



(19) 대한민국특허청(KR)
(12) 공개특허공보(A)

(11) 공개번호 10-2019-0069047
(43) 공개일자 2019년06월19일

(51) 국제특허분류(Int. Cl.)

A61B 5/00 (2006.01)

(52) CPC특허분류

A61B 5/7275 (2013.01)

G16H 50/20 (2018.01)

(21) 출원번호 10-2017-0169399

(22) 출원일자 2017년12월11일

심사청구일자 2017년12월11일

(71) 출원인

한림대학교 산학협력단

강원도 춘천시 한림대학길 1, 한림대학교(옥천동)

(72) 발명자

김의직

강원도 춘천시 지석로 29, 514동 1103호 (석사동, 휴먼타운아파트)

이솔비

강원도 춘천시 후석로326번길 31, 303동 501호(후평동, 현대3차아파트)

권정혁

강원도 춘천시 스무숲1길 32-9, 202호 (석사동)

(74) 대리인

유민규

전체 청구항 수 : 총 9 항

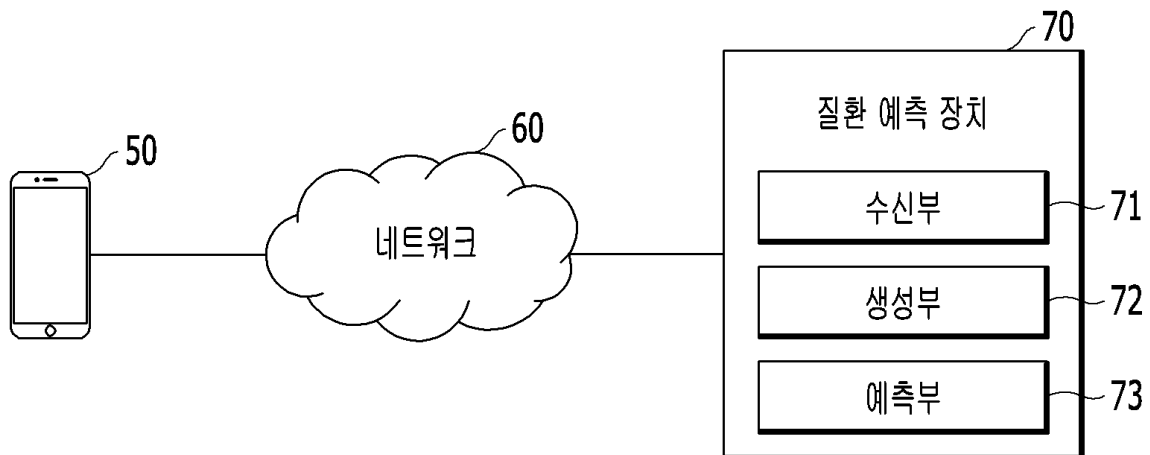
(54) 발명의 명칭 질환 예측 장치 및 방법

(57) 요약

질환 예측 방법에 관한 것이며, 질환 예측 방법은, 예측 대상자의 건강 데이터를 수신하는 단계; 전자건강기록(electronic health records, EHR) 및 모바일 기기를 통해 획득된 모바일 기반의 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하는 단계; 및 상기 질환 예측 모델에 기초하여 상기 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측하는 단계를 포함할 수 있다.

대표도 - 도3

200



이 발명을 지원한 국가연구개발사업

과제고유번호 NRF-2017R1D1A1B03031055

부처명 교육부

연구관리전문기관 한국연구재단

연구사업명 이공학개인지초연구지원사업(기본연구)

연구과제명 산업환경 안전을 위한 사물인터넷 빅데이터 통합 프레임워크 개발

기 여 율 1/1

주관기관 한림대학교 산학협력단

연구기간 2017.06.01 ~ 2018.02.28

명세서

청구범위

청구항 1

질환 예측 방법에 있어서,

예측 대상자의 건강 데이터를 수신하는 단계;

전자건강기록(electronic health records, EHR) 및 모바일 기기를 통해 획득된 모바일 기반의 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하는 단계; 및

상기 질환 예측 모델에 기초하여 상기 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측하는 단계,

를 포함하는 질환 예측 방법.

청구항 2

제1항에 있어서,

상기 생성하는 단계는,

상기 건강 데이터가 수신된 이후에 상기 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 상기 질환 예측 모델을 생성하는 것인, 질환 예측 방법.

청구항 3

제1항에 있어서,

상기 생성하는 단계는,

상기 통합 데이터셋에 포함된 데이터를 기반으로 복수 사용자의 특성이 고려된 복수의 질환 예측 모델을 생성하고, 상기 건강 데이터가 수신된 이후에 상기 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 상기 복수의 질환 예측 모델 중 상기 예측 대상자에 최적화된 최적 질환 예측 모델을 선택하는 단계를 포함하고,

상기 예측하는 단계는,

상기 최적 질환 예측 모델에 기초하여 상기 예측 대상자에 대한 질환 가능성을 예측하는 것인, 질환 예측 방법.

청구항 4

제1항에 있어서,

상기 질환 예측 모델은, 상기 통합 데이터셋에 포함된 데이터의 복수의 속성과 관련된 복수의 규칙을 포함하고,

상기 복수의 규칙의 조합 수 및 조합 순서는 상기 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 결정되는 것인, 질환 예측 방법.

청구항 5

제1항에 있어서,

상기 질환 예측 모델은,

상기 통합 데이터셋에 포함된 데이터에 대하여 의사결정트리(decision tree)를 적용함으로써 생성되는 것인, 질환 예측 방법.

청구항 6

제1항에 있어서,

상기 통합 데이터셋은,

상기 전자건강기록을 포함하는 제1 데이터셋 및 상기 모바일 기기를 통해 획득되는 개인건강기록을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하고, 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 생성되는 것인, 질환 예측 방법.

청구항 7

질환 예측 장치에 있어서,

예측 대상자의 건강 데이터를 수신하는 수신부;

전자건강기록(electronic health records, EHR) 및 모바일 기기를 통해 획득된 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하는 생성부; 및

상기 질환 예측 모델에 기초하여 상기 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측하는 예측부,

를 포함하는 질환 예측 장치.

청구항 8

질환 예측 시스템에 있어서,

예측 대상자의 건강 데이터를 제공하는 제1 단말 기기; 및

상기 제1 단말 기기로부터 상기 건강 데이터를 수신하고, 전자건강기록(electronic health records, EHR) 및 모바일 기기를 통해 획득된 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하고, 상기 질환 예측 모델에 기초하여 상기 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측하고, 예측 결과를 제공하는 질환 예측 장치,

를 포함하는 질환 예측 시스템.

청구항 9

제1항 내지 제6항 중 어느 한 항의 방법을 컴퓨터에서 실행하기 위한 프로그램을 기록한 컴퓨터에서 판독 가능한 기록 매체.

발명의 설명

기술 분야

[0001]

본원은 질환 예측 장치 및 방법에 관한 것이다.

배경 기술

[0002]

최근 하드 카피 형식으로 저장된 많은 양의 의료 데이터가 의료 산업에서 빠르게 디지털화되어 축적되고 있다. 또한, 사물 인터넷(Internet of Things, IoT)을 기반으로 하는 의료 기기 및 플랫폼의 급속한 개발과 함께 모바일 기기의 보급이 증가함에 따라 개인 사용자를 위한 의료 빅데이터 수집이 가능하게 되었다. 많은 조직과 병원 은 디지털화된 전자건강기록(electronic health records, EHR)의 방대한 수집으로부터 가치 있는 정보를 얻을 수 있다. 최근 몇 년 동안 의미 있는 정보를 얻기 위해 EHR을 이용한 다양한 의학적 연구가 수행되었다. 즉, 의료 빅데이터 분석의 중요성에 대한 인식이 확산되고 있으며, 이에 따라 데이터 마이닝과 같은 기술을 사용하여 방대한 양의 데이터에서 의미 있는 정보를 효과적으로 찾아야 할 필요가 있다.

[0003]

대부분의 의료기관에서 간호의 질과 환자의 건강을 향상시키기 위해 사용되는 EHR은 전자적으로 기록된 환자의 전반적인 건강 관리와 관련된 건강 정보(데이터)로 이루어진다. 이러한 EHR과 같은 의료용 빅데이터(big data)는 다양한 데이터 유형과 고차원 데이터의 존재로 인해 기존의 방법으로 분석하기에 어려운 문제가 있다. 더욱이, EHR은 후술할 mPHR에 비해 상대적으로 정적이기 때문에 EHR을 사용하는 것만으로는 환자의 현재 건강 상태를 파악하기 어려운 단점이 있다. 그럼에도 불구하고, EHR에는 의료 전문가가 제공한 진료 및 처방 정보가 포함

되어 있기 때문에, 환자로 하여금 의료 제공자가 보다 높은 수준의 의료 서비스를 제공할 수 있도록 한다는 점에서 큰 장점이 있다.

[0004] 한편, 모바일 기반 개인건강기록(mobile personal health records, mPHR)은 모바일 의료 기기를 통해 수집되기 때문에 데이터의 정확성과 신뢰성을 확인하기 어려운 문제가 있다. 이는 mPHR 데이터가 의료 전문가에 의해 수집되거나 병원의 의료 기기를 사용하여 수집되지 않기 때문이라 할 수 있다. 이러한 mPHR은 질병과 관련된 심층 분석에 사용하기 어려운 측면이 있다. 그러나, mPHR의 경우에는 환자의 상태가 EHR에 비해 상대적으로 짧은 기간에 측정되어 업데이트됨에 따라 환자의 현재 상태가 보다 정확하게 반영된다는 점에서 큰 장점이 있다. 종래에는 이러한 EHR 및 mPHR이 갖는 한계를 고려하여 향상된 의료 서비스를 제공하기 위해 EHR 및 mPHR과 관련된 광범위한 연구가 다양한 측면에서 수행된 바 있다.

[0005] 일례로, 논문 ["A hybrid outlier detection method for health care big data", Ke Yan, Xiaoming You, Xiaobo Ji, Guangqiang Yin, Fan Yang, 2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom) (2016), pp: 157-162]에서는 프루닝(pruning) 기반의 K-Nearest Neighbor(PB-KNN)이라 불리는 새로운 hybrid outlier detection 방법을 제안했다. 이 방법은 밀도 기반, 클러스터 기반 방법 및 KNN 알고리즘을 통합함으로써 많은 양의 데이터, 다양한 데이터 유형 및 고차원의 데이터가 포함되어 있는 의료 분야에서의 데이터 분석의 어려움을 극복한다. 그러나 선행 논문에서 제안하는 방법은 EHR만을 사용하기 때문에 사용자의 현재 상태에 대한 반영이 어려워 분석에 한계가 있다.

[0006] 또한, 논문 ["Outlier detection for patient monitoring and alerting", Milos Hauskrecht, Iyad Batal, Michal Valko, Shyam Visweswaran, Gregory F Cooper, Gilles Clermont, Journal of Biomedical Informatics, 46권 1호, (2013) pp: 47-55]에서는 새로운 데이터 기반 모니터링 및 경고 프레임워크에 대하여 제안했다. 상기 논문에서는 EHR에 저장된 과거 환자 사례를 사용하여 의료 이상치 정보를 검출하는 것을 목적으로 한다. 그런데, 상기 논문의 기술에서는 이전에 기록된 EHR만을 활용하기 때문에 환자의 실제 상태(actual status)를 확인하는 데에 어려움이 있다. 즉, 상기의 논문 또한 사용자의 현재 상태에 대한 반영이 어려워 분석에 한계가 있다.

발명의 내용

해결하려는 과제

[0007] 본원은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 의료 서비스를 제공함에 있어서 EHR 활용 시 환자의 현재의 상태를 확인하는데 어려웠던 문제와 mPHR 활용 시 데이터의 정확성과 신뢰성을 확인하기 어려웠던 문제를 해소할 수 있는 EHR과 mPHR이 통합된 통합 데이터셋을 제공하려는 것을 목적으로 한다.

[0008] 본원은 전술한 종래 기술의 문제점을 해결하기 위한 것으로서, 사용자의 현재 상태를 반영하면서 보다 정확하고 신뢰성 있는 건강 관련 진단 내지 분석이 이루어질 수 있도록 하는 통합 데이터셋 생성 장치 및 방법과 질환 예측 장치 및 방법을 제공하려는 것을 목적으로 한다.

[0009] 다만, 본원의 실시예가 이루고자 하는 기술적 과제는 상기된 바와 같은 기술적 과제들로 한정되지 않으며, 또 다른 기술적 과제들이 존재할 수 있다.

과제의 해결 수단

[0010] 상기한 기술적 과제를 달성하기 위한 기술적 수단으로서, 본원의 제1 측면에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 방법은, 전자건강기록(electronic health records, EHR)을 포함하는 제1 데이터셋 및 모바일 기기를 통해 획득되는 모바일 기반 개인건강기록(mobile personal health records, mPHR)을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하는 단계; 및 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 상기 전처리된 제1 데이터셋의 데이터 및 상기 전처리된 제2 데이터셋의 데이터가 통합된 통합 데이터셋을 생성하는 단계를 포함할 수 있다.

[0011] 또한, 상기 전처리를 수행하는 단계는, 상기 제1 데이터셋 및 상기 제2 데이터셋 각각에 포함된 데이터에 대하여, 누락 값 또는 이상치가 존재하는 것으로 판단되는 경우, 상기 누락 값 또는 이상치가 속한 데이터셋 내에서 상기 누락 값 또는 이상치에 대응하는 속성과 동일 속성에 속하는 속성 값들의 평균값으로 대체하는 전처리를 수행할 수 있다.

- [0012] 또한, 상기 생성하는 단계는, 상기 제1 데이터셋의 데이터 속성과 상기 제2 데이터셋의 데이터 속성 간의 동일 여부를 고려한 상기 매칭을 통해 상기 통합 데이터셋을 생성할 수 있다.
- [0013] 또한, 상기 생성하는 단계는, 상기 제1 데이터셋의 데이터 속성과 상기 제2 데이터셋의 데이터 속성이 동일하지 않는 것으로 판단되는 경우, 상기 제2 데이터셋의 데이터 속성을 상기 제1 데이터셋의 데이터 속성과 결합시키고, 상기 제1 데이터셋의 데이터 속성과 상기 제2 데이터셋의 데이터 속성이 동일한 것으로 판단되는 경우, 상기 제1 데이터셋의 데이터 속성을 상기 제2 데이터셋의 데이터 속성으로 덮어씌울 수 있다.
- [0014] 또한, 본원의 제1 측면에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 방법은, 상기 전처리를 수행하는 단계 이전에, 상기 제1 데이터셋 및 상기 제2 데이터셋을 생성하는 단계를 더 포함할 수 있다.
- [0015] 또한, 상기 제1 데이터셋 및 제2 데이터셋을 생성하는 단계는, UCI(University of California-Irvine) 기계 학습 저장소(Machine Learning Repository)에서 제공하는 데이터로서 서로 다른 장소를 갖는 복수의 데이터셋에 속한 데이터를 통합하여 상기 제1 데이터셋을 생성하고, 모바일 기기를 통해 획득되는 데이터의 속성이 상기 제1 데이터셋의 속성과 매치되도록 상기 제2 데이터셋을 생성할 수 있다.
- [0016] 본원의 제2 측면에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 장치는, 전자건강기록(electronic health records, EHR)을 포함하는 제1 데이터셋 및 모바일 기기를 통해 획득되는 개인건강기록(mobile personal health records, mPHR)을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하는 전처리부; 및 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 상기 전처리된 제1 데이터셋의 데이터 및 상기 전처리된 제2 데이터셋의 데이터가 통합된 통합 데이터셋을 생성하는 매칭부를 포함할 수 있다.
- [0017] 본원의 제3 측면에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 시스템은, 복수의 사용자에게 대한 전자건강기록(electronic health records, EHR)을 제공하는 복수의 의료기관의 장치; 복수의 사용자 각각에 대한 모바일 기반의 개인건강기록(mobile personal health records, mPHR)을 측정하여 제공하는 복수의 모바일 기기; 및 상기 전자건강기록을 포함하는 제1 데이터셋 및 상기 모바일 기기를 통해 획득되는 상기 개인건강기록을 포함하는 제2 데이터셋을 생성하여 상기 제1 데이터셋 및 상기 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하고, 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 상기 전처리된 제1 데이터셋의 데이터 및 상기 전처리된 제2 데이터셋의 데이터가 통합된 통합 데이터셋을 생성하는 데이터셋 생성 장치를 포함할 수 있다.
- [0018] 본원의 제4 측면에 따른 컴퓨터 프로그램은, 본원의 제1 측면에 따른 데이터셋 생성 방법을 실행시키기 위하여 기록 매체에 저장되는 것일 수 있다.
- [0019] 본원의 제5 측면에 따른 질환 예측 방법은, 예측 대상자의 건강 데이터를 수신하는 단계; 전자건강기록(electronic health records, EHR) 및 모바일 기기를 통해 획득된 모바일 기반의 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하는 단계; 및 상기 질환 예측 모델에 기초하여 상기 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측하는 단계를 포함할 수 있다.
- [0020] 또한, 상기 생성하는 단계는, 상기 건강 데이터가 수신된 이후에 상기 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 상기 질환 예측 모델을 생성할 수 있다.
- [0021] 또한, 상기 생성하는 단계는, 상기 통합 데이터셋에 포함된 데이터를 기반으로 복수 사용자의 특성이 고려된 복수의 질환 예측 모델을 생성하고, 상기 건강 데이터가 수신된 이후에 상기 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 상기 복수의 질환 예측 모델 중 상기 예측 대상자에 최적화된 최적 질환 예측 모델을 선택하는 단계를 포함하고, 상기 예측하는 단계는, 상기 최적 질환 예측 모델에 기초하여 상기 예측 대상자에 대한 질환 가능성을 예측할 수 있다.
- [0022] 또한, 상기 질환 예측 모델은, 상기 통합 데이터셋에 포함된 데이터의 복수의 속성과 관련된 복수의 규칙을 포함하고, 상기 복수의 규칙의 조합 수 및 조합 순서는 상기 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 결정될 수 있다.
- [0023] 또한, 상기 질환 예측 모델은, 상기 통합 데이터셋에 포함된 데이터에 대하여 의사결정트리(decision tree)를 적용함으로써 생성될 수 있다.
- [0024] 또한, 상기 통합 데이터셋은, 상기 전자건강기록을 포함하는 제1 데이터셋 및 상기 모바일 기기를 통해 획득되

는 개인건강기록을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하고, 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 생성될 수 있다.

[0025] 본원의 제6 측면에 따른 질환 예측 장치는, 예측 대상자의 건강 데이터를 수신하는 수신부; 전자건강기록(electronic health records, EHR) 및 모바일 기기를 통해 획득된 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하는 생성부; 및 상기 질환 예측 모델에 기초하여 상기 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측하는 예측부를 포함할 수 있다.

[0026] 본원의 제7 측면에 따른 질환 예측 시스템은, 예측 대상자의 건강 데이터를 제공하는 제1 단말 기기; 및 상기 제1 단말 기기로부터 상기 건강 데이터를 수신하고, 전자건강기록(electronic health records, EHR) 및 모바일 기기를 통해 획득된 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하고, 상기 질환 예측 모델에 기초하여 상기 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측하고, 예측 결과를 제공하는 질환 예측 장치를 포함할 수 있다.

[0027] 본원의 제8 측면에 따른 컴퓨터 프로그램은, 본원의 제5 측면에 따른 질환 예측 방법을 실행시키기 위하여 기록 매체에 저장되는 것일 수 있다.

[0028] 상술한 과제 해결 수단은 단지 예시적인 것으로서, 본원을 제한하려는 의도로 해석되지 않아야 한다. 상술한 예시적인 실시예 외에도, 도면 및 발명의 상세한 설명에 추가적인 실시예가 존재할 수 있다.

발명의 효과

[0029] 전술한 본원의 과제 해결 수단에 의하면, EHR과 mPHR이 통합된 통합 데이터셋을 제공함으로써, 의료 서비스를 제공함에 있어서 EHR 활용시 환자의 현재의 상태를 확인하는데 어려웠던 문제와 mPHR 활용시 데이터의 정확성과 신뢰성을 확인하기 어려웠던 문제를 해소할 수 있다.

[0030] 전술한 본원의 과제 해결 수단에 의하면, 통합 데이터셋 및 통합 데이터셋을 이용하여 생성된 질환 예측 모델로 하여금, 사용자의 현재 상태를 반영하면서 보다 정확하고 신뢰성 있는 의료 서비스의 제공(즉, 건강 관련 진단 내지 분석)이 이루어질 수 있도록 할 수 있다.

[0031] 다만, 본원에서 얻을 수 있는 효과는 상기된 바와 같은 효과들로 한정되지 않으며, 또 다른 효과들이 존재할 수 있다.

도면의 간단한 설명

[0032] 도 1은 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 시스템의 개략적인 구성을 나타낸 도면이다.

도 2는 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 장치에서 고려되는 EHR의 속성을 나타낸 도면이다.

도 3은 본원의 일 실시예에 따른 질환 예측 시스템의 개략적인 구성을 나타낸 도면이다.

도 4는 본원의 일 실시예에 따른 통합 데이터셋에 의사결정트리를 적용함에 따라 생성된 질환 예측 모델의 예를 나타낸 도면이다.

도 5는 종래의 EHR에 의사결정트리를 적용함에 따라 생성된 질환 예측 모델의 예를 나타낸 도면이다.

도 6은 종래의 mPHR에 의사결정트리를 적용함에 따라 생성된 질환 예측 모델의 예를 나타낸 도면이다.

도 7은 EHR 또는 mPHR 대비 본원의 일 실시예에 따른 통합 데이터셋 기반의 질환 예측 모델의 정확도를 나타낸 도면이다.

도 8은 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 방법에 대한 동작 흐름도이다.

도 9는 본원의 일 실시예에 따른 질환 예측 방법에 대한 동작 흐름도이다.

발명을 실시하기 위한 구체적인 내용

[0033] 아래에서는 첨부한 도면을 참조하여 본원이 속하는 기술 분야에서 통상의 지식을 가진 자가 용이하게 실시할 수

있도록 본원의 실시예를 상세히 설명한다. 그러나 본원은 여러 가지 상이한 형태로 구현될 수 있으며 여기에서 설명하는 실시예에 한정되지 않는다. 그리고 도면에서 본원을 명확하게 설명하기 위해서 설명과 관계없는 부분은 생략하였으며, 명세서 전체를 통하여 유사한 부분에 대해서는 유사한 도면 부호를 붙였다.

- [0034] 본원 명세서 전체에서, 어떤 부분이 다른 부분과 "연결"되어 있다고 할 때, 이는 "직접적으로 연결"되어 있는 경우뿐 아니라, 그 중간에 다른 소자를 사이에 두고 "전기적으로 연결" 또는 "간접적으로 연결"되어 있는 경우도 포함한다.
- [0035] 본원 명세서 전체에서, 어떤 부재가 다른 부재 "상에", "상부에", "상단에", "하에", "하부에", "하단에" 위치하고 있다고 할 때, 이는 어떤 부재가 다른 부재에 접해 있는 경우뿐 아니라 두 부재 사이에 또 다른 부재가 존재하는 경우도 포함한다.
- [0036] 본원 명세서 전체에서, 어떤 부분이 어떤 구성 요소를 "포함"한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성 요소를 제외하는 것이 아니라 다른 구성 요소를 더 포함할 수 있는 것을 의미한다.
- [0037] 본원은 전자건강기록(electronic health records, EHR)과 모바일 기기를 통해 획득되는 모바일 기반 개인건강기록(mobile personal health records, mPHR)을 포함하는 의료 서비스를 위한 통합 데이터셋의 생성 기술에 대하여 제안한다.
- [0038] 구체적인 설명에 앞서 간단히 살펴보면, EHR은 높은 수준의 진료를 용이하게 하는 920개의 레코드로 이루어질 수 있다. 이러한 기록에는 의료기관에서 측정되는 심장 질환과 관련된 많은 특성이 포함될 수 있다. 그러나 EHR은 mPHR에 비해 상대적으로 정적이라 할 수 있으며, 이는 사용자(환자)가 병원에서 진료를 받을 때에만 정보가 업데이트되기 때문이라 할 수 있다. 그러므로 EHR 만 사용하여 사용자의 건강 상태 관련 분석 내지 진단을 수행하는 경우에는 환자의 현재 건강 상태를 정확하게 진단하기가 어려울 수 있다.
- [0039] 따라서, 본원에서는 이러한 제한을 극복하기 위해 EHR과 mPHR이 통합된 통합 데이터셋을 생성하는 기술에 대하여 제안한다. 여기서, mPHR은 모바일 기기가 주기적으로 측정하는 데이터의 집합(데이터셋)을 의미할 수 있으며, 사용자의 건강 상태 관련 분석 내지 진단 수행 시 mPHR을 이용하는 경우에는 사용자의 현재 상태에 대한 정확한 분석/진단이 이루어질 수 있다. 통합 데이터셋의 생성 방법에 대한 구체적인 설명은 다음과 같다.
- [0040] 도 1은 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋(data set) 생성 시스템(이하 '본 데이터셋 생성 시스템'이라 함)의 개략적인 구성을 나타낸 도면이다.
- [0041] 도 1을 참조하면, 본 데이터셋 생성 시스템(100)은 복수의 의료기관의 장치(10), 복수의 모바일 기기(20) 및 헬스케어 서비스 제공을 위한 데이터셋 생성 장치(30, 이하 '본 데이터셋 생성 장치'라 함)를 포함할 수 있다.
- [0042] 복수의 의료기관의 장치(10)는 복수의 사용자에 대한 전자건강기록(electronic health records, EHR)을 제공할 수 있다. 복수의 의료기관의 장치(10)는 복수의 사용자(환자)의 건강과 관련된 정보(데이터)를 EHR로서 전자적으로 기록하여 제공할 수 있다. 즉, EHR은 복수의 의료기관의 장치(10)에 의하여 제공될 수 있다. EHR로부터 획득 가능한 사용자의 건강 관련 정보(데이터)로는 나이, 성별, 콜레스테롤, 혈당, 혈압 등의 정보가 포함될 수 있으나, 이에만 한정되는 것은 아니다. 또한, 복수의 의료기관의 장치(10)에 대응하는 의료기관으로는 일예로 병원, 한방병원, 치과병원, 대학 병원, 조산원, 의원, 한의원, 치과의원, 요양병원, 종합병원 등이 포함될 수 있으며, 이에만 한정되는 것은 아니다.
- [0043] 복수의 모바일 기기(20)는 복수의 사용자 각각에 대한 모바일 기반의 개인건강기록(mobile personal health records, mPHR)을 측정하여 제공할 수 있다. 즉, 모바일 기반의 개인건강기록인 mPHR은 사용자 각각이 소지한 모바일 기기를 통해 측정되어 제공될 수 있다. mPHR로부터 획득 가능한 사용자의 건강 관련 정보(데이터)로는 혈압, 심장 박동수, 혈당, 활동 칼로리 등의 정보가 포함될 수 있으나, 이에만 한정되는 것은 아니다.
- [0044] 모바일 기기(20)는 휴대성과 이동성이 보장되는 이동 통신 장치로서, 예를 들면, PCS(Personal Communication System), GSM(Global System for Mobile communication), PDC(Personal Digital Cellular), PHS(Personal Handyphone System), PDA(Personal Digital Assistant), IMT(International Mobile Telecommunication)-2000, CDMA(Code Division Multiple Access)-2000, W-CDMA(WCode Division Multiple Access), Wibro(Wireless Broadband Internet) 단말, 스마트폰(Smartphone), 스마트패드(SmartPad), 태블릿 PC, 노트북, 웨어러블 디바이스 등과 같은 모든 종류의 무선 통신 장치를 포함할 수 있으며, 이에 한정되는 것은 아니다.
- [0045] 본 데이터셋 생성 장치(30)는 복수의 의료기관의 장치(10)로부터 획득되는 전자건강기록(EHR)을 포함하는 제1 데이터셋 및 복수의 모바일 기기(20)로부터 획득되는 모바일 기반의 개인건강기록(mPHR)을 포함하는 제2 데이터

셋을 생성하여 제1 데이터셋 및 상기 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하고, 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터가 통합된 통합 데이터셋을 생성할 수 있다.

[0046] 여기서, 본 데이터셋 생성 장치(30)가 복수의 의료기관의 장치(10)로부터 획득하는 EHR 및 복수의 모바일 기기(20)로부터 획득하는 mPHR은 네트워크(40)를 통해 획득될 수 있다.

[0047] 네트워크(40)는 일예로 3GPP(3rd Generation Partnership Project) 네트워크, LTE(Long Term Evolution) 네트워크, WIMAX(World Interoperability for Microwave Access) 네트워크, 인터넷(Internet), LAN(Local Area Network), Wireless LAN(Wireless Local Area Network), WAN(Wide Area Network), PAN(Personal Area Network), 블루투스(Bluetooth) 네트워크, NFC(Near Field Communication) 네트워크, 위성 방송 네트워크, 아날로그 방송 네트워크, DMB(Digital Multimedia Broadcasting) 네트워크 등이 포함될 수 있으며, 이에 한정된 것은 아니다.

[0048] 이러한 본 데이터셋 생성 장치(30)는, EHR의 경우 의료 전문가가 제공한 진료 및 처방 정보가 포함되어 있어 사용자에게 대한 신뢰성 있는 의료 정보를 제공할 수 있는 데에 반해 mPHR에 비해 상대적으로 정적이기 때문에 EHR을 사용하는 것만으로는 사용자의 현재 건강 상태를 정확하게 파악하기 어려운 문제가 있고, mPHR의 경우 EHR에 비해 사용자의 건강 관련 정보가 상대적으로 짧은 기간에 측정되어 업데이트됨에 따라 사용자의 현재 상태를 정확하게 파악할 수 있는 데에 반해 획득된 정보가 의료 전문가에 의해 수집되거나 병원의 의료 기기를 사용하여 수집된 것이 아님에 따라 신뢰성이 다소 떨어진다는 점을 고려하여, EHR과 mPHR이 통합된 통합 데이터셋을 생성하는 기술에 대하여 제안한다. 본 데이터셋 생성 장치(30)에 대한 보다 구체적인 설명은 다음과 같다.

[0049] 이하 본 데이터셋 생성 장치(30)에 대해 설명함에 있어서, 본 데이터셋 생성 장치(30)는 사용자의 현재 상태를 반영하여 일예로 심장 질환이 존재하는지 여부에 대한 정확한 진단이 이루어질 수 있도록 하는 통합 데이터셋을 생성할 수 있다. 즉, 본원은 다양한 유형의 질환 중 일예로 심장 질환에 초점을 맞춘 분석이 이루어질 수 있도록 하는 통합 데이터셋의 생성 기술에 대해서 설명하며, 다만 이에만 한정되는 것은 아니고, 다양한 유형의 질환에 대한 분석이 용이하도록 하는 통합 데이터셋의 생성 또한 가능하다.

[0050] 본 데이터셋 생성 장치(30)는 데이터셋 생성부(31), 전처리부(32) 및 매칭부(33)를 포함할 수 있다.

[0051] 데이터셋 생성부(31)는 전자건강기록(electronic health records, EHR)을 포함하는 제1 데이터셋 및 모바일 기기를 통해 획득되는 개인건강기록(mobile personal health records, mPHR)을 포함하는 제2 데이터셋을 생성할 수 있다.

[0052] 데이터셋 생성부(31)는 UCI(University of California-Irvine) 기계 학습 저장소(Machine Learning Repository)에서 제공하는 데이터로서 서로 다른 장소를 갖는 복수의 데이터셋에 속한 데이터를 통합하여 제1 데이터셋을 생성할 수 있다. 또한, 데이터셋 생성부(31)는 모바일 기기(20)를 통해 획득되는 데이터(즉, mPHR 데이터)의 속성이 제1 데이터셋의 속성과 매치되도록 제2 데이터셋을 생성할 수 있다.

[0053] 구체적으로, 앞서 말한 바와 같이 일예로 심장 질환에 초점을 맞춘 분석이 이루어질 수 있도록, 데이터셋 생성부(31)는 UCI 기계 학습 저장소의 심장 질환 데이터셋 디렉토리의 데이터를 사용하여 EHR을 포함하는 제1 데이터셋을 생성할 수 있다. 심장 질환 데이터셋 디렉토리에는 심장 질환 진단과 관련된 4가지의 데이터셋이 존재할 수 있다. 이때, 4가지의 데이터셋 각각은, 4개의 장소(즉, Cleveland Clinic Foundation, OH, USA, Hungarian Institute of Cardiology, Budapest, Hungary, Veterans Affairs Medical Center, CA, USA, and University Hospital, Zurich, Switzerland) 중 어느 하나에서 수집함으로써 구성(생성)될 수 있다. 또한, 상기 4가지의 데이터셋 각각은 동일한 인스턴스 형식으로 된 14 개의 원시 속성(raw attributes)으로 구성될 수 있다. 데이터셋 생성부(31)는 일예로 서로 다른 장소에서 수집된 상기 4가지의 데이터셋을 통합함으로써 제1 데이터셋을 생성할 수 있다.

[0054] 도 2는 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 장치에서 고려되는 EHR의 속성을 나타낸 도면이다.

[0055] 도 2를 참조하면, 본 데이터셋 생성 장치(30)에서 고려되는 EHR의 속성은 14가지가 포함될 수 있다. EHR의 14가지 속성(attributes)으로는 나이(age), 성별(sex), 흉통 유형(chest pain type, cp), 안정혈압(resting blood pressure, trestbps), 혈청 콜레스테롤(serum cholesterol, chol), 공복혈당(fasting blood sugar, fbs), 안정 시 심전도 결과(resting electrocardiographic results, restecg), 최대 심박수(maximum heart rate achieved, thalach), 운동 유발성 협심증(exercise induced angina, exang), 휴식과 관련된 활동에 의해 유도된 ST 우울

증(ST depression induced by exercise relative to rest, oldpeak), 피크 운동 ST 세그먼트의 슬롭(the slope of the peak exercise ST segment, slope), 플루오로스코피에 의해 채색된 주요 혈관의 수(number of major vessels colored by flourosopy, ca), 지중해빈혈(thalassemia, thal), 심장 질환의 존재(existence of heart disease, num)가 포함될 수 있다.

- [0056] EHR의 모든 속성에 대한 값(즉, 속성 값)은 숫자 값으로 표현될 수 있다. 일예로, 성별(sex)은 남자인 경우 1, 여자인 경우 0으로 표현될 수 있다. 흉통 유형(cp)은 전형적인 협심증(typical angina)인 경우 1, 비전형 협심증(atypical angina)인 경우 2, 비-협심 통증(non-anginal pain)인 경우 3, 무증상(asymptomatic)인 경우 4로 표현될 수 있다. 공복혈당(fbs)은 120 mg/dl를 초과하는 경우 1, 120 mg/dl를 초과하지 않는 경우 0으로 표현될 수 있다. 안정 시 심전도 결과(restecg)는 정상인 경우(normal) 0, ST-T파를 가지는 경우(having ST-T wave abnormality) 1, 좌심실비대(left ventricular hypertrophy)로 확진되거나 가능성이 있는 경우 2로 표현될 수 있다. 운동 유발성 협심증(exang)은 존재하는 경우 1, 존재하지 않는 경우 0으로 표현될 수 있다. 피크 운동 ST 세그먼트의 슬롭(slope)은 업슬로핑(upsloping)인 경우 1, 평평한(flat) 경우 2, 다운 슬로핑(ownsloping)인 경우 3으로 표현될 수 있다. 지중해빈혈(thal)은 정상(normal)인 경우 3, 고정된 결함(fixed defect)인 경우 6, 심장병으로 진단(diagnosis of heart disease)된 경우 7로 표현될 수 있다. 심장 질환의 존재(num)는 존재하지 않는 경우 0, 존재하는 경우 1로 표현될 수 있다.
- [0057] 이에 따르면, 데이터셋 생성부(31)를 통해 생성된 제1 데이터셋은 도 2에 도시된 14가지의 속성에 대한 정보를 포함할 수 있다.
- [0058] 또한, 데이터셋 생성부(31)는 통합 데이터셋 생성시 모바일 기기(20)를 통해 획득되는 데이터의 속성(즉, mPHR의 속성)이 제1 데이터셋의 속성(즉, EHR의 속성)과 매치되도록 제2 데이터셋을 생성할 수 있다.
- [0059] 이때, 데이터셋 생성부(31)를 통해 생성된 제2 데이터셋은 일예로 심장 질환과 관련하여 4가지의 속성을 포함할 수 있다. 제2 데이터셋에 포함된 4가지의 속성, 달리 말해 mPHR과 관련하여 모바일 기기로부터 획득하는 데이터의 속성 정보로는 일예로, 혈압, 심장 박동수, 혈당 및 활동 칼로리가 포함될 수 있다. 이에 따르면, 제1 데이터셋, 제2 데이터셋 및 후술할 통합 데이터셋에 포함된 데이터는 일예로 심장 질환 관련 데이터일 수 있다.
- [0060] 또한, 데이터셋 생성부(31)를 통해 생성된 제1 데이터셋 및 제2 데이터셋 각각은 일예로 920개의 레코드로 이루어질 수 있다. 이때 920개의 레코드라 함은 920명의 사용자에게 대한 건강 관련 기록을 의미할 수 있으며, 그 개수는 본원의 이해를 돕기 위한 하나의 예시일 뿐 이에만 한정되는 것은 아니다.
- [0061] 전처리부(32)는 제1 데이터셋과 제2 데이터셋이 생성된 이후 통합 데이터셋의 생성을 위해, 전자건강기록(EHR)을 포함하는 제1 데이터셋 및 모바일 기기를 통해 획득되는 개인건강기록(mPHR)을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행할 수 있다. 이에 따르면, 제1 데이터셋 및 제2 데이터셋은 전처리가 수행되기 이전에 생성될 수 있다.
- [0062] 구체적으로, 전처리부(32)는 제1 데이터셋과 제2 데이터셋 내의 각 개별 레코드에 대하여 전처리를 수행할 수 있다. 즉, 전처리부(32)는 제1 데이터셋과 제2 데이터셋 각각에 포함된 개별 레코드에 대하여 전처리를 수행할 수 있다. 다시 말해, 전처리부(32)는 제1 데이터셋에 포함된 EHR 관련 개별 레코드 각각에 대하여 전처리를 수행하고, 제2 데이터셋에 포함된 mPHR 관련 개별 레코드 각각에 대하여 전처리를 수행할 수 있다.
- [0063] 전처리부(32)는 전처리 수행 시 제1 데이터셋 및 제2 데이터셋 각각에 포함된 데이터(즉, 레코드)에 대하여, 누락 값 또는 이상치가 존재하는 것으로 판단되는 경우, 누락 값 또는 이상치가 속한 데이터셋 내에서 누락 값 또는 이상치에 대응하는 속성과 동일 속성에 속하는 속성 값들의 평균값으로 대체하는 전처리를 수행할 수 있다. 이러한 전처리 과정은 클렌징(cleansing) 과정이라 달리 표현될 수 있다.
- [0064] 데이터셋 내의 누락된 값이나 이상치는 분석의 정확성을 떨어뜨리므로, 본 데이터셋 생성 장치(30)는 전처리부(32)를 통해 제1 데이터셋 및 제2 데이터셋 내의 누락 값 또는 이상치를 속성 값들의 평균값으로 대체하는 전처리를 수행함으로써, 분석의 정확성을 향상시킬 수 있다. 즉, 전처리를 통해 제1 데이터셋과 제2 데이터셋에 포함된 누락 값 및 이상치를 제거(즉, 속성 값들의 평균값으로 대체)함으로써, 본 데이터셋 생성 장치(30)가 생성하는 통합 데이터셋에 기반한 분석 수행 시 보다 정확한 분석이 이루어지도록 제공할 수 있다.
- [0065] 예를 들어, 제1 데이터셋 내의 920개의 레코드 중 어느 한 레코드(즉, 어느 한 사용자에게 대한 건강 관련 기록)에 있어서, 흉통 유형(cp) 속성에 대한 값(즉, 흉통 유형 속성 값)이 누락되어 있다고 가정하자. 또한, 상기 어느 한 레코드를 제외한 919개의 레코드에 있어서는 흉통 유형 속성에 대한 값이 모두 존재한다고 가정하자. 이러한 경우, 전처리부(32)는 누락된 상기 흉통 유형 속성의 속성 값을, 해당 누락 값이 속해 있는 제1 데이터셋

내에서 누락된 값에 대응하는 속성인 흉통 유형 속성과 동일 속성에 속하는 속성 값들의 평균값으로 대체하는 전처리를 수행할 수 있다. 즉, 전처리부(32)는 누락된 흉통 유형의 속성 값을, 제1 데이터셋 내의 919개의 레코드에서 흉통 유형 속성에 속하는 919개의 흉통 유형 속성 값들의 평균값으로 대체할 수 있다.

[0066] 또한, 전처리부(32)는 매칭부(33)를 통한 매칭이 수행되기 이전에 통합 데이터셋의 포맷(통합 포맷)을 설정할 수 있다. 즉, 전처리부(32)는 통합 데이터셋을 생성함에 있어서 EHR과 mPHR을 통합시키기에 적합한 포맷을 설정할 수 있다. 이때, 통합 데이터셋의 포맷은 EHR을 포함하는 제1 데이터셋의 포맷과 mPHR을 포함하는 제2 데이터셋의 포맷을 고려하여 설정될 수 있으며, 또는 사용자에게 의하여 설정될 수 있다.

[0067] 매칭부(33)는 제1 데이터셋과 제2 데이터셋 각각에 대하여 전처리가 수행된 이후, 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터가 통합된 통합 데이터셋을 생성할 수 있다. 즉, 통합 데이터셋은 전처리된 제1 데이터셋의 데이터(개별 레코드) 및 전처리된 제2 데이터의 데이터(개별 레코드) 간의 매칭에 의해 생성될 수 있다.

[0068] 매칭부(33)는 제1 데이터셋과 제2 데이터셋 내의 각 개별 레코드에 대하여 매칭을 수행하되, 제1 데이터셋의 개별 레코드와 제2 데이터셋의 개별 레코드 간에 매칭을 수행할 수 있다.

[0069] 매칭부(33)는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성 간의 동일 여부를 고려한 매칭을 통해 통합 데이터셋을 생성할 수 있다.

[0070] 구체적으로, 매칭부(33)는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일하지 않는 것으로 판단되는 경우, 제2 데이터셋의 데이터 속성을 제1 데이터셋의 데이터 속성과 결합시킬 수 있다. 또한, 매칭부(33)는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일한 것으로 판단되는 경우, 제1 데이터셋의 데이터 속성을 제2 데이터셋의 데이터 속성으로 덮어씌울 수 있다(달리 표현하여, 업데이트할 수 있다). 보다 자세하게는, 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일한 것으로 판단되는 경우, 제1 데이터셋의 데이터 속성 값이 제2 데이터셋의 데이터 속성 값으로 덮어 씌워질 수 있다.

[0071] 예를 들어, 제2 데이터셋의 데이터 속성으로는 혈압, 심장 박동수, 혈당 및 활동 칼로리가 존재하는데, 여기서 활동 칼로리를 제외한 혈압, 심장 박동수 및 혈당에 대한 속성은 제1 데이터셋의 데이터 속성(즉, 제1 데이터셋 내의 데이터 속성인 혈압(trestbps), 심장 박동수(thalach), 및 혈당(fbs)과 중복(동일)된다. 이에 따라, 매칭부(33)는 두 데이터셋 간의 속성 매칭 수행 시 혈압, 심장 박동수 및 혈당에 대하여 매칭이 이루어진 경우, 제1 데이터셋의 데이터 속성을 제2 데이터셋의 데이터 속성으로 덮어씌울 수 있다. 한편, 두 데이터셋 간의 속성 매칭 수행 시 활동 칼로리에 대하여 매칭이 이루어진 경우, 매칭부(33)는 제2 데이터셋의 데이터 속성을 제1 데이터셋의 데이터 속성과 결합시킬 수 있다.

[0072] 이때, 매칭부(33)는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성 간의 매칭 수행 시 전처리부(32)에서 설정된 통합 데이터셋의 포맷(통합 포맷)을 고려하여 속성을 결합시키거나 속성을 덮어씌울 수 있다. 달리 말해, 매칭부(33)는 매칭을 통해 EHR 및 mPHR 내의 개별 속성(특성)을 통합 데이터셋의 통일된 형식(즉, 통합 데이터셋의 통합 포맷)으로 덮어 씌우거나 결합시킬 수 있다. 구체적인 일례로, 매칭부(33)는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일한 것으로 판단되는 경우, 제1 데이터의 속성 값을 통합 데이터셋의 포맷을 고려하여 제2 데이터의 속성 값으로 대체하여 적용할 수 있다.

[0073] 본 데이터셋 생성 장치(30)는 매칭 과정을 수행함에 있어서, 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일한 경우 제1 데이터셋의 데이터 속성을 제2 데이터셋의 데이터 속성으로 덮어씌우고, 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일하지 않은 경우 제2 데이터셋의 데이터 속성을 제1 데이터셋의 데이터 속성과 결합시킴으로써 통합 데이터셋을 생성할 수 있다.

[0074] 이러한 과정을 통해, 본 데이터셋 생성 장치(30)는 EHR이 갖는 이점(즉, 신뢰성 있는 의료 정보의 제공)과 mPHR이 갖는 이점(즉, 사용자의 현재 상태를 정확하게 파악)을 함께 유지하여 EHR과 mPHR로부터 상호 보완적인 정보의 접근이 이루어질 수 있도록 하는 통합 데이터셋(즉, EHR과 mPHR이 통합된 통합 데이터셋)을 생성할 수 있다. 즉, 본원은 생성된 통합 데이터셋으로 하여금 사용자의 현재 상태를 반영하면서 보다 정확하고 신뢰성 있는 건강 관련 진단 내지 분석이 이루어질 수 있도록 제공할 수 있다.

[0075] 매칭부(33)에 의하여 생성된 통합 데이터셋은 일례로 심장 질환과 관련하여 15가지의 속성과 920개의 레코드를 포함하도록 이루어질 수 있다. 여기서, 15가지의 속성은 EHR의 속성인 14가지의 항목과 mPHR의 속성인 활동 칼로리의 항목이 함께 고려된 것을 의미할 수 있다.

- [0076] 또한, 통합 데이터셋은 특정 사용자(환자)의 현재 건강 상태에 대한 정확한 분석(판단)이 이루어질 수 있도록 제2 데이터셋으로부터 미리 설정된 주기로 업데이트될 수 있다. 달리 말해, 통합 데이터셋 내의 데이터는 특정 사용자의 현재 건강 상태에 대한 실시간 분석 및 정확한 분석이 이루어질 수 있도록, 제2 데이터셋에 대응하는 모바일 기반의 개인건강기록인 mPHR로부터 미리 설정된 주기로 업데이트될 수 있다. 여기서 미리 설정된 주기는 사용자에게 의하여 설정될 수 있으며, 일예로 시간(time), 일(day) 등의 단위로 설정될 수 있다.
- [0077] 또한, 통합 데이터셋 내의 레코드 수가 많을수록, 통합 데이터셋에 기초한 사용자의 건강 상태 진단 시 그 정확성이 향상될 수 있다.
- [0078] 이러한 본 데이터셋 생성 장치(30)는 EHR과 mPHR이 통합(병합)된 통합 데이터셋을 생성하여 제공할 수 있으며, 통합 데이터셋으로 하여금 고정밀 질병 진단이 가능하도록 할 수 있다. 여기서, EHR은 사용자(환자)의 과거 의료 기록을 포함하여 환자의 전반적인 건강 상태와 관련된 데이터를 포함할 수 있으며, 이에 따라 사용자(환자)의 과거 병력, 생체 신호, 약물, 방사선 기록 등을 확인할 수 있다. mPHR은 사용자의 개인 단말(모바일 기기)에 의해 기록된 데이터로서, 시간 경과에 따른 사용자의 실시간 건강 상태 관련 정보를 제공할 수 있으며, 이에 따라 사용자(환자)가 의료기관 외부에서 자신의 건강 상태의 추적이 가능하도록 하여 사용자의 현재 건강 상태에 대한 진단이 효과적으로 이루어지도록 할 수 있다.
- [0079] 이를 고려하여, 본 데이터셋 생성 장치(30)는 통합 데이터셋으로 하여금 EHR과 mPHR을 통해 정확한 건강 진단 서비스를 가능하게 하여 사용자(환자)의 질병 진단에 보완적인 역할이 이루어지도록 할 수 있다. 다시 말해, 본원은 통합 데이터셋에 의하여 사용자에게 대한 정확한 건강 진단 정보를 제공하며, 사용자의 질병 여부에 대한 진단/예측/분석 시 EHR과 mPHR로부터 상호 보완적인 정보의 접근이 가능하도록 할 수 있다. 일예로, 본원의 통합 데이터셋에 의하면, 심장 질환의 존재 여부에 대한 정확한 진단/예측/분석이 이루어질 수 있도록 할 수 있다.
- [0080] 이하에서는 앞서 설명된 내용에 기초하여, 본 데이터셋 생성 장치(30)에 의하여 생성된 통합 데이터셋을 이용하여 사용자의 질환 가능성을 예측하는 기술에 대하여 설명하기로 한다.
- [0081] 도 3은 본원의 일 실시예에 따른 질환 예측 시스템(200)의 개략적인 구성을 나타낸 도면이다. 참고로, 본원에서 질환 예측이라 함은 질환이 존재할 가능성을 예측하거나 질환이 존재 여부에 대한 검출 또는 측정 등의 넓은 의미로 이해될 수 있다.
- [0082] 도 3을 참조하면, 본원의 일 실시예에 따른 질환 예측 시스템(200)은 제1 단말 기기(50), 질환 예측 장치(70)를 포함할 수 있다.
- [0083] 제1 단말 기기(50)는 예측 대상자의 건강 데이터를 제공할 수 있다. 즉, 제1 단말 기기(50)는 질환 가능성을 예측하고자 하는 대상인 예측 대상자의 건강 데이터를 측정하여 질환 예측 장치(70)로 제공할 수 있다. 여기서, 건강 데이터는 예측 대상자의 건강 상태와 관련된 데이터로서, 이는 제1 단말 기기(50)를 통해 측정되는 데이터로서 mPHR을 의미할 수 있다.
- [0084] 또한, 제1 단말 기기(50)는 앞서 도 1에서 설명한 복수의 모바일 기기(20) 중 어느 하나의 모바일 기기를 의미할 수 있다. 따라서, 이하 생략된 내용이라 하더라도 모바일 기기(20)에 대하여 설명된 내용은 제1 단말 기기(50)에 대한 설명에도 동일하게 적용될 수 있다.
- [0085] 또한, 제1 단말 기기(50)와 질환 예측 장치(70) 간에 데이터 송수신은 네트워크(60)를 통해 이루어질 수 있다. 여기서, 네트워크(60)는 앞서 설명한 네트워크(40)와 동일한 네트워크를 의미할 수 있다. 따라서, 이하 생략된 내용이라 하더라도 네트워크(40)에 대하여 설명된 내용은 네트워크(60)에 대한 설명에도 동일하게 적용될 수 있다.
- [0086] 질환 예측 장치(70)는 수신부(71), 생성부(72) 및 예측부(73)를 포함할 수 있다.
- [0087] 수신부(71)는 제1 단말 기기(50)로부터 예측 대상자의 건강 데이터를 수신할 수 있다.
- [0088] 생성부(72)는 복수의 사용자에게 대한 전자건강기록(electronic health records, EHR) 및 복수의 사용자에게 대한 모바일 기기를 통해 획득된 모바일 기반의 개인건강기록(mobile personal health records, mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성할 수 있다.
- [0089] 여기서, 통합 데이터셋은 전자건강기록을 포함하는 제1 데이터셋 및 모바일 기반의 개인건강기록을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하고, 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 생성될 수 있다. 이러한 통합 데이터셋은 앞서 설명한 데이터셋 생성 장치

(30)에 의하여 생성될 수 있다. 통합 데이터셋의 생성 과정에 대해서는 앞서 자세히 설명했으므로 이하 구체적인 설명을 생략하기로 한다.

- [0090] 일례로, 생성부(72)는 수신부(71)를 통해 건강 데이터가 수신된 이후에 예측 대상자의 특성(상태, 유형) 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 질환 예측 모델을 생성할 수 있다. 이후 예측부(73)는 생성된 질환 예측 모델에 기초하여 수신부(71)에서 수신한 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측할 수 있다.
- [0091] 여기서, 예측 대상자의 특성(상태, 유형)이라 함은 건강 데이터를 기반으로 한 예측 대상자의 현재의 건강 상태를 의미할 수 있다. 예를 들면, 질환 예측 모델 생성시 고려되는 예측 대상자의 특성으로는 예측 대상자의 mPHR과 관련하여 현재 사용자의 혈압, 심장 박동수, 혈당 및 활동 칼로리가 어느 정도의 수준을 나타내는지에 대한 상태 정보를 의미할 수 있다. 또한, 예측하고자 하는 질환 유형이라 함은 예측 대상자에 대하여 어떤 유형의 질환 가능성을 예측하고자 하는지에 대한 정보를 의미할 수 있다. 일례로 예측하고자 하는 질환 유형으로는 심장 질환, 피부 질환, 뇌혈관 질환 등이 포함될 수 있으나, 이에만 한정되는 것은 아니다. 이러한 예측하고자 하는 질환 유형은 사용자에게 의하여 설정될 수 있다.
- [0092] 다른 일례로, 생성부(72)는 통합 데이터셋에 포함된 데이터를 기반으로 복수 사용자의 특성이 고려된 복수의 질환 예측 모델을 생성하고, 수신부(71)를 통해 건강 데이터가 수신된 이후에 예측 대상자의 특성(상태) 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 복수의 질환 예측 모델 중 예측 대상자에 최적화된 최적 질환 예측 모델을 선택할 수 있다. 이후 예측부(73)는 생성부(72)에서 선택된 최적 질환 예측 모델에 기초하여 수신부(71)에서 수신한 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측할 수 있다.
- [0093] 즉, 일례로 예측 대상자에 대한 질환 가능성의 예측을 위해 이용되는 질환 예측 모델은, 건강 데이터가 수신된 이후에, 수신된 건강 데이터에 기초한 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려함으로써 생성될 수 있다. 다른 일례로, 질환 예측 장치(70)는 건강 데이터가 수신되기 이전에 통합 데이터셋에 기초하여 복수의 질환 예측 모델을 생성할 수 있다. 이후, 예측 대상자에 대한 질환 가능성의 예측을 위해 이용되는 질환 예측 모델은, 건강 데이터가 수신되면 예측 대상자의 특성(상태) 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 복수의 질환 예측 모델 중 최적 질환 예측 모델을 선택함으로써 선정될 수 있다.
- [0094] 또한, 생성부(72)를 통해 생성되는 질환 예측 모델은, 통합 데이터셋에 포함된 데이터의 복수의 속성과 관련된 복수의 규칙을 포함할 수 있다. 즉, 질환 예측 모델은 일례로 통합 데이터셋에 포함된 15가지의 속성과 관련된 복수의 규칙을 포함할 수 있다. 여기서, 통합 데이터셋에 포함된 15가지의 속성에 대한 설명은 앞서 자세히 설명했으므로, 이하 생략하기로 한다.
- [0095] 또한, 질환 예측 모델이 생성됨에 있어서, 생성되는 질환 예측 모델의 복수의 규칙의 조합 수 및 조합 순서는 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 결정될 수 있다. 또한, 생성되는 질환 예측 모델의 복수의 규칙의 조합 수 및 조합 순서는 통합 데이터셋에 포함된 복수의 속성에 대하여 의료 관련 업계에서 정의(규정)된 속성 값 기준을 고려하여 결정될 수 있다. 또한, 생성되는 질환 예측 모델의 복수의 규칙의 조합 수 및 조합 순서는 자동으로 결정될 수 있다.
- [0096] 또한, 질환 예측 모델은, 통합 데이터셋에 포함된 데이터에 대하여 의사결정트리(decision tree)를 적용함으로써 생성될 수 있다. 생성부(72)는 질환 예측 모델 생성시 일례로, IF-THEN 규칙을 제공하는 의사결정트리를 이용할 수 있으며, 이에만 한정되는 것은 아니다.
- [0097] 예측부(73)는 질환 예측 모델에 기초하여 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측(또는 질환 존재 여부를 측정, 검출)할 수 있다. 구체적인 예로, 예측부(73)는 건강 데이터의 수신 시 생성된 질환 예측 모델에 기초하여 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측할 수 있다. 또는 예측부(73)는 복수의 질환 예측 모델 중 선택된 최적 질환 예측 모델에 기초하여 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측할 수 있다.
- [0098] 이러한 본원의 일 실시예에 따른 질환 예측 장치(70)는 EHR과 mPHR이 통합된 통합 데이터셋을 이용하여 질환 예측 모델을 생성하고, 생성된 질환 예측 모델에 기초하여 사용자의 건강 상태 관련 질환 가능성을 예측하므로, 종래에 EHR 만을 이용하거나 mPHR만을 이용하는 것 대비 예측 정확도가 향상될 수 있다.
- [0099] 이하에서는 종래에 EHR 만을 이용하거나 mPHR만을 이용하여 예측하는 것과 대비하여 본원에서 제안하는 통합 데이터셋 기반의 질환 예측의 우수성을 입증하기 위한 실험 결과에 대하여 설명하기로 한다.

- [0100] 본원의 우수성 입증을 위한 본원의 일 실험예에서는 R 버전 3.4.1을 사용하여 실험을 수행했다.
- [0101] 또한, 본원의 일 실험예에서는 통합 데이터셋의 성능 비교를 위해 데이터 분류를 위한 IF-THEN 규칙을 제공하는 의사결정트리를 사용하여 심장 질환과 관련된 질환 예측 모델을 생성한다. 또한, 본원의 일 실험예에서는 통합 데이터셋 내의 'num' 속성을 이용하여 사용자(예를 들어, 예측 대상자 또는 통합 데이터셋 내에 기록된 사용자)의 심장 질환의 존재 여부(달리 말해, 심장 질환의 가능성 여부)를 '예(존재함, 존재할 가능성이 있음)' 또는 '아니오(존재하지 않음, 존재할 가능성이 없음)'로 분류할 수 있다.
- [0102] 이하에서는 EHR과 mPHR을 포함하는 본원의 통합 데이터셋에 기초하여 생성된 질환 예측 모델과 종래의 EHR 또는 mPHR 기반의 질환 예측 모델에 대한 성능(정확성) 비교의 실험 예에 대하여 보다 자세히 설명하기로 한다.
- [0103] 도 4는 본원의 일 실시예에 따른 통합 데이터셋에 의사결정트리를 적용함에 따라 생성된 질환 예측 모델의 예를 나타낸 도면이다.
- [0104] 도 4를 참조하면, 본원의 일 실험예에 따라 생성된 질환 예측 모델은 일예로 15 개의 규칙을 가질 수 있다. 이 규칙에 의하면 예측 대상자(또는 통합 데이터셋 내에 기록된 사용자)가 질환 가능성이 있는지 여부(달리 말해, 질환의 존재 여부)에 대해 '예' 또는 '아니오'로 분류될 수 있다. 이하 본원의 일 실험예에서는 질환 가능성 예측과 관련하여 일예로 심장 질환 가능성의 예측이 이루어질 수 있으며, 이에만 한정되는 것은 아니고, 다양한 질환에 대한 존재 가능성의 예측이 이루어질 수 있다.
- [0105] 이에 따르면, 통합 데이터셋 내에 기록된 모든 사용자들 또는 수신부(71)를 통해 수신한 예측 대상자(일예로, 신규 예측 대상자)에 대한 심장 질환 가능성의 여부는 의사결정트리 기반의 질환 예측 모델을 통해 예측될 수 있다.
- [0106] 예를 들어, 본원의 일 실험예에 따라 생성된 질환 예측 모델에 의하면, 첫번? 규칙과 관련하여 특정 사용자(이는 신규 예측 대상자 또는 통합 데이터셋 내에 기록된 사용자 중 어느 하나의 사용자를 의미할 수 있음)의 건강 데이터의 흉통 유형을 나타내는 'cp' 값이 1, 2 또는 3이 아닌 경우, 두 번째 규칙과 관련하여 상기 특정 환자의 안정 혈압을 나타내는 'trestbps'에 대한 확인이 이루어질 수 있다. 일예로 두 번째 규칙에서 'trestbps'의 값이 132보다 작은 경우, 상기 특정 환자의 건강 데이터에서 세 번째 규칙과 관련된 속성에 대한 확인이 이루어질 수 있다. 즉, 세 번째 규칙과 관련된 속성으로서 일예로 콜레스테롤 'chol'에 대한 확인이 이루어질 수 있다. 여기서, 세 번째 규칙에서 특정 사용자의 'chol' 속성 값이 430보다 크다고 판단된 경우, 특정 사용자의 심장 질환 가능성 여부(즉, 심장 질환의 존재 여부)는 '예'로 예측될 수 있다. 참고로, 도 4의 도면 상에서 'hr' 속성은 mPHR과 관련된 심박수(심장 박동수)를 의미할 수 있다.
- [0107] 도 5는 종래의 EHR에 의사결정트리를 적용함에 따라 생성된 질환 예측 모델의 예를 나타낸 도면이다.
- [0108] 도 5를 참조하면, EHR 기반의 질환 예측 모델에서는 일예로 'cp'가 첫 번째 규칙으로 고려되어 질환 가능성의 예측이 이루어질 수 있다. 그런데, EHR 기반의 질환 예측 모델의 경우에는 통합 데이터셋 기반의 질환 예측 모델에 비해 규칙의 수가 적기 때문에 본원의 기술 대비 질환 예측 모델의 정확성이 낮다는 문제가 있다.
- [0109] 도 6은 종래의 mPHR에 의사결정트리를 적용함에 따라 생성된 질환 예측 모델의 예를 나타낸 도면이다.
- [0110] 도 6을 참조하면, mPHR 기반의 질환 예측 모델에서는 일예로 5가지의 규칙이 존재할 수 있다. 이는 mPHR의 경우 심장 질환과 관련하여 4가지의 속성(즉, 혈압, 심장 박동수, 혈당 및 활동 칼로리)만 포함되어 있기 때문이라 할 수 있다. 이에 따르면, mPHR 기반의 질환 예측 모델 또한 EHR 기반의 질환 예측 모델과 마찬가지로 통합 데이터셋 기반의 질환 예측 모델에 비해 규칙의 수가 적기 때문에, 본원의 기술 대비 질환 예측 모델의 정확성이 낮다는 문제가 있다.
- [0111] 이하에서는 EHR, mPHR 및 통합 데이터셋 각각의 의사결정트리를 분석하여 의사결정트리에서 파생된 예측 결과의 정확성에 대하여 평가한다.
- [0112] 도 7은 EHR 또는 mPHR 대비 본원의 일 실시예에 따른 통합 데이터셋 기반의 질환 예측 모델의 정확도를 나타낸 도면이다.
- [0113] 도 7을 참조하면, 질환 예측 모델의 정확도를 계산하기 위해 n 과 p 에 대한 값이 다음과 같이 설정될 수 있다. n 은 실제 값이 '아니오'일 때 질환 예측 모델에 의한 예측 값이 '아니오'로 예측된 데이터 요소의 수를 나타낸다. p 는 실제 값이 '예'일 때 질환 예측 모델에 의한 예측 값이 '예'로 예측된 데이터 요소의 수를 나타낸다. 질환 예측 모델(즉, 의사결정트리)의 정확도는 $(n + p) / (\text{데이터셋의 총 레코드 수})$ 에 기초하여 산출될

수 있다.

- [0114] 정확도 산출 결과, 본원의 일 실시예에 따른 통합 데이터셋 기반의 질환 예측 모델(integrated dataset)의 정확도(accuracy)는 0.82로 나타났다. 또한, EHR을 포함하는 데이터셋 기반의 질환 예측 모델의 정확도는 0.79로 나타났다. 또한 mPHR을 포함하는 데이터셋 기반의 질환 예측 모델의 정확도는 0.78로 나타났다.
- [0115] 이에 따르면, 본원의 통합 데이터셋 기반의 질환 예측 모델은 기존의 EHR 또는 mPHR 기반의 질환 예측 모델에 비해 가장 높은 정확도를 가짐을 확인할 수 있다. 특히, 본원의 통합 데이터셋 기반의 질환 예측 모델은 기존의 EHR 또는 mPHR 기반의 질환 예측 모델에 비해 약 3 % 및 4 % 더 높은 정확도를 보임을 확인할 수 있다. 즉, 본원의 통합 데이터셋 기반의 질환 예측 모델에 의하면 EHR 또는 mPHR 대비 사용자의 질병을 예측하는데 더 높은 정확성을 나타냄을 확인할 수 있다.
- [0116] 이러한 본원은 보다 효과적이고 정확한 의료 서비스의 제공이 이루어지도록 하는 통합 데이터셋을 제공할 수 있다. 즉, 본원은 통합 데이터셋 기반으로 의미 있는 정보의 추출이 이루어지도록 하여 보다 양질의 의료 서비스를 제공할 수 있다.
- [0117] 이하에서는 상기에 자세히 설명된 내용을 기반으로, 본원의 동작 흐름을 간단히 살펴보기로 한다.
- [0118] 도 8는 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 방법에 대한 동작 흐름도이다.
- [0119] 도 8에 도시된 데이터셋 생성 방법은 앞서 설명된 데이터셋 생성 장치(30)에 의하여 수행될 수 있다. 따라서, 이하 생략된 내용이라고 하더라도 데이터셋 생성 장치(30)에 대하여 설명된 내용은 데이터셋 생성 방법에 대한 설명에도 동일하게 적용될 수 있다.
- [0120] 도 8을 참조하면, 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 방법은 단계S11에서 전자건강기록(EHR)을 포함하는 제1 데이터셋 및 모바일 기기를 통해 획득되는 모바일 기반 개인건강기록(mPHR)을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행할 수 있다.
- [0121] 또한, 단계S11에서는, 제1 데이터셋 및 상기 제2 데이터셋 각각에 포함된 데이터에 대하여, 누락 값 또는 이상치가 존재하는 것으로 판단되는 경우, 누락 값 또는 이상치가 속한 데이터셋 내에서 상기 누락 값 또는 이상치에 대응하는 속성과 동일 속성에 속하는 속성 값들의 평균값으로 대체하는 전처리를 수행할 수 있다.
- [0122] 다음으로, 단계S12에서는 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터가 통합된 통합 데이터셋을 생성할 수 있다.
- [0123] 또한, 단계S12에서는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성 간의 동일 유무를 고려한 매칭을 통해 통합 데이터셋을 생성할 수 있다.
- [0124] 또한, 단계S12에서는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일하지 않는 것으로 판단되는 경우, 제2 데이터셋의 데이터 속성을 제1 데이터셋의 데이터 속성과 결합시킬 수 있다. 또한, 단계S12에서는 제1 데이터셋의 데이터 속성과 제2 데이터셋의 데이터 속성이 동일한 것으로 판단되는 경우, 제1 데이터셋의 데이터 속성을 제2 데이터셋의 데이터 속성으로 덮어씌울 수 있다.
- [0125] 또한, 본원의 일 실시예에 따른 헬스케어 서비스 제공을 위한 데이터셋 생성 방법은 단계S11 이전에, 제1 데이터셋 및 제2 데이터셋을 생성하는 단계를 포함할 수 있다.
- [0126] 이때, 제1 데이터셋 및 제2 데이터셋을 생성하는 단계에서는, UCI(University of California-Irvine) 기계 학습 저장소(Machine Learning Repository)에서 제공하는 데이터로서 서로 다른 장소를 갖는 복수의 데이터셋에 속한 데이터를 통합하여 제1 데이터셋을 생성할 수 있다. 또한, 제1 데이터셋 및 제2 데이터셋을 생성하는 단계에서는 모바일 기기를 통해 획득되는 데이터의 속성이 제1 데이터셋의 속성과 매치되도록 제2 데이터셋을 생성할 수 있다.
- [0127] 상술한 설명에서, 단계 S11 내지 S12는 본원의 구현예에 따라서, 추가적인 단계들로 더 분할되거나, 더 적은 단계들로 조합될 수 있다. 또한, 일부 단계는 필요에 따라 생략될 수도 있고, 단계 간의 순서가 변경될 수도 있다.
- [0128] 도 9는 본원의 일 실시예에 따른 질환 예측 방법에 대한 동작 흐름도이다.
- [0129] 도 9에 도시된 질환 예측 방법은 앞서 설명된 질환 예측 장치(70)에 의하여 수행될 수 있다. 따라서, 이하 생략

된 내용이라고 하더라도 질환 예측 장치(70)에 대하여 설명된 내용은 질환 예측 방법에 대한 설명에도 동일하게 적용될 수 있다.

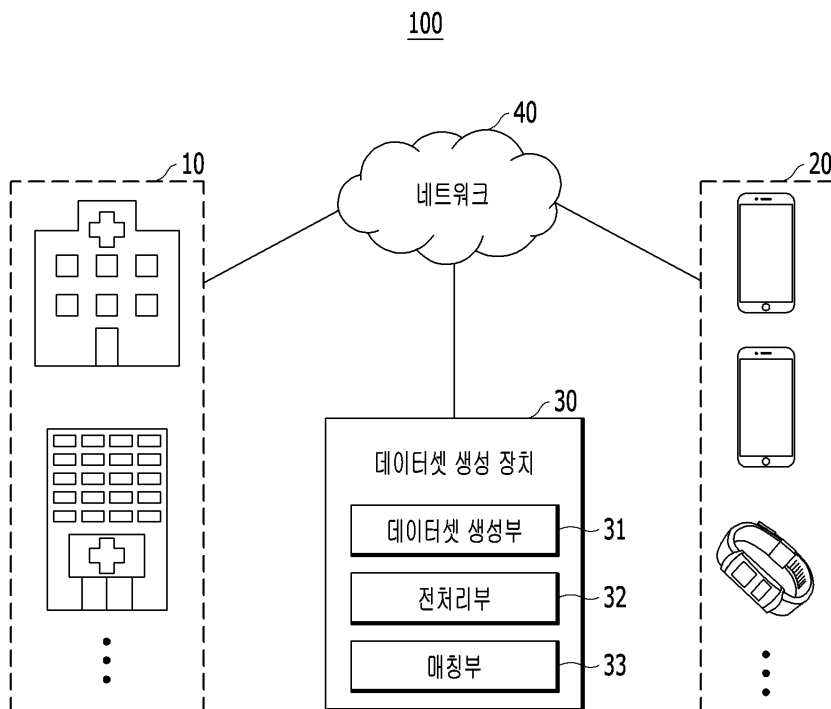
- [0130] 도 9를 참조하면, 본원의 일 실시예에 따른 질환 예측 방법은 단계S21에서 예측 대상자의 건강 데이터를 수신할 수 있다.
- [0131] 다음으로, 단계S22에서는 전자건강기록(EHR) 및 모바일 기기를 통해 획득된 모바일 기반의 개인건강기록(mPHR)에 기초하여 생성된 통합 데이터셋을 이용하여 질환 예측 모델을 생성할 수 있다.
- [0132] 또한, 단계S22에서는 건강 데이터가 수신된 이후에 예측 대상자의 특성(상태) 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 질환 예측 모델을 생성할 수 있다.
- [0133] 또한, 단계S22에서는 통합 데이터셋에 포함된 데이터를 기반으로 복수 사용자의 특성이 고려된 복수의 질환 예측 모델을 생성하고, 건강 데이터가 수신된 이후에 예측 대상자의 특성(상태) 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 복수의 질환 예측 모델 중 예측 대상자에 최적화된 최적 질환 예측 모델을 선택할 수 있다. 이후, 단계S23에서는 최적 질환 예측 모델에 기초하여 예측 대상자에 대한 질환 가능성을 예측할 수 있다.
- [0134] 또한, 단계S22에서의 질환 예측 모델은, 통합 데이터셋에 포함된 데이터의 복수의 속성과 관련된 복수의 규칙을 포함할 수 있다. 또한, 복수의 규칙의 조합 수 및 조합 순서는 예측 대상자의 특성 및 예측하고자 하는 질환 유형 중 적어도 하나를 고려하여 결정될 수 있다.
- [0135] 또한, 질환 예측 모델은, 통합 데이터셋에 포함된 데이터에 대하여 의사결정트리(decision tree)를 적용함으로써 생성될 수 있다.
- [0136] 또한, 단계S22에서의 통합 데이터셋은, 전자건강기록을 포함하는 제1 데이터셋 및 모바일 기기를 통해 획득되는 개인건강기록을 포함하는 제2 데이터셋 각각에 포함된 데이터에 대한 전처리를 수행하고, 전처리된 제1 데이터셋의 데이터 및 전처리된 제2 데이터셋의 데이터 간에 매칭을 통해 생성될 수 있다.
- [0137] 다음으로, 단계S23에서는 질환 예측 모델에 기초하여 건강 데이터에 대응하는 예측 대상자에 대한 질환 가능성을 예측할 수 있다.
- [0138] 상술한 설명에서, 단계 S21 내지 S23은 본원의 구현예에 따라서, 추가적인 단계들로 더 분할되거나, 더 적은 단계들로 조합될 수 있다. 또한, 일부 단계는 필요에 따라 생략될 수도 있고, 단계 간의 순서가 변경될 수도 있다.
- [0139] 본원의 일 실시 예에 따른 데이터셋 생성 방법 및 질환 예측 방법은 다양한 컴퓨터 수단을 통하여 수행될 수 있는 프로그램 명령 형태로 구현되어 컴퓨터 판독 가능 매체에 기록될 수 있다. 상기 컴퓨터 판독 가능 매체는 프로그램 명령, 데이터 파일, 데이터 구조 등을 단독으로 또는 조합하여 포함할 수 있다. 상기 매체에 기록되는 프로그램 명령은 본 발명을 위하여 특별히 설계되고 구성된 것들이거나 컴퓨터 소프트웨어 당업자에게 공지되어 사용 가능한 것일 수도 있다. 컴퓨터 판독 가능 기록 매체의 예에는 하드 디스크, 플로피 디스크 및 자기 테이프와 같은 자기 매체(magnetic media), CD-ROM, DVD와 같은 광기록 매체(optical media), 플롭티컬 디스크(floptical disk)와 같은 자기-광 매체(magneto-optical media), 및 롬(ROM), 램(RAM), 플래시 메모리 등과 같은 프로그램 명령을 저장하고 수행하도록 특별히 구성된 하드웨어 장치가 포함된다. 프로그램 명령의 예에는 컴파일러에 의해 만들어지는 것과 같은 기계어 코드뿐만 아니라 인터프리터 등을 사용해서 컴퓨터에 의해서 실행될 수 있는 고급 언어 코드를 포함한다. 상기된 하드웨어 장치는 본 발명의 동작을 수행하기 위해 하나 이상의 소프트웨어 모듈로서 작동하도록 구성될 수 있으며, 그 역도 마찬가지이다.
- [0140] 전술한 본원의 설명은 예시를 위한 것이며, 본원이 속하는 기술분야의 통상의 지식을 가진 자는 본원의 기술적 사상이나 필수적인 특징을 변경하지 않고서 다른 구체적인 형태로 쉽게 변형이 가능하다는 것을 이해할 수 있을 것이다. 그러므로 이상에서 기술한 실시예들은 모든 면에서 예시적인 것이며 한정적이 아닌 것으로 이해해야만 한다. 예를 들어, 단일형으로 설명되어 있는 각 구성 요소는 분산되어 실시될 수도 있으며, 마찬가지로 분산된 것으로 설명되어 있는 구성 요소들도 결합된 형태로 실시될 수 있다.
- [0141] 본원의 범위는 상기 상세한 설명보다는 후술하는 특허청구범위에 의하여 나타내어지며, 특허청구범위의 의미 및 범위 그리고 그 균등 개념으로부터 도출되는 모든 변경 또는 변형된 형태가 본원의 범위에 포함되는 것으로 해석되어야 한다.

부호의 설명

- [0142]
- 100: 데이터셋 생성 시스템
 - 30: 데이터셋 생성 장치
 - 31: 데이터셋 생성부
 - 32: 전처리부
 - 33: 매칭부
 - 200: 질환 예측 시스템
 - 70: 질환 예측 장치
 - 71: 수신부
 - 72: 생성부
 - 73: 예측부

도면

도면1



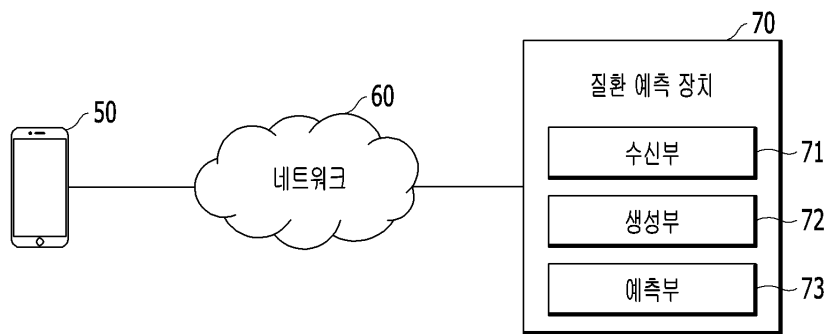
도면2

Description of EHR attributes.

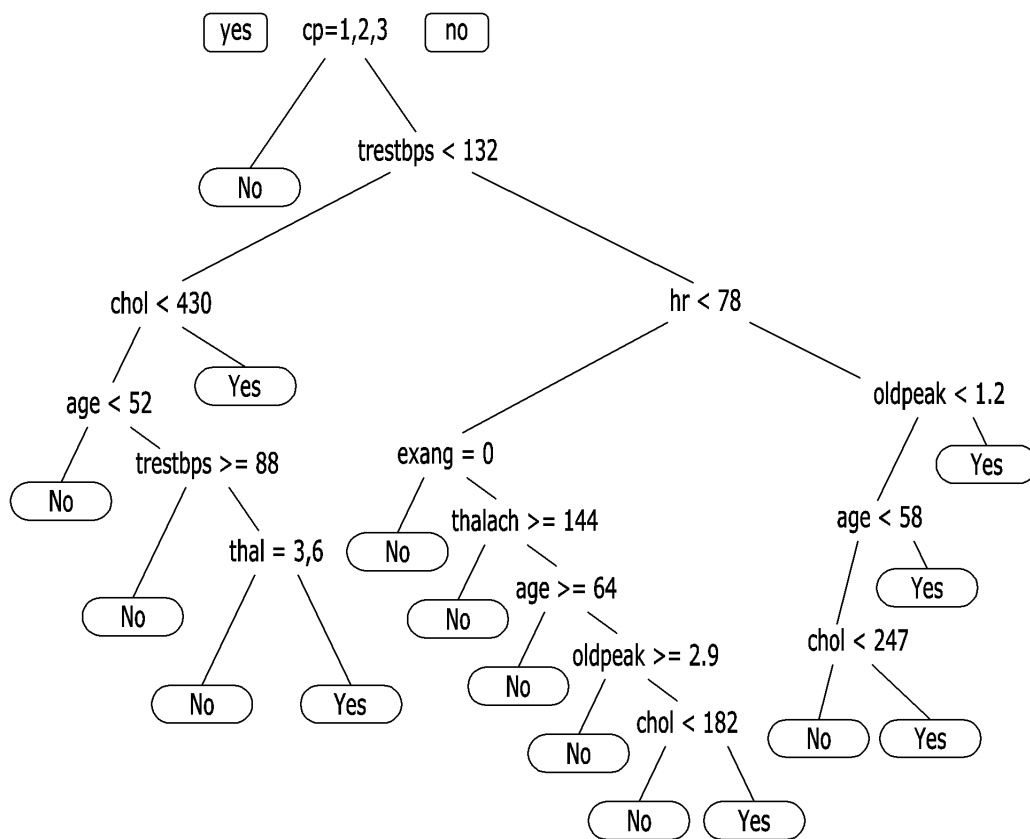
Attribute	Description	Example of value
age	age in years	63
sex	sex (1=male, 0=female)	1
cp	chest pain type (1=typical angina, 2=atypical angina, 3=non-anginal pain, 4=asymptomatic)	1
trestbps	resting blood pressure	145
chol	serum cholesterol	233
fbs	fasting blood sugar > 120 mg/dl (1=true, 0=false)	1
restecg	resting electrocardiographic results (0=normal, 1=having ST-T wave abnormality, 2=showing probable or definite left ventricular hypertrophy)	2
thalach	maximum heart rate achieved	150
exang	exercise induced angina (1=yes, 0=no)	0
oldpeak	ST depression induced by exercise relative to rest	2.3
slope	the slop of the peak exercise ST segment (1=upsloping, 2=flat, 3=downsloping)	3
ca	number of major vessels colored by flourosopy	0
thal	thalassemia (3=normal, 6=fixed defect, 7=diagnosis of heart disease)	6
num	existence of heart disease (0=no, 1=yes)	0

도면3

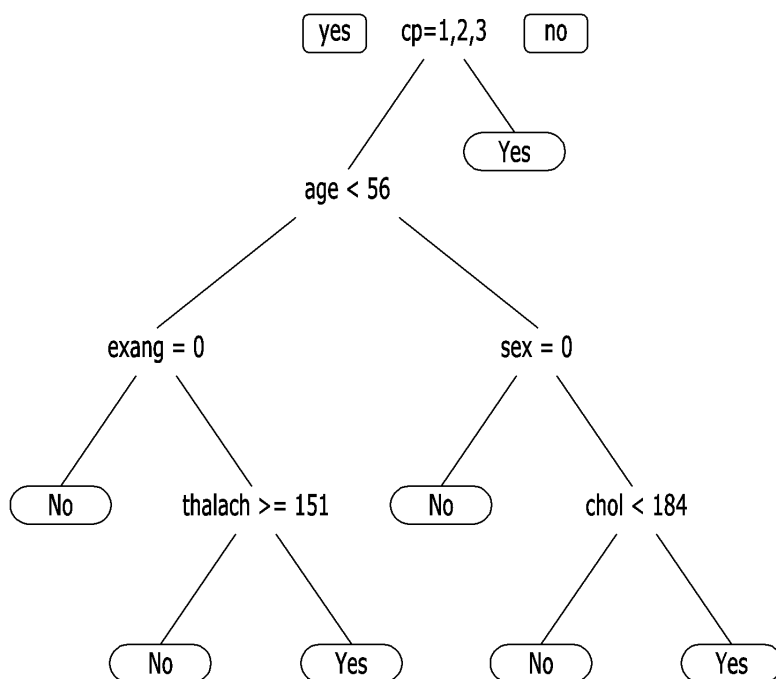
200



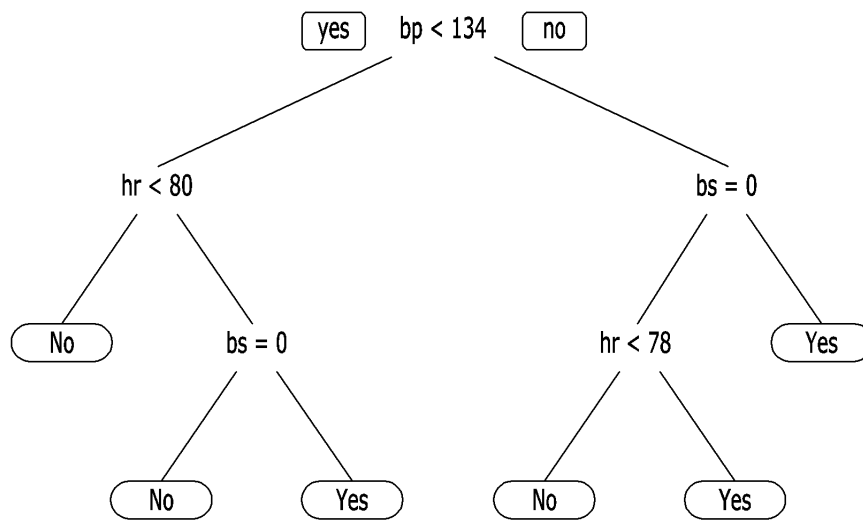
도면4



도면5



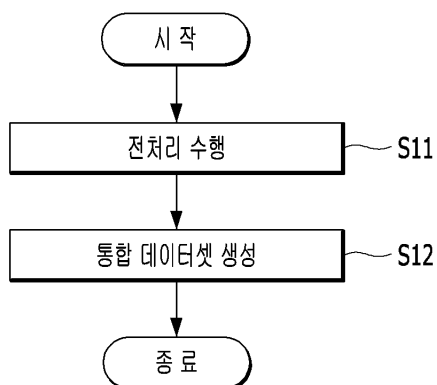
도면6



도면7

	integrated dataset	EHR	mPHR
accuracy	0.82	0.79	0.78

도면8



도면9

