Deep Learning Analysis to Automatically Detect the Presence of Penetration or Aspiration in Videofluoroscopic Swallowing Study

Jeoung Kun Kim
Yeungnam University
Yoo Jin Choo
Yeungnam University
Gyu Sang Choi
Yeungnam University
Hyunkwang Shin
Yeungnam University
Min Cheol Chang
Yeungnam University
Donghwi Park (bdome@hanmail.net)
Yeungnam University

Research Article

Keywords: Deep learning, VFSS, deglutition, swallowing reflex

Posted Date: September 20th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-884186/v1

Version of Record: A version of this preprint was published at Journal of Korean Medical Science on January 1st, 2022. See the published version at https://doi.org/10.3346/jkms.2022.37.e42.

Abstract

Background: Videofluoroscopic swallowing study (VFSS) is currently considered the gold standard to precisely diagnose and quantitatively investigate dysphagia. However, VFSS interpretation is complex and requires consideration of several factors.

Purpose: Therefore, considering the expected impact on dysphagia management, this study aimed to apply deep learning to detect the presence of penetration or aspiration in VFSS of patients with dysphagia automatically.

Materials and Methods: The VFSS data of 190 participants with dysphagia were collected. A total of 10 frame images from one swallowing process were selected (five high-peak images and five low-peak images) for the application of deep learning in a VFSS video of a patient with dysphagia. We applied a convolutional neural network (CNN) for deep learning using the Python programming language. For the classification of VFSS findings (normal swallowing, penetration, and aspiration), the classification was determined in both high-peak and low-peak images. Thereafter, the two classifications determined through high-peak and low-peak images were integrated into a final classification.

Results: The area under the curve (AUC) for the validation dataset of the VFSS image for the CNN model was 0.946 for normal findings, 0.885 for penetration, and 1.000 for aspiration. The average AUC was 0.962.

Conclusion: This study demonstrated that deep learning algorithms, particularly the CNN, could be applied for detecting the presence of penetration and aspiration in VFSS of patients with dysphagia.

Summary

Summary: The deep learning algorithms, particularly the CNN, could be applied for detecting the presence of penetration and aspiration in VFSS of patients with dysphagia.

Key Results: The area under the curve (AUC) for the validation dataset of the VFSS image for the CNN model was 0.946 for normal findings, 0.885 for penetration, and 1.000 for aspiration. The average AUC was 0.962.

Introduction

The swallowing process includes the coordinated contraction and relaxation of the muscles of the tongue, pharynx, larynx, and esophagus, which is controlled by the central nervous system (CNS) from the brain cortex to the brainstem [1–3]. Any lesion in the path from the CNS to the swallowing muscles can cause difficulty in swallowing, which is referred as dysphagia [4, 5]. Dysphagia is a common clinical symptom in patients with cerebrovascular, neuromuscular, and neurodegenerative diseases and with head and neck cancers [6–8]. The videofluoroscopic swallowing study (VFSS) is currently considered the

gold standard to accurately diagnose and quantitatively analyze dysphagia [9]. Clinicians repeatedly perform a frame-by-frame analysis of spatiotemporal and quantitative parameters in a recorded VFSS video to determine the cause of dysphagia and the appropriate diet [10–13]. Therefore, despite being able to objectively observe the entire process of swallowing through VFSS, its interpretation is complex and needs consideration of several factors [9].

Recently, deep learning, a technique in artificial intelligence wherein the system learns rules and patterns from the given information, has been increasingly studied in the medical field [14]. Deep learning has several advantages in terms of detecting the possible interactions between attributes or variables; hence, it may be useful in diagnosis and prediction [15].

The application of the recent developments in deep learning research could reduce the burden over clinicians caused by the complexity of VFSS interpretation. Moreover, to date, no research pertaining to deep learning has been directed to detect the presence of penetration or aspiration in VFSS of patients with dysphagia. Therefore, considering the expected impact on dysphagia management, this study aimed to apply deep learning to detect penetration or aspiration in VFSS of patients with dysphagia automatically.

Materials And Methods

This study was approved by the Institutional Review Board of Yeungnam University hospital. It was approved by the Institutional Review Board of Yeungnam University hospital that informed consent was not required due to the retrospective nature of the study and the use of anonymous clinical data. All procedures were carried out in accordance with the relevant guidelines and regulations. We included patients who visited the outpatient clinic of the rehabilitation department, who were admitted to the rehabilitation department of one of the two university hospitals (Ulsan university hospital and Yeungnam university hospital) because of dysphagia, or who were diagnosed using VFSS between January 2009 and April 2020. The steps of the modeling process applied in this study are shown in Fig. 1.

Data collection

The VFSS data of 190 participants with dysphagia were collected. The exclusion criteria were as follows: (1) patients of age less than 20 years; (2) patients who had undergone tracheostomy; (3) patients with facial or cranial anomalies; and (4) patients having metal plate in the cervical spine or facial bone that could develop an artifact.

Analysis of VFSS

When the VFSS was performed, the patients were instructed to seat upright under a videofluoroscopy machine with the head in a neutral position. Boundaries for the frame of videofluoroscopy included the incisors anteriorly, cervical vertebrae posteriorly, nasal border of the soft palate superiorly, and cervical

esophagus inferiorly [16, 17]. The fluoroscopic images of swallows were digitally recorded and stored at 30 frames/s [16, 17].

Each VFSS was performed using a bolus of “thin” fluid (1–50 cP). Each patient received a 5-ml bolus delivered using a 10-ml syringe [16, 17].

In the analysis of VFSS, the presence of penetration was determined when the contrast material passed above the true vocal cord, and not below [18]. The presence of aspiration was determined when the contrast material passed below the true vocal cord [18]. Based on the above criteria, the presence or absence of penetration or aspiration in the dynamic fluoroscopic images was reviewed by two rehabilitation medicine specialists with more than 10 years of clinical experience in dysphagia. Based on the VFSS, patients were classified into normal (without penetration and aspiration), penetration, and aspiration groups.

VFSS image selection

To analyze VFSS by deep learning, we selected five consecutive frame images (at 0.33-s intervals) from the VFSS, back and forth, when the hyoid bone reached the peak (the highest position of the hyoid bone; high-peak image), and another five consecutive frame images from the VFSS when the hyoid bone completely descended from the peak (the lowest position of the hyoid bone; low-peak image) (Fig. 1). Therefore, 10 frame images were selected from one swallowing process (five high-peak images and five low-peak images) for the application of deep learning in the VFSS video of a patient with dysphagia (Fig. 1).

Deep learning analysis

We applied a convolutional neural network (CNN) for deep learning using the Python programming language. TensorFlow 2.4, the Keras framework, and scikit-learn toolkit 0.24.1 were used to train the CNN model. To achieve better learning outcomes, we employed a pre-trained CNN model with fine-tuning. The details and performance of the model are described in Table 2. A CNN consists of one or more convolutional layers, often with a subsampling layer; the convolutional layers are followed by one or more fully connected layers, similar to that in a standard neural network [19]. The deep learning models were trained using VFSS images as inputs to classify patients with dysphagia into normal (no penetration and aspiration), penetration, or aspiration groups. Of the study population (total 190 patients), 70% (n = 133), 20.53% (n = 39), and 9.47% (n = 18) were included in the training, validation, and test sets, respectively. Additionally, of the 950 images each for high-peak and low-peak images, 70% (665 images), 20.53% (195 images), and 9.47% (90 images) were used for training, validation, and test, respectively.

For obtaining the classification model according to VFSS findings (normal, penetration, and aspiration), the classification was initially conducted in both high-peak and low-peak images. We applied the following classification criteria: 1) normal: ≥ 4 normal images (of five images [separately for high-peak and low-peak images]); 2) penetration: < 4 normal images and no aspiration image; and 3) aspiration: < 4

normal images and ≥ 1 aspiration images. The two classifications from the high-peak and low-peak images were integrated into a final classification according to the following criteria: 1) normal: normal in both high-peak and low-leak images; 2) penetration: ≤ 1 normal (in the two classification results) and no aspiration; and 3) aspiration: ≤ 1 normal and ≥ 1 aspiration (Table 3).

Statistical analysis

Statistical analyses were performed using Python 3.7.9 and scikit-learn version 0.24.1. Receiver operating characteristic curve analysis was performed, and the area under the curve (AUC) was calculated. The confidence interval for the average AUC was calculated as bias-corrected and accelerated using the R 4.0.5 and multiROC 1.1.1 package [20].

Results

A total of 190 patients (mean age, 66.83 ± 15.47 years; 92 men, 88 women) were included in this study (Table 1). Of the 190 patients, 113 (59.47%) patients were classified in the normal group (no penetration and aspiration), 32 (16.84%) patients in the penetration group, and 45 (23.68%) patients in the aspiration group (Table 2). Additionally, of the 950 high-peak images of 190 patients, 590 images (62.11%) were normal, and 147 (15.47%) and 213 images (22.42%) showed penetration and aspiration, respectively. Of the 950 low-peak images of 190 patients, 700 (73.68%), 40 (4.21%), and 210 (22.11%) showed normal, penetration, and aspiration findings, respectively.

Table 1
Characteristics of patients with dysphagia who were included in this study.

Characteristics	Numbers
Age (years)	66.83 ± 15.47
Sex (male : female)	92 : 88
Normal : Penetration : Aspiration	113 (59.47%) : 32 (16.84%) : 45 (23.68%)
Cause	
stroke	92 (48.42%)
spinal cord injury, cervical level	16 (8.42%)
parkinson`s disease	15 (7.89%)
motor neuron disease	19 (10.00%)
dementia	23 (12.11%)
deconditioning	25 (13.16%)

Table 2
Performances of the deep-learning model.

Sample size (patients) **Sample ratio(patients)**	**133, 70% for training, 39, 20.53% for validation, 18, 9.47% for test, total 190** **Normal: 113, 59.47%; penetration: 32, 16.84%; aspiration: 45, 23.68%**	
Sample size (images)	665, 70% for training, 195, 20.53% for validation, 90, 9.47% for test, total 950 each for high-peak and low-peak images	
Sample ratio(images)	Normal: 590, 62.11%; penetration: 147, 15.47%; aspiration: 213, 22.42% for high-peak images Normal: 700, 73.68%; penetration: 40, 4.21%; aspiration: 210, 22.11% for low-peak images	
DNN model	Model for high-peak images	Model for low-peak images
	- MobileNet CNN model with SGD optimizer - Training accuracy: 100% - Validation accuracy: 95.38% - Test accuracy: 94.44%	- MobileNet CNN model with SGD optimizer - Training accuracy: 100% - Validation accuracy: 95.90% - Test accuracy: 92.22%
VFSS classifier	Classifier of high-peak images for individual patient - Training accuracy: 100% - Validation accuracy: 94.87% - Test accuracy: 94.44%	Classifier of low-peak images for individual patient - Training accuracy: 100% - Validation accuracy: 97.44% - Test accuracy: 94.44%
VFSS integrated classifier	- Training accuracy: 100%, validation accuracy: 94.87%, test accuracy: 100% - Validation ROC AUC for normal 0.946, penetration 0.885, aspiration 1.000 - Validation micro average ROC AUC 0.962, CI [0.943–0.992] - Test ROC AUC for normal 1.000, penetration 1.000, aspiration 1.000 - Test average ROC AUC 1.000	

ML, machine learning; DNN, deep neural network; CNN, convolutional neural network; SGD, stochastic gradient descent; ROC, receiver operating characteristics; AUC, area under the curve; CI, confidence interval

Table 3
The criteria for the integration of the classification results of high-peak and low-peak images.

Classification model	Dysphagia classification criteria
Initial classifier in each high-peak and low-peak images	Normal : NI > = 4 Penetration : NI < 4 and AI = 0 Aspiration : NI < 4 and AI > = 1
Integrated classifier (final decision)	Normal : N = 2 Penetration : N < = 1 and A = 0 Aspiration : N < = 1 and A > = 1
NI, normal image; AI, aspiration image, N, normal decision ; A, aspiration decision	

The AUC of the validation dataset of the VFSS images for the CNN model was 0.946 for normal findings, 0.885 for penetration, and 1.000 for aspiration. For calculating the average AUC, the micro average AUC was employed because the class imbalance was high. The micro average AUC was 0.962. However, the AUCs of the test dataset of the VFSS images for the CNN model were 1.000 for normal, penetration, and aspiration findings (Table 2; Fig. 2).

Discussion

To the best of our knowledge, this study is the first to use deep learning to detect the presence of penetration or aspiration in VFSS of patients with dysphagia. The results of this study are promising, and the study has high accuracy. Considering that AUCs of 0.7–0.8, 0.8–0.9, and > 0.9 are generally considered acceptable, excellent, and outstanding, respectively, the ability of deep learning models used in this study to detect normal swallowing, penetration, or aspiration is outstanding [21].

While neural networks and other pattern detection methods have been utilized for the past 50 years, recently, there has been a significant development in the field of CNN [14]. The multiple convolutional layers of the CNN model may be more appropriate for classifying the clinical outcome based on radiologic or other image-based data because of the characteristics of the model such as ruggedness to shifts and distortion in images, limited memory requirement, and easier and better training [19]. Detection of a particular finding using CNN has been reported to be rugged to distortions such as changes in shape caused by different poses, lighting conditions, and camera angles, presence of partial occlusions, and horizontal and vertical shifts, if a considerable amount of data set is sufficiently trained [19]. Moreover, in the convolutional layer of the CNN, the same coefficients are used across different locations in space; hence, the memory requirement is drastically reduced [19].

Several methods of deep learning-based VFSS analysis have been reported in previous studies (9, 10, 22). Using the single-shot multi-box detector, one of the state-of-the-art deep learning methods for object

detection, Zhang et al. [22] developed a tracking system for the detection of the hyoid bone. However, the analysis of motion or action in VFSS videos is difficult using this method, because the technique focuses on the detection of a spatial region on a single image rather than on the analysis of a sequence of images from videographic data. Lee et al. [9, 10] reported a state-of-the-art video analysis method using an integrated three-dimensional convolutional network for the detection of the pharyngeal phase and for analyzing the swallowing reflex in a VFSS video without manual spatial annotations. While the detection of the pharyngeal phase and analysis of the swallowing reflex are useful for shortening the time required for VFSS by the clinician, they have limitations in that both require further analysis to d etermine the status of the patients.

To date, most VFSS-based deep learning studies have focused on tracking anatomical structures such as hyoid bones, analyzing the pharyngeal phase, or recording the swallowing reflex time. However, in clinical settings, the most important implication of VFSS is detection of the presence of penetration or aspiration. Therefore, unlike previous studies, the deep learning program developed in this research would be useful to physicians in clinical settings.

There are a few limitations to this study. We could not input the entire video of VFSS for deep learning analysis; we trained the CNN model only by selecting two sets of five consecutive frame images from VFSS of patients with dysphagia. However, in VFSS, we believe that penetration or aspiration usually develops in two phases. If the primary cause of penetration or aspiration is delayed swallowing reflex or reduced laryngeal elevation, the penetration or aspiration usually develops when the hyoid bone is at the high-peak. In the low-peak, over-flow penetration or aspiration can also develop when the amount of pyriformis or vallecular sinus residue increases while the hyoid bone descends (at the end of the swallowing process). Therefore, five consecutive VFSS images in both positions of the hyoid bone (high-peak and low-peak) include considerable moments of penetration and aspiration in VFSS video. This hypothesis was proven correct according to the results of this study, using VFSS with deep learning by means of a CNN, which showed high accuracy. However, for more accurate analysis, deep learning analysis of complete VFSS video images will be necessary in the future.

Conclusion

This study demonstrated that deep learning algorithms, particularly the CNN, could be applied for detecting the presence of penetration and aspiration in VFSS of patients with dysphagia.

Abbreviations

VFSS; Videofluoroscopic swallowing study, CNN; convolutional neural network, AUC; area under the curve, CNS; central nervous system,

Declaration

Author Contribution Statement

Jeoung Kun Kim : Data acquisition & analysis, wrote the main manuscript text

Yoo Jin Choo : Data acquisition

Gyu Sang Choi : Data acquisition & analysis

Hyunkwang Shin : Data acquisition & analysis

Min Cheol Chang : Wrote the main manuscript text

Donghwi Park : Data acquisition & wrote the main manuscript text

References

1. Steele CM, Miller AJ. Sensory input pathways and mechanisms in swallowing: a review. Dysphagia. 2010;25:323-33.
2. Ertekin C. Electrophysiological evaluation of oropharyngeal Dysphagia in Parkinson's disease. Journal of movement disorders. 2014;7:31-56.
3. Park S, Cho JY, Lee BJ, Hwang JM, Lee M, Hwang SY, et al. Effect of the submandibular push exercise using visual feedback from pressure sensor: an electromyography study. Sci Rep. 2020;10:11772.
4. Shaker R, Geenen JE. Management of Dysphagia in stroke patients. Gastroenterol Hepatol (N Y). 2011;7:308-32.
5. Finestone HM, Greene-Finestone LS. Rehabilitation medicine: 2. Diagnosis of dysphagia and its nutritional management for stroke patients. CMAJ. 2003;169:1041-4.
6. Yeom J, Song YS, Lee WK, Oh BM, Han TR, Seo HG. Diagnosis and Clinical Course of Unexplained Dysphagia. Ann Rehabil Med. 2016;40:95-101.
7. Chang MC, Park JS, Lee BJ, Park D. Effectiveness of pharmacologic treatment for dysphagia in Parkinson's disease: a narrative review. Neurol Sci. 2021;42:513-9.
8. Yu KJ, Park D. Clinical characteristics of dysphagic stroke patients with salivary aspiration: A STROBE-compliant retrospective study. Medicine (Baltimore). 2019;98:e14977.
9. Lee JT, Park E, Hwang JM, Jung TD, Park D. Machine learning analysis to automatically measure response time of pharyngeal swallowing reflex in videofluoroscopic swallowing study. Sci Rep. 2020;10:14735.
10. Lee JT, Park E, Jung TD. Automatic Detection of the Pharyngeal Phase in Raw Videos for the Videofluoroscopic Swallowing Study Using Efficient Data Collection and 3D Convolutional Networks (dagger). Sensors (Basel). 2019;19.

11. Park D, Lee HH, Lee ST, Oh Y, Lee JC, Nam KW, et al. Normal contractile algorithm of swallowing related muscles revealed by needle EMG and its comparison to videofluoroscopic swallowing study and high resolution manometry studies: A preliminary study. J Electromyogr Kinesiol. 2017;36:81-9.
12. Park D, Oh Y, Ryu JS. Findings of Abnormal Videofluoroscopic Swallowing Study Identified by High-Resolution Manometry Parameters. Archives of physical medicine and rehabilitation. 2016;97:421-8.
13. Park D, Shin CM, Ryu JS. Effect of Different Viscosities on Pharyngeal Pressure During Swallowing: A Study Using High-Resolution Manometry. Archives of physical medicine and rehabilitation. 2017;98:487-94.
14. Lundervold AS, Lundervold A. An overview of deep learning in medical imaging focusing on MRI. Z Med Phys. 2019;29:102-27.
15. Ahmed Z, Mohamed K, Zeeshan S, Dong X. Artificial intelligence with multi-functional machine learning platform development for better healthcare and precision medicine. Database (Oxford). 2020;2020.
16. Martin-Harris B, Jones B. The videofluorographic swallowing study. Phys Med Rehabil Clin N Am. 2008;19:769-85, viii.
17. Pauloski BR, Rademaker AW, Lazarus C, Boeckxstaens G, Kahrilas PJ, Logemann JA. Relationship between manometric and videofluoroscopic measures of swallow function in healthy adults and patients treated for head and neck cancer with various modalities. Dysphagia. 2009;24:196-203.
18. Uhm KE, Yi SH, Chang HJ, Cheon HJ, Kwon JY. Videofluoroscopic swallowing study findings in full-term and preterm infants with Dysphagia. Ann Rehabil Med. 2013;37:175-82.
19. Yamashita R, Nishio M, Do RKG, Togashi K. Convolutional neural networks: an overview and application in radiology. Insights Imaging. 2018;9:611-29.
20. DeLong ER, DeLong DM, Clarke-Pearson DL. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. Biometrics. 1988;44:837-45.
21. Mandrekar JN. Receiver operating characteristic curve in diagnostic test assessment. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer. 2010;5:1315-6.
22. Zhang Z, Coyle JL, Sejdic E. Automatic hyoid bone detection in fluoroscopic images using deep learning. Sci Rep. 2018;8:12310.

Figures

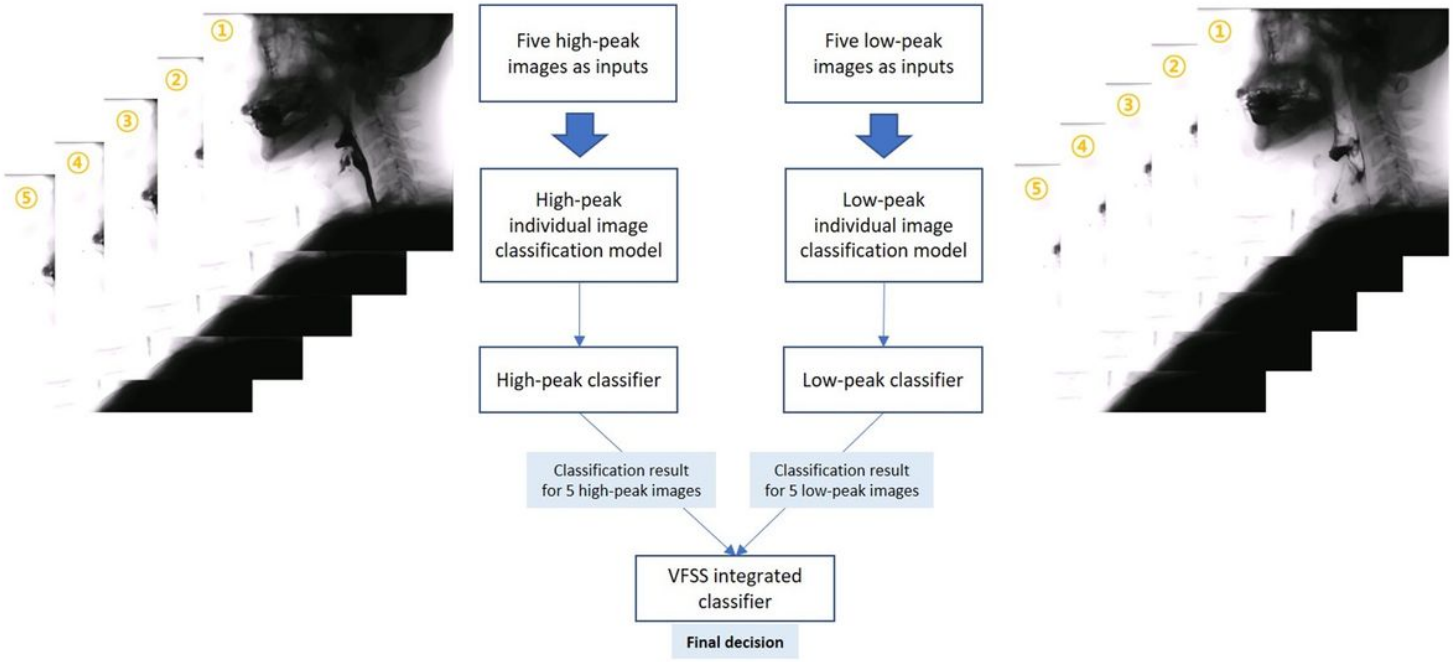

Figure 1

The overall modeling process of the study.

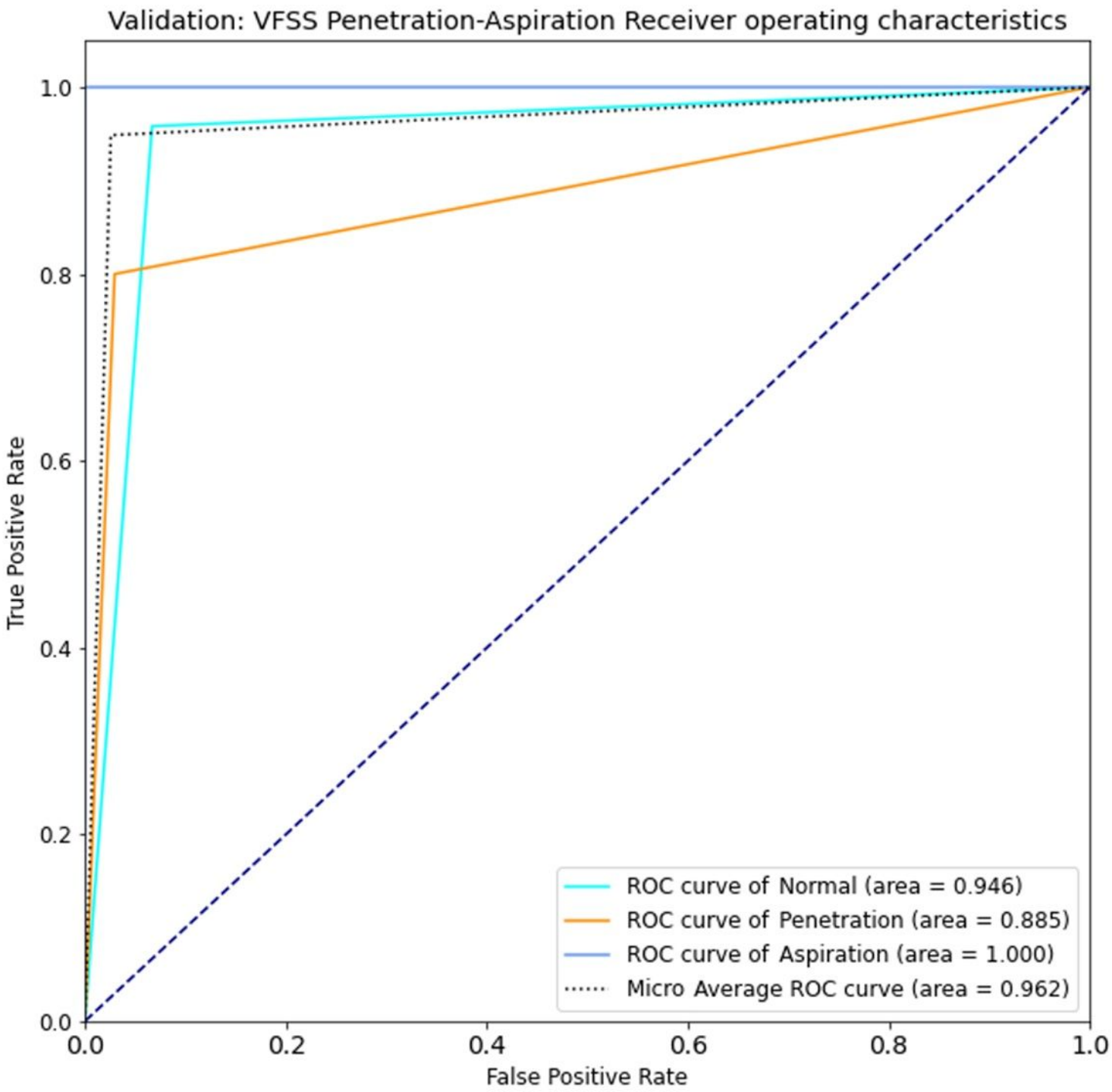

Figure 2

Receiver operating characteristic curve for the data validation models. The AUC of the validation dataset of the videofluoroscopic swallowing study image for the convolutional neural network model was 0.946 for normal findings, 0.885 for penetration, and 1.000 for aspiration. The average AUC was 0.962. AUC: area under the curve