scientific reports

Check for updates

OPEN

머신 러닝 (Machine Learning)은 비트로 이전의 라이브 비트 발생을 예측합니다

Ashish Goyal, Maheshwar Kuchana & Kameswari Prasada Rao ay Yagari ✉

내보다 비트로 수정 (IVF)은 자궁 내막증, 난자 품질 불량, 어머니 또는 아버지의 유전자 질환, 배란 문제, 정자 또는 계란에 해를 끼치는 항체 문제, 정자가 침투하거나 생존 할 수없는 것과 같은 합병증을 해결하는 인기있는 방법입니다. 자궁 경부 점액과 낮은 정자 수에서 인간 불임이 발생합니다. 그럼에도 불구하고 IVF는 수정에서의 성공을 보장하지 않습니다. IVF를 선택하는 것은 결과에서 높은 비용과 불확실성의 이유로 부담이됩니다. IVF 과정에서 합병증과 수정 요인이 많기 때문에 이는 번거로운 작업입니다.

다산 의사는 성공적인 출생을 정확하게 예측합니다. 인공 지능 (AI) 은이 연구에서 라이브 비버 발생을 예측하기 위해 사용되었습니다. 이 작품은 주로 기증자가 아닌 부부로부터 배아가 형성 될 때 라이브 바로 발생을 예측하는 데 중점을 둡니다. 여기, 우리는 인간 수정 및 배아 당국 (HFEA)이 제공하는 공개적으로 이용 가능한 데이터 세트에 대한 고전적인 기계 학습, 딥 러닝 아키텍처 및 알고리즘 앙상블을 포함한 다양한 AI 알고리즘을 비교합니다. 혼동 행렬, F1 스코어, 정밀, 리콜, 수신기 작동 특성 (ROC) 곡선과 같은 데이터 및 메트릭에 대한 통찰력은 후속 섹션에서 시연됩니다. 교육 프로세스에는 분류기 모델을 훈련시키기 위해 두 가지 설정 ***Without feature selection***와 ***With feature selection***가 있습니다. 머신 러닝, 딥 러닝, 앙상블 모델 분류 패러다임은 두 설정 모두에서 교육을 받았습니다. 임의의 산림 모델은 76.49%의 가장 높은 F1 스코어를 달성합니다.

without feature selection 설정에서. 동일한 모델의 경우 ROC 곡선 (ROC AUC) 점수에 따른 정밀, 리콜 및 면적은 각각 77%, 76%및 84.60%입니다. 임신의 성공은 남성과 여성의 특성과 생활 조건에 따라 다릅니다. 이 연구는 비트로 수정에서 임상 적으로 관련된 매개 변수를 통해 성공적인 임신을 예측합니다. 따라서 인공 지능
진단, 예후, 치료 등을 뒷받침하는 의사 결정 과정에서 유망한 역할을합니다.

8 천만 명이 넘는 커플이 불임의 영향을받습니다. 불임 [1]으로 인해 보호되지 않은 성관계를 갖는 거의 12 개월 후에 '실패한 개념'이 발생할 수 있습니다. 실패한 개념의 수를 줄이려면 수컷의 암컷 난소와 정자의 난자가 신체 밖에서 융합됩니다. 즉, 실험실에서 배아가 발생하여 암컷의 난소에 배치됩니다. 수정 (IVF). 경우에 따라 인공 수정은 자궁에 직접 정자를 주사함으로써 임신을 초래합니다. 전 세계 IVF에서 5 백만 명 이상의 아기가 태어 났다고보고되었습니다. IVF는 성별의 생식 특성과 관련된 다양한 문제로 인한 남성과 여성 불임을 극복하는 데 사용됩니다. IVF는 다양한 의료 및 수술 절차를 결합하여 수정에 도움이됩니다. 전체 과정에는 한 번 이상의 라운드가 있으며 임신을하는 데 몇 달이 걸릴 수 있습니다. IVF 처리의 쉬운 접근성은 IVF의 사용을 상승 시켰지만 불임 커플 [2]에 의한 것은 아닙니다. IVF 치료를 선택하는 것도 높은 비용, 성공 보장 및 치료 스트레스 [3,4]으로 인해 매우 어려운 작업으로 간주됩니다. 환자는 일반적으로 치료의 신체적, 심리적 부담으로 인해 IVF 치료를 중단합니다 (v11}.

몇몇 의료 종사자들은 전문 지식을 통해 시행 착오 방법으로 임신 가능성을 예측하고 있습니다. 따라서, 기존의 예측 방법은 체계적인 통계적 접근법을 사용하지 않는 개별 의료 전문가의 경험 수준에 의존한다. 따라서 그들은 더 주관적입니다. 의료 종사자와 환자는 IVF 치료에 대한 의사 결정을위한 측정을 간절히 찾고 있습니다. 인공 지능 (AI), 머신 러닝 (ML), 딥 러닝 (DL)과 같은 기술의 최근 발전은 통계 데이터 중심 접근 방식의 발병 문제를 해결하는 데 약속합니다. AI에 의해 주도되는 매우 정확한 분석은 의미있는 방식으로 해석하여 방대한 양의 데이터로 대부분의 과제를 근본적으로 해결할 수 있습니다. 통계적 접근 방식이 있습니다

BML Munjal University, Gurugram 122413, India. ✉email: akprasada@yahoo.com

의료 종사자가 할 수있는 생식력에 대한 예측 모델을 개발하기 위해 연구원의 관심을 끌었습니다. IVF 설정에서 성공적인 출생이 발생하면 정확한 예측을 얻으십시오.

기계 학습은 컴퓨터/시스템이 비슷한 방식으로 생각하도록 가르치고 과거 경험 [7]에 대한 학습/훈련을 통해 예측을 출력하는 연구 분야입니다. 그것은 인간의 의사 결정 능력을 모방하기 위해 시스템의 견고성을 제공하는 의미 있고 패턴 지향적 인 방식으로 데이터를 탐색합니다. 딥 러닝은 인간 신경 네트워크 [8]의 원리에 작용하는 ML의 하위 집합입니다. 많은 매개 변수로 거대한 데이터 레코드를 분석하면 인간이 데이터의 중요한 패턴을 놓칠 수 있습니다. ML과 DL은 이러한 측면에서 당신을 덮고 이러한 패턴을 쉽게 식별 할 수 있으므로 의사 결정에 인간이 도움이됩니다.

ML의 지속적인 개선으로, 개인화 된 치료, 수술 시뮬레이션, 약물 발견 및 질병 진단 가속화 [9–12]를 포함하여 몇 가지 의료 도메인이 이미이를 채택했습니다. 생식 과학에서, ML은 IVF [13]에서 배반포 전이 후 이식 예측을 예측하기 위해 적용되었다. ML은 또한 라이브 배아 출생 예측 변수를 평가하고 쌍둥이 (쌍둥이 [14–16]를 예측하기 위해 배아 선택에 대한 예측 모델을 구축하기 위해 적용되었습니다. 치명적인 심장 임신과 인간 배반포 선택을 예측하기위한 DL 기술의 현대적 사용 [17,18]도 목격되었습니다. 따라서 ML은 임상 데이터 세트에 적용하여 위험 평가, 진단, 예후 모델을 개발하고 환자 의료를 개선 할 수 있습니다.

기계 학습 과거의 일부 연구는 ML 기술을 사용하여 IVF 치료를받는 여성의 생생한 기회를 예측합니다. 초기에 가장 많이 받아 들여지는 예측 모델 중 하나는 McLernon 모델 [19,20]으로, 이산 로지스틱 회귀 만 활용하여 최대 6 개의 완전한 IVF주기를 가진 커플의 라이브 시트의 가능성을 예측합니다. 전처리 모델 (IVF 처리가 시작되기 전 예측)과 치료 후 모델 (배아 전달에 대한 첫 번째 시도 후 살아있는 생산 가능성을 예측) 인 두 가지 예측 모델이 개발됩니다. 데이터는 1999 년부터 2008 년까지 영국에서 IVF 치료를 시작한 253,417 명의 여성의 인간 수정 및 배아 당국 (HFEA)에서 자신의 알과 파트너 정자를 사용하여 수집되었습니다. C- 인덱스는 두 모델의 성능을 평가하는 데 사용되었습니다. 전처리 모델에 대한 C- 인덱스는 0.69 (0.68 – 0.69)이고 치료 후 모델의 C- 인덱스는 0.76 (0.75 – 0.77)이었다.

Rafiul Hassan et al. [21]는 IVF 임신을 더 정확하게 분석하고 예측하기 위해 5 가지 ML 모델을 갖춘 Hill-Climbing 기능 선택 알고리즘을 제안합니다. 이 연구에 대한 데이터는 2005 년 3 월부터 2008 년 1 월까지 약 3 년 동안 터키 이스탄불의 불임 클리닉에서 수집되었으며 1048 명의 환자의 불임 치료로 구성되었습니다. 이 연구는 연령, 진단, Antral Follicle Counts (AFC), 정자 품질 등과 같은 27 가지 속성을 사용했습니다. 나이는 임신 결과에 영향을 미치는 가장 영향력있는 IVF 속성이라는 것이 밝혀졌습니다. 언덕 등반 기능 선택 기술 (클래스 피어의 중요한 기능 만 선택)이 사용될 때 모든 분류기의 성능이 향상되었습니다. 전반적으로, SVM (Support Vector Machine)은 19 개의 IVF 속성을 고려하여 98.38%, F1-점수 98.4%, AUC 점수 99.5%의 정확도를 달성합니다.

ML 모델, 즉 SVM, 의사 결정 트리, 순진한 베이, K 근처 (KNN) 등의 Guvenir et al. [22]의 설문 조사에 따르면 다양한 모델이 다양한 기능을 필요로하는 것으로 나타났습니다. 잘 수행하십시오. 연령, 체질량 지수 (BMI), 정자 수 등과 같은 환자 속성은 이러한 모델을 훈련시키는 데 사용되었습니다. 이 설문 조사에서 SVM은 최대 64 개의 기능을 고려하여 84%의 정확도로 간주되는 반면, 다른 것들은 Kaufmann et al.의 인공 신경망 (ANN)과 같은 5-6 기능으로 간주되어 59 개를 초래했습니다. % 정확도 [24]. 부부가 IVF 처리 여부를 결정하는 데 도움이되는 모델 개발에 중점을 둡니다. 이 조사에서는 두 가지 문제가 해결됩니다. 하나는 IVF 치료에서 임신 가능성을 확인하고 다른 하나는 의사가 가장 실용적인 배아를 선택하도록 돕는 데 있습니다.

JIAHUI QIU et al. [25]은 로지스틱 회귀, 임의의 숲, 극한 그라디언트 부스트 (XGBOOST) 및 SVM의 네 가지 모델을 사용하여 IVF 전에 생생한 출생을 예측합니다. 데이터는 2014-2018 년 중국 의과 대학의 Shengjing 병원 의료 센터에서 첫 IVF 치료를받은 7188 명의 여성으로부터 수집됩니다. 연령, AMH, BMI, 불임 기간, 이전 출생, 이전 유산 등 및 불임의 유형 (관, 남성 인자, 무사, 설명 할 수없는 등)과 같은 속성이 고려됩니다. 교정 및 수신기 작동 특성 (ROC) 곡선은 성능 메트릭으로 사용됩니다. Xgboost는 검증 데이터 세트에서 ROC AUC (Roc Curve) 점수 0.73에서 최고 영역을 달성했으며 모든 모델 중에서 가장 좋은 교정 모델을 보여주었습니다.

생생한 발생 예측은 이진 분류 문제에 속합니다.
여성은 주어진 IVF 매개 변수에 따라 출산 여부를 제공합니다. 현재의 작업은 비교하는 것을 목표로합니다
완전한 IVF 사이클 후 라이브 제보 발생을 예측하는 다양한 모델. 작업은 주로 중점을 둡니다
배아가 기증자가 아닌 부부로부터 형성 될 때 살아있는 생산을 예측합니다. 완전한
IVF 사이클은 신선한 사이클과 한 번의 난소 자극에서 다음과 같은 동결 - 해동 사이클을 나타냅니다.
불임을 유발하는 남성과 여성과 관련된 몇 가지 생식 특성이 있습니다. 그랬어요
여성과 유사한 연령 (계란의 양 및 질 감소), 생리 장애, 생리 장애,
자궁 요인, 자궁 경부 요인, 이전 임신, 불임 기간, 여성 1 차 (환자가 불가능한 경우
최소 1 년 후 임신), 여성 2 차 (환자가 적어도 한 번 임신 할 수있는 경우
할 수없고 설명 할 수없는 요인은 불임 [26–30]을 유발하는 데 큰 영향을 미칩니다. 남성과 관련된 요인,
정액 농도, 정액 운동성, 정액 형태, 정자 양 및 정액 수와 같은 것이 필수적입니다.
남성의 불임을 테스트하기 위해 [31,32]. 남성과 여성의 위에서 구매 한 모든 중요한 생식 특성
이 연구에서 고려되었습니다. 공개 데이터 세트 [33]는
인간 수정 및 배아 당국은 세계에서 가장 오랫동안 실행되는 다산 치료 데이터베이스입니다.
이 연구에서는 94 개의 임상 특징을 가진 495,630 개의 기록 데이터가 2010 년부터 2016 년까지 인수되었습니다.
영국 전역의 IVF 센터. 데이터 청소를 수행 한 후 25 개의 필수 임상 기능을 갖춘 141,160 레코드
긍정적 클래스와 부정적인 클래스 모두 각각 70,580 개의 레코드를 포함하는 훈련 및 테스트로 간주됩니다.

ML, DL, 앙상블 학습 이이 연구에서 사용됩니다. 로지스틱 회귀 [34], K 가장 가까운 이웃 [35], 다중 계층 퍼셉트론 [36], 의사 결정 트리 [37], 1-d 딥 러닝 모델 [38]와 같은 모델은 교육 목적에 사용됩니다. Ensemble Learning [39]는 위에서 구사 한 모델의 예측에 대한 집단적 결정을 내리는 데 사용됩니다. 랜덤 포레스트 [40], adaboost [41], 투표 분류기 [42] 이이 전략에 사용됩니다. 주로 두 개의 설정이 이어집니다

이 연구에서 : 기능 선택이없는 하나의 훈련 모델과 다른 하나는 두 가지 기능 선택 기술을 사용합니다. 선형 지지대 벡터 분류기 (선형 SVC) 기반 모델과 같은 기능 선택 기술 모델 [43], 트리 기반 기능 선택 [44]가 고려됩니다. 이 두 설정에서 모델의 성능을 측정했습니다. F1- 점수, 정밀, 리콜 및 ROC-AUC 곡선과 같은 메트릭을 기반으로합니다.

이 기사의 나머지 부분은 다음과 같이 구성됩니다. 방법론 섹션에서는 기차 준비 데이터 및 모델을 훈련시키는 데 사용되는 데이터 세트, 사전 처리 기술을 설명합니다. 다양한 모델의 성능 지표 간의 포괄적 인 비교는 결과 및 토론에서 입증됩니다. 결론 섹션에서는이 연구에서 도출 될 수있는 통찰력과 향후 작업에 대해 설명합니다.

방법론

데이터 세트 설명. 이 연구에 사용 된 데이터 세트는 인간 수정 및 배아 당국 [33]에서 얻은 2010 -2016 년부터 수집 한 익명의 등록 데이터입니다. 그것은 환자 치료를 개선하기 위해 전 세계에서 가장 긴 실행되는 데이터베이스 데이터베이스 레지스터를 보유하고 환자, 기증자 및 자손 기밀을 신뢰할 수있는 보호를 보장합니다. 이 데이터 세트에는 2010 년과 2016 년 사이에 다양한 환자 연구에서 수집 된 치료주기에 대한 94 개의 특징이있는 495,630 개의 환자 기록이 포함되어 있습니다.이 연구의 목표에 중점을 둔이 연구의 목표에 중점을 둔 데이터 세트에는 141,160 개의 환자 기록이 포함되어 있습니다. 여기에는 수치, 범주 및 텍스트 인 세 가지 유형의 데이터가 포함됩니다. 이 작업에 대한 부분의 행동 및 생물 의학적 일상에는 인한 적 개입이 없었습니다. 또한이 작업에는 검플의 데이터만 분석하는 것이 포함됩니다. 따라서 기관 검토 위원회 (IRB)에서 허가를받지 못했습니다. 연구를 위해 모든 관련 지침이 따랐습니다.

많은 요인들이 IVF 치료 후 라이브 생성 (목표 변수)에 영향을 미칩니다. 이 연구에서 원래 데이터 세트에는 94 개의 기능이 포함되어 있지만 모든 기능이 결과에 큰 영향을 미치는 것은 아닙니다. 따라서 30 개의 기능만 고려됩니다. 따라서 기능 엔지니어링은 Bharti Bansal 박사와 Nice Clinical Guideline (National Institute of Health and Care Excellence, UK)이 권장하는 주제 지식에 따라 수행되었습니다. 이 연구에서는 신선한 사이클 및 IVF (ICSI 포함)를받는 여성에 대한 동일한 자극으로 인한 다음 동결 - 해동주기와 같은 특징이 고려됩니다. 기증자 난 모세포/정자 사이클 및 PGD/PCS 사이클은 제외됩니다. 이 연구에서 선택된 특징은 연령, 이전주기 총 수, 이전 IVF 임신의 총 수, 파트너 정자와 혼합 된 계란 수,이주기에서 전달 된 배아 수, 남성 요인, 여성을 포함하는 불임의 유형 및 원인입니다. 요인, 배란, 자궁 내막증, 튜브, 자궁 경부 등. 표 1은 접근 방식에서 고려 된 데이터 세트 기능에 대한 자세한 설명을 요약 한 것입니다.

데이터 세트의 사전 처리. 원시 데이터 세트에는 94 개의 속성이 포함되어 있으며, 그 중 라이브 제작 예측에 큰 영향을 미치지 않는 사람은 거의 없습니다. 데이터 세트의 여과는 사용된 자극, 정자 소스, 계란 소스 기능에 따라 다릅니다. 정자와 계란의 근원이 동일한 부부, 즉 파트너와 환자에서 나온 경우, 환자 기록이 고려되고 나머지는 제거됩니다. IVF에서, 여포 자극 호르몬 (FSH) 및 루테인니 화 호르몬 (LH)을 포함하는 주사 가능한 약물은 암컷에 주사하여 한 번에 하나 이상의 난자를 자극하여 [45]. 데이터 세트에서 "자극"으로 설명되고; 이 연구는 "자극이 확장되는" 환자 기록만 고려합니다. "치료시 환자 연령" 분야에서 제거 된 값 999를 포함하는 환자 기록은 거의 없습니다. 텍스트와 나이 범위는 범주 형 데이터로 변환됩니다. 예를 들어, "치료시 환자 연령"현장에서.

1. 18 – 34가 0으로 변환
되었습니다. 5로 변환됩
니다

"라이브 베르트 발생"필드는 목표 변수이며, 여기에는 숫자이며 0에서 5 범위의 값이 포함되어 있으며, 여기서 0은 출생 (음의 클래스)을 나타내지 않고 1보다 큰 1은 출생 발생 (양성 클래스)을 나타냅니다. 분류 바이너리를 만들기 위해 라이브 비트 발생이 1보다 더 중요한 모든 환자 기록은 1로 설정되고 0으로 남아 있습니다.

상기 여과 후, 음성 샘플은 데이터 불균형을 만드는 양성 샘플보다 5 × 더 많다. 데이터 세트에서 불균형 문제를 겪기 위해 음성 샘플의 환자 기록은 거의 제거되지 않습니다. 이제 데이터 세트에는 141,160 개의 환자 기록과 25 개의 기능이 포함되어 있으며 각 클래스마다 70,580 개의 샘플을 배포합니다. 정자 공급원, 계란 소스, 불임 파트너의 원인 정자 면역 인자, 사용 된 자극, 여성 불임 인자의 원인은 긍정적 인 경우와 부정적인 경우 모두에서 균질 한 경우 (동일한 값을 함유 함)와 같은 분야가 거의 없다는 것이 밝혀졌습니다. 분류 이므로이 필드는 제거됩니다. 그런 다음 샘플 또는 환자 기록은 검증에서 34%, 교육 세트에서 66%로 분할됩니다.

- 훈련 세트에는 93,165 개의 샘플이 포함되어 있습니다
- 검증 세트에는 47,995 개의 샘플이 포함되어 있습니다

데이터는 광범위한 정수 범위에 분포 된 값을 포함하므로 정규화됩니다. 매우 유사한 트렌드가있는 기능도 매우 유사한 정보를 가지고있을 것입니다. 이 경우 그중 하나만 머신 러닝 모델을 공급하기에 충분합니다. 구성된 상관 행렬은 그림 1에는 개발 된 모델에서 각 매개 변수의 중요성을 나타내는 25 가지 기능을 포함합니다. 여기서는 상관 관계를 계산합니다

Field	Type	Description
Patient age at treatment	Categorical	Patient age at treatment, banded as follows: 18–34, 35–37, 38–39, 40–42, 43–44, 45–50
Total number of previous cycles	Numerical	How many treatment cycles of IVF the patient has previously had
Total number of IVF pregnancies	Numerical	How many patients have been pregnant through IVF
Total number of live births- conceived through IVF	Numerical	How many live births the patients have had through IVF
Type of infertility—female primary	Categorical	1 if the patient unable to get pregnant after at least 1 year, 0 otherwise
Type of Infertility—female secondary	Categorical	1 if the patient able to get pregnant at least once but now unable to, 0 otherwise
Type of infertility—male primary	Categorical	1 if the leading cause of the infertility is patient, 0 otherwise
Type of infertility—male secondary	Categorical	1 if the secondary cause of infertility is due to the patient, 0 otherwise
Type of infertility—couple primary	Categorical	1 if the leading cause of the infertility is patient/partner, 0 otherwise
Type of infertility—couple secondary	Categorical	1 if the secondary cause of infertility is due to the patient/partner, 0 otherwise
Cause of infertility—tubal disease	Categorical	1 if there is damage in the fallopian tubes that prevents sperm from reaching the ovary, 0 otherwise
Cause of infertility—ovulatory disorder	Categorical	1 if the primary cause of this infertility is due to ovulation disorder, 0 otherwise
Cause of infertility—male factor	Categorical	1 if the primary cause of this infertility is due to male patients, 0 otherwise
Cause of infertility—patient unexplained	Categorical	1 if the primary cause of infertility in the patient is unknown, 0 otherwise
Cause of infertility—endometriosis	Categorical	1 if the primary cause of this infertility is due to endometriosis, 0 otherwise
Cause of infertility—cervical factors	Categorical	1 if the primary cause of this infertility is due to the Cervical factor, 0 otherwise
Cause of infertility—female factors	Categorical	1 if the primary cause of this infertility is due to female factors, 0 otherwise
Cause of infertility—partner sperm concentration	Categorical	1 if the primary cause of this infertility is due to low sperm count, 0 otherwise
Cause of infertility—partner sperm morphology	Categorical	1 if the primary cause of this infertility is an abnormality in sperm morphology, 0 otherwise
Cause of infertility—partner sperm motility	Categorical	1 if the primary cause of this infertility is poor sperm motility, 0 otherwise
Cause of infertility—partner sperm immunological factors	Categorical	1 if the primary cause of this infertility is due to sperm immunological factors, 0 otherwise
Stimulation used	Categorical	1 if the stimulation medication is used, 0 otherwise
Egg source	Text	Indicates whether the eggs used in this cycle came from Patient (P) or a Donor (D)
Sperm source	Text	Indicates whether the eggs used in this cycle came from Patient (P) or a Donor (D)
Fresh cycle	Categorical	1 if this cycle using fresh embryos, 0 otherwise
Frozen cycle	Categorical	1 if the cycle used from frozen embryos, 0 otherwise
Eggs thawed	Numerical	If this cycle frozen eggs, the number of eggs thawed
Fresh eggs collected	Numerical	The number of eggs collected in this cycle
Eggs mixed with partner sperm	Numerical	The number of eggs mixed with sperm from the partner
Embryos transferred	Numerical	The number of embryos transferred into the patient in this cycle

표 1. 데이터 세트의 IVF 속성.

계수와 피어슨의 카이 제곱 [46] 값으로서 수치 및 공칭 열 사이의 계수. 임계 값보다 높은 상관 계수를 갖는 열 쌍은 하나로 줄어 듭니다.

모델 교육. 개요. 이 연구에서는 기계 학습, 딥 러닝 및 앙상블 모델이 목적을 위해 교육을 받았습니다. 로지스틱 회귀 (LR), K 가장 가까운 이웃 (KNN), 다층 퍼셉트론 (MLP), 의사 결정 트리와 같은 기계 학습 ML 모델은 교육에 사용됩니다. 딥 러닝 모델, 특히 출력 층에 Sigmoid 활성화 뉴런이있는 1D 신경망이 제안됩니다. 앙상블 학습은 또한 기계 학습 모델 목록에서 구체적인 결정을 내리기 위해 사용됩니다. 임의의 숲, Adaboost, 투표 분류기 하드/소프트는 앙상블 학습 기술에서 고려됩니다.

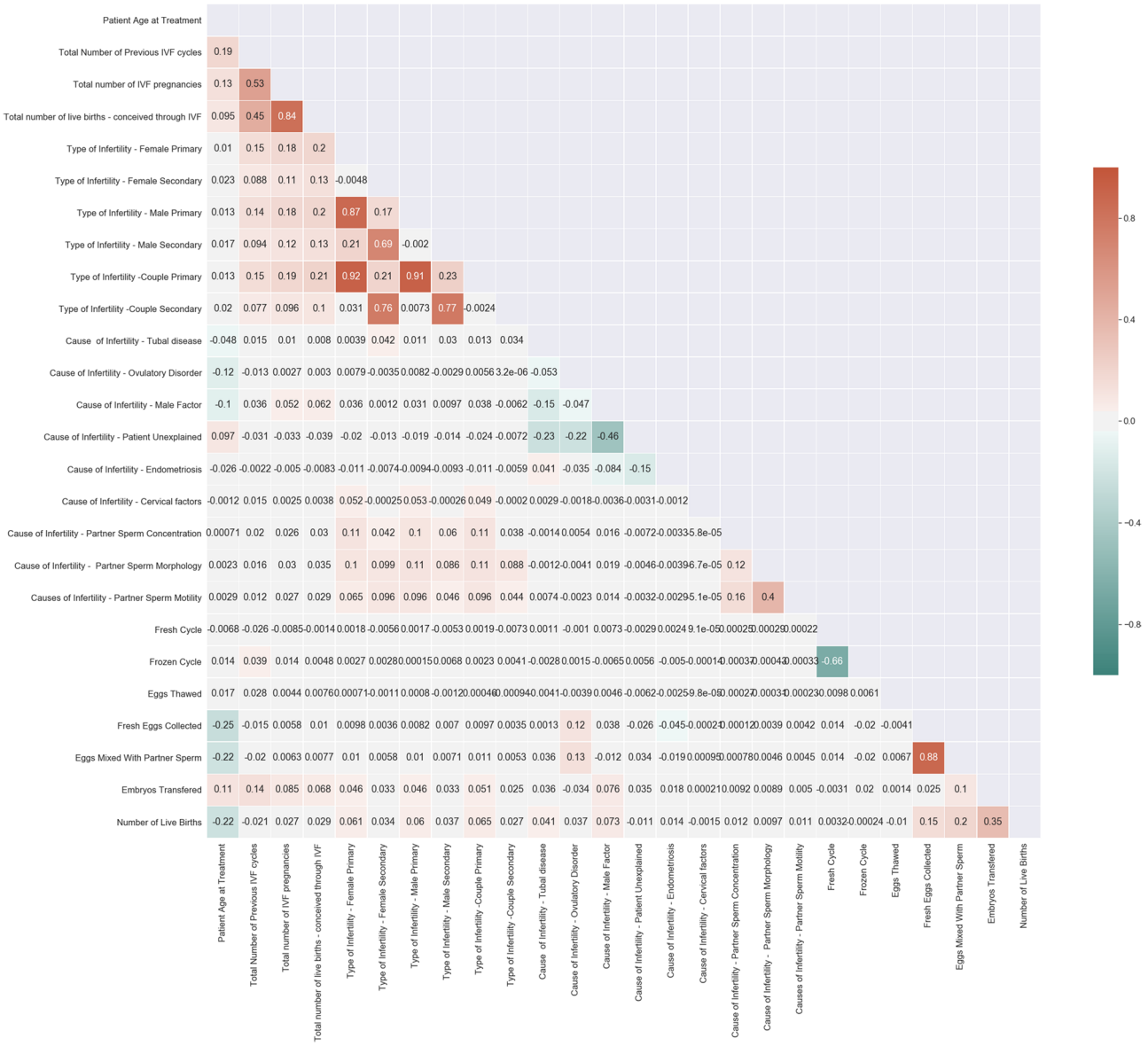


그림 1. 25 가지 기능의 상관 행렬.

두 가지 설정. 이 연구의 교육 모델은 기능 선택이없는 두 가지 설정으로 진행됩니다. 기능 선택 기술은 데이터 세트에서 필수 기능을 얻기 위해 적용됩니다. 기능 선택 설정이없는 상태에서 모든 기능 (총 25 개)이 교육 프로세스에 사용됩니다. 기능 선택 설정에서는 기술을 기반으로 교육 프로세스에 특정 기능 (중요한 기능) 만 사용됩니다. 이 접근법은 기능 선택 유무에 관계없이 모델이 데이터에 대해 교육을받은 결과에 대한 포괄적 인 분석을 제공합니다. 이 연구에 사용 된 훈련 프로토콜의 흐름도는 그림 2에 자세히 설명되어 있습니다. 기능 선택과 기능 선택 설정이없는 모델은 그림 2에 설명 된 것과 동일하게 유지됩니다.

도메인 전문가 Bharti Bansal 박사의 제안을받은 후 우리는 필수 기능을 선택했지만 통계 기능 중요성은 나중에 사용됩니다. 따라서 다양한 기능 선택 알고리즘을 탐색하면 전반적인 성능 요소를 개선하는 데 많은 도움이 될 수 있습니다. 선택된 기능 선택 기술이 있습니다

1. 선형지지 벡터 분류기 (선형 SVC) + 모델 2에서 선택. 트리 기반 기능 선택

선형 svc + 모델에서 선택하십시오. L1 Norm으로 처벌 된 선형 모델에는 드문 솔루션이 있습니다. 추정 계수 중 다수는 0입니다. 목표가 데이터의 차원을 줄이고 다른 분류기를 사용하는 것이라면 Scikit-Learn의 feaction_selection.selectfripmodel과 함께 사용하여 0이 아닌 계수를 선택할 수 있습니다. 이 목적에 유용한 희소 추정기는 회귀, 로지스틱 회귀 및 선형 SVC [43]에 대한 레이소입니다. 이 방법에 사용 된 드문 추정기는 로지스틱 회귀, 의사 결정 트리, 임의 포레스트, K 가장 가까운 이웃 분류기입니다. 기능 공간은이 기술을 사용하여 25 ~ 20로 감소합니다.

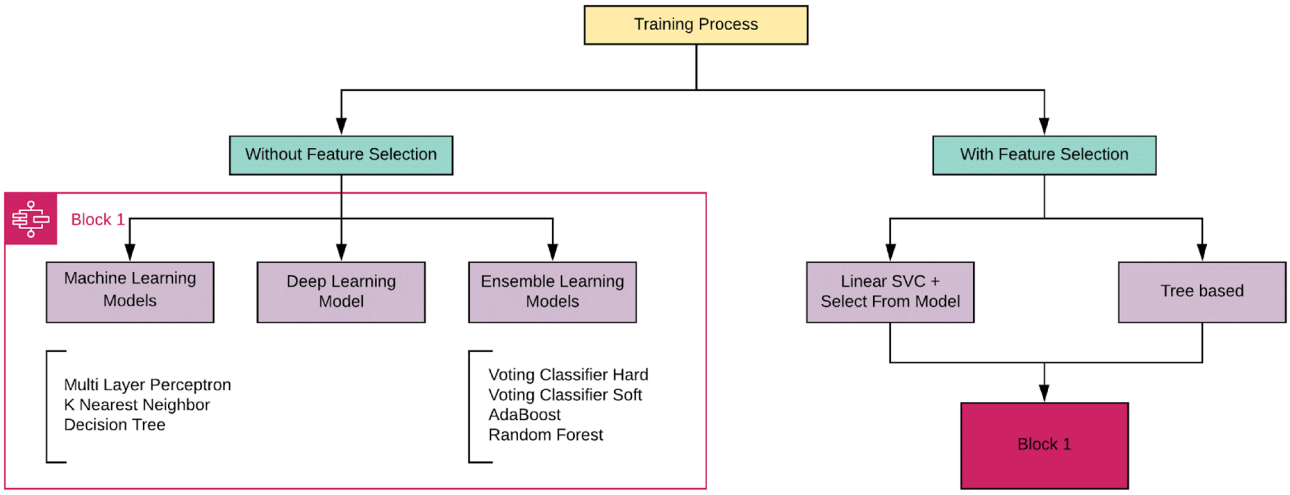


그림 2. 훈련 과정의 흐름도.

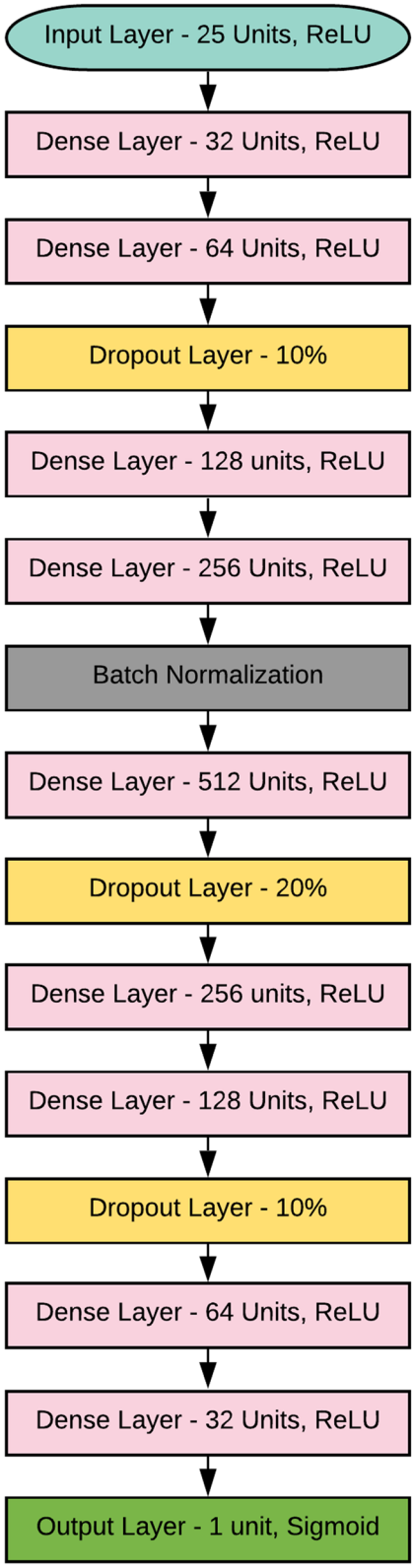
	Model	Precision (%)	Recall (%)	F1-score (%)	ROC AUC score (%)
Machine learning models	Multi-layer Perceptron	74	72	72.98	77.90
	K Nearest Neighbours	71	71	71.00	77.60
	Decision Tree	76	76	76.00	83.30
Deep learning model	DL Classifier	73	72	72.49	78.00
Ensemble Learning models	Voting—hard classifier	75	73	73.98	73.10
	Voting—soft classifier	77	75	75.98	83.20
	Random forest	77	76	76.49	84.60
	AdaBoost	74	72	72.98	77.40

기능 선택 모델이없는 분류 지표 간의 비교.

선형 SVC + 트리 기반 기능 선택. 임의의 숲과 같은 트리 기반 추정기는 일단 훈련 된 후에 각 기능의 중요성을 계산하여 피처 공간을 필터링하고 줄일 수 있습니다. 임의의 산림의 모든 기능은이 측정을 사용하여 GINI 불순물 또는 정보 이득/엔트로피로 훈련을받는 반면, 우리는 기능 중요성 [44]를 계산합니다. 기능 공간을 줄인 후이 새로운 세트에서 다른 추정기 또는 분류기를 훈련시킬 수 있습니다. 이 연구의 희소 추정치는 로지스틱 회귀, 의사 결정 트리, 선형 분사 분석, 임의의 산림, K 가장 가까운 이웃입니다. 기능 공간은이 기술을 사용하여 25 ~ 5로 감소합니다.

딥 러닝 : 맞춤형 심해 네트워크. ML 모델과 함께 딥 러닝 분류기 (DL) 보관소가 동일한 데이터에 대해 교육을 받았습니다. 신경망은 입력으로 수치 값 (크기 25 배열)을 취합니다. 따라서 건축 적 관점 (1-D 모델)에서 1 차원입니다. 출력 층은 이진 출력 (출생 발생 여부)을 제공하기 위해 Sigmoid 활성화 기능을 갖는 하나의 뉴런을 함유한다. 아키텍처에는 총 9 개의 조밀 한 층이 포함되어 있으며, 각 뉴런 (모든 조밀 한 층) 출력 값은 정류 된 선형 유닛 (Relu) [47] 활성화 함수를 통해 전달됩니다. DL 분류기의 전반부에서, 각 층의 뉴런은 이전 층의 정확하게 2 배 증가합니다. 이것은이 데이터 세트의 성능으로 인해 균일하게 유지됩니다. 후반은 층당 두 개의 뉴런의 감소 속도를 따릅니다. Adam Optimizer [48]는 심층 신경망을 훈련하는 동안 손실 값을 최적화하는 데 사용됩니다. 딥 러닝 애플리케이션에 대한 광범위한 채택과 Adagrad 및 RMSProp 알고리즘의 최상의 속성을 결합하여 시끄러운 문제에 대한 희소 기울기를 처리 할 수있는 최적화 알고리즘을 제공하기 때문에 Adam Optimizer가 선택됩니다. Adam Optimizer의 성능에 대한 이론 직관뿐만 아니라이 데이터 세트의 성능은 확률 적 구배 하강, RMSProp, Adagrad와 같은 다른 최적화 알고리즘에서 확인되며 Adam Optimizer 성능이 다른 것보다 낫다는 것을 알 수 있습니다.

DL 분류기를 훈련시키기위한 총 에포크 수는 50입니다.이 모델은이 데이터 세트를 과잉으로하여 드롭 아웃 및 배치 정규화 [49]와 같은 과적으로 정규화 된 기술을 방지 할 수 있습니다. [49]. 뉴런의 수가 증가함에 따라, 노이즈 생성 확률은 DL 분류기의 밀도가 높은 중간 층 중에서 더 높아질 것이므로 중간 밀도 층 (512 단위) 후에 드롭 아웃의 20%가 도입됩니다. 이진 크로스-엔트로피 손실은 딥 러닝 기술에 대한 훈련시 이진 분류에 가장 적합합니다. 전체 데이터 세트에 대한 기울기를 계산하는 데 비용이 많이 들며, 따라서 128 개의 샘플의 배치 크기가 에포크 당 훈련을 받았습니다. 맞춤형 딥 러닝 아키텍처를 엿볼 수 있습니다.



Training Parameter	Value
Input Size	(, 25)
Output Size	(,1)
Total Dense Layers	9
Regularization	Batch Normalization, Dropout
Batch_Size	128
Optimizer	Adam
Loss	Binary crossentropy
Epochs	50
Callbacks used	Early Stopping (patience = 5)

그림 3. 교육 매개 변수와 함께 딥 러닝 아키텍처가 설명되었습니다.

앙상블 학습. 앙상블 방법은 우수한 성능으로 인해 ML 도메인에서 가장 유명한 학습 알고리즘입니다. 이러한 방법의 전체 요점은 많은 ML 알고리즘을 결합하여 정확한 결정을 내린다는 것입니다. 이 연구에서는 다음과 같은 알고리즘이 훈련됩니다.

1. Random Forest 2. Adaboost
3. 투표 분류기 - 소프트/하드

투표 분류기. 개별 모델 예측에서 투표하여 최종 결정을 내리는 많은 분류 모델의 포장지입니다. 예를 들어, 5 개의 이진 분류 모델이 보이지 않는 샘플로 훈련되고 쿼리되면 개별 모델의 예측은 투표 시스템에 대한 입력으로 최종 예측을 선언합니다. 보충 그림 1. 투표 분류기를 보여줍니다.

투표 시스템은 두 가지 다른 전략, 즉 하드 투표, 소프트 투표에서 작동합니다. 하드 투표는 또한 개별 모델 세트에서 가장 많은 투표권을 얻는 클래스가 [42]를 선택하는 대다수 투표라고도합니다. NC가 클래스의 투표 수이고 Y1, Y2, y3,⋯

$$y_{final} = Max\left(N_c(y_1), N_c(y_2), N_c(y_3), \ldots, N_c(y_n)\right) \quad (1)$$

소프트 투표는 확률이 개별 클래스의 벡터로 벡터로 입력을 취하고 나중에 다른 모든 분류기와 합산합니다. 나중에 평균 [42]. 최종 출력 클래스는 최대 확률 점수를 얻는 클래스입니다. $p_1, p_2, \cdots, p_n$은 n 서로 다른 분류기의 확률 점수 인 경우, 소프트 투표를위한 공식은 식 (v9} 2)에 의해 제공됩니다.

$$y_{final} = Max\left(\frac{1}{n}\left(\sum\left(p_1, p_2, \ldots, p_n\right)\right)\right) \quad (2)$$

이 투표 분류기에 사용되는 분류기는 로지스틱 회귀, 의사 결정 트리, 선형 판별 분석, 임의 포리스트 및 K 가장 가까운 이웃입니다. 다음 섹션에서는 모델이 실험 데이터 세트로 검증되었습니다.

결과와 토론

이 연구에서는 딥 러닝 분류기를 훈련시키기위한 Keras 백엔드가 장착 된 텐서 플로우 라이브러리 및 ML 분류기의 Scikit-Learn이 사용됩니다. 이 연구에서 비교 된 메트릭은 F1- 점수, 정밀, 리콜, ROC AUC 점수 및 다양한 모델 간의 곡선입니다. 이 섹션에서는 기능 선택이 있거나없는 훈련 된 모델의 결과를 보여줍니다. 비교하여 ML 기반, DL 기반, 앙상블 기반과 같은 더 넓은 범주가 표시됩니다. 표 2 세부 사항 기능 선택없이 훈련 된 모델 간의 비교.

기능 선택 설정이없는 결과. 표 2는 앙상블 학습 모델 카테고리가 리콜, F1- 스코어 및 ROC AUC 점수에서 더 나은 분류 성능을 초래한다고 설명합니다. 임의의 Forest 점수는 76.49%의 가장 높은 F1 점수를 기록합니다. 임의의 산림에 의해 달성 된 리콜 값은 다른 훈련 된 모델, 즉 76%에서 눈에 띄게 나타납니다. 그림 4a는 기능 선택없이 훈련 된 모델의 ROC AUC 곡선을 보여줍니다. 랜덤 포레스트의 AUC 점수는 84.6%입니다.

기능 선택 설정에 대한 결과. 방법 : 선형 svc + selectfromomel. 표 3은 앙상블 학습 모델 범주가 더 나은 분류를 가지고 있음을 설명합니다. 다층 퍼셉트론과 아다 부스트는 72.98%의 F1 점수가 가장 높습니다. Adaboost, 즉 77.60%는 최대 ROC AUC 점수를 달성합니다.

그림 4b는 기능 선택 설정, 즉 선형 SVC + 모델에서 모델에서 훈련 된 모델의 ROC AUC 곡선을 보여줍니다. 이 방법에서 Adaboost의 AUC 점수는 77.60%입니다. 이러한 결과가 표 2의 이전 결과와 비교되면 기능 선택 방법의 영향이있어 ROC AUC 점수, F1- 스코어 및 리콜과 같은 메트릭의 전체 성능이 감소했습니다.

방법 : 선형 SVC + 트리 기반 기능 선택. 이 방법은 트리 기반 기능 추출기를 추가 트리 분류기로 사용했으며, 이는 임의의 숲과 약간 다릅니다. 여분의 트리 분류기는 부모 노드를 두 개의 임의의 자식 노드로 나누기 위해 임의의 분할을 선택하기 때문에 다릅니다.

표 4는 기계 학습 기반 분류가 더 좋습니다. 73.46%는이 기능 선택 방법에 의해 달성 된 가장 높은 F1 점수이며, 이는 이전 방법보다 적습니다. 여기서 달성 된 최대 리콜 값은 72%이며 이전 방법과 동일합니다. 그러나 ROC AUC 점수는 딥 러닝 분류기 및 Adaboost를 제외한 이전 방법에서 증가했습니다. 그림 4c. 기능 선택 설정, 즉 선형 SVC + 추가 트리 분류기로 훈련 된 모델의 ROC AUC 곡선을 묘사합니다. Adaboost, MLP, DL 분류기는 AUC 점수가 가장 높습니다.

위의 세 가지 방법의 결과를 비교할 때, 기능 선택 방법, 특히 임의의 산림 (앙상블 학습 방법)을 사용하지 않고 정기적 인 기능은 76.49%의 정확도와 84.6%의 정확도가 향상되는 것이 명확하고 권장됩니다. 따라서 실시간 결과를 위해이 모델을 생산에 사용하는 것이 바람직합니다.

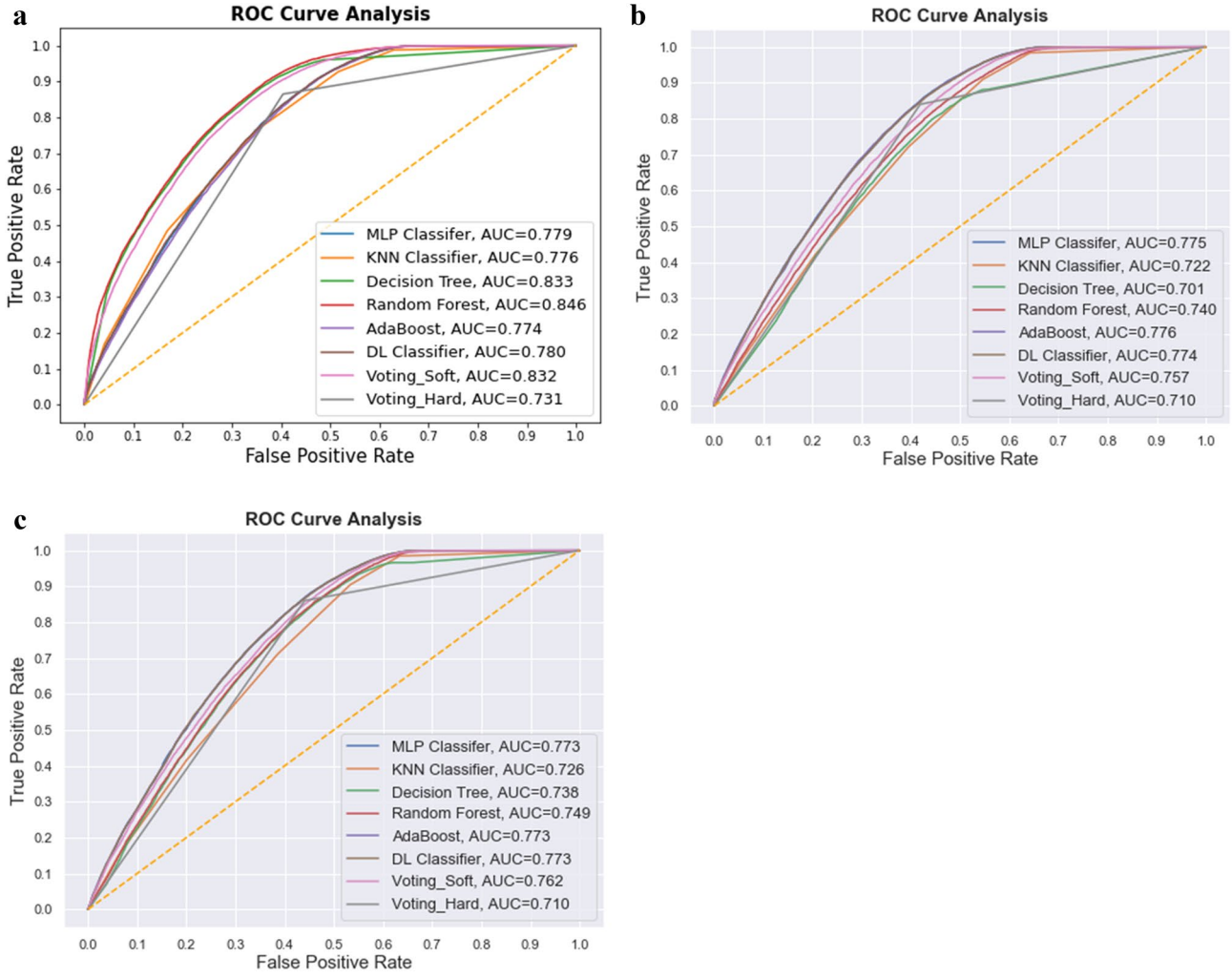


그림 4. (a) 기능 선택없이 훈련 된 모델의 ROC 곡선 분석. (b) 기능 선택 설정, 즉 선형 SVC +에서 다른 모델의 ROC 곡선 분석. (c) 기능 선택 방법, 즉 선형 SVC + 엑스트라 트리 분류기로 훈련 된 모델의 ROC 곡선 분석.

	Model	Precision (%)	Recall (%)	F1-score (%)	ROC AUC score (%)
Machine Learning models	Multi-layer Perceptron	74	72	72.98	77.50
	K Nearest Neighbours	67	66	66.49	72.20
	Decision Tree	67	67	67.00	70.10
Deep learning model	DL Classifier	74	72	72.98	77.40
Ensemble Learning models	Voting—Hard classifier	73	71	71.98	71.70
	Voting—Soft classifier	71	70	70.49	75.70
	Random Forest	69	68	68.49	74.00
	AdaBoost	74	72	72.98	77.60

표 3. 기능 선택 설정과 함께 다른 모델의 분류 메트릭 비교, 즉 선형 SVC + 모델에서 선택.

결론

클리닉에서 의료 종사자는 경험이나 생식 센터의 성공률에 따라 생명에 대한 상담을 제공 할 수 있으며, 경우에 따라 부적절 할 수 있습니다. 이 연구는 환자와 의료 종사자 모두가 환자의 자연적인 측정 가능한 예측 변수를 기반으로 성공적이거나 실패한 IVF 치료를 예측하는 도구에 의존하는 구체적인 결정을 내리는 데 도움이 될 것입니다. 이 도구는 부부에게 비용이 많이 들고 번거로운 IVF를 겪기 전에 감정적으로 준비 할 가능성에 대해 커플에게 상담을 제공 할 것입니다.

	Model	Precision (%)	Recall (%)	F1-score (%)	ROC AUC score (%)
Machine Learning models	Multi-layer Perceptron	75	72	73.46	77.30
	K Nearest Neighbours	66	66	66.00	72.60
	Decision Tree	70	69	69.49	73.80
Deep learning model	DL Classifier	75	71	72.94	77.30
Ensemble learning models	Voting—hard classifier	73	71	71.98	71.00
	Voting—soft classifier	72	70	70.98	76.20
	Random Forest	71	70	70.49	74.90
	AdaBoost	74	71	72.46	77.30

표 4. 기능 선택 설정과의 다른 모델의 분류 메트릭 비교, 즉 선형 SVC + Extra Trees 분류기.

치료. 기능 선택이없는 랜덤 포레스트 모델은 다른 모델과 비교하여 AUC 점수 84.60% 및 76.49% F1-점수를 달성 한 최상의 결과를 보여주었습니다. 그러나 데이터가 단일 소스에서 수신되므로 현재 의사 결정을 위해이 도구에만 의존하는 것은 아니므로 모든 인구에게 일반화되지는 않습니다. 이 모델은 제한된 요인에 대해 교육을 받았으며, 알코올, 흡연, 카페인 소비, 고혈압 및 임신 예측에 중대한 영향을 미치는 기타 라이프 스타일 요인과 같은 몇 가지 중요한 요소는 데이터 세트의 한계로 인해 고려되지 않았습니다.

미래의 작업의 범위는 데이터를 전 세계의 많은 종족에 대한 정보를 포함하여 다양한 지리적 위치의 다양한 IVF 클리닉에서 수집 할 수 있다는 것입니다. 이러한 세부 사항이 간접적으로 다산에 영향을 미치기 때문에 개인의 생활 방식에 대한 매개 변수는 거의 고려해야합니다. 다양한 인종 및 연령 그룹의 다양한 데이터가 수집되면 AI 성능을 향상시킬 수 있습니다. IVF에서 각 기능의 중요성을 강조하는 성공적인 다산에 관한 연구도 수행 할 수 있습니다. 다양한 기능 선택 및 차원 감소 방법을 사용하여 모델 성능을 향상시킬 수 있습니다.

받은 : 2020 년 5 월 1 일; 허용 : 2020 년 10 월 19 일
Published online: 01 December 2020

참조

1. Gurunath, S., Pandian, Z., Anderson, A. R. & Bhattacharya, S. 불임 정의 - 유병률 연구에 대한 체계적인 검토. 험. 재판. 업데이트. 17, 575 – 588. https : //doi.org/10.1093/humup d/dmr01 5 (2011). 2. Loendersloot, V. L., Repping, S., Bossuyt, P.M. M., Veen, F. V. D. & Wely, M. V. 시험 관내 수정의 예측 모델; 우리는 어디에 있습니까? 미니 리뷰. J. adv. 해안 5, 295 – 301. https : //doi.org/10.1016/j.jare.2013.05.002 (2014). 3. Zarinara, A. et al. 불임 치료의 성공을 예측하는 모델 : 체계적인 검토. J. 재판. 불임. 17, 68 – 81 (2016). 4. Cooper, S. G. 불임 치료 비용 분석. 오전. J. 공중 보건. 76, 1018 – 1019. https : //doi.org/10.2105/ ajph.76.8.1018 (1986). 5. Gameiro, S., Boivin, J., Peronace, L. & Verhaak, C. M. 왜 환자가 다산 치료를 중단합니까? 생식 치료의 중단에 대한 이유와 예측 자에 대한 체계적인 검토. 험. 재판. 업데이트. 18, 652 – 669. https : //doi.org/10.1093/humup d/dms031 (2012). 6. Ranjbar, F., Moghadam, Z. B., Borimnejad, L., Saeed, R. G. & Akhondi, M. M.이란에서 임신 지원을 추구하는 불임 여성의 경험 : 질적 연구. J. 재판. 불임. 16, 221 – 228 (2015). 7. Smola, A. & Vishwanathan, S. V. N. 기계 학습 소개 3 – 9 (Cambridge University Press, Cambridge, 2008). 8. Goodfellow, I. et al. 딥 러닝 151 – 161 (The Mit Press, Cambridge, 2016). 9. Deo, R. C. 의학에서 기계 학습. 순환 132, 1920 – 1930 (2015). 10. Handelman, G. S. et al. 기계 학습과 의학의 미래. J. 인턴 메드. 284, 603 – 619. https : //doi.org/10.1111/ joim.12822 (2018). 11. Rahimian, F. et al. 기계 학습을 통한 응급 입학 위험 예측 : 연결된 전자 건강 기록을 사용한 개발 및 검증. PLOS MED. 11, E1002695. https : //doi.org/10.1371/journ al.pmed.10026 95 (2018). 12. Davenport, T. & Kalakota, R. 건강 관리의 인공 지능의 잠재력. 퓨트. 건강. J. 6 (2), 94 – 98. https : // doi. org/10.7861/futur ehosp.6-2-94 (2019). 13. Blank, C. et al. 시험 관내 수정에서 배반포 전이 후 이식의 예측 : 기계 학습 관점. 비료. 살균. 111, 318 – 326. https : //doi.org/10.1016/j.fertn Stert .2018.10.030 (2019). 14. Barash, O., Ivani, K., Weckstein, L. & Hinckley, M. 단일 배아 전달을 갖는 IVF PGT주기에서 배아 선택을위한 고 정확한 기계 학습 예측 모델. 비료. 살균. 110, E372. https : //doi.org/10.1016/j.fertn Stert .2018.07.1038 (2018). 15. Vaughan, D. et al. 프로그래밍 된 해동 사이클에서 라이브 비트의 예측 변수를 평가하기위한 기계 학습 방법의 적용. 비료. 살균. 112, E273. https : //doi.org/10.1016/j.fertn Stert .2019.07.808 (2019). 16. John, R. IVF에서 기계 학습의 적용 : 쌍둥이 예측을위한 분류 알고리즘의 정확성 비교. Gynecol. 오스켓. 9, 497. https : //doi.org/10.4172/2161-0932.10004 97 (2019). 17. Tran, D., Cooke, S., Illingworth, P. J. & Gardner, D. K. 시간 경과 배양 및 배반포 전달 후 태아 심장 임신을위한 예측 도구로서의 딥 러닝. 험. 재판. 34, 1011 – 1018. https : //doi.org/10.1093/humre p/dez06 4 (2019). 18. Khosravi, P. et al. 딥 러닝은 시험 관내 수정 후 인간 배반포의 강력한 평가 및 선택을 가능하게합니다. NPJ Digit Med. 2, 21. https : //doi.org/10.1038/S4174 6-019-0096-Y (2019). 19. McLernon, D.J., Steyerberg, E., Velde, E. R. T., Lee, A. J. & Bhattacharya, S. 에서 하나 이상의 완전한주기 후 살아있는 바로의 가능성을 예측하는 : 인구 기반 연구의 인구 기반 연구 113873 명의 여성의 사이클 데이터. BMJ 355, i5735. https : // doi. org/10.1136/bmj.i5735 (2016). 20. McLernon, D.J., Steyerberg, E., Velde, E. R. T., Lee, A. J. & Bhattacharya, S. 하나 이상의 완전한 사이클 후 살아있는 기회를 예측할 때 범죄 능력을 평가하는 데 사용되는 방법의 개선 에서의 시험 관내 수정. BMJ 363, K3598. https : //doi.org/10.1136/bmj.k3598 (2018).

21. Hassan, R., Al-Insaif, S., Hossain, M. I. & Kamruzzaman, J. IVF 치료 후 임신 결과 예측을위한 기계 학습 접근법. Springer 3 2, 2283 – 2297. https : //doi.org/10.1007/s0052 1-018-3693-9 (2020). 22. Guvenir, A. H. et al. 순위 알고리즘을 사용하여 IVF 처리의 성공 가능성을 추정합니다. Springer 53, 911 – 920. https : // doi.org/10.1007/s1151 7-015-1299-2 (2015). 23. Kaufmann, S. J. et al. 투시 트로 수정의 결과를 예측할 때 신경망의 적용. 험. 재판. 12 (7), 1454 – 1457. https : //doi.org/10.1093/humre p/12.7. 1454 (1997). 24. 바비타. 성공률을 향상시키기 위해 IVF 처리에 사용 된 기계 학습 기술에 대한 설문 조사. 7 (2019). 25. Qiu, J. et al. 시험 관내 수정 치료에서 첫 번째 생명기의 개인화 된 예측 : 기계 학습 방법. J. 번역. 메드. 17, 317. https : //doi.org/10 .1186/S1296 7-019-2062-5 (2019). 26. Olmedo, S. B., Chillik, C. & Kopelman, S. 불임의 정의 및 원인. 재판. 생물체. 온라인 2, 4 1 – 53. https : // doi. org/10.1016/s1472 -6483 (10) 62187 -6 (2001). 27. Cedars, M. & Jaffe, R. B. 불임과 여성. J. Clin. 엔도 크리놀. 대사. 90, 4. https : //doi.org/10.1210/jcem.90.4.9997 (2005). 28. Masoumi, Z. S. et al. 하마단의 Fatemieh Hospital에있는 불임 센터를 언급 한 환자의 불임의 원인에 대한 역학적 조사. 이란 J. 재판. 메드. 13, 513 – 516 (2015). 29. Templeton, A., Morris, K. J. & Parslow, W. Vitro 수정 결과에 영향을 미치는 요인. Lancet 348, 1402 – 1406. https : // doi. org/10.1016/s0140 -6736 (96) 05291 -9 (1996). 30. Bhattarcharya, S. et al. 여성 불임. 클린. 증거. 11, 819 (2010). 31. Lackner, J. et al. 도시 인구의 불임 남성의 정자 농도의 지속적인 감소 : 18 년 이상의 경험. 비료. 살균. 84, 1657 – 1661. https : //doi.org/10.1016/j.fertn Stert .2005.05.0 49 (2005). 32. Cooper, T. G. et al. 인간 정액 특성에 대한 세계 보건기구 참조 가치. 험. 재판. 업데이트. 16, 231 – 245. https : // doi.org/10.1093/humup d/dmp04 8 (2010). 33. 인간 수정 및 배아 당국. https : //www.hfea.gov.uk/media /2667/ar-2015-2016-xlsb. xlsb (2016). 34. Dreiseitl, S. & Machado, L. O. 로지스틱 회귀 및 인공 신경망 분류 모델 : 방법론 검토. J. 바이오 메드. 알리다. 35, 5-6. https : //doi.org/10.1016/s1532 -0464 (03) 00034 -0 (2002). 35. T. Cover & P. Hart. 가장 가까운 이웃 패턴 분류. IEEE 트랜스. 알리다. 이론, 13, 21 – 27. doi : https : //doi.org/10.1109/ tit.1967.10539 64 (1967) 36. Almeida, L. B. 다층 퍼셉트론. 신경 계산 핸드북 (Oxford University Press, Oxford, 1997). 37. Breiman, L., Friedman, J., Stone, C. J. & Olshen, R. A. 분류 및 회귀 나무 (CRC Press, Boca Raton, 1984). 38. Anderson, J. A. et al. 말하는 그물 : 신경망의 구전 역사 (The Mit Press, Cambridge, 2000) . 39. Dietterich, T. G. 기계 학습의 앙상블 방법. 다중 분류기 시스템에 관한 국제 워크숍 (Springer, Berlin, 2000). 40. Breiman, L. 랜덤 숲. 스프링거 45, 5 – 32. https : //doi.org/10.1023/a : 10109 33404 324 (2001). 41. Freund, Y. & Schapire, R. E. 새로운 부스트 알고리즘을 통한 실험. 국제 기계 학습에 관한 국제 회의 (ICML'96) 148 – 156 (1996) 42. Raschka, S. Python Machine Learning (Packt Publishing, Birmingham, 2015). 43. Moon, M. & Nakai, K. 바이오 마커 발견을위한 앙상블 L 1-Norm지지 벡터 기계를 기반으로 한 안정적인 기능 선택. BMC 게놈. 17, 1026. https : //doi.org/10.1186/S1286 4-016-3320-Z (2016). 44. Uddin, M . T. & Uddiny, M. A. 매우 무작위 화 된 나무를 사용한 웨어러블 센서의 인간 활동 인식. 전기 공학 및 정보 통신 기술에 관한 전국 회의 (ICEICT), 다카, 1-6. https : // doi. org/10.1109/iceei ct.2015.73073 84 (2015) 45. Ahemmed, B. et al. 보조 생식 기술을 위한 통제 된 난소 자극에 대한 실습을 향상시키기위한 인도 전문가 패널의 결과 및 권장 사항. int. J. 재판. 메드. 2017, 94512 35. https : //doi.org/10.1155/2017/94512 35 (2017). 46. Rakesh, R. & Richa, S. Chi-Square 테스트 및 가설 테스트 응용. J. Pract. 심근. 공상 과학. 1, 69 – 71. https : // doi. org/10.4103/2395-5414.15757 7 (2015). 47. Vinod N., & Geoffrey, E. Hinton. 정류 된 선형 유닛은 제한된 볼츠 만 기계를 개선합니다. 기계 학습에 관한 국제 회의 (ICML)에 관한 27 차 전국 회의 절차. Omnipress, Madison, WI, USA, 807 – 814 (2017) 48. Kingma, DP, & BA, J. Adam : 확률 적 최적화를위한 방법. Corr, https : // arxiv .org/abs/ 1412.6980 (2015) 49. Ioffe, S., & Szegedy, C. 배치 정규화 : 내부 공변량 이동을 줄임으로써 깊은 네트워크 훈련을 가속화합니다. https : // arxiv .org/abs/1502.03167 (2015)

감사의 말

우리는 Bharti Bansal 박사 (MS- 산부인과, -Jawaharlal Nehru Medical College, Aligarh —2006)에게 감사의 말씀을 전합니다. 분류기.

저자 기여

A.G. : 데이터 수집, 계산 방법론, 모델 생성, 검증, 결과 생성, 초기 초안 작성. M.K : 연구 및 조사 프로세스, AI 컴퓨팅, 딥 러닝 모델 생성, 작성 및 초안 편집. K.P.R.A. : 개념화, 최종 원고 편집 및 쓰기 및 전반적인 감독.

경쟁 관심사

저자는 경쟁 관심사를 선언하지 않습니다.

추가 정보

이 백서에서는 https : //doi.org/10.1038/s4159 8-020-76928 -Z에서 보충 정보가 제공됩니다.

자료에 대한 서신 및 요청은 K.P.R.A.

재 인쇄 및 권한 정보는 www.nature.com/reprints에서 확인할 수 있습니다.