



Accuracy of the Information on Sudden Sensorineural Hearing Loss From Chat Generated Pre-Trained Transformer

Sang Jae Lee¹ , Hyung Gyun Na¹ , Yoon Seok Choi¹ , Si-Youn Song¹ ,
Yong-Dae Kim^{1,2} , and Chang Hoon Bae¹

¹Department of Otorhinolaryngology-Head and Neck Surgery, College of Medicine, Yeungnam University, Daegu; and

²Regional Center for Respiratory Diseases, Yeungnam University Medical Center, Daegu, Korea

Chat Generated Pre-Trained Transformer를 통해 얻은 돌발성 감각신경성 난청 정보의 정확성

이상재¹ · 나형균¹ · 최윤석¹ · 송시연¹ · 김용대^{1,2} · 배창훈¹

¹영남대학교 의과대학 이비인후-두경부외과학교실, ²영남대학교병원 권역 호흡기 전문질환센터

Received March 11, 2023

Revised April 3, 2023

Accepted April 14, 2023

Address for correspondence

Chang Hoon Bae, MD, PhD
Department of Otorhinolaryngology-
Head and Neck Surgery,
College of Medicine,
Yeungnam University,
170 Hyeonchung-ro, Nam-gu,
Daegu 42415, Korea
Tel +82-53-620-3784
Fax +82-53-628-7884
E-mail baich@med.yu.ac.kr

Background and Objectives Chat generated pre-trained transformer (ChatGPT) is a conversational artificial intelligence model, which has recently attracted worldwide attention by enabling natural conversations based on huge information from deep learning in various fields. Several studies have reported usefulness and reliability of medical information obtained by ChatGPT, but there are no studies explaining reliability and accuracy in the field of otorhinolaryngology. On this regard, we investigated the accuracy of information on sudden sensorineural hearing loss obtained by ChatGPT.

Materials and Method Twenty-five questions and answers related to sudden sensorineural hearing loss were recorded from ChatGPT based on textbook from Korean Society of Otorhinolaryngology-Head and Neck Surgery and Clinical Guidelines of American Academy of Otolaryngology-Head and Neck Surgery. Answers were shown to one specialist in otorhinolaryngology in a blind test and asked to assess their accuracy. Each question was rated as 'accurate' or 'inaccurate.' If the contents were not found in the textbook or the guidelines, it was rated as 'unreliable.'

Results Of the 25 questions, 19 (76%) were identified as 'accurate,' 6 (24%) were 'inaccurate,' and 0 (0%) were 'unreliable.' Questions about definition, prevalence, hearing rehabilitation, diagnosis and treatment were found to be more accurate than the average, while causes and prognosis were less accurate than the average.

Conclusion The information on sudden sensorineural hearing loss obtained from ChatGPT was quite accurate. It is expected to provide substantial help to patients and doctors. As medicine and medical artificial intelligence develop together, further research is needed on reliability and accuracy in various diseases and fields of otorhinolaryngology-head and neck surgery.

Korean J Otorhinolaryngol-Head Neck Surg 2024;67(2):74-8

Keywords Artificial intelligence; Data accuracy; Hearing loss, sudden.

서론

Chat generated pre-trained transformer (ChatGPT)는 2022년 11월에 open AI 회사에서 개발된 대화형 인공지능 (artificial intelligence) 모델이다. 서비스 공개 후 5일 만에 하루 이용자가 100만 명을 넘었고 의학을 포함한 다양한 주제에 대해 자연어 처리와 기계학습 기술을 이용하여 이용자와 대화하고 질문에 답변을 할 수 있어 전 세계적인 관심을 얻고 있다. 손쉽게 많은 양의 정보를 얻을 수 있지만 심층 학습(deep learning)을 통해 기존의 정보를 학습하고 이를 바탕으로 대답하는 인공지능의 특성상 아직까지 신뢰도에 대한 의문점이 있으며 특히 환자들에게 잘못된 의학적인 정보를 제공하면 치명적인 결과를 초래할 수 있어 주의가 필요하다.¹⁾

돌발성 감각신경성 난청(sudden sensorineural hearing loss)은 확실한 원인 없이 갑자기 발생하는 감각신경성 난청으로 10만 명당 10명 이상 발병하는 것으로 보고되고 있다. 대부분의 경우 원인을 찾을 수 없고 갑작스럽게 발생하며 예후가 좋지 않기 때문에 충분한 설명을 해도 환자들이 쉽게 이해하는데 어려움이 있다. 돌발성 감각신경성 난청은 초기에 발견하여 치료하는 것이 예후에 많은 영향을 미치므로 이 과학 분야에서 응급한 질환들 중 하나로 볼 수 있다.^{2,3)}

ChatGPT에 의해 얻어진 의학적 정보의 유용성과 신뢰도에 대한 연구들이 보고되고 있다.⁴⁻⁷⁾ 하지만 아직 이비인후과 분야에서 신뢰도와 정확도를 조사한 연구는 없으며, 특히 돌발성 감각신경성 난청의 응급도와 이해의 어려움을 고려하였을 때 이에 대한 연구가 필요할 것으로 생각된다.

이에 본 문헌을 통해 ChatGPT를 통해 얻은 돌발성 감각신경성 난청에 대한 정보의 정확도에 대해 조사하고 보고하고자 한다.

재료 및 방법

저자들은 돌발성 감각신경성 난청의 정의(1번 문항), 유병률(2번 문항), 원인(3번 문항), 진단(4-9번 문항), 치료(10-23번 문항), 예후(24번 문항), 청력재활(25번 문항)에 관련한 25가지 질문들을 2018년도에 대한이비인후과학회에서 발행한 이비인후과학 교과서 개정2판 34장 돌발성 난청⁸⁾과 2019년도에 배포된 American Academy of Otolaryngology-Head And Neck Surgery (AAO) 임상 가이드라인³⁾을 기준으로 만들었다(Table 1). 각 질문들을 ChatGPT Feb 13 version (OpenAI, San Francisco, CA, USA)에 한국어로 3차례 입력하여 질문에 대한 대답들을 종합하여 기록하였다(Supplementary Material). 이 기록들을 ChatGPT에 대한 언급

없이 맹검시험(blind test)을 통해 이비인후과 전문의 1명에게 보여주고 정확도를 평가하도록 하였다. 평가자는 각 질문에 대한 반응을 2018년도 대한이비인후과학회에서 발행한 이비인후과학 교과서 개정2판 34장 돌발성 난청⁸⁾과 2019년 AAO 가이드라인³⁾을 바탕으로 '정확하다', '부정확하다'의 두 가지로 나누고, 교과서와 가이드라인에 없는 내용일 경우 '신뢰할 수 없다'로 평가하였다. '정확하다'는 문맥의 내용이 교과서와 가이드라인에 부합하는 경우, '부정확하다'는 문맥의 내용 중 2가지 이상 잘못된 것이 있을 때로 간주하였다.

결과

25가지 질문에서 19가지 항목(76%)에서 '정확하다', 6가지 항목(24%)에서 '부정확하다', 0가지 항목(0%)에서 '신뢰할 수 없다'로 확인되었다(Fig. 1).

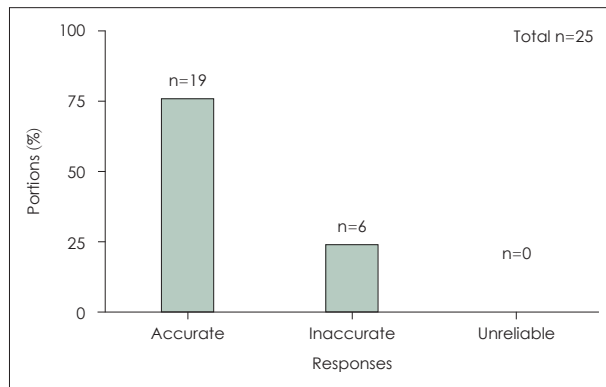
각 세부항목으로 나누면 정의는 1가지 질문 중 1가지 항목(100%)에서 '정확하다'로 정확도가 전체 질문의 평균(76%)보다 높았다. 유병률에서는 1가지 질문 중 1가지 항목(100%)에서 '정확하다'로 정확도가 전체 질문의 평균(76%)보다 높았다. 원인은 1가지 질문 중 1가지 항목(100%)에서 '부정확하다'로 정확도가 전체 질문의 평균(76%)보다 낮았다. 진단은 6가지 질문 중 5가지 항목(83%)에서 '정확하다', 1가지 항목(17%)에서 '부정확하다'로 정확도가 전체 질문의 평균(76%)보다 높았다. 치료는 14가지 질문 중 11가지 항목(79%)에서 '정확하다', 3가지 항목(21%)에서 '부정확하다'로 확인되어 정확도가 질문의 평균(76%)보다 높았다. 예후는 1가지 질문 중 1가지 항목(100%)에서 '정확하다'로 정확도가 전체 질문의 평균(76%)보다 낮았다. 청력재활에서는 1가지 질문 1가지 항목(100%)에서 '정확하다'로 정확도가 전체 질문의 평균(76%)보다 높았다(Fig. 2).

고찰

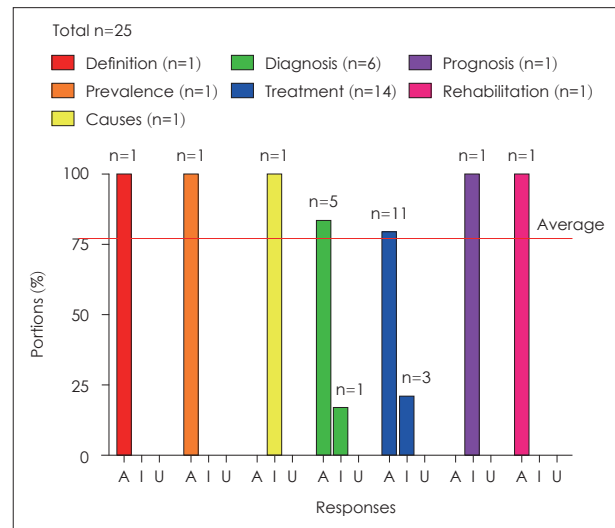
Sarraj et al.⁴⁾이 ChatGPT를 통해 얻은 심혈관질환의 예방에 대한 의학 정보에서 84%의 높은 정확도를 보였고, Johnson 등⁷⁾은 다양한 분야의 의학 관련 180개의 질문에서 57.8%의 정확도를 보였다. 이 연구를 통해서 ChatGPT가 돌발성 감각신경성 난청에 대해 70% 이상의 높은 정확도를 가진 의학 정보를 준다는 것을 확인하였다. 배포된 지 얼마 되지 않은 이 의료인공지능 도구가 이렇게 높은 정확도를 보이는 것은 시사하는 바가 크다. 첫 번째는 의료인공지능이 환자들에게 주는 신뢰도가 올라가고 있다. 많은 연구들에서 의료인공지능의 신뢰도와 한계에 대한 지적들이 있었으나 실제로 사용

Table 1. Questions to chat generated pre-trained transformer about the sudden sensorineural hearing loss

번호	분류	질문	정확도
1	정의	돌발성 감각신경성 난청의 정의	정확하다
2	유병률	돌발성 감각신경성 난청의 유병률	정확하다
3	원인	돌발성 감각신경성 난청의 원인	부정확하다
4	진단	돌발성 감각신경성 난청의 진단	정확하다
5	진단	돌발성 감각신경성 난청에서 청력검사의 필요성	정확하다
6	진단	돌발성 감각신경성 난청에서 CT의 필요성	부정확하다
7	진단	돌발성 감각신경성 난청에서 MRI의 필요성	정확하다
8	진단	돌발성 감각신경성 난청에서 뇌간유발반응검사의 필요성	정확하다
9	진단	돌발성 감각신경성 난청에서 혈액검사의 필요성	정확하다
10	치료	돌발성 감각신경성 난청의 치료	정확하다
11	치료	돌발성 감각신경성 난청의 치료가능 기간	정확하다
12	치료	돌발성 감각신경성 난청에서 전신적 스테로이드 사용의 효과	정확하다
13	치료	돌발성 감각신경성 난청에서 고실 내 스테로이드 사용의 효과	정확하다
14	치료	돌발성 감각신경성 난청에서 고압산소치료의 효과	부정확하다
15	치료	돌발성 감각신경성 난청에서 혈관확장제의 효과	정확하다
16	치료	돌발성 감각신경성 난청에서 항응고제의 효과	정확하다
17	치료	돌발성 감각신경성 난청에서 항바이러스제의 효과	정확하다
18	치료	돌발성 감각신경성 난청에서 성상신경차단술의 효과	정확하다
19	치료	돌발성 감각신경성 난청에서 이노제의 효과	부정확하다
20	치료	돌발성 감각신경성 난청에서 저염식의 효과	부정확하다
21	치료	돌발성 감각신경성 난청에서 입원치료의 효과	정확하다
22	치료	돌발성 감각신경성 난청에서 수술적 치료의 효과	정확하다
23	치료	돌발성 감각신경성 난청 치료에서 추적 청력검사 기간	정확하다
24	예후	돌발성 감각신경성 난청의 예후	부정확하다
25	청력재활	돌발성 감각신경성 난청에서 청력재활의 필요성	정확하다

**Fig. 1.** Accuracy of the responses of sudden sensorineural hearing loss from chat generated pre-trained transformer. n, number of questions.

되는 도구에서 전문의에게 정보의 정확도를 확인할 수 있었다.⁹⁻¹¹⁾ 두 번째는 환자에게 직접적으로 도움을 줄 수 있다. 전통적으로 의료인공지능이 의사의 진단을 도와주고 효율성을 높여주지만 환자에게 직접적인 도움을 주지는 못했다. Amann 등¹²⁾은 의료 인공지능의 실현가능성을 다양한 측면에서 살펴보았을 때 환자를 치료 과정에서 적극적인 파트너로 간주하여 의료결정에서 선택권과 통제권에 대해 강조하였

**Fig. 2.** Accuracy of the responses of sudden sensorineural hearing loss from chat generated pre-trained transformer based on the classifications. Red line indicates the average in whole questions. A, accurate; I, inaccurate; U, unreliable; n, number of questions.

다. 환자가 실제로 의사뿐만 아니라 ChatGPT 등의 의료 인공지능에서 충분한 정보를 제공받고 의료 결정 과정에 함께 참여할 수 있을 것으로 기대된다. 세 번째는 촉박한 진료 시

간으로 인한 환자에게 부족한 설명을 채워주고 알 권리를 충족시킬 수 있을 것으로 기대된다.

정의, 유병률, 진단, 치료, 청력 재활의 정보는 평균보다 정확도가 높은 반면, 원인과 예후에 대한 정보는 평균보다 정확도가 낮았다. 돌발성 감각신경성 난청의 원인과 예후에 대한 내용은 아직까지 명확히 밝혀진 것이 없고, 자료마다 결과가 달라서 정확도가 떨어지는 것으로 추정된다. 또한 Johnson 등⁷⁾의 연구에 따르면 질문의 난이도에 따라 어려운 질문들이 쉬운 질문들에 비해 정확도가 낮은 것으로 보고하였다.

24%의 '부정확하다' 평가를 받은 6가지 항목에 대해서 수정할 필요가 있다. 첫 번째, 질문 3의 돌발성 감각신경성 난청의 원인에서, 돌발성 감각신경성 난청은 대부분의 경우 원인을 찾을 수 없으며 감염, 혈관장애, 와우막파열, 내림프수종, 외상, 이독성, 종양 등 다양하게 나타나고, 여러 개의 원인이 복합적으로 작용하였을 가능성이 높다.⁸⁾ 두 번째, 질문 6의 돌발성 감각신경성 난청에서 전산화단층촬영(CT)의 필요성에서, 전산화단층촬영은 돌발성 감각신경성 난청의 진단에 중요한 부분을 차지하지 않으며 일반적으로 시행하는 것은 강하게 권하지 않는 것으로 보고되고 있다.³⁾ 세 번째로, 질문 14의 돌발성 감각신경성 난청에서 고압산소치료의 효과에서, 돌발성 감각신경성 난청에서 고압산소치료는 2019년 AAO 임상 가이드라인에서 초기 치료와 구제요법으로 시행해볼 수 있다고 나타내고 있고, 권고 수준은 선택적(option)으로 제시하고 있다.³⁾ 고압산소치료가 돌발성 감각신경성 난청의 치료에 '효과적입니다'는 표현은 고압산소치료가 반드시 필요한 것으로 오해가 될 소지가 있어 '도움이 될 수 있습니다'의 완곡한 표현으로 고치는 것이 좋겠다. 네 번째와 다섯 번째, 질문 19와 20의 돌발성 감각신경성 난청에서 이노제와 저염식의 효과에서, 이노제와 저염식의 효과는 혈압을 감소 시켜 증상을 개선하는 것이 아니라 돌발성 감각신경성 난청에서 내림프 수종이 동반되는 경우가 있고 청력저하가 처음 나타나는 경우에는 메니에르병과 감별할 수 없기 때문에 이노제와 저염식을 통해 함께 치료하는 것으로 수정이 필요하겠다.^{8,13)} 여섯 번째로, 질문 24의 돌발성 감각신경성 난청의 예후에서, 돌발성 감각신경성 난청의 예후는 치료 시기와 방법 등 여러 인자에 의해 다양하게 나타나는 것은 맞지만 예후는 좋지 않은 질환이다. '대부분의 환자들은 적극적인 치료를 받으면 호전되는 경우가 많다'는 표현은 환자에게 잘못된 정보를 제공하고 치료 실패에 대한 실망감을 줄 수 있어서 반드시 수정이 필요하다.^{2,8)}

ChatGPT가 심층 학습을 기반으로 하고 있기 때문에 가이드라인과 교과서에 없는 내용을 제공하지는 않았고 '신뢰할 수 없다'의 항목은 나타나지 않았다. 하지만 더 높은 정확도

를 가지기 위해서는 수 많은 정보들 속에서 양질의 정보를 구분하는 능력이 더 필요할 것으로 생각된다. 특히, 의학 분야에서는 근거 수준(quality of evidence)과 권고의 등급(grading of recommendation)을 반영하여 신뢰도 높은 정보를 주는 것이 중요하다.

돌발성 감각신경성 난청은 불안과 우울이 연관되어 있는 경우가 많았고 적절한 정신과적 개입을 하는 것이 치료에 대한 반응률과 삶의 질을 높이는 데 도움이 된다.^{14,15)} 의사들도 환자들이 ChatGPT에서 얻은 정보에 대해서 많은 질문을 할 수 있어 이에 대한 대비가 필요하다. 적절한 답변과 충분한 정보제공으로 환자에게 신뢰관계가 형성하고 적절한 정신과적 개입을 하는 것이 치료 결과에 좋은 영향을 미칠 수 있을 것으로 기대된다.

대부분의 질문의 마지막 문단에는 즉시 의료전문가의 진료를 받는 것이 중요하고 의료 전문가와 상담이 필요하다고 설명하고 있다. 환자들은 ChatGPT의 정보에만 의존하여 환자 스스로 파악하고 이해하는 것 보다는 반드시 이비인후과 의사에게 방문하여 진찰과 검사를 받아보는 것이 중요하다.

ChatGPT의 몇 가지 한계점들이 존재하였다. 미국에서 만들어진 인공지능이고 한국어 질문에 대해 번역을 하여 답변을 해서 오타와 오역 등의 문제가 있을 수 있다. 실제로 같은 질문을 영어로 하였을 때 더 정확한 정보를 얻을 수 있었다. 그리고 2021년까지의 데이터를 기반으로 하고 있어 급변하는 의학에서 최신 지견을 반영하는 데 어려움이 있다. 같은 질문에 대해 항상 일관된 답변을 주는 것이 아니라 정보가 조금씩 변경되는 것도 개선이 필요하다.

이 연구의 제한점들로는 첫 번째가 전체 질문의 수가 25개로 적었으며, 정의, 유병률, 원인, 예후, 청력재활의 항목에는 1가지의 질문밖에 없어 분류에 따른 정확도 평가와 분석이 부족할 수 있다. 추후 연구에서는 질문의 난이도를 고려하고, 질문의 숫자가 충분하면서 균등하게 배분된다면 분류에 따른 정확도를 더 정밀하게 비교할 수 있을 것으로 생각된다. 두 번째로는 정확도 평가를 이비인후과 전문의 1명에게만 시행한 점과 2018년도 대한이비인후과학회에서 발행한 이비인후과학 교과서 개정2판 34장 돌발성 난청과 2019년 AAO 가이드라인 2개만을 기반으로 한 점이다. 추후 연구에서는 평가자를 이비인후과 전문의 다수로 하고 공신력 있는 최신 논문과 저서를 기반으로 정확도를 평가한다면 더 신뢰도 높은 결과를 가질 수 있었을 것으로 생각된다. 그럼에도 불구하고 이 연구는 돌발성 감각신경성 난청에 대해서 ChatGPT가 제시하는 정보의 신뢰도와 정확도를 조사하는 연구에 많은 정보를 제공할 것으로 생각된다.

결론적으로, 의학과 의료 인공지능 모두 발전하고 있는 분

아이며 다양한 인자들이 포함되어 있어 일반인 수준으로 쉽게 표현하는 것이 어렵지만 ChatGPT를 통해 얻은 돌발성 감각신경성 난청에 대한 정보의 정확도는 비교적 우수하였다. 앞으로 의료인공지능이 발전하여 돌발성 감각신경성 난청에 대해 환자와 의사에게 실질적인 도움을 줄 수 있을 것으로 기대되며, 돌발성 감각신경성 난청뿐만 아니라 다른 이비인후과 질환들과 분야들까지 인공지능의 신뢰성 및 발전가능성에 대한 연구가 꾸준히 필요하다.

Supplementary Materials

The Data Supplement is available with this article at <https://doi.org/10.3342/kjorl-hns.2023.00248>.

Acknowledgments

None

Author Contribution

Conceptualization: Sang Jae Lee, Chang Hoon Bae. Data curation: Sang Jae Lee, Chang Hoon Bae. Formal analysis: Hyung Gyun Na, Yoon Seok Choi. Investigation: Chang Hoon Bae, Si-Youn Song. Methodology: Sang Jae Lee, Chang Hoon Bae. Supervision: Chang Hoon Bae, Yong-Dae Kim. Visualization: Sang Jae Lee, Hyung Gyun Na. Writing—original draft: Sang Jae Lee. Writing—review & editing: Sang Jae Lee, Chang Hoon Bae.

ORCIDs

Sang Jae Lee	https://orcid.org/0000-0003-2368-4524
Hyung Gyun Na	https://orcid.org/0000-0003-4017-4796
Yoon Seok Choi	https://orcid.org/0000-0002-0616-7122
Si-Youn Song	https://orcid.org/0000-0001-6642-8841
Yong-Dae Kim	https://orcid.org/0000-0003-0501-966X
Chang Hoon Bae	https://orcid.org/0000-0002-0835-4060

REFERENCES

- 1) Open AI. ChatGPT: optimizing language models for dialogue. [cited 2023 Feb 13]. Available from: URL: <https://openai.com/blog/chatgpt/>.
- 2) Kuhn M, Heman-Ackah SE, Shaikh JA, Roehm PC. Sudden sensorineural hearing loss: A review of diagnosis, treatment, and prognosis. *Trends Amplif* 2011;15(3):91-105.
- 3) Chandrasekhar SS, Tsai Do BS, Schwartz SR, Bontempo LJ, Faucett EA, Finestone SA, et al. Clinical practice guideline: Sudden hearing loss (update). *Otolaryngol Head Neck Surg* 2019;161(1 Suppl):S1-45.
- 4) Sarraju A, Bruemmer D, Van Iterson E, Cho L, Rodriguez F, Laffin L. Appropriateness of cardiovascular disease prevention recommendations obtained from a popular online chat-based artificial intelligence model. *JAMA* 2023;329(10):842-4.
- 5) Salvagno M, Taccone FS, Gerli AG. Can artificial intelligence help for scientific writing? *Crit Care* 2023;27(1):75.
- 6) Kung TH, Cheatham M, Medenilla A, Sillos C, De Leon L, Elepaño C, et al. Performance of ChatGPT on USMLE: Potential for AI-assisted medical education using large language models. *PLOS Digit Health* 2023;2(2):e0000198.
- 7) Johnson D, Goodman R, Patrinely J, Stone C, Zimmerman E, Donald R, et al. Assessing the accuracy and reliability of AI-generated medical responses: An evaluation of the Chat-GPT model. *Research Square* [Preprint]. 2023 [cited 2023 Feb 28]. Available at: <https://doi.org/10.21203/rs.3.rs-2566942/v1>.
- 8) Park MS, Song JJ. Sudden sensorineural hearing loss. In: *Korean Society of Otorhinolaryngology-Head and Neck Surgery*, editor. *Otorhinolaryngology-Head and Neck Surgery*. 2nd ed. Paju: Koonja Publishing Inc.;2018. p.631-40.
- 9) Hatherley JJ. Limits of trust in medical AI. *J Med Ethics* 2020; 46(7):478-81.
- 10) Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. *BMC Med* 2019;17(1):195.
- 11) Wu E, Wu K, Daneshjou R, Ouyang D, Ho DE, Zou J. How medical AI devices are evaluated: Limitations and recommendations from an analysis of FDA approvals. *Nat Med* 2021;27(4):582-4.
- 12) Amann J, Blasimme A, Vayena E, Frey D, Madai VI, Precise4Q Consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. *BMC Med Inform Decis Mak* 2020;20(1):310.
- 13) Liu H, Zhou K, Zhang X, Peng KA. Fluctuating sensorineural hearing loss. *Audiol Neurotol* 2019;24(3):109-16.
- 14) Sun B, Liu L, Ren X, Wang Z. Psychological state of patients with sudden deafness and the effect of psychological intervention on recovery. *J Int Med Res* 2020;48(9):300060520957536.
- 15) Kim JY, Lee JW, Kim M, Kim MJ, Kim DK. Association of idiopathic sudden sensorineural hearing loss with affective disorders. *JAMA Otolaryngol Head Neck Surg* 2018;144(7):614-21.