iMRI 2025;29(2):96-99
https://doi.org/10.13104/imri.2025.0005
eISSN 2384-1109

Original Article

Received: February 5, 2025
Revised: March 5, 2025
Accepted: April 7, 2025

Correspondence
Hee Woo Cho, MD
Department of Radiology, Research Institute of Radiological Science, and Center for Clinical Imaging Data Science (CCIDS), Yongin Severance Hospital, Yonsei University College of Medicine, 363 Dongbaekjukjeon-daero, Giheung-gu, Yongin 16995, Korea.
E-mail: hughj@yuhs.ac

Revolutionizing Radiology Reporting With Artificial Intelligence-Based Voice Recognition: A Pilot Study on Lumbar Spine Magnetic Resonance Imaging

Minwook Lee[1], Hee Woo Cho[1], and Young Han Lee[2]

[1]Department of Radiology, Research Institute of Radiological Science, and Center for Clinical Imaging Data Science (CCIDS), Yongin Severance Hospital, Yonsei University College of Medicine, Yongin, Korea
[2]Department of Radiology, Research Institute of Radiological Science, and Center for Clinical Imaging Data Science (CCIDS), Yonsei University College of Medicine, Seoul, Korea

Purpose: This study evaluated whether the interpretation speed of routine lumbar spine magnetic resonance imaging (MRI) increases when using an artificial intelligence (AI)-based voice recognition system compared with the conventional keyboard typing method.
Materials and Methods: We retrospectively reviewed 527 routine lumbar spine MRI images performed between November 2022 and February 2023. Two radiologists interpreted 292 and 235 images using conventional keyboard typing and dictation with an AI-based voice recognition system, respectively. Interpretation time, report character count, and turnaround time for the two methods were compared.
Results: Interpretation time was significantly reduced by 21.7% using dictation with the AI-based voice recognition method compared with that using the conventional keyboard typing method ($p < 0.05$). However, no statistically significant differences were observed in the reported character count or turnaround time ($p > 0.05$).
Conclusion: AI-based voice recognition system for interpreting lumbar spine MRI significantly reduced interpretation time compared with the conventional keyboard typing method, suggesting enhanced efficiency for radiologists.

Keywords: Interpretation time; Voice recognition; Artificial intelligence; Spine; Magnetic resonance imaging

INTRODUCTION

The integration of artificial intelligence (AI) into radiology has resulted in significant advancements in image interpretation and workflow efficiency. Although AI has primarily been utilized for image analysis, its application in voice recognition systems has emerged as a transformative tool for enhancing report generation efficiency [1]. Conventional reporting methods often involve manual keyboard typing or transcription by hired typists, both of which are time consuming and error-prone [2-4]. Recent advancements in AI-based voice recognition systems have enabled accurate real-time transcription of dictated reports, thereby addressing these challenges [5].

Earlier studies on voice recognition systems often focused on non-AI-based technologies, which exhibited high error rates, requiring radiologists to spend additional time correcting transcription errors [2,3,6]. Furthermore, these systems are limited by their inability to adapt to the unique speech patterns of individual users [7]. To date, no study has specifically evaluated the efficiency of AI-based voice recognition systems in interpreting lumbar spine magnetic resonance imaging (MRI), which is one of the most commonly performed imaging studies in radiology. Therefore, this study aimed to fill this gap by examining whether such systems can improve interpretation speed and workflow efficiency in routine radiology practice.

MATERIALS AND METHODS

Case Selection and Process

A total of 540 routine lumbar spine MRI were performed at the Yongin Severance Hospital over a 4-month period between November 2022 to February 2023. Cases were excluded if the interpretation time was prolonged owing to external factors, such as interruptions or personal circumstances. After excluding 13 cases, 527 examinations (201 males and 326 females with a mean age of 62.6 years [range, 20–93 years]) were included.

Two fellowship-trained musculoskeletal radiologists alternated between keyboard typing and dictation using an AI-based voice recognition system (VUNO Med-DeepASR). The radiologists had approximately 2 years of experience using AI-based voice recognition systems. Within 2 years of deep learning and frequent updates, the voice recognition system was highly proficient in accurately recognizing the radiologists' speech and converting it into text. One radiologist read the MRI images performed at our hospital for 1 month, and the other radiologist read the images for the next month using only the keyboard typing method. At the end of this period, the MRI images were read using the dictation method with an AI-based voice recognition system for 1 month each. Interpretation time was defined as the duration from initiation of image opening to completion of the interpretation. Report character count was defined as the total number of characters in the interpretation report. Turnaround time was defined as the period from uploading of the MRI images to completion of the final report. This retrospective study was approved by the Institutional Review Board of Yonsei University College of Medicine, Yongin Severance Hospital (no. 2023-0101). Informed patient consent was not required due to the retrospective nature of the study.

Speech Recognition Technology

The voice recognition technology employed in this study combines the Time-Delay Neural Network (TDNN) and Hidden Markov Model (HMM) techniques. The HMM technique was first proposed in 1989 [8], and the hybrid TDNN-HMM model was later discussed and expanded upon in research directions by Bourlard and Morgan [9] in 1998. This model has since been widely applied across various fields owing to its adaptability and robustness [10]. The TDNN-HMM framework utilizes the Lattice-Free Maximum Mutual Information (LF-MMI) criterion to optimize recognition accuracy [11]. This approach enables the system to effectively extract features from speech signals and convert them into phoneme representations with high precision [12]. Additionally, the system achieves greater contextual understanding and transcription accuracy by incorporating n-grams into language modeling [13].

Statistical Analysis

Data were presented as mean ± standard deviation for continuous variables. Independent two-sample t-tests were performed to compare the differences between the two methods, with statistical significance set at $p < 0.05$.

RESULTS

Reader 1 (Lee) interpreted 159 and 118 MRI images using the conventional typing and dictation methods, respectively. Reader 2 (Cho) interpreted 133 and 117 MRI images using the conventional typing and dictation methods, respectively. The results are summarized in Table 1.

Average interpretation time for Reader 1 was 248.5 seconds using the typing method and 176.2 seconds using the dictation method, while for Reader 2, it was 436.1 seconds using the typing method and 347.9 seconds using the dictation method. Both readers had shorter interpretation times with the dictation method, and this difference was statistically significant ($p < 0.001$). Both readers showed a 21.7% reduction in interpretation time with the dictation method (average, 261.7 seconds) compared with that when using the typing method (average, 333.9 seconds), which was statistically significant ($p < 0.001$).

Average character count of the interpretation reports for Reader 1 was 471.9 using the typing method and 478.6 using the dictation method. For Reader 2, the average character count was 540.7 for the typing method and 563.2 for the dictation method. Difference in the average character count of the interpretation reports between the two methods was not statistically significant for both readers ($p > 0.05$).

Although the turnaround time was shorter with the dicta-

Table 1. Comparison between using conventional typing and dictation method with AI-based voice recognition system

	Typing (95% CI)	Dictation (95% CI)	p
Reader 1-Lee			
No. of exams	159	118	
Avg. characters per report	471.9 ± 186.4	478.6 ± 196.5	0.774
Avg. interpretation time (s)	248.5 ± 152.4	176.2 ± 86.9	<0.001
Avg. turnaround time (h)	14.1 ± 12.4	11.4 ± 18.3	0.156
Reader 2-Cho			
No. of exams	133	117	
Avg. characters per report	540.7 ± 225.1	563.2 ± 234.3	0.439
Avg. interpretation time (s)	436.1 ± 244.5	347.9 ± 293.4	0.010
Avg. turnaround time (h)	37.5 ± 26.7	31.6 ± 26.3	0.078
Readers all			
No. of exams	292	235	
Avg. characters per report	503.2 ± 207.4	520.7 ± 219.8	0.350
Avg. interpretation time (s)	333.9 ± 220.2	261.7 ± 232.0	<0.001
Avg. turnaround time (h)	24.7 ± 23.3	21.5 ± 24.7	0.119

Data are presented as number only or mean ± standard deviation.
AI, artificial intelligence; CI, confidence interval.

tion method for both readers, the difference was not statistically significant ($p > 0.05$).

DISCUSSION

This study highlights the significant efficiency gains achieved using an AI-based voice recognition system for interpreting lumbar spine MRI. Unlike previous studies utilizing conventional voice recognition systems, this AI-based voice recognition system demonstrated a reduced interpretation time owing to its advanced accuracy developed through continuous updates and adaptation to individual speech patterns [14]. Conventional voice recognition systems often exhibit high error rates and require additional time for error correction, which negates the intended time-saving benefits [15].

In contrast, AI-based systems leverage robust acoustic modeling techniques, such as the TDNN-HMM and LF-MMI techniques, ensuring greater transcription accuracy and reliability. These improvements align with a prior study showing that AI-based systems outperform traditional models in various speech recognition tasks, particularly in environments with complex acoustic conditions [16]. Advanced models, such as those integrating recurrent neural networks, are promising in improving acoustic modeling [17].

Although the turnaround time improvements were not statistically significant in this study, the observed reduction in typing fatigue emphasized the potential benefits of voice recognition systems in clinical practice. Similar observations have been reported that voice recognition systems reduce turnaround time compared with the conventional typing method [18,19].

There were results of similar character counts in the interpretation reports, without a significant difference between the two methods used. This suggests that the level of details of the interpretation reports when using the dictation method does not appear to be compromised compared with that when using the typing method, as no significant differences were found in the description quality.

This study has a few limitations. First, there may have been a process of checking for typographical errors after using the speech recognition system; however, this time was not included in the total reading time. However, as deep learning was used to reduce recognition errors, we believe that it did not have a significant impact on the results. Second, this pilot study focused only on routine lumbar spine MRI. This choice was made because lumbar spine MRI is one of the most common MRI examinations performed at our institution, making it easier to select participants. Furthermore, the format and content of interpretation reports for routine lumbar spine MRI are relatively standardized, making them a suitable choice for implementing the dictation method. Nonetheless, future studies that include more imaging modalities should be conducted to determine the generalizability of our findings. Finally, an interesting experience during this study was that while using the AI-based speech recognition system, readers tended to choose words that the AI could understand better, rather than trying to teach the AI to recognize new words with their unfamiliar pronunciation. This shift may affect the reproducibility of the study in different settings where controlling typographical or transcription errors might be more challenging. However, it is important to note that prior to this study, the researchers had operated an AI-based voice recognition system for approximately 2 years, which resulted in a level of speech recognition accuracy with almost no errors.

In conclusion, the application of an AI-based voice recognition system significantly reduced the interpretation time for lumbar spine MRI, indicating its potential in enhancing workflow efficiency in radiology. By reducing the cognitive burden of manual typing and minimizing error correction, these systems are promising tools for optimizing reporting workflows, particularly in high-demand clinical environments. Future studies should evaluate its broader application across various imaging modalities and long-term impact on the efficiency of radiology practice.

Availability of Data and Material

The datasets generated or analyzed during the study are not publicly available due to institution data protection but are available from the corresponding author on reasonable request.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: Hee Woo Cho. Data curation: Minwook Lee, Hee Woo Cho. Formal analysis: Minwook Lee. Investigation: Minwook Lee, Hee Woo Cho. Methodology: all authors. Project administration: Hee Woo Cho. Validation: Minwook Lee, Hee Woo Cho. Writing—original draft: Minwook Lee, Hee Woo Cho. Writing—review & editing: all authors.

ORCID iDs

Minwook Lee	https://orcid.org/0000-0003-2822-0489
Hee Woo Cho	https://orcid.org/0000-0002-5079-6954
Young Han Lee	https://orcid.org/0000-0002-5602-391X

Funding Statement

This study was supported by a grant for the Korean Society of MSK MRI by the Korean Society of Magnetic Resonance in Medicine.

Acknowledgments

None

REFERENCES

1. Luo Y, Chen Z, Hershey JR, Le Roux J, Mesgarani N. DEEP clustering and conventional networks for music separation: stronger together. Proc IEEE Int Conf Acoust Speech Signal Process 2017;2017:61-65.
2. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. J Am Med Inform Assoc 2016;23:e169-e179.
3. Hammana I, Lepanto L, Poder T, Bellemare C, Ly MS. Speech recognition in the radiology department: a systematic review. Health Inf Manag 2015;44:4-10.
4. Chen Z, Luo Y, Mesgarani N. Deep attractor network for single-microphone speaker separation. Proc IEEE Int Conf Acoust Speech Signal Process 2017;2017:246-250.
5. Pandey A, Wang D. On cross-corpus generalization of deep learning based speech enhancement. IEEE/ACM Trans Audio Speech Lang Process 2020;28:2489-2499.
6. Tan K, Wang D. Towards model compression for deep learning based speech enhancement. IEEE/ACM Trans Audio Speech Lang Process 2021;29:1785-1794.
7. Wang ZQ, Wang P, Wang D. Multi-microphone complex spectral mapping for utterance-wise and continuous speech separation. IEEE/ACM Trans Audio Speech Lang Process 2021;29:2001-2014.
8. Rabiner LR. A tutorial on hidden Markov models and selected applications in speech recognition. Proc IEEE 1989;77:257-286.
9. Bourlard H, Morgan N. Hybrid HMM/ANN systems for speech recognition: overview and new research directions. In: Giles CL, Gori M, eds. Adaptive processing of sequences and data structures (vol. 1387). 1st ed. Berlin, Heidelberg: Springer, 1998. pp.389-417.
10. Graves A, Mohamed AR, Hinton G. Speech recognition with deep recurrent neural networks. Accessed on Jan 22, 2025. Available at: https://doi.org/10.1109/ICASSP.2013.6638947.
11. Hinton G, Deng L, Yu D, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. IEEE Signal Process Mag 2012;29:82-97.
12. Xiong W, Wu L, Alleva F, Droppo J, Huang X, Stolcke A. The Microsoft 2017 conversational speech recognition system. Accessed on Jan 22, 2025. Available at: https://doi.org/10.1109/ICASSP.2018.8461870.
13. Dahl GE, Yu D, Deng L, Acero A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. IEEE Trans Audio Speech Lang Process 2012;20:30-42.
14. Maas AL, Hannun AY, Ng AY. Rectifier nonlinearities improve neural network acoustic models. Proc 30th Int Conf Mach Learn 2013;28. Accessed on Jan 22, 2025. Available at: https://ai.stanford.edu/~amaas/papers/relu_hybrid_icml2013_final.pdf.
15. Mohamed AR, Dahl GE, Hinton G. Acoustic modeling using deep belief networks. IEEE Trans Audio Speech Lang Process 2012;20:14-22.
16. Veselý K, Hannemann M, Burget L. Semi-supervised training of deep neural networks. Accessed on Jan 22, 2025. Available at: http://doi.org/10.1109/ASRU.2013.6707741.
17. Seide F, Li G, Chen X, Yu D. Feature engineering in context-dependent deep neural networks for conversational speech transcription. Accessed on Jan 22, 2025. Available at: http://doi.org/10.1109/ASRU.2011.6163899.
18. Kauppinen T, Koivikko MP, Ahovuo J. Improvement of report workflow and productivity using speech recognition--a follow-up study. J Digit Imaging 2008;21:378-382.
19. Krishnaraj A, Lee JK, Laws SA, Crawford TJ. Voice recognition software: effect on radiology report turnaround time at an academic medical center. AJR Am J Roentgenol 2010;195:194-197.