



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2024-0067186
(43) 공개일자 2024년05월16일

(51) 국제특허분류(Int. Cl.)

G16H 50/20 (2018.01) G06N 3/0442 (2023.01)
G06N 3/0464 (2023.01) G10L 25/18 (2013.01)
G10L 25/24 (2013.01) G10L 25/66 (2013.01)
G16H 10/60 (2018.01)

(52) CPC특허분류

G16H 50/20 (2018.01)
A61B 5/4205 (2013.01)

(21) 출원번호 10-2022-0147315

(22) 출원일자 2022년11월07일

심사청구일자 없음

(71) 출원인

주식회사 케이티

경기도 성남시 분당구 불정로 90(정자동)

연세대학교 산학협력단

서울특별시 서대문구 연세로 50 (신촌동, 연세대학교)

(뒷면에 계속)

(72) 발명자

이주강

서울특별시 서초구 서초대로64길 36, 101동 505호(서초동, 대성유니드아파트)

박윤길

서울특별시 강남구 언주로30길 13, B-1204(도곡동, 대림아크로빌, 대림아크로텔)

(뒷면에 계속)

(74) 대리인

특허법인충정

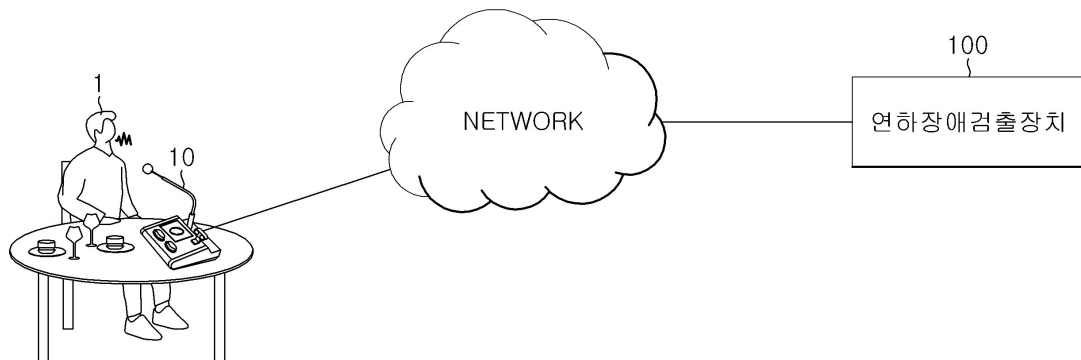
전체 청구항 수 : 총 20 항

(54) 발명의 명칭 음성신호 기반의 연하장애검출장치 및 연하장애검출방법

(57) 요약

본 발명은 음성신호 기반의 연하장애검출장치 및 연하장애검출방법에 관한 것으로, 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출방법은 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와, 상기 복수의 정상인 및 연하장애환자에 대해 수행한 비디오 투시 연하검사(VFSS: Video Fluoroscopic swallowing study)의 검사결과를 인공지능모델로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성하는 단계; 검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신하는 단계; 및 상기 대상 음성신호를 분석모델에 입력하여, 상기 검사대상자에 대한 연하장애를 판정하여 판단결과를 도출하는 단계를 포함할 수 있다.

대표도



(52) CPC특허분류

A61B 5/4803 (2013.01)
G06N 3/0442 (2023.01)
G06N 3/0464 (2023.01)
G10L 25/18 (2013.01)
G10L 25/24 (2013.01)
G10L 25/66 (2013.01)
G16H 10/60 (2021.08)
G16H 50/50 (2018.01)

(71) 출원인

재단법인 아산사회복지재단
 서울특별시 송파구 올림픽로43길 88 (풍납동)
가천대학교 산학협력단
 경기도 성남시 수정구 성남대로 1342 (복정동)
(의료)길의료재단
 인천광역시 남동구 남동대로774번길 21 (구월동)
울산대학교 산학협력단
 울산광역시 남구 대학로 93(무거동)

(72) 발명자

최경효
 서울특별시 강남구 압구정로 201, 79동 904호(압구정동, 현대아파트)
강상익
 경기도 성남시 분당구 불정로 90(정자동)
금명철
 경기도 성남시 분당구 불정로 90(정자동)
조경일
 경기도 성남시 분당구 불정로 90(정자동)
고훈석
 경기도 성남시 분당구 불정로 90(정자동)
권도완
 경기도 성남시 분당구 불정로 90(정자동)
권승현
 경기도 성남시 분당구 불정로 90(정자동)
이재호
 경기도 성남시 분당구 불정로 90(정자동)
박진영
 서울특별시 강남구 언주로63길 20 (역삼동)

명세서

청구범위

청구항 1

연하장애검출장치를 이용한 음성신호 기반의 연하장애검출방법에 있어서,

검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신하는 단계; 및

상기 대상음성신호를 분석모델에 입력하여, 상기 검사대상자에 대한 연하장애를 판정하여 판단결과를 도출하는 단계를 포함하는 것으로,

상기 분석모델은

복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와, 상기 복수의 정상인 및 연하장애 환자에 대해 수행한 비디오 투시 연하검사(VFSS: Video Fluoroscopic swallowing study)의 검사결과를 인공지능모델로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 것인 음성신호 기반의 연하장애검출방법.

청구항 2

제1항에 있어서, 상기 분석모델은

상기 음성신호에 대한 전처리를 수행하여 1차원 특징정보를 생성하고, 상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리(LSTM: Long Short Term Memory) 모델로 학습하여 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 3

제2항에 있어서, 상기 1차원 특징정보는

상기 음성신호의 기본주파수, 지터(jitter), 시머(shimmer) 및 배음 대 소음비(HNR: Harmonics to Noise Ratio) 중 적어도 어느 하나를 포함하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 4

제1항에 있어서, 상기 분석모델은

상기 음성신호에 대한 전처리를 수행하여 2차원 이미지를 생성하고, 상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를, 합성곱 신경망(CNN: Convolutional Neural Network) 모델, 비전 트랜스포머(ViT: Vision Transformer) 모델 및 오토인코더(Auto-encoder) 모델 중 어느 하나로 학습하여 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 5

제4항에 있어서, 상기 2차원 이미지는

상기 음성신호에 단시간 푸리에 변환(STFT: Short-Time Frequency Transform) 또는 멜 주파수 켈프스트럼 계수(MFCC: Mel-Frequency Cepstral Coefficient)를 적용하여 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 6

제1항에 있어서, 상기 분석모델은

상기 음성신호에 대한 전처리를 수행하여 1차원 특징정보 및 2차원 이미지를 생성하고,

상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리 모델로 학습하는 제1 모델, 상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를 합성곱 신경망 모델, 비전 트랜스포머 모델 및 오토인코더 모델 중 적어도 어느 하나로 학습하는 제2 모델 및 상기 제1 모델로부터 추출한 제1 특징벡터와 상기 제2 모델로부터 추출한 제2 특징벡터를 결합한 결합벡터에 적용하는 분류기(classifier) 함수를 생성한 후,

상기 제1 모델, 제2 모델 및 분류기 함수를 학습하여 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 7

제6항에 있어서, 상기 분석모델은

상기 정상인과 연하장애환자의 인적정보를 수집하고, 상기 인적정보로부터 생성한 제3 특징벡터를, 상기 제1 특징벡터 및 제2 특징벡터와 더 결합하여 상기 결합벡터를 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 8

제7항에 있어서,

상기 인적정보는, 상기 연하장애환자의 성별, 나이 및 기저질환 중 적어도 어느 하나를 포함하고,

상기 제3 특징 벡터는, 상기 기저질환을 나타내는 원-핫 벡터(1-hot vector)에 상기 성별 또는 나이를 나타내는 정보를 추가하여 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 9

제6항에 있어서, 상기 제2 모델은,

상기 2차원 이미지가 상기 오토인코더 모델에 입력되면, 상기 오토인코더 모델에 의해 생성된 복원 2차원 이미지를, 상기 2차원 이미지와 채널 방향으로 결합(concatenation)하여 결합 2차원 이미지를 생성하고,

상기 결합 2차원 이미지가 상기 합성곱 신경망 모델에 입력되면, 상기 결합 2차원 이미지에 대응하는 특징맵을 생성하며,

상기 특징맵을 풀링(pooling)하여 복수의 비전 트랜스포머 모델에 각각 풀링된 특징맵을 입력하면, 각각의 비전 트랜스포머 모델에서 생성한 비전 트랜스포머 특징벡터들을 결합(concatenation)하여 제2 특징벡터를 생성하도록, 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 10

제1항에 있어서, 상기 판단결과를 도출하는 단계는

상기 검사대상자의 연하장애를 정상 또는 비정상으로 판정하여 상기 판단결과를 도출하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 11

제1항에 있어서, 상기 판단결과를 도출하는 단계는

상기 검사대상자의 연하장애를 침습, 흡인 및 정상 중 어느 하나로 판정하거나, 상기 연하장애를 침습-흡인 척도(PAS: Penetration-aspiration score)에 따른 복수의 단계 중 어느 하나로 판정하여 상기 판단결과를 도출하는 것을 특징으로 하는 음성신호 기반의 연하장애검출방법.

청구항 12

하드웨어와 결합하여, 제1항 내지 제11항 중 어느 하나의 항에 따른 음성신호 기반의 연하장애검출방법을 수행하기 위하여 매체에 저장된 컴퓨터 프로그램.

청구항 13

복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와, 상기 복수의 정상인 및 연하장애환자에 대해 수행한 비디오 투시 연하검사(VFSS: Video Fluoroscopic swallowing study)의 검사결과를 인공지능모델로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성하는 모델생성부;

검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신하는 음성수신부; 및

상기 대상음성신호를 분석모델에 입력하여, 상기 검사대상자에 대한 연하장애를 판정하여 판단결과를 도출하는 판정부를 포함하는 음성신호 기반의 연하장애검출장치.

청구항 14

제13항에 있어서, 상기 모델생성부는

상기 음성신호에 대한 전처리를 수행하여, 1차원 특징정보를 생성하는 전처리부; 및

상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리(LSTM: Long Short Term Memory) 모델로 학습하여, 상기 분석모델을 생성하는 학습부를 포함하는 것을 특징으로 하는 음성신호 기반의 연하장애검출장치.

청구항 15

제13항에 있어서, 상기 모델생성부는

상기 음성신호에 대한 전처리를 수행하여, 2차원 이미지를 생성하는 전처리부; 및

상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를, 합성곱 신경망(CNN: Convolutional Neural Network) 모델, 비전 트랜스포머(ViT: Vision Transformer) 모델 및 오토인코더(Auto-encoder) 모델 중 어느 하나로 학습하여, 상기 분석모델을 생성하는 학습부를 포함하는 것을 특징으로 하는 음성신호 기반의 연하장애검출장치.

청구항 16

제13항에 있어서, 상기 모델생성부는

상기 음성신호에 대한 전처리를 수행하여, 1차원 특징정보 및 2차원 이미지를 생성하는 전처리부; 및

상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리 모델로 학습하는 제1 모델, 상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를 합성곱 신경망 모델, 비전 트랜

스포머 모델 및 오토인코더 모델 중 적어도 어느 하나로 학습하는 제2 모델, 및 상기 제1 모델로부터 추출한 제1 특징벡터와 상기 제2 모델로부터 추출한 제2 특징벡터를 결합한 결합벡터에 적용하는 분류기 함수를 생성한 후, 상기 제1 모델, 제2 모델 및 분류기 함수를 학습하여, 상기 분석모델을 생성하는 학습부를 포함하는 것을 특징으로 하는 음성신호 기반의 연하장애검출장치.

청구항 17

제16항에 있어서, 상기 학습부는

상기 정상인과 연하장애환자의 인적정보를 수집하고, 상기 인적정보로부터 생성한 제3 특징벡터를, 상기 제1 특징벡터 및 제2 특징벡터와 더 결합하여 상기 결합벡터를 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출장치.

청구항 18

제16항에 있어서, 상기 학습부는

상기 2차원 이미지가 상기 오토인코더 모델에 입력되면, 상기 오토인코더 모델에 의해 생성된 복원 2차원 이미지를, 상기 2차원 이미지와 채널 방향으로 결합하여 결합 2차원 이미지를 생성하고,

상기 결합 2차원 이미지가 상기 합성곱 신경망 모델에 입력되면, 상기 결합 2차원 이미지에 대응하는 특징맵을 생성하며,

상기 특징맵을 풀링하여 복수의 비전 트랜스포머 모델에 각각 풀링된 특징맵을 입력하면, 각각의 비전 트랜스포머 모델에서 생성한 비전 트랜스포머 특징벡터들을 결합하여 제2 특징벡터를 생성하도록, 상기 제2 모델을 생성하는 것을 특징으로 하는 음성신호 기반의 연하장애검출장치.

청구항 19

제13항에 있어서,

사용자 단말로부터 검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신하고, 상기 판단결과를 상기 사용자 단말로 전송하는 통신부를 더 포함하는 것을 특징으로 하는 음성신호 기반의 연하장애검출장치.

청구항 20

제13항에 있어서, 상기 음성수신부는

상기 연하장애검출장치에 포함된 마이크로폰으로부터 상기 대상음성신호를 수신하는 것을 특징으로 하는 음성신호 기반의 연하장애검출장치.

발명의 설명

기술 분야

[0001] 본 발명은 음식물 섭취 전후의 음성을 활용하여 연하장애를 진단할 수 있는 음성신호 기반의 연하장애검출장치 및 연하장애검출방법에 관한 것이다.

배경 기술

[0003] 연하(swallowing)는 음식물을 입을 통해 목구멍으로 넘기는 행위이며, 말초 및 중추 신경계의 체계적인 역할을 기반으로 감각 및 운동기관들이 유기적으로 움직여야 그 기능이 원활하게 동작한다. 연하 과정은 크게 구강 준

비, 구강 운반, 인-후두, 식도 단계, 총 4 단계로 구분된다. 만일 각 단계와 관련된 기능과 구조 등에 문제가 생기거나, 각 단계 사이의 매끄러운 협응이 불가하면 연하장애(dysphagia)가 발생하게 된다. 이것으로 인하여 섭취하는 음식의 질이 떨어지고 종류가 제한되어 영양실조 혹은 탈수증 등과 같은 질환이 유발될 수 있다. 또한 연하장애를 늦게 인지하여 조기에 치료하지 못하면 흡인성 폐렴 등의 합병증으로 사망에까지 이를 수 있다. 이러한 연하장애는 뇌졸중 환자에게 많이 발병되는 질환이지만 노화의 진행에 따라 근육 능력의 저하로 인하여 삼킴 조절 능력이 떨어져 관련 기저질환이 없는 노인들도 발생할 가능성이 높아지고 있다.

[0004] 연하장애를 진단하기 위한 검사로는 VFSS(Video Fluoroscopic Swallowing Study)가 대표적인데, 이는 방사선을 이용하여 다양한 식이(요거트, 죽, 물, 밥)에 조영제(황산바륨 등)를 일정 비율로 섞어 환자에게 순차적으로 제공하면서 환자가 삼키는 동안 검사자가 실시간으로 촬영되는 모니터를 보며 침투 흡인 척도(Penetration aspiration scale, PAS) 등급을 매기는 검사이다. 하지만 VFSS는 환자와 직원 모두에게 생물학적 위험(X-선 노출)이 있는 침습적 진단검사 방법이다. 또한 검사자가 영상을 보고 직접 판단하므로, 각각의 검사자의 전문성에 대한 의존도도 존재한다.

[0005] 최근 이러한 문제를 해결하기 위해 목넘김 음향, 인두부의 진동 등의 신호를 기반으로 연하장애 여부를 검출하는 연구가 진행되었다. 하지만 목넘김 음향과, 인두부의 진동 신호 기반의 연구의 경우 목에 밀착되는 마이크와 가속도 센서가 결합된 별도의 기구가 필요하며, 밀착 정도와 부착된 기구의 위치에 따른 변수가 많아 실제 환경에 적용하기에 어려운 문제점을 가지고 있다.

선행기술문헌

특허문헌

[0007] (특허문헌 0001) 대한민국 공개특허공보 제10-2014-0134443호

발명의 내용

해결하려는 과제

[0008] 본 발명은, 음식물 섭취 전후의 목소리 차이를 인공지능 모델로 분석하여, 연하장애를 진단할 수 있는 음성신호 기반의 연하장애검출장치 및 연하장애검출방법을 제공하기 위한 것이다.

[0009] 본 발명은, 가정 등에서 검사대상자의 목소리만으로 연하장애를 검사할 수 있도록 하여, 연하장애를 조기진단할 수 있는 음성신호 기반의 연하장애검출장치 및 연하장애검출방법을 제공하기 위한 것이다.

[0010] 본 발명은, 방사능 노출이 있는 VFSS 검사와 달리 인체에 무해한 검사방식으로 연하장애를 진단할 수 있는 음성신호 기반의 연하장애검출장치 및 연하장애검출방법을 제공하기 위한 것이다.

과제의 해결 수단

[0012] 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출방법은, 연하장애검출장치를 이용하는 것으로, 검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신하는 단계; 및 상기 대상음성신호를 분석모델에 입력하여, 상기 검사대상자에 대한 연하장애를 판정하여 판단결과를 도출하는 단계를 포함하며, 상기 분석모델은 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와, 상기 복수의 정상인 및 연하장애환자에 대해 수행한 비디오 투시 연하검사(VFSS: Video Fluoroscopic swallowing study)의 검사결과를 인공지능 모델로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 것일 수 있다.

[0013] 여기서 상기 분석모델은 상기 음성신호에 대한 전처리를 수행하여 1차원 특징정보를 생성하고, 상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리(LSTM: Long Short Term Memory) 모델로 학습하여 생성하는 것일 수 있다.

[0014] 여기서 상기 1차원 특징정보는, 상기 음성신호의 기본주파수, 지터(jitter), 시머(shimmer) 및 배음 대 소음비(HNR: Harmonics to Noise Ratio) 중 적어도 어느 하나를 포함할 수 있다.

- [0015] 여기서 상기 분석모델은 상기 음성신호에 대한 전처리를 수행하여 2차원 이미지를 생성하고, 상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를, 합성곱 신경망(CNN: Convolutional Neural Network) 모델, 비전 트랜스포머(ViT: Vision Transformer) 모델 및 오토인코더(Auto-encoder) 모델 중 어느 하나로 학습하여 생성하는 것일 수 있다.
- [0016] 여기서 상기 2차원 이미지는, 상기 음성신호에 단시간 푸리에 변환(STFT: Short-Time Frequency Transform) 또는 멜 주파수 cepstrum 계수(MFCC: Mel-Frequency Cepstral Coefficient)를 적용하여 생성할 수 있다.
- [0017] 여기서 상기 분석모델은 상기 음성신호에 대한 전처리를 수행하여, 1차원 특징정보 및 2차원 이미지를 생성하고, 상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리 모델로 학습하는 제1 모델, 상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를 합성곱 신경망 모델, 비전 트랜스포머 모델 및 오토인코더 모델 중 적어도 어느 하나로 학습하는 제2 모델 및 상기 제1 모델로부터 추출한 제1 특징벡터와 상기 제2 모델로부터 추출한 제2 특징벡터를 결합한 결합벡터에 적용하는 분류기(classifier) 함수를 생성한 후, 상기 제1 모델, 제2 모델 및 분류기 함수를 학습하여 생성하는 것일 수 있다.
- [0018] 여기서, 상기 분석모델은 상기 정상인과 연하장애환자의 인적정보를 수집하고, 상기 인적정보로부터 생성한 제3 특징벡터를, 상기 제1 특징벡터 및 제2 특징벡터와 더 결합하여 상기 결합벡터를 생성하는 것으로, 상기 인적정보는, 상기 연하장애환자의 성별, 나이 및 기저질환 중 적어도 어느 하나를 포함하고, 상기 제3 특징 벡터는, 상기 기저질환을 나타내는 원-핫 벡터(1-hot vector)에 상기 성별 또는 나이를 나타내는 정보를 추가하여 생성하는 것일 수 있다.
- [0019] 여기서 상기 제2 모델은, 상기 2차원 이미지가 상기 오토인코더 모델에 입력되면, 상기 오토인코더 모델에 의해 생성된 복원 2차원 이미지를, 상기 2차원 이미지와 채널 방향으로 결합(concatenation)하여 결합 2차원 이미지를 생성하고, 상기 결합 2차원 이미지가 상기 합성곱신경망 모델에 입력되면, 상기 결합 2차원 이미지에 대응하는 특징맵을 생성하며, 상기 특징맵을 풀링(pooling)하여 복수의 비전 트랜스포머 모델에 각각 풀링된 특징맵을 입력하면, 각각의 비전 트랜스포머 모델에서 생성한 비전 트랜스포머 특징벡터들을 결합(concatenation)하여 제2 특징벡터를 생성하도록, 생성한 것일 수 있다.
- [0020] 여기서 상기 판단결과를 도출하는 단계는, 상기 검사대상자의 연하장애를 정상 또는 비정상으로 판정하여 상기 판단결과를 도출할 수 있다.
- [0021] 여기서 상기 판단결과를 도출하는 단계는, 상기 검사대상자의 연하장애를 침습, 흡인 및 정상 중 어느 하나로 판정하거나, 상기 연하장애를 침습-흡인 척도(PAS: Penetration-aspiration score)에 따른 복수의 단계 중 어느 하나로 판정하여 상기 판단결과를 도출할 수 있다.
- [0022] 본 발명의 일 실시예에 의하면, 하드웨어와 결합하여, 상술한 음성신호 기반의 연하장애검출방법을 수행하기 위하여 매체에 저장된 컴퓨터 프로그램이 존재할 수 있다.
- [0023] 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출장치는, 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와, 상기 복수의 정상인 및 연하장애환자에 대해 수행한 비디오 투시 연하검사의 검사결과를 인공지능모델로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성하는 모델생성부; 검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신하는 음성수신부; 및 상기 대상음성신호를 분석모델에 입력하여, 상기 검사대상자에 대한 연하장애를 판정하여 판단결과를 도출하는 판정부를 포함할 수 있다.
- [0024] 여기서 상기 모델생성부는, 상기 음성신호에 대한 전처리를 수행하여, 1차원 특징정보를 생성하는 전처리부; 및 상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리 모델로 학습하여, 상기 분석모델을 생성하는 학습부를 포함할 수 있다.
- [0025] 여기서 상기 모델생성부는, 상기 음성신호에 대한 전처리를 수행하여, 2차원 이미지를 생성하는 전처리부; 및 상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를, 합성곱 신경망 모델, 비전 트랜스포머 모델 및 오토인코더 모델 중 어느 하나로 학습하여, 상기 분석모델을 생성하는 학습부를 포함할 수 있다.
- [0026] 여기서 상기 모델생성부는, 상기 음성신호에 대한 전처리를 수행하여, 1차원 특징정보 및 2차원 이미지를 생성하는 전처리부; 및 상기 1차원 특징정보에 대응하는 상기 비디오 투시 연하검사의 검사결과를 장단기 메모리 모델로 학습하는 제1 모델, 상기 2차원 이미지에 대응하는 상기 비디오 투시 연하검사의 검사결과를 합성곱 신경망 모델, 비전 트랜스포머 모델 및 오토인코더 모델 중 적어도 어느 하나로 학습하는 제2 모델 및 상기 제1 모

델로부터 추출한 제1 특징벡터와 상기 제2 모델로부터 추출한 제2 특징벡터를 결합한 결합벡터에 적용하는 분류기 함수를 생성한 후, 상기 제1 모델, 제2 모델 및 분류기 함수를 학습하여 상기 분석모델을 생성하는 학습부를 포함할 수 있다.

[0027] 여기서 상기 학습부는 상기 정상인과 연하장애환자의 인적정보를 수집하고, 상기 인적정보로부터 생성한 제3 특징벡터를, 상기 제1 특징벡터 및 제2 특징벡터와 더 결합하여 상기 결합벡터를 생성하는 것일 수 있다.

[0028] 여기서 상기 학습부는, 상기 2차원 이미지가 상기 오토인코더 모델에 입력되면, 상기 오토인코더 모델에 의해 생성된 복원 2차원 이미지를, 상기 2차원 이미지와 채널 방향으로 결합하여 결합 2차원 이미지를 생성하고, 상기 결합 2차원 이미지가 상기 합성곱 신경망 모델에 입력되면, 상기 결합 2차원 이미지에 대응하는 특징맵을 생성하며, 상기 특징맵을 풀링하여 복수의 비전 트랜스포머 모델에 각각 풀링된 특징맵을 입력하면, 각각의 비전 트랜스포머 모델에서 생성한 비전 트랜스포머 특징벡터들을 결합하여 제2 특징벡터를 생성하도록, 상기 제2 모델을 생성할 수 있다.

[0029] 여기서, 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출장치는, 사용자 단말로부터 검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신하고, 상기 판단결과를 상기 사용자 단말로 전송하는 통신부를 더 포함할 수 있다.

[0030] 여기서 상기 음성수신부는, 상기 연하장애검출장치에 포함된 마이크로폰으로부터 상기 대상음성신호를 수신할 수 있다.

[0031] 덧붙여 상기한 과제에의 해결수단은, 본 발명의 특징을 모두 열거한 것이 아니다. 본 발명의 다양한 특징과 그에 따른 장점과 효과는 아래의 구체적인 실시형태를 참조하여 보다 상세하게 이해될 수 있을 것이다.

발명의 효과

[0033] 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출장치 및 연하장애검출방법에 의하면, 음식물 섭취 전후의 목소리 차이를 인공지능 모델로 분석하여, 연하장애를 진단할 수 있다.

[0034] 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출장치 및 연하장애검출방법에 의하면, 가정 등에서 검사대상자의 목소리만으로 연하장애를 검사할 수 있으므로, 연하장애를 조기진단할 수 있다.

[0035] 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출장치 및 연하장애검출방법에 의하면, 방사능 노출이 있는 VFSS 검사와 달리 인체에 무해한 검사방식으로 연하장애를 진단할 수 있다.

[0036] 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출장치 및 연하장애검출방법에 의하면, 복수의 인공지능 모델들을 결합한 멀티-모달 모델을 이용하여 연하장애를 진단할 수 있으므로, 보다 정확한 연하장애 진단을 수행하는 것이 가능하다.

도면의 간단한 설명

[0038] 도1은 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출시스템을 나타내는 개략도이다.

도2는 본 발명의 일 실시예에 의한 연하장애검출장치를 나타내는 블록도이다.

도3은 본 발명의 일 실시예에 의한 연하장애 질환정도를 나타내는 PAS(Panetration-Aspiration score)를 나타내는 표이다.

도4는 본 발명의 일 실시예에 의한 연하장애검출장치의 모델생성부를 나타내는 블록도이다.

도5는 본 발명의 일 실시예에 의한 LSTM 모델을 나타내는 예시도이다.

도6은 본 발명의 일 실시예에 의한 CNN 모델을 나타내는 예시도이다.

도7은 본 발명의 일 실시예에 의한 ViT 모델을 나타내는 예시도이다.

도8 및 도9는 본 발명의 일 실시예에 의한 오토인코더 모델을 나타내는 예시도이다.

도10은 본 발명의 일 실시예에 의한 멀티 모달 모델을 나타내는 개략도이다.

도11은 본 발명의 일 실시예에 의한 인적정보를 나타내는 표이다.

도12는 본 발명의 다른 실시예에 의한 음성신호 기반의 연하장애검출장치를 나타내는 개략도이다.

도13은 본 발명의 다른 실시예에 의한 연하장애검출장치를 나타내는 블록도이다.

도14는 본 발명의 일 실시예에 의한 음성기반의 연하장애검출방법을 나타내는 순서도이다.

도15 내지 도17은 본 발명의 일 실시예에 의한 분석모델의 생성방법을 나타내는 순서도이다.

발명을 실시하기 위한 구체적인 내용

- [0039] 이하, 첨부된 도면을 참조하여 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자가 본 발명을 용이하게 실시할 수 있도록 바람직한 실시예를 상세히 설명한다. 다만, 본 발명의 바람직한 실시예를 상세하게 설명함에 있어, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략한다. 또한, 유사한 기능 및 작용을 하는 부분에 대해서는 도면 전체에 걸쳐 동일한 부호를 사용한다.
- [0040] 덧붙여, 명세서 전체에서, 어떤 부분이 다른 부분과 '연결'되어 있다고 할 때, 이는 '직접적으로 연결'되어 있는 경우뿐만 아니라, 그 중간에 다른 소자를 사이에 두고 '간접적으로 연결'되어 있는 경우도 포함한다. 또한, 어떤 구성요소를 '포함'한다는 것은, 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있다는 것을 의미한다. 또한, 명세서에 기재된 "~부", "모듈" 등의 용어는 적어도 하나의 기능이나 동작을 처리하는 단위를 의미하며, 이는 하드웨어 또는 소프트웨어로 구현되거나 하드웨어와 소프트웨어의 결합으로 구현될 수 있다.
- [0042] 도1은 본 발명의 일 실시예에 의한 음성신호 기반의 연하장애검출시스템을 나타내는 개략도이다.
- [0043] 도1을 참조하면, 본 발명의 일 실시예에 의한 연하장애검출시스템은 사용자 단말(10) 및 연하장애검출장치(100)를 포함할 수 있다.
- [0044] 이하 도1을 참조하여 본 발명의 일 실시예에 의한 연하장애검출시스템을 설명한다.
- [0046] 사용자 단말(10)은 검사대상자(1)의 음성신호를 녹음할 수 있으며, 녹음한 음성신호를 통신망을 통하여 연하장애검출장치(100)로 전송할 수 있다.
- [0047] 검사대상자(1)는 연하(Swallowing) 장애 여부를 진단하고자 하는 사람일 수 있으며, 검사대상자(1)는 음식물 섭취를 전후하여, 사용자 단말(10) 앞에서 미리 정해진 음절이나 단어, 문장 등을 발화하도록 요청받을 수 있다. 예를들어, 검사대상자(1)는 /a/를 발음하거나, "시계", "안녕하세요" 등을 발음하도록 요청받을 수 있다. 검사대상자(1)가 미리 정해진 음절이나 단어 등을 발화하면, 사용자 단말(10)은 구비된 마이크로폰으로 검사대상자(1)의 음성을 수집하여 음성신호를 생성할 수 있다.
- [0048] 이후, 사용자 단말(10)은 통신망을 통해 음성신호를 연하장애검출장치(100)로 전송할 수 있으며, 이때 통신망은 유선 통신 네트워크, 근거리 또는 원거리 무선 통신 네트워크, 이들이 혼합된 네트워크 등 데이터를 전달하는 모든 형태의 통신 네트워크를 포함할 수 있다.
- [0049] 한편, 사용자 단말(10)은 메모리(memory)와 프로세서(processor)를 구비하여 연산처리 능력을 갖춘 장치로, 스마트폰, 태블릿 PC, PDA, 디지털 카메라, 스마트워치, 디지털/아날로그 녹음기 등일 수 있다. 실시예에 따라서는 녹음기능과 통신기능을 포함하는 연하장애검출을 위한 전용의 단말장치일 수 있다.
- [0050] 사용자 단말(10)에는 연하장애를 검출하기 위한 전용의 어플리케이션 등이 설치될 수 있으며, 어플리케이션 등을 통하여 연하장애검출장치(100)와 통신을 수행할 수 있다. 검사대상자(1)는 사용자 단말(10)에 설치된 전용의 어플리케이션을 이용하여 음성을 녹음할 수 있으며, 전용의 어플리케이션을 통하여 연하장애검출장치(100)로부터 수신한 연하장애 판단결과를 제공받을 수 있다.
- [0051] 연하장애검출장치(100)는 사용자 단말(10)로부터 검사대상자(1)의 음성신호를 수신할 수 있으며, 음성신호를 이용하여 검사대상자(1)의 연하장애를 판정할 수 있다. 실시예에 따라서는, 연하장애검출장치(100)를 클라우드 서버 내에 구축할 수 있으며, 사용자 단말(10) 내 설치된 전용의 어플리케이션을 통하여 사용자 단말(10)과의 통

신을 수행할 수 있다.

- [0052] 연하장애는 음식물이 구강에서 식도로 넘어가는 과정에서 식도가 아닌 기도로 넘어가는 증상으로, 종래에는 연하장애를 판정하기 위하여 비디오 투시 연하검사(VFSS: Video Fluoroscopic swallowing study)를 활용하여 진단하였다. 그러나, VFSS 검사는 검사식이에 조영제를 일정 비율로 섞어 환자에게 제공한 후, 검사대상자가 검사식이를 삼키는 동안 실시간으로 방사선을 통해 촬영되는 모니터를 보면서 연하장애를 판정하는 것으로, 검사대상자나 검사자 모두 방사선 노출에 대한 위험이 존재하고, 검사자의 개별 전문성에 의한 의존도가 높으며, 검사를 위한 전문 장비가 필요하다는 점에서 제약이 존재한다.
- [0053] 이로 인하여, 최근에는 VFSS 검사를 대신하여 물 삼키기 검사(water swallowing test) 등을 통해 연하장애 평가를 진행하기도 하지만, 민감도 및 특이도 측면에서 VFSS 검사에 비해 정확도가 떨어지는 단점이 존재한다.
- [0054] 반면에, 본 발명의 일 실시예에 의한 연하장애검출장치(100)는 검사대상자(1)의 음성신호를 인공지능 모델을 이용하여 분석함으로써, 연하장애 여부를 검출하는 것으로, 검사대상자(1)의 음성신호만으로 신속하고 정확하게 연하장애를 검출하는 것이 가능하다. 즉, 음식물이 기도로 넘어가게 되면, 기도 안에 들어간 음식물 또는 기도 내 습도의 변화 등으로 인하여, 목소리의 주파수적 특성의 변화를 초래하게 된다. 따라서, 연하장애검출장치(100)는 검사대상자(1)의 음식물 섭취 전후 음성신호의 변화를 이용하여 검사대상자(1)의 연하장애 여부를 판별할 수 있다. 이하, 도2를 참조하여 본 발명의 일 실시예에 의한 연하장애검출장치(100)를 설명한다.
- [0056] 도2는 본 발명의 일 실시예에 의한 연하장애검출장치(100)를 나타내는 블록도이다.
- [0057] 도2를 참조하면 본 발명의 일 실시예에 의한 연하장애검출장치(100)는 모델생성부(110), 음성수신부(120), 판정부(130) 및 통신부(140)를 포함할 수 있다.
- [0058] 모델생성부(110)는 음성신호로부터 연하장애여부를 판정하기 위한 인공지능 모델 기반의 분석모델을 생성할 수 있다. 즉, 복수의 정상인과 연하장애환자로부터 음식물 섭취 전후의 음성신호를 수집하고, 동일한 정상인과 연하장애환자들에 대한 VFSS 검사결과를 수집한 후, 이를 학습데이터로 이용하여 분석모델을 학습시킬 수 있다. 따라서, 모델생성부(110)에서 학습한 분석모델을 이용하면, 검사대상자(1)의 대상음성신호에 대응하는 연하장애를 판정하는 것이 가능하다. 모델생성부(110)의 구체적인 동작은 이후 도4 내지 도11을 참조하여 설명한다.
- [0059] 음성수신부(120)는 검사대상자(1)의 음식물 섭취 전후에 녹음한 대상음성신호를 수신할 수 있다. 음식물은 요거트, 물, 죽, 밥 등 다양한 종류로 제공될 수 있으며, 검사대상자(1)는 음식물의 섭취 전에 특정 음을 발성하여 녹음하고, 음식물 섭취 후에도 동일한 특정 음을 발성하여 녹음할 수 있다. 여기서, 특정 음은 /a/와 같은 단모음일 수 있으나 이에 한정되는 것은 아니고, 자음, 단어, 간단한 문장 등을 포함할 수 있다. 음성수신부(120)는 통신부(140)를 통하여, 사용자 단말(10)로부터 전송되는 대상음성신호를 수신받을 수 있다.
- [0060] 판정부(130)는 검사대상자(1)의 대상음성신호를 분석모델에 입력하여, 검사대상자(1)에 대한 연하장애를 판정하여 판단결과를 도출할 수 있다. 인공지능 모델 기반의 분석모델은 입력받은 대상음성신호에 대응하는 판단결과를 제공할 수 있으며, 판단결과에 따라 검사대상자(1)의 연하장애를 진단할 수 있다.
- [0061] 여기서, 분석모델에 이진 분류기가 구비되는 경우에는, 판정부(130)가 검사대상자(1)의 연하장애를 정상 또는 비정상의 둘 중 하나로 판정하여 판단결과를 도출할 수 있다. 즉, 검사대상자에게 침습이나 흡인 등의 연하장애가 존재하는 경우에는 비정상으로 판정하고, 연하장애가 존재하지 않는 경우는 정상으로 판정할 수 있다.
- [0062] 또한, 실시예에 따라서는 분석모델에 다중 분류기가 구비될 수 있으며, 이 경우, 판정부(130)는 연하장애의 정도에 따라 2 이상으로 판정할 수 있다. 예를들어, 판정부(130)는 도3에 도시한 바와 같이, 검사대상자의 연하장애를 침습, 흡인 및 정상 중 어느 하나로 판정하거나, 연하장애를 침습-흡인 척도(PAS: Penetration-aspiration score)에 따른 8단계 중 어느 하나로 판정하여 판단결과를 도출할 수 있다. 여기서, PAS는 VFSS 검사의 결과로 기도로 음식이 들어가는지를 검사자가 확인한 후, 연하장애 질환의 정도를 8단계로 표현한 것이다. 1단계는 정상, 2 내지 5단계는 침습, 6 내지 8단계는 흡인으로 구별할 수 있으며, 정상은 음식이 기도로 들어가지 않는 경우, 침습은 음식이 기도로 들어가나 성대 아래를 지나지 않는 경우, 흡인은 음식이 기도로 들어가서 성대 아래까지 지나는 경우에 해당한다.
- [0063] 통신부(140)는 사용자 단말(10)과 통신망을 통하여 통신을 수행할 수 있다. 통신부(140)는 사용자 단말(10)로부터 검사대상자(1)의 음식물 섭취 전후에 녹음한 대상음성신호를 수신할 수 있으며, 판정부(130)에서의 판단결과

를 사용자 단말(10)로 전송할 수 있다.

- [0065] 한편, 도4를 참조하면 본 발명의 일 실시예에 의한 모델생성부(110)는 수집부(111), 전처리부(112) 및 학습부(113)를 포함할 수 있다.
- [0066] 수집부(111)는 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와 VFSS 검사결과를 수집할 수 있다. 여기서, 수집부(111)는 각각의 음성신호와 그에 대응되는 VFSS 검사결과가 매칭되도록 수집할 수 있다. 즉, 각각의 음성신호에 정상/비정상, 정상/침습/흡입, PAS의 8단계와 같은 VFSS 검사결과가 매칭되도록 수집할 수 있다. 이 경우, 각각의 음성신호에 대하여 분석모델을 통하여 판정하고자 하는 연하장애의 판단결과에 대한 정답이 레이블링된 것이므로, 이들을 분석모델 학습을 위한 학습데이터로 활용할 수 있다.
- [0067] 전처리부(112)는 음성신호에 대한 전처리를 수행하여, 음성신호에 대한 1차원 특징정보 또는 2차원 이미지를 생성할 수 있다. 즉, 전처리부(112)는 음성신호로부터 특성을 추출하기 위한 전처리를 수행할 수 있으며, 분석모델에 적용할 인공지능 모델의 종류에 따라, 1차원 특징정보 또는 2차원 이미지를 생성할 수 있다.
- [0068] 1차원 특징정보에는 음성신호로부터 추출하는 기본주파수(fundamental frequency), 지터(jitter), 시머(shimmer), 배음 대 소음비(HNR: Harmonics to Noise Ratio) 등이 포함될 수 있다. 기본주파수는 성대 진동의 주파수로, 지각적으로는 음높이(pitch)에 해당한다. 지터는 음성신호 내 진동의 주기가 얼마나 일정한지를 나타내는 주파수 변동률이고, 시머는 연속 주기의 진폭을 평균 진폭으로 나눈 것으로 음성신호 내 진동의 진폭이 얼마나 일정한지를 나타내는 진폭 변동률에 해당한다. HNR은 70~4,500Hz 사이에 존재하는 배음과 1,500~4,500Hz 사이에 존재하는 비정상 배음 간의 비율 평균치이며, 그 값이 클수록 소음의 비율이 높음을 의미한다.
- [0069] 2차원 이미지는 음성신호를 시각화한 스펙트로그램(Spectrogram)으로, 2차원 이미지는 시간에 따른 음성신호의 주파수, 진폭 등의 특성변화를 나타내는 것일 수 있다. 실시예에 따라서는, 음성신호에 단시간 푸리에 변환(STFT: Short-Time Frequency Transform) 또는 멜 주파수 켈스트럼 계수(MFCC: Mel-Frequency Cepstral Coefficient)를 적용하여 2차원 이미지를 생성할 수 있다.
- [0070] 학습부(113)는 전처리부(112)로부터 수신한 1차원 특징정보 또는 2차원 이미지와, VFSS 결과를 이용하여 인공지능 모델을 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성할 수 있다. 여기서, 인공지능 모델은 장단기 메모리(LSTM: Long Short-Term Memory) 모델, 합성곱 신경망(CNN: Convolutional Neural Network) 모델, 비전 트랜스포머(ViT: Vision Transformer) 모델, 오토인코더(Auto-encoder) 모델 등일 수 있으며, 이외에도, 의사결정트리(Decision Tree) 모델, 로지스틱 회귀 분석(Logistic Regression) 모델, 서포트 벡터 머신(Support Vector Machine) 모델, 랜덤 포레스트(Random Forest) 모델, RNN(Recurrent Neural Network) 모델, GAN(Generative Adversarial Network) 모델 등 다양한 모델을 적용할 수 있다.
- [0071] 일 실시예에 의하면, 학습부(113)는 1차원 특징정보와, 1차원 특징정보에 대응하는 VFSS 검사결과를 LSTM 모델로 학습하여, 분석모델을 생성할 수 있다. 즉, 학습데이터에는, 1차원 특징정보인 기본주파수, 지터, 시머, HNR 등에 따른 각각의 정답이 레이블링되어 있으므로, 학습데이터로 LSTM 모델을 학습시켜, LSTM 모델 기반의 분석모델을 생성할 수 있다. 이 경우, 분석모델은 수신한 대상음성신호의 1차원 특징정보로부터, 대응하는 연하장애를 판정하는 것이 가능하다. 실시예에 따라서는, 도5와 같이 LSTM 모델을 구현할 수 있다.
- [0072] 다른 실시예에 의하면, 학습부(113)는 2차원 이미지와, 2차원 이미지에 대응하는 VFSS 검사결과를 CNN 모델, ViT 모델 및 오토인코더 모델 중 어느 하나로 학습하여, 분석모델을 생성할 수 있다.
- [0073] CNN 모델을 이용하는 경우, 학습데이터에는 각각의 2차원 이미지에 대응하는 정답이 레이블링되어 있으므로, 학습데이터로 CNN 모델을 학습시켜, CNN 모델 기반의 분석모델을 생성할 수 있다. 도6에 도시한 바와 같이, CNN 모델은 입력된 이미지에 컨볼루션(convolution) 연산 및 서브샘플링을 반복하여 해당 이미지에 대한 특징맵을 추출할 수 있으며, 특징맵을 기반으로 이미지를 분류할 수 있다. 즉, CNN 모델을 이용하는 분석모델의 경우, 수신한 대상음성신호의 2차원 이미지로부터, 대응하는 연하장애를 판정할 수 있다.
- [0074] 또한, ViT 모델을 이용하는 경우에도, 학습데이터에는 각각의 2차원 이미지에 대응하는 정답이 레이블링되어 있으므로, 학습데이터로 ViT 모델을 학습시켜 ViT 모델 기반의 분석모델을 생성할 수 있다. 이 경우, 분석모델은 수신한 대상음성신호의 2차원 이미지로부터 대응하는 연하장애를 판정할 수 있으며, ViT 모델은 도7에 도시한 바와 같이 구현가능하다.
- [0075] 한편, 오토인코더 모델은, 도8에 도시한 바와 같이, 입력데이터를 압축한 후, 압축데이터로부터 입력데이터를

다시 복원하는 신경망으로, 입력데이터를 압축하는 인코더와, 압축데이터로부터 입력데이터를 복원하는 디코더를 포함할 수 있다. 여기서, 오토인코더 모델은 출력데이터를 입력데이터와 동일하게 복원하도록 설계될 수 있으며, 특히 학습부(113)는 오토인코더 모델을 VFSS 검사결과 "정상"에 해당하는 2차원 이미지로만 학습하여 분석모델을 생성할 수 있다. 이 경우, "정상"에 해당하는 2차원 이미지의 입력시에는 동일하게 복원될 수 있으나, VFSS 검사결과 침습이나 흡인 등 "비정상"에 해당하는 2차원 이미지들이 입력되는 경우에는 상이하게 복원될 수 있다. 즉, 오토인코더 모델은 "정상"에 해당하는 2차원 이미지만을 정확하게 복원하도록 학습되어 있으므로, "비정상"에 해당하는 2차원 이미지들은 상대적으로 부정확하게 복원될 수 있다. 따라서, 분석모델에 입력된 2차원 이미지들을 압축 및 복원한 후, 입력된 2차원 이미지와 복원된 2차원 이미지의 차이를 통하여 "비정상"에 해당하는 2차원 이미지들을 구별할 수 있다. 구체적으로, 도9에 도시한 바와 같이, 입력된 2차원 이미지와 복원된 2차원 이미지 사이의 차이를 이용하여 이상 점수(anomaly score)를 구할 수 있으며, 이상 점수가 문턱값보다 크면 해당 2차원 이미지는 "비정상"으로 판별할 수 있다.

[0076] 추가적으로, 실시예에 따라서는, 학습부(113)가 복수의 모델들을 결합한 멀티-모달 모델을 이용하여 분석모델을 생성하는 것도 가능하다. 즉, 연하장애판정의 정확도를 높이기 위하여, 복수의 인공지능 모델들을 결합한 멀티-모달 모델을 분석모델로 구현할 수 있다.

[0077] 도10을 참조하면, 학습부(113)는 1차원 특징정보(F1)와, 1차원 특징정보(F1)에 대응하는 VFSS 검사결과를 LSTM 모델(M1)로 학습하는 제1 모델을 생성할 수 있다.

[0078] 또한, 학습부(113)는 CNN 모델(M2), ViT 모델(M3) 및 오토인코더 모델(M1)을 결합하여, 제2 모델을 생성할 수 있으며, 도10에 도시한 바와 같이 오토인코더 모델(M2), CNN 모델(M3) 및 ViT 모델(M3)을 순차적으로 연결하여 제2 모델을 생성할 수 있다. 이 경우, 오토인코더 모델(M2)은 2차원 이미지(F2)를 복원한 복원 2차원 이미지를 생성하고, 복원 2차원 이미지와 원본 2차원 이미지를 채널 방향으로 결합(concatenation)하여 결합 2차원 이미지를 생성할 수 있다. 이때 오토인코더 모델(M2)은 미리 학습(pre-trained)되어 있을 수 있다.

[0079] 이후, 결합 2차원 이미지는 CNN 모델(M3)에 입력될 수 있으며, CNN 모델(M3)은 입력된 결합 2차원 이미지에 대응하는 특징맵을 생성할 수 있다. 특징맵은 풀링(pooling)을 통하여 복수의 ViT 모델(M4)에 각각 입력될 수 있으며, 각각의 ViT 모델(M4)에서 생성한 비전 트랜스포머 특징벡터(이하, ViT 특징벡터)들을 결합하여 제2 특징벡터를 생성하도록, 제2 모델을 구현할 수 있다.

[0080] 한편, 인적정보(F3)에는, 도11에 도시한 바와 같이 연하장애환자의 성별, 나이 및 기저질환 등을 포함될 수 있으며, 학습부(113)는 인적정보(F3)를 기반으로 제3 특징벡터를 생성할 수 있다. 즉, 기저질환을 원-핫 벡터(1-hot vector) 형태로 나타내고, 원-핫 벡터에 나머지 성별이나 나이를 나타내는 정보 등을 추가하여 제3 특징벡터를 생성할 수 있다. 예를들어, 도11에 나타난 바와 같이, 기저질환으로 뇌졸중과 고혈압이 존재하나 성대 결정은 존재하지 않는 경우, 원-핫 벡터로 [1 0 1]을 생성할 수 있다. 또한, 원-핫 벡터에 남성을 나타내는 0과 나이 56세를 나타내는 0.28을 더 포함하여, 제3 특징벡터로 [0 0.28 1 0 1]을 생성할 수 있다. 이 경우, 분석모델 내에 연하장애환자의 성별, 나이, 기저질환 등에 관한 정보가 반영되므로, 보다 정확하게 연하장애를 판정하는 것이 가능하다.

[0081] 도10을 참조하면, 멀티-모달(Multi-modal) 모델은 제1 모델로부터 추출한 제1 특징벡터, 제2 모델로부터 추출한 제2 특징벡터 및 인적정보로부터 생성한 제3 특징벡터를 결합(concatenation)한 결합벡터에 적용하는 분류기 함수(C)를 포함할 수 있다. 분류기 함수(C)에는 완전연결(fully connected) 블록과 시그모이드(sigmoid) 블록이 포함될 수 있으며, 분류기 함수(C)를 통하여 결합벡터로부터 최종적으로 연하장애를 판정할 수 있다. 다만, 도10에는 시그모이드 블록을 포함하여 연하장애를 정상과 비정상의 이진분류하는 것으로 나타나 있으나, 실시예에 따라서는 시그모이드 블록을 대신하여 Softmax 블록 등 다중 분류기를 포함하여 정상/침습/흡인으로 분류하거나 PAS의 8단계로 분류하도록 하는 것도 가능하다.

[0082] 즉, 학습부(113)는 제1 모델, 제2 모델 및 분류기 함수를 포함하는 분석모델을 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성할 수 있다.

[0083] 추가적으로, 도10에서는 오토인코더 모델(M1), CNN 모델(M2), ViT 모델(M3)을 모두 결합하였으나, 실시예에 따라서는 다양한 방식으로 인공지능 모델을 선택하는 것도 가능하다. 예를들어, 오토인코더 모델(M1), CNN 모델(M2), ViT 모델(M3) 중 어느 하나만 선택하거나, 세 개 중 어느 두개만을 선택하여 결합하여 제2 모델을 형성하는 것도 가능하다.

- [0085] 도12는 본 발명의 다른 실시예에 의한 음성신호 기반의 연하장애검출장치를 나타내는 개략도이다. 도12를 참조하면, 연하장애검출장치(100)가 직접 검사대상자(1)의 음성을 수신하여 녹음할 수 있으며, 수신한 음성신호를 기반으로 연하장애를 판정하여 판정결과를 제공할 수 있다. 연하장애검출장치(100)는 메모리(memory)와 프로세서(processor)를 구비하여 연산처리 능력을 갖춘 장치일 수 있으며, 스마트폰, 태블릿 PC, PDA, 디지털 카메라, 스마트워치, 디지털/아날로그 녹음기, 연하장애검출을 위한 전용의 단말장치 등일 수 있다. 예를들어, 연하장애검출장치(100)는 스마트폰일 수 있으며, 스마트폰에 설치된 전용의 어플리케이션을 통하여 검사대상자(1)의 음성을 녹음할 수 있다. 이후, 전용의 어플리케이션 내에 내장된 분석모델을 이용하여, 대상음성신호로부터 검사대상자(1)의 연하장애 여부를 판정하고, 판정결과를 검사대상자(1)에게 제공할 수 있다. 즉, 외부의 서버와의 별도의 통신을 수행할 필요없이, 연하장애검출장치(100)가 수신한 음성신호에 대한 연하장애여부를 자체적으로 판별할 수 있다.
- [0086] 여기서, 전용의 어플리케이션은 별도의 서버(예를들어, 앱스토어, 플레이스토어 등)로부터 전송받아 연하장애검출장치(100) 내에 설치되거나, 업데이트될 수 있다.
- [0087] 도13을 참조하면, 본 발명의 다른 실시예에 의한 음성신호 기반의 연하장애검출장치는, 모델생성부(110), 음성수신부(120), 판정부(130) 및 마이크로폰(150)을 포함할 수 있다. 여기서, 모델생성부(110), 음성수신부(120) 및 판정부(130)는 앞서 설명하였으므로, 구체적인 설명을 생략한다.
- [0088] 마이크로폰(150)은 검사대상자(1)의 음성을 수집하여 음성신호를 생성할 수 있다. 검사대상자(1)는 음식물 섭취를 전후하여, 마이크로폰(150) 앞에서 미리 정해진 음절이나 단어, 문장 등을 발화할 수 있다. 예를들어, 검사대상자(1)는 /a/를 발음하거나, "시계", "안녕하세요" 등을 발음할 수 있다. 이후 마이크로폰(150)은 검사대상자(1)의 음성을 수집하여 음성신호를 생성할 수 있으며, 생성한 음성신호는 음성수신부(120)로 전송할 수 있다.
- [0090] 도14는 본 발명의 일 실시예에 의한 음성기반의 연하장애검출방법을 나타내는 순서도이다. 여기서, 각 단계들은 본 발명의 일 실시예에 의한 연하장애검출장치에 의하여 실행될 수 있다.
- [0091] 도14을 참조하면, 연하장애검출장치는 검사대상자의 음식물 섭취 전후에 녹음한 대상음성신호를 수신할 수 있다(S10). 음식물은 요거트, 물, 죽, 밥 등 다양한 종류로 제공될 수 있으며, 검사대상자는 음식물의 섭취 전에 특정 음을 발성하여 녹음하고, 음식물 섭취 후에도 동일한 특정 음을 발성하여 녹음할 수 있다. 여기서, 특정 음은 /a/와 같은 단모음일 수 있으나 이에 한정되는 것은 아니고, 자음, 단어, 간단한 문장 등을 포함할 수 있다.
- [0092] 이후, 연하장애검출장치는 대상음성신호를 분석모델에 입력하여, 검사대상자에 대한 연하장애를 판정하여 판정결과를 도출할 수 있다(S20). 인공지능 모델 기반의 분석모델은 입력받은 대상음성신호에 대응하는 판정결과를 제공할 수 있으며, 판정결과에 따라 검사대상자의 연하장애를 진단할 수 있다. 여기서, 분석모델에 이진 분류기가 구비되는 경우에는, 검사대상자의 연하장애를 정상 또는 비정상의 둘 중 하나로 판정하여 판정결과를 도출할 수 있다. 반면에 다중 분류기가 구비되는 경우에는, 연하장애의 정도에 따라 2 이상으로 판정할 수 있다. 예를들어, 검사대상자의 연하장애를 침습, 흡인 및 정상 중 어느 하나로 판정하거나, 연하장애를 PAS에 따른 8단계 중 어느 하나로 판정하여 판정결과를 도출할 수 있다.
- [0093] 여기서, 분석모델은 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와, 상기 복수의 정상인 및 연하장애환자에 대해 수행한 VFSS 검사결과를 인공지능모델로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 것일 수 있다. 즉, 복수의 정상인과 연하장애환자로부터 음식물 섭취 전후의 음성신호를 수집하고, 동일한 정상인과 연하장애환자들에 대한 VFSS 검사결과를 수집한 후, 이를 학습데이터로 이용하여 분석모델을 학습시킬 수 있다. 따라서, 분석모델을 이용하면, 검사대상자의 대상음성신호에 대응하는 연하장애를 판정하는 것이 가능하다.
- [0094] 분석모델은 도15 내지 도17에 도시한 바와 같이 생성할 수 있다.
- [0095] 먼저, 도15를 참조하면, 분석모델을 생성하기 위하여, 연하장애검출장치는 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와 VFSS 검사결과를 수집할 수 있다(S110). 이?, 연하장애검출장치는 각각의 음성신호와 그에 대응되는 VFSS 검사결과가 매칭되도록 수집할 수 있다. 즉, 각각의 음성신호에 정상/비정상, 정상/침습/흡인, PAS의 8단계와 같은 VFSS 검사결과가 매칭되도록 수집할 수 있다. 이 경우, 각각의 음성신호에 대하여 분석모델을 통하여 판정하고자 하는 연하장애의 판정결과에 대한 정답이 레이블링된 것이므로, 이들을 분석모델 학습을 위한 학습데이터로 활용할 수 있다.

- [0096] 연하장애검출장치는 음성신호에 대한 전처리를 수행하여, 음성신호에 대한 1차원 특징정보를 생성할 수 있다(S120). 즉, 연하장애검출장치는 음성신호로부터 특성을 추출하기 위한 전처리를 수행할 수 있으며, 분석모델에 적용할 인공지능 모델의 종류에 따라, 1차원 특징정보를 생성할 수 있다. 여기서, 1차원 특징정보에는 음성신호로부터 추출하는 기본주파수, 지터, 시머, HNR 등이 포함될 수 있다.
- [0097] 연하장애검출장치는 1차원 특징정보와, 1차원 특징정보에 대응하는 VFSS 검사결과를 LSTM 모델로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성할 수 있다(S130). 즉, 학습데이터에는, 1차원 특징정보인 기본주파수, 지터, 시머, HNR 등에 따른 각각의 정답이 레이블링되어 있으므로, 학습데이터로 LSTM 모델을 학습시켜, LSTM 모델 기반의 분석모델을 생성할 수 있다. 이 경우, 분석모델은 수신한 대상음성신호의 1차원 특징정보로부터, 대응하는 연하장애를 판정하는 것이 가능하다.
- [0099] 도16을 참조하면, 분석모델을 생성하기 위하여, 연하장애검출장치는 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호와 VFSS 검사결과를 수집할 수 있으며(S210), 음성신호에 대한 전처리를 수행하여, 음성신호에 대한 2차원 이미지를 생성할 수 있다(S220). 즉, 연하장애검출장치는 음성신호로부터 특성을 추출하기 위한 전처리를 수행할 수 있으며, 분석모델에 적용할 인공지능 모델의 종류에 따라, 2차원 이미지를 생성할 수 있다. 2차원 이미지는 음성신호를 시각화한 스펙트로그램으로, 시간에 따른 음성신호의 주파수, 진폭 등의 특성변화를 나타내는 것일 수 있다. 실시예에 따라서는, 음성신호에 STFT 또는 MFCC를 적용하여 2차원 이미지를 생성할 수 있다.
- [0100] 연하장애검출장치는 2차원 이미지와, 2차원 이미지에 대응하는 VFSS 검사결과를, CNN 모델, ViT 모델 및 오토인코더 모델 중 어느 하나로 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성할 수 있다(S230).
- [0101] CNN 모델이나 ViT 모델을 이용하는 경우, 학습데이터에는 각각의 2차원 이미지에 대응하는 정답이 레이블링되어 있으므로, 학습데이터로 CNN 모델 또는 ViT 모델을 학습시켜, 분석모델을 생성할 수 있다. 이 경우, 분석모델은 수신한 대상음성신호의 2차원 이미지로부터, 대응하는 연하장애를 판정할 수 있다.
- [0102] 한편, 오토인코더 모델은 입력데이터를 압축하는 인코더와, 압축데이터로부터 입력데이터를 복원하는 디코더를 포함하는 것으로, VFSS 검사결과 "정상"에 해당하는 2차원 이미지로만 학습하여 분석모델을 생성할 수 있다. 이 경우, "정상"에 해당하는 2차원 이미지의 입력시에는 동일하게 복원될 수 있으나, VFSS 검사결과 침습이나 흡인 등 "비정상"에 해당하는 2차원 이미지들이 입력되는 경우에는 상이하게 복원될 수 있다. 따라서, 입력된 2차원 이미지와 복원된 2차원 이미지 사이의 차이를 이용하여 이상 점수(anomaly score)를 구할 수 있으며, 이상 점수가 문턱값보다 크면 해당 2차원 이미지는 "비정상"으로 판별할 수 있다.
- [0104] 도17를 참조하면, 분석모델을 생성하기 위하여, 연하장애검출장치는 복수의 정상인 및 연하장애환자로부터 녹음한 음식물 섭취 전후의 음성신호, VFSS 검사결과, 및 정상인과 연하장애환자의 인적정보를 수집할 수 있다(S310). 여기서, 인적정보에는 연하장애환자의 성별, 나이 및 기저질환 등이 포함될 수 있다.
- [0105] 연하장애검출장치는 음성신호에 대한 전처리를 수행하여, 음성신호에 대한 1차원 특징정보 및 2차원 이미지를 생성할 수 있으며(S320), 연하장애검출장치는 1차원 특징정보와, 1차원 특징정보에 대응하는 VFSS 검사결과를, LSTM 모델로 학습하는 제1 모델을 생성할 수 있다(S330).
- [0106] 또한, 연하장애검출장치는 2차원 이미지와, 2차원 이미지에 대응하는 VFSS 검사결과를, CNN 모델, ViT 모델 및 오토인코더 모델 중 적어도 어느 하나로 학습하는 제2 모델을 생성할 수 있다(S340). 예를들어, 오토인코더 모델, CNN 모델, ViT 모델 중 어느 하나만 선택하거나, 세 개 중 어느 두개만을 선택하여 결합하여 제2 모델을 형성할 수 있으며, 세 개의 모델 모두를 결합하여 제2 모델을 형성하는 것도 가능하다.
- [0107] 여기서, 세 개의 모델을 모두 결합하는 경우, 연하장애검출장치는 먼저 2차원 이미지가 오토인코더 모델에 입력되면, 오토인코더 모델에 의해 생성된 복원 2차원 이미지를, 2차원 이미지와 채널 방향으로 결합하여 결합 2차원 이미지를 생성하도록 할 수 있다. 또한, 결합 2차원 이미지가 CNN 모델에 입력되면, 결합 2차원 이미지에 대응하는 특징맵을 생성하도록 하고, 특징맵을 풀링하여 복수의 ViT 모델에 각각 풀링된 특징맵을 입력하면, 각각의 ViT 모델에서 생성한 ViT 특징벡터들을 결합하여 제2 특징벡터를 생성하도록, 제2 모델을 생성할 수 있다.
- [0108] 이후, 연하장애검출장치는 제1 모델로부터 추출한 제1 특징벡터, 제2 모델로부터 추출한 제2 특징벡터 및 인적

정보로부터 생성한 제3 특징벡터를 결합한 결합벡터에 적용하는 분류기 함수를 생성할 수 있다(S350). 실시예에 따라서는, 제1 특징벡터와 제2 특징벡터를 결합하여 결합벡터를 생성하거나, 제1 특징벡터와 제3 특징벡터, 제2 특징벡터와 제3 특징벡터를 결합하여 각각 결합벡터를 생성하고, 이들에 대한 분류기 함수를 생성하는 것도 가능하다. 여기서 제3 특징벡터는 기저질환을 원-핫 벡터 형태로 나타내고, 원-핫 벡터에 나머지 성별이나 나이를 나타내는 정보 등을 추가하여 생성하는 것일 수 있다. 즉, 제3 특징벡터에 의하여 분석모델 내에 연하장애환자의 성별, 나이, 기저질환 등에 관한 정보가 반영될 수 있으므로, 보다 정확하게 연하장애를 판정하는 것이 가능하다.

[0109] 연하장애검출장치는 제1 모델, 제2 모델 및 분류기 함수를 포함하는 분석모델을 학습하여, 수신한 음성신호에 대응하는 연하장애를 판정하는 분석모델을 생성할 수 있다(S360).

[0111] 전술한 본 발명은, 프로그램이 기록된 매체에 컴퓨터가 읽을 수 있는 코드로서 구현하는 것이 가능하다. 컴퓨터가 읽을 수 있는 매체는, 컴퓨터로 실행 가능한 프로그램을 계속 저장하거나, 실행 또는 다운로드를 위해 임시 저장하는 것일 수도 있다. 또한, 매체는 단일 또는 수개 하드웨어가 결합된 형태의 다양한 기록수단 또는 저장수단일 수 있는데, 어떤 컴퓨터 시스템에 직접 접속되는 매체에 한정되지 않고, 네트워크 상에 분산 존재하는 것일 수도 있다. 매체의 예시로는, 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체, CD-ROM 및 DVD와 같은 광기록 매체, 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical medium), 및 ROM, RAM, 플래시 메모리 등을 포함하여 프로그램 명령어가 저장되도록 구성된 것이 있을 수 있다. 또한, 다른 매체의 예시로, 애플리케이션을 유통하는 앱 스토어나 기타 다양한 소프트웨어를 공급 내지 유통하는 사이트, 서버 등에서 관리하는 기록매체 내지 저장매체도 들 수 있다. 따라서, 상기의 상세한 설명은 모든 면에서 제한적으로 해석되어서는 아니되고 예시적인 것으로 고려되어야 한다. 본 발명의 범위는 첨부된 청구항의 합리적 해석에 의해 결정되어야 하고, 본 발명의 등가적 범위 내에서의 모든 변경은 본 발명의 범위에 포함된다.

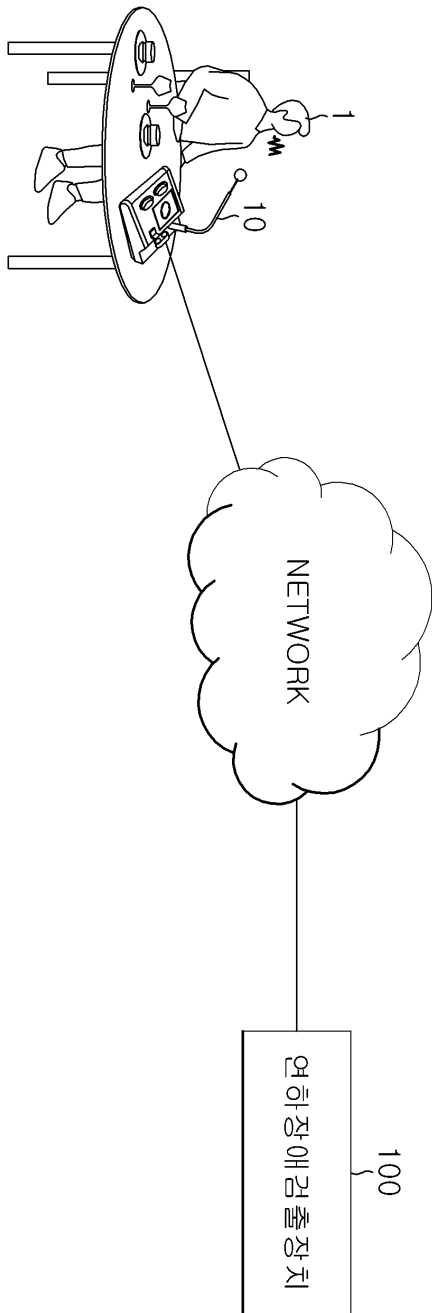
[0113] 본 발명은 전술한 실시예 및 첨부된 도면에 의해 한정되는 것이 아니다. 본 발명이 속하는 기술분야에서 통상의 지식을 가진 자에게 있어, 본 발명의 기술적 사상을 벗어나지 않는 범위 내에서 본 발명에 따른 구성요소를 치환, 변형 및 변경할 수 있다는 것이 명백할 것이다.

부호의 설명

[0115] 1: 검사대상자 10: 사용자 단말
100: 연하장애검출장치 110: 모델생성부
120: 음성수신부 130: 판정부
140: 통신부 150: 마이크로폰

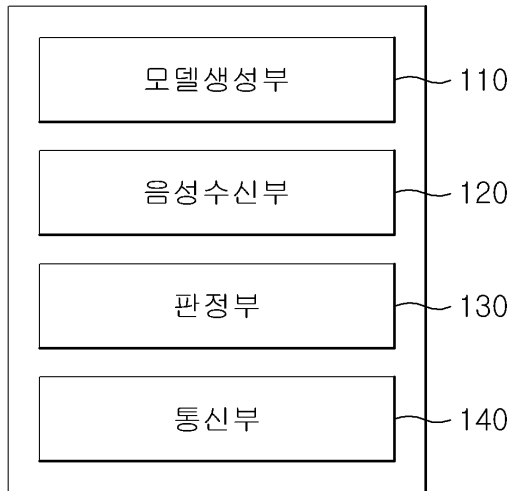
도면

도면1



도면2

100

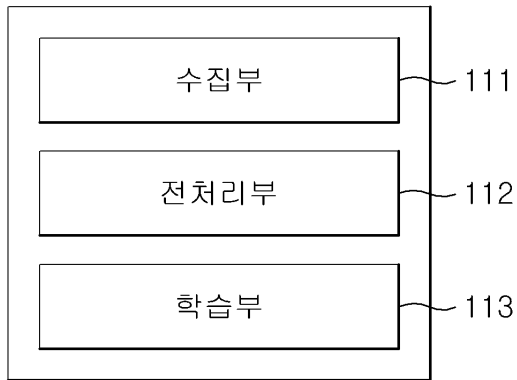


도면3

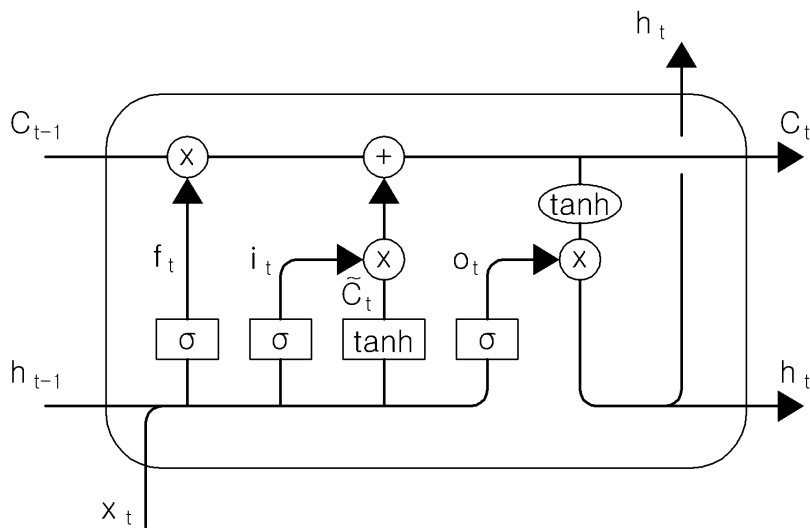
척도	기준	결과
1	음식이 기도로 들어가지 않음	정상
2	음식이 기도로 들어가고, 성대 위에 남아 있음. 기도 밖으로 나옴	침습
3	음식이 기도로 들어가고, 성대 위에 남아 있음. 기도 밖으로 나오지 않음	침습
4	음식이 기도로 들어가고, 성대에 닿음, 기도 밖으로 나옴	침습
5	음식이 기도로 들어가고, 성대에 닿음, 기도 밖으로 나오지 않음	침습
6	음식이 기도로 들어가고, 성대 아래를 지남, 후두 또는 기도 밖으로 나옴	흡인
7	음식이 기도로 들어가고, 성대 아래를 지남, 후두 또는 기도 밖으로 나오지 않음	흡인
8	음식이 기도로 들어가고, 성대 아래를 지남, 밖으로 나오려는 작용이 없음	흡인

도면4

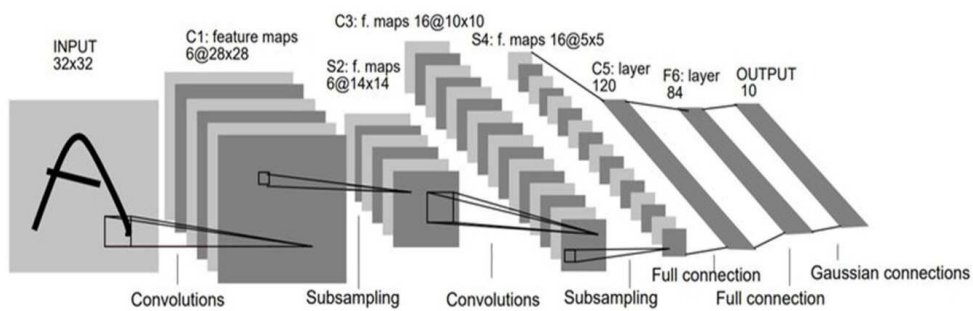
110



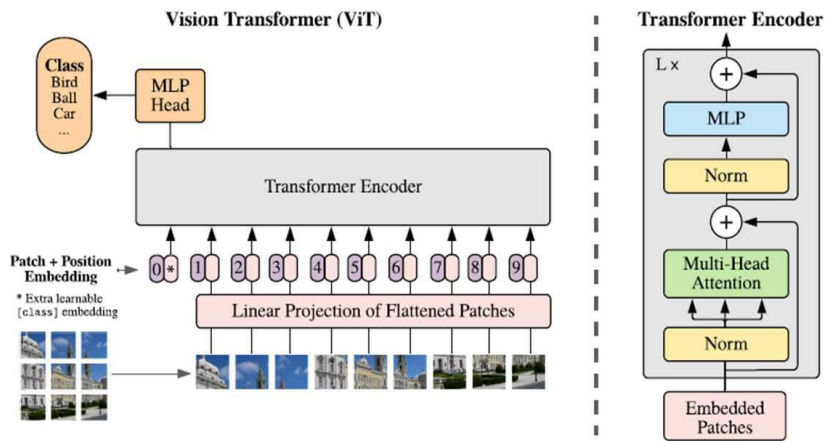
도면5



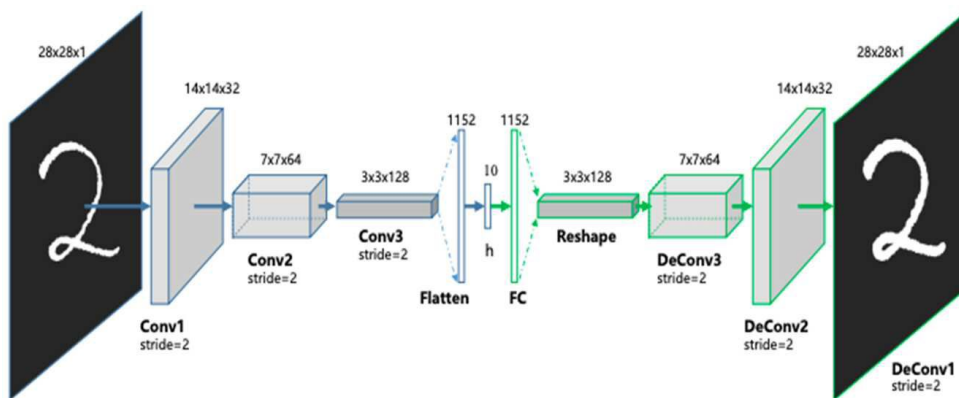
도면6



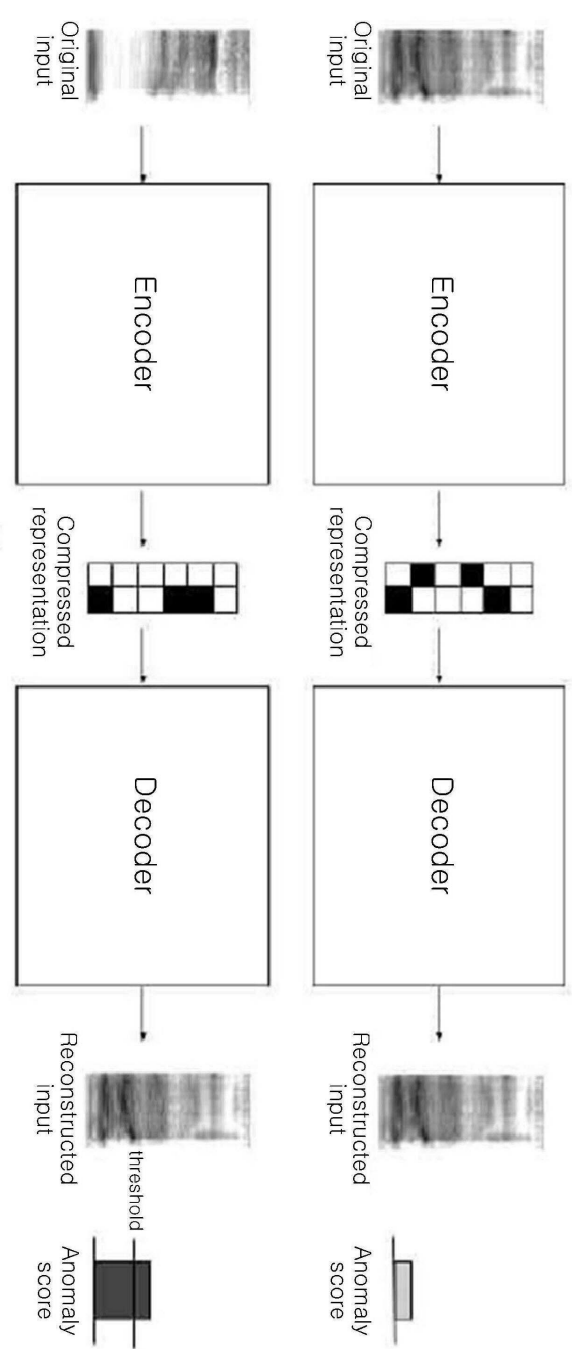
도면7



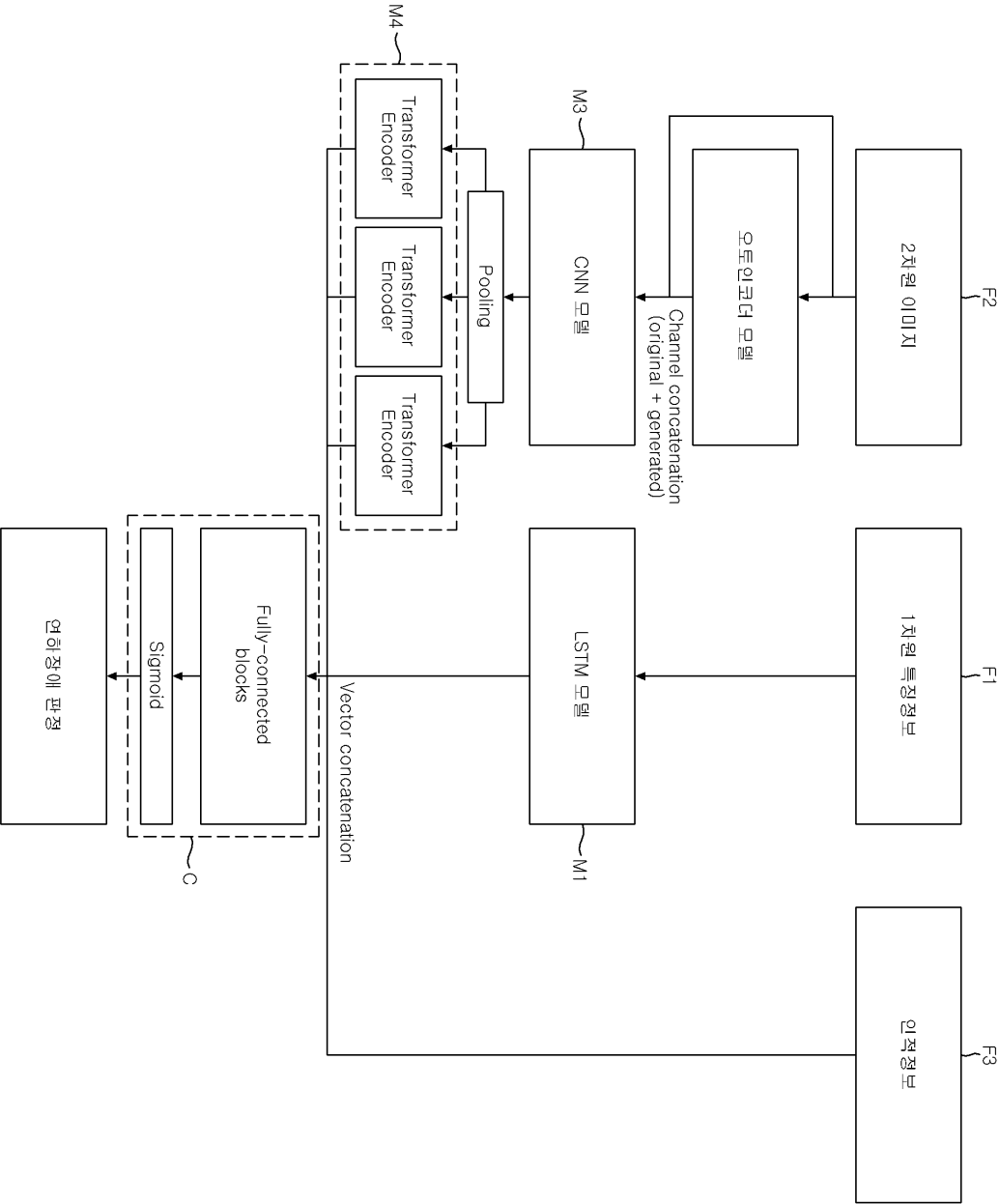
도면8



도면9



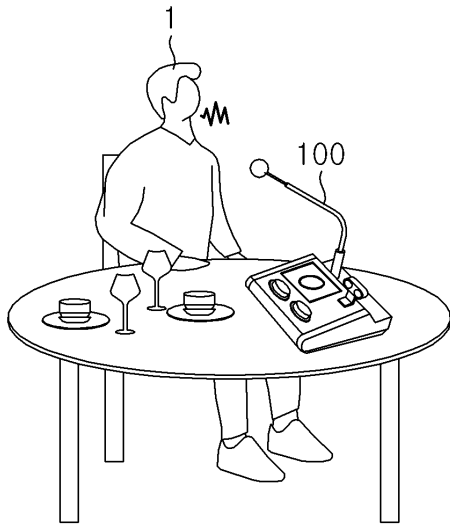
도면10



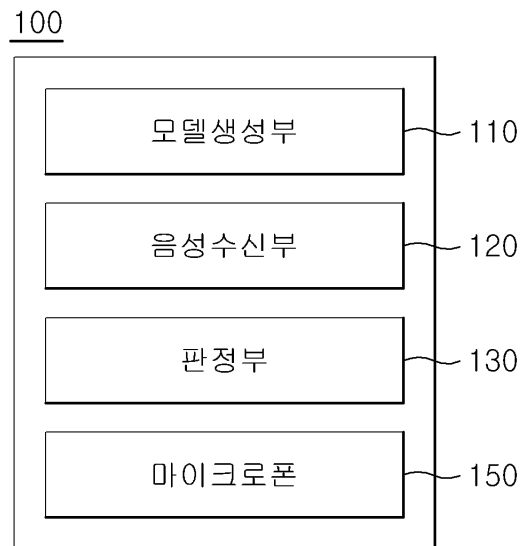
도면11

	성별	나이	기저 질환			
			뇌졸중	성대 결절	고혈압	...
환자 정보	남자	56	유	무	유	...
Scalar feature	0	0.28	1	0	1	...

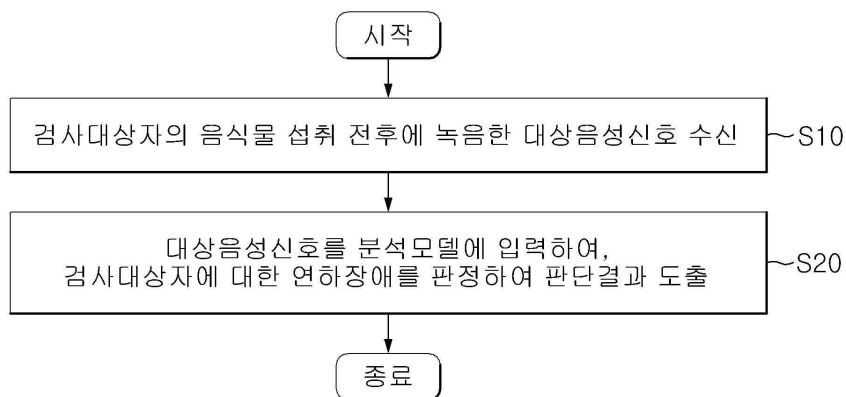
도면12



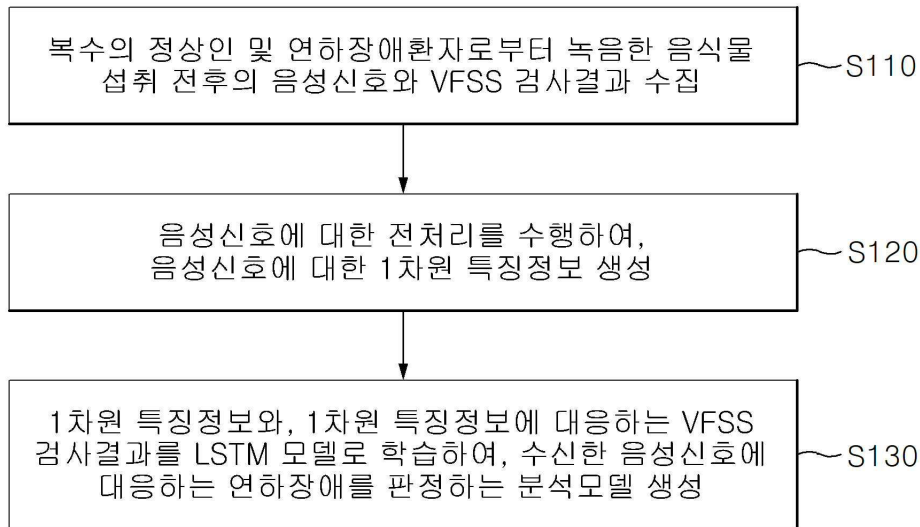
도면13



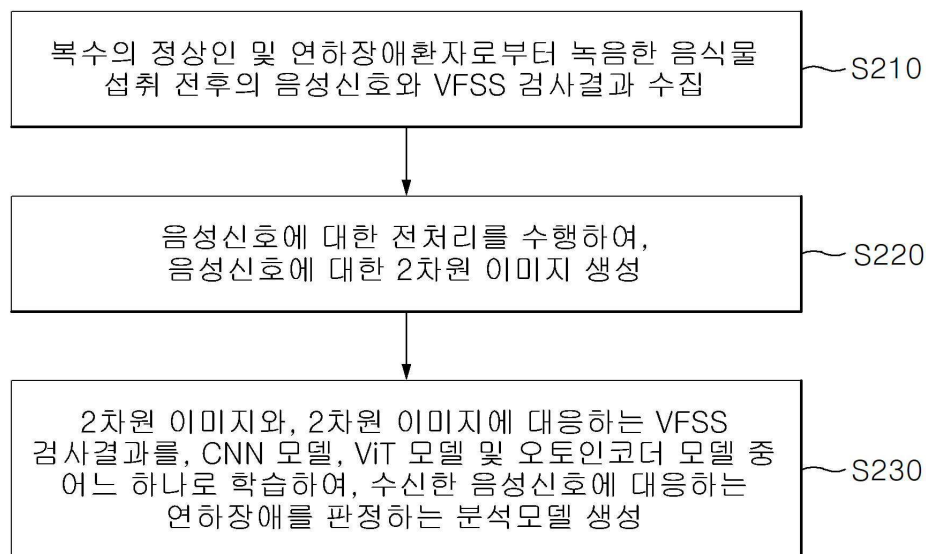
도면14



도면15



도면16



도면17

