

(19) 대한민국특허청(KR)
(12) 공개특허공보(A)(11) 공개번호 10-2022-0075700
(43) 공개일자 2022년06월08일

(51) 국제특허분류(Int. Cl.)

G16B 40/20 (2019.01) G16B 20/20 (2019.01)

G16H 10/60 (2018.01) G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

(52) CPC특허분류

G16B 40/20 (2019.02)

G16B 20/20 (2019.02)

(21) 출원번호 10-2020-0164072

(22) 출원일자 2020년11월30일

심사청구일자 없음

(71) 출원인

울산과학기술원

울산광역시 울주군 언양읍 유니스트길 50

고신대학교 산학협력단

부산광역시 영도구 와치로 194(동삼동)

(72) 발명자

이정혜

울산광역시 울주군 언양읍 유니스트길 50

한석주

울산광역시 울주군 언양읍 유니스트길 50

(뒷면에 계속)

(74) 대리인

특허법인태백

전체 청구항 수 : 총 4 항

(54) 발명의 명칭 gPRS를 이용한 제2형 당뇨병질환 예측 시스템

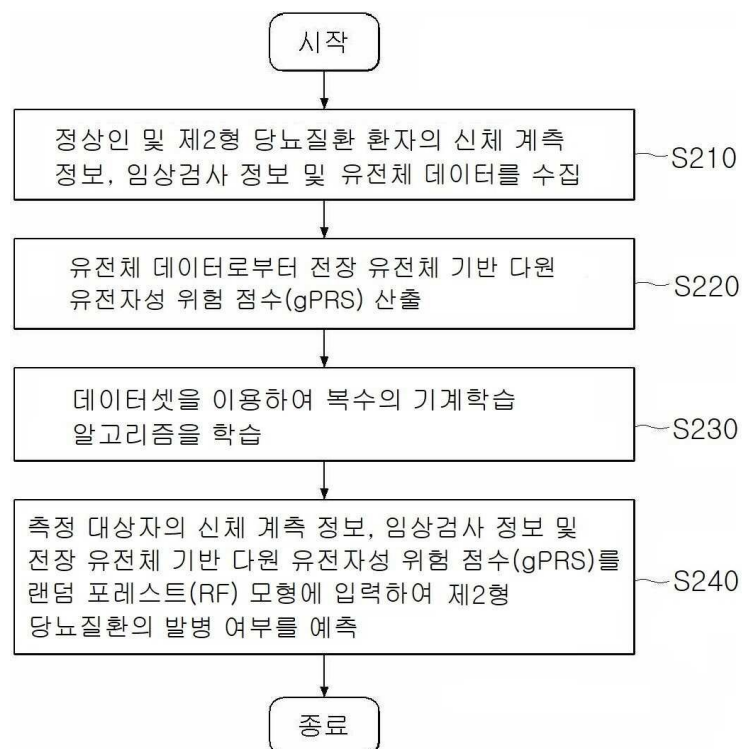
(57) 요약

본 발명은 gPRS를 이용한 제2형 당뇨병질환 예측 시스템에 대한 것이다.

본 발명에 따른 제2형 당뇨병질환 예측 시스템은 코호트 자료를 이용하여 당뇨병질환을 가진 환자와 정상인으로부터 신체 계측 정보, 임상 검사 정보 및 유전체 데이터를 획득하는 정보수집부, 상기 유전체 데이터를 이용하여 당뇨

(뒷면에 계속)

대표도 - 도2



질환에 대한 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 획득하는 제어부, 상기 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 교차 조합하여 생성된 데이터셋을 이용하여 랜덤 포레스트(RF) 모형을 학습시키는 학습부, 그리고 상기 랜덤 포레스트(RF) 모형에 측정 대상자의 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 입력하여 제2형 당뇨병 발생 여부를 예측하는 예측부를 포함한다.

이와 같이 본 발명에 따르면, 유전체 데이터를 기반으로 산출된 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 변수로 이용함으로써 예측 모형의 제2형 당뇨병 발생 예측 정확도를 향상시킬 수 있다.

(52) CPC특허분류

G16H 10/60 (2021.08)

G16H 50/50 (2018.01)

G16H 50/70 (2018.01)

강지훈

부산광역시 부산진구 진남로356번길 78, 605호(전포동, 전포화신거화2차 아파트)

(72) 발명자

김수현

울산광역시 울주군 언양읍 유니스트길 50

이 발명을 지원한 국가연구개발사업

과제고유번호	1711105477
과제번호	2019M3E5D1A02070863
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	혁신형의사과학자공동연구사업(R&D)
연구과제명	유전체 빅데이터 기반의 당뇨예측 모델 및 맞춤 치료상품 개발
기 여 율	1/1
과제수행기관명	고신대학교
연구기간	2020.01.01 ~ 2020.12.31

명세서

청구범위

청구항 1

gPRS를 이용한 제2형 당뇨병 예측 시스템에 있어서,

코호트 자료를 이용하여 당뇨병을 가진 환자와 정상인으로부터 신체 측정 정보, 임상 검사 정보 및 유전체 데이터를 획득하는 정보수집부,

상기 유전체 데이터를 이용하여 당뇨병에 대한 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 획득하는 제어부,

상기 신체 측정 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 교차 조합하여 생성된 데이터셋을 이용하여 랜덤 포레스트(RF) 모형을 학습시키는 학습부, 그리고

상기 랜덤 포레스트(RF) 모형에 측정 대상자의 신체 측정 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 입력하여 제2형 당뇨병 발생 여부를 예측하는 예측부를 포함하는 제2형 당뇨병 예측 시스템.

청구항 2

제1항에 있어서,

상기 제어부는,

상기 추출된 유전체 데이터를 이용하여 제2형 당뇨병과 연관성이 있는 단일 염기 다형성(SNP)을 획득하고, 획득한 단일 염기 다형성(SNP)을 점수화하여 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 획득하는 제2형 당뇨병 예측 시스템.

청구항 3

제1항에 있어서,

상기 학습부는,

상기 신체 측정 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 이용하여 데이터셋을 형성하고, 형성된 데이터셋을 랜덤 포레스트(RF) 모형에 입력하여 학습시키는 제2형 당뇨병 예측 시스템.

청구항 4

제1항에 있어서,

상기 학습부는,

상기 신체 측정 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 이용하여 데이터셋을 형성하고, 형성된 데이터셋을 랜덤 포레스트(RF) 모형에 입력하여 학습시키는 제2형 당뇨병 예측 시스템.

발명의 설명

기술 분야

[0001] 본 발명은 gPRS를 이용한 제2형 당뇨병 예측 시스템에 관한 것으로서, 더욱 상세하게는 코호트 자료를 기반으로 산출된 전장 유전체 기반 다원 유전자성 위험 점수(genome-wide Polygenic Risk Score, gPRS)를 이용하여 제2형 당뇨병을 예측하는 당뇨병 예측 시스템에 관한 것이다.

배경 기술

- [0002] 당뇨병은 발생하는 기전에 따라 제1형과 제2형으로 나뉜다.
- [0003] 제1형 당뇨병은 췌장에서 분비되는 인슐린의 감소에 따라 주로 청소년기 이전에 발생하는 것으로, 소아 당뇨병이라고도 한다. 반면에 제2형 당뇨병은 췌장의 인슐린 분비 능력은 비교적 유지되고 있으나 비만 등의 다른 요인으로 인해 인슐린이 작용하지 못하여 발생하는 것으로, 성인형 당뇨병이라고도 한다.
- [0004] 제2형 당뇨병은 초기 증상이 약하게 드러나는 편으로 환자가 자각하지 못하는 경우가 대부분이다.
- [0005] 기존에는 가족력, 과체중 유무 및 생활 습관 등을 통해 제 2형 당뇨병의 발병 여부를 예측하였으나, 최근에는 유전적 정보에 기반하여 제2형 당뇨병의 위험성 예측의 개선에 초점을 둔 연구가 점진적으로 진행되는 추세이다.
- [0006] 본 발명의 배경이 되는 기술은 대한민국 등록특허공보 제10-21722107호(2017.03.27. 공고)에 개시되어 있다.

발명의 내용

해결하려는 과제

- [0007] 본 발명이 이루고자 하는 기술적 과제는 코호트 자료를 기반으로 산출된 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 이용하여 제2형 당뇨병을 예측하는 당뇨병 예측 시스템을 제공하기 위한 것이다.

과제의 해결 수단

- [0008] 이러한 기술적 과제를 이루기 위한 본 발명의 실시예에 따른 gPRS를 이용한 제2형 당뇨병 예측 시스템에 있어서, 코호트 자료를 이용하여 당뇨병을 가진 환자와 정상인으로부터 신체 계측 정보, 임상 검사 정보 및 유전체 데이터를 획득하는 정보수집부, 상기 획득한 유전체 데이터를 이용하여 제2형 당뇨병에 대한 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 획득하는 제어부, 상기 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 교차 조합하여 생성된 데이터셋을 이용하여 랜덤 포레스트(Random Forest, RF) 모델을 학습시키는 학습부, 그리고 상기 랜덤 포레스트(RF) 모델에 측정 대상자의 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 입력하여 당뇨병 발생 여부를 예측하는 예측부를 포함한다.
- [0009] 상기 제어부는, 상기 추출된 유전체 데이터를 이용하여 제2형 당뇨병과 연관성이 있는 단일 염기 다형성(Single Nucleotide polymorphism, SNP)을 획득하고, 획득한 단일 염기 다형성(SNP)을 점수화하여 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 획득할 수 있다.
- [0010] 상기 학습부는, 상기 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 이용하여 데이터셋을 생성하고, 생성된 데이터셋을 랜덤 포레스트(RF) 모델에 입력하여 학습시킬 수 있다.
- [0011] 상기 학습부는, 상기 신체 계측 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 이용하여 데이터셋을 생성하고, 생성된 데이터셋을 랜덤 포레스트(RF) 모델에 입력하여 학습시킬 수 있다.

발명의 효과

- [0012] 이와 같이 본 발명에 따르면, 유전체 데이터를 기반으로 산출된 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 변수로 이용함으로써 예측 모형의 제2형 당뇨병 예측 정확도를 향상시킬 수 있다. 또한, 제2형 당뇨병 예측 정확도가 높은 기계학습모형을 추출하고, 추출된 기계학습모형에 측정 대상자의 유전체 기반 데이터를 입력하여 제2형 당뇨병 발생 여부를 예측함으로써 예측 성능을 더욱 향상시킬 수 있다.

도면의 간단한 설명

- [0013] 도 1은 본 발명의 실시예에 따른 제2형 당뇨병 예측 시스템을 설명하기 위한 도면이다.
- 도 2는 본 발명의 실시예에 따른 제2형 당뇨병 예측 시스템을 이용한 제2형 당뇨병 예측방법을 설명하기 위한 순서도이다.
- 도 3은 도 2에 도시된 S210단계에서 데이터를 획득하는 방법에 대해 설명하기 위한 도면이다.
- 도 4는 도 2에 도시된 S230단계에서 진행된 성능평가 결과를 나타내는 도면이다.

도 5는 도 2에 도시된 S230단계에서 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)가 제2형 당뇨병 질환 발생 여부 예측에 있어서의 효용성을 나타내는 도면이다.

발명을 실시하기 위한 구체적인 내용

- [0014] 이하 첨부된 도면을 참조하여 본 발명에 따른 바람직한 실시 예를 상세히 설명하기로 한다. 이 과정에서 도면에 도시된 선들의 두께나 구성요소의 크기 등은 설명의 명료성과 편의상 과장되게 도시되어 있을 수 있다.
- [0015] 또한 후술되는 용어들은 본 발명에서의 기능을 고려하여 정의된 용어들로서, 이는 사용자, 운용자의 의도 또는 관례에 따라 달라질 수 있다. 그러므로 이러한 용어들에 대한 정의는 본 명세서 전반에 걸친 내용을 토대로 내려져야 할 것이다.
- [0016] 먼저, 도 1을 통해 본 발명의 실시 예에 따른 제2형 당뇨병질환 예측 시스템에 대하여 설명한다.
- [0017] 도 1은 본 발명의 실시예에 따른 제2형 당뇨병질환 예측 시스템을 설명하기 위한 도면이다.
- [0018] 도 1에서와 같이 본 발명의 실시 예에 따른 제2형 당뇨병질환 예측 시스템(100)은 정보수집부(110), 제어부(120), 학습부(130) 및 예측부(140)를 포함한다.
- [0019] 먼저, 정보수집부(110)는 제2형 당뇨병질환을 가진 환자와 정상인으로부터 획득한 신체 계측 정보, 임상 검사 정보 및 유전체 데이터를 수집한다. 부연하자면, 정보수집부(110)는 질병관리본부 국립보건연구원에 기 구축된 코호트 자료를 이용하여 건강 및 생활습관 관련 설문조사와 검진을 통한 신체 계측 정보, 임상 검사 정보 및 유전체 데이터를 수집한다.
- [0020] 제어부(120)는 수집된 유전체 데이터를 이용하여 제2형 당뇨병질환에 대한 유전체 기반 다원 유전자성 위험 점수(gPRS)를 획득한다.
- [0021] 이를 다시 설명하면, 제어부(120)는 코호트로부터 획득한 유전체 데이터를 이용하여 단일 염기 다형성(SNP)에 대한 분석 정보를 획득한다. 그리고, 제어부(120)는 획득한 단일 염기 다형성(SNP)를 이용하여 제2형 당뇨병질환 발생과 연관성이 있는 유전자를 추출하고, 추출된 유전자의 위험도에 따라 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 산출한다.
- [0022] 학습부(130)는 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 이용하여 데이터셋을 생성하고, 생성된 데이터셋을 이용하여 랜덤 포레스트(RF) 모형을 학습시킨다.
- [0023] 마지막으로 예측부(140)는 학습이 완료된 랜덤 포레스트(RF) 모형에 측정 대상자의 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 입력하여 제2형 당뇨병질환 발생 여부를 예측한다.
- [0024] 이하에서는 도 2 내지 도 5를 이용하여 본 발명의 실시예에 따른 제2형 당뇨병질환 예측 시스템(100)을 이용하여 제2형 당뇨병질환 예측방법에 대해 더욱 상세하게 설명한다.
- [0025] 도 2는 본 발명의 실시예에 따른 제2형 당뇨병질환 예측 시스템을 이용한 제2형 당뇨병질환 예측방법을 설명하기 위한 순서도이고, 도 3은 도 2에 도시된 S210단계에서 데이터를 획득하는 방법에 대해 설명하기 위한 도면이다.
- [0026] 도 2에 도시된 바와 같이, 본 발명의 실시예에 따른 제2형 당뇨병질환 예측 시스템(100)은 정상인 및 제2형 당뇨병질환 환자의 신체 계측 정보, 임상 검사 정보 및 유전체 데이터를 수집한다(S210).
- [0027] 부연하자면, 정보수집부(110)는 코호트를 통해 신체 계측 정보, 임상 검사 정보 및 유전체 데이터를 획득한다. 여기서 코호트는 지역사회기반 코호트(안산, 안성), 도시기반 코호트(HEXA) 및 농촌기반 코호트(CAVAS)를 포함한다.
- [0028] 도 3에 도시된 바와 같이, 정보수집부(110)는 2001년에서부터 2016년까지 공개된 코호트를 기반으로 총 5,493명의 자료를 획득한다. 그리고, 정보수집부(110)는 획득한 자료 중에서 하기의 기준을 이용하여 일부 자료를 제외한다.
- [0029] 자세히는 2001년에서 2002년 사이에 획득한 코호트 자료를 기준으로 당뇨병질환 질환을 진단받았거나, 당뇨병질환 질환을 치료하였거나, 인슐린을 이용하여 치료를 받았거나, 측정된 공복 혈당 수치를 이용하여 806명의 자료를 제외한다.
- [0030] 그 다음, 정보수집부(110)는 4,687명의 자료 중에서 2001년에서 2002년 사이에 획득한 코호트 자료를 기준으로 당뇨병질환 질환을 진단받았거나, 당뇨병질환 질환을 치료하였거나, 인슐린을 이용하여 치료를 받았거나, 측정된 공

복 혈당 수치를 이용하여 3,585명의 자료를 제외한다.

- [0031] 마지막으로 정보수집부(110)는 1,102명의 자료 중에서 결측치가 존재하는 자료, 약 23%를 제외하고 최종적으로 획득한 1,095명의 자료를 이용하여 분석을 수행한다. 그리고 정보수집부(110)는 분석에 따라 신체 계측 정보, 임상 검사 정보 및 유전체 정보를 획득한다.
- [0032] 여기서 신체 계측 정보는 성별, 나이, 고혈압 여부, 신체질량지수(BMI), 가족력, 흡연 여부, 섭취 알코올량 중에서 적어도 하나를 포함한다.
- [0033] 그리고, 임상 검사 정보는 당화혈색소(HbA1C)(%), 혈당(glucose) (mg/dL), 총 콜레스테롤(mg/dL), HDL 콜레스테롤(mg/dL), 중성 지방(Triglycerides)(mg/dL), 메타세포의 분비기능(HOMA-B)(mg/dL) 및 인슐린 저항성(HOMA-IR)(mg/Dl)중에서 적어도 하나를 포함한다.
- [0034] 제어부(120)는 획득한 자료를 기반으로 유전형 분석을 수행하여 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 산출한다(S220).
- [0035] 제어부(120)는 코호트를 통해 획득한 유전체 데이터를 로지스틱 회귀 (Logistic Regression, LR) 모델을 통해 분석하여 계층 차원의 제2형 당뇨병 유전자좌를 추출한다.
- [0036] 그 다음, 제어부(120)는 본페로니 보정(Bonferroni correction)을 이용하여 제2형 당뇨병과 연관성이 있는 단일 염기 다형성(SNP)을 분석한다.
- [0037] 이때, 본페로니 보정을 위해 제2형 당뇨병과 계층 전체 연관성의 유의 임계값을 $P < 5 \times 10^{-8}$ 로 설정하고, 위양성 오류의 임계치는 $P < 0.05$ 로 설정한다.
- [0038] 그리고, 제어부(120)는 각각의 단일 염기 다형성(SNP)별 영향력을 고려하여 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 계산한다.
- [0039] S220단계가 완료되면, 학습부(130)는 생성된 데이터셋을 이용하여 복수의 기계학습 알고리즘을 학습시킨다(S230).
- [0040] 학습부(130)는 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 교차 조합하여 복수의 데이터셋을 생성한다. 부연하자면, 학습부(130)는 신체 계측 정보로부터 추출된 데이터를 이용하여 제1 타입 데이터셋을 생성하고, 신체 계측 정보와 임상 검사 정보로부터 추출된 데이터를 이용하여 제2 타입 데이터셋을 생성한다. 또한, 학습부(130)는 신체 계측 정보와 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)로부터 추출된 데이터를 이용하여 제3 타입 데이터셋을 생성하고, 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)로부터 추출된 데이터를 이용하여 제4 타입 데이터셋을 생성한다.
- [0041] 그 다음, 학습부(130)는 생성된 제1 타입 데이터셋 내지 제4 타입 데이터셋을 이용하여 로지스틱 회귀(LR) 모델, 랜덤 포레스트(RF) 모델 및 심층 신경망(Deep Neural Network, DNN) 모델을 각각 학습시킨다.
- [0042] 그리고 학습부(130)는 학습이 완료된 로지스틱 회귀(LR) 모델, 랜덤 포레스트(RF) 모델 및 심층 신경망(DNN) 모델의 성능평가를 진행한다. 이때, 복수의 기계 학습 알고리즘에 대한 성능평가는 곡선하면적(area under the curve, AUC) 값을 이용하여 획득한다. 한편, 곡선하면적(AUC) 값이 1에 가까울수록 예측 모형의 정확도가 높다고 판단한다.
- [0043] 도 4는 도 2에 도시된 S230단계에서 진행된 성능평가 결과를 나타내는 도면이고, 도 5는 도 2에 도시된 S230단계에서 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)가 제2형 당뇨병 발생 여부 예측에 있어서 효용성을 나타내는 도면이다.
- [0044] 도 4에 도시된 바와 같이, 제1 타입 데이터셋을 이용하여 학습시켰을 경우에는 심층 신경망(DNN) 모형의 곡선하면적(AUC) 값이 가장 크게 산출되고, 제2 타입 데이터셋을 이용하여 학습시켰을 경우에는 랜덤 포레스트(RF) 모형의 곡선하면적(AUC) 값이 가장 크게 산출되었다. 또한, 제3 타입 데이터셋 및 제4 타입 데이터셋을 이용하여 학습시켰을 경우에는 랜덤 포레스트(RF) 모형의 곡선하면적(AUC) 값이 가장 크게 산출되었다.
- [0045] 한편, 데이터셋은 제1 타입 데이터셋에서 제4 타입 데이터셋으로 갈수록 변수가 추가되는 형태로 구성되며, 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)가 포함되었을 경우에 기계 학습 알고리즘을 이용한 제2형 당뇨병 질환의 예측 정확도가 향상되었다.
- [0046] 또한, 도 5에 도시된 바와 같이, 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)는 실제로 당뇨 발병 환자군

과 비 발병 환자군을 구별하는 효용성을 가진다.

[0047] 마지막으로 예측부(140)는 측정 대상자의 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 랜덤 포레스트(RF) 모형에 입력하여 제2형 당뇨병의 발병 여부를 예측한다(S240).

[0048] 부연하자면, 예측부(140)는 로지스틱 회귀(LR) 모형, 랜덤 포레스트(RF) 모형 및 심층 신경망(DNN) 모형 중에서 예측 정확도가 가장 높은 랜덤 포레스트(RF) 모형을 이용하여 측정 대상자의 제2형 당뇨병 발생 여부를 예측한다.

[0049] 이때, 예측부(140)는 측정 대상자의 신체 계측 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 랜덤 포레스트(RF) 모형에 입력하여 제2형 당뇨병 발생 여부를 예측할 수도 있고, 신체 계측 정보, 임상 검사 정보 및 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 랜덤 포레스트(RF) 모형에 입력하여 제2형 당뇨병 발생 여부를 예측할 수도 있다.

[0050] 본 발명의 실시예에서의 예측부는 랜덤 포레스트(RF) 모형을 이용하여 측정 대상자의 제2형 당뇨병 발생 여부를 예측하였으나, 이에 한정하지 않고 로지스틱 회귀(LR) 모형 또는 심층 신경망(DNN) 모형을 통해서 측정 대상자의 제2형 당뇨병 발생 여부를 예측할 수도 있다.

[0051] 이와 같이 본 발명에 따른 당뇨병 예측 장치는 유전체 데이터를 기반으로 산출된 전장 유전체 기반 다원 유전자성 위험 점수(gPRS)를 변수로 이용함으로써 예측 모형의 제2형 당뇨병 예측 정확도를 향상시킬 수 있다. 또한, 제2형 당뇨병 예측 정확도가 높은 기계학습모형을 추출하고, 추출된 기계학습모형에 측정 대상자의 유전체 기반 데이터를 입력하여 제2형 당뇨병 발생 여부를 예측함으로써 예측 성능을 더욱 향상시킬 수 있다.

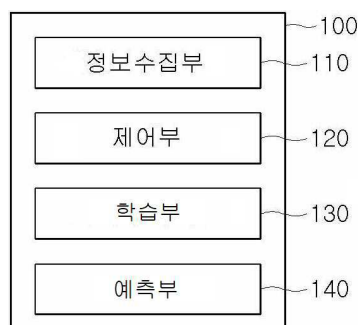
[0053] 본 발명은 도면에 도시된 실시 예를 참고로 하여 설명되었으나 이는 예시적인 것에 불과하며, 당해 기술이 속하는 분야에서 통상의 지식을 가진 자라면 이로부터 다양한 변형 및 균등한 타 실시 예가 가능하다는 점을 이해할 것이다. 따라서 본 발명의 진정한 기술적 보호범위는 아래의 특허청구범위의 기술적 사상에 의하여 정해져야 할 것이다.

부호의 설명

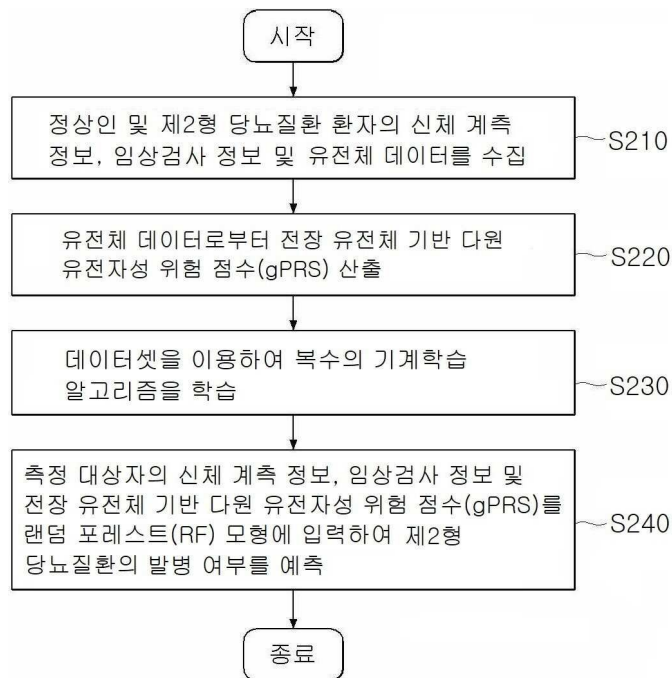
- [0054] 100 : 제2형 당뇨병 예측 시스템
110 : 정보수집부
120 : 제어부
130 : 학습부
140 : 예측부

도면

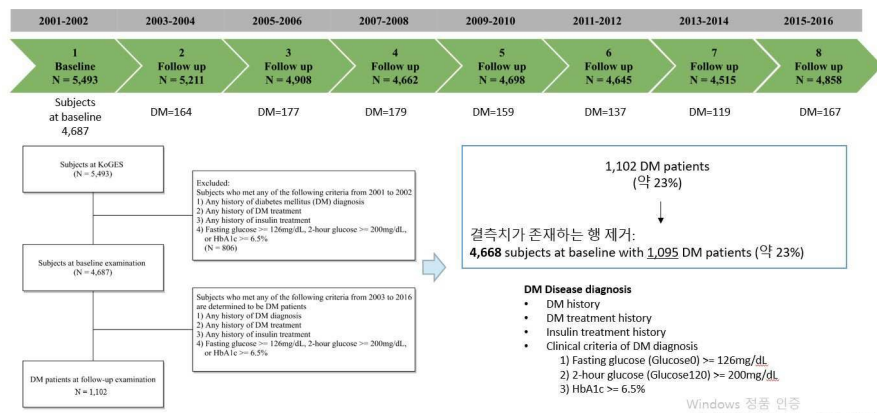
도면1



도면2



도면3



도면4

Type	Model	Accuracy	Sensitivity	Specificity	Precision	F1	Brier	Logloss	Threshold	AUC	P-value
Type 1 (Demo.)	LR	0.618	0.647	0.610	0.342	0.440	0.238	0.670	0.499	0.626±0.065	
	RF	0.600	0.650	0.585	0.325	0.430	0.237	0.667	0.504	0.614±0.050	
	DNN	0.564	0.694	0.524	0.309	0.428	0.171	0.522	0.232	0.645±0.042	
Type 2 (Demo. + Clin.)	LR	0.799	0.831	0.790	0.570	0.667	0.160	0.486	0.531	0.846±0.045	$p<0.001$ (compared Type1)
	RF	0.793	0.798	0.792	0.552	0.646	0.141	0.434	0.455	0.851±0.040	$p<0.001$ (compared Type1)
	DNN	0.695	0.694	0.695	0.411	0.516	0.147	0.459	0.245	0.767±0.037	$p<0.001$ (compared Type1)
Type 3 (Demo. + gPRS)	LR	0.875	0.686	0.932	0.780	0.719	0.140	0.438	0.664	0.845±0.037	$p<0.001$ (compared Type1)
	RF	0.863	0.689	0.916	0.731	0.702	0.124	0.408	0.490	0.849±0.034	$p<0.001$ (compared Type1)
	DNN	0.833	0.616	0.899	0.652	0.634	0.101	0.344	0.194	0.824±0.037	$p<0.001$ (compared Type1)
Type 4 (Demo. + Clin. + gPRS)	LR	0.864	0.837	0.872	0.674	0.743	0.110	0.354	0.519	0.911±0.028	$p<0.001$ (compared Type3)
	RF	0.867	0.828	0.879	0.694	0.748	0.096	0.323	0.432	0.917±0.029	$p<0.001$ (compared Type3)
	DNN	0.863	0.699	0.913	0.712	0.705	0.092	0.314	0.243	0.863±0.034	$p<0.001$ (compared Type3)

도면5

