



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2023-0089752
(43) 공개일자 2023년06월21일

(51) 국제특허분류(Int. Cl.)
G16H 50/80 (2018.01) G06N 3/08 (2023.01)
G16H 50/50 (2018.01) G16H 50/70 (2018.01)
(52) CPC특허분류
G16H 50/80 (2018.01)
G06N 3/08 (2023.01)
(21) 출원번호 10-2021-0178401
(22) 출원일자 2021년12월14일
심사청구일자 2021년12월14일

(71) 출원인
인천대학교 산학협력단
인천광역시 연수구 아카데미로 119 (송도동)
(72) 발명자
전광길
인천광역시 연수구 해송로 70, 209동 805호(송도동, 송도웹카운티2단지)
(74) 대리인
특허법인충정

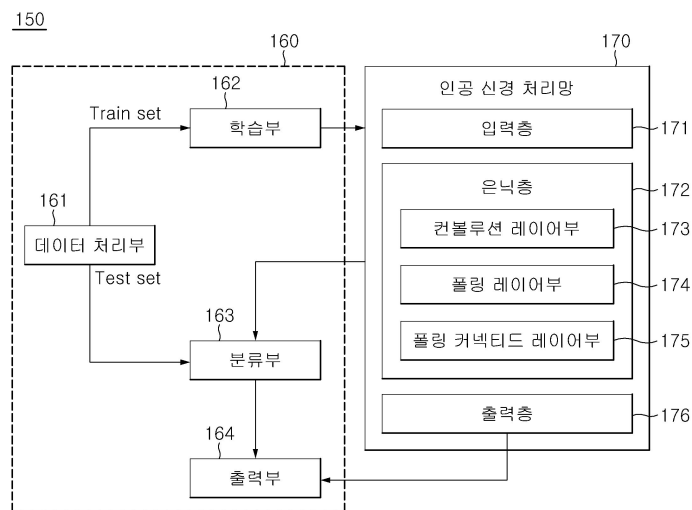
전체 청구항 수 : 총 10 항

(54) 발명의 명칭 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템 및 방법

(57) 요약

빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템 및 방법은 코로나 바이러스 감염증과 질병 증상 간의 상관 관계를 분석하여 복수의 질병 증상을 인공 신경 처리망에 입력하여 코로나 19 바이러스 감염 결과 여부를 예측하며, 신경망 모델을 기반으로 전염병을 진단하고 예측하도록 설계되어 99%의 정확도를 달성할 수 있는 효과가 있다.

대표도



(52) CPC특허분류

G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

명 세 서

청구범위

청구항 1

다양한 연령대와 성별에 대하여 코로나 19 바이러스와 관련된 질병 증상을 복수개로 이루어진 데이터 세트로 수집하는 데이터 수집부;

상기 코로나 바이러스 감염증과 상기 질병 증상 간의 상관 관계를 분석하는 진단 분석부; 및

상기 복수의 질병 증상을 인공 신경 처리망에 입력하고, 상기 인공 신경 처리망의 응답으로 질병 증상 정보에 대응하는 코로나 19 바이러스 감염 결과 여부를 예측하는 예측 분석부를 포함하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템.

청구항 2

청구항 1에 있어서,

상기 데이터 수집부로부터 수신한 데이터 세트에서 다양한 연령대의 환자(남성과 여성)에 대한 코로나 19 바이러스의 분석, 예측 및 탐지를 위해 상기 복수개의 질병 증상을 분석하고, 상기 분석한 질병 증상의 통계를 계산하고, 상기 계산된 통계를 퍼센트율로 시각화하여 그래프로 표시하는 기술 분석부를 더 포함하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템.

청구항 3

청구항 2에 있어서,

상기 진단 분석부는 상기 복수개의 질병 증상의 통계가 기설정된 기준 임계값 이상인 질병 증상을 상기 코로나 바이러스 감염증과 연관성이 있는 양성으로 판단하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템.

청구항 4

청구항 2에 있어서,

상기 예측 분석부는 상기 각각의 질병 증상을 분석하여 통계를 퍼센트율로 계산하고, 상기 계산한 퍼센트율을 기초로 상기 각각의 질병 증상의 가중치를 다르게 설정하고, 상기 다르게 설정된 가중치를 입력 벡터인 상기 각각의 질병 증상과 복수의 연산을 수행하여 입력층, 은닉층, 출력층을 통과하면, 상기 코로나 19 바이러스 감염 결과를 출력 벡터로 생성하도록 지도 학습을 통해 학습되고, 상기 입력층에서 상기 출력층으로 방향으로 계산하고, 반대로 상기 출력층에서 상기 입력층 방향으로 계산하는 작업을 반복하면서 가중치를 수정하여 오차를 최소화하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템.

청구항 5

청구항 4에 있어서,

상기 예측 분석부는 테스트 데이터인 복수의 질병 증상 정보를 상기 인공 신경 처리망의 딥러닝 기반 분류기 모델을 이용하여 감염된 코로나 19 바이러스 또는 비감염된 코로나 19 바이러스 결과 여부를 판단하고, 상기 복수의 질병 증상 정보에 대응하여 상기 감염된 코로나 19 바이러스인 제1 출력값 확률이 제1 퍼센트율인 경우, 상기 복수의 질병 증상 정보에 대응하는 상기 비감염된 코로나 19 바이러스인 제2 출력값 확률이 (100 - 제1 퍼센트율)인 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템.

청구항 6

청구항 5에 있어서,

상기 예측 분석부는 상기 제1 출력값 확률이 기설정된 기준 확률값 이상인 경우, 상기 제1 출력값 확률의 입력 벡터인 상기 복수의 질병 증상 정보를 감염된 코로나 19 바이러스에 대응하는 질병 증상으로 설정하고, 상기 기

준 확률값을 필요에 따라 조정하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템.

청구항 7

데이터 수집부는 다양한 연령대와 성별에 대하여 코로나 19 바이러스와 관련된 질병 증상을 복수개로 이루어진 데이터 세트로 수집하는 단계;

기술 분석부는 상기 복수개의 질병 증상을 분석하여 통계를 계산하고, 상기 계산된 통계를 퍼센트율로 시각화하여 표시하는 단계;

진단 분석부는 상기 코로나 바이러스 감염증과 상기 질병 증상 간의 상관 관계를 분석하는 단계; 및

예측 분석부는 상기 복수의 질병 증상을 인공 신경 처리망에 입력하고, 상기 인공 신경 처리망의 응답으로 질병 증상 정보에 대응하는 코로나 19 바이러스 감염 결과 여부를 예측하는 단계를 포함하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 방법.

청구항 8

청구항 7에 있어서,

상기 코로나 19 바이러스 감염 결과 여부를 예측하는 단계는,

상기 예측 분석부는 상기 각각의 질병 증상을 분석하여 통계를 퍼센트율로 계산하는 단계;

상기 예측 분석부는 상기 계산한 퍼센트율을 기초로 상기 각각의 질병 증상의 가중치를 다르게 설정하는 단계;

상기 예측 분석부는 상기 다르게 설정된 가중치를 입력 벡터인 상기 각각의 질병 증상과 복수의 연산을 수행하여 입력층, 은닉층, 출력층을 통과하면, 상기 코로나 19 바이러스 감염 결과를 출력 벡터로 생성하도록 지도 학습을 통해 학습되는 단계; 및

상기 예측 분석부는 상기 입력층에서 상기 출력층으로 방향으로 계산하고, 반대로 상기 출력층에서 상기 입력층 방향으로 계산하는 작업을 반복하면서 가중치를 수정하여 오차를 최소화하는 단계를 더 포함하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 방법.

청구항 9

청구항 8에 있어서,

상기 예측 분석부는 테스트 데이터인 복수의 질병 증상 정보를 상기 인공 신경 처리망의 딥러닝 기반 분류기 모델을 이용하여 감염된 코로나 19 바이러스 또는 비감염된 코로나 19 바이러스 결과 여부를 판단하는 단계; 및

상기 예측 분석부는 상기 복수의 질병 증상 정보에 대응하여 상기 감염된 코로나 19 바이러스인 제1 출력값 확률이 제1 퍼센트율인 경우, 상기 복수의 질병 증상 정보에 대응하는 상기 비감염된 코로나 19 바이러스인 제2 출력값 확률 (100 - 제1 퍼센트율)을 계산하는 단계를 더 포함하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 방법.

청구항 10

청구항 9에 있어서,

상기 예측 분석부는 상기 제1 출력값 확률이 기설정된 기준 확률값 이상인 경우, 상기 제1 출력값 확률의 입력 벡터인 상기 복수의 질병 증상 정보를 감염된 코로나 19 바이러스에 대응하는 질병 증상으로 설정하는 단계; 및

상기 기준 확률값을 필요에 따라 조정하는 단계를 더 포함하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 방법.

발명의 설명

기술 분야

본 발명은 분석 시스템에 관한 것으로서, 더욱 상세하게는 코로나 바이러스 감염증과 질병 증상 간의 상관 관계를 분석하여 복수의 질병 증상을 인공 신경 처리망에 입력하여 코로나 19 바이러스 감염 결과 여부를 예측하는

[0001]

빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템 및 방법에 관한 것이다.

배경 기술

- [0002] IoT(사물 인터넷) 장치와 스마트 센서는 산업, 비즈니스, 감시, 의료, 운송, 통신 등 다양한 생활 분야에서 사용된다.
- [0003] 최근 신종 코로나 바이러스 감염증(COVID-19)이 인류의 건강, 생명, 생산, 사회적 상호 작용, 국제 관계를 심각하게 위협하고 있다. 이러한 상황에서 IoT와 빅데이터 기술은 전염병에 맞서 싸우는데 필수적인 역할을 했다.
- [0004] 응용 프로그램에는 빅데이터의 빠른 수집, 전염병 정보 시각화, 전염병 위험 분석, 확인 사례 추적, 예방 수준 추적, COVID-19 예방 및 통제에 대한 적절한 평가가 포함될 수 있다.
- [0005] 사물 인터넷 장치 및 시스템의 수는 나날이 증가하고 있다. 이러한 장치와 센서는 산업, 비즈니스, 감시, 의료, 운송, 통신 등과 같은 다양한 생활 부문에서 사용되는 많은 데이터를 생성한다. 의료 분야에서는 다양한 장치, 센서 및 시스템을 통해 수집되는 정보의 양이 증가함에 따라 빅데이터 분석 및 IoT를 활용하는 부분이 매우 적거나 전혀 없는 영역도 있다.
- [0006] 코로나 바이러스 감염증은 빠른 확산 속도로 인하여 코로나 분석, 예측, 모니터링 분야가 점차 중요해지고 있으며, 추후 사물 인터넷 장치와 스마트 센서를 이용한 영역이 확대될 가능성이 높아지고 있다.

선행기술문헌

특허문헌

- [0007] (특허문헌 0001) 한국 등록특허번호 제10-2239163호

발명의 내용

해결하려는 과제

- [0008] 이와 같은 문제점을 해결하기 위하여, 본 발명은 코로나 바이러스 감염증과 질병 증상 간의 상관 관계를 분석하여 복수의 질병 증상을 인공 신경 처리망에 입력하여 코로나 19 바이러스 감염 결과 여부를 예측하는 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템 및 방법을 제공하는데 그 목적이 있다.

과제의 해결 수단

- [0009] 상기 목적을 달성하기 위한 본 발명의 특징에 따른 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템은,
- [0010] 다양한 연령대와 성별에 대하여 코로나 19 바이러스와 관련된 질병 증상을 복수개로 이루어진 데이터 세트로 수집하는 데이터 수집부;
- [0011] 상기 코로나 바이러스 감염증과 상기 질병 증상 간의 상관 관계를 분석하는 진단 분석부; 및
- [0012] 상기 복수의 질병 증상을 인공 신경 처리망에 입력하고, 상기 인공 신경 처리망의 응답으로 질병 증상 정보에 대응하는 코로나 19 바이러스 감염 결과 여부를 예측하는 예측 분석부를 포함한다.
- [0014] 본 발명의 특징에 따른 빅데이터 분석을 이용한 전염병 예측을 위한 분석 방법은,
- [0015] 데이터 수집부는 다양한 연령대와 성별에 대하여 코로나 19 바이러스와 관련된 질병 증상을 복수개로 이루어진 데이터 세트로 수집하는 단계;
- [0016] 기술 분석부는 상기 복수개의 질병 증상을 분석하여 통계를 계산하고, 상기 계산된 통계를 퍼센트율로 시각화하여 표시하는 단계;
- [0017] 진단 분석부는 상기 코로나 바이러스 감염증과 상기 질병 증상 간의 상관 관계를 분석하는 단계; 및
- [0018] 예측 분석부는 상기 복수의 질병 증상을 인공 신경 처리망에 입력하고, 상기 인공 신경 처리망의 응답으로 질병

증상 정보에 대응하는 코로나 19 바이러스 감염 결과 여부를 예측하는 단계를 포함한다.

발명의 효과

[0019] 전술한 구성에 의하여, 본 발명은 신경망 모델을 기반으로 전염병을 진단하고 예측하도록 설계되어 99%의 정확도를 달성할 수 있는 효과가 있다.

[0020] 본 발명은 빅 데이터 분석과 IoT를 통합하여 새로운 질병을 분석하고 예측하여 코로나 19 바이러스의 분석 및 예측을 위한 건강 모니터링 프레임워크를 개발하고, 이에 따라 예측된 결과를 의료진에 제공하여 바이러스를 감지하고 진단하는데 도움이 되는 효과가 있다.

도면의 간단한 설명

[0021] 도 1은 본 발명의 실시예에 따른 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템의 구성을 간략하게 나타낸 블록도이다.

도 2는 본 발명의 실시예에 따른 예측 분석부의 내부 구성을 나타낸 블록도이다.

도 3은 본 발명의 실시예에 따른 다양한 속성 분포를 나타낸 도면이다.

도 4는 본 발명의 실시예에 따른 데이터 세트에서 코로나 19 바이러스의 증상 분포를 나타낸 도면이다.

도 5는 본 발명의 실시예에 따른 양성 및 음성인 남성과 여성 환자의 분포를 나타낸 도면이다.

도 6은 본 발명의 실시예에 따른 환자의 나이와 성별 간의 관계를 나타낸 도면이다.

도 7은 본 발명의 실시예에 따른 서로 다른 속성 간의 관계를 나타낸 도면이다.

도 8은 본 발명의 실시예에 따른 피드포워드와 역전파의 두 단계로 신경망 모델의 훈련 과정을 나타낸 도면이다.

발명을 실시하기 위한 구체적인 내용

[0022] 명세서 전체에서, 어떤 부분이 어떤 구성요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.

[0023] 도 1은 본 발명의 실시예에 따른 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템의 구성을 간략하게 나타낸 블록도이고, 도 2는 본 발명의 실시예에 따른 예측 분석부의 내부 구성을 나타낸 블록도이고, 도 3은 본 발명의 실시예에 따른 다양한 속성 분포를 나타낸 도면이고, 도 4는 본 발명의 실시예에 따른 데이터 세트에서 코로나 19 바이러스의 증상 분포를 나타낸 도면이고, 도 5는 본 발명의 실시예에 따른 양성 및 음성인 남성과 여성 환자의 분포를 나타낸 도면이고, 도 6은 본 발명의 실시예에 따른 환자의 나이와 성별 간의 관계를 나타낸 도면이고, 도 7은 본 발명의 실시예에 따른 서로 다른 속성 간의 관계를 나타낸 도면이고, 도 8은 본 발명의 실시예에 따른 피드포워드와 역전파의 두 단계로 신경망 모델의 훈련 과정을 나타낸 도면이다.

[0024] 본 발명의 실시예에 따른 빅데이터 분석을 이용한 전염병 예측을 위한 분석 시스템(100)은 데이터 수집부(110), 기술 분석부(120), 진단 분석부(130), 제어부(140), 예측 분석부(150), 디스플레이부(180)를 포함한다.

[0025] 데이터 수집부(110)는 다양한 연령대의 다양한 환자(남성과 여성)에 대한 거의 26000건의 기록이 데이터 세트를 저장하고 있다. 팬데믹의 분석, 예측 및 탐지를 위해 다양한 속성(질병 증상)이 사용된다. 데이터 세트의 대상 속성은 실험실 결과 및 바이러스 진단 후 환자의 생존이다.

[0026] 데이터 수집부(110)는 다양한 연령대와 성별에 대하여 코로나 19 바이러스와 관련된 질병 증상을 복수개로 이루어진 데이터 세트로 수집한다.

[0027] 각 속성에 대한 자세한 설명은 하기의 표 1에 나와 있다. 대부분의 속성이 독감, 발열, 인후통, 기침 등과 같은 질병의 일반적인 증상이라는 것을 표 1에서 알 수 있다. 속성은 질병 증상을 나타낸다.

[0028] 대부분의 데이터 세트 속성은 범주형이다. 예를 들어, yes는 1이고 no는 0이다. 실험실 결과의 경우 긍정적인 실험실 테스트에 1을 사용하고 부정적인 결과에 0을 사용했다. age 속성은 숫자이고, 최종 환자 상태는 categorical 속성이다.

표 1

S. No.	Attribute	Attribute information
1	Age	In Years
2	Gender	Male & Female
3	Is Patient symptomatic?	Yes & No
4	Flu	Yes & No
5	Fever	Yes & No
6	Fever History	Yes & No
7	Sore Throat	Yes & No
8	Cough	Yes & No
9	Diarrhea	Yes & No
10	Breathing issue	Yes & No
11	Headache	Yes & No
12	Cardiovascular & hypertension	Yes & No
14	Diabetes	Yes & No
15	Any travel history in 14 days while symptomatic?	Yes & No
16	Foreign Travel History	Yes & No
17	Specimen information	Nasopharyngeal & Oropharyngeal swab
18	Lab Results	Positive & Negative
19	Is Home Quarantine?	Yes & No
20	ICU Admission	Yes & No
21	Put On Ventilator	Yes & No
22	Patient Final Status	Expired & Recovered & Active

[0029]

[0030] 기술 분석부(120)는 데이터 수집부(110)로부터 수신한 데이터 세트의 속성의 수, 각 속성의 기능 및 속성의 모집단 크기를 포함하여 데이터 세트에 있는 모든 속성에 대한 자세한 정보를 제공한다.

[0031] 시각적 그래프를 사용하여 사람이 해석할 수 있는 원시 데이터의 요약 또는 자세한 설명을 만든다.

[0032] 기술 분석은 마찬가지로 가장 간단한 유형의 분석으로 알려져 있다. 의료에서는 단순히 데이터의 표준 통계를 식별한다.

[0033] 기술 분석부(120)는 데이터 수집부(110)로부터 수신한 데이터 세트에서 다양한 연령대의 환자(남성과 여성)에 대한 코로나 19 바이러스의 분석, 예측 및 탐지를 위해 다양한 속성(질병 증상)을 분석한다.

[0034] 기술 분석부(120)는 데이터 수집부(110)로부터 수신한 데이터 세트에서 다양한 연령대의 환자(남성과 여성)에 대한 코로나 19 바이러스의 분석, 예측 및 탐지를 위해 복수개의 질병 증상을 분석하고, 분석한 질병 증상의 통계를 계산하고, 계산된 통계를 퍼센트율로 시각화하여 디스플레이부(180)에 출력한다(그래프 등).

[0035] 예를 들어, 수행된 검사실 검사의 총 횟수, 환자의 평균 연령, 특정 질병으로 고통받는 남성과 여성의 수, 환자의 평균 입원 기간/기간 또는 특정 질병에서 회복된 사람의 수, 기술 분석의 경우 막대 또는 세로 막대형 차트 및 표, 파이 차트 또는 서면 내러티브가 사용된다.

[0036] 기술 분석의 예는 도 3과 도 4에 나와 있다. 수집된 데이터 세트의 다양한 속성 분포를 표시하기 위해 파이 그래프를 사용했다.

[0037] 도 3의 (a)와 같이, 실험실 결과는 양성(1) 95.2%, 음성(0) 4.8%이고, 도 3의 (b)와 같이, 남성(0) 74.3%, 여성(1) 25.7%이며, 도 3의 (c)와 같이, 회복된 사례(0) 29.2%, 만료된 사례(1) 25.7%, 활성화된 사례(2) 4.6%이다.

[0038] 대부분의 환자가 데이터 세트에서 긍정적인 실험실 결과를 가지고 있음을 도 3에서 알 수 있다.

- [0039] 유사하게 성별 분포 정보도 도 3의 (b)에 나타나 있으며, 대부분의 환자가 남성인 것으로 분석될 수 있다.
- [0040] 최종 환자 상태도 도 3의 (c)에 나와 있다. 대부분의 환자는 데이터 세트에서 활성 상태를 가지고 있다.
- [0041] 환자의 회복 상태는 만료된 환자의 상태에 비해 상대적으로 높다. 독감, 발열 이력, 인후통, 기침 및 호흡 문제를 포함한 대유행의 다양한 증상 속성의 분포도 도 3에 나와 있다.
- [0042] 대부분의 환자가 이러한 증상을 겪고 있음을 도 4에서 알 수 있다. 이러한 그래프의 시각화는 의료 전문가가 데이터 내부에 무엇이 있는지 확인하는 데 유용하다.
- [0043] 대유행 기간 동안 병원을 방문하는 환자들은 다음과 같은 일반적인 증상을 가지고 있음을 알 수 있다.
- [0044] 도 4의 (a)에서 대부분의 환자는 병원에 보고될 때 증상이 나타난다. 도 4에서 발열, 두통, 독감, 기침, 인후통 및 호흡 문제가 있는 환자의 인구 규모를 쉽게 분석할 수 있다.
- [0045] 도 4는 데이터 세트에서 유행성 증상의 분포로 전체 데이터 세트에서 대부분의 기록은 증상이 있는 환자에 대한 것이고, 환자들은 대부분 두통, 발열, 발열, 인후통, 기침, 호흡곤란 등 다양한 증상으로 고통받고 있다.
- [0046] 도 4의 (a)와 같이, 증상 환자 없음(0) 64.7%, 증상 환자 있음(1) 35.3%이고, 도 4의 (b)와 같이, 두통 증상 없음(0) 8.0%, 두통 증상 있음(1) 92.0%이고, 도 4의 (c)와 같이, 감기 증상 없음(0) 28.5%, 감기 증상 있음(1) 71.5%이고, 도 4의 (d)와 같이, 열 증상 없음(0) 5.3%, 열 증상 있음(1) 94.7%이고, 도 4의 (e)와 같이, 열 이력 없음(0) 5.3%, 열 이력 있음(1) 94.7%이고, 도 4의 (f)와 같이, 목 쓰림 없음(0) 8.2%, 목 쓰림 있음(1) 91.8%이고, 도 4의 (g)와 같이, 기침 없음(0) 7.6%, 기침 있음(1) 92.4%이고, 도 3의 (h)와 같이, 호흡 문제 없음(0) 7.6%, 호흡 문제 있음(1) 92.4%이다.
- [0047] 진단 분석부(130)는 데이터 검색, 데이터 마이닝 및 상관 기술로 데이터의 통찰력을 분석하고 이벤트 및 행동의 원인을 해석한다.
- [0048] 진단 분석부(130)는 정보를 탐색하고 다른 속성 정보를 사용하여 상관 관계를 형성한다. 예를 들어, 고열, 마른 기침, 독감, 피로와 같은 환자의 모든 증상이 동일한 바이러스 인자를 가리키는지 확인하는데 도움이 될 수 있다. 진단 분석에서는 질병의 증상과 원인을 탐색한다.
- [0049] 진단 분석부(130)는 데이터 세트에서 환자의 모든 증상(속성)을 분석하고, 서로 다른 코로나 바이러스 감염증의 증상 간의 상관 관계를 분석한다.
- [0050] 진단 분석부(130)는 복수개의 질병 증상의 통계가 기설정된 기준 임계값 이상인 질병 증상을 코로나 바이러스 감염증과 연관성이 있는 양성으로 판단한다.
- [0051] 여기서, 기준 임계값은 도 3과 같이, 양성을 판단할 수 있는 기준 퍼센트율을 의미하고, 70%, 80%, 90% 등 다양하게 설정할 수 있으며, 필요에 따라 기준 임계값을 변경할 수 있다.
- [0052] 진단 분석 예는 도 5에 나와 있습니다. 진단 분석부(130)는 도 4와 같이, 다양한 속성에 대해 Lab 결과를 분석했다.
- [0053] 도 5에 도시된 바와 같이, 검사 결과의 양성 상태(1)가 높기 때문에 질병이 남성 환자에서 대부분 발견됨을 알 수 있다.
- [0054] 도 5의 (a)로부터, 남성 환자에 대한 대부분의 결과가 여성에 비해 양성임을 알 수 있다.
- [0055] 도 5의 (b)에서 환자의 증상 속성은 Lab 결과에 대해 분석된다. 시각화 결과, 증상이 없는 환자에서도 양성 사례가 많기 때문에 증상이 있다고 해서 코로나 바이러스 감염증을 겪는 것은 아니라는 분석을 했다.
- [0056] 시각화 결과에서 대부분의 경우 환자의 결과가 긍정적인임을 알 수 있다. 도 5의 (a)에서 발열과 양성 및 음성 검사 결과 사이의 관계를 살펴본다.
- [0057] 시각화 결과를 보면 발열을 앓고 있는 환자들 중 실험실 검사에서 양성 반응을 보인 것을 알 수 있다. 또한, 도 5에서 독감, 기침 및 인후통으로 고통받는 환자에서 양성 실험실 테스트가 보고되었다.
- [0058] 과거 발열 병력이 있는 환자도 양성 대유행 증상을 보고했다. 시각화 결과, 발열, 독감, 두통, 호흡곤란, 기침, 인후통을 호소하는 환자들의 결과가 대부분 양성인 것을 관찰했다.
- [0059] 도 5는 검사실 결과가 양성 및 음성인 남성과 여성 환자의 분포를 보여준다.

- [0060] y축에는 케이스의 수가 표시되고 x축에는 연령 분포를 볼 수 있다. 도 6의 (a)와 도 6의 (b)의 가시화 결과를 보면, 실험실 결과가 양성인 남성과 여성이 대부분 40 내지 60세임을 분석할 수 있다.
- [0061] 마찬가지로 도 6의 (c)와 도 6의 (d)에서 20세 내지 40세 사이의 젊은 남성 및 여성 환자는 실험실 결과가 음성이다.
- [0062] 데이터 세트의 변수 간에 복잡하고 알려지지 않은 관계가 있을 수 있다. 데이터 세트 변수가 서로 의존하는 정도를 발견하고 정량화하는 것이 필수적이다.
- [0063] 서로 다른 속성 간의 관계를 표시하기 위해서 도 7에 표시된 상관 그래프를 사용했다. 두 변수가 같은 방향으로 움직이면 상관관계가 양수이고 한 변수의 값이 증가하면 다른 변수의 값이 감소하면 음의 상관관계가 있을 수 있다.
- [0065] 본 발명의 실시예에 따른 예측 분석부(150)는 복수의 질병 증상을 인공 신경 처리망에 입력하고, 인공 신경 처리망의 응답으로 질병 증상 정보에 대응하는 코로나 19 바이러스 감염 결과 여부를 예측한다.
- [0066] 예측 분석부(150)는 질병 증상 데이터의 주요 패턴과 추세를 고려하는 머신 또는 딥 러닝 모델에 입력된다.
- [0067] 데이터 세트는 훈련 세트와 테스트 세트로 더 나뉜다. 훈련 세트는 머신 러닝 또는 딥 러닝 모델에 제공된다.
- [0068] 본 발명의 실시예에 따른 예측 분석부(150)는 학습 모델부(160) 및 인공 신경 처리망(170)을 포함한다.
- [0069] 학습 모델부(160)는 데이터 처리부(161), 학습부(162) 및 분류부(163)를 포함한다.
- [0070] 인공 신경 처리망(170)은 입력층(171), 컨볼루션 레이어부(173), 풀링 레이어부(174) 및 풀리 커넥티드 레이어부(175)로 이루어진 은닉층(172), 출력층(176)을 포함한다.
- [0071] 데이터 처리부(161)는 데이터 수집부(110)로부터 데이터 세트에서 복수의 질병 증상을 이루어진 훈련 세트(Train Set)를 수신하여 인공 신경 처리망(170)로 전송한다. 훈련 세트는 학습 데이터를 나타낸다.
- [0072] 데이터 처리부(161)는 데이터 수집부(110)로부터 데이터 세트에서 복수의 질병 증상을 이루어진 테스트 세트(Test Set)를 수신하여 분류부(163)로 전송한다.
- [0073] 데이터 처리부(161)는 분산 병렬 처리가 가능한 데이터베이스부로 형성될 수 있다.
- [0074] 인공 신경 처리망(170)은 복수의 질병 증상을 이루어진 훈련 세트(Train Set)를 인공 신경 처리망(170)에 입력하여 적용시켜 오류를 수정하고, 수정된 오류를 이용하여 감염된 코로나 19 바이러스 또는 비감염된 코로나 19 바이러스 결과 여부를 출력한다.
- [0075] 이때, 인공 신경 처리망(170)은 심층 컨볼루션 신경망(Deep Convolutional Neural Networks, CNNs)을 이용하고, 입력층(171), 은닉층(172), 출력층(176)을 포함할 수 있다.
- [0076] 인공 신경 처리망(170)은 예측 분석을 위해 신경망 기반 모델을 사용한다.
- [0077] 신경망 모델의 전체 아키텍처는 도 6에 설명되어 있다. 인공 신경 처리망(170)은 입력층(171) x, 출력층(176) y 및 4개의 뉴런을 포함하는 임의의 양의 은닉층(172)을 포함한다.
- [0078] 각 계층은 출력층(176)을 제외하고 밴드 W로 표시되는 편향 및 가중치 세트로 구성됩니다. 각 은닉층의 활성화 함수로 시그모이드 함수를 사용한다. 모델의 예측 점수를 향상시키기 위해 입력 데이터의 편향 및 가중치 미세조정이 수행됩니다. 훈련 과정에서 각 반복에는 다음 단계가 포함됩니다.
- [0079] 도 8은 예측된 출력층(176) y의 계산을 포함하는 피드포워드(Feed-forward)와, 가중치와 편향을 업데이트하는 역전파(Back-propagation)의 두 단계로 구성된 신경망 모델의 훈련 과정을 보여준다.
- [0080] 출력층(176) y는 는 하기의 수학식 1과 같이 나타낸다.

수학식 1

$$y = \sigma(W_2 \sigma(W_1 x + b_1) + b_2)$$

여기서, σ 는 활성화함수(Activation Function)이고, W 는 가중치이고 b 는 출력 y 의 값에 영향을 미치는 유일한 변수인 편향을 나타낸다.

인공 신경 처리망(170)은 손실 함수 추정을 위해 하기의 수학식 2의 제곱 오차의 합으로 계산된 단순 손실 함수를 사용한다. 실제값과 각 예측값의 차이의 합이다.

수학식 2

$$\text{Sum of square Error} = \sum_{i=1}^n (y - y')^2$$

인공 신경 처리망(170)은 예측 손실의 오차를 측정하기 위해 역전파가 수행하고, 예측 오차(손실)를 측정한다.

편향과 가중치에 대한 손실 함수의 미분은 가중치와 편향을 조정하기 위해 사용된다. 경사하강법(gradient descent)이라고도 하며, 하기의 수학식 3으로 계산된다.

수학식 3

$$\text{Loss}(y, y') = \sum_{i=1}^n (y - y')^2$$

체인 법칙(Chain Rule)은 손실 함수의 도함수를 추정하는데 사용되며, 수학적으로 (수학식 4식에서 $z = Wx + b$) 수학식 4로 정의된다.

수학식 4

$$\frac{\delta \text{Loss}(y, y')}{\delta W} = \frac{\delta \text{Loss}(y, y')}{\delta y} * \frac{\delta y}{\delta z} * \frac{\delta y'}{\delta W}$$

입력층(171)은 데이터 처리부(161)에 저장된 학습 데이터를 획득하고, 획득한 학습 데이터를 특징맵을 가지는 레이어로 저장한다. 여기서, 특징맵은 다수의 노드들이 2차원으로 배열된 구조를 가짐으로써 후술되는 은닉층(172)과의 연결을 용이하게 할 수 있다.

은닉층(172)은 상위 계층에 위치한 레이어의 특징맵을 획득하고, 획득한 특징맵으로부터 점차 높은 수준의 특징을 추출한다. 은닉층(172)은 하나 또는 그 이상으로 형성될 수 있으며 컨볼루션 레이어부(173), 풀링 레이어부(174) 및 풀리 커넥티드 레이어부(175)를 포함한다.

컨볼루션 레이어부(173)는 학습 데이터로부터 컨볼루션 연산을 수행하는 구성으로서, 복수의 입력 특징맵과 연결되는 특징맵을 포함한다.

풀링 레이어부(174)는 컨볼루션 레이어부(173)의 출력을 입력으로 받아 컨볼루션 연산, 즉 서브 샘플링 연산을 수행하는 구성이고, 은닉층(172)의 하위 계층에 위치한 컨볼루션 레이어부(173)가 가지는 입력 특징맵의 수와

동일한 수의 특징맵을 포함하며, 각각의 특징맵은 입력 특징맵과 일대일로 연결된다.

- [0094] 폴리 커넥티드 레이어부(175)는 컨볼루션 레이어부(173)의 출력을 입력으로 받아 출력층(130)에서 출력되는 각 카테고리별 출력에 맞게 학습하는 구성이고, 학습된 국소적 정보, 즉 특징들을 종합하여 추상적인 내용을 학습한다.
- [0095] 이때, 은닉층(172)이 풀링 레이어부(172)를 구비할 경우, 풀링 커넥티드 레이어부(175)는 풀링 레이어부(174)와 연결되며, 풀링 레이어부(174)의 출력으로부터 특징들을 종합하여 추상적인 내용을 학습한다.
- [0096] 출력층(176)은 소프트 맥스(soft-max) 등의 함수를 이용하여 분류하고자 원하는 각 카테고리별 출력을 확률값으로 매핑한다. 이때, 출력층(176)에서 출력된 결과는 학습부(162) 또는 분류부(163)로 전달되어 오류역전파를 수행하거나 응답 데이터로 출력될 수도 있다.
- [0097] 학습부(162)는 지도 학습을 수행하는 것으로, 지도 학습은 학습 데이터에 기계학습 알고리즘을 적용하여 함수를 추론하고, 그 추론된 함수를 통해 해답을 찾는다.
- [0098] 학습부(162)는 지도 학습을 통해서 학습 데이터를 대표하는 선형 모델을 생성하고, 그 선형 모델을 통해 미래의 사건을 예측할 수 있다.
- [0099] 학습부(162)는 이전까지 학습된 데이터를 근거로 새로운 데이터가 기존에 학습된 데이터에 어떻게 분류되는지 판단한다.
- [0100] 학습부(162)는 데이터 처리부(161)로부터 데이터 세트에서 복수의 질병 증상을 이루어진 훈련 세트(Train Set)를 인공 신경 처리망(170)의 학습을 수행하고, 유형별 딥러닝 특징값을 이용하여 감염된 코로나 19 바이러스 인식 여부를 학습한다.
- [0101] 학습부(162)는 유형별 딥러닝 특징값을 이용하여 감염된 코로나 19 바이러스 인식 여부를 인공 신경 처리망(170)에서 학습한다.
- [0102] 본 발명의 일실시예에서 인공 신경 처리망(170)의 학습은 지도 학습(supervised-learning)으로 이루어진다.
- [0103] 지도 학습은 학습 데이터와 그에 대응하는 출력 데이터를 함께 인공 신경 처리망(170)에 입력하고, 학습 데이터에 대응하는 출력 데이터가 출력되도록 연결된 간선들의 가중치를 업데이트 하는 방법이다. 일례로, 본 발명의 인공 신경 처리망(170)은 델타 규칙 및 오류역전파 학습 등을 이용하여 인공뉴런들 사이의 연결 가중치를 업데이트 할 수 있다.
- [0104] 오류역전파(Error-back-propagation) 학습은 주어진 학습 데이터에 대해 전방계산(Feed-Forward)으로 오류를 추정한 후, 출력 레이어에서 시작하여 은닉층(172)과 입력층(171) 방향인 역방향으로 추정한 오류를 전파하고, 오류를 줄이는 방향으로 인공 뉴런들 사이의 연결 가중치를 업데이트한다.
- [0105] 인공 신경 처리망(170)은 입력층(171) - 은닉층(172) - 풀링 커넥티드 레이어부(175) - 출력층(176)을 통해 획득된 결과로부터 오차를 계산하고, 계산된 오차를 보정하기 위해 다시 출력층(176) - 풀링 커넥티드 레이어부(175) - 은닉층(172) - 입력층(171)의 순서로 오류를 전파하여 연결 가중치를 업데이트할 수 있다.
- [0106] 학습부(162)는 복수의 질병 증상을 이루어진 훈련 세트(Train Set)를 입력 벡터가 되며, 입력층(171), 은닉층(172), 출력층(176)을 통과하면, 복수의 질병 증상에 대응하는 감염된 코로나 19 바이러스 인식 여부를 출력 벡터로 생성하도록 지도 학습을 통해 학습된다.
- [0107] 학습부(162)는 인공 신경 처리망(170)을 이용하여 입력되는 복수의 질병 증상 정보의 특징값들이 입력 벡터가 되며, 입력층(171), 은닉층(172), 출력층(176)을 통과하면, 질병 증상에 대응하는 감염된 코로나 19 바이러스 결과를 출력 벡터로 생성하도록 지도 학습을 통해 학습된다.
- [0108] 학습부(162)는 복수의 질병 증상과, 질병 증상에 대응하는 감염된 코로나 19 바이러스 결과를 학습 데이터로 하여 인공 신경 처리망(170)과 연동하여 인공지능에 학습한다.
- [0109] 인공 신경 처리망(170)은 입력값(복수의 질병 증상 정보)이 입력되면, 출력값(질병 증상 정보에 대응하는 코로나 19 바이러스 감염 결과)이 나와야 하는지 미리 알고 있다.
- [0110] 분류부(163)는 학습부(162)에서의 오류역전파를 통해 업데이트된 연결 가중치를 가지는 인공 신경 처리망(170)의 출력 데이터를 응답 데이터로 출력할 수 있다.

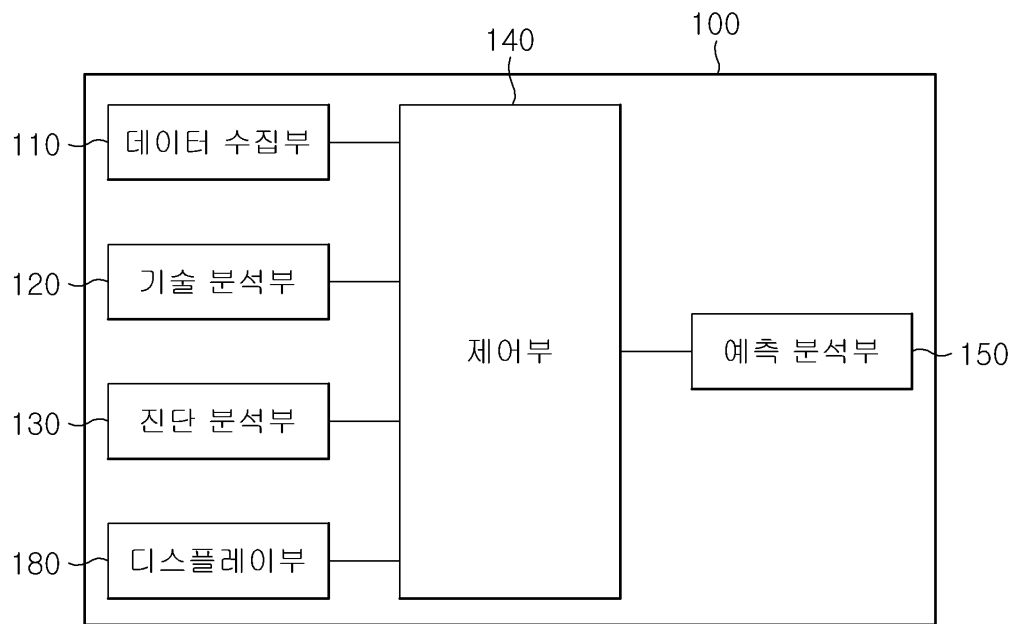
- [0111] 분류부(163)는 업데이트된 연결 가중치를 가지는 인공 신경 처리망(170)에 학습 데이터, 테스트 데이터 또는 학습에 사용되지 않은 새 데이터가 입력되면, 입력층(171) - 은닉층(172) - 풀링 커넥티드 레이어부(175) - 출력층(176)을 통해 출력된 결과를 획득하여 응답 데이터로 출력할 수 있다.
- [0112] 인공 신경 처리망(170)은 입력된 복수의 질병 증상 정보와, 질병 증상 정보에 대응하는 코로나 19 바이러스 감염 결과 여부를 기반으로 최적화를 통해 딥러닝 기반 분류기 모델을 생성한다.
- [0113] 학습부(162)는 각각의 질병 증상을 분석하여 통계를 퍼센트율로 계산하고, 계산한 퍼센트율을 기초로 각각의 질병 증상의 가중치를 다르게 설정하고, 다르게 설정된 가중치를 입력 벡터인 각각의 질병 증상과 복수의 연산을 수행하여 입력층(171), 은닉층(172), 출력층(176)을 통과한다.
- [0114] 학습부(162)는 코로나 19 바이러스 감염 결과를 출력 벡터로 생성하도록 지도 학습을 통해 학습되고, 입력층(171)에서 출력층(176)으로 방향으로 계산하고, 반대로 출력층(176)에서 입력층(171) 방향으로 계산하는 작업을 반복하면서 가중치를 수정하여 오차를 최소화한다.
- [0115] 분류부(163)는 테스트 데이터인 복수의 질병 증상 정보를 인공 신경 처리망(170)의 딥러닝 기반 분류기 모델을 이용하여 응답 데이터의 결과값(감염된 코로나 19 바이러스 또는 비감염된 코로나 19 바이러스)으로 출력한다.
- [0116] 분류부(163)는 복수의 질병 증상 정보에 대응하여 감염된 코로나 19 바이러스 또는 비감염된 코로나 19 바이러스 결과 여부를 판단한다.
- [0117] 분류부(163)는 테스트 데이터인 복수의 질병 증상 정보를 인공 신경 처리망(170)의 딥러닝 기반 분류기 모델을 이용하여 감염된 코로나 19 바이러스 또는 비감염된 코로나 19 바이러스 결과 여부를 판단한다.
- [0118] 분류부(163)는 복수의 질병 증상 정보에 대응하여 감염된 코로나 19 바이러스인 제1 출력값 확률이 제1 퍼센트율인 경우, 복수의 질병 증상 정보에 대응하는 비감염된 코로나 19 바이러스인 제2 출력값 확률 (100 - 제1 퍼센트율)을 계산한다.
- [0119] 분류부(163)는 복수의 질병 증상 정보에 대응하여 감염된 코로나 19 바이러스인 제1 출력값이 0.87인 경우, 복수의 질병 증상 정보에 대응하는 비감염된 코로나 19 바이러스인 제2 출력값이 0.13이 된다.
- [0120] 이러한 경우, 감염된 코로나 19 바이러스일 확률이 87%이고, 비감염된 코로나 19 바이러스일 확률이 13%가 된다.
- [0121] 출력부(164)는 분류부(163)로부터 수신된 질병 증상 정보에 대응하는 코로나 19 바이러스 감염 결과 여부를 표시한다.
- [0122] 출력부(164)는 인공 신경 처리망(170)에 적용된 결과를 바탕으로 복수의 질병 증상 정보, 코로나 19 바이러스 감염 결과 여부를 표시한다.
- [0123] 제어부(140)는 예측 분석의 결과를 더 나은 평가 및 예측 목적으로 추가로 사용된다. 예를 들어, 결과에서 독감, 발열, 기침, 인후통 등의 증상이 있는 환자는 COVID-19에 걸릴 확률이 더 높다는 결론을 내릴 수 있다.
- [0124] 이상에서 본 발명의 실시예에 대하여 상세하게 설명하였지만 본 발명의 권리범위는 이에 한정되는 것은 아니고 다음의 청구범위에서 정의하고 있는 본 발명의 기본 개념을 이용한 당업자의 여러 변형 및 개량 형태 또한 본 발명의 권리범위에 속하는 것이다.

부호의 설명

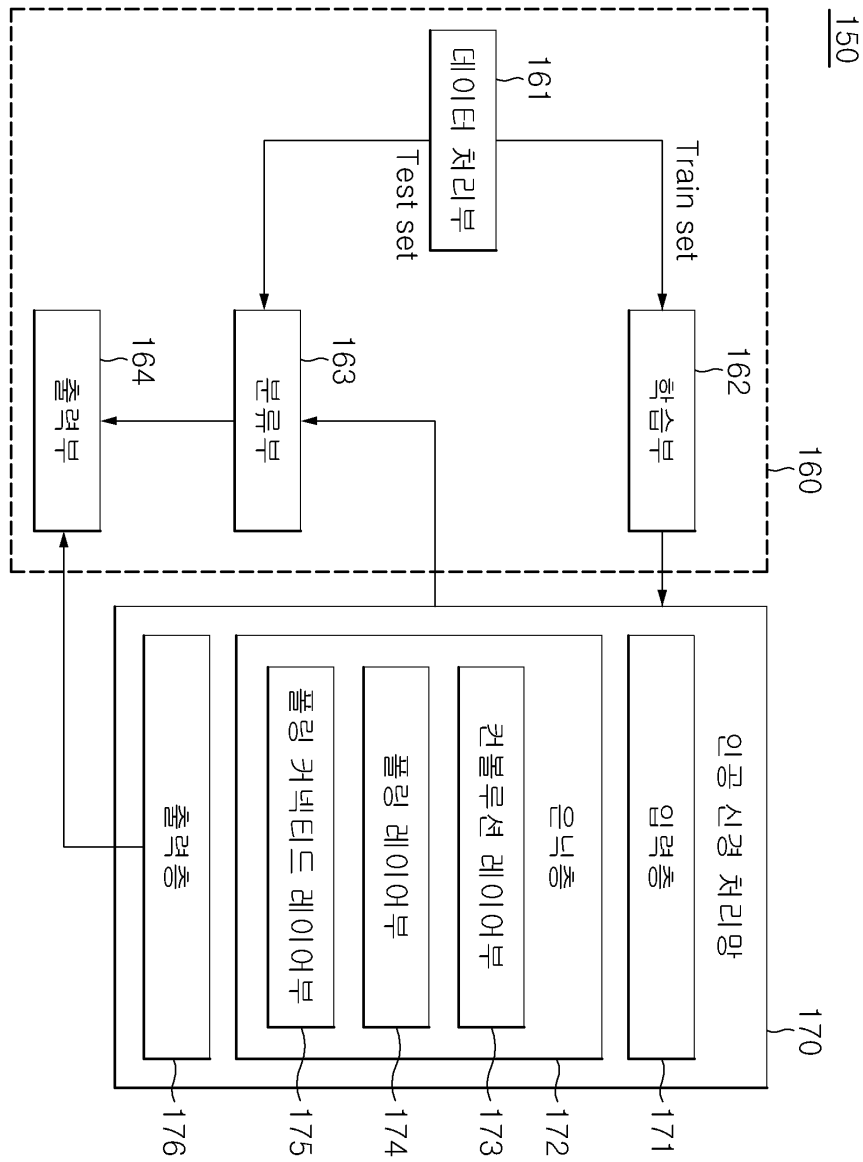
- [0125] 100: 분석 시스템
110: 데이터 수집부
120: 기술 분석부
130: 진단 분석부
140: 제어부
150: 예측 분석부
180: 디스플레이부

도면

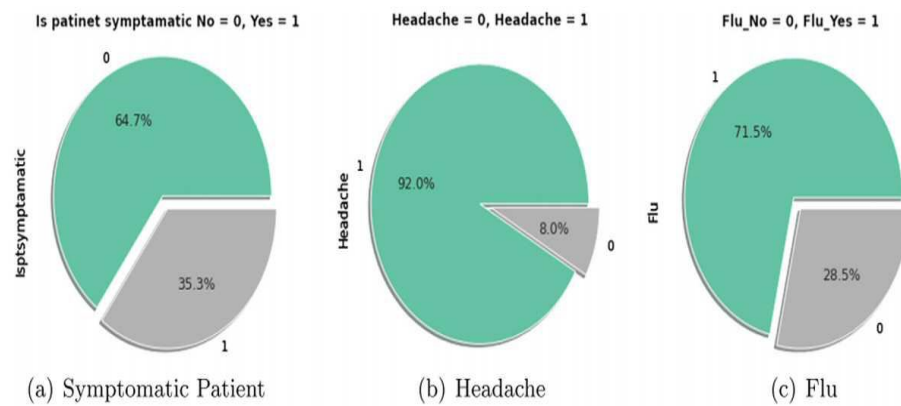
도면1



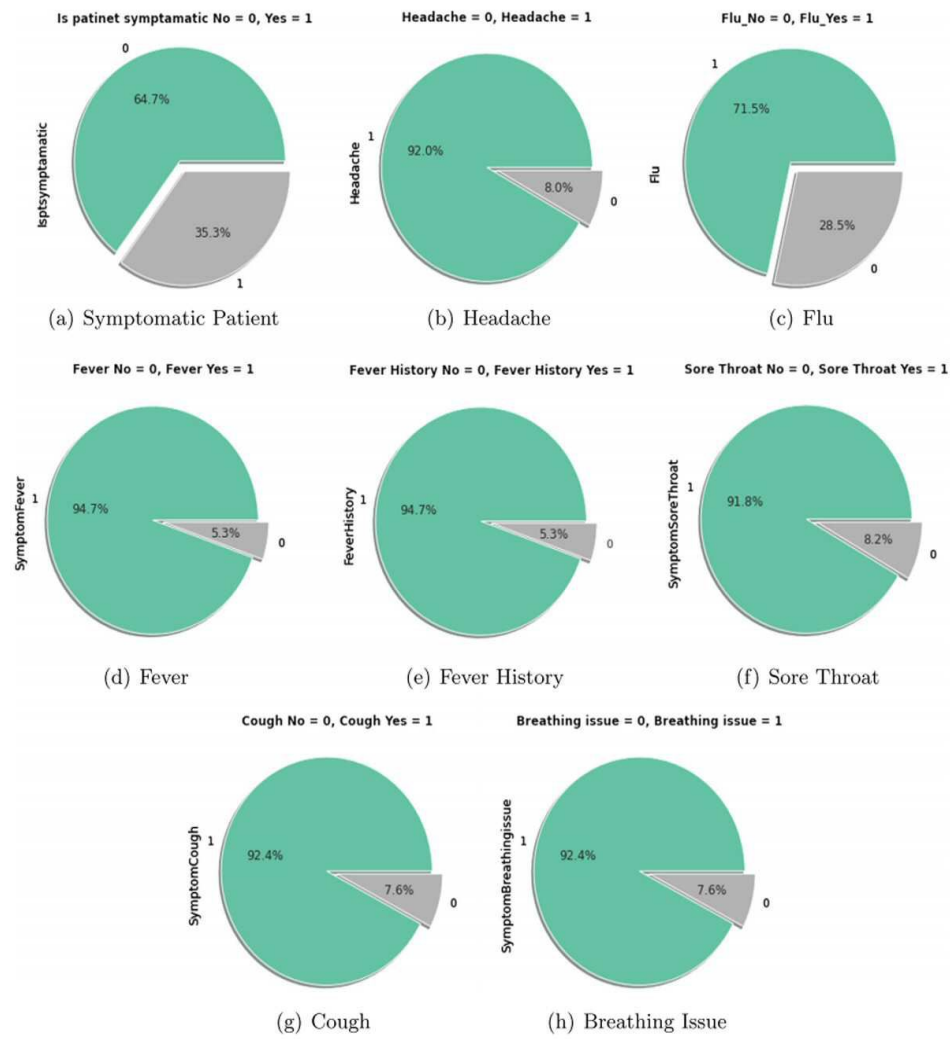
도면2



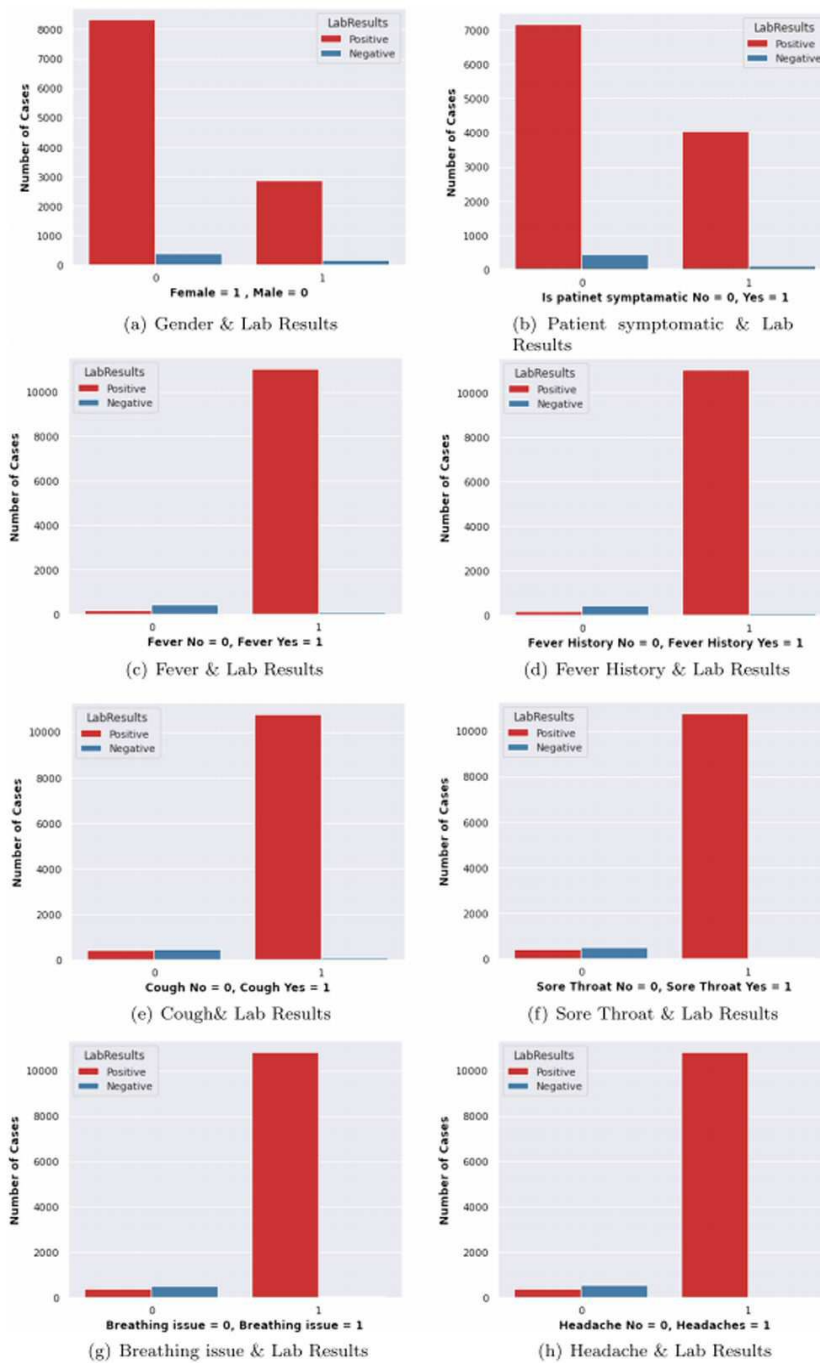
도면3



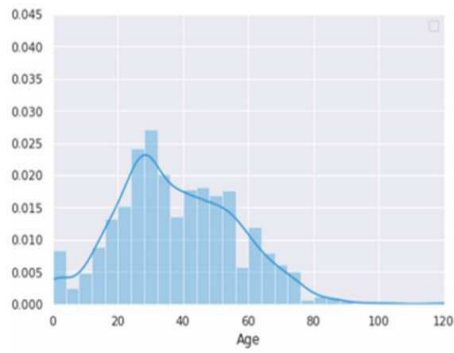
도면4



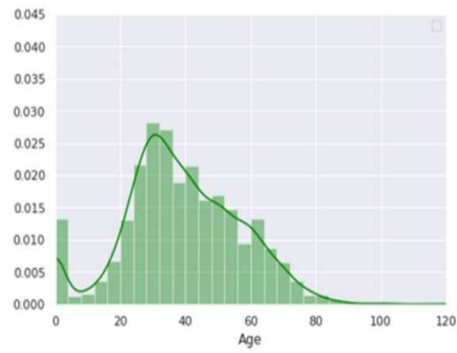
도면5



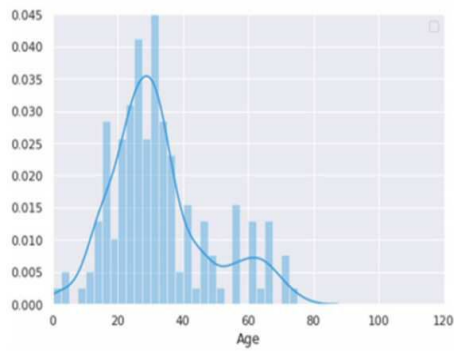
도면6



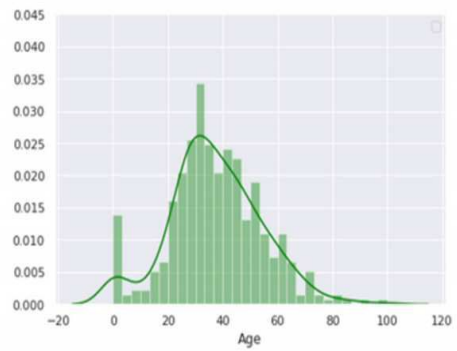
(a) Female with Positive Lab results



(b) Male with Positive Lab results

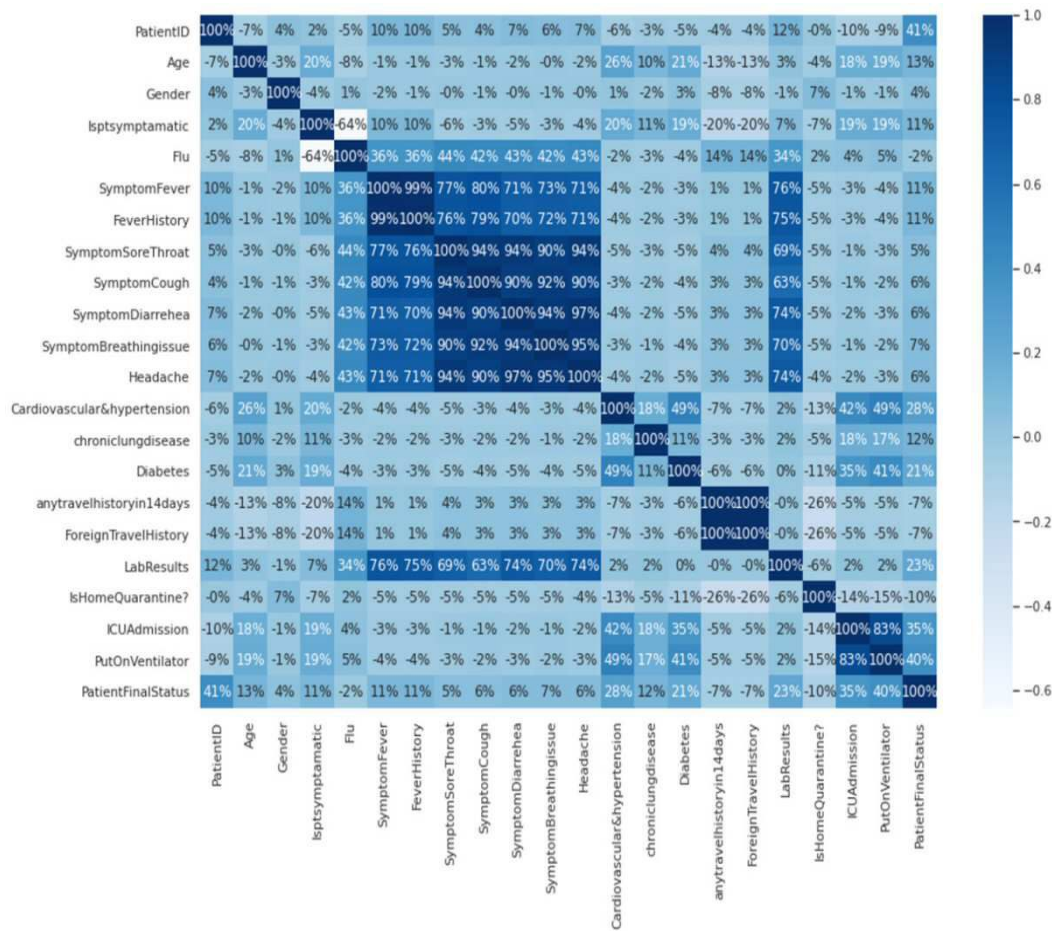


(c) Female with Negative Lab results



(d) Male with Negative Lab results

도면7



도면8

