



Opinion

Artificial Intelligence in the Interpretation of Videofluoroscopic Swallow Studies: Implications and Advances for Speech–Language Pathologists

Anna M. Girardi ^{1,2,*} , Elizabeth A. Cardell ^{1,2,3} and Stephen P. Bird ^{1,2}

¹ School of Health and Medical Sciences, University of Southern Queensland, Ipswich, QLD 4305, Australia; elizabeth.cardell@unisq.edu.au (E.A.C.); stephen.bird@unisq.edu.au (S.P.B.)

² Centre for Health Research, University of Southern Queensland, Toowoomba, QLD 4350, Australia

³ School of Medicine and Dentistry, Griffith University, Gold Coast, QLD 4222, Australia

* Correspondence: anna.girardi@unisq.edu.au

Abstract: Radiological imaging is an essential component of a swallowing assessment. Artificial intelligence (AI), especially deep learning (DL) models, has enhanced the efficiency and efficacy through which imaging is interpreted, and subsequently, it has important implications for swallow diagnostics and intervention planning. However, the application of AI for the interpretation of videofluoroscopic swallow studies (VFSS) is still emerging. This review showcases the recent literature on the use of AI to interpret VFSS and highlights clinical implications for speech–language pathologists (SLPs). With a surge in AI research, there have been advances in dysphagia assessments. Several studies have demonstrated the successful implementation of DL algorithms to analyze VFSS. Notably, convolutional neural networks (CNNs), which involve training a multi-layered model to recognize specific image or video components, have been used to detect pertinent aspects of the swallowing process with high levels of precision. DL algorithms have the potential to streamline VFSS interpretation, improve efficiency and accuracy, and enable the precise interpretation of an instrumental dysphagia evaluation, which is especially advantageous when access to skilled clinicians is not ubiquitous. By enhancing the precision, speed, and depth of VFSS interpretation, SLPs can obtain a more comprehensive understanding of swallow physiology and deliver a targeted and timely intervention that is tailored towards the individual. This has practical applications for both clinical practice and dysphagia research. As this research area grows and AI technologies progress, the application of DL in the field of VFSS interpretation is clinically beneficial and has the potential to transform dysphagia assessment and management. With broader validation and inter-disciplinary collaborations, AI-augmented VFSS interpretation will likely transform swallow evaluations and ultimately improve outcomes for individuals with dysphagia. However, despite AI's potential to streamline imaging interpretation, practitioners still need to consider the challenges and limitations of AI implementation, including the need for large training datasets, interpretability and adaptability issues, and the potential for bias.

Keywords: dysphagia; artificial intelligence; videofluoroscopic swallow study; deep learning; machine learning; imaging; speech–language pathology



Citation: Girardi, A.M.; Cardell, E.A.; Bird, S.P. Artificial Intelligence in the Interpretation of Videofluoroscopic Swallow Studies: Implications and Advances for Speech–Language Pathologists. *Big Data Cogn. Comput.* **2023**, *7*, 178. <https://doi.org/10.3390/bdcc7040178>

Academic Editor: Min Chen

Received: 17 October 2023

Revised: 16 November 2023

Accepted: 22 November 2023

Published: 28 November 2023



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

1. Introduction

Dysphagia (impaired swallowing function) is a difficulty in moving a food or liquid bolus from the mouth to the stomach [1] and can emerge from age-related changes, neurological changes (e.g., stroke, neurodegenerative diseases), structural changes or anomalies (e.g., cancer, fistulae), and cognitive decline (e.g., dementia) [2]. Dysphagia can negatively impact quality of life and well-being and result in social isolation, particularly as the severity of the dysphagia increases [3,4]. Further, individuals with dysphagia experience an increased risk of aspiration, pneumonia, choking, malnutrition, and dehydration [5].

Dysphagia symptomology can be difficult to extricate from other respiratory conditions, and therefore, a timely and accurate diagnosis is vital to mitigate adverse effects and ensure prompt treatment planning. Management of dysphagia is typically complex, given its multi-faceted nature and concordance with other medical conditions. While many health professionals can be involved in the assessment and management of dysphagia, in clinical practice, the management of dysphagia is primarily aligned with the scope of practice of speech–language pathologists (SLPs) [6,7].

Following referral for dysphagia assessment, a speech–language pathologist (SLP) will typically conduct a Clinical Swallow Evaluation (CSE). A CSE enables the SLP to gather information on the pre-oral and oral phases of swallowing and form hypotheses about the pharyngeal phase of swallowing [8]. A CSE involves obtaining a thorough case history, observing the person’s oral structures and oral motor movements (including lips, tongue, and jaw), and assessing the person’s ability to manage varying consistencies of food and fluids through controlled trials. A swallow screening may also include the observation of cough occurrence during or following water swallows and changes in voice quality, as a potential marker for laryngeal penetration or aspiration [9]. Following such an assessment, the SLP can make an informed decision on the person’s ability to manage oral intake and decide on intervention strategies if required. It is difficult to detect or exclude aspiration with a CSE or other non-instrumental methods, with prior studies reporting the sensitivity of a CSE in identifying aspiration of 54% to 77% [10–12]. Therefore, an important function of a CSE is to determine the need for an instrumental examination of swallowing [13,14].

Videofluoroscopic swallow studies (VFSS) are the gold standard clinical assessment for dysphagia, using x-ray imaging to examine swallow function [15]. Trained SLPs, along with a multidisciplinary team comprised of radiologists and radiographers, observe the bolus trajectory from the oral cavity to the stomach, using a radiocontrast agent (i.e., barium sulfate). Swallow kinetics and pathophysiological processes are analyzed frame by frame, which is often a time-consuming process, and although standardized, it is susceptible to human error [16]. In response to this, research has utilized computerized image analysis programs to track components of VFSS [17–19]. Using computerized image analysis to track hyoid bone movement, Kellen et al. [20] reported a high correlation with manual analysis of hyoid bone movement during swallowing ($r = 0.97$ and above). However, the clinical usefulness of computerized analyses has been restricted because these models often necessitate manual identification of anatomical landmarks. Recently, deep learning (DL), a component of artificial intelligence (AI), has increased the accuracy and efficiency of VFSS interpretation [21,22]. To date, the available studies have not been described or appraised to determine the clinical applicability for SLPs. To this end, this review showcases the recent literature investigating AI-aided interpretation of VFSS. Given this is an emerging field of research, the current manuscript will provide a brief review and discuss implications for SLPs. Sections 2 and 3 provide an overview of AI and describe the research methodology. Subsequent sections describe and summarize relevant studies, with the consideration of key findings, study limitations, and the clinical applicability for dysphagia management.

2. Background

2.1. Overview of Artificial Intelligence

AI is the simulation of human intelligence by machines and computer systems. With rapid advances in technology, AI has established itself as a transformative force across various fields, including medical imaging and diagnostics. AI, through ML and DL algorithms, has the potential to analyze complex data, identify patterns, and offer diagnostic and prognostic insights that surpass human capability in both speed and accuracy [23,24]. In medical imaging, AI applications have been successful in detecting abnormalities in radiographic images and reducing the manual workload [24]. Hence, AI is becoming increasingly recognized as a powerful tool in the field of radiology. AI advancements have been occurring for many years. The use of AI for imaging emerged in the 1980’s and 1990’s with the concept of computer-aided diagnosis (CAD), and this developed into further diag-

nostic applications in the following decade. Rapid developments have occurred in the last ten years, which have resulted in increased clinical implementation and commercialization. AI implementation in healthcare involves data collection, cleaning, and pre-programming so that relevant variables can be extracted for training and identification by AI models. While AI offers opportunities for improved efficiency and accuracy of assessment interpretation, the use of AI in healthcare presents unique challenges and has important ethical considerations. Practitioners need to conform to industry guidelines and principles to guide AI implementation into practice and ensure patients are protected from data leakage, bias, and interpretation error.

2.2. Overview of Machine Learning and Deep Learning

Under the umbrella of AI, ML enables a program to learn and improve from exposure to data, without being explicitly programmed [25]. Put simply, ML is about prediction, i.e., feeding data into an algorithm to make predictions without a pre-defined rule [26]. The process involves training a model on a dataset, and an algorithm refines its projections until it reaches an acceptable level of accuracy. DL, a subfield of ML, is inspired by the structure and function of the human brain and utilizes neural networks within many layers (deep neural networks) to analyze data. In essence, DL networks have a self-learning ability [27], and algorithms are inspired by the structure and function of the human brain. Artificial neural networks (ANNs) consist of multiple layers, each performing a discrete task and passing its output to the next layer. A recurrent neural network (RNN) is a type of ANN that analyzes sequential or time-series data and has various applications, including speech recognition and image captioning. There are three primary categories of ML architecture, (1) supervised, (2) unsupervised, and (3) reinforcement learning [25]. Supervised learning involves training a model on a labeled dataset, meaning each input is paired with the correct output. An example of a supervised DL architecture is convolutional neural networks (CNNs). CNNs have layers that can abstract data from videos or imaging and another fully connected layer that provides the desired output. Figure 1 contains a depiction of the interconnected framework of the DL approach, with CNNs and RNNs.

Interconnected Framework of Deep Learning, CNNs and RNNs

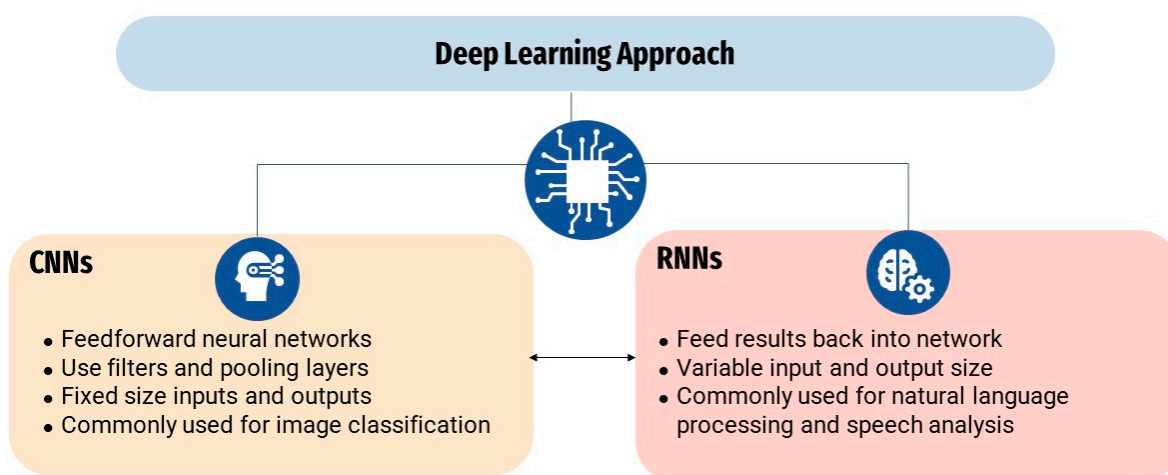


Figure 1. The interconnected framework of deep learning CNNs and RNNs.

2.3. DL Applications for Medical Imaging

One area in which AI has been increasingly utilized is medical imaging. In various forms, AI has been a component of imaging for decades, with the first research regarding the use of AI in imaging occurring in the 1990's. More recently, CNNs have been applied to chest x-rays for detecting diseases like pneumonia, tuberculosis, and lung cancer, often

out-performing human interpreters [28]. Similarly, DL algorithms have been explored in the analysis of aspects of VFSS that are difficult or time-consuming, using AI models.

There is potential for AI to improve dysphagia management by providing diagnostic precision and targeted, individualized treatment planning; however, there remains a paucity of literature on this emerging tool [21]. Further, given the complex nature of AI applications in imaging interpretation, there is a need for studies that focus specifically on clinical implementation of AI strategies. In VFSS interpretation, DL algorithms are trained to identify pertinent aspects of the swallow mechanism. This process involves image selection and segmentation, training of the AI model via image classification, and quantification of outputs. While the research in this area is developing, no available studies have described the current available literature or discussed the clinical implications, recommendations, and limitations for SLPs.

3. Aims and Methodology

This study aimed to investigate how AI is being used to interpret VFSS. Specifically, we explore the elements of VFSS that are currently being interpreted by AI and how these methods can be used to inform and extend clinical SLP practice. A literature review methodology was used to enable synthesis of the literature in the field [29]. This methodology was selected as it can provide a useful summary of evidence for practitioners to guide decision making and work practices in the absence of a substantial research base.

Literature Selection

Following narrative review guidelines, as described by Ferrari [30] for the literature selection process, inclusion and exclusion criteria were defined, and data, keywords used, and number of records retrieved were recorded during each search. Manual searches were conducted until the saturation point was reached. Each article was then critically assessed, and articles with the most pertinent contributions to the current topic were included. A literature search of full-text, peer-reviewed manuscripts, published in English between January 2018 and July 2023 was conducted on 10 August 2023 and 23 August 2023 utilizing PubMed (MEDLINE), Web of Science (WoS), SCOPUS, and Google Scholar. The search terms were “videofluoroscopic swallow study”; “VFSS”; “deep learning”; “artificial intelligence”; and “AI”. The initial search returned 162 articles. Titles and abstracts were reviewed, and 33 were removed. The 129 remaining studies were reviewed in full, and 118 were removed as they did not fit the aim of the research. A total of 11 studies were selected for final inclusion. Figure 2 is a flowchart of the literature selection process.

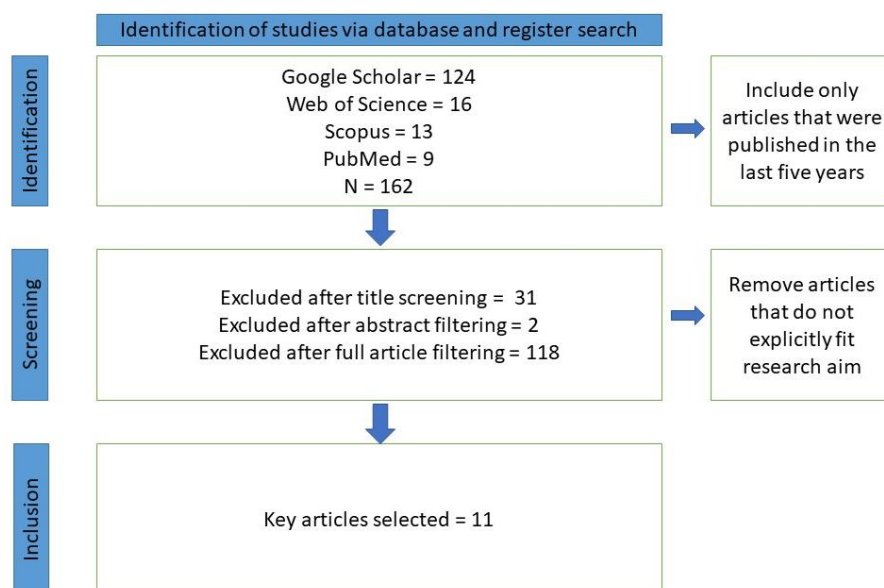


Figure 2. Flowchart depicting the literature selection process.

4. Results

The studies that we identified for inclusion in this review, and the clinical applications of the findings for SLPs, are presented below. Following a thematic reduction, three broad themes were derived from the identified studies, and these are presented in Sections 4.1–4.3.

4.1. Detection of Aspiration

Aspiration refers to the entry of food or fluids into the trachea, and potentially, the lungs. Aspiration can lead to serious health complications, including aspiration pneumonia, which is a leading cause of morbidity and mortality in dysphagia patients [31]. An important function of VFSS is the ability to visualize whether laryngeal penetration or aspiration occurs and examine the contributing physiological factors. Studies examining the use of VFSS for swallowing diagnostics have used DL to identify the presence or absence of aspiration. Kim et al. [21] used a CNN to identify the presence of aspiration in 190 participants with dysphagia with high accuracy. Similarly, Iida et al. [32] used CNNs to detect aspiration in 18,333 images and showed that DL has the potential to detect aspiration with precision. Lee et al. [33] used a DL model to detect airway invasion from VFSS images, without clinician input, with 97.2% accuracy in classifying image frames and 93.2% in classifying video files. Kim et al. [34] used the same DL model and found moderate to substantial inter-rater agreement between the machine and human. However, these studies highlight a pertinent limitation in the use of AI for aspiration detection. While the DL model presented could detect the presence or absence of aspiration, aspiration is the result of a sequence of events and requires skilled interpretation of the concurrent pathophysiological patterns to be clinically useful. Table 1 summarizes key studies detecting laryngeal penetration or aspiration.

Table 1. Summary of key studies detecting laryngeal penetration or aspiration.

| Ref | Sample | Algorithm | Findings |
|------|---|--|---|
| [21] | 190 participants with dysphagia | CNN | The AUC of the validation dataset of the VFSS images for the CNN model was 0.942 for normal findings, 0.878 for penetration, and 1.000 for aspiration |
| [32] | 54 participants with aspiration, 75 participants without aspiration | Three CNNs; Simple-Layer, Multiple-Layer, and Modified LeNet | The AUC values at epoch 50 were 0.973, 0.890, and 0.950, respectively, with statistically significant differences between AUC values |
| [33] | 106 participants with dysphagia | Deep CNN using U-Net | Detected airway invasion with an overall accuracy of 97.2% in classifying image frames and 93.2% in classifying video files |
| [34] | 49 participants with dysphagia | Deep CNN using U-Net | Kappa coefficients indicate moderate to substantial interrater agreement between AI and human raters in identifying laryngeal penetration or aspiration |

Abbreviations: AI = artificial intelligence, AUC = area under the curve, CNN = convolutional neural network, VFSS = videofluoroscopic swallow study.

4.2. Temporal Parameters of Swallowing Function

To evaluate clinical features and determine rehabilitation strategies of dysphagia, it is crucial to measure the exact response time of the pharyngeal swallowing reflex in a VFSS. Swallowing involves a sequence of precisely coordinated physiological events. Any delay in the oral, pharyngeal, or esophageal phases, or premature initiation, can cause an incomplete or inefficient bolus transfer. Bandini et al. [35] examined time points in the pharyngeal phase, namely, the ‘bolus pass mandible’ (BPM; where the leading edge of the bolus touches or crosses the shadow of the ramus of the mandible) and the upper esophageal sphincter closure (UESC; where the upper esophageal sphincter (UES) achieves

closure behind the bolus tail). CNN-based approaches were able to detect these measures with high accuracy, which is congruent with research by Lee et al. [22], who automatically detected the response time for the pharyngeal swallowing reflex with high accuracy. Jeong et al. [36] measured seven temporal parameters with relatively high accuracy; however, they encountered difficulty measuring pharyngeal delay and laryngeal vestibule closure, presumably due to the innate variability of these phases in the swallow mechanism. In addition, their AI model was trained on a small amount of thin liquid (2 mL) ingested, which limits pharyngeal residue, and has limited extension to cases with varying liquid viscosities and volumes. Extension to larger and more diverse cohorts is required to enhance the clinical applicability of these models. Table 2 summarizes key studies analyzing temporal measures of swallowing function.

Table 2. Summary of key studies measuring temporal parameters of swallowing.

| Ref | Sample | Algorithm | Findings |
|------|---|----------------------------------|--|
| [22] | 27 participants with subjective dysphagia | 3D CNN | Average success rate of detection during the pharyngeal phase of 97.5% |
| [35] | 78 healthy participants | Compared multiple CNN algorithms | Pearson's correlation coefficient of 0.951 for BPM and 0.996 for UESC |
| [36] | 547 VFSS video clips from patients with dysphagia | 3D CNN | Average accuracy of 0.864 to 0.981 |

Abbreviations: BPM = bolus pass mandible, CNN = convolutional neural network, UESC = upper esophageal sphincter closure, VFSS = videofluoroscopic swallow study.

4.3. Hyoid Bone Movement

The hyoid bone is in the anterior neck, suspended by ligaments and muscles. Its localization and movement during swallowing are important for various reasons. The swallow reflex is initiated by touch receptors in the pharynx, which results in the forward and upward movement of the hyoid [37]. The upward movement of the hyoid assists in pulling open the UES, allowing the bolus to move from the pharynx into the esophagus. If the hyoid does not move correctly, there can be incomplete UES opening, leading to pharyngeal residue and an increased risk of aspiration [38]. Four reviewed studies investigated hyoid bone detection and tracking through DL frameworks. Hsiao et al. [39] examined 409 videos from 233 patients using fully automated hyoid bone localization and tracking. They found excellent inter-rater reliability of hyoid bone detection between the algorithm and a group of three human annotators. Similarly, Kim et al. [40] utilized automated hyoid bone tracking and designed a network that can detect salient objects in VFSS images. Zhang et al. [41] also presented a model that could automatically detect the hyoid bone; however, inaccuracies in tracking were a limitation in this study. Lee et al. [42] proposed two types of DL networks for tracking the hyoid in VFSS images with high levels of accuracy. However, a limitation of this study was that the user was required to specify the hyoid location in the first instance, to enable ongoing tracking. Another limitation was the erroneous detection of the mandible instead of the hyoid bone. These limitations pose challenges for clinical implementation and applicability. Table 3 summarizes key studies using DL-aided hyoid bone movement detection.

Table 3. Summary of key studies using AI hyoid bone movement detection.

| Ref | Sample | Algorithm | Findings |
|------|---|-------------------------------|--|
| [39] | 44 participants with dysphagia | CNN; Cascaded Pyramid Network | Excellent inter-rater reliability for hyoid bone detection, good-to-excellent inter-rater reliability for displacement and the average velocity of the hyoid bone in horizontal or vertical directions, moderate-to-good reliability in calculating the average velocity in horizontal direction |
| [40] | 207 participants with dysphagia | CNN; U-Net | mAP of 91% for hyoid bone detection |
| [41] | 265 participants with dysphagia | CNN; SSD | mAP of 89.14% for hyoid bone detection |
| [42] | 77 participants; healthy individuals and individuals with Parkinson’s Disease and stroke. | CNN; MDNet | DSC results for the proposed method were 0.87 for healthy individuals, 0.88 for patients with Parkinson’s Disease, 0.85 for patients with stroke, and a total of 0.87. |

Abbreviations: CNN = convolutional neural network, DSC = dice similarity coefficient, mAP = mean average precision, SSD = single-shot detector.

5. Implications for Speech–Language Pathology

AI has been increasingly integrated into various healthcare fields and has the potential to increase clinical efficiency and accuracy and improve patient outcomes. AI also has the potential to capture diverse data, with enhanced precision, for research. This research can then be used to inform targeted interventions. In the field of SLP, VFSS are considered the gold standard for dysphagia assessment. However, VFSS are time-consuming to conduct and can be laborious to interpret, particularly for inexperienced clinicians. Further, VFSS interpretation is prone to human error [31,34]. DL frameworks have the potential to improve the accuracy and speed with which VFSS are interpreted, and thus, they have several clinical applications depending on the method of detection used. Figure 3 presents five key areas of AI-aided VFSS interpretation for SLPs. The use of AI in VFSS interpretation has the potential to improve diagnostic accuracy, enabling clinicians to make informed decisions around swallow treatment. Another major benefit of AI use in VFSS interpretation is the potential for improved workflow and efficiency of VFSS analysis, and this is currently a time-consuming component of the assessment. The ability to provide tailored and real-time feedback to patients is another benefit, given that specific VFSS components could be isolated and reported on with a higher accuracy. This has positive implications for treatment and care planning, given that treatment outcomes can be monitored with greater precision. Lastly, the generation of VFSS data via AI provides a platform for dysphagia research and will enable the investigation of swallow function with greater accuracy and a correlation with clinical outcomes.

The elements of VFSS that have been investigated in key AI studies identified in this review comprise three broad groups: laryngeal penetration and aspiration detection, temporal aspects of swallowing, and hyoid bone detection and localization. Each of these areas has clinical implications for SLPs and individuals with dysphagia.

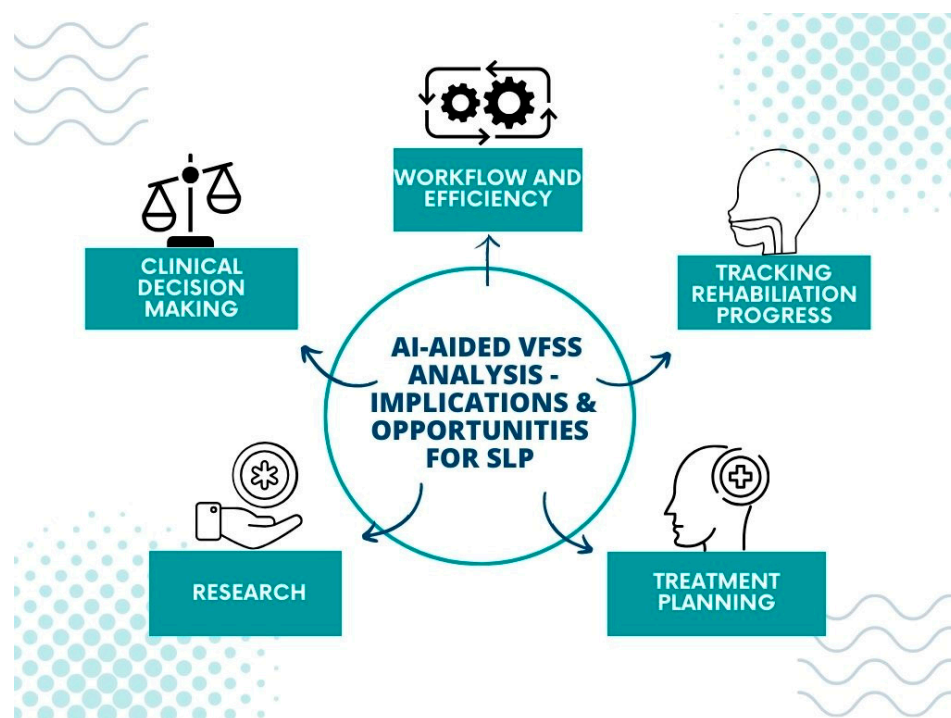


Figure 3. AI-aided VFSS interpretation—implications and opportunities for SLP.

5.1. Clinical Applications of AI Detection of Laryngeal Penetration and Aspiration

Studies have shown that AI can detect the presence of laryngeal penetration or aspiration with accuracy [21,32–34]. The accurate detection of aspiration has the potential to improve patient outcomes. As highlighted by Kim et al. [21], rather than identifying and tracking anatomical structures, aspiration detection is an important function of VFSS, and therefore, it is a clinically applicable AI capability. Further, if persistent and undetected, aspiration can result in adverse health outcomes, and it is a significant cause of mortality and morbidity in vulnerable clinical populations [43]. The accurate and timely detection of aspiration also has the potential to improve informed decision making. When healthcare providers, people with dysphagia, and their families are fully informed about the presence and extent of aspiration, they can make reasoned decisions about their health and healthcare. This may include the modification of their diet or fluids or alternative feeding or hydration methods.

5.2. Clinical Applications of AI Measurement of Temporal Parameters of Swallowing

Swallowing involves rapid, sequential physiologic components. The late oral and early pharyngeal components of swallowing are arguably the most crucial from a safety perspective. When the bolus reaches the oropharynx, the pharyngeal swallow is initiated. While the onset of the pharyngeal swallow is variable relative to the bolus position [44], once initiated, there are a series of laryngeal and pharyngeal events that protect the airway and clear ingested material from the pharynx [45]. While an intricate description of the physiology of swallowing is outside of the scope of this review, the reader will appreciate the alignment of the pharyngeal motion with physiological events, for a successful swallow function. AI frameworks can be clinically useful tools for estimating the absence, or delayed response time, of the swallowing reflex in patients with dysphagia and improving inter-rater reliability of the response time evaluation of the pharyngeal swallowing reflex between expert and unskilled clinicians. The frameworks in these studies can be used to provide considerable clinical information for dysphagia treatment. Clinically, this DL application can also be expanded to other spatiotemporal parameters in VFSS.

5.3. Clinical Applications of Hyoid Bone Movement Detection Using AI

The hyoid bone is a salient anatomical feature commonly monitored during the analysis of VFSS [46]. The anterior–superior movement of the hyoid bone plays a significant role in preventing aspiration and opening the UES to enable the food bolus to move into the esophagus [47]. Thus, evaluating hyoid bone movement during VFSS is an important factor in clinical dysphagia management. Manual tracking of hyoid movement is considered the gold standard for SLPs in dysphagia management; however, it is of course prone to human error and is time-consuming. Zhang et al. [41] and Lee et al. [42] used DL models to track the hyoid bone with precision. During swallowing, the hyoid bone moves upwards and forwards, at times becoming obscured by the shadow of the mandible, thus becoming difficult to detect by human reviewers. Both models could detect the hyoid bone, even when obscured, and therefore appear to be promising as a widely applicable pre-processing step for dysphagia research and, eventually, clinically [41,42].

Collectively, the studies illustrate the efficacy of AI in VFSS interpretation, but they also introduce clinical challenges for SLPs. Figure 4 highlights the challenges for SLPs in applying AI to VFSS interpretation and includes recommendations for application to clinical practice.

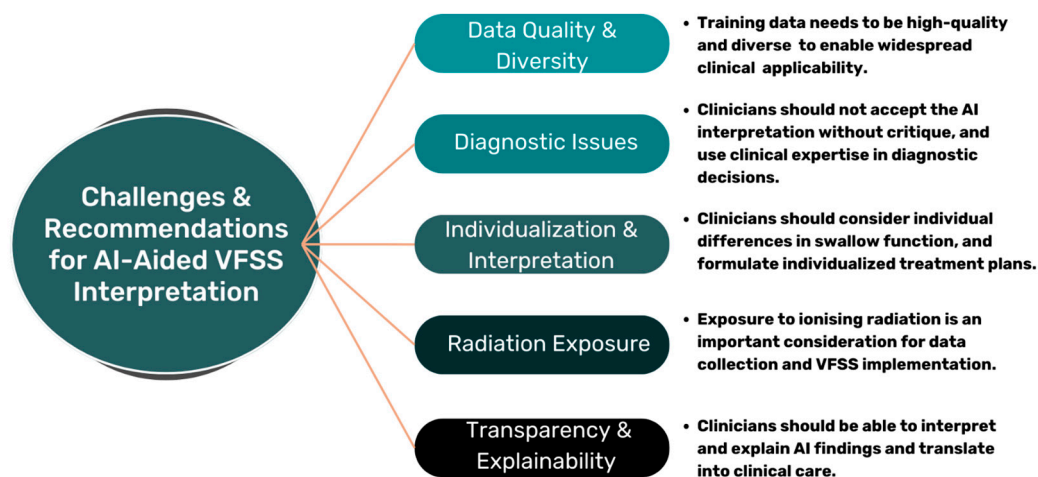


Figure 4. Challenges and recommendations for AI-aided VFSS interpretation.

6. Limitations and Future Directions

While the benefits of using AI in VFSS interpretation are likely to outweigh the risks, there are various limitations to consider in its implementation. First, the accuracy of AI depends largely on the quality and diversity of the training data. Studies require access to large datasets to train ML and DL models and refine algorithms for their intended purpose. Additionally, if certain patient demographics or groups are not represented in the training data, the widespread application may be limited, and bias may influence model development and predictive power. Another limitation is that the nuances and individual differences in swallowing function that a trained clinician may identify may not be elucidated by AI models. Notably, the studies reviewed in this paper utilized heterogenous participant groups, with varying dysphagia presentation and severity both within and between studies. A series of studies with homogeneity around dysphagia etiology would be useful to train AI models to identify the diagnostic variability within specific populations. Additionally, while VFSS is deemed the gold standard for swallowing assessment, patient exposure to ionizing radiation is an inherent risk of this technique that should be considered in any data collection involving VFSS. Further, to enable the collection of quality VFSS data for training DL models, participants will need to be optimally positioned. This may be difficult to achieve in some cohorts and may introduce challenges in data collection and utilization of VFSS datasets.

The wider clinical context will need to remain at the center of AI interpretation for VFSS. A clinical evaluation of swallowing involves the consideration of not only the instrumental evaluation of swallowing but the person's medical history, presentation, and individual needs and goals. AI interpretation will provide another tool that must be applied within the larger context of client-centered care. As the use of AI becomes widespread, there is a risk of clinicians becoming reliant on AI interpretation. When clinicians accept the AI interpretation without critique or the use of clinical expertise, ethical and diagnostic issues may arise [48,49]. A fundamental challenge of AI implementation is the lack of transparency of AI models, thus impacting the clinician's understanding of the model process. Healthcare is a field that relies on transparent decision making, and AI has the potential to remove that aspect. Explainability or transparency of AI models is an important consideration to ensure clinicians can be confident implementing AI findings into clinical care and explaining outcomes to patients.

7. Conclusions

The integration of AI into VFSS interpretation holds promise for enhancing diagnostic accuracy, automating routine components of analysis, and assisting clinicians with what can be a time-consuming task. In areas where access to expert clinicians is limited, AI can enable rapid and accurate assessment of swallowing and facilitate targeted intervention for individuals, regardless of the skill level of the treating SLP. AI, at least in its initial implementation in the field, should be viewed as a complement to human expertise in swallow diagnostics. It is anticipated that with broader clinical validation and interdisciplinary collaborations, AI-augmented VFSS interpretation will become the cornerstone of dysphagia management in the future.

Author Contributions: Conceptualization, A.M.G. and S.P.B.; methodology, A.M.G. and E.A.C.; data curation, A.M.G.; writing—original draft preparation, A.M.G.; writing—review and editing, E.A.C. and S.P.B.; project administration, A.M.G., E.A.C. and S.P.B. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

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

References

1. O'Rourke, F.; Vickers, K.; Upton, C.; Chan, D. Swallowing and Oropharyngeal Dysphagia. *Clin. Med.* **2014**, *14*, 196–199. [CrossRef]
2. Azer, S.A.; Kanugula, A.K.; Kshirsagar, R.K. Dysphagia. In *StatPearls*; StatPearls Publishing: Treasure Island, FL, USA, 2023.
3. Ekberg, O.; Hamdy, S.; Woisard, V.; Wuttge-Hannig, A.; Ortega, P. Social and Psychological Burden of Dysphagia: Its Impact on Diagnosis and Treatment. *Dysphagia* **2002**, *17*, 139–146. [CrossRef] [PubMed]
4. Smith, R.; Bryant, L.; Hemsley, B. The True Cost of Dysphagia on Quality of Life: The Views of Adults with Swallowing Disability. *Int. J. Lang. Commun. Disord.* **2023**, *58*, 451–466. [CrossRef] [PubMed]
5. Altman, K.W.; Yu, G.; Schaefer, S.D. Consequence of Dysphagia in the Hospitalized Patient: Impact on Prognosis and Hospital Resources. *Arch. Otolaryngol. Head Neck Surg.* **2010**, *136*, 784–789. [CrossRef] [PubMed]
6. Speech Pathology Australia Clinical Guidelines. Available online: https://www.speechpathologyaustralia.org.au/SPAweb/Members/Clinical_Guidelines/spaweb/Members/Clinical_Guidelines/Clinical_Guidelines.aspx?hkey=f66634e4-825a-4f1a-910d-644553f59140 (accessed on 28 August 2023).
7. Scope of Practice in Speech-Language Pathology. Available online: <https://www.asha.org/policy/sp2016-00343/> (accessed on 28 August 2023).
8. Barnaby-Mann, G.; Lenius, K. The Bedside Examination in Dysphagia. *Phys. Med. Rehabil. Clin. N. Am.* **2008**, *19*, 747–768. [CrossRef]
9. Garand, K.L.F.; McCullough, G.; Crary, M.; Arvedson, J.C.; Dodrill, P. Assessment across the Life Span: The Clinical Swallow Evaluation. Available online: https://pubs.asha.org/doi/epdf/10.1044/2020_AJSLP-19-00063 (accessed on 23 August 2023).

10. Lynch, Y.T.; Clark, B.J.; Macht, M.; White, S.D.; Taylor, H.; Wimbish, T.; Moss, M. The Accuracy of the Bedside Swallowing Evaluation for Detecting Aspiration in Survivors of Acute Respiratory Failure. *J. Crit. Care* **2017**, *39*, 143–148. [CrossRef]
11. DePippo, K.L.; Holas, M.A.; Reding, M.J. Validation of the 3-oz Water Swallow Test for Aspiration Following Stroke. *Arch. Neurol.* **1992**, *49*, 1259–1261. [CrossRef]
12. McCullough, G.H.; Rosenbek, J.C.; Wertz, R.T.; McCoy, S.; Mann, G.; McCullough, K. Utility of Clinical Swallowing Examination Measures for Detecting Aspiration Post-Stroke. *J. Speech Lang. Hear. Res.* **2005**, *48*, 1280–1293. [CrossRef]
13. Desai, R.V. Build a Case for Instrumental Swallowing Assessments in Long-Term Care. Available online: <https://leader.pubs.asha.org/doi/10.1044/leader.OTP.24032019.38> (accessed on 23 August 2023).
14. Warner, H.; Coutinho, J.M.; Young, N. Utilization of Instrumentation in Swallowing Assessment of Surgical Patients during COVID-19. *Life* **2023**, *13*, 1471. [CrossRef]
15. Costa, M.M.B. Videofluoroscopy: The Gold Standard Exam for Studying Swallowing and Its Dysfunction. *Arq. Gastroenterol.* **2010**, *47*, 327–328. [CrossRef]
16. Kerrison, G.; Miles, A.; Allen, J.; Heron, M. Impact of Quantitative Videofluoroscopic Swallowing Measures on Clinical Interpretation and Recommendations by Speech-Language Pathologists. *Dysphagia* **2023**, *38*, 1528–1536. [CrossRef] [PubMed]
17. Hossain, I.; Roberts-South, A.; Jog, M.; El-Sakka, M.R. Semi-Automatic Assessment of Hyoid Bone Motion in Digital Videofluoroscopic Images. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2014**, *2*, 25–37. [CrossRef]
18. Natarajan, R.; Stavness, I.; Pearson, W. Semi-Automatic Tracking of Hyolaryngeal Coordinates in Videofluoroscopic Swallowing Studies. *Comput. Methods Biomech. Biomed. Eng. Imaging Vis.* **2017**, *5*, 379–389. [CrossRef]
19. Lee, W.H.; Chun, C.; Seo, H.G.; Lee, S.H.; Oh, B.-M. STAMPS: Development and Verification of Swallowing Kinematic Analysis Software. *Biomed. Eng. OnLine* **2017**, *16*, 120. [CrossRef] [PubMed]
20. Kellen, P.M.; Becker, D.L.; Reinhardt, J.M.; Van Daele, D.J. Computer-Assisted Assessment of Hyoid Bone Motion from Videofluoroscopic Swallow Studies. *Dysphagia* **2010**, *25*, 298–306. [CrossRef] [PubMed]
21. Kim, J.K.; Choo, Y.J.; Choi, G.S.; Shin, H.; Chang, M.C.; Chang, M.; Park, D. Deep Learning Analysis to Automatically Detect the Presence of Penetration or Aspiration in Videofluoroscopic Swallowing Study. *J. Korean Med. Sci.* **2022**, *37*, e42. [CrossRef] [PubMed]
22. Lee, J.T.; Park, E.; Hwang, J.-M.; Jung, T.-D.; Park, D. Machine Learning Analysis to Automatically Measure Response Time of Pharyngeal Swallowing Reflex in Videofluoroscopic Swallowing Study. *Sci. Rep.* **2020**, *10*, 14735. [CrossRef] [PubMed]
23. Jiang, F.; Jiang, Y.; Zhi, H.; Dong, Y.; Li, H.; Ma, S.; Wang, Y.; Dong, Q.; Shen, H.; Wang, Y. Artificial Intelligence in Healthcare: Past, Present and Future. *Stroke Vasc. Neurol.* **2017**, *2*, e000101. [CrossRef]
24. Litjens, G.; Kooi, T.; Bejnordi, B.E.; Setio, A.A.A.; Ciompi, F.; Ghafoorian, M.; van der Laak, J.A.W.M.; van Ginneken, B.; Sánchez, C.I. A Survey on Deep Learning in Medical Image Analysis. *Med. Image Anal.* **2017**, *42*, 60–88. [CrossRef]
25. Janiesch, C.; Zschech, P.; Heinrich, K. Machine Learning and Deep Learning. *Electron. Mark.* **2021**, *31*, 685–695. [CrossRef]
26. Jordan, M.I.; Mitchell, T.M. Machine Learning: Trends, Perspectives, and Prospects. *Science* **2015**, *349*, 255–260. [CrossRef] [PubMed]
27. Busnatu, S.; Niculescu, A.G.; Bolocan, A.; Petrescu, G.E.D.; Păduraru, D.N.; Năstasă, I.; Lupuşoru, M.; Geantă, M.; Andronic, O.; Grumezescu, A.M.; et al. Clinical Applications of Artificial Intelligence—An Updated Overview. *J. Clin. Med.* **2022**, *11*, 2265. [CrossRef] [PubMed]
28. Rajpurkar, P.; Zhu, K.; Yang, B.; Mehta, H.; Duan, T.; Ding, D.; Bagul, A.; Langlotz, C.; Shpanskaya, K.; et al. CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning. *arXiv* **2017**, arXiv:1711.05225.
29. Paré, G.; Kitsiou, S. Chapter 9—Methods for Literature Reviews. In *Handbook of eHealth Evaluation: An Evidence-Based Approach [Internet]*; Lau, F., Kuziemsky, C., Eds.; University of Victoria: Victoria, BC, USA, 27 February 2017. Available online: <https://www.ncbi.nlm.nih.gov/books/NBK481583/> (accessed on 1 October 2023).
30. Ferrari, R. Writing narrative style literature reviews. *Med. Writ.* **2015**, *24*, 4. [CrossRef]
31. Langmore, S.E.; Terpenning, M.S.; Schork, A.; Chen, Y.; Murray, J.T.; Lopatin, D.; Loesche, W.J. Predictors of Aspiration Pneumonia: How Important Is Dysphagia? *Dysphagia* **1998**, *13*, 69–81. [CrossRef]
32. Iida, Y.; Näppi, J.; Kitano, T.; Hironaka, T.; Katsumata, A.; Yoshida, H. Detection of Aspiration from Images of a Videofluoroscopic Swallowing Study Adopting Deep Learning. *Oral Radiol.* **2023**, *39*, 553–562. [CrossRef]
33. Lee, S.J.; Ko, J.Y.; Kim, H.I.; Choi, S.I. Automatic Detection of Airway Invasion from Videofluoroscopy via Deep Learning Technology. *Appl. Sci.* **2020**, *10*, 6179. [CrossRef]
34. Kim, Y.; Kim, H.I.; Park, G.S.; Kim, S.Y.; Choi, S.I.; Lee, S.J. Reliability of Machine and Human Examiners for Detection of Laryngeal Penetration or Aspiration in Videofluoroscopic Swallowing Studies. *J. Clin. Med.* **2021**, *10*, 2681. [CrossRef]
35. Bandini, A.; Steele, C.M. The Effect of Time on the Automated Detection of the Pharyngeal Phase in Videofluoroscopic Swallowing Studies. In Proceedings of the 2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), Mexico City, Mexico, 1–5 November 2021; pp. 3435–3438. [CrossRef]
36. Jeong, S.Y.; Kim, J.M.; Park, J.E.; Baek, S.J.; Yang, S.N. Application of Deep Learning Technology for Temporal Analysis of Videofluoroscopic Swallowing Studies. *Res. Sq.* **2022**. [CrossRef]
37. Matsuo, K.; Palmer, J.B. Anatomy and Physiology of Feeding and Swallowing—Normal and Abnormal. *Phys. Med. Rehabil. Clin. N. Am.* **2008**, *19*, 691–707. [CrossRef]

38. Cook, I.J.; Dodds, W.J.; Dantas, R.O.; Massey, B.; Kern, M.K.; Lang, I.M.; Brasseur, J.G.; Hogan, W.J. Opening Mechanisms of the Human Upper Esophageal Sphincter. *Am. J. Physiol.* **1989**, *257*, G748–G759. [[CrossRef](#)] [[PubMed](#)]
39. Hsiao, M.Y.; Weng, C.H.; Wang, Y.C.; Cheng, S.H.; Wei, K.C.; Tung, P.Y.; Chen, J.Y.; Yeh, C.Y.; Wang, T.G. Deep Learning for Automatic Hyoid Tracking in Videofluoroscopic Swallow Studies. *Dysphagia* **2023**, *38*, 171–180. [[CrossRef](#)] [[PubMed](#)]
40. Kim, H.I.; Kim, Y.; Kim, B.; Shin, D.Y.; Lee, S.J.; Choi, S.I. Hyoid Bone Tracking in a Videofluoroscopic Swallowing Study Using a Deep-Learning-Based Segmentation Network. *Diagnostics* **2021**, *11*, 1147. [[CrossRef](#)] [[PubMed](#)]
41. Zhang, Z.; Coyle, J.L.; Sejdic, E. Automatic Hyoid Bone Detection in Fluoroscopic Images Using Deep Learning. *Sci. Rep.* **2018**, *8*, 12310. [[CrossRef](#)]
42. Lee, D.; Lee, W.H.; Seo, H.G.; Oh, B.-M.; Lee, J.C.; Kim, H.C. Online Learning for the Hyoid Bone Tracking during Swallowing with Neck Movement Adjustment Using Semantic Segmentation. *IEEE Access* **2020**, *8*, 157451–157461. [[CrossRef](#)]
43. Shin, D.; Lebovic, G.; Lin, R.J. In-Hospital Mortality for Aspiration Pneumonia in a Tertiary Teaching Hospital: A Retrospective Cohort Review from 2008 to 2018. *J. Otolaryngol. Head Neck Surg.* **2023**, *52*, 23. [[CrossRef](#)]
44. Martin-Harris, B.; Brodsky, M.B.; Michel, Y.; Lee, F.-S.; Walters, B. Delayed Initiation of the Pharyngeal Swallow: Normal Variability in Adult Swallows. *J. Speech Lang. Hear. Res.* **2007**, *50*, 585–594. [[CrossRef](#)]
45. Martin-Harris, B.; Jones, B. The Videofluorographic Swallowing Study. *Phys. Med. Rehabil. Clin. N. Am.* **2008**, *19*, 769–785. [[CrossRef](#)]
46. Donohue, C.; Mao, S.; Sejdic, E.; Coyle, J.L. Tracking Hyoid Bone Displacement during Swallowing without Videofluoroscopy Using Machine Learning of Vibratory Signals. *Dysphagia* **2021**, *36*, 259–269. [[CrossRef](#)]
47. Wei, K.C.; Hsiao, M.Y.; Wang, T.G. The Kinematic Features of Hyoid Bone Movement during Swallowing in Different Disease Populations: A Narrative Review. *J. Formos. Med. Assoc.* **2022**, *121*, 1892–1899. [[CrossRef](#)]
48. Goisau, M.; Abadía, M. Ethics of AI in Radiology: A Review of Ethical and Societal Implications. *Front. Big Data* **2022**, *5*, 850383. [[CrossRef](#)] [[PubMed](#)]
49. Naik, N.; Hameed, B.M.Z.; Shetty, D.K.; Swain, D.; Shah, M.; Paul, R.; Aggarwal, K.; Ibrahim, S.; Patil, V.; Smriti, K.; et al. Legal and Ethical Consideration in Artificial Intelligence in Healthcare: Who Takes Responsibility? *Front. Surg.* **2022**, *9*, 266. [[CrossRef](#)] [[PubMed](#)]

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