experts classified the image data set per subject as showing "no anomalies" (N), a "tendency towards deformation" (T), or a "visible deformation" (D) as exemplary shown in Fig. 1.

To evaluate the statistical evidence of the expert's classifications we calculate the non-parametric interobserver variability. Based on the Fleiss' approach [12] we observe a kappa value of 0.29 for all examiners when distinguishing between the three categories, which refers to a fair agreement [13]. The score shows a high variability in the classification and rather subjective assessments. If only considering the rating of N and D, a Fleiss' kappa of 0.83 is obtained, showing that a binary classification is more accurate and appropriate for training the model. To address the observer variability, we define that there must be at least four matching expert opinions for N and at least three matching experts for D to assign the respective data to the class. This ensures that normal patients are classified as normal by the vast majority, while the threshold for deformed patients is set lower to be more sensitive to deformity. The remaining data sets that are classified as having a tendency or where the opinions do not match are put aside as uncertain patients. The characteristics of the two classes and the uncertain data are shown in Tab. 1.

Tab. 1: Class characteristics. The information on subject age are given in months.

Class	Patients	Pictures	Age range	Mean age
Normal	17	2599	0-21	8.2
Uncertain	18	3111	0-35	11.7
Deformed	8	2594	14-80	33.1

2.2 Deep Learning Methods

We design a deep learning approach for the classification of palate deformation using different pre-trained CNN architectures. Given the endoscopic images of one patient as input, the trained network should predict the probability of the two deformation classes. In a preliminary study, we investigate the classification performance for four architectures that have different numbers of layers and parameters, namely ResNet-18, ResNet-50, DenseNet, and EfficientNet. Based on the resulting performance, EfficientNet is selected for the cross-validation described in the following, as it is particularly well suited to performing this classification task due to its significantly larger number of layers.

To evaluate the ability of the network to classify the clinical data, a five-fold cross-validation is carried out with individual patients in the test data set. Thereby, overfitting of the data is avoided and the differences between various probands and their effects on the performance of the learning are investigated. As there are big differences in the size of the patients' data sets, groups (Norm1-Norm7, Def1-Def6) are created to generate approximately equal quantities while respecting the ground truth annotations as well as the probands' age (i.e. Norm1 patients are of age 0m and Norm7 of ages 16 to 21 months). The networks are trained for 200 epochs using the Adam optimizer and a learning rate 0.0001. The EfficientNet is pre-trained on ImageNet and the implementation is done in Python using Pytorch Version 2.2.2. and Keras Version 2.15.0.

For the last step, we identify the best model by choosing the model with the best combination of the metrics test accuracy, precision, recall and f1-score. The entire process is illustrated in Fig. 2. To compare the clinicians' ratings with the network's classification, we then show all data, including the uncertain set, to the corresponding model. As the medical professionals only had the option to rate between three categories N, T and D, we need to convert the scoring into a binary probability.

Therefore the probability of clinical deformation is defined as

$$P_{D,med} = \frac{1}{5} \cdot (0 \cdot N_N + 0.5 \cdot N_T + 1 \cdot N_D),$$

where N_i is the number of votes for each category (N: Normal, T: Tendency, D: Deformed). As five clinicians evaluated the data the maximum score would be 1 (all clinicians see deformation) and the minimum score would be 0 (all clinicians rate the patient as normal).

3 Results

The results of the cross-validation over different patient groups are depicted in Tab. 2. The comparison of the clinical probability and the network's evaluation is shown in Fig. 3. The associated Pearson coefficient of the two ratings is 0.76.

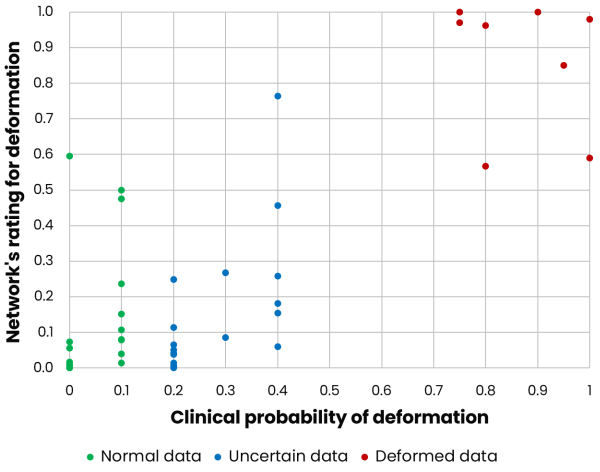


Fig. 3: Comparison of clinicians' score with the classification rate of the network (0: No deformation, 1: Clear deformation)

Tab. 2: Results of cross-validation for testing on separated patient groups from each class using the EfficientNet.

(a) Normal

	Accuracy	Precision	Recall	F1
Norm1	0.76±0.07	0.80±0.05	0.76±0.07	0.75±0.09
Norm2	0.72±0.07	0.74±0.06	0.72±0.07	0.70±0.09
Norm3	**0.80±0.09**	**0.84±0.06**	**0.80±0.09**	**0.79±0.11**
Norm4	0.71±0.08	0.73±0.07	0.71±0.08	0.70±0.09
Norm5	0.78±0.08	0.80±0.07	0.78±0.08	0.78±0.10
Norm6	0.76±0.07	0.78±0.06	0.76±0.07	0.75±0.08
Norm7	0.77±0.06	0.78±0.05	0.77±0.06	0.77±0.06

(b) Deformed

	Accuracy	Precision	Recall	F1
Def1	0.77±0.03	0.79±0.03	0.77±0.03	0.77±0.04
Def2	0.65±0.04	0.71±0.05	0.65±0.04	0.62±0.06
Def3	0.80±0.06	0.82±0.05	0.80±0.06	0.79±0.07
Def4	**0.85±0.05**	**0.85±0.04**	**0.85±0.05**	**0.85±0.05**
Def5	0.75±0.09	0.76±0.08	0.75±0.09	0.74±0.10
Def6	0.75±0.10	0.76±0.10	0.75±0.10	0.74±0.10

4 Discussion

The evaluation of the cross-validation according to different patient groups in Tab. 2 shows that there are differences in the detectability of the individual classes as well as for the individual patient groups within a class. The mean values of the metrics are between 0.70 and 0.84 for class normal and between 0.62 and 0.85 for class deformed. It should be noted that the general variability in the metrics is higher for the deformed patients, implying different grades and variations of the disease while healthy patients show a more uniform picture.The impact of age on categorization appears inconclusive, as the performance metrics for older patients do not show significant improvement or deterioration. Nonetheless, it is important to note, as demonstrated in Tab. 1, that the patients with deformities have a higher average age because the deformation becomes more clearly visible beyond a certain age threshold.

When evaluating all patients separately using the best-performing model (Fig. 3), it is noticeable that there is an evident linear relationship between the network's rating and the clinicians' classification. This is further reflected in the high value of the Pearson correlation. However, some outliers indicate, that the network does not always agree with the clinicians' opinions. An example is the point at (0.57 | 0.8), which represents one of the patients from the Def2 group shown above. This does not appear to be perceived as deformed by the network, as already shown by the poorer performance in Tab. 2. As this patient was also only rated by three experts as deformed and by two experts as a tendency, the network's rating of 0.57 reflects the observer variability.

Overall, it can be concluded that it is possible to classify the available data using suitable networks. The classification shows very satisfactory results given the limitations of the small data set and the variance of data quantity between the patients. However, the uncertainty of annotations by the clinical experts should be addressed in future studies with a more standardized evaluation scheme.

5 Conclusion

This paper showed that an automatic assessment of palatal deformity from endoscopic images can be achieved using deep learning methods. Our results demonstrate that a classification in normal and deformation is possible. Thus, the method has great potential to assist in clinical detection, especially in settings where multiple expert opinions are not available. It is an important step towards preventing swallowing disorders in SMA patients and thus improving their chances and duration of survival. In future work, a more standardized clinical classification procedure could be developed to minimize the uncertainties in annotation by specialists. Once a much larger data set is available, anomaly detection seems to be a promising alternative method for this problem, as it has already been successfully applied for medical image analysis [14], particularly for endoscopy images [15].

Author Statement

Research funding: This work was partially funded by the i^3 initiative of the Hamburg University of Technology, and partially by the Interdisciplinary Competence Center for Interface Research (ICCIR) on behalf of the University Medical Center Hamburg-Eppendorf and the Hamburg University of Technology. The data collection as part of the OSMA study was funded by the initiative "Forschung und Therapie für die Spinale Muskelatrophie". The local ethics committee approved the study (2022-100827_2-BO-ff). Conflict of interest: Authors state no conflict of interest. Informed consent: Informed consent has been obtained from all individuals included in this study. Ethical approval: The research related to human use complies with all the relevant national regulations, institutional policies and was performed in accordance with the tenets of the Helsinki Declaration.

References

[1] Verhaart IE, Robertson A, Leary R, McMacken G, König K, Kirschner J, et al. A multi-source approach to determine SMA incidence and research ready population. Journal of neurology. 2017;264:1465-1473.

[2] Kolb SJ, Kissel JT. Spinal Muscular Atrophy. Neurol Clin. 2015 Nov;33(4):831-46. doi: 10.1016/j.ncl.2015.07.004. PMID: 26515624; PMCID: PMC4628728.

[3] Zang J, Johannsen J, Denecke J, Weiss D, Koseki JC, Nießen A, et al. Flexible endoscopic evaluation of swallowing in children with type 1 spinal muscular atrophy. European Archives of Oto-Rhino-Laryngology. 2023;280(3):1329-1338.

[4] Audic F, Barnerias C. Spinal muscular atrophy (SMA) type I (Werdnig-Hoffmann disease). Archives de Pédiatrie. 2020;27(7):7S15-7S17.

[5] Ducloyer M, Wargny M, Medo C, Gourraud PA, Clement R, Levieux K, et al. The Ogival Palate: A New Risk Marker of Sudden Unexpected Death in Infancy? Front Pediatr. 2022 Apr 18;10:809725. doi: 10.3389/fped.2022.809725. PMID: 35509830; PMCID: PMC9058094.

[6] Hohoff A, Stamm T, Meyer U, Wiechmann D, Ehmer U. Objective growth monitoring of the maxilla in full term infants. Archives of Oral Biology. 2006;51(3):222-235.

[7] Croquet B, Matthews H, Mertens J, Fan Y, Nauwelaers N, Mahdi S, et al. Automated landmarking for palatal shape analysis using geometric deep learning. Orthodontics & Craniofacial Research. 2021;24:144-152.

[8] Wang X, Pastewait M, Wu TH, Lian C, Tejera B, Lee YT, et al. 3D morphometric quantification of maxillae and defects for patients with unilateral cleft palate via deep learning-based CBCT image auto-segmentation. Orthodontics & Craniofacial Research. 2021;24:108-116.

[9] Rourke R, Weinberg SM, Marazita ML, Jabbour N. Diagnosing subtle palatal anomalies: Validation of video-analysis and assessment protocol for diagnosing occult submucous cleft palate. International Journal of Pediatric Otorhinolaryngology. 2017;100:242-246.

[10] Farhad M, Masud MM, Beg A, Ahmad A, Ahmed L. A Review of Medical Diagnostic Video Analysis Using Deep Learning Techniques. Applied Sciences. 2023;13(11):6582.

[11] Choi J, Shin K, Jung J, Bae H J, Kim D H, Byeon J S, & Kim N (2020). Convolutional neural network technology in endoscopic imaging: artificial intelligence for endoscopy. Clinical endoscopy, 53(2), 117-126.

[12] Falotico R, Quatto P. Fleiss' kappa statistic without paradoxes. Quality & Quantity. 2015;49:463-470.

[13] Zühlke D, Geweniger T, Heimann U, & Villmann T (2009). Fuzzy Fleiss-kappa for Comparison of Fuzzy Classifiers. The European Symposium on Artificial Neural Networks.

[14] Behrendt F, Bengs M, Rogge F, Krüger J, Opfer R, & Schlaefer A (2022, March). Unsupervised Anomaly Detection in 3D Brain MRI using Deep Learning with impured training data. In 2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI) (pp. 1-4). IEEE.

[15] Hu E, Nosato H, Sakanashi H, Murakawa M. Anomaly detection for capsule endoscopy images using higher-order local auto correlation features. In: 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC); October 2012. IEEE; pp. 2289-2293.