JKMS

Original Article
Medical Informatics

Check for updates

Address for Correspondence:
Jinwook Choi, MD, PhD
Department of Biomedical Engineering, Seoul
National University College of Medicine, 103
Daehak-ro, Jongno-gu, Seoul 03080, Korea.
Email: jinchoi@snu.ac.kr

ORCID iDs
Seung-Hwa Lee
https://orcid.org/0000-0001-5508-7519
Jungchan Park
https://orcid.org/0000-0002-7794-3547
Kwangmo Yang
https://orcid.org/0000-0002-7176-4935
Jeongwon Min
https://orcid.org/0000-0001-8412-5545
Jinwook Choi
https://orcid.org/0000-0002-9424-9944

Accuracy of Cloud-Based Speech Recognition Open Application Programming Interface for Medical Terms of Korean

Seung-Hwa Lee [iD],[1,2] Jungchan Park [iD],[3,4] Kwangmo Yang [iD],[5] Jeongwon Min [iD],[6,7] and Jinwook Choi [iD] [2,8]

[1]Rehabilitation and Prevention Center, Heart Vascular Stroke Institute, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
[2]Department of Biomedical Engineering, Seoul National University College of Medicine, Seoul, Korea
[3]Department of Anesthesiology and Pain Medicine, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
[4]Department of Biomedical Sciences, Ajou University Graduate School of Medicine, Suwon, Korea
[5]Center for Health Promotion, Samsung Medical Center, Sungkyunkwan University School of Medicine, Seoul, Korea
[6]Interdisciplinary Program in Bioengineering, Graduate School, Seoul National University, Seoul, Korea
[7]Integrated Major in Innovative Medical Science, Seoul National University Graduate School, Seoul, Korea
[8]Institute of Medical and Biological Engineering, Medical Research Center, Seoul National University, Seoul, Korea

ABSTRACT

Background: There are limited data on the accuracy of cloud-based speech recognition (SR) open application programming interfaces (APIs) for medical terminology. This study aimed to evaluate the medical term recognition accuracy of current available cloud-based SR open APIs in Korean.

Methods: We analyzed the SR accuracy of currently available cloud-based SR open APIs using real doctor–patient conversation recordings collected from an outpatient clinic at a large tertiary medical center in Korea. For each original and SR transcription, we analyzed the accuracy rate of each cloud-based SR open API (i.e., the number of medical terms in the SR transcription per number of medical terms in the original transcription).

Results: A total of 112 doctor–patient conversation recordings were converted with three cloud-based SR open APIs (Naver Clova SR from Naver Corporation; Google Speech-to-Text from Alphabet Inc.; and Amazon Transcribe from Amazon), and each transcription was compared. Naver Clova SR (75.1%) showed the highest accuracy with the recognition of medical terms compared to the other open APIs (Google Speech-to-Text, 50.9%, $P < 0.001$; Amazon Transcribe, 57.9%, $P < 0.001$), and Amazon Transcribe demonstrated higher recognition accuracy compared to Google Speech-to-Text ($P < 0.001$). In the sub-analysis, Naver Clova SR showed the highest accuracy in all areas according to word classes, but the accuracy of words longer than five characters showed no statistical differences (Naver Clova SR, 52.6%; Google Speech-to-Text, 56.3%; Amazon Transcribe, 36.6%).

Conclusion: Among three current cloud-based SR open APIs, Naver Clova SR which manufactured by Korean company showed highest accuracy of medical terms in Korean, compared to Google Speech-to-Text and Amazon Transcribe. Although limitations are existing in the recognition of medical terminology, there is a lot of rooms for improvement of this promising technology by combining strengths of each SR engines.

Generated by xmlinkpress

Visualization: Lee SH. Writing - original draft: Lee SH, Park J. Writing - review & editing: Lee SH, Min J, Choi J.

INTRODUCTION

Speech recognition (SR) systems for healthcare services have been commercially available since the 1980s.[1] SR has been a promising technology for clinical documentation, considered the most time-consuming and costly aspect of using electronic health record (EHR) systems.[2] With the recent rapid development of artificial intelligence (AI) and the use of cloud computing technology, the performance of SR systems has greatly improved.[3,4] Currently, SR systems are widely used and studied in hospitals and several clinical settings, such as emergency departments, pathology departments, and radiology departments, in the United States.[1,5]

Cloud-based SR open application programming interfaces (APIs) can save a lot of time, effort, and money when building a voice recognition application system and are being applied to various fields, such as movie subscription and real-time translation.[6] Despite these advances, only a small number of APIs for healthcare services are currently available.

Sufficient and comprehensive collection of a patient's medical history is crucial in clinical medicine.[7] Longer consultation times improve health outcomes and reduce the number of drug prescriptions.[8] The clinician–patient contact time is on the decline due to the increased burden of deskwork after EHR implementation.[9] If a cloud-based SR open API recognizes medical terms correctly, it can be applied to clinician–patient consultations with high-quality accuracy, and might help to reduce the burden of deskwork and increase in-person clinical face-to-face time. Therefore, we sought to evaluate the accuracy of cloud-based SR open APIs in discerning medical terminology presented in Korean, a non-Latin-based language, using records and transcriptions of real doctor–patient conversations, and we compared the accuracy between several cloud-based SR open APIs. Our findings might help to evaluate the possibility of direct adaptation of current cloud-based SR open APIs to medical documentation.

METHODS

Clinical setting and data collection

Patients who visited the outpatient cardiology clinic of Samsung Medical Center were eligible for participation. Eligible patients were those who 1) were older than 20 years of age; 2) were on their first visit to the clinic; and 3) agreed to be recorded. Those who met the following exclusion criteria did not participate: 1) younger than 20 years of age; 2) could not speak or hear; 3) had a cognitive disorder, such as Alzheimer's disease; or 4) who refused to be recorded. From April 2021 to July 2021, a total of 112 patients were enrolled.

Recordings were performed with a PCM-A10 recorder (Sony, Tokyo, Japan) in the outpatient clinic office. In order to prevent the recording of private information, all recordings were started after the patient gave their name, patient number, and date of birth and after confirming informed consent. The recording mode was linear pulse code modulated audio with 96 kHz/24 bit, which is a method for digitally encoding uncompressed audio information, and audio was saved as .wav files.
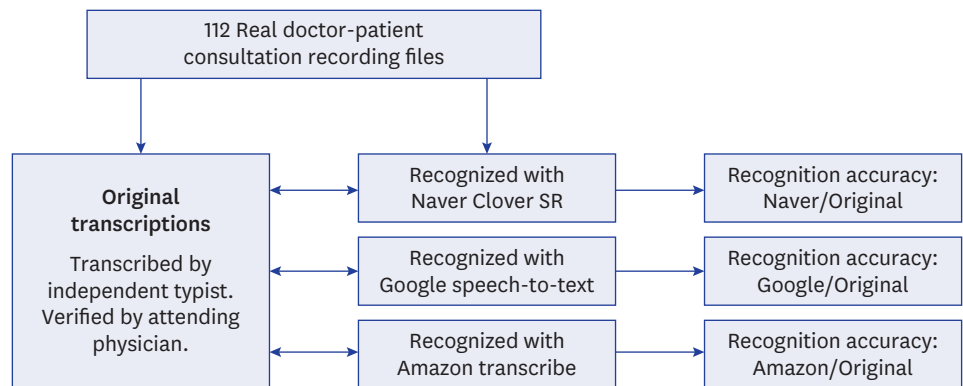
Fig. 1. Study flowchart.
SR = speech recognition.

Transcriptions and SR

The study flowchart is described in **Fig. 1**. Gold standard is the original transcriptions of each recording and were created by an independent nursing student and physician (JP) and validated by an attending outpatient clinic physician (SHL). We selected three cloud-based SR open APIs, Naver Clova SR (Naver, Seongnam, Korea), Google Speech-to-Text (Alphabet Inc., Mountain View, CA, USA), and Amazon Transcribe (Amazon, Seattle, WA, USA). The first API is from a domestic company, and the other two are from international companies located in the United States. All recording files were uploaded to each cloud and underwent SR via their API (Naver Clova SR and Amazon Transcribe) or Python 3 (Google Speech-to-Text). Recognized transcriptions were saved as.txt files, and the uploaded audio files were removed from the cloud immediately after the SR process.

Extraction, annotation definition, and process

We extracted any medical terms which were nouns from each original transcription. Each extracted medical term was defined as one of seven classes (**Table 1**). The total number of words and the frequency of their occurrence, the length of the word, and whether the word was Korean or from a foreign language were also evaluated. According to the definition, medical terms in SR transcription were also extracted, and typos of medical terms in each SR transcription were also evaluated. The typos were defined by three classes: 1) omission (deletion of the word); 2) spelling mistake (incorrect spelling but one could still understand the original meaning of the word); and 3) wrong word: (different spelling which results in a completely different meaning).[10] An annotation of the typos was performed by two physicians and cross-checked. For each original and SR transcription, we analyzed the accuracy rate of each cloud-based SR open API (i.e., the number of medical terms in the SR transcription per number of medical terms in the original transcription). Additionally, we analyzed the recognition accuracy according to word class, length, and non-Korean origin.

Table 1. Definition of medical terms and examples

Class	Example
Department	Cardiology (순환기내과), Surgery (외과), Urology (비뇨의학과)
Symptom, disease	Chest pain (흉통), hypertension (고혈압), cancer (암), bleeding (출혈)
Organ, location	Heart (심장), chest (가슴), blood (피)
Test	Blood test (피검사), electrocardiogram (심전도), endoscopy (내시경)
Treatment	Operation (수술), admission (입원), radiation therapy (방사선치료)
Medication	Drug (약), antiplatelet agent (항혈전제), antihypertensive agent (혈압약)
Specific name of a medication	Aspirin (아스피린), omega-3 (오메가쓰리), lipitor (리피토), clopidogrel (클로피도그렐)

Statistical analysis

For continuous data, differences were compared using the t-test or the Mann-Whitney U test, as applicable, and data were presented as mean ± standard deviation values or median with interquartile range values. Categorical data were presented as number (percentage) values and compared using the χ^2 or Fisher's exact test. Word counts from original transcriptions were defined as the reference values, and we compared the accuracy rate of each SR open API. All statistical analyses were performed using Python 3 and R version 4.1.0 (R Foundation for Statistical Computing, Vienna, Austria). All tests were two-tailed, and $P < 0.050$ was considered to be statistically significant.

Ethics statement

The study was approved by the institutional review board at Samsung Medical Center (2021-03-123-001), and written informed consent was obtained from all involved patients. This prospective study was conducted in accordance with the principles of the Declaration of Helsinki.

RESULTS

Table 2 shows the baseline characteristics of the original transcriptions. Since the specialty of the attending physicians was preoperative cardiac evaluation and prevention, 79 patients were seen for preoperative visits and the others were seen for diagnosis or management of cardiologic diseases. The mean recording time was 328 seconds.

Accuracy according to SR open API is described in **Table 3**. Naver Clova SR showed the highest overall accuracy, but its accuracy rate was still relatively low (75.1%). In the analysis according to the class of medical terms, Naver Clova SR had the greatest accuracy throughout all classes, but statistically significant differences existed in two classes, symptom or disease and organ or location. In the aspects of accuracy according to word length, Naver Clova SR demonstrated the highest accuracy for words shorter than three letters, while the difference was diminished as the word length increased. The recognition accuracy of non-Korean words was also numerically higher with Naver Clova SR (58.6%) compared to Google Speech-to-Text (35.5%) and Amazon Transcribe (30.9%) but still statistically insignificant.

In the comparison between Google Speech-to-Text and Amazon Transcribe, the overall recognition accuracy was higher with the latter (57.9% vs. 50.9%; $P < 0.001$), but the numerical difference was not statistically significant. Amazon Transcribe showed better accuracy in the class of symptom or disease (64.7% vs. 53.5%; $P = 0.008$) and poorer accuracy in the class of specific name of the medication (22.8% vs. 33.0%; $P = 0.010$) compared to Google Speech-to-Text. In addition, the accuracy for words measuring shorter than three letters was higher with Amazon Transcribe than Google Speech-to-Text, but the accuracy for words longer than four letters was higher with Google Speech-to-Text. For sensitivity

Table 2. Baseline characteristics of the original transcriptions

Characteristics	Value
Preoperative visit	79 (70.5)
Recording time (seconds)	328 ± 161
Extracted medical vocabularies	25.30 ± 7.48
Total word count	65.40 ± 26.89
Non-Korean words	1.88 ± 1.71

Data are presented as number (%) or mean ± standard deviation values.

Table 3. Accuracy according to cloud-based speech recognition open application programming interface

Characteristics	Total words	Naver	Google	Amazon	Naver vs. Google	Naver vs. Amazon	Google vs. Amazon
Total	7,319	5,493 (75.1)	3,726 (50.9)	4,237 (57.9)	< 0.001	< 0.001	< 0.001
Class							
Department	276	145 (52.5)	141 (51.1)	128 (46.4)	0.320	0.140	0.630
Symptom, disease	1,343	1,060 (78.9)	718 (53.5)	869 (64.7)	< 0.001	0.005	0.008
Organ, location	1,935	1,627 (84.1)	1,104 (57.1)	1,410 (72.9)	< 0.001	0.003	0.700
Test	1,160	799 (68.9)	601 (51.8)	587 (50.6)	0.140	0.190	0.930
Treatment	1,251	944 (75.5)	522 (41.7)	605 (48.4)	0.110	0.060	0.780
Medication	1,139	840 (73.7)	569 (50.0)	589 (51.7)	0.330	0.330	0.980
Specific name of a medication	215	79 (36.7)	71 (33.0)	49 (22.8)	0.005	0.760	0.010
Word length							
1	1,108	894 (80.7)	542 (48.9)	658 (59.4)	< 0.001	0.030	0.002
2	3,695	3,049 (82.5)	1,874 (50.7)	2,387 (64.6)	< 0.001	< 0.001	< 0.001
3	1,468	955 (65.1)	749 (51.0)	740 (50.4)	0.290	0.100	0.540
4	659	408 (61.9)	337 (51.1)	305 (46.3)	0.900	0.080	0.090
5	325	171 (52.6)	183 (56.3)	119 (36.6)	0.430	0.110	0.010
6	61	15 (24.6)	39 (36.9)	27 (44.3)	0.250	0.670	0.460
7	1	1 (100.0)	1 (100.0)	1 (100.0)	-	-	-
8	2	1 (50.0)	1 (50.0)	0	-	-	-
Non-Korean terms	459	269 (58.6)	163 (35.5)	142 (30.9)	0.990	0.100	0.090

Data are presented as number (%).

analysis, we extracted the 10 most frequently appearing medical words and analyzed the recognition accuracy; here, Naver Clova SR showed significantly greater accuracy compared to the other two open APIs, while the difference between Google Speech-to-Text and Amazon Transcribe was limited (**Supplementary Table 1**).

We additionally analyzed the type of typos according to open APIs (**Supplementary Table 2**). The proportion of wrong words was the highest with Naver Clova SR, compared to Google Speech-to-Text and Amazon Transcribe (69.0% vs. 34.2% vs. 30.8%, $P < 0.001$, respectively), but there was no statistically significant difference between Naver Clova SR and Google Speech-to-Text ($P = 0.180$). Google Speech-to-Text and Amazon Transcribe showed a higher omission rates compared to Naver Clova SR (13.5% vs. 61.0% vs. 55.6%, $P < 0.001$, respectively).

DISCUSSION

The main findings of this study are as follows: 1) among cloud-based SR open APIs, the one that was manufactured by a domestic company showed the highest accuracy rate; 2) the recognition accuracy of cloud-based SR open APIs in recognizing medical terms is relatively low to apply in healthcare services; and 3) the SR performance of each open API showed strengths in different areas of medical terminology. The results of the current study may provide insights into the possibilities and obstacles of cloud-based SR open API adaptation to healthcare services.

SR technology is becoming widely used in healthcare services in areas ranging from simple remote symptom checks for individuals self-quarantining due to coronavirus disease 2019 to the documentation of medical records and classification of emergent patients.[11,12] Most of the time spent in EHR is allotted to the documentation of health records, and there were definite benefits in documentation speed seen when using SR technology, and this time-saving affects other benefits, work productivity, and cost-effectiveness.[1,13,14] Despite this, evidence about the benefits of SR usage for clinical documentation is implausible and relatively neutral due to its own barriers, such as the requirement of training to use SR and

interoperability with existing systems.[1,15] Most current medical SR systems target physicians' medical documentation, such as radiologic or pathologic reports, and a system for narrative documentation between the doctor and patient is lacking.[1] In addition, the implementation costs of an SR system may offset the clinical benefits. Compared to a conventional SR system, cloud-based SR open APIs have several promising benefits. The application of cloud-based SR open APIs to clinical documentation may alleviate the cost burden and the need for training on the SR system. Cloud-based SR open APIs have significantly improved with the development of AI and cloud computing technology and saved on the time and costs necessary to develop an applied SR system.[6] Moreover, the convenient implantation nature of cloud-based SR open APIs could reduce the time and cost of system development and implementation into different EHR systems and may ultimately contribute to the reduction of physician burnout in the present disastrous pandemic situation.[16]

In the present study, we compared the accuracy of recognition of medical terms between widely used cloud-based SR open APIs in Korea, and the results showed that the open API built by a domestic company (Naver Clova SR) had the highest accuracy between three open APIs. The performance of AI largely depends upon the quality and quantity of training databases, and domestic companies might have strengths in the aspects of collecting data in the native language. A previous study showed that domestic companies of Korea achieved greater performance in SR of Korean compared to international companies.[17] Despite this, the overall performance of SR with medical terms was relatively low at less than 80%, which means that the accuracy of cloud-based SR open APIs in collecting real doctor–patient conversations need to be improved. We could consider several explanations for the low accuracy. First, quality control of recording files in the outpatient clinic could not be maintained. During a consultation, noises from the physical examination; computer usage during the consultation; and individuals other than the patient and doctor who participate during the consultation, such as an attending nurse or a caretaker, also affect the quality. Moreover, patient–doctor speech is not a simple conversation organized in a question-and-answer format. During this conversation, interruptions by both the patient and doctor may occur, and the current cloud-based SR open APIs cannot discriminate between the voices of multiple speakers appropriately. Our own nature of medical terms could be another explanation for the low accuracy. The training of AI requires a number of databases, but medical terms are relatively less commonly used during ordinary discussion between individuals. Hence, training on medical terms might be a hard task. Moreover, although Naver Clover SR showed the highest recognition accuracy, the others showed greater accuracy in the recognition of longer words or some specific words. For example, "순환기내과 (sunhwanginaegwa, cardiology)" was recognized as "술 한잔 (sulhanjan, a drink of alcohol)," and "심방 (simbang, atrium)" as "신방 (sinbang, bridal room) by Naver Clova SR, and this may reflect that medical terms were not much included in the Naver Clova SR training data. The exact Korean pronunciation of each word is presented in **Supplementary Table 3** via a Google translate link. As a result, among 57 appearances of the word, Naver Clova SR only recognized it two times; meanwhile, Google Speech-to-Text recognized it 19 and Amazon Transcribe recognized it 20 times, respectively. The current result reflects the inherent bias of the training datasets of deep learning algorithms.[18] In addition, this draws the necessity to construct cloud-based SR open APIs dedicated to healthcare services, such as Amazon Medical Transcribe.[19] Moreover, each cloud-based SR open API showed different strengths according to vocabulary or word length. In this regard, improvements in performance are anticipated if we could combine the strengths of each API. To enhance the accuracy of cloud-based SR open APIs in healthcare services, these points should be considered.

The difference between our native language (Korean) and Latin-based languages needs to be considered. In Korea, conversations between caregivers, such as doctor–doctor or doctor–nurse conversations, generally take place with original medical terminologies in English. However, in the patient–doctor conversation, both participants use Korean except when using simple or famous words, such as "CT" or "aspirin." Hence, to construct well-established cloud-based SR open APIs that function in the area of healthcare services, not only recognition of native words but also foreign ones that are frequently used between caregivers needs to be guaranteed, especially in countries that use non-Latin-based languages. In addition, Korean is an agglutinative language distinctly different from Latin-based languages; its words can contain two or more morphemes to determine meanings semantically, and this makes SR of Korean more complicated.[20] Hence, there is a lot of room to improve the application of SR technology to healthcare services, even in Korea, one of the leading countries in the information technology industry. Many countries that use different language systems will not be able to share the benefits of technological advances of SR equally. This raises another important issue that AI should contribute to improving health inequalities.[19] The anticipation of AI's contributions to improving patient treatment and the doctor–patient relationship may not be equally distributed in this circumstance, especially in non-developed countries, and we need to have a serious discussion about this.[21,22] Improving AI translation algorithm from other languages into English may be another answer for this issue. If English translation quality is guaranteed, available medical SR systems, such as Amazon Transcribe Medical could be applied to patients using other languages.

One intriguing finding of the current study is the difference in typo patterns between APIs of domestic and international companies. Naver Clova SR showed a higher proportion of wrong words, while the other two showed more omissions. This pattern reflects the distinct features of the SR algorithm. We could assume that, when recognizing inaccurate or confusing input, Naver Clova SR seemed to try to match a similar word, but the others just skipped the word. Changing words into a different word may cause more serious problems than just skipping it in the area of healthcare services. For example, "신장 (sinjang)" means "kidney" in Korean and "심장 (simjang)" means "heart"; here, a single-character change leads to a significantly clinically different word. Future development of SR algorithms for healthcare services should consider this issue.

The following limitations should be considered when interpreting the results of this study. First, this study was performed in Korea; hence, SR performance in other languages could not be evaluated. Second, SR open APIs could not discriminate between multiple speakers appropriately. Third, the study was performed by a single cardiologist; hence, the disease spectrum was limited to the cardiologic area. However, more than half of patients visited for preoperative cardiac evaluations, so relatively various medical terms other than cardiologic ones were presented. Finally, the results of current study could not provide the most suitable SR open APIs for medical documentation owing to its substantial low accuracy. Despite these limitations, we provide information about the recognition capabilities of cloud-based SR open APIs during real doctor–patient conversations, and the results of the current study may inspire insight about the application of cloud-based SR open APIs to healthcare services.

In conclusion, among three current cloud-based SR open APIs, Naver Clova SR which manufactured by domestic company showed highest accuracy of medical terms in Korean, compared to Google Speech-to-Text and Amazon Transcribe. Although limitations are

existing in the recognition of medical terminology, there is a lot of room for improvement of this promising technology by combining strengths of each SR engines.

SUPPLEMENTARY MATERIALS

Supplementary Table 1
Accuracy according to top 10 frequent words

Click here to view

Supplementary Table 2
Error rate according to the classification of typos

Click here to view

Supplementary Table 3
URL of Korean pronunciation presented in the manuscript

Click here to view

REFERENCES

1. Hodgson T, Coiera E. Risks and benefits of speech recognition for clinical documentation: a systematic review. *J Am Med Inform Assoc* 2016;23(e1):e169-79.
   PUBMED | CROSSREF

2. Zhou L, Blackley SV, Kowalski L, Doan R, Acker WW, Landman AB, et al. Analysis of errors in dictated clinical documents assisted by speech recognition software and professional transcriptionists. *JAMA Netw Open* 2018;1(3):e180530.
   PUBMED | CROSSREF

3. Hinton G, Deng L, Yu D, Dahl G, Mohamed A, Jaitly N, et al. Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups. *IEEE Signal Process Mag* 2012;29(6):82-97.
   CROSSREF

4. Chiu CC, Sainath TN, Wu Y, Prabhavalkar R, Nguyen P, Chen Z, et al. State-of-the-art speech recognition with sequence-to-sequence models. In: *Proceeding of the 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*; 2018 April 15–20; Calgary, Canada. Manhattan, NY, USA: IEEE; 2018, 4774-8.

5. Johnson M, Lapkin S, Long V, Sanchez P, Suominen H, Basilakis J, et al. A systematic review of speech recognition technology in health care. *BMC Med Inform Decis Mak* 2014;14:94.
   PUBMED | CROSSREF

6. Yoo HJ, Seo S, Im SW, Gim GY. The performance evaluation of continuous speech recognition based on Korean phonological rules of cloud-based speech recognition open API. *Int J Networked Distrib Comput* 2021;9(1):10-8.
   CROSSREF

7. Spinazze P, Aardoom J, Chavannes N, Kasteleyn M. The computer will see you now: overcoming barriers to adoption of computer-assisted history taking (CAHT) in primary care. *J Med Internet Res* 2021;23(2):e19306.
   PUBMED | CROSSREF

8. Elmore N, Burt J, Abel G, Maratos FA, Montague J, Campbell J, et al. Investigating the relationship between consultation length and patient experience: a cross-sectional study in primary care. *Br J Gen Pract* 2016;66(653):e896-903.
   PUBMED | CROSSREF

9. Sinsky C, Colligan L, Li L, Prgomet M, Reynolds S, Goeders L, et al. Allocation of physician time in ambulatory practice: a time and motion study in 4 specialties. *Ann Intern Med* 2016;165(11):753-60.
   PUBMED | CROSSREF

10. Basma S, Lord B, Jacks LM, Rizk M, Scaranelo AM. Error rates in breast imaging reports: comparison of automatic speech recognition and dictation transcription. *AJR Am J Roentgenol* 2011;197(4):923-7.
   PUBMED | CROSSREF

11. Kim D, Oh J, Im H, Yoon M, Park J, Lee J. Automatic classification of the Korean triage acuity scale in simulated emergency rooms using speech recognition and natural language processing: a proof of concept study. *J Korean Med Sci* 2021;36(27):e175.
   PUBMED | CROSSREF

12. Kim S, Lee J, Choi SG, Ji S, Kang J, Kim J, et al. Building a Korean conversational speech database in the emergency medical domain. *Phon Speech Sci* 2020;12(4):81-90.
   CROSSREF

13. Kauppinen T, Koivikko MP, Ahovuo J. Improvement of report workflow and productivity using speech recognition--a follow-up study. *J Digit Imaging* 2008;21(4):378-82.
   PUBMED | CROSSREF

14. Blackley BV, Schubert VD, Goss FR, Al Assad W, Garabedian PM, Zhou L. Physician use of speech recognition versus typing in clinical documentation: a controlled observational study. *Int J Med Inform* 2020;141:104178.
   PUBMED | CROSSREF

15. Ghatnekar S, Faletsky A, Nambudiri VE. Digital scribe utility and barriers to implementation in clinical practice: a scoping review. *Health Technol (Berl)* 2021;11(4):803-9.
   PUBMED | CROSSREF

16. Swayamsiddha S, Prashant K, Shaw D, Mohanty C. The prospective of artificial intelligence in COVID-19 pandemic. *Health Technol (Berl)* 2021;11(6):1311-20.
   PUBMED | CROSSREF

17. Choi SJ, Kim JB. Comparison analysis of speech recognition open APIs' accuracy. *Asia Pac J Multimed Serv Converg Art Humanit Sociol* 2017;7(8):411-8.
   CROSSREF

18. Kaushal A, Altman R, Langlotz C. Geographic distribution of US cohorts used to train deep learning algorithms. *JAMA* 2020;324(12):1212-3.
   PUBMED | CROSSREF

19. Amazon. Guide of Amazon Transcribe Medical. https://aws.amazon.com/ko/transcribe/medical/. Updated 2022. Accessed January 18, 2022.

20. Choi DH, Park IN, Shin M, Kim EG, Shin DR. Korean erroneous sentence classification with Integrated Eojeol Embedding. *IEEE Access* 2021;9:81778-85.
   CROSSREF

21. Leslie D, Mazumder A, Peppin A, Wolters MK, Hagerty A. Does "AI" stand for augmenting inequality in the era of COVID-19 healthcare? *BMJ* 2021;372:n304.
   PUBMED | CROSSREF

22. Aminololama-Shakeri S, López JE. The doctor-patient relationship with artificial intelligence. *AJR Am J Roentgenol* 2019;212(2):308-10.
   PUBMED | CROSSREF