



저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:



저작자표시. 귀하는 원저작자를 표시하여야 합니다.



비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.



변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 [이용허락규약\(Legal Code\)](#)을 이해하기 쉽게 요약한 것입니다.

[Disclaimer](#)

의학박사 학위논문

한국 성인에서 심뇌혈관 질환 위험도
예측: Cox 비례위험모형과
프라이버시 보전 딥러닝

Prediction of Cardiovascular Risk in Korean Adults:

Cox proportional hazards model
and privacy-preserving deep learning

2022 년 4 월

서울대학교 대학원

의학과 의공학 전공

최 승 혁

한국 성인에서 심뇌혈관 질환 위험도
예측: Cox 비례위험모형과
프라이머시 보전 답러닝

지도 교수 윤 형 진

이 논문을 의학박사 학위논문으로 제출함

2022 년 4 월

서울대학교 대학원

의학과 의공학 전공

최 승 혁

최승혁의 의학박사 학위논문을 인준함

2022 년 7 월

위 원 장 김 희 찬 (인)

부위원장 윤 형 진 (인)

위 원 최 진 욱 (인)

위 원 이 해 영 (인)

위 원 김 중 현 (인)

국문 초록

한국 성인에서 심뇌혈관 질환 위험도

예측: Cox 비례위험모형과

프라이버시 보전 딥러닝

최승혁

의학과 의공학 전공

서울대학교 대학원

인공지능 (Artificial Intelligence; 이하 AI)이 여러 분야에서 광범위한 잠재력을 보여 주고 있지만 의료 분야에서 인공지능 모형을 개발하는 것은 개인정보보호에 대한 제약으로 인해 제한되고 있다. 프라이버시 보전 딥러닝 (Privacy Preserving Deep Learning; 이하 PPDL)은 개인정보를 보호하는 동시에 AI 모형을 훈련할 수 있으며, 그의 일종인 연합학습 (Federated Learning; 이하 FL)은 PPDL의 해답으로서 각광을 받고 있다. 하지만, FL은 자동화된 플랫폼을 요하는 모형 동기화에 기반을 두고 있어 폐쇄 네트워크를 지닌 의료 기관에서 이를 활용하기 어렵다. 반면, 전이반복학습 (Transfer Iterative Learning; 이하 TIL)은 전이 학습 (Transfer Learning)에 기반, 탈중앙화된 PPDL 훈련 알고리즘으로서 더 적은 통신

비용으로 FL에 비견할만한 모형의 성능을 달성할 수 있다.

따라서 본 연구의 주요 목적은 건강보험공단의 데이터를 이용하여 기존의 통계 방법론을 응용한 예측 모형을 예비 분석으로 하고, 단일 기관의 공통 데이터 모형 (Common Data Model; 이하 CDM)에 구현된 환자 데이터를 다기관 데이터로 시뮬레이션하여 FL과 TIL을 구현하고 성능을 비교하는 것이다.

임상적 의미를 지닌 예측 모형 작성을 위하여, 서구권 인구에 대해 개발되고 검증된 심뇌혈관질환에 대한 예측 모형이 많지만, 한국을 포함한 아시아 인구를 대상으로 개발된 모형은 진행이 더디다. 2009년 1월에서 2016년 12월까지 건강보험공단의 국가 건강검진을 시행 받은 한국인 중 심뇌혈관질환의 발병을 후향적으로 분석하였다. 이 데이터셋은 기저에 심뇌혈관 질환이 없는 40세에서 79세 사이의 21,581,796명 (남성: 10,412,947명, 여성: 11,168,849명)의 성인으로 구성되었다. 일차 결과인 심뇌혈관질환은 급성 관상동맥증후군, 뇌경색과 뇌출혈의 발생으로 정의하였으며 이는 건강보험공단의 명세서 데이터를 이용하였다. 예측 모형은 Cox 비례 위험 모형으로 제작되었으며 10회 교차 검증으로 검증하였다. 모형의 성능은 Harrell's C-index와 Brier score로 평가하였다. 모형의 구분 능력 (discrimination)은 수신자 작동특성곡선 분석의 곡선하 면적 (Area Under the Receiver Operating Characteristics curve; 이하 AUROC)으로 분석하였다. 모형의 AUROC는 남성에서 0.762, 여성에서 0.811를 보였으며, Brier score는 남성에서 0.018, 여성에서 0.010이었다. 본 연구의 모형은 한국인에 대해서 Framingham risk score (FRS)와 pooled

cohort equation (PCE)보다 좋은 성능을 보였다.

데이터를 공유할 수 없는 환경에서 심뇌혈관질환을 예측하기 위한 방법론 중 FL과 TIL의 성능을 실험하기 위해, 40,507명의 서울대학교병원의 CDM에 구현된 2형 당뇨 환자로 구성된 데이터를 사용하여 심뇌혈관질환의 예측 모형을 작성하였다. 다기관을 시뮬레이션하기 위해 5개의 상이한 분포를 지닌 가상의 기관으로 나누었고, TIL 모형은 모든 데이터를 통합해 제작된 모형과 비교하여 가장 낮은 경우 97.6%, 가장 좋은 경우 99.6%의 상대적 효율성 (Relative Efficiency; 이하 RE)을 보였다. FL은 99.4%의 RE를 보였으며, 협력 없이 자신의 데이터만으로 학습한 지역 학습 모형은 가장 낮은 경우 85.1%, 가장 좋은 경우 99.4%의 RE를 보였다. 이를 종합하면 FL 프레임워크를 사용할 수 없는 의료 기관에 대해 FL의 대안으로서 TIL이 통신에 효율적인 PDDL 훈련 알고리즘으로 제안될 수 있다.

주요어 : 전이반복학습, 프라이버시 보전 딥러닝, 예측모형, 심뇌혈관질환, 연합학습, 공통데이터모형

학 번 : 2016-30548

목 차

제 1 부 Cox 비례위험모형을 이용한 심뇌혈관 질환의 예측

제 1 장 서론.....	1
제 1 절 심뇌혈관 질환 연구의 배경.....	1
제 2 절 문제 제기	4
제 3 절 관련 연구 동향.....	7
제 4 절 연구의 필요성.....	9
제 5 절 연구의 내용, 목적 및 의의	10
제 6 절 성능 평가를 위한 지표 상세.....	12
제 2 장 심뇌혈관질환 예측 모형의 방법론.....	15
제 1 절 연구에 포함된 인구 집단의 정의.....	15
제 2 절 설명변수의 정의.....	18
제 3 절 결과변수의 정의.....	21
제 4 절 Cox 비례 위험 모형의 이론적 배경	22
제 5 절 통계 분석	25
제 3 장 심뇌혈관질환 예측 모형의 결과	26
제 1 절 심뇌혈관질환 예측 모형.....	29
제 2 절 심뇌혈관질환 예측 모형의 성능.....	44
제 4 장 심뇌혈관질환 예측 모형의 고찰	46

제 2 부 프라이버시 보전 딥러닝을 이용한 심뇌혈관 질환의 예측

제 1 장 서 론.....	53
제 1 절 프라이버시 보전 딥러닝 연구의 배경	53
제 2 절 문제 제기	55
제 3 절 관련 연구 동향.....	58
제 4 절 연구의 필요성.....	60
제 5 절 연구의 내용, 목적 및 의의	61
제 6 절 실험에 사용된 방법론과 성능 평가를 위한 지표.....	62
제 2 장 프라이버시 보전 딥러닝 모형의 방법론.....	65
제 1 절 데이터 확보.....	65
제 2 절 데이터 전처리.....	69
제 3 절 가상의 이질적 의료기관 생성	72
제 4 절 훈련 알고리즘과 평가	73
제 5 절 훈련 알고리즘의 상세	80
제 6 절 실험 설계	83
제 7 절 기관 설정	85
제 8 절 모형의 평가.....	86
제 3 장 프라이버시 보전 딥러닝 모형의 결과	88
제 1 절 프라이버시 보전 딥러닝의 기계학습 파이프라인.....	88
제 2 절 프라이버시 보전 딥러닝에서 훈련 알고리즘.....	90
제 3 절 가상 기관의 기저 특성.....	92
제 4 절 혼합 기관 환경 하에서 TIL 의 성능.....	107
제 5 절 혼합 동질 환경 하에서 TIL 의 성능.....	125
제 6 절 작은 기관 환경 하에서 TIL 의 성능.....	140
제 7 절 FL 과 TIL 의 통신 비용의 비교.....	152
제 4 장 프라이버시 보전 딥러닝 예측 모형의 고찰.....	157
결 론.....	161
참고문헌.....	165
Abstract	175

표 목 차

제 1 부 Cox 비례위험모형을 이용한 심뇌혈관 질환의 예측

표 2-1 WHO 기준 일 평균 음주량의 위험 분류	18
표 2-2 WHO 기준 체질량 지수의 구분	19
표 3-1 성별 연구 참여자의 기저 특성	26
표 3-2 심뇌혈관질환 예측 모형에 포함된 변수의 보정된 위험비	29
표 3-3 확장된 Cox 모형에 포함된 변수의 보정된 위험비	32
표 3-4 심뇌혈관질환 예측 모형의 chi-square 와 p값	34
표 3-5 확장된 Cox 모형의 chi-square 와 p값	36
표 3-6 Framingham risk score 와 pooled cohort equation (PCE), 개정된 PCE, 심뇌혈관 예측 모형의 AUROC, Harrell's C-index, Brier score 의 10회 교차 검증의 결과	44

제 2 부 프라이버시 보전 딥러닝을 이용한 심뇌혈관 질환의 예측

표 3-1 2형 당뇨 환자의 기저 특성	93
표 3-2 서울대학교병원 공통 데이터 모형에서 추출되어 프라이버시 보전 딥러닝 모형에 사용된 변수 목록	67
표 3-3 (전역 성능) 전이반복학습 모형의 여러 평가 척도 결과	109
표 3-4 (지역 성능) 전이반복학습 모형의 여러 평가 척도 결과	114
표 3-5 (전역 성능) 전이반복학습 모형의 여러 평가 척도 결과	127
표 3-6 (지역 성능) 전이반복학습 모형의 여러 평가 척도 결과	127
표 3-7 (전역 성능) 전이반복학습 모형의 여러 평가 척도 결과	127
표 3-8 (지역 성능) 전이반복학습 모형의 여러 평가 척도 결과	127
표 3-9 FL의 실험 당 평균 경과 라운드 수 및 평균 통신 비용	152
표 3-10 동기화 epoch 수에 따른 평균 경과 라운드 및 평균 AUROC 의 변화	154

그림 목 차

제 1 부 Cox 비례위험모형을 이용한 심뇌혈관 질환의 예측

그림 2-1 심뇌혈관질환 예측 모형을 위한 연구 참여자의 코호트 정의...16	16
그림 3-1 성별에 따른 (남성) Cox 모형에 대한 시간 대 계수의 Schoenfeld 잔차도.....40	40
그림 3-2 성별에 따른 (여성) Cox 모형에 대한 시간 대 계수의 Schoenfeld 잔차도.....43	43

제 2 부 프라이버시 보전 딥러닝을 이용한 심뇌혈관 질환의 예측

그림 1-1 기존의 AI와 분산 학습의 연구 과정.....55	55
그림 1-2 암호화된 추론 과정과 프레임워크의 개요.....56	56
그림 2-1 결과 변수의 정의.....66	66
그림 2-2 서울대학교병원의 공통 데이터 모형에서 추출한 2형 당뇨 환자의 코호트정의.....67	67
그림 2-3 프라이버시 보전 딥러닝 연구의 과정.....70	70
그림 2-4 Federated learning의 학습 과정.....73	73
그림 2-5 Transfer iterative learning의 학습 과정.....78	78
그림 2-6 전역 성능 점수와 지역 성능 점수.....79	79
그림 2-7 TIL의 내림차순 다중 라운드 훈련 과정.....81	81
그림 2-8 실험에 사용된 딥러닝 모형의 모식도.....83	83
그림 3-1 프라이버시 보전 딥러닝 환경에서 훈련 알고리즘.....90	90
그림 3-2 가상의 의료 기관의 데이터 분포.....93	93
그림 3-3 혼합된 분포의 가상 기관 환경에서 훈련 알고리즘의 전역 성능.....107	107
그림 3-4 혼합된 분포의 가상 기관 환경에서 훈련 알고리즘의 지역 성능.....112	112
그림 3-5 혼합 동질 분포 환경에서 훈련 알고리즘의 전역 성능.....125	125
그림 3-6 혼합 동질 분포 환경에서 훈련 알고리즘의 지역 성능.....129	129
그림 3-7 작은 기관 환경에서 훈련 알고리즘의 전역 성능.....140	140
그림 3-7 작은 기관 환경에서 훈련 알고리즘의 지역 성능.....144	144

축약어 목록

AI Artificial Intelligence

AUROC Area Under the Receiver Operating Characteristics curve

CDM Common Data Model

CVDs Cardiovascular Diseases

EHR Electronic Health Record

FL Federated Learning

FRS Framingham Risk Score

HDL high-density lipoprotein

LDL low-density lipoprotein

PCE Pooled Cohort Equation

PPDL Privacy Preserving Deep Learning

RE Relative Efficiency

TIL Transfer Iterative Learning

제 1 부 Cox 비례위험모형을 이용한 심뇌혈관질환의 예측

제 1 장 서 론

제 1 절 심뇌혈관 질환 연구의 배경

제 1 항 심뇌혈관 질환의 임상적 의미

심뇌혈관질환 (Cardiovascular diseases; 이하 CVD)은 세계적으로 주요한 사망 원인¹으로 2015년 100,000명당 285.5명이 사망하였다². 이는 한국에서도 주요한 문제로 2019년 질병관리청의 통계청 자료를 이용한 조사 결과 심혈관 질환은 100,000명당 60.4명이 사망했으며, 이는 주요 만성질환 중 암을 제외하고 사망률이 가장 높았으며, 뇌혈관 질환은 42.0명으로 폐렴 다음이었다³. 이처럼 CVD는 보건 분야에서 사회경제적으로 큰 부담으로 작용하고 있다. 고혈압 (Hypertension), 당뇨병 (Diabetes Mellitus), 이상지질혈증 (Dyslipidemia)은 CVD의 선행질환으로 개인 생활습관 교정과 더불어 의학적 관리를 통해 심각한 합병증인 심근경색 (Myocardial Infarction), 뇌졸중 (Stroke)으로 이어지는 것을 예방할

수 있다⁴. 이는 효과적인 생활 습관 교정 또는 적절한 약품 처방을 목표로 하기 위해 위험이 높은 사람들을 시기 적절하게 조기에 식별해야 한다는 것을 의미한다⁵. 이를 가능하게 하는 방법 중 하나는 위험 예측모형을 통해 고위험군을 식별함으로써 CVD를 예방할 수 있다⁶.

제 2 항 한국의 국민건강보험공단 데이터의 특성

모든 한국인은 단일 보험 체계에 의해 보장 받으며, 모든 명세서 데이터는 국민건강보험공단 (National Health Insurance Service - NHIS)에 의해 디지털화 되어 수집되고 저장된다. NHIS는 수진자의 데이터를 보관, 관리하며 합당한 연구자에게 제공하고 있다. 저장된 데이터는 NHIS 웹사이트 (<http://nhiss.nhis.or.kr>)를 통해 접속할 수 있다⁷. 연구 제안서 제출 시 연구자의 기관 심의위원회의 윤리 승인이 필요하며, 이는 데이터를 제공하기 전에 공단의 심의위원회에서 심의된다. 공공데이터의 제공 및 이용촉진에 관한 법률에 따른 정책 및 학술연구를 지원하기 위해 데이터가 제공되며 현재 국내 연구자만 사용 가능하다.

NHIS는 국가 건강검진 데이터를 모든 건강검진 시행 병원으로부터 수집하며 매년 혹은 2년마다 건강검진을 받을 것을 권고하고 있으며 이는 무료로 진행되고 있다⁸. 2014년 건강검진에 참여한 사람은 전체 건강검진 대상 인구의 74.8%⁷로 한국의 전 인구를 대변한다 할 수 있다.

제공하는 데이터베이스 (database, 이하 DB)로 자격 DB, 국가검진 DB, 의료이용 DB 등이 있으며, 자격 DB에 크게 socio-demographics domain과

vital statistics domain을 포함하며 성별, 연령, 거주지, 생년월일, 사망일 등의 변수가 포함되어 있다. 국가검진 DB에는 일반 건강 검진과 생애 전환기 검진, 국가 암검진이 포함되어 있으며, 이 중 일반 건강 검진에 건강 행태, 진찰 소견, laboratory test에 대한 검진 결과가 포함된다. 제공되는 변수로 흡연, 음주, 신체 활동, 과거력, 가족력 등의 생활 습관에 관련한 변수와 키, 몸무게, 허리 둘레, 체질량지수, 혈압의 일반적인 진찰 소견 변수, 혈색소, 공복 혈당, 총 콜레스테롤, HDL 콜레스테롤, triglyceride, 간기능, 신기능에 관련한 혈액 검사, 뇨단백을 위한 소변 검사 결과 변수가 포함된다. 의료이용 DB는 대상자의 진료 내역에 대해 요양 기관이 요양 급여를 청구하여 심사 결정된 후 지급이 결정된 자료로 국민건강보험공단 심사평가원의 명세서 서식 구분에 따라 각각 명세서, 진료내역, 상병내역, 처방전교부상세내역의 세부 테이블로 구성되어 있다.

진료 테이블은 요양기관 (병원, 의원, 약국 등)에서 대상자에게 진료, 조제 등을 하고 청구 방법에 따라 작성한 청구단위 명세서 공통내역이 포함된 명세서 (T20) 테이블, 진료 (조제 등 포함) 행위 및 그에 따른 청구금액 발생 상세 자료가 포함된 진료내역 (T30) 테이블, 수진자 상병내역 (진료 받은 모든 상병을 관리) 관련 상세 자료가 포함된 상병내역 (T40), 처방전교부건별 원외처방내역 관련 상세 자료가 포함된 처방전교부상세내역 (T60) 테이블이 포함되어 있다⁹.

제 2 절 문제 제기

제 1 항 Framingham Heart Study 를 이용한 심뇌혈관질환 예측 모형

모든 CVD 와 그 구성 질환으로 진행될 위험을 예측하기 위한 다변수 위험 함수가 2008 년 Framingham Heart Study (FHS)를 이용하여 제작되었다¹⁰. 이 연구는 Cox 비례 위험 함수를 30 에서 74 세 사이의 기저에 CVD 가 없는 일반 건강 검진에 참여한 8,491 명 (평균 연령 49 세; 4,522 명의 여성)의 Framingham 연구 참여자의 첫 CVD 발생을 평가하기 위해 사용하였다. 성별 다변수 위험 함수 (일반 CVD 알고리즘)는 연령, 총 콜레스테롤과 HDL 콜레스테롤, 수축기 혈압, 고혈압 치료, 흡연과 당뇨를 포함하여 제작되었다. 또한 개별 CVD 의 발생 (관상동맥 심질환, 뇌졸중, 말초 동맥 질환, 또는 심부전)의 예측에 대한 일반 CVD 알고리즘의 성능을 평가하였다. 12 년이 넘는 추적 기간 동안 1,174 명 (456 명의 여성)이 첫 CVD 발생을 겪었다. 이 알고리즘은 C-statistics 에서 0.763 (남성), 0.793 (여성)으로 좋은 성능을 보여 주었다. 이에 저자는 CVD 의 위험을 평가하기 위해 간편하게 사용 가능하며 추정된 CVD 발생 비율은 CVD 의 위험을 계량하고 예방하는 데 도움이 될 것이라 주장했다.

하지만, 1976년 Framingham Heart Study를 사용해 개발된 예측 모형¹¹ 이후로 CVD의 유병률과 사망률은 미국에서 점차 감소되어 왔고¹², 비록 이처럼 Framingham Risk Score (FRS)가 최근까지 갱신되어 왔지만¹⁰, 서구

국가에서 이를 자국의 여러 민족 집단에게 적용했을 때 CVD가 과대평가되는 문제가 꾸준히 제기되어 왔다¹³.

FRS의 한계를 극복하기 위해, Pooled Cohort Equation (PCE)가 미심장학회 (American College of Cardiology/American Heart Association - ACC/AHA)에 의해 제안되었다¹⁴.

제 2 항 ACC/AHA 심뇌혈관질환 예측 모형

2014 년 ACC/AHA 는 Work Group 을 결성하여 전문 교육 및 연구를 통해 CVD 가 발병한 환자들의 관리를 개선하고 최적의 환자 관리를 제공하고 심혈관 건강을 촉진하는 지침, 표준 및 정책을 개발하고자 하는 것에 목표를 두었다.

Work Group 은 백인과 아프리카계 미국인의 미국 인구를 광범위하게 대표하는 지역 사회 기반 코호트에서 risk equation 을 도출했으며, 그룹은 최초의 CVD 발생 (비 치명적 심근경색, 관상동맥 심질환으로 인한 사망, 치명적 혹은 비 치명적 뇌졸중)의 추정치 환자와 의료 제공자 모두에게 더 큰 관련이 있을 것으로 간주되어 관심결과로서 이에 초점을 맞추었다.

Risk equation 의 작성에 포함된 최종 pooled cohort 에는 ARIC (Atherosclerosis Risk in Communities) 연구¹⁵, Cardiovascular Health Study¹⁶, 및 CARDIA (Coronary Artery Risk Development in Young Adults) 연구¹⁷, Framingham Original 및 Offspring Study 코호트의 해당 데이터와 이를 결합한 연구^{18,19} 를 포함하여 여러 대규모, 인종 및 지리적으로

다양한 현대 NHLBI (National Heart, Lung, and Blood Institution) 후원 코호트 연구를 포함하였다.

Work Group 은 백인 및 아프리카계 미국인 남녀의 10년 CVD 위험에 대한 성별 및 인종 별 추정치를 제공하는 PCE 를 도출하고 내부적으로 검토하였다. 40 세에서 79 세 사이의 심근경색, 관상동맥질환, 뇌혈관질환을 예측하였으며 위험 평가 방정식에 포함된 변수로 연령, 총 콜레스테롤, HDL 콜레스테롤, 수축기 혈압, 당뇨병 및 현재 흡연 상태를 포함하였고, 일차 진료용 단순화된 점수 체계로 공중에게 공개되어 있다.

하지만, Work Group이 PCE를 작성 시 CVD의 추정 10년 위험은 일반적으로 히스패닉계 미국인 및 아시아계 미국인 인구에서 더 낮아²⁰ 과대추정의 위험이 있어 사용에 유의할 것을 선제적으로 경고 하고 있다.

PCE는 미국²¹에서 뿐 아니라, 여러 아시아 국가^{22,23}에서 검증되었으며, 가장 최근 2018년 개정 되었다²⁴. 아시아인은 미국인에 비해 CVD로 인한 사망이 2/3에 그치는 것으로 알려져 있으며², 아시아 국가 또한 고령 사회로 진입하여²⁵ CVD로 인한 사망과 장애가 증가하고 있어²⁶, 신뢰할만한 CVD 예측 모형이 필요한 상황이다.

제 3 절 관련 연구 동향

제 1 항 Korean Heart Study 를 이용한 심뇌혈관질환 예측 모형과 Framingham risk score 의 비교

Jee 등²⁷은 1996년부터 2001년까지 한국의 18개 건강증진센터의 기저에 관상 동맥 심장 질환 (CHD)이 없는 30세에서 74세 사이의 한국인 268,315명의 데이터를 이용하여 CHD 위험 모형을 개발하고 이를 Framingham Risk Score와 비교하였다.

1997년과 2011년 사이의 치명적이지 않거나 치명적인 CHD 사건의 발생을 outcome으로 하였으며, 11.6년의 중앙값 추적 기간 동안 2,596건의 CHD 발생 (1,903건의 비 치명적, 693건의 치명적)이 있었다. 기본 CHD 모형에 HDL 콜레스테롤, LDL 콜레스테롤과 중성지방을 추가하여 최적의 CHD 모형을 생성하고 AUROC와 net reclassification index (NRI)를 이용하여 모형을 평가하였다.

최적 모형의 AUROC는 남성의 경우 0.764, 여성의 경우 0.815로 좋은 성능을 보였으며, FRS를 이 코호트에 적용하였을 때 관찰된 실제 발생보다 3~6배 많은 발생을 예측하였다. 이에 연구자는 recalibration을 통해 FRS의 성능을 향상시켰다.

이를 통해 CHD 발병률이 낮은 한국 인구에서 FRS가 CHD의 위험을 과대 평가함을 제시하고 있다.

제 2 항 Korean Heart Study 를 이용한 심뇌혈관질환 예측 모형과 Pooled Cohort Equation 의 비교

ACC/AHA 의 PCE 를 한국의 KHS 에 적용해 성능을 평가하고 죽상동맥경화성 심혈관 질환 (Atherosclerotic cardiovascular disease - ASCVD)의 발생의 예측 모형 작성을 하기 위해 연구가 수행되었다²⁸. KHS 코호트에는 기저에 ASCVD 가 없는 40 에서 79 세 한국 성인이 200,010 명 포함되었으며, 10 년 ASCVD 위험을 예측하는 PCE 의 discrimination, calibration 과 recalibration 을 평가하였다. 예측 모형은 Cox 비례 위험 모형을 이용하였다.

판별 분석에서 PCE 의 백인과 아프리카계 미국인 모형은 KHS 코호트에서 환자군과 대조군의 경우를 적절히 구별하였다. 남성의 경우 AUROC 는 백인 모형 0.727, 아프리카계 미국인 모형 0.725 및 한국인 모형은 0.741 이었으며, 여성의 경우 각각 0.738, 0.739 및 0.745 였다.

KHS 코호트에서 남성의 10 년 ASCVD 위험은 56.5% (백인 모형) 및 74.1% (아프리카계 미국인 모형) 과대 추정된 반면 여성의 위험은 27.9% (백인 모형) 및 29.1% (아프리카계 미국인 모형) 과소 추정되었다.

이처럼 서구 인구를 대상으로 하여 이전에 개발된 모형은 아시아 국가에서 CVD의 위험을 과대 추정한다는 비판이 지속적으로 있어 왔으며²⁹, 이에, 연구자들은 자발적인 건강 검진을 수행한 한국인을 포함한 연구²⁸와 전향적 코호트에 등록된 중국인을 포함한 연구³⁰에서 예측모형을 구축하고 이를 검증하였다.

제 4 절 연구의 필요성

백인과 아프리카계 미국인에 기반한 PCE 를 아시아인에게 적용함으로써 예측 성능은 한국의 위험군에 따라 과대 혹은 과소 추정되었기에 악화되었을 것으로 예상된다. 위의 연구자들은 그 이유로 인구 구조, 추적 관찰 기간, 코호트에 따른 결과 정의의 차이에 따라 위험의 수준이 상이하야 이러한 결과가 나타났을 것으로 상정하였다²⁸.

또한, 앞서 한국인을 대상으로 선행된 연구^{27,28} 는 건강증진센터의 임상 데이터를 이용하였으나 이 데이터는 혈압 및 기타 의학적 검사 결과의 1 회 측정된 데이터만을 사용하여 측정 오류의 가능성이 상존한다.

이런 이유로, 동일 검진 수신자에 대한 여러 번 검진을 시행한 시간에 대한 가변효과를 고려한 반복 측정 데이터에 대한 분석이 필요하며, 한국인에게 정교히 적용될 수 있는 개선된 CVD 예측 모형이 필요하다.

제 5 절 연구의 내용, 목적 및 의의

한국의 국민건강보험공단의 주관으로 건강검진을 시행 받은 21,581,796명 (남성: 10,412,947명, 여성: 11,168,849명)의 검진 데이터와 명세서 데이터를 이용하여 생존분석 중 Cox 비례 위험 모형과 시간에 대한 가변 효과를 고려한 확장된 Cox 비례 위험 모형을 제작하고 한국인의 8년간의 CVD 발병을 예측하는 모형을 작성하였다. 또한 이를 현재 임상에서 널리 이용되고 있는 CVD의 예측 모형인 FRS, PCE와 비교하였다.

- 2009년 1월에서 2016년 12월 사이 건강 검진 데이터의 첫 검진 데이터를 사용
- 8년간의 CVD의 발병 확률을 계산하는 Cox 비례 위험 모형 작성
- 시간 가변 효과를 보기 위해 확장된 Cox 비례 위험 모형 작성
- 우리 비례 위험 모형의 성능과 FRS, PCE를 우리 데이터에 적용했을 때의 성능을 비교

FRS와 PCE는 연령과 혈압, 총 콜레스테롤, 흡연 상태, 당뇨 여부, HDL 콜레스테롤, LDL 콜레스테롤, 중성지방 등의 CVD의 위험 요인으로 알려진 변수를 이용하여 CVD 예측에 이용하였으나, 본 연구의 경우 앞의 변수들을 모두 포함하며, 건강 검진 수신 시 기본으로 시행되는 체질량지수, 허리 둘레, 공복 혈당을 비롯하여 개인이 지니고 있는 질환의 치료 여부, 운동량, 음주량, 심장 질환 및 뇌졸중의 가족력 등을 추가로 고려하여 예측 모형을

작성하였다.

이를 통해 현재 한국인의 생활 습관의 현황을 알 수 있으며, 반복된 검진을 통해 생활 습관의 교정 시 얻을 수 있는 CVD의 발병 확률 변화를 알 수 있게 설계되었다.

또한, 검진 데이터가 모두 국민건강보험공단에 저장 및 관리 된다는 특성으로 인해 검진 시행 시 자동화된 계산이 가능하여 수신자에게 제공될 수 있다는 점에서 의의를 가진다 할 수 있다.

제 6 절 성능 평가를 위한 지표 상세

제 1 항 Area Under Receiver Operating Characteristic curve (AUROC)

Receiver Operating Characteristic (이하 ROC) 곡선³¹은 식별 임계값이 변할 때 이진 분류기 시스템의 진단 능력을 나타내는 그림으로, 다양한 임계 값 설정에서 False Positive Rate (FPR)에 대해 True Positive Rate (TPR)을 도표화 하여 생성된다.

TPR 은 recall 과 동의어이며 아래와 같이 정의한다:

$$TPR = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

FPR 은 아래와 같이 정의한다:

$$FPR = \frac{\text{False Positive}}{\text{False Positive} + \text{True Negative}}$$

분류 임계 값을 낮추면 더 많은 항목이 양성으로 분류되어 False Positive 와 True Positive 모두 증가하게 된다. AUC 는 가능한 모든 분류 임계 값에 대한 종합적인 성능 측정값을 제공하여 예측이 모두 틀리면 0.0 의 값을, 모두 정확히 맞추면 1.0 의 값을 갖도록 한다.

제 2 항 Brier Score

Brier score³²의 간소화된 공식은 다음과 같다:

$$BS = \frac{1}{N} \sum_{t=1}^N (f_t - o_t)^2$$

여기서, f_t 는 예측된 확률, o_t 는 인스턴스 t 에서 실제 이벤트의 발생 (발생하지 않은 경우 0, 발생한 경우 1), 그리고 N 은 예측 인스턴스의 수이다. 실사용에서 예측의 평균 제곱 오차라고 할 수 있다. 이 공식은 이진 분류 형태의 사건에서 사용되며 위의 공식은 이진 분류에 대해서만 적절하다.

제 3 항 Harrell's C-index

데이터가 중도 절단될 수 있는 생존 분석에서 위험 모형을 평가하는 데 일반적으로 사용하며 위험 점수를 생성하는 모형에 대한 적합도 측정 방법이다³³.

환자 i 에 대해서, 위험 모형은 위험 점수 η_i 을 할당한다. 위험 모형이 적절히 구성된 경우, 질환으로 진행될 때까지 걸리는 시간이 짧은 환자가 더 높은 위험 점수를 받게 될 것이다. 즉, 역으로 치환하면 위험 점수가 더 높은 환자는 질병에 걸리는 시간이 더 짧아야 한다.

모든 환자 쌍 i 와 j 에 대해 ($i \neq j$), 그의 위험 점수와 발생 시간 (T)을 고려해야 한다. Harrell's C 는 일치 확률 $P(\eta_j > \eta_i | T_i > T_j)$ 를 추정하도록 설계되었으며³⁴, 이는 두 개의 독립적인 생존 시간 쌍의 순위 T_i, T_j 및 예측 η_i, η_j 를 비교한다. 일치 확률은 η_i 의 큰 값이 T_i 의 작은 값과 연관되는지 여부를 평가하고 그 반대의 경우도 마찬가지로 평가한다. Harrell' C 는 시간에 따른 ROC 곡선 아래 면적의 요약 측정치로 해석될 수도 있다. C = 0.5의 값은 정보가 없는 예측 규칙에 해당하는 반면 C = 1은 완벽한 연관에 해당된다. 따라서 Harrell의 C는 관찰된 생존 시간의 전체 범위를 설명하는 해석하기 쉬운 계수라고 할 수 있다.

제 2 장 심뇌혈관질환 예측 모형의 방법론

이 연구는 서울대학교 Institutional Review Board (IRB)의 승인을 받아 수행되었다 (1606-016-768). 이 연구의 특성에 의해 제공된 동의와 승인은 IRB에 의해 면제되었다. 모든 방법과 절차는 관련된 가이드라인과 규제에 맞추어 수행되었다.

제 1 절 연구에 포함된 인구 집단의 정의

연구는 2009년 1월에서 2016년 12월 사이 수행된 30,613,756명의 99,585,141건의 건강검진 데이터를 검토하는 것으로 시작하였다. 검진 당시 40세 미만 혹은 79세 초과인 대상자에게 시행된 건강검진 29,546,912건은 제외하였다. 설문과 건강검진 결과에 결측이 있던 2,840,127건은 제외되었다.

22,259,625명 중 첫 건강검진 시 자가 보고 표준 설문지에 CVD로 진단 혹은 치료 받은 이력이 있는 677,829명은 제외하였다. 최종 데이터셋으로 21,581,796명의 건강검진 기록을 분석하였다 (그림 2-1).

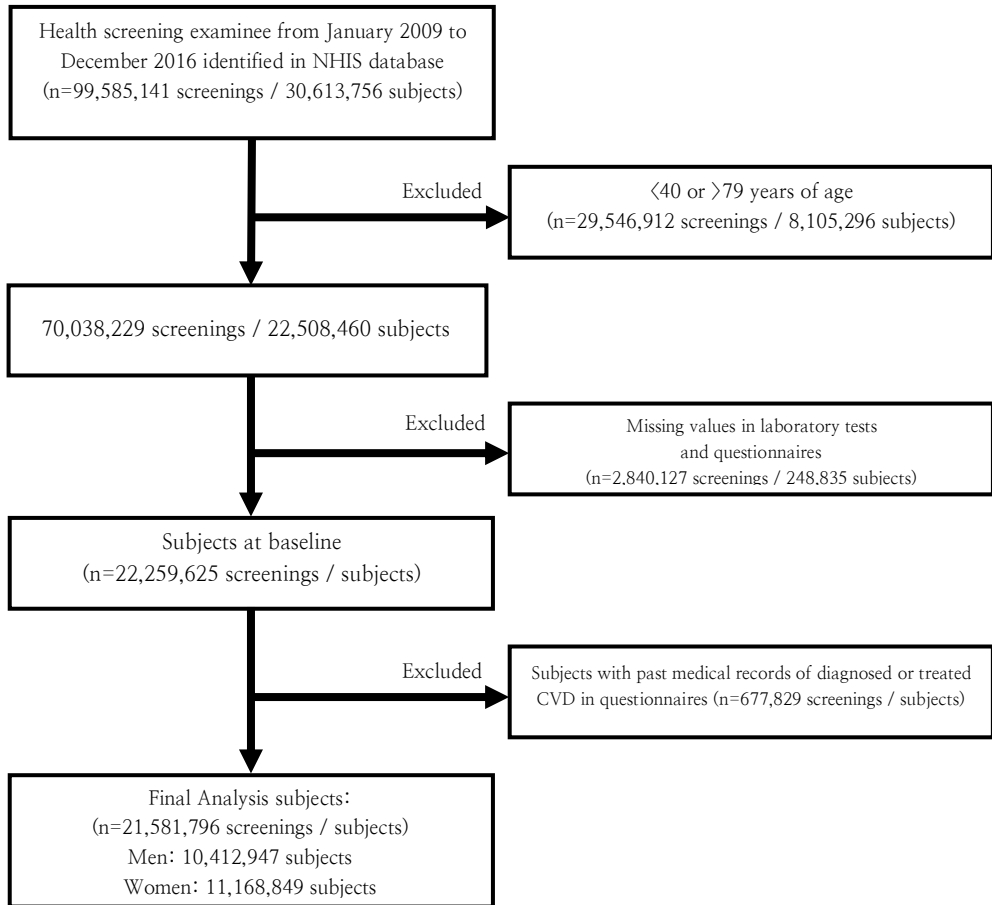


그림 2-1 심뇌혈관질환 예측 모형을 위한 연구 참여자의 코호트 정의

2017년 한국의 통계청³⁵에 따르면 40에서 79세의 인구는 약 2,570만 (1,260만 여명이 남성, 1,300만 여명이 여성)이었다. 본 연구의 연구 대상자는 이 연령군의 84.0%에 해당한다.

건강검진 데이터는 (1) 나이, 성별, 거주지와 같은 인구학적 변수; (2) 흡연 상태/량/기간, 주당 음주 빈도와 음주량, 주당 신체 활동의 횟수, 약물 복용력, 질환 가족력과 같은 건강 행태에 관련한 변수; (3) 키, 몸무게, 혈압, 허리둘레와 같은 신체 계측 결과; (4) 공복 혈당 (fasting serum glucose), 지단백검사 (lipid profile), 혈색소 (hemoglobin), 혈청 크레아티닌 (serum

creatinine), 간 효소 (liver enzymes), 단백뇨 검출을 위한 소변 검사를 포함한 각종 검사 수치를 포함한다.

제 2 절 설명변수의 정의

연구 대상자는 흡연 상태에 따라 비흡연, 과거 흡연, 현재 흡연자로 분류되었다. 과거 그리고 현재 흡연의 총량은 갑년 (pack-year)으로 환산하였다. 일일 평균 음주량은 세계보건기구 (World Health Organization - WHO)³⁶의 음주량의 정의에 대한 공식으로 일 평균 음주량을 계산하였으며 표준 음주 행위는 50cc 소주잔을 기준으로 하였다.

표 2-1 WHO 기준 일 평균 음주량의 위험 분류

	Males	Females
Low Risk	1 to 40g	1 to 20g
Medium Risk	41 to 60g	21 to 40g
High Risk	61 to 100g	41 to 60g
Very High Risk	101+g	61+g

일평균 음주량 (g/day) = [음주 빈도 (주당 음주한 횟수) * 음주량 (하루 음주량) * 잔의 부피 (cc) * 알코올의 부피 (0.2) * 알코올의 밀도 (0.785)] / 7 (days/week)의 수식을 이용하였다. 이후 평균 음주량을 네 범주로 나눴는데³⁷, (1) 금주, (2) 저위험 음주 - 일평균 음주량이 남성에서 < 41.0 g/day, 여성에서 < 21.0 g/day; (3) 중간 위험 음주: 일평균 음주량이 남자에서 41.0에서 60.0 g/day, 여성에서 21.0에서 40.0 g/day; (4) 고위험 음주 - 일평균 음주량이 남성에서 ≥ 61.0 g/day, 여성에서 ≥ 41.0

g/day인 경우로 구분하였다.

신체 활동의 경우, 건강검진의 표준 설문지에 따라 지난 일주일 동안 격렬한 활동 (예를 들어, 달리기, 에어로빅, 빠르게 자전거 타기 혹은 등산 같은 활동)을 하루에 20분 이상한 횟수, 중간 정도의 활동 (예를 들어, 빠르게 걷기, 복식 테니스, 일반적인 속도로 자전거 타기, 걸레질)을 하루에 30분 이상한 횟수, 하루에 적어도 30분 동안 걸은 횟수를 설문하였다. International Physical Activity Questionnaire (IPAQ)³⁸에서 제안한 점수 체계를 따라 총 일주일 당 분으로 환산한 신진대사 해당치 (total physical activity metabolic equivalent task - METs * min/week)는 걷기 (3.3 METs), 중간 강도 활동 (4.0 METs), 고강도 활동 (8.0 METs)의 빈도의 합으로 계산하였다. Victoria 등³⁹이 제안한 바에 따라 이를 저강도 (< 600 METs * min/week), 중간 강도 (600-2999 METs * min/week), 고강도 (≥ 3000 METs * min/week) 활동 군으로 분류하였다.

표 2-2 WHO 기준 체질량 지수의 구분

BMI	Nutritional status
Below 18.5	Underweight
18.5-24.9	Normal weight
25.0-29.9	Pre-obesity
30.0+	Obesity

체질량지수 (Body Mass Index - BMI)의 측정 결과는 WHO⁴⁰에서 제안한 네 가지 분류로 나뉘었다. BMI < 18.5 kg/m²인 경우 저체중, 18.5-24.9 kg/m²인 경우 적정 체중, 25.0-29.9 kg/m²의 경우 과체중, 비만은 ≥ 30.0 kg/m²로 정의하였다.

추정된 사구체 여과율 (estimated Glomerular Filtration Rate - eGFR)은 혈청 크레아티닌 수지에 기반한 Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) 공식⁴¹을 이용하여 계산하였으며 비정상 eGFR은 < 60 mL/min/1.73m²로 정의하였다.

CKD-EPI 공식은 다음과 같다:

$$eGFR = 141 \times \min(S_{Cr}/\kappa, 1)^\alpha \times \max(S_{Cr}/\kappa, 1)^{-1.209} \times 0.993^{Age} \\ \times 1.018[if Female] \times 1.159[if Black]$$

S_{Cr} 은 혈청 크레아티닌 (mg/dL)이며, κ 는 여성의 경우 0.7, 남성의 경우 0.9이다. α 는 여성의 경우 -0.329, 남성의 경우 -0.411이다. \min 은 S_{Cr}/κ 의 최소치를 의미하며 \max 는 S_{Cr}/κ 의 최대치를 의미한다.

단백뇨 검출을 위한 요 시험지 검사 (urine dipstick test)의 결과는 negative, trace, 1+부터 4+로 보고되었으며, 1+ 이상을 단백뇨 양성으로 정의하였다.

제 3 절 결과변수의 정의

CVD의 발생으로서 결과 사건을 정의하였으며, 이는 급성 관상동맥 증후군 (Acute Coronary Syndrome - ACS), 뇌경색 (Cerebral Infarction), 뇌출혈 (Cerebral Hemorrhage)을 포함한다. ACS는 International Classification of Disease (ICD-10)⁴²에 따라 I20 또는 I21과 명세서에 경피관상동맥중재시술, 관상동맥우회술의 시행 여부, 또는 혈전용해제의 처방이 있는 경우로 정의하였다. 뇌경색은 I63과 명세서에 뇌자기공명영상 또는 acetylsalicylic acid의 처방이 있는 경우로 정의하였다. 뇌출혈은 I60, I61 또는 I62와 뇌 단층촬영영상의 처방이 있는 경우로 정의하였다.

제 4 절 Cox 비례 위험 모형의 이론적 배경

Cox 비례 위험 모형은 위험 함수에 설명 변수에 대한 회귀식을 포함하는 모형으로 다음과 같은 특징이 있어 널리 이용되고 있다⁴³.

- 기저분포에 대한 가정이 필요하지 않음
- 시간 가변 설명 변수를 회귀 모형에 포함시킬 수 있음
- 추정된 회귀 계수는 설명 변수와 위험 함수 간의 관계를 나타냄

제 1 항 Cox 비례 위험 모형

비례 위험 모형은 다음과 같은 모형을 사용한다⁴³:

$$\lambda(t; Z_i) = \lambda_0(t) \exp(Z_i' \beta)$$

여기서 $\lambda_0(t)$ 는 분포 가정이 주어지지 않은 기저 위험 함수 (baseline hazard function)이며, $\beta = (\beta_1, \dots, \beta_p)$ 는 $p \times 1$ 벡터로 설명 변수의 효과를 추정하는 회귀 계수로 설명 변수 Z_{ik} 가 한 단위 증가할 때마다 $\exp(\beta_k)$ 만큼 위험률이 증가하게 된다. 예를 들어 $p = 1$ 일 때, 즉, 한 개의 설명 변수에 대해 두 개체 i 와 j 의 사건 발생 위험비는

$$\frac{\lambda(t; Z_i)}{\lambda(t; Z_j)} = \frac{\lambda_0(t) \exp(Z_i \beta)}{\lambda_0(t) \exp(Z_j \beta)} = \exp[(Z_i - Z_j) \beta]$$

이 된다. 위 식의 좌변은 시간의 함수지만 우변의 결과를 통해 위험률의 비는 시간에 의존하지 않고 항상 일정하다는 것을 알 수 있다. 이러한 효과는 시간이 변하더라도 불변함을 의미한다. 이러한 성질 때문에 비례위험 모형이라고 한다.

제 2 항 확장된 Cox 비례 위험 모형

일정 기간 동안 결과 변수를 측정하는 종단 연구에서는 설명 변수의 값이 변할 경우 이를 시간 가변 (time-varying) 설명 변수라고 한다⁴³. 이를 이용해 다음과 같이 위험률을 정의할 수 있다.

$$\lambda_i(t) = \lambda_0(t) \exp(\beta' z_i(t))$$

β 추정을 위해 부분 우도 함수는

$$PL(\beta) = \prod_i^m \left[\frac{\exp(\beta' z_i(t_i))}{\sum_{l \in R_i} \exp[\beta' z_l(t_i)]} \right]^{\delta_i}$$

이며, R_i 는 $t_{(i)}$ 시점에서 위험 그룹에 속한 개체의 모임이다.

제 3 항 Schoenfeld 잔차

i 번째 관측 개체의 j 번째 설명 변수를 이용하여 다음의 잔차가 정의된다⁴⁴.

$$\bar{r}_{ij} = \delta_i(z_{ij} - e_{ij}) \text{ where } e_{ij} = \frac{\sum_{l \in R_i} \exp(z_l \hat{\beta}) z_{lj}}{\sum_{l \in R_i} \exp(z_l \hat{\beta})}$$

여기서 R_i 는 t_i 시점에서의 위험군을 표시한다. \bar{r}_{ij} 는 $\delta_i = 1$ 인 개체에 대해서만 정의되며 가장 큰 값을 가지는 생존자료는 $\bar{r}_{ij} = 0$ 이 된다. 관측 개체마다 설명 변수의 개수만큼 p 개의 Schoenfeld 잔차가 정의된다. 이 잔차는 부분 우도 함수로부터 구할 수 있기 때문에 종종 부분 편차 (partial residual)라고도 한다. 큰 표본에서 \bar{r}_{ij} 는 서로 독립이며, 기대 값은 0이다.

비례 위험 모형의 시간에 따라 설명 변수의 효과가 고정된다는 가정을 위배한다면 이는 효과가 시간에 따라 변하는 함수 형태임을 의미한다. 이를 잔차도로 표현하여, 시간 축에 평행함을 보였다.

제 5 절 통계 분석

심뇌혈관질환 예측 모형은 Cox 비례 위험 모형을 사용해 구축되었다. 비례 위험 가정은 통계 검정과 Schoenfeld 잔차에 기반한 그래픽 진단 (그림 3-1, 그림 3-2)을 이용하여 검정하였다. Cox 모형의 시간에 대한 가변 효과 (time-varying effect)를 구축된 Cox 예측 모형과 비교하기 위해 참여자의 건강검진 횟수를 5회까지 확장하여 확장된 Cox 모형 (extended Cox model)을 고안하고 검정하였다 (표 3-3 ~ 표 3-5). 모형에 포함된 공변량은 stepwise로 선택되었으며 10회 교차 검증이 모형 제작과 검증에 사용되었다. Harrell 's C-index로 예측 모형의 성능을 계산하였고, Cox 모형과 PCE의 판별 능력을 비교하기 위해 수신자 작동특성곡선 분석의 곡선하 면적 (AUROC)와 Brier score를 계산하였다. 기저 생존 함수를 맞추기 위해 PCE의 위험 확률을 1.25로 나눠 건강검진 데이터의 8년 추정 위험과 맞추었다⁴⁵. 분석을 위해 R을 사용하였고 양측 < 0.05 를 유의하다고 판단하였다.

제 3 장 심뇌혈관질환 예측 모형의 결과

표 3-1 성별 연구 참여자의 기저 특성. 범주형 변수의 경우 백분율로 표시; 연속형 변수의 경우 평균 ± 표준편차로 표기하였음. 범주형 변수의 경우 Pearson's chi-squared test 와 연속형 변수의 경우 t-test 로 계산하였음. BP, blood pressure; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; eGFR, estimated glomerular filtration rate; CVD, cardiovascular disease. ^a≥1+ in urine dipstick for proteinuria.

Variables	Men (n = 10,412,947)	Women (n = 11,168,849)	P-value
Age (year)	51.5 ± 10.1	52.3 ± 10.5	<0.0001
Systolic BP (mmHg)	125.6 ± 14.6	121.0 ± 15.7	<0.0001
Diastolic BP (mmHg)	78.7 ± 10.1	75.0 ± 10.2	<0.0001
Waist circumference (cm)	84.5 ± 7.8	77.4 ± 8.7	<0.0001
Fasting serum glucose (mg/dL)	103.1 ± 29.1	97.1 ± 22.5	<0.0001
Gamma-glutamyl transferase (U/L)	54.9 ± 71.6	24.1 ± 30.5	<0.0001
Serum total cholesterol (mg/dL)	197.8 ± 37.0	200.1 ± 37.5	<0.0001
HDL (mg/dL)	51.8 ± 13.1	58.0 ± 14.0	<0.0001
LDL (mg/dL)	115.2 ± 35.3	118.8 ± 34.6	<0.0001
Triglycerides (mg/dL)	156.9 ± 103.3	115.5 ± 71.1	<0.0001
eGFR <60 mL/min/1.73 m ²	3.8%	4.8%	<0.0001
Proteinuria ^a	3.0%	2.4%	<0.0001
Body mass index (kg/m ²)			<0.0001
<18.5	2.0%	3.3%	
≥18.5, <25.0	58.2%	66.2%	
≥25.0, <30.0	36.0%	26.3%	
≥30.0	3.8%	4.3%	

Smoking			<0.0001
Nonsmoker	28.9%	94.1%	
Ex-smoker	28.1%	1.8%	
Smoker	42.9%	4.1%	
Pack-years (among ever-smokers)	11.7±13.7	0.4 ± 2.7	<0.0001
Alcohol drinking			<0.0001
No drinking	34.4%	77.0%	
Low risk	57.3%	20.8%	
Moderate risk	5.2%	1.8%	
High risk	3.2%	0.5%	
Activity group			<0.0001
Low activity	36.4%	44.4%	
Moderate activity	51.4%	46.3%	
High activity	12.2%	9.3%	
Patients with hyperlipidemia	2.0%	3.0%	<0.0001
Patients with hypertension	16.3%	16.6%	<0.0001
Patients with diabetes	6.8%	5.3%	<0.0001
Family history			
Heart disease	3.3%	3.9%	<0.0001
Stroke	6.5%	6.8%	<0.0001
Incident CVD events	195,022 (1.9%)	111,546 (1.0%)	<0.0001

표 3-1 은 성별에 따른 연구 참여자의 기저 특성을 보여준다. 남성에서 평균 연령 ± 표준편차는 51.5 ± 10.1 세였으며, 여성은 52.3 ± 10.5 세였다. 인-년 (person-years)은 남성에서 53,955,629, 여성에서 59,078,475 였다.

흡연, 음주를 하는 인구의 비율과 중간 강도에서 고강도 신체 활동을 하는 비율은 남성에서 여성보다 높았다. CVD 의 가족력이 있는 여성의 비율은 남성보다 약간 높았다.

평균 혈압, 공복 혈당, 혈청 중성지방 (triglycerides), Gamma-glutamyl transferase (GGT) 수치는 남성에서 약간 높았지만 고지혈증과 고혈압, 총 콜레스테롤, 고밀도 및 저밀도 지단백 콜레스테롤 수치는 여성에서 조금 더 높았다. 비정상 eGFR 은 여성에서 남성보다 흔했다.

제 1 절 심뇌혈관질환 예측 모형

8 년 간 추적 관찰 기간 동안 남성에서 195,022 (1.9%)명, 여성에서 111,546 (1.0%)명에서 CVD 가 발생하였다. CVD 의 incidence density 는 남성에서 100,000 명-년 (person-years) 당 361.5 건, 여성에서 188.8 건이었다. 현재 흡연 [남성: 보정된 위험률 (hazard ratio; 이하 HR) 1.463, 95% 신뢰구간 (confidence interval; 이하 CI) 1.443-1.484; 여성: HR 1.886, 95% CI 1.821-1.953], 당뇨 약제 [남성: HR 1.414, 95% CI 1.394-1.434; 여성: HR 1.504, 95% CI 1.477-1.532], 신체 활동 [(중간 강도 신체 활동군) 남성: HR 0.916, 95% CI 0.907-0.924 (고강도 신체 활동군) 남성: HR 0.910, 95% CI 0.897-0.923; (중간 강도 신체 활동군) 여성: HR 0.914, 95% CI 0.902-0.925 (고강도 신체 활동군) 여성: HR 0.893, 95% CI 0.874-0.913], 단백뇨 [남성: HR 1.409, 95% CI 1.384-1.435; 여성: HR 1.446, 95% CI 1.408-1.484]가 유의하였다 (표 3-2).

표 3-2 심뇌혈관질환 예측 모형에 포함된 변수의 보정된 위험비 (adjusted hazard ratio). $S_0(8)$ of men = 0.9798, $S_0(8)$ of women = 0.9916. Covariates included in Cox models were selected by stepwise procedures.

HR, hazard ratio; CI, confidence interval; BP, blood pressure; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol;

eGFR, estimated glomerular filtration rate ^a $\geq 1+$ in urine dipstick for proteinuria.

Variables	Men	Women
	Adjusted HR (95% CI)	Adjusted HR (95% CI)
Age (year)	1.061 (1.060–1.061)	1.074 (1.073–1.075)
Systolic BP (mmHg)	1.010 (1.010–1.010)	1.010 (1.010–1.011)
Diastolic BP (mmHg)	1.007 (1.006–1.008)	1.009 (1.008–1.010)
Waist circumference (cm)	1.004 (1.003–1.004)	1.006 (1.005–1.007)
Fasting serum glucose (mg/dL)	1.003 (1.003–1.003)	1.003 (1.003–1.004)
Gamma-glutamyl transferase (U/L)	1.001 (1.001–1.001)	1.001 (1.001–1.001)
Serum total cholesterol (mg/dL)	1.005 (1.005–1.005)	1.003 (1.002–1.004)
HDL (mg/dL)	0.988 (0.988–0.989)	0.991 (0.990–0.992)
LDL (mg/dL)	1.001 (1.000–1.001)	0.999 (0.998–1.000)
Triglycerides (mg/dL)	1.000 (1.000–1.000)	1.000 (1.000–1.001)
eGFR <60 mL/min/1.73 m ²	1.174 (1.156–1.192)	1.181 (1.161–1.202)
Proteinuria ^a	1.409 (1.384–1.435)	1.446 (1.408–1.484)
Body mass index (kg/m ²)		
<18.5	0.982 (0.951–1.014)	1.141 (1.095–1.188)
≥ 25.0 , <30.0	1.021 (1.009–1.033)	0.958 (0.944–0.973)
≥ 30.0	1.013 (0.984–1.042)	0.867 (0.841–0.895)
Smoking		
Ex-smoker	0.889 (0.876–0.901)	1.176 (1.113–1.242)
Smoker	1.463 (1.443–1.484)	1.886 (1.821–1.953)
Pack-year	1.005 (1.005–1.005)	1.004 (1.002–1.006)
Alcohol drinking		
Low risk	0.795 (0.787–0.803)	0.990 (0.971–1.010)
Moderate risk	0.760 (0.743–0.777)	1.126 (1.064–1.191)
High risk	0.783 (0.762–0.804)	1.154 (1.049–1.269)
Activity group		
Moderate activity	0.916 (0.907–0.924)	0.914 (0.902–0.925)
High activity	0.910 (0.897–0.923)	0.893 (0.874–0.913)
Current medication		
Blood glucose-lowering drugs	1.414 (1.394–1.434)	1.504 (1.477–1.532)
Antihypertensive drugs	1.183 (1.170–1.196)	1.240 (1.223–1.257)
Lipid-modifying drugs	0.996 (0.969–1.023)	0.822 (0.798–0.846)
Family history of heart disease	1.331 (1.297–1.365)	1.076 (1.037–1.117)
Family history of stroke	1.167 (1.146–1.187)	1.159 (1.133–1.186)

흡연의 분석에서 과거 흡연 범주가 남성에서 CVD 에 대해 위험을 낮추었지만 (0.889, 0.876-0.901) 여성에서 그런 효과가 없었다 (1.176, 1.113-1.242). 여성 중, CVD 의 위험은 음주 총량이 클 수록 증가하였으나 (중간 정도 위험: 1.126, 1.064-1.191, 고위험 음주: 1.154, 1.049-1.269), 남성의 경우 CVD 의 위험은 음주의 어떤 수준에서도 증가하지 않았다. 심뇌혈관 질환의 가족력은 양성 모두에서 CVD 의 위험을 증가하는 것과 연관되었다 (표 3-2).

남성에서 CVD 의 위험은 BMI 가 과체중일 경우 (1.021, 1.009-1.033) 증가되었지만 여성은 저체중인 경우가 가장 위험했다 (1.141, 1.095-1.188). 또한 과체중 (0.958, 0.944-0.973)과 비만 (0.867, 0.841-0.895)에서는 CVD 의 위험이 감소되는 것으로 관찰 되었다 (표 3-2).

제안된 CVD 예측 모형과 확장된 Cox 모형을 비교했을 때, 단백뇨의 계수가 남성 (1.409, 1.384-1.435 에서 1.512, 1.485-1.539), 여성 (1.446, 1.408-1.484 에서 1.573, 1.533-1.614) 모두에서 커지는 것을 관찰 하였다. 하지만 다른 변수에서 주목할만한 변화는 없었다 (표 3-3 ~ 표 3-5, 그림 3-1, 그림 3-2).

표 3-3 확장된 Cox 모형에 포함된 변수의 보정된 위험비 (adjusted hazard ratios). $S_0(8)$ of men = 0.9798, $S_0(8)$ of women = 0.9916. Covariates included in Cox models were selected by stepwise procedures.

HR, hazard ratio; CI, confidence interval; BP, blood pressure; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; eGFR, estimated glomerular filtration rate ^a $\geq 1+$ in urine dipstick for proteinuria.

Variables	Men	Women
	Adjusted HR (95% CI)	Adjusted HR (95% CI)
Age (year)	1.062 (1.061–1.063)	1.075 (1.075–1.076)
Systolic BP (mmHg)	1.011 (1.011–1.011)	1.012 (1.011–1.012)
Diastolic BP (mmHg)	1.007 (1.006–1.008)	1.010 (1.009–1.011)
Waist circumference (cm)	1.003 (1.002–1.004)	1.005 (1.004–1.006)
Fasting serum glucose (mg/dL)	1.003 (1.003–1.003)	1.003 (1.003–1.004)
Gamma-glutamyl transferase (U/L)	1.001 (1.000–1.001)	1.001 (1.001–1.002)
Serum total cholesterol (mg/dL)	1.005 (1.005–1.006)	1.004 (1.003–1.005)
HDL (mg/dL)	0.985 (0.985–0.986)	0.989 (0.988–0.990)
LDL (mg/dL)	1.000 (1.000–1.001)	0.999 (0.998–1.000)
Triglycerides (mg/dL)	1.000 (1.000–1.000)	1.000 (1.000–1.000)
eGFR <60 mL/min/1.73 m ²	1.187 (1.169–1.206)	1.194 (1.174–1.215)
Proteinuria ^a	1.512 (1.485–1.539)	1.573 (1.533–1.614)
Body mass index (kg/m ²)		
<18.5	0.995 (0.963–1.027)	1.186 (1.141–1.234)
≥ 25.0 , <30.0	1.006 (0.994–1.017)	0.948 (0.933–0.962)
≥ 30.0	1.005 (0.977–1.034)	0.852 (0.826–0.879)
Smoking		
Ex-smoker	0.890 (0.878–0.903)	1.199 (1.135–1.267)
Smoker	1.456 (1.435–1.476)	1.913 (1.847–1.982)
Pack-year	1.005 (1.005–1.005)	1.004 (1.002–1.006)
Alcohol drinking		
Low risk	0.803 (0.795–0.811)	0.992 (0.972–1.011)
Moderate risk	0.780 (0.763–0.797)	1.138 (1.076–1.204)
High risk	0.808 (0.787–0.830)	1.178 (1.072–1.296)
Activity group		

Moderate activity	0.909 (0.900–0.917)	0.904 (0.893–0.915)
High activity	0.898 (0.885–0.911)	0.877 (0.858–0.897)
Current medication		
Blood glucose-lowering drugs	1.394 (1.375–1.414)	1.515 (1.487–1.542)
Antihypertensive drugs	1.181 (1.169–1.194)	1.245 (1.229–1.262)
Lipid-modifying drugs	0.972 (0.946–0.999)	0.801 (0.778–0.824)
Family history of heart disease	1.336 (1.303–1.371)	1.076 (1.037–1.117)
Family history of stroke	1.167 (1.147–1.188)	1.155 (1.129–1.183)

표 3-4 심뇌혈관질환 예측 모형의 chi-square 와 p 값. BP, blood pressure; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; eGFR, estimated glomerular filtration rate ^a≥1+ in urine dipstick for proteinuria.

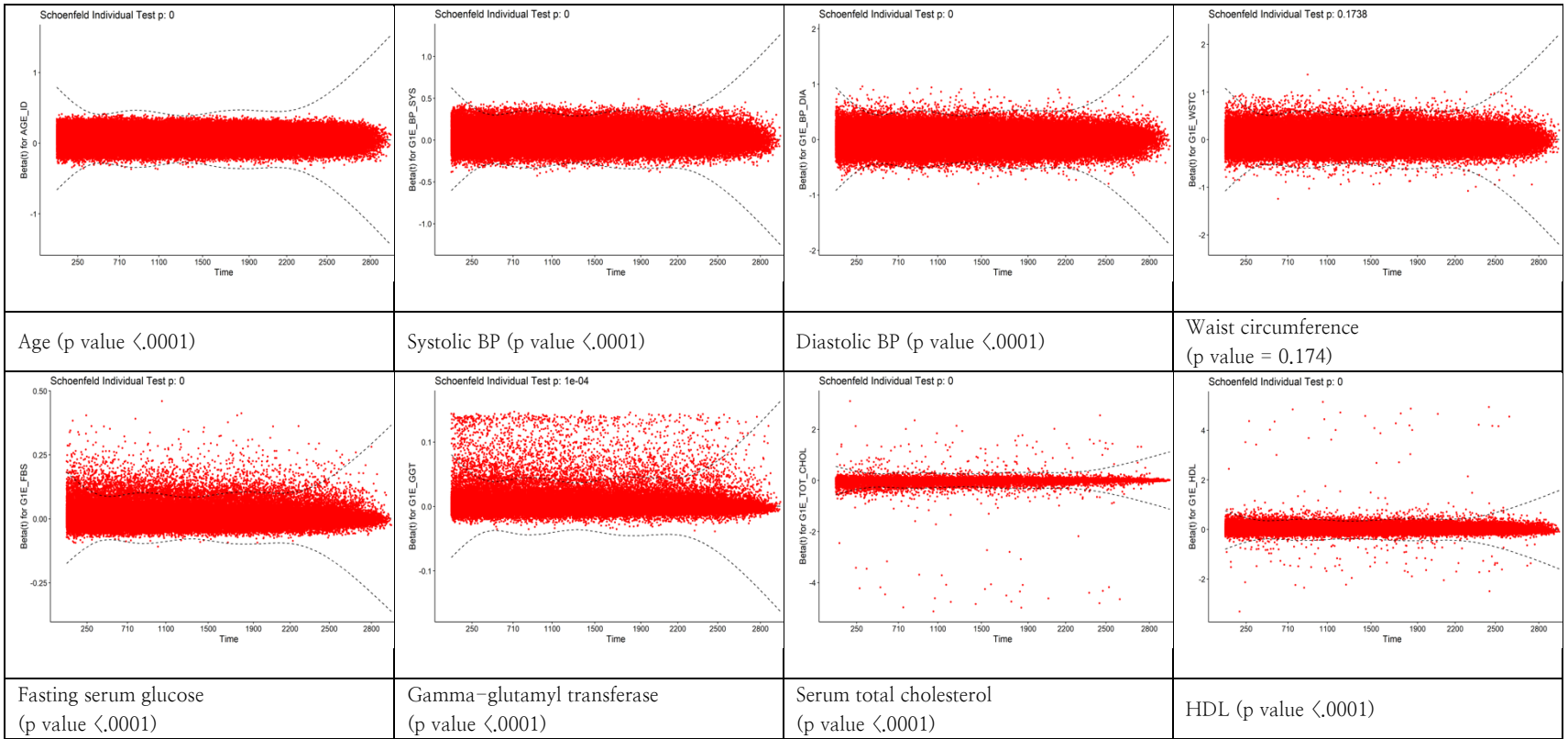
Variables	Men			Women		
	df	Chi-square	P-value	df	Chi-square	P-value
Age (year)	1	52241.770	<.0001	1	41162.186	<.0001
Systolic BP (mmHg)	1	2129.351	<.0001	1	1473.649	<.0001
Diastolic BP (mmHg)	1	471.399	<.0001	1	479.748	<.0001
Waist circumference (cm)	1	75.678	<.0001	1	133.694	<.0001
Fasting serum glucose (mg/dL)	1	2666.102	<.0001	1	1335.252	<.0001
Gamma-glutamyl transferase (U/L)	1	338.417	<.0001	1	350.305	<.0001
Serum total cholesterol (mg/dL)	1	664.332	<.0001	1	49.742	<.0001
HDL (mg/dL)	1	2017.570	<.0001	1	287.448	<.0001
LDL (mg/dL)	1	7.207	<.0001	1	3.446	0.063
Triglycerides (mg/dL)	1	0.562	0.453	1	11.142	0.001
eGFR <60 mL/min/1.73 m ²	1	423.001	<.0001	1	363.726	<.0001
Proteinuria ^a	1	1337.072	<.0001	1	756.070	<.0001
Body mass index (kg/m ²)						
<18.5	1	1.229	0.268	1	40.239	<.0001
≥25.0, <30.0	1	11.754	0.001	1	30.453	<.0001
≥30.0	1	0.721	0.396	1	78.838	<.0001
Smoking						
Ex-smoker	1	264.124	<.0001	1	33.362	<.0001
Smoker	1	2838.992	<.0001	1	1255.534	<.0001
Pack-year	1	1005.263	<.0001	1	18.300	<.0001
Alcohol drinking						
Low risk	1	2034.126	<.0001	1	1.037	0.308
Moderate risk	1	586.868	<.0001	1	17.065	<.0001
High risk	1	318.656	<.0001	1	8.704	0.003
Activity group						
Moderate activity	1	321.094	<.0001	1	203.536	<.0001
High activity	1	169.663	<.0001	1	103.508	<.0001

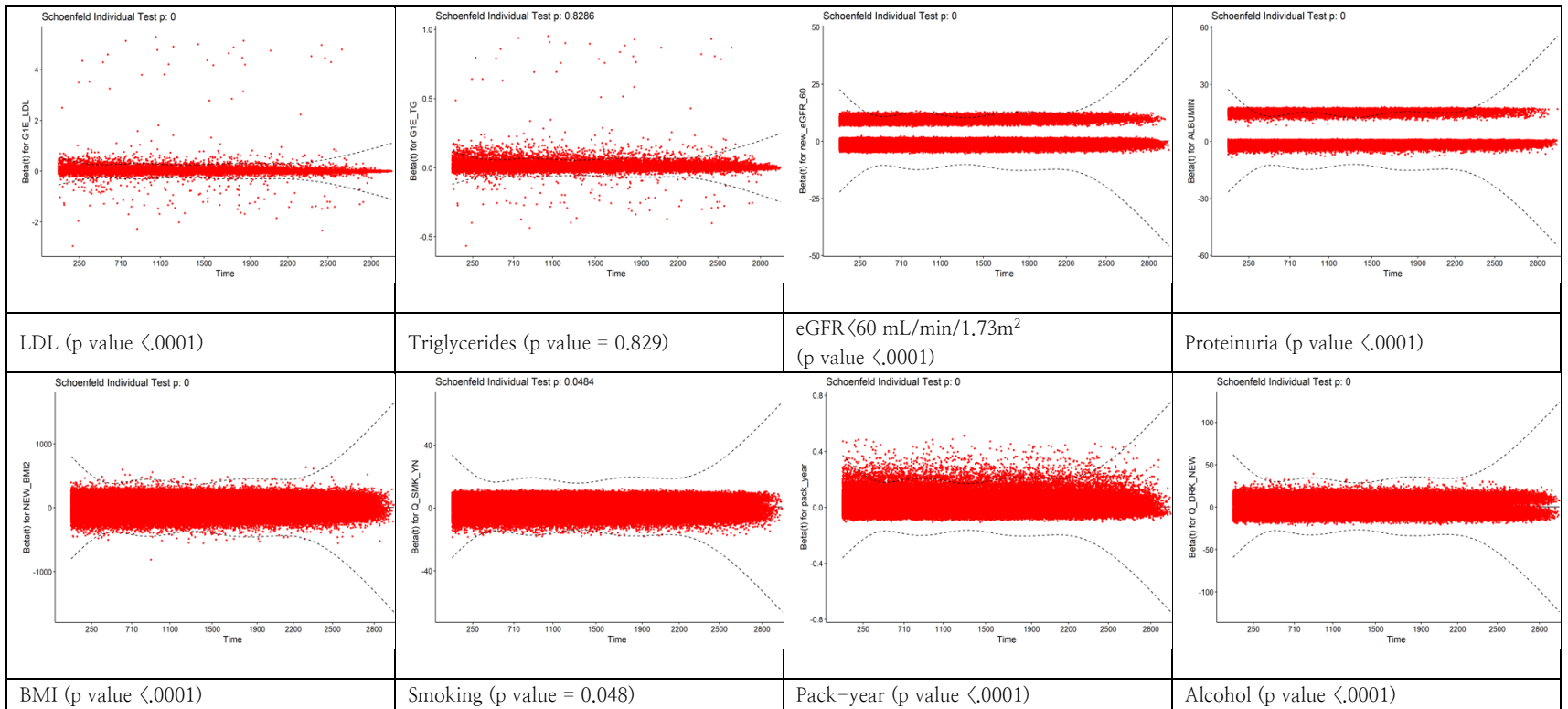
Current medication							
Blood glucose-lowering drugs	1	2277.645	<.0001	1	1940,170	<.0001	
Antihypertensive drugs	1	932.099	<.0001	1	974,284	<.0001	
Lipid-modifying drugs	1	0.099	0.753	1	175,497	<.0001	
Family history of heart disease	1	477.716	<.0001	1	15,151	<.0001	
Family history of stroke	1	299,543	<.0001	1	156,294	<.0001	

표 3-5 확장된 Cox 모형의 chi-square 와 *p* 값. BP, blood pressure; HDL, high-density lipoprotein cholesterol; LDL, low-density lipoprotein cholesterol; eGFR, estimated glomerular filtration rate ^a≥1+ in urine dipstick for proteinuria.

Variables	Men			Women		
	df	Chi-square	P-value	df	Chi-square	P-value
Age (year)	1	52555.092	<.0001	1	40267.906	<.0001
Systolic BP (mmHg)	1	2623.143	<.0001	1	1999.573	<.0001
Diastolic BP (mmHg)	1	465.172	<.0001	1	552.518	<.0001
Waist circumference (cm)	1	64.584	<.0001	1	119.794	<.0001
Fasting serum glucose (mg/dL)	1	3008.243	<.0001	1	1361.690	<.0001
Gamma-glutamyl transferase (U/L)	1	363.384	<.0001	1	445.855	<.0001
Serum total cholesterol (mg/dL)	1	490.968	<.0001	1	62.826	<.0001
HDL (mg/dL)	1	2547.729	<.0001	1	371.748	<.0001
LDL (mg/dL)	1	1.585	0.2081	1	4.171	0.0411
Triglycerides (mg/dL)	1	15.555	<.0001	1	6.337	0.0118
eGFR <60 mL/min/1.73 m ²	1	486.744	<.0001	1	413.976	<.0001
Proteinuria ^a	1	2080.883	<.0001	1	1191.305	<.0001
Body mass index (kg/m ²)						
<18.5	1	0.116	0.7339	1	73.186	<.0001
≥25.0, <30.0	1	0.863	0.3529	1	48.230	<.0001
≥30.0	1	0.111	0.7387	1	99.475	<.0001
Smoking						
Ex-smoker	1	254.747	<.0001	1	42.068	<.0001
Smoker	1	2762.972	<.0001	1	1313.720	<.0001
Pack-year	1	1013.516	<.0001	1	18.479	<.0001
Alcohol drinking						
Low risk	1	1867.009	<.0001	1	0.681	0.4092
Moderate risk	1	487.998	<.0001	1	20.558	<.0001
High risk	1	243.018	<.0001	1	11.498	0.0007
Activity group						
Moderate activity	1	379.096	<.0001	1	254.878	<.0001
High activity	1	221.749	<.0001	1	139.094	<.0001

Current medication							
Blood glucose-lowering drugs	1	2092.561	<.0001	1	1998.526	<.0001	
Antihypertensive drugs	1	917.405	<.0001	1	1018.164	<.0001	
Lipid-modifying drugs	1	4.225	0.0398	1	225.756	<.0001	
Family history of heart disease	1	492.181	<.0001	1	15.111	0.0001	
Family history of stroke	1	149.586	<.0001	1	149.586	<.0001	





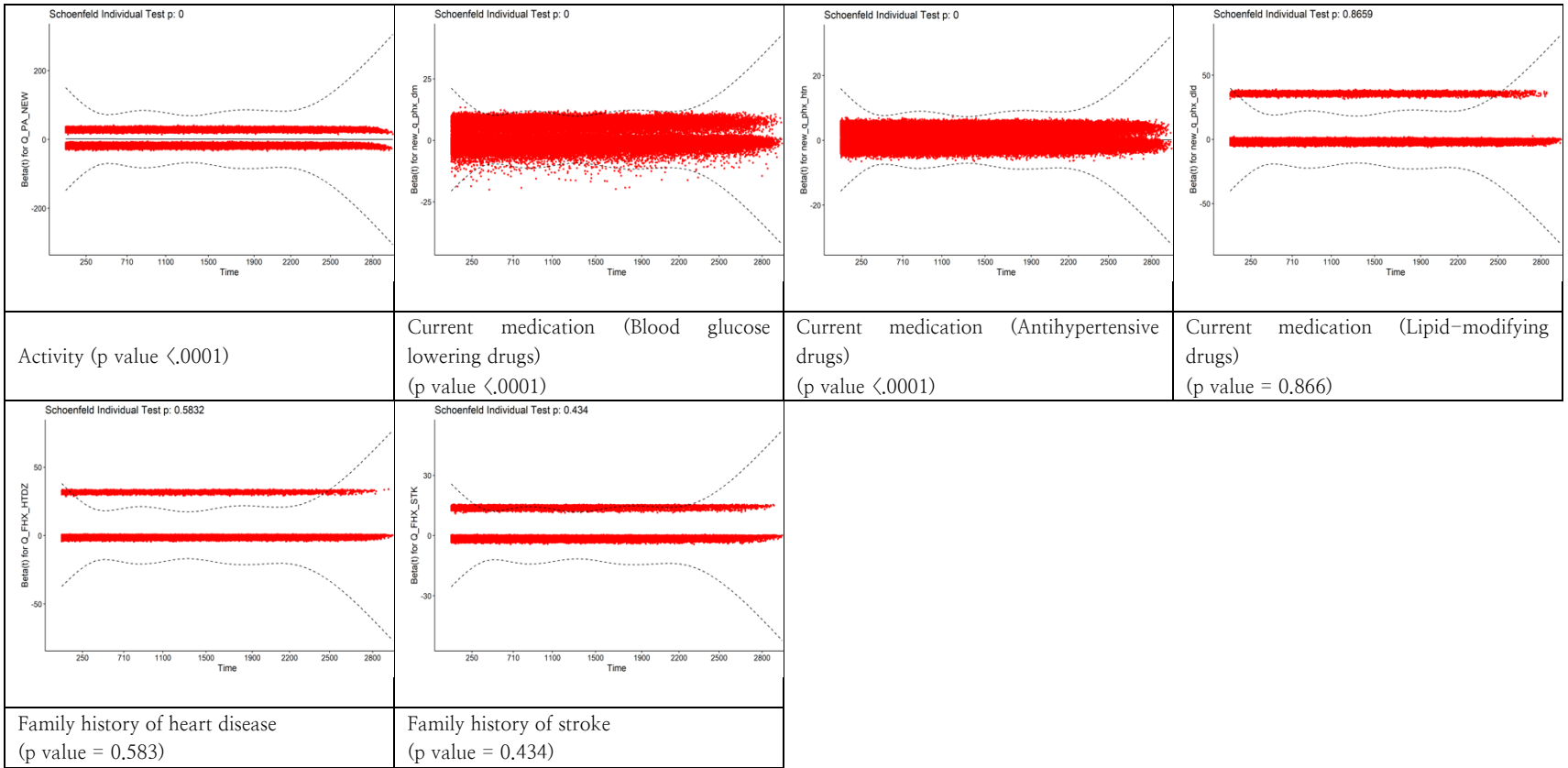
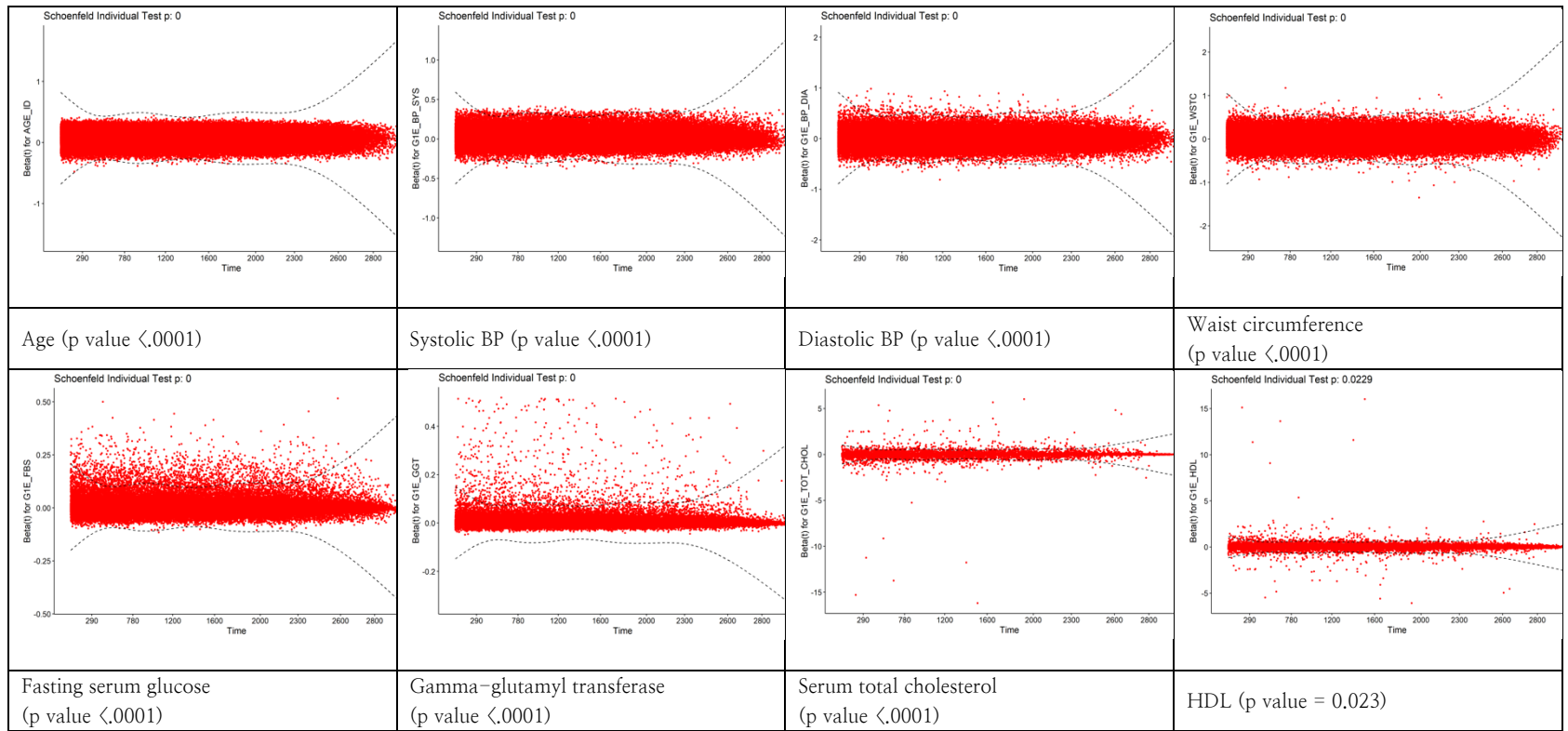
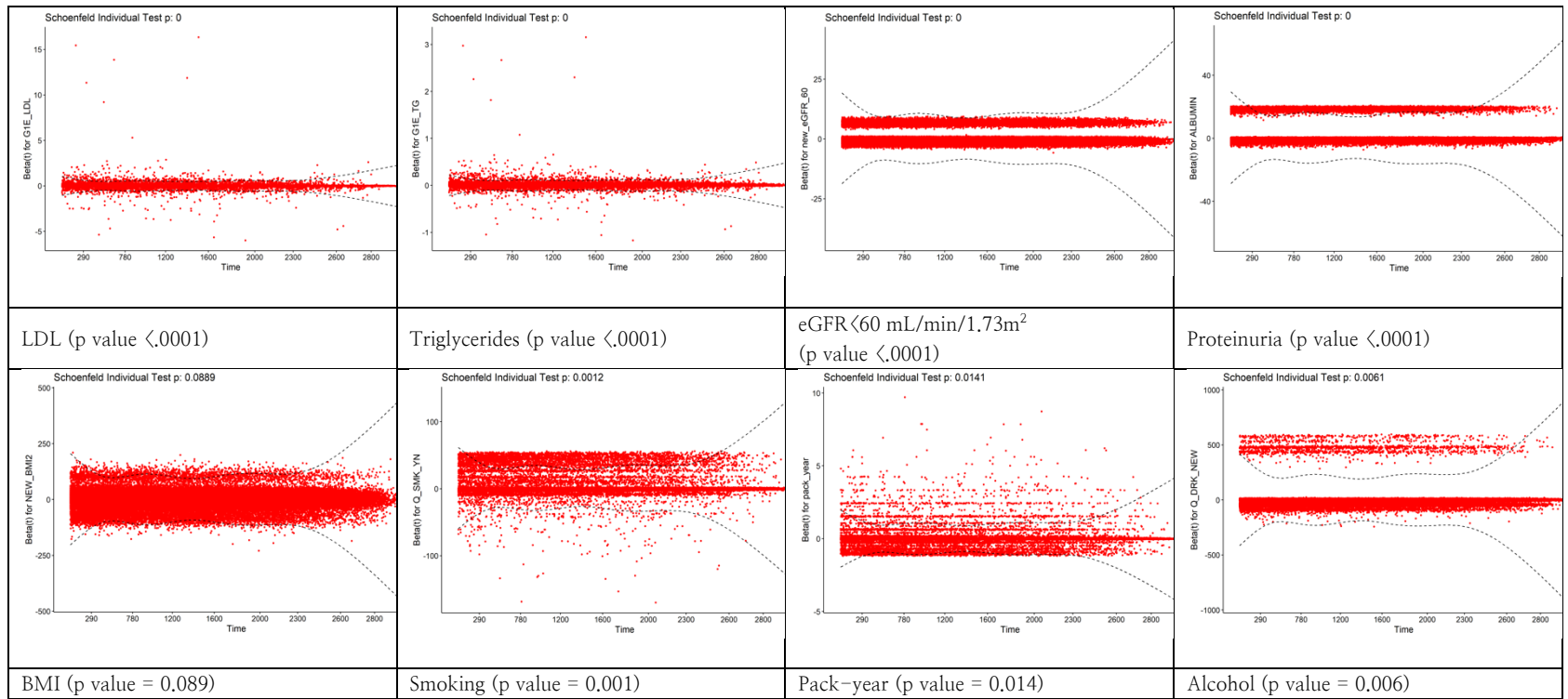


그림 3-1 성별에 따른 (남성) Cox 모형에 대한 시간 대 계수의 Schoenfeld 잔차도





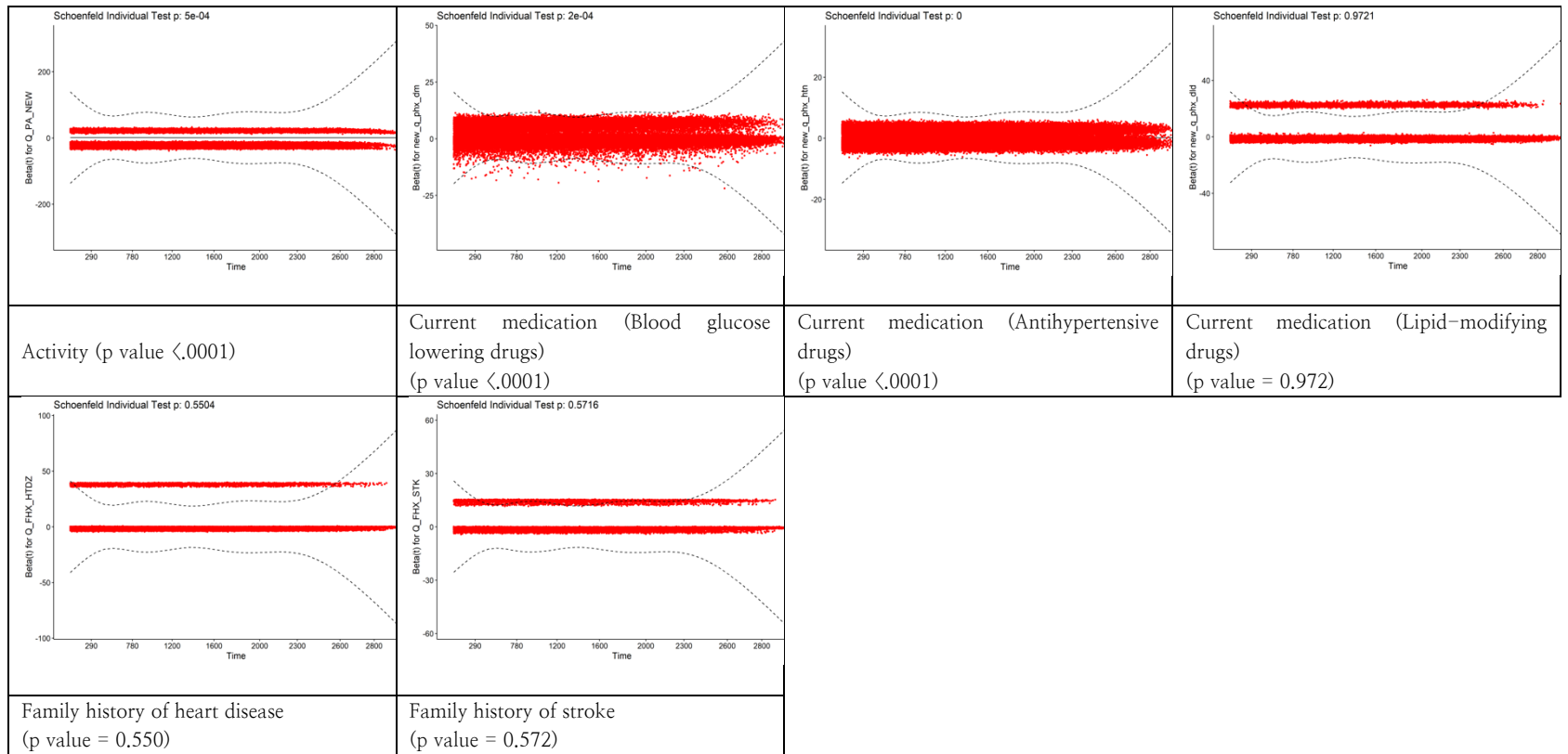


그림 3-2 성별에 따른 (여성) Cox 모형에 대한 시간 대 계수의 Schoenfeld 잔차도

제 2 절 심뇌혈관질환 예측 모형의 성능

표 3-6 Framingham risk score 와 pooled cohort equation (PCE), 개정된 PCE, 심뇌혈관 예측 모형의 AUROC, Harrell's C-index, Brier score 의 10 회 교차 검증의 결과

	AUROC	Harrell's C	Brier score
Men			
Framingham risk score	0.728±0.002	0.715±0.003	0.032±0.003
PCE (whites)	0.730±0.002	0.715±0.003	0.031±0.000
PCE (African-Americans)	0.723±0.002	0.710±0.003	0.026±0.000
Revised PCE	0.748±0.002	0.737±0.002	0.024±0.000
Our model	0.762±0.002	0.749±0.002	0.018±0.000
Women			
Framingham risk score	0.741±0.001	0.723±0.002	0.013±0.000
PCE (whites)	0.753±0.002	0.734±0.003	0.016±0.000
PCE (African-Americans)	0.734±0.001	0.712±0.001	0.025±0.000
Revised PCE	0.808±0.002	0.792±0.002	0.011±0.000
Our model	0.811±0.002	0.795±0.002	0.010±0.000

표 3-6 은 성별 심뇌혈관질환 예측 모형의 성능을 요약한 것이다. 두 모형 모두 모형 판별 능력에서 좋은 성능을 보였다. 평균 Harrell's C 는 남성에서 0.749 ± 0.002 , 여성에서 0.795 ± 0.002 이었다. AUROC 의 교차 검증 결과는 그림 3-3 에 보였다. 개정된 PCE 의 AUROC 와 비교하여, 제안된 모형은 남성에서 0.748 ± 0.002 에서 0.762 ± 0.002 로, 여성에서 0.808 ± 0.002 에서 0.811 ± 0.002 로 상승하였다. 비교를 위해, 백인과 아프리카 계 미국인 모형을 제안된 모형에 적용했을 때 남성의 경우 0.723

± 0.002 에서 0.730 ± 0.002 , 여성에서 0.734 ± 0.753 의 범위를 보였다. Brier score 의 경우 제안된 모형이 가장 좋은 성능을 보여 남성에서 0.018, 여성에서 0.010 이었으며 다른 모형들의 경우 남성에서 0.024 에서 0.032, 여성에서 0.011 에서 0.025 의 범위를 보였다.

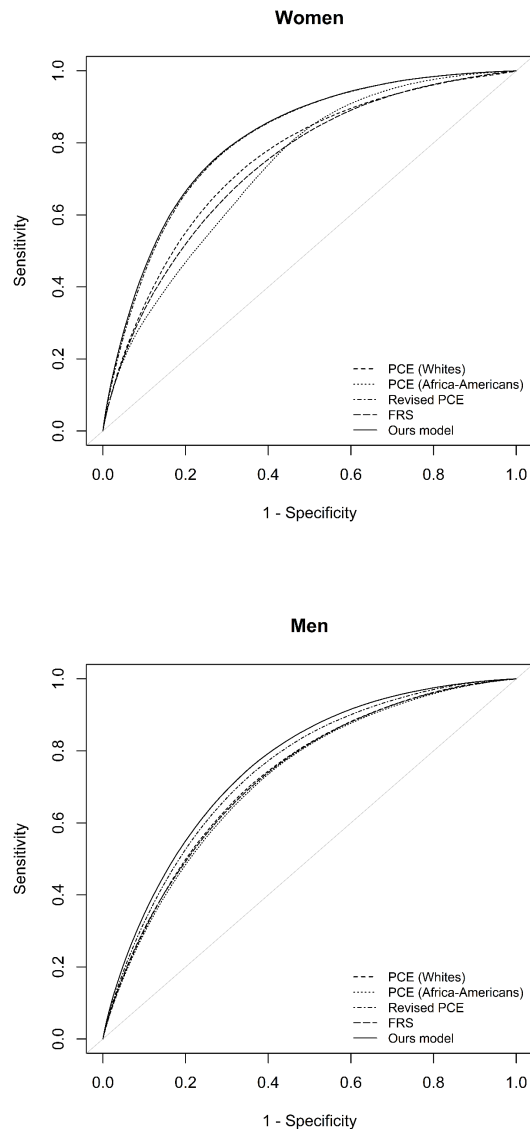


그림 3-3 Framingham risk score (FRS), pooled cohort equation (PCE)와 심뇌혈관질환 예측 모형의 AUROC 비교

제 4 장 심뇌혈관질환 예측 모형의 고찰

분산인공지능을 이용한 연구의 전단계로 종래의 통계 방법론을 응용하여 한국의 건강보험공단의 건강검진 데이터를 통해 예측 모형을 만들었고, 이는 한국인에서 PCE 보다 더 나은 성능을 보였다. 이 모형은 모든 한국인에 대해 검진에서 수행된 생활습관, 가족력과 혈액, 뇨검사를 예측 변수로 포함하였다. 자발적 참여자에 기반하여 비교한 연구 결과이기 때문에 전체 인구의 구성과 CVD 에 연관된 검진과 설문지의 결과가 상이할 수 있다. 그러나 우리의 연구는 대다수의 한국인에서 나온 데이터로 결과를 예측하였고 한국의 전반적인 양상을 잘 표현했다 할 수 있다.

관상동맥의 죽상동맥경화와 뇌혈관계의 동맥경화가 연관이 되어 있다는 것은 잘 알려져 있고 ⁴⁶, 혈압, 흡연, 총 콜레스테롤, 고밀도 지단백, 저밀도 지단백 농도가 죽상동맥경화에 대한 교정가능한 위험요인이라는 것은 충분히 논의되어 왔으며 ⁴⁷, 임상 데이터는 이 요인이 cluster 되어 있고 중첩되어 있는 상호작용을 하고 있다는 것을 보여주고 있다. 그러므로 결과 변수들을 하나로 합치고 관상동맥과 뇌혈관 질환들을 동일선 상에서 보아 전체 위험을 추정하고 위험 요인을 교정하는 것은 정당하다 할 수 있을 것이다.

이들 위험 요인의 상호작용에 의해 예측된 위험은 인구 소그룹의 특성에 의존하여 다양할 수 있다. 이 연구에서 흡연, 음주, BMI 와 CVD 의 발병의 연관성은 성별에 따라 다른 양상을 보였다. 현재 흡연자 중 여성은

남성보다 CVD 위험이 더 높았으며 남성 흡연자에 비해 여성 흡연자에서 관상동맥질환에 대한 위험이 25% 더 높았다는 것이 75 개 코호트의 메타 분석에 의해 분석 되었다⁴⁸. 과거 흡연자의 경우, 이전 연구는 비록 어떤 나이에서던지 금연이 유의하게 CVD 의 이환율 (morbidity)을 줄인다고 되어 있으며 여성은 금연을 한 이후에도 CVD 의 위험이 증가되었다고 한다⁴⁹. 아마 금연 기간이 길수록 더 큰 이득이 있을 것이지만 우리 연구의 추적기간이 금연의 이득을 관찰할만큼 충분히 길지 않았을 가능성이 있다.

남녀 모두에서 CVD 에 대한 BMI 의 다른 양상이 관찰되었다. 남성은 CVD 에 대항하여 보호되는 측면이 있지만 이 관계는 정상 BMI 에 비해 저체중 범주에서 유의하지 않았다. 그러나 여성에서 정상 범주에 비해 저체중 범주에서 가장 위험했다. 게다가 여성은 BMI 가 증가할수록 CVD 에 대항해 보호 받았다. 몇가지 가능한 설명이 있는데⁵⁰, 여성의 경우 metabolically health obese (MHO)의 가능성에 대한 것이다. MHO 의 정의는 연구에 따라 다양하지만, 엄격한 정의에 따르면 대사 증후군의 요소가 없으며 이전 CVD 진단 이력도 없어야 한다. 최근 코호트 연구는 MHO 의 유병률이 비만 연구 대상자 중 남성에서 3.3%에서 32.1%까지, 여성에서 11.4%에서 43.3%까지⁵¹ 일 것으로 추측한다. 더 많은 연구로 성별에 따라 실제로 MHO 에 의해 CVD 가 영향을 받는 것인지 증명되어야 한다.

음주량의 분석에서, 여성은 중등도에서 고위험 음주일수록 CVD 의 증가된 위험이 있었으나 남성은 음주량은 모든 범주에서 보호되었다. 성별에 따른 이런 변이는 생태적 정신사회문화적 감수성에 따라 다를 것으로 사료된다⁵².

1996년부터 2001년까지 30세에서 74세 사이의 한국인 268,315명의 Korean Heart Study (KHS) 데이터를 이용하여 개발한 CHD 위험 모형²⁷과 본 연구의 CVD 예측 모형의 연구 환경을 비교하고, 모형의 설명 변수의 계수 중 총 콜레스테롤, LDL 콜레스테롤, HDL 콜레스테롤 및 중성 지방의 계수를 비교하였다. KHS 모형의 경우 11년, 본 연구의 경우 8년의 추적 관찰을 가지며, 기저 인구학적 및 콜레스테롤 검사를 포함한 주요 변수들의 성별 평균 값은 유사하였다.

KHS 모형은 연령, 범주화된 수축기 혈압 및 총 콜레스테롤, 흡연 상태, 당뇨 여부를 포함한 basic model 에 범주화된 HDL 콜레스테롤을 포함한 model 1, 범주화된 LDL 콜레스테롤을 포함한 model 2, 범주화된 중성지방을 포함한 model 3 을 각각 작성하였으며 그 중 AUROC 가 가장 높았던 model 1 을 최종 모형으로 선택하였다.

이에 비해 CVD 예측 모형은 개인의 생활 습관인 음주량, 운동량, 개인의 질환 및 가족력을 포함하여 더 많은 설명 변수가 포함되어 각 계수의 effect size 가 KHS 모형에 비해 작을 가능성이 높다. 또한, 여성 모형의 경우 KHS 와 CVD 예측 모형 모두 LDL 콜레스테롤이 통계적으로 유의하지 않았다.

CVD 예측 모형은 총 콜레스테롤을 비롯하여 혈액 검사 결과를 모두 연속형 변수로 입력하였고, KHS 모형의 범주화된 변수와 비교를 위해 CVD 예측 모형의 연속형 계수에 KHS 모형의 각 범주의 범위의 중앙값과 CVD 예측 모형의 참조 값의 차를 제곱 항에 넣어 계산하면 KHS 모형과 같이 증가하는 추세를 보인다.

또한, 중성지방의 경우 KHS 모형에서 범주에 따라 위험비가 높아지다가 최종 구간에서 하락하는 뒤집혀진 U 모양을 보이나 CVD 예측 모형의 경우 완만하게 상승하는 추세를 보이는데, 이는 CVD 예측 모형의 경우 연속형으로 모형에 적합하는 과정에서 선형을 취하게 되어 차이가 발생했을 것으로 추정할 수 있다.

다른 국가에서 많은 연구는 CVD 의 위험을 여러 방법론과 다양한 변수들로 예측하였다^{53,54}. 최근, 5 개의 미국의 주요 역학적 연구에 기반한 연구를 토대로 작성된 인종별, 성별 예측 모형인 PCE 는, CVD 의 발생을 나이, 총 콜레스테롤, HDL 콜레스테롤, 수축기 혈압 측정치, 흡연 여부, 당뇨 진단 여부를 이용하여 10 년 이내에 발생할 확률로 예측하였다. 모형은 계속해서 검증되었으며²⁹, 검증 결과는 한국인과 중국인³⁰에 적용되었다. 그러나 PCE 는 KHS 를 이용해 한국인에 적용하였을 때 CVD 위험을 과대추정하였다²⁸. 총 192,605 명의 연구에 따르면 절대적인 10 년 CVD 의 남성에 대한 위험은 백인 모형로 적용했을 때 56.5% 아프리카계 미국인 모형로 적용했을 때 74.1% 과대 추정하였으며, 여성은 27.9% (백인), 29.1% (아프리카계 미국인)을 과소 추정하였다. 이를 기반으로 추가로 한국인에 대한 위험 예측 모형을 개발하였으며, 평균 12.8 년 추적 관찰 기간동안, 12,327 건의 CVD 가 발생 하였고 AUROC 는 남성에서 0.741 여성에서 0.745 를 보였다. 이 연구를 PCE 에 적용했을 때 더 낮은 AUROC, 즉, 0.727 (백인), 0.725 (아프리카계 미국인), 여성에서 0.738 (백인), 0.739 (아프리카계 미국인)을 얻었다. 한국인 남성에서 비록 AUROC 는 두 모형에서 유사했지만 CVD 의 발생은 두 배 넘게 과대 추정되었다. 이런

관찰 결과는 서구 인구에 대해 개발된 모형을 그대로 적용하는 것보다 아시아 인에 대한 특화된 위험 추정의 필요를 강조한다.

우리 연구에서 CVD 의 발생은 남성에서 1.9% 여성에서 1.0%였고 PCE 의 데이터 분석에 대해 사용한 코호트에 의하면 10 년 CVD 발생은 인종과 성별에 따라 1.0%에서 28.5% (남성), 0.0%에서 23.0% (여성) 이었다 ¹⁴. 코호트에서 인종 별 발생을 보면, CVD 의 발생은 비 히스패닉 미국인에 비해서 히스패닉계와 아시아계 미국인에서 더 위험이 낮았지만, 미국 원주민의 경우 높았다. 이를 아시아에 적용하면, 아시아인 또한 다인종 및 다민족으로 구성되어 있으며, 생활습관과 같은 건강 특성도 다를 것이 예상된다 ². 그러므로 CVD 의 위험 또한 인종과 민족에 따라 다를 것이다. 한국인에서 관찰된 것과 같이 아시아 국가에서 관찰된 CVD 의 절대적인 위험이 서구 국가에서 관찰된 위험과 상이하여 아시아만의 예측 모형이 필요하다.

Cox 비례 위험 모형의 개발에서 비례 위험 가정을 탐색하였으며 이는 Cox 비례 위험 모형의 핵심적인 측면 중 하나이다. Schoenfeld 잔차는 시간에 따라 도시 하였고 적합된 곡선의 기울기를 평가하였다. 테스트 세트에서 제공된 p 값에 기반하여 작성된 모형에 사용된 대다수의 변수는 매우 작은 p 값을 보여 비례 위험 가정을 위반하는 것으로 판명되었다. 하지만 이는 chi 제곱 검정이라는 검정 방법이 큰 샘플에서 매우 예민하게 반응하기 때문인 것으로 보인다 ⁵⁵. 이런 이유로 Klein 과 Moeschberger 는 그래프를 이용한 접근을 제안했다 ⁵⁶. 그림 3-1, 그림 3-2 는 적합된 직선이 거의 $y = 0$ 인 수평선에 거의 가깝고 공변량이 비례 위험 가정을 따를

것으로 사료된다. 추가적으로 시간에 대한 가변 효과를 고려한 확장된 Cox 모형을 적합하였으며 이는 비례 위험 가정을 더 검증하기 위해서 였고, 그 결과는 표 3-5 에 제시하였다. 그 결과를 보면 기 작성된 Cox 비례 위험 모형의 계수와 추정된 위험비가 저명한 차이를 보이지 않았다. 이는 설명 변수-시간 상호작용이 없어서 제안된 CVD 예측 모형과 큰 차이가 없었다는 것을 의미한다. 이 결과에 근거하여 제안된 Cox 비례 위험 모형을 확장된 Cox 모형보다 다음과 같은 이유로 사용하기로 하였다. (1) 비례 위험 가정을 맞췄다. (2) 확장된 모형은 예측 검정력과 통계적 효율성을 증진시키지 못한다. (3) 임상 활용 면에서 실용적이지 못하다.

비록 이 연구는 큰 데이터를 사용했지만 몇 가지 제한점을 지닌다. 첫째로, 비록 CVD 의 발생이 ICD-10 과 명세서에 의해 결정되었지만 실제보다 과대하게 보고되었을 가능성을 배제할 수 없다. 둘째로, 비록 검진이 전국민에게 추천되지만 몇몇 치명적인 그룹이 포함되지 않아 대표성이 저하되었을 수 있다. 하지만 연구를 디자인할 때 기저에 CVD 가 없었던 검진 수진자를 대상으로 하였으며 치명적인 그룹의 비대표성에 의한 바이어스는 매우 적을 가능성이 높다. 이런 한계에도 불구하고, 이 연구는 몇가지 이점이 있다. 첫째로 이 예측 모형은 큰 샘플 사이즈 (21,581,796)의 전 한국인의 대부분을 포함하는 전체 건강검진 데이터를 기반해 모형을 제작하였다. 또한, 관심 결과의 발생 전에 표준화된 프로토콜을 이용해 데이터를 수집하였다. 둘째로, PCE 의 변수에 추가해 이 모형은 신체 활동의 강도, 음주량, 흡연과 개인 및 가족력을 포함하였다.

한국의 국가 건강 검진을 이용하여 CVD 를 예측하는 새로운 모형을 만들었으며 한국인에 대해 PCE 보다 성능이 좋았다. 앞으로의 연구는 다른 아시아 국가에 대해 이 모형을 검증하고 임상적 활용을 시험해야 할 것이다.

제 2 부 프라이버시 보전 딥러닝을 이용한

심뇌혈관질환의 예측

제 1 장 서 론

제 1 절 프라이버시 보전 딥러닝 연구의 배경

의료 분야를 포함하여 많은 분야에서 인공지능 (artificial intelligence; 이하 AI)의 출현은 여러 어려운 임무를 혁신하였다⁵⁷⁻⁵⁹. 여러 방면에 응용 가능한 기계학습 모형의 일종인 딥러닝은^{60,61} 최근 의료 인공지능의 최전선에 서 있다⁶²⁻⁶⁵. 의료 분야에서 기대되는 잠재력에도 불구하고⁶⁶⁻⁶⁹, 성능이 좋은 딥러닝 모형의 발전은 충분히 큰 데이터의 확보와 개인정보보호라는 의료 AI 발전과 관련된 두 가지 충돌하는 주제에 의해 제한되어 있다.

요구 조건 중 하나는 질과 양 모두의 측면에서 훌륭한 데이터에 기술적인 접근에 대한 문제이다⁷⁰. 충분히 큰 데이터는 새로이 유입된 데이터에 대한 예측의 일반화를 위한 AI 모형의 훈련과 평가를 수행하기 위해 필요하다^{69,71,72}. 따라서, 다양한 의료 기관 간 연구 협력을 통해 충분한 데이터의 집적이 바이어스 (bias)가 최소화된 의료 AI 모형을 구축하는 데에

필수적인 과정이다⁷³⁻⁷⁵. 그러나, 의료 데이터를 공유하는 것은 강력한 사생활 데이터 보호 제한에 의해 규제되고 있다^{76,77}. 의료 데이터가 환자의 민감한 사생활 정보를 포함하고 있다는 태생적 속성에 의해, 법률로 환자의 사생활을 보호하기 위해 의료 데이터의 집적을 제한하고 있다⁷⁸⁻⁸⁰. 의료 기관의 외부에서 의료 데이터를 공유하려면 이를 극복하기 위한 동기 부여가 필요하며, 경제적, 정치적, 법적, 그리고 윤리적 우려⁸¹⁻⁸³가 완화되어야 하지만 현재 상황에서 의료 AI 발전을 위한 적절한 데이터셋을 구축하는 것은 난제이다.

충분히 큰 데이터에 대한 수요와 개인정보보호 사이에서 프라이버시 보전 딥러닝 (Privacy Preserving Deep Learning; 이하 PPDL)에 대한 연구⁸⁴⁻⁸⁶가 주목을 받고 있다. PPDL은 암호화 (encryption)⁸⁷⁻⁸⁹, 훈련 알고리즘^{85,90}, 차분 개인정보 (differential privacy)^{89,91-94}와 같은 다양한 접근 방법을 응용해 데이터의 개인정보를 보호하면서 AI 모형을 훈련시키는 데 목적이 있다. 연합 학습 (Federated Learning; 이하 FL)⁹⁵⁻⁹⁸은 저명한 PPDL 훈련 알고리즘이며 모형의 매개변수나 기울기를 교환함으로써 모형을 훈련시키고, 각 데이터 원천에서 참여한 여러 기관의 데이터 세트를 보호한다. 영상의학^{91,94,99-102}, 조직학¹⁰³, 전자의무기록 (Electronic Health Record; 이하 EHR)^{93,104}을 이용한 연구를 포함하여 여러 의료 분야의 응용을 통해 종전의 데이터 공유 방법을 이용한 모형 작성과 FL을 이용한 모형 작성을 비교하여 유사한 성능을 보였다.

제 2 절 문제 제기

기존의 AI 연구 및 개발 과정은 각 데이터의 단일 소유자만으로 모형을 제작 시 과적합의 문제가 있을 수 있으며, 데이터의 절대적인 양의 부족으로 유병률이 낮은 질환에 대해 예측이 어려운 문제가 있다. 이를 해결하기 위해 하나의 DB 로 데이터를 수집하는 과정이 반드시 선행되어야 하는 과정이지만 앞서 설명되었듯이 이는 법적, 윤리적인 문제로 인해 현재 상황에서 점차 어려워지고 있다.

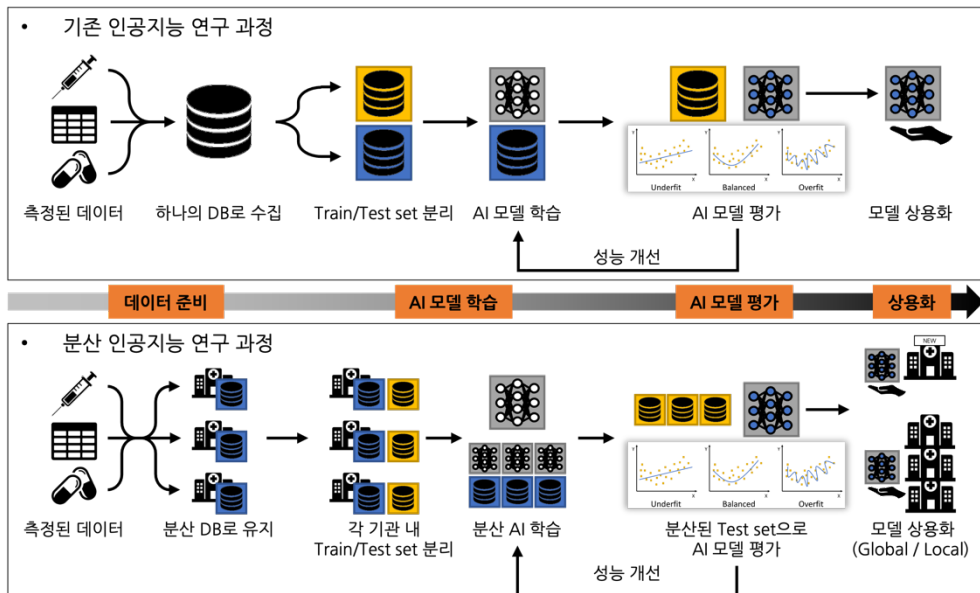


그림 1-1 기존의 AI와 분산 학습의 연구 과정

개인정보보호와 데이터 보안을 충족하는 AI 기술 개발이 필요한 상태이며, 이를 현실화하기 위해 개인정보보호가 가능한 분산 DB로 데이터를 각 데이터의 소유자에게 두고 기계 학습 및 딥러닝을 통한 모형

개발 및 평가 방법, 폐쇄된 네트워크 환경에서도 충분히 개발이 가능한 여러 방법이 제시되고 있다.

이 중 FL이 의료 AI의 개발과 개인정보보호에 대한 기대되는 해법이지만, 실사용에 몇 가지 제한이 있다. FL의 훈련 프로세스 동안, 개인정보에 민감하지 않은 매개변수는 미리 정의된 업데이트나 에포크(epoch) 수마다 동기화된다⁹⁵. 즉, 데이터를 한 곳에 모은 센터 최적화의 관점에서 통신 비용은 상대적으로 낮고, 계산 비용이 지배적이지만 최근에는 GPU를 사용하여 이러한 비용을 낮추는 데 초점을 두고 있다. 대조적으로, 연합 학습을 위한 최적화에서 통신 비용이 지배적으로 일반적으로 1MB/s 이하의 업로드 대역폭으로 제한된다.

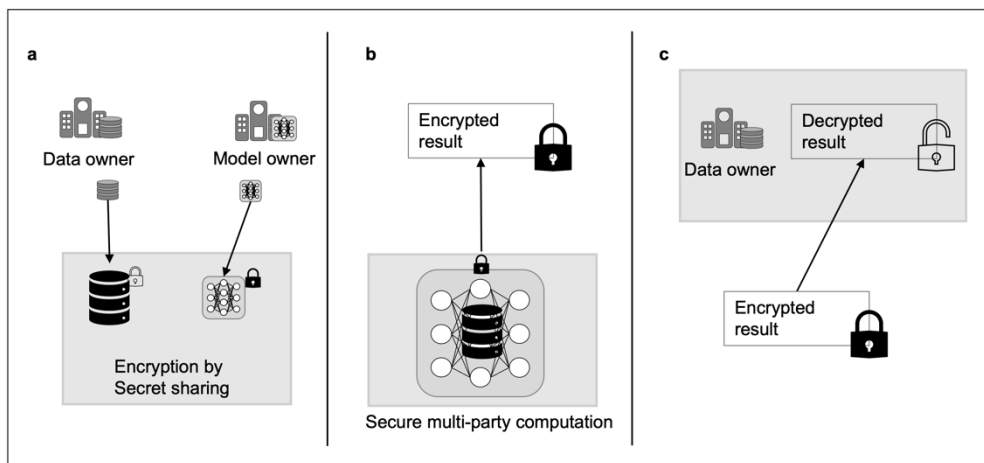


그림 1-2 암호화된 추론 과정과 프레임워크의 개요

데이터 소유자는 인터넷을 통해 모형의 추론 결과를 요청하지만 자신이 보유한 민감한 환자 데이터가 비밀로 유지되기를 희망한다. 마찬가지로, 모형의 소유자는 추론을 서비스로서 제공하지만 모형을 기밀로 유지하고자

한다. (a) 초기에 데이터 소유자와 모형 소유자는 각각 데이터와 모델을 secret sharing을 통하여 암호화한다. 이 프로세스는 데이터/모형을 공유로 분할하는데 의존하고, 그 공유 자체는 어떤 사용 가능한 정보를 포함하지 않으므로 타 데이터 소유자와 공유될 수 있다. (b) 추론은 신경망 추론 절차를 공동으로 계산하여 수행된다. (c) 데이터 소유자는 자신만 해독할 수 있는 암호화된 결과를 받아 추론을 완성한다.

따라서, FL을 이용하여 모형을 고안하는 것은 모형 동기화에 필요한 충분한 통신 비용을 감내할 수 있는 자동화된 연합학습 프레임워크^{87,91,92,105}를 요한다.

그러나, 위에 서술한 프레임워크가 자동으로 외부 서버와 파일을 교환할 때, 보안을 해치고 데이터 유출로 이어질 수 있어 보안 프로토콜 때문에 의료 기관에서 이 프레임워크를 사용하는 것은 어렵다¹⁰⁶.

추가로, 작은 의료 기관은 이런 FL 프레임워크를 지탱할 기술적 하부구조가 충분하지 않을 수 있다¹⁰⁷.

제 3 절 관련 연구 동향

제 1 항 의료 영상에 대한 기관 간 분산 딥러닝 연구

2018 년 Chang 등¹⁰⁸ 은 환자 데이터 공유에 대한 대안으로 딥러닝 모형을 분산 네트워크 방식으로 훈련하기 위한 방법을 제안하였다.

Kaggle 의 총 6 천명의 당뇨 망막병증 영상 데이터를 이용하여 데이터를 한 곳에 모아서 훈련 시키는 centrally hosted 방식, 4 개의 가상 기관으로 나누는 후 각 기관에서 개별적으로 훈련된 모형을 제작 후 출력을 평균화한 앙상블 방식으로 최종 모형을 제작한 ensemble single institution models, 검증 손실의 정체기에 도달할 때까지 단일 기관에서 모형을 훈련한 다음 모형을 다음 기관으로 이전하는 single weight transfer 방식으로 이는 한 번만 각 기관에 전송되었다. 마지막으로 모형을 다음 기관으로 이전하기 전 미리 결정된 가중치 이전 빈도 (epoch) 동안 각 기관에서 모형을 훈련시키는 cyclical weight transfer 방식으로 합성곱신경망 (convolutional neural network)을 이용하여 각 모형의 성능을 정확도로 실험하였다.

그 결과 데이터를 모두 모아 놓은 방식의 경우 78.7%의 테스트셋에 대한 정확도를 보였고, ensemble 방법의 경우 60.0%, single weight transfer 는 68.1%, cyclical weight transfer 는 20 epoch 마다 가중치를 전달했을 때 76.1%의 정확도를 보여 centrally hosted 방식과 대등한 성능을 보였다.

제 2 항 의무 기록 데이터를 이용한 기관 간 분산 딥러닝 연구

Informatics for Biomedical Computing 에서 주최한 Informatics for Integrating Biology and Bedside (i2b2) 2014 비식별화 챌린지의 임상 기록을 이용하여 진행한 실험에서, 저자는 각 기관에 산재된 훈련 세트를 학습 프로세스로 통합하여 collaborative learning 을 실현하고자 하였다¹⁰⁹.

이를 위해 개인정보보호 장치가 없이 중앙에서 데이터를 다루는 nonprotective training 방식, 한 기관에서 모형을 훈련한 후 훈련된 파라미터를 옆 기관으로 전달하여 모형을 개선하는 round robin training 방식, 이는 앞서 서술한 cyclical weight transfer 와 유사하다. 마지막으로 각 기관에서 모형을 선택적으로 확률적 경사하강법을 이용하여 참가한 기관이 로컬 매개변수에 대한 정보가 공유되는 양을 결정할 수 있도록 훈련한 후 모형을 중앙 서버로 보내는 privacy-preserving distributed selective stochastic gradient descent (DSSGD) 방식을 고안하였다.

그 결과 nonprotective training 방식의 F1 값은 0.962, round robin 방식의 값은 0.961, 정보 공개의 양을 조절하여 수행한 DSSGD 방식에서 0.955 와 0.840 를 보여 round robin 방식이 종전의 중앙 집중된 데이터를 이용한 방식과 큰 차이를 보이지 않았다.

제 4 절 연구의 필요성

무거운 통신을 요하는 FL의 대안으로서 전이학습 (transfer learning)에 기반한 탈중앙화된 PDDL 훈련 알고리즘^{110,111}은 현저하게 적은 통신을 요한다. 이 훈련 알고리즘은 논문에 따라 cyclic weight transfer^{85,108,112}, iterative institutional learning^{101,113}, round robin method¹⁰⁹, cyclic transfer learning¹¹⁴ 등의 다양한 이름으로 불리고 있다.

위의 학습 방법은 공통적으로 전이 학습의 개념에 기반하고 의료 기관을 따라 반복적으로 모형 전이가 수행되기 때문에 본 연구에서 용어를 통일하고자 이 훈련 알고리즘의 이름으로 “전이반복학습 (transfer iterative learning; 이하 TIL)”으로 명명하였다. TIL은 이전 연구인 의료 영상 데이터^{101,108,112,113}, 의료 자연어 데이터^{109,114}에서 FL과 동등한 성능을 보이며 FL의 모형 동기화와 비교하여 모형 전이에 대해 훨씬 적은 통신을 요한다. 그러므로 TIL은 수동 전이를 이용하여 비동기로 고안되어 자동화된 프레임워크에 덜 의존할 수 있다.

제 5 절 연구의 내용, 목적 및 의의

이 연구는 EHR 데이터를 사용하여 CVD 예측 모형에 TIL을 포함한 PDDL의 여러 상황을 가정하여 훈련 알고리즘의 성능을 비교하였다. 서울대학교병원의 common data model (CDM)¹¹⁵을 이용하여 40,507명의 당뇨 환자 코호트를 구성하였다. 이 코호트를 k-means clustering을 이용하여 5개의 가상의 기관으로 분리한 후 다양한 기관들의 가상의 상황으로 TIL의 성능을 평가하기 위해 상이한 혹은 동질의 데이터 분포를 지닌 세 가지 다른 실험 환경이 사용되었다. PDDL의 환경 하에서 평가 결과는 TIL을 통한 모형의 타 모형에 대한 경쟁적인 성능으로 표현하였다.

이 연구에서 TIL은 FL에 비해 현저히 적은 통신 비용을 사용하며 FL에 비견할만한 성능을 보여 TIL이 의료 영역에서 충분히 사용할 가치가 높은 PDDL 훈련 알고리즘임을 보이고자 하였다.

이로서 FL보다 통신 비용을 줄이며 전 데이터를 한 곳에 모아 훈련 시킨 것과 유사한 성능을 낼 수 있는 TIL을 이용하여 데이터가 작은 기관도 민감한 데이터의 유출에 대한 우려 없이 큰 병원과 협력하여 개인정보를 보호하며 모형을 제작하여 모형에 대한 일반화를 높일 수 있다..

제 6 절 실험에 사용된 방법론과 성능 평가를 위한 지표

제 1 항 k-means clustering

k-means clustering 은 주어진 데이터를 k 개의 클러스터로 묶는 알고리즘으로, 각 클러스터와 거리 차이의 분산을 최소화하는 방식으로 동작한다. 이는 label 이 달려 있지 않은 입력 데이터에 label 을 달아주는 역할로 사용될 수 있다.

입력 값

- 1) k : 클러스터 수
- 2) D : n 개의 데이터를 포함하는 집합

출력 값: k 개의 클러스터

알고리즘

- 1) 데이터의 집합 D 에서 k 개의 데이터를 임의로 추출하고, 이 데이터들을 각 클러스터의 중심 (centroid)으로 설정한다.
- 2) 집합 D 의 각 데이터들에 대해 k 개의 클러스터 중심과 거리를 각각 구하고, 각 데이터가 어느 중심점과 가장 유사도가 가장 높은 지 알아낸다. 그리고 그렇게 찾아낸 중심점으로 각 데이터를 할당한다.
- 3) 클러스터의 중심점을 다시 계산한다.

4) 각 데이터의 소속 클러스터가 바뀌지 않을 때까지 2), 3)을 반복한다.

본 실험에서 이 방법을 사용하여 전체 코호트를 이질적인 특성을 갖는 5 개의 가상의 기관으로 분할하였다.

제 2 항 Accuracy

Accuracy 는 이분형 분류 테스트가 얼마나 정확하게 label 을 예측하는 지 혹은 오답을 배제하는지를 보기 위한 통계적 지표로 사용된다. 즉, accuracy 는 총 조사한 데이터 중 정확한 예측의 비율 (true positive 혹은 true negative)이다. 이를 계산하기 위한 식은 아래와 같다:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

여기서, *TP*는 True positive; *FP*는 False positive; *TN*는 True negative; *FN*는 False negative 를 이른다.

제 3 항 Relative efficiency

두 절차의 상대적 효율성은 효율성의 비율이지만 종종 이 개념은 주어진 절차와 개념적 최상의 절차를 비교할 때 사용된다. 두 절차의 효율성과 상대적 효율성은 이론적으로 주어진 절차에 사용할 수 있는 표본 크기에 따라 다르지만 종종 점근적 상대 효율성을 주요 비교 지표로 사용하는 것이 가능하다.

본 연구에서는 이 아이디어를 차용하여 전체 데이터셋을 통합한 중앙 딥러닝 모형 (collaborative data sharing; 이하 CDS)의 AUROC 를 분모에 두고 개별 성능 척도를 분자에 두어 상대적 효율성 (relative efficiency; 이하 RE)¹¹⁶이 계산되었다¹⁰¹. RE는 상한인 CDS의 AUROC에 비해 개별 PPDL 알고리즘의 상대적 성능을 보여준다.

RE는 다음과 같은 식을 이용하였다:

$$RE = score / score_{CDS}$$

제 2 장 프라이버시 보전 딥러닝 모형의 방법론

제 1 절 데이터 확보

연구 코호트는 서울대학교병원의 CDM으로부터 구축되었고 이를 데이터로 2형 당뇨병으로 진단된 환자에서 CVD를 예측하고자 하였다. 2형 당뇨병은 ICD-10의 E11 질병 코드를 가지고 있으며 glutamic acid decarboxylase antibody (GAD65 Ab)가 1.0 미만 혹은 C-peptide 농도가 0.9 이상인 경우로 정의하였다.

CVD 발생 여부의 판정은 이분형 분류 (binary classification)로 급성 관상동맥증후군, 뇌경색 혹은 뇌출혈 중 하나의 발생으로 정의하였다. ACS는 ICD-10의 I20 또는 I21의 코드를 가지며 동맥조영술을 포함한 경피관상동맥중재시술, 동맥 내막절제술 (endarterectomy), 또는 관상동맥우회술 (coronary artery bypass graft; CABG)를 시행한 경우, 뇌경색은 I63 코드를 가지며 뇌 자기공명영상의 기록을 가진 경우, 뇌출혈은 I60, I61, 또는 I62 코드를 가지며 뇌 단층촬영영상의 시행 기록이 있는 경우로 정의했다.

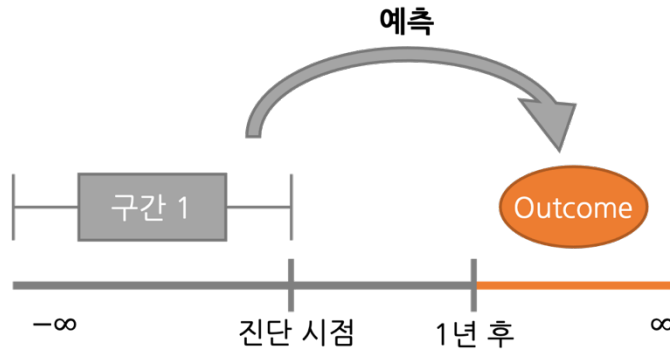


그림 2-1 결과 변수의 정의. 2형 당뇨 진단 시점 이전까지의 데이터를 이용하여 진단 1년 후 이후의 CVD 발병 여부를 예측함

최초 연구에 포함된 인구는 2004년 1월부터 2020년 12월 사이 2형 당뇨병으로 진단된 52,718명이었다. 2형 당뇨병 진단 이전 CVD로 진단된 환자는 제외되었다. CVD는 여러 위험 요인¹⁷에 의해 발생할 수 있으므로 CVD에 대한 2형 당뇨병의 영향을 분석하기 위해 2형 당뇨병 진단의 1년 이내 CVD로 진단된 환자는 제외되었다. 또한 기록이 없는 환자가 제외되어 40,507명의 환자가 최종 코호트로 남았다 (그림 2-2).

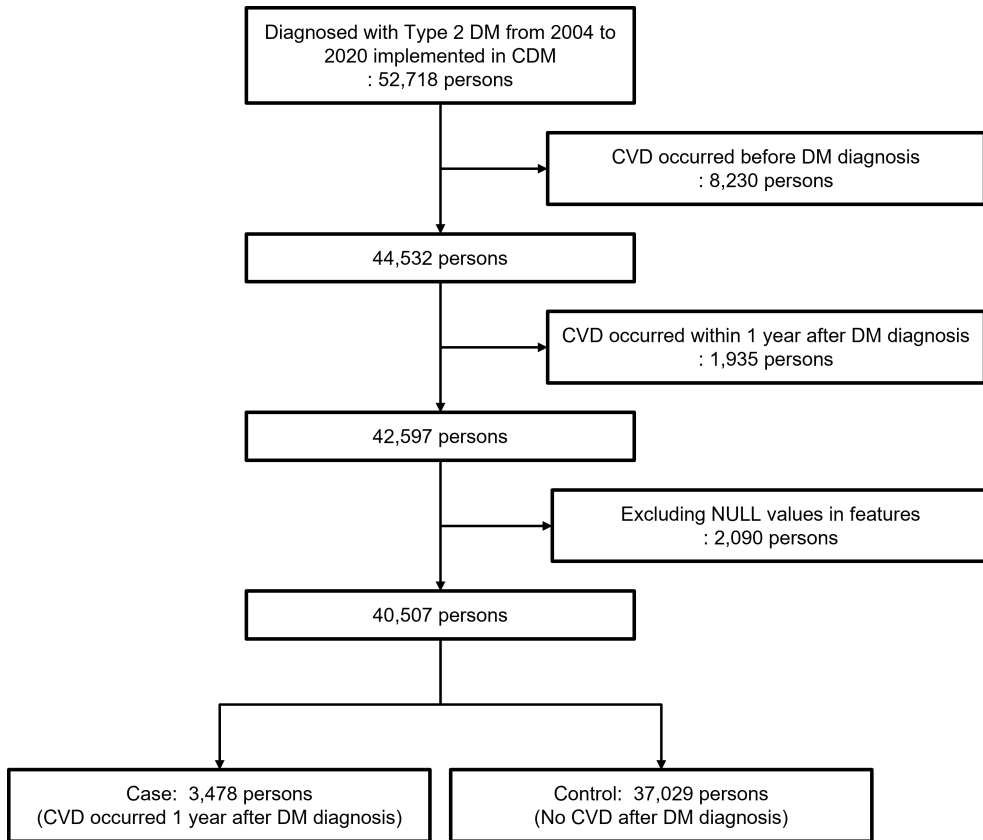


그림 2-2 서울대학교병원의 공통 데이터 모형에서 추출한 2형 당뇨병 환자의 코호트 정의. DM은 2형 당뇨병을 의미하며 CVD는 심뇌혈관질환을 의미함

서울대학교병원의 CDM에서 처음 추출된 변수는 당화혈색소 (hemoglobin A1c), 혈청 포도당 (serum glucose), 혈청 크레아티닌 (serum creatinine) 등의 27개의 측정 (measurement) 변수, 고혈압, 고지혈증, 만성 콩팥병과 같은 5,466개의 상태 (condition) 변수, 당뇨 약제, 고혈압 약제, 고지혈증 약제와 같은 RxNorm¹¹⁸에서 추출한 78개의 약제 복용력이었다. 환자의 3% 이상에서 갖고 있는 다빈도 상태 변수만 사용하였다. 약제 복용력은 환자의 1% 초과된 연구 대상자에서 복용 중인 약품으로

추출하였고 이를 약물의 효과에 근거하여 다섯 개의 범주로 묶었다. 그리고 기록이 없는 변수는 제외하였다. 마지막으로 두 가지 인구학적 특징 변수, 160 개의 상태 변수, 27 개의 측정 변수, 5 개의 약제 복용력 변수 (표 5-2)가 포함되었다. 이 연구는 IRB 2008-013-1145 의 승인 하에 이뤄졌다.

제 2 절 데이터 전처리

주어진 EHR 데이터로 개별 환자에 대해 단일 특성 벡터로 만들기 위해 전처리를 진행하였다. 개별 환자의 기록 내 변수들은 연속형, 이분형, 범주형 변수를 포함한다. 연속형 변수는 측정과 인구학적 변수가 포함되었고, 같은 연속형 변수에 대해 여러 기록이 있을 경우 대푯값으로 마지막 기록을 선택하였으며, 개별 변수에 대해 통계량을 사용하여 표준화하였다. 결측치는 연속형 변수의 경우 0으로 대체되었고, 이는 평균값과 동등하다.

대치된 값이 표준화를 위해 통계량이 재계산 되는 것을 막기 위해 표준화 후 누락된 기록을 대체하였다.

성별, 상태 변수와 약제 복용력과 같은 범주형 변수는 개별 범주에 상응하는 다중 이분형 변수로 전환되었다. 환자에 대한 기록이 존재하면 1의 값으로 없으면 0으로 입력되었다. 그러므로 표준화된 연속형 변수와 1과 0으로 채워진 이분형 변수로 최종 특성 벡터가 구성되었다.

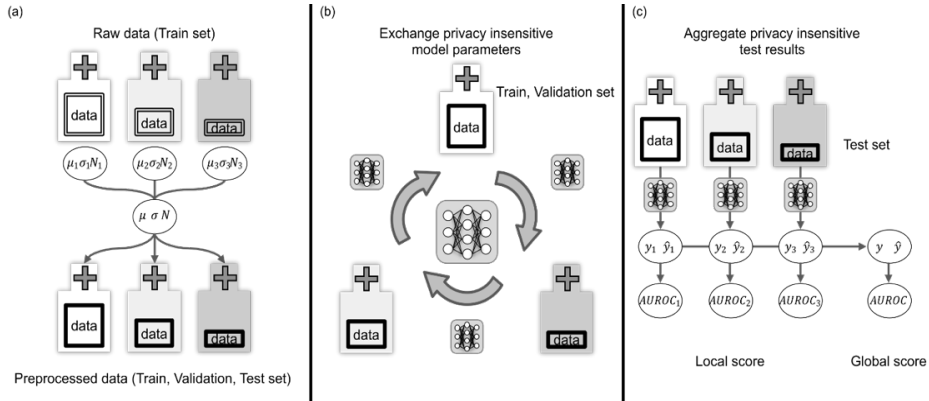


그림 2-3 프라이버시 보전 딥러닝 연구의 과정. 각 기계학습 파이프라인의 단계는 개별 기관에 저장된 개인 데이터를 공유하는 것을 허용하지 않는 것으로 조정되었다. (a) 전처리 단계 지역 통계량은 데이터 공유 없이 전역 통계량을 계산할 수 있도록 집계되었다. 전역 통계량은 전자 의무 기록에서 연속형 변수를 표준화하는 데에 사용되었다. (b) 훈련 단계 각기 다른 프라이버시 보전 딥러닝 훈련 알고리즘이 훈련 단계에서 적용되었다. 비민감 개인정보 매개변수는 민감 개인정보 데이터가 개별 기관 내 안전하게 저장되어 있는 동안 공유된다. (c) 모형 평가 단계 평가될 모형은 모형이 예측할 개별 기관으로 복사된다. 개별 기관 내의 예측 확률과 정답 label 이 지역 점수를 계산하기 위해 사용되고 합산된 예측 확률과 정답 label 은 전역 점수를 계산하기 위해 사용된다.

연속형 변수를 표준화할 때 각 PDDL 알고리즘에 따라 다른 통계량이 적용되었다. 지역 학습 방법의 경우 의료 기관 사이 협업이 없을 것이라고 가정했기 때문에 각 가상 기관의 통계량을 사용하였다. 모든 데이터의 전역 통계량은 CDS, FL, TIL에 사용되었다. 데이터가 각 기관 외부로 공유되지

않기 때문에 전역 통계량은 개별 지역 통계량을 수식을 이용해 합산된 후 도출되었다 (그림 2-3 a).

$$N = \sum_i^k N_i$$

$$\mu = \frac{\sum_i^k N_i \mu_i}{\sum_i^k N_i}$$

$$\sigma = \frac{\sum_i^k N_i (\mu_i^2 + \sigma_i^2)}{N} - \mu^2$$

$$= \frac{\sum_i^k N_i \left(\mu_i^2 + \left(\frac{1}{N_i} \sum_j^{N_i} x_{i,j}^2 - \mu_i^2 \right) \right)}{N} - \mu^2$$

$$= \frac{\sum_i^k \sum_j^{N_i} x_{i,j}^2}{N} - \mu^2$$

$$= E(X^2) - E(X)^2$$

수식에서 k 는 기관의 수, 아래 첨자 i 는 기관 번호, $x_{i,j}$ 는 i 번째 기관의 j 번째 데이터를 의미한다. μ 는 평균, σ 는 표준 편차, 그리고 N 은 데이터의 수를 의미한다. μ , σ , N 은 전역 통계량을 의미하며, μ_i , σ_i , N_i 는 기관 i 의 지역 통계량을 의미한다.

제 3 절 가상의 이질적 의료기관 생성

PPDL의 환경에서 훈련 알고리즘을 실험하기 위해 다중의 기관 데이터가 필요하다. 전체 데이터셋에 k-means clustering을 적용하여 전처리된 데이터를 이질적인 데이터 분포로 5개의 가상 기관으로 나눴다 (그림 5-2). 데이터를 클러스터링 할 때 전역 통계량을 사용하여 연속형 변수를 표준화하였다.

훈련, 검증, 테스트셋은 기관들이 분리된 후 개별 가상 기관 내에서 분화-분리되었다.

마지막으로 지역 통계량은 개별 가상 기관 내 훈련 셋을 이용해 다시 계산되었고 훈련 셋의 전역 통계량은 앞 절의 수식을 이용해 재계산되었다. 재계산된 훈련 셋의 지역 통계량으로부터 도출된 전역 통계량으로 훈련, 검증, 테스트 데이터를 표준화하였다.

이 절차는 개별 가상 기관 간 데이터 분포가 이질적이지만, 개별 가상 기관 내 훈련, 검증, 테스트 셋의 데이터 분포는 동질하게 유지되도록 할 수 있다. 게다가, 훈련 데이터의 지역 통계량을 검증과 테스트 셋을 표준화하는 데에 사용 하는 것은 기계 학습 파이프라인을 실세계 환경에 적용 가능성을 염두에 둔 것이기도 하다.

가상 기관 간 통계적 차이를 비교하기 위해 이분형 변수에 대해 Pearson 카이 제곱 검정을, 연속형 변수에 대해 일방향 ANOVA를 수행하였다.

제 4 절 훈련 알고리즘과 평가

제 1 항 Federated learning

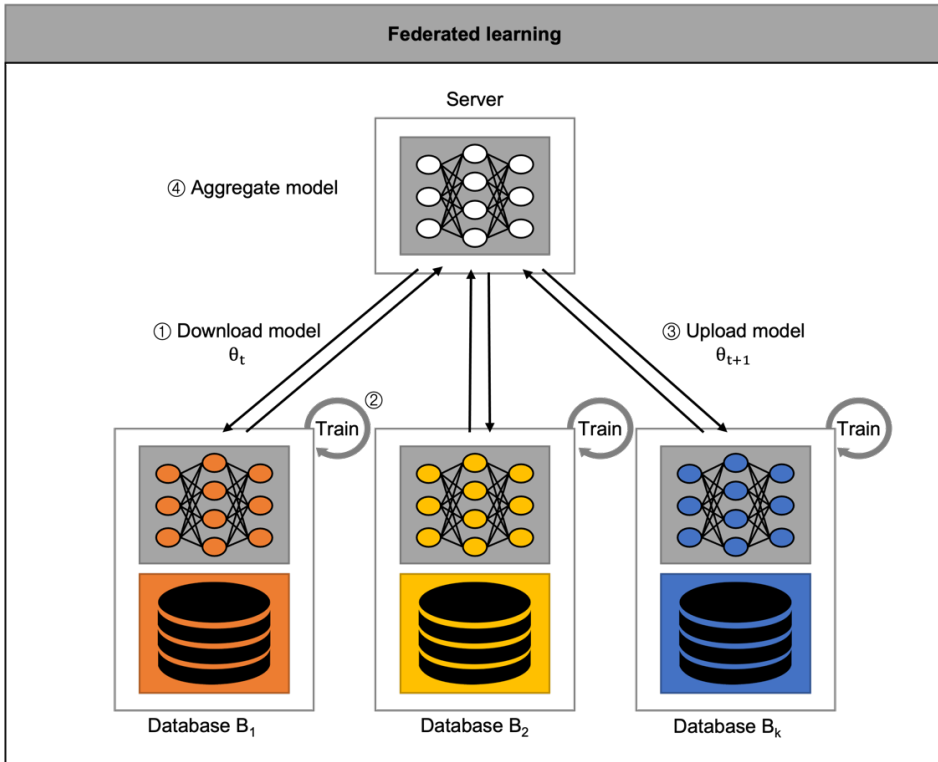


그림 2-4 Federated learning 의 학습 과정

- 1) 훈련의 시작에서 중앙 서버는 훈련되지 않은 모형을 훈련을 위해 병원 혹은 데이터 소유자 같은 각 계산 노드, 이 실험에서 각 개별 가상 기관으로 전송한다.
- 2) 수렴이 될 때까지, 모형은 각 가상 기관에서 지역적으로 각자 훈련된다.

- 3) 정해진 하이퍼파라미터에 따라 모형은 암호화되어 집적된다. 집적 과정은 가상 기관의 데이터 소유자 간에만 일어난다. 업데이트된 모형은 훈련의 다른 라운드를 위해 재분배 된다.
- 4) 마지막 순환 후, 중앙 모형은 전체 학습된 암호화된 집적 모형이 되고 유추에 사용될 수 있다.

제 2 항 Federated learning 모형의 업데이트에 사용된 알고리즘

최근 많은 딥러닝 모형의 구조는 손실 함수를 채택하여 간단한 기울기 기반 방법에 의한 최적화에 근거를 두고 구축되고 있어 확률적 경사 하강법 (Stochastic gradient descent; 이하 SGD)으로 알고리즘의 구성을 시작하는 것이 바람직하다.

SGD 는 연합 최적화 문제에 바로 적용될 수 있다. 여기서 단일 배치 기울기 계산 (예를 들어 무작위로 선택된 단일 가상 기관)이 통신 라운드 당 수행된다. 이 방법은 계산상 효율적이지만, 좋은 모형을 생성하려면 매우 많은 훈련 라운드가 필요하다.

FL 의 실제 데이터에 적용하기 위해 각 가상 기관의 개별 라운드에서 C - fraction 을 선택하고 이 가상 기관에 할당되어 있는 모든 데이터에 대한 기울기를 계산하였다. 그러므로 C 값은 전역 배치 크기를 조절하며, $C = 1$ 은 전체 배치의 기울기 감쇠에 해당한다. 이를 FederatedSGD (FedSGD) 알고리즘이라 하였다¹¹⁹.

$C = 1$ 이고 고정된 학습률 η 의 전형적인 FedSGD 의 과정은 개별 가상 기관 k 가 $g_k = \nabla F_k(w_t)$ 를 계산하며, 현재 모형 w_t 에서 지역 데이터의 평균 기울기를 계산하며, 중앙 서버는 이 기울기를 집적하여 업데이트된 $w_{t+1} \leftarrow w_t - \eta \sum_{k=1}^K \frac{n_k}{n} g_k$ 의 과정을 통해 모형이 업데이트 된다.

즉, 개별 가상 기관은 가상 기관의 데이터를 사용하여 현재 모형에 대한 단계의 경사 하강을 수행하며 서버는 이후 결과 모형의 가중 평균을 갖게 된다.

알고리즘이 이 방식으로 작성되면 평균화 단계 전에 지역 업데이트 ($w_k \leftarrow w_k - \eta \nabla F_k(w_k)$)를 여러 번 반복하여 각 개별 가상 기관에 더 많은 계산을 수행할 수 있다. 이를 FederatedAveraging (FedAvg) 알고리즘이라 한다.

계산의 총량은 세 개의 매개변수로 결정된다. C 는 개별 라운드에서 계산을 수행하는 개별 가상 기관의 비율; E 는 개별 라운드에서 지역 데이터셋에 대해 수행하는 훈련 통과 수; B 는 개별 가상 기관의 업데이트에 사용된 지역 미니배치의 크기이다⁹⁵. 이를 의사 코드로 나타내면 아래와 같다.

FederatedAveraging (FedAvg) 알고리즘. K 가상 기관은 k 로 인덱스 되었다; B 는 지역 미니 배치 크기, E 는 지역 epoch 수, η 는 학습률을 의미한다.

Server executes:

```
initialize  $w_0$ 
for each round  $t = 1, 2, \dots$  do
     $m \leftarrow \max(C \cdot K, 1)$ 
     $S_t \leftarrow$  (random set of  $m$  clients)
    for each client  $k \in S_t$  in parallel do
         $w_{t+1}^k \leftarrow \text{ClientUpdate}(k, w_t)$ 
     $w_{t+1} \leftarrow \sum_{k=1}^K \frac{n_k}{n} w_{t+1}^k$ 
```

ClientUpdate(k, w): //Run on client k

```
 $\mathcal{B} \leftarrow$  (split  $\mathcal{P}_k$  into batches of size  $B$ )
for each local epoch  $i$  from 1 to  $E$  do
    for batch  $b \in \mathcal{B}$  do
         $w \leftarrow w - \eta \nabla \ell(w; b)$ 
return  $w$  to server
```

제 3 항 Transfer iterative learning

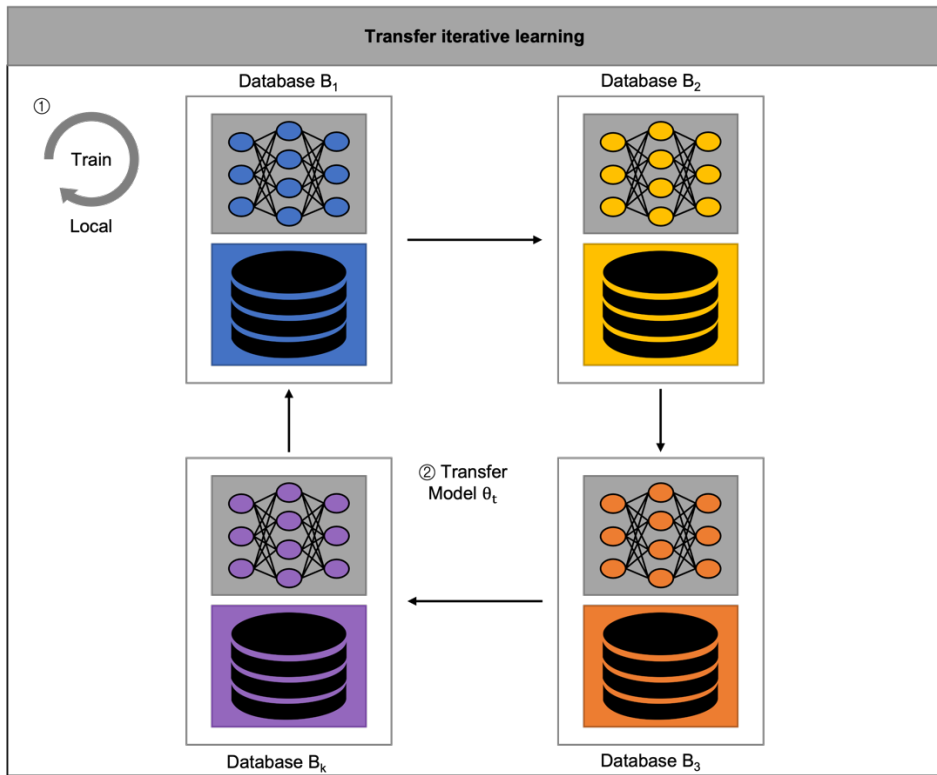


그림 2-5 Transfer iterative learning 의 학습 과정

- 1) 첫 개별 가상 기관부터 초기화 되는 것으로 TIL 의 훈련 과정은 시작 된다.
- 2) 첫 개별 가상에서 훈련이 완료된 후 연속적으로 모형은 반복하여 가상 기관 간 모형 매개변수를 전이하여 개별 기관 데이터를 훈련한다.
- 3) 모든 기관을 한 회 순회한 것을 단일 라운드라고 하며 통신 비용은 그러므로 기관의 수와 동일하다.

제 4 항 전역 및 지역 성능의 평가

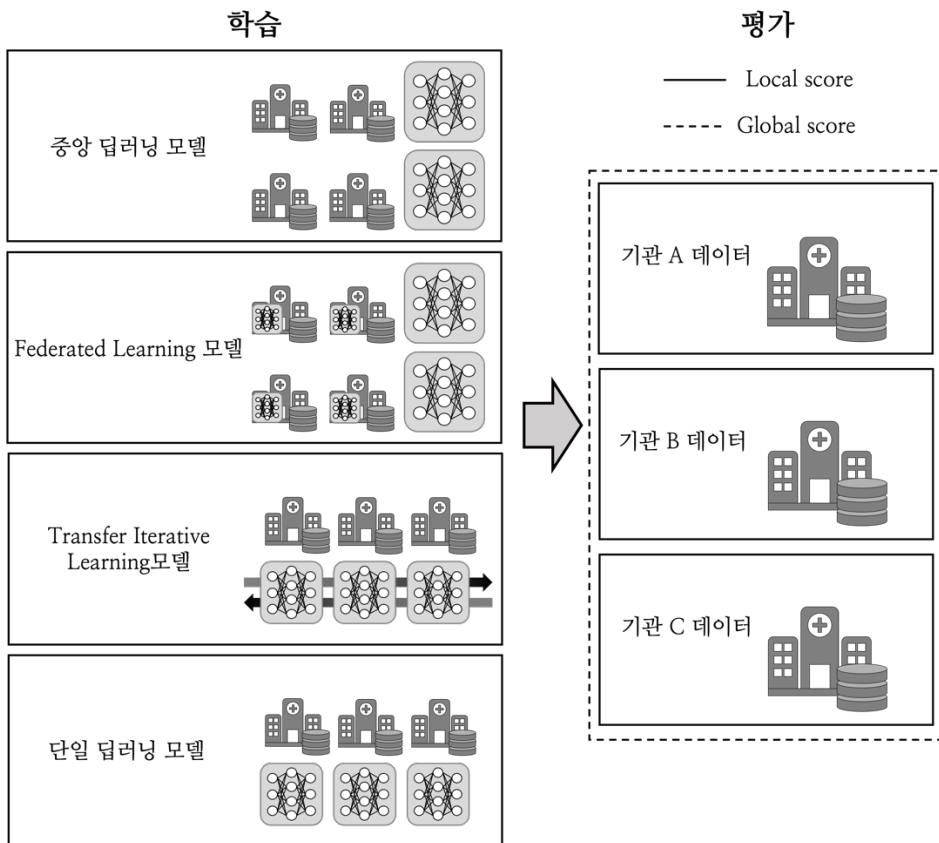


그림 2-6 전역 성능 점수와 지역 성능 점수

- 1) 지역 성능 점수 (local score): 개별 가상 기관의 데이터를 이용하여 모형을 작성 후 개별 기관의 테스트셋에 모형의 성능을 비교
- 2) 전역 성능 점수 (global score): 개별 가상 기관의 데이터를 이용하여 모형을 작성 후 모든 가상 기관의 테스트셋을 사용하여 모형의 성능을 측정

제 5 절 훈련 알고리즘의 상세

두 가지 형태의 PPDL 훈련 알고리즘을 사용하였다 (그림 5-1). 첫 훈련 알고리즘은 TIL이었고 TIL로 모형을 훈련할 때, 딥러닝 모형은 첫 가상 기관부터 초기화 되었다. 첫번째 가상 기관에서 훈련된 모형은 연속적, 반복적으로 이웃으로 가상 기관 간 모형 매개변수를 전이하여 개별 가상 기관의 데이터를 훈련한다. 이 학습 방법은 전이 학습과 유사하다^{110,111}. 모든 가상 기관 간 반복 전이는 단일 라운드를 의미한다. 개별 가상 기관에서 훈련은 비동기적으로 수행되고, 모형 전이로부터 발생한 통신 비용은 개별 라운드에 대한 가상 기관의 수와 동일하다.

각 가상 기관으로부터 TIL 모형은 조기 종료의 적용과 분리된 가상 기관의 최선 모형을 유지함으로써 생성된다. 타 가상 기관의 검증 데이터를 사용할 수 없기 때문에, 조기 종료는 가상 기관 자신의 검증 데이터를 사용하여 수행되었다. 가상 기관의 검증 데이터에 일반화하는 것이 전역 데이터 분포의 일반화에 대해 필요한 상황이기 때문에 가상 기관의 검증 데이터가 훈련 데이터 보다 더 좋게 일반화 되는 것이 더 요망된다.

가상 기관에서 나온 마지막 모형으로서 선택된 검증 점수를 비교함으로써 가상 기관이 모형을 한 라운드 이상 훈련할 때 최선의 모형이 가상 기관에 유지가 된다. 최선의 가상 기관 모형보다 더 많은 데이터로 훈련된 최신 훈련 모형은 일반화를 확신하기 위해 다음 가상 기관으로 전이된다. 그러므로, TIL로 모형을 훈련하면 개별 기관으로부터 다중의 모형이 생성된다 (그림 3-3).

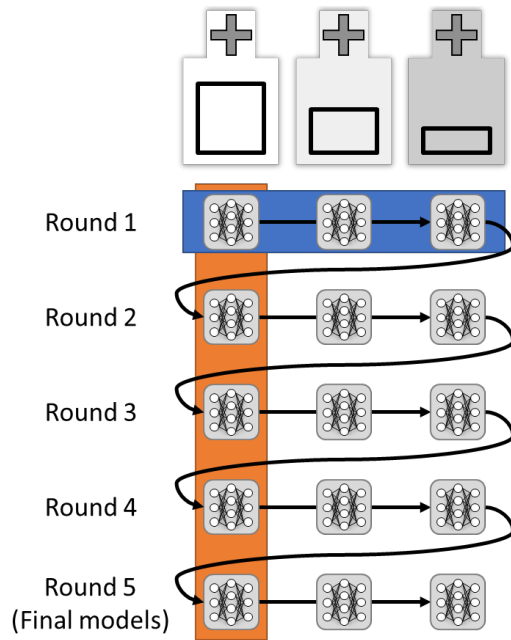


그림 2-7 TIL 의 내림차순 다중 라운드 훈련 과정. 이 예시에서 training 중 총 3개의 기관 * 5 라운드의 15개의 모형이 생성됨. 가로 방향 강조: 각 가상 기관 별 데이터의 이질성으로 개별 비교하기 보다 전역 점수로 비교, 세로 방향 강조: 라운드 수가 많을 수록 성능이 좋아질 것이 자명하므로 동일 가상 기관 간 다른 라운드의 비교는 시행하지 않음.

다음 PPDL 알고리즘은 FL로 같은 모형이, 개별 가상 기관의 데이터로 모형을 훈련하여 개별 가상 기관으로 복사되는 학습 방법을 일컫는다. 개별 라운드로 미리 정의된 기간 동안 모형이 훈련된 후 모형 매개변수는 동기화를 위해 집적된다. 모형 분포와 집적으로부터 통신 비용은 업로드와 다운로드를 포함하기에 최소 가상 기관의 수의 두 배이며, 훈련 공정은 동기화되어 수행된다. 다중 연합 라운드는 모형이 훈련될 때까지 수행된다.

모형 매개변수가 동기화될 때 개별 가상 기관의 데이터 크기에 따라 가중 평균⁹⁵ (FedAvg 알고리즘) 된다. FL로 모형을 훈련할 때, 조기 종료는 모든 가상 기관으로부터 나온 검증 데이터에 대해 모형 성능을 평가함으로써 조기 종료가 수행된다. 테스트 공정은 개인정보보호 방식으로 이뤄졌다 (그림 3-2 c). FL로 모형을 훈련하면 단일 서버 모형 형태로 모형이 하나가 출력 된다.

PPDL 환경에서 훈련 알고리즘의 성능을 비교하기 위해, 두 기저 훈련 알고리즘이 시험되었다 (그림 5-1). 첫 기저 훈련 알고리즘은 중앙 딥러닝 모형인 CDS 이며, 이는 모든 가상 기관으로부터 데이터가 집적되어 단일 데이터셋을 형성하는 종래의 기계 학습 환경을 일컫는다. 모든 가상 기관에서 온 데이터가 사용 가능하기 때문에 CDS 는 PPDL 환경 하에서 모형 성능의 상한으로 생각할 수 있다. 조기 종료는 집적된 검증 데이터에 대해 평가함으로써 수행되었다. CDS 방법은 단일 서버 모형 하나를 출력한다.

두 번째 기저 훈련 알고리즘은 지역 학습으로 개별 가상 기관이 서로 협력하지 않고 가상 기관 자신의 데이터만 이용하여 분리된 모형을 훈련하는 것을 말한다. 그러므로 지역 학습은 PPDL 훈련 알고리즘이 반드시 우월한 성능을 보여야 하는 하한으로 상정할 수 있다. 개별 기관의 모형을 훈련할 때 조기 종료는 가상 기관 자신의 검증 데이터만을 이용해 수행되었고 개별 가상 기관 마다 모형을 생성하므로 가상 기관 수 만큼 모형이 생성된다.

제 6 절 실험 설계

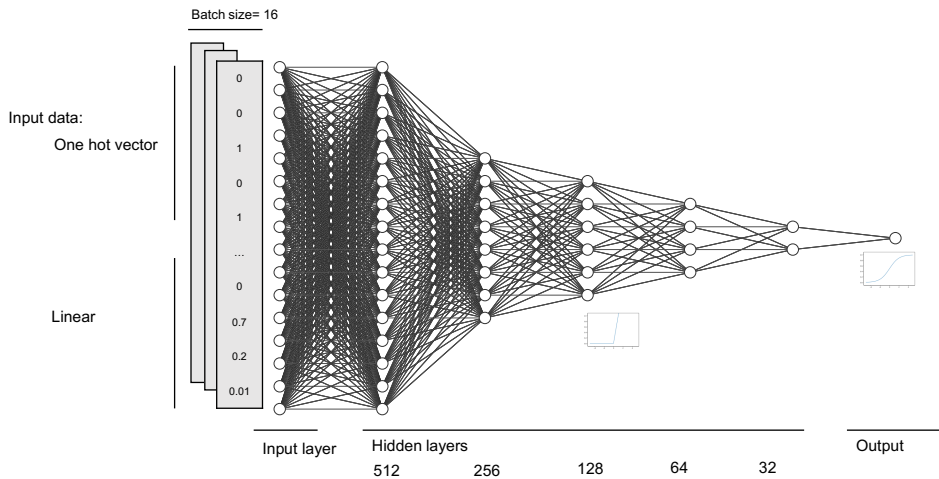


그림 2-8 실험에 사용된 딥러닝 모형의 모식도

hyperparameters는 grid search 결과를 분석한 후 선정되었고, 성능과 안정성을 고려하였다. 모형은 전체 연결된 신경망으로 (fully connected neural network) 각각 512, 256, 128, 64, 32개의 뉴런으로 이루어진 5개의 은닉층으로 구성되었으며 단일 출력 뉴런으로 이루어 졌다. 정류 선형 단위 (rectified linear unit - ReLU) 함수가 각 은닉층에 적용되었고 출력층에는 시그모이드 함수가 적용되었다. 배치 크기는 16, 학습률은 0.0001, 가중치 감쇠는 0.001이 사용되었으며 Adam optimizer가 모형 매개변수를 최적화하기 위해 사용되었다. State 초기화가 optimizer state을 전이하는 것과 비교했을 때 약간의 성능 향상을 보였기 때문에 optimizer state는 FL과 TIL에서 매 라운드마다 초기화 되었다¹¹³.

훈련, 검증, 테스트셋은 각 0.7, 0.15, 0.15의 비율로 개별 가상 기관

내에서 무작위로 분리되었지만 결과 비율은 보존하였다. 각 환경은 모형 매개변수의 다른 초기화를 적용하여 5번씩 테스트 되었다. CVD 미발생군과 발생군 사이 모형 예측의 bias를 균형 잡기 위해, 발생군의 데이터는 미발생군의 수와 유사하게 업 샘플 되었다. 이분형 교차 엔트로피 (binary cross-entropy) 함수가 손실 함수로 사용되었고, 분류 가중치는 발생군과 미발생군 사이 예측의 균형을 위해 업 샘플 한 후 훈련 데이터 내에서 정답 label의 수에 반비례를 적용하였다. 다른 분류 가중치가 FL, TIL, 지역 학습에 대해 각 기관에 사용되었다. 같은 hyperparameter가 모든 실험에 대해 사용되었다.

PPDL 훈련 알고리즘을 위한 특정 hyperparameter가 유사한 학습 상태를 산출하기 위해 선택되었다. 모형은 CDS, 지역 학습, TIL을 위해 100 epoch를 훈련하였다. 개별 가상 기관의 개별 모형은 지역 학습과 TIL을 위해 100 epoch, 단일 서버 모형은 CDS를 위해 100 epoch를 훈련하였다.

FL은 1개의 가상 기관에서 자신의 데이터로 1 epoch의 훈련을 마치고 생성된 모형을 중앙 서버로 업로드, 중앙 서버에서 업데이트, 그 업데이트된 모형의 다운로드의 순환을 1 라운드로 할 수 있다. 그러므로 FL은 개별 라운드에서 한 기관의 epoch로 100 라운드의 동기화로 훈련되었다. 조기 종료는 검증 데이터에 대한 AUROC에 따라 수행되었고 검증 데이터에 대한 평가는 CDS, 지역 학습, TIL에 대해 20번 업데이트마다 수행되었다.

TIL은 네 가지 버전으로 훈련하였는데, 단일 라운드와 10 라운드로 훈련, 데이터 크기의 순서에 따라 내림차순과 오름차순으로 방문하며 훈련하였다.

제 7 절 기관 설정

가상의 기관은 혼합 기관 환경으로서 사용되었다. 혼합 기관 환경 중 k-means clustering 에 의해 이질적으로 생성된 분포가 이질적인 가상 기관의 집합이 아닌 최대한 동질적인 분포를 지닌 혼합 동질 기관 환경을 추가로 생성하였다.

작은 기관 환경은 가장 작은 데이터를 가진 기관 (기관 5 의 866 개)의 크기에 맞춰 각 기관의 데이터를 다운 샘플 하여 생성되었다. 작은 기관 환경은 작은 기관에 따른 TIL 의 기능을 평가하기 위해 제작되었다. 데이터는 10 가지 다른 random seed 로 다운 샘플 되었고, 결과 비율은 보존되었다. 모든 가상 기관은 작은 기관 환경에서 같은 데이터 총량을 지닌다.

모든 훈련, 검증, 테스트 세트는 데이터가 다운 샘플 한 후 층화 하여 분리되었고, 모든 훈련 알고리즘은 혼합, 작은 기관 환경 모두에서 동일한 초매개변수를 이용하여 테스트되었다.

작은 기관 환경의 TIL 모형은 기관의 데이터 크기가 모두 같기에 항상 기관 1 에서 5 의 순서로 훈련하였다.

제 8 절 모형의 평가

두 가지 형태의 평가 척도가 사용되었다. 우선, 모형의 성능 점수는 이분형 분류 평가를 사용하여 측정되었고, AUROC 가 정답 label y 와 모형 예측 확률 \hat{y} 으로부터 계산되었다. 그 다음, 이분 모형의 예측 결과는 특이도가 0.75 보다 큰 점을 모형 예측 확률 기준으로 하여 계산되었다. 정확도, f1 점수, 민감도, 특이도 역시 계산되었다.

두번째로 CDS 의 성능에 따라 개별 성능 척도의 상대적 효율성 (relative efficiency; 이하 RE)¹¹⁶ 가 계산되었다¹⁰¹. RE 는 상한에 비해 개별 PPDL 알고리즘의 상대적 성능을 보여준다.

$$\text{Relative Efficiency} = \text{score}/\text{score}_{CDS}$$

추가로, 이분 분류 척도와 그 RE 는 PPDL 환경 하의 두 가지 형태의 점수를 사용하여 평가되었다 (그림 3-2 c). 첫번째 형태는 지역 성능 점수로 이는 개별 기관으로부터 데이터를 사용하여 측정된 성능 척도를 의미한다.

다음 형태는 전역 성능 점수로 이는 모든 기관으로부터 데이터를 사용하여 측정된 같은 척도를 의미한다. 테스트 데이터는 각 기관의 외부로 공유될 수 없기 때문에 정답 label y 와 모형 예측 확률 \hat{y} 이 전역 성능 점수를 계산하기 위해 각 기관의 외부로 공유되었다. 입력 데이터인 x 와 달리, y 와 \hat{y} 이 개인정보에 민감한 정보를 담지 않았으므로, 기관 외부로 공유될 수 있었다.

전역 성능 점수는 AUROC 를 제외하고, Sheller 등¹¹³이 보고한 대로 가중 평균 점수와 동등하다. 전역 AUROC 점수는 지역 AUROC 점수의 가중 합과 다르며 모든 데이터에 걸쳐 예측 확률을 요한다.

제 3 장 프라이버시 보전 딥러닝 모형의 결과

제 1 절 프라이버시 보전 딥러닝의 기계학습 파이프라인

개인정보 데이터를 안전하게 보호해야 하는 개인정보보호 환경에서 기계 학습 파이프라인은 크게 나누면, 전처리, 훈련, 평가 단계에서 조정이 필요하다 (그림 2-3).

전처리 단계에서 연속형 변수의 전역 통계량은 표준화가 필요하지만 전역 통계량을 계산하기 위해 외부로 반출되어 합산할 수 없다. 그러므로 개별 가상 기관의 지역 통계량이 전역 통계량을 간접적으로 도출하기 위해 집계되었다 (그림 2-3 a). 지역 통계량은 민감한 개인정보가 포함된 것이 아니기 때문에, 이는 가상 기관 밖으로 공유가 가능하다.

개별 가상 기관 내에서 전처리를 한 후 딥러닝 모형은 PPDL 훈련 알고리즘을 통해 훈련되었다 (그림 2-3 b, 그림 3-1 b 과 그림 3-1 d). 각 훈련 알고리즘이 각자 알고리즘에 따라 작동하는 동안, 각 가상 기관 내에서 모든 훈련 데이터를 보호하고 있으며 개인정보에 민감하지 않은 모형의 매개변수를 공유할 뿐이다.

모형이 훈련을 끝냈을 때, 그 성능을 평가하기 위해 개별 가상 기관의 테스트 데이터에 대해 평가하였다 (그림 2-3 c). 개별 가상 기관에 훈련된

모형을 복사하고 개별 가상 기관 내 테스트 데이터를 예측함으로써 평가가 수행된다.

성능 점수는 개별 가상 기관 내에서 계산되며 이를 “지역 성능 점수”로 부른다. 이분형 분류에서 정답 label y 와 모형 예측 확률 \hat{y} 은 민감한 개인정보를 포함하지 않으므로 y 와 \hat{y} 은 각 기관으로부터 합산되어 “전역 성능 점수”라고 부르는 모든 데이터에 대한 성능 점수를 계산할 수 있다. 전역 성능 점수는 모든 기관에서부터 집적된 데이터를 사용한 모형을 평가한 것과 동등한 성능 점수이다.

제 2 절 프라이버시 보전 딥러닝에서 훈련 알고리즘

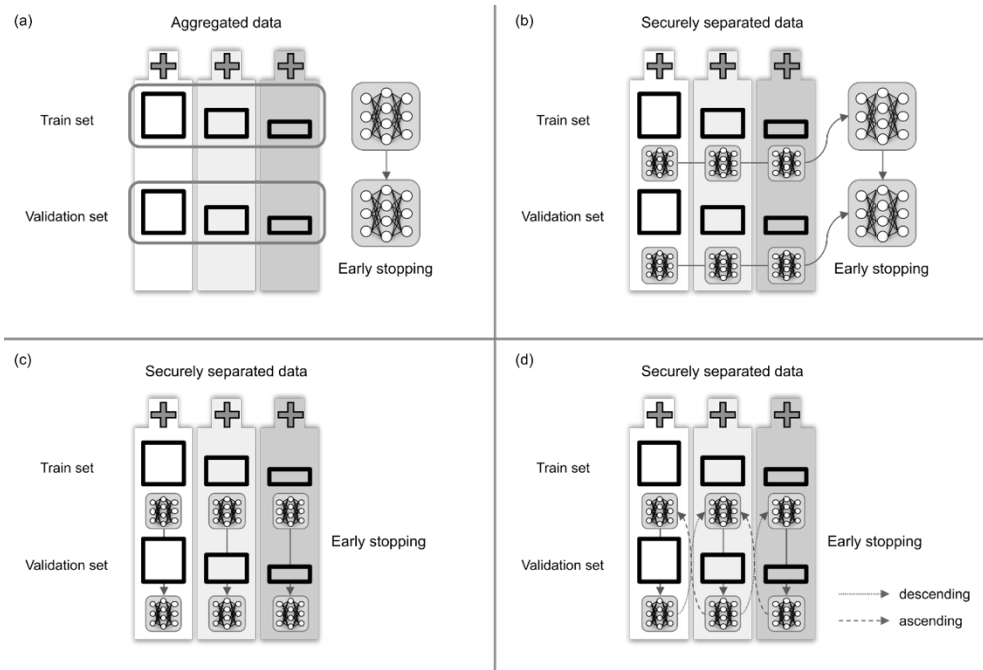


그림 3-1 프라이버시 보전 딥러닝 환경에서 훈련 알고리즘 (a) collaborative data sharing (CDS), (b) 연합 학습 (federated learning; FL), (c) 지역 학습 (local learning), (d) 전이반복학습 (transfer iterative learning; TIL). CDS와 FL은 단일 서버 모형로 결과가 나오며, TIL, 지역 학습은 개별 기관 별로 모형을 생성한다.

두 가지 형태의 PPDL 훈련 알고리즘과 두 가지 형태의 기저 학습 방법이 이 연구에서 비교 되었다 (그림 3-1). 두 PPDL 훈련 알고리즘은 TIL 과 FL 이었으며, TIL 은 데이터 크기의 순으로 내림차순과 오름차순으로 기관을 방문하고 설정한 횟수의 라운드를 돌려 수행하였다. 설정한 횟수의

라운드 (1 라운드와 10 라운드)와 방문의 순서 (데이터 크기에 따른 내림차순과 오름차순)은 줄여서 TIL-1-asc, TIL-1-desc, TIL-10-asc, TIL-10-desc 로 표기하였다. 각 라운드에서 모형 전이의 횟수는 TIL 에서 기관의 수와 동일 하였고 FL 에서 최소 가상 기관 수의 두 배와 같았다. 추가적으로 TIL 은 FL 보다 적은 수의 라운드를 필요로 하며 모형 전이가 덜 일어난다.

사용된 기저 학습 방법은 CDS¹¹³ 와 지역 학습이다. 하지만, CDS 는 민감한 개인정보 데이터를 집적하기 때문에, 실 세계에서 수행할 수 없는 방법이며 이것이 시뮬레이션 연구라 수행이 가능하였다.

개별 학습 방법은 다른 개수의 모형을 생성하는데, CDS 와 FL 은 단일 서버 모형 하나로 출력 되었고, 반면에 TIL 과 지역 학습은 개별 가상 기관마다 모형을 생성한다.

각 훈련 알고리즘에서 생성된 모든 모형은 지역 성능 점수와 전역 성능 점수를 만들기 위해 모든 기관에서 모은 테스트 데이터로 평가되었다.

제 3 절 가상 기관의 기저 특성

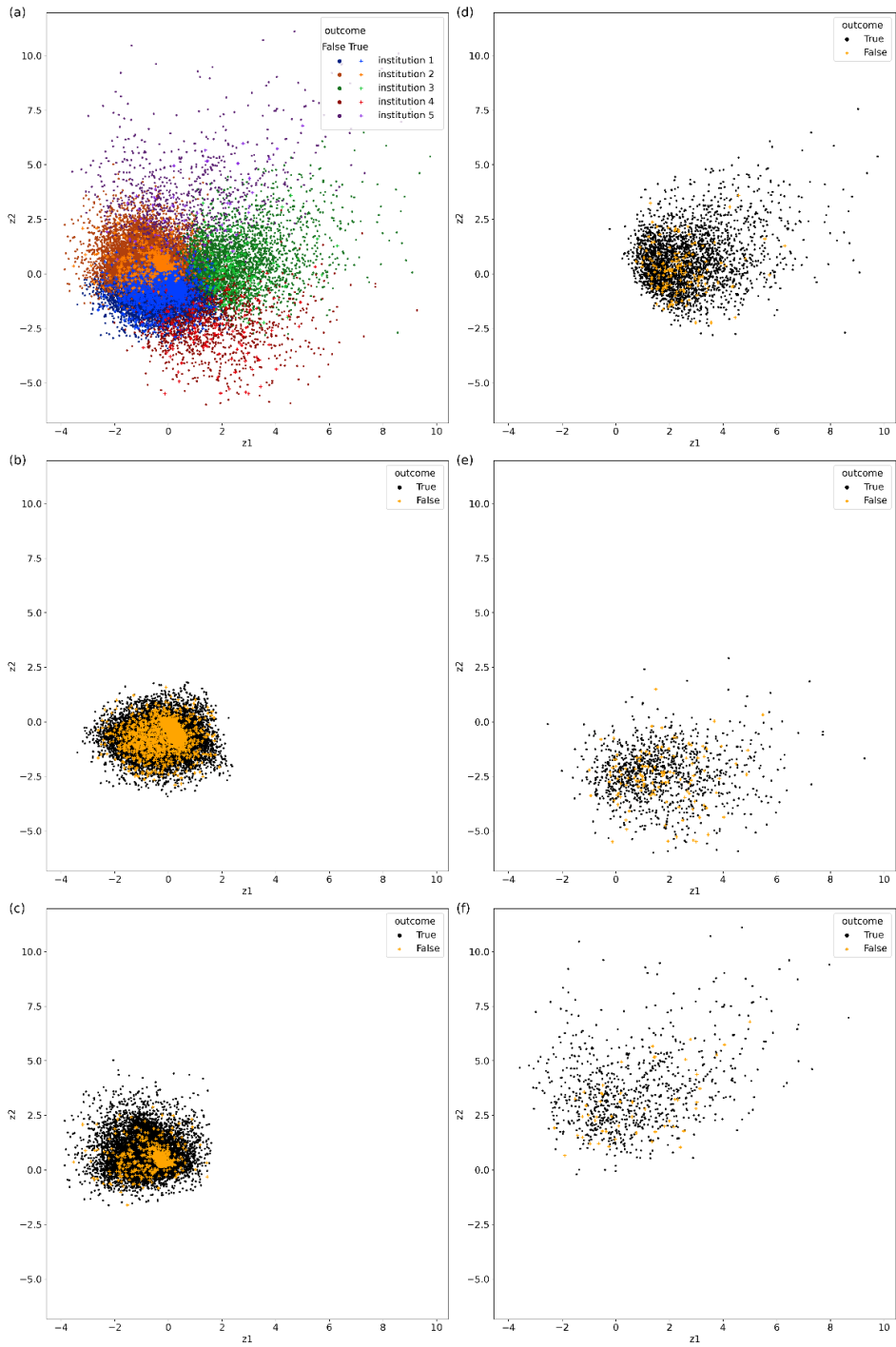


그림 3-2 가상의 의료 기관의 데이터 분포. 각 가상 기관들은 전체 전처리된 데이터의 군집에 의해 다섯 군으로 생성되었다. 산포도는 주성분분석을 사용하여 2 차원으로 데이터의 차원으로 줄여 생성되었다. z_1 과 z_2 는 첫번째와 두번째 주성분을 의미한다. (a) 모든 기관의 데이터 분포, (b-f)는 개별 가상 기관 1 부터 5 까지의 데이터 분포를 의미한다.

표 3-1 2 형 당뇨 환자의 기저 특성. 중요도 순서로 내림차순으로 상위 20 개 변수가 제시되었다. 중요도는 전체 전처리된 데이터로 훈련한 로지스틱 회귀 모형의 계수로부터 측정되었다. 각 열은 가상 기관 1~5 와 그 가상 기관을 모두 합한 overall, 그리고 각 가상 기관 사이 p-value 를 제시하였다.: 1) serum CO₂: serum carbon dioxide, 2) aPTT: activated Partial Thromboplastin Clotting Time, 3) HbA1c: glycated hemoglobin, 4) ALT: Alanine aminotransferase, 5) AST: Aspartate aminotransferase, 6) GAD65 Ab: Glutamic Acid Decarboxylase Antibody, 7) BPH: Benign Prostate Hypertrophy

Variable		Overall	Institution _1	Institution _2	Institution _3	Institution _4	Institution _5	P-Value
Population		40507	21532	13981	2924	1204	866	
Outcomes	n (%)	3478 (8.6)	2517 (11.7)	649 (4.6)	145 (5.0)	116 (9.6)	51 (5.9)	<0.05
Male	n (%)	20616 (50.9)	13783 (64.0)	4948 (35.4)	1047 (35.8)	472 (39.2)	366 (42.3)	<0.05
Female	n (%)	19891 (49.1)	7749 (36.0)	9033 (64.6)	1877 (64.2)	732 (60.8)	500 (57.7)	<0.05
Age by year	mean (SD)	59.58 (12.54)	66.82 (6.95)	47.50 (9.75)	63.35 (10.95)	65.81 (12.13)	53.22 (14.39)	<0.05
Diabetes medication	n (%)	23515 (58.1)	12197 (56.6)	8268 (59.1)	1864 (63.7)	707 (58.7)	479 (55.3)	<0.05
Hyperlipidemia medication	n (%)	10282 (25.4)	6095 (28.3)	3268 (23.4)	431 (14.7)	339 (28.2)	149 (17.2)	<0.05
Anticoagulant	n (%)	1740 (4.3)	835 (3.9)	330 (2.4)	379 (13.0)	140 (11.6)	56 (6.5)	<0.05
Antihypertension	n (%)	11701 (28.9)	7009 (32.6)	2653 (19.0)	1119 (38.3)	682 (56.6)	238 (27.5)	<0.05
Antiplatelet agent	n (%)	2204 (5.4)	1611 (7.5)	369 (2.6)	97 (3.3)	105 (8.7)	22 (2.5)	<0.05
serum Sodium (mmol/L)	mean (SD)	139.1 (3.3)	140.4 (2.5)	139.2 (2.6)	135.7 (3.8)	138.1 (3.9)	137.8 (3.9)	<0.05
serum CO ₂ (mmol/L) ¹⁾	mean (SD)	26.7 (3.3)	27.4 (2.9)	27.1 (3.0)	25.3 (3.5)	23.7 (4.0)	26.1 (3.4)	<0.05
Fibrinogen (mg/dL)	mean (SD)	366.9 (116.3)	349.1 (97.6)	343.2 (98.6)	417.6 (139.6)	396.2 (123.6)	343.6 (123.9)	<0.05
aPTT (sec) ²⁾	mean (SD)	32.7 (5.4)	32.2 (4.5)	32.4 (4.4)	34.2 (7.8)	32.9 (6.2)	33.6 (6.1)	<0.05
HbA1c (%) ³⁾	mean (SD)	7.4 (1.1)	7.4 (1.0)	7.3 (1.1)	7.4 (1.3)	7.3 (1.2)	7.1 (1.2)	<0.05
ALT (IU/L) ⁴⁾	mean (SD)	33.8 (31.1)	26.2 (15.4)	36.1 (22.0)	27.5 (18.6)	23.4 (19.2)	139.6 (62.3)	<0.05

AST (IU/L) ⁵⁾	mean (SD)	29.2 (23.8)	24.6 (10.9)	27.3 (13.1)	27.8 (16.8)	22.5 (13.0)	108.9 (58.4)	<0.05
C-peptide (ng/mL)	mean (SD)	2.5 (1.8)	2.4 (1.6)	2.4 (1.7)	2.4 (2.1)	4.0 (2.9)	3.3 (2.2)	<0.05
GAD65 (U/ml) ⁶⁾	Ab mean (SD)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.1 (0.1)	0.1 (0.1)	0.61
Essential hypertension	n (%)	3115 (7.7)	2131 (9.9)	768 (5.5)	94 (3.2)	63 (5.2)	59 (6.8)	<0.05
Cerebral atherosclerosis	n (%)	39 (0.1)	32 (0.1)	2 (0.0)	1 (0.0)	2 (0.2)	2 (0.2)	<0.05
BPH ⁷⁾	n (%)	156 (0.4)	133 (0.6)	6 (0.0)	6 (0.2)	8 (0.7)	3 (0.3)	<0.05

2004년 1월부터 2020년 12월까지 서울대학교병원에 구현된 CDM에서 추출된 2형 당뇨병 환자 40,507명을 분석하였다. 3,478명 (8.6%)의 환자가 CVD의 조작적 정의에 기반해 CVD로 진단되었다. 평균 연령은 59.6 ± 12.5 세였다. 23,515명 (58.1%)의 환자가 2형 당뇨병 약제를 복용 중이었으며 평균 당화혈색소 (HbA1c)는 $7.4 \pm 1.1\%$ 였다. 3,115명 (7.7%)의 환자가 본태성 고혈압으로 진단 받았으며 39명 (0.1%)의 환자가 뇌혈관의 죽상동맥경화로 진단 받았다.

서울대학교병원 CVD 데이터로부터 이질적인 데이터 분포를 지닌 5개의 가상 기관이 생성되었으며 (그림 3-2) 개인정보보호 의료 기관 환경을 시뮬레이션하였다. 가상 기관들은 데이터 크기에 따라 내림차순으로 번호 붙였다. 표 3-1은 각 가상 기관의 환자의 기저 특성을 보여준다. 기관 간 유의한 차이 ($p < 0.05$)를 보이는 CVD의 발생을 포함한 117개의 변수를 분석 및 모형에 사용하였다 (표 3-2).

표 3-2 서울대학교병원 공통 데이터 모형에서 추출되어 프라이버시 보전 딥러닝 모형에 사용된 변수 목록

concept_name		category	Missing	Overall	snuh_0	snuh_1	snuh_2	snuh_3	snuh_4	P-Value	
total counts		characteristics		40507	21532	13981	2924	1204	866		
male	n (%)	characteristics		20616 (50.9)	13783 (64.0)	4948 (35.4)	1047 (35.8)	472 (39.2)	366 (42.3)	<0.001	
female	n (%)	characteristics		19891 (49.1)	7749 (36.0)	9033 (64.6)	1877 (64.2)	732 (60.8)	500 (57.7)	<0.001	
age (year)	mean (SD)	characteristics		59.58 (12.54)	66.82 (6.95)	47.50 (9.75)	63.35 (10.95)	65.81 (12.13)	53.22 (14.39)	<0.001	
3007263	mean (SD)	Hemoglobin A1c/Hemoglobin.total in Blood by calculation	measurement	3010	7.4 (1.1)	7.4 (1.0)	7.3 (1.1)	7.4 (1.3)	7.3 (1.2)	7.1 (1.2)	<0.001
3015377	mean (SD)	Calcium in Serum or Plasma	measurement	19087	9.2 (0.6)	9.3 (0.4)	9.4 (0.4)	8.5 (0.6)	8.9 (0.7)	9.2 (0.6)	<0.001
3014576	mean (SD)	Chloride in Serum or Plasma	measurement	21993	102.7 (3.7)	103.7 (3.0)	102.2 (3.0)	100.6 (4.6)	104.1 (5.3)	101.7 (3.9)	<0.001
3028833	mean (SD)	Bilirubin.total in Blood	measurement	18767	0.8 (0.5)	0.8 (0.3)	0.9 (0.4)	0.9 (0.7)	0.6 (0.4)	1.4 (1.3)	<0.001
3015632	mean (SD)	Carbon dioxide, total in Serum or Plasma	measurement	24557	26.7 (3.3)	27.4 (2.9)	27.1 (3.0)	25.3 (3.5)	23.7 (4.0)	26.1 (3.4)	<0.001
3042781	mean (SD)	Aspartate aminotransferase in Serum or Plasma	measurement	19300	29.2 (23.8)	24.6 (10.9)	27.3 (13.1)	27.8 (16.8)	22.5 (13.0)	108.9 (58.4)	<0.001
42870365	mean (SD)	C reactive protein in Blood by High sensitivity method	measurement	26141	1.5 (3.0)	0.7 (1.5)	0.8 (1.8)	4.5 (4.9)	1.4 (2.7)	1.6 (2.9)	<0.001
3013682	mean (SD)	Urea nitrogen in Serum or Plasma	measurement	18896	16.0 (6.5)	16.0 (4.6)	14.2 (4.2)	14.0 (5.7)	33.1 (8.4)	14.6 (5.4)	<0.001
3004501	mean (SD)	Glucose in Serum or Plasma	measurement	19778	165.7 (62.0)	151.2 (51.7)	175.1 (62.7)	189.6 (72.7)	166.8 (73.0)	180.6 (66.8)	<0.001
3035995	mean (SD)	Alkaline phosphatase in Serum or Plasma	measurement	18816	81.5 (40.0)	73.2 (25.1)	76.1 (29.7)	108.1 (62.1)	85.1 (42.9)	128.4 (71.7)	<0.001

3019550	mean (SD)	Sodium in Serum or Plasma	measurement	21819	139.1 (3.3)	140.4 (2.5)	139.2 (2.6)	135.7 (3.8)	138.1 (3.9)	137.8 (3.9)	<0.001
3023103	mean (SD)	Potassium in Serum or Plasma	measurement	21837	4.3 (0.4)	4.3 (0.4)	4.3 (0.4)	4.2 (0.5)	4.7 (0.6)	4.3 (0.5)	<0.001
44787076	mean (SD)	Cholesterol in Serum or Plasma	measurement	19060	181.6 (44.6)	180.6 (39.6)	200.7 (42.2)	145.1 (39.3)	164.9 (44.7)	191.4 (49.5)	<0.001
3016723	mean (SD)	Creatinine in Serum or Plasma	measurement	18834	1.0 (0.4)	0.9 (0.2)	0.9 (0.2)	0.9 (0.3)	2.1 (0.9)	0.9 (0.2)	<0.001
3003458	mean (SD)	Phosphate in Serum or Plasma	measurement	19127	3.5 (0.6)	3.6 (0.6)	3.6 (0.6)	3.1 (0.7)	3.7 (0.8)	3.5 (0.7)	<0.001
3006923	mean (SD)	Alanine aminotransferase in Serum or Plasma	measurement	19310	33.8 (31.1)	26.2 (15.4)	36.1 (22.0)	27.5 (18.6)	23.4 (19.2)	139.6 (62.3)	<0.001
3024561	mean (SD)	Albumin in Serum or Plasma	measurement	18752	4.1 (0.5)	4.2 (0.4)	4.4 (0.4)	3.3 (0.5)	3.7 (0.6)	4.0 (0.6)	<0.001
3002417	mean (SD)	Prothrombin time in Blood by Coagulation assay	measurement	26288	54.5 (44.5)	55.4 (45.6)	56.4 (46.1)	49.3 (38.3)	55.3 (45.3)	51.7 (42.2)	<0.001
3032080	mean (SD)	INR in Blood by Coagulation assay	measurement	26291	1.0 (0.2)	1.0 (0.2)	1.0 (0.1)	1.2 (0.3)	1.0 (0.2)	1.1 (0.2)	<0.001
3013466	mean (SD)	aPTT in Blood by Coagulation assay	measurement	27678	32.7 (5.4)	32.2 (4.5)	32.4 (4.4)	34.2 (7.8)	32.9 (6.2)	33.6 (6.1)	<0.001
3004077	mean (SD)	Glucose in Capillary blood	measurement	32029	168.4 (64.8)	157.9 (56.0)	167.5 (62.7)	180.3 (70.3)	174.9 (75.5)	176.7 (70.0)	<0.001
3010084	mean (SD)	C peptide in Serum or Plasma	measurement	25473	2.5 (1.8)	2.4 (1.6)	2.4 (1.7)	2.4 (2.1)	4.0 (2.9)	3.3 (2.2)	<0.001
46236952	mean (SD)	Glomerular filtration rate in Serum, Plasma or Blood by Creatinine-based formula (MDRD)	measurement	27980	84.8 (26.8)	81.1 (19.3)	94.2 (21.5)	97.6 (32.0)	34.3 (16.2)	94.6 (25.1)	<0.001
3016407	mean (SD)	Fibrinogen in Platelet poor plasma by Coagulation assay	measurement	31710	366.9 (116.3)	349.1 (97.6)	343.2 (98.6)	417.6 (139.6)	396.2 (123.6)	343.6 (123.9)	<0.001
3020148	mean (SD)	Glutamate decarboxylase 65 Ab in Serum	measurement	36574	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.2 (0.1)	0.1 (0.1)	0.1 (0.1)	0.606
3022250	mean (SD)	Lactate dehydrogenase in Serum or Plasma by Lactate to pyruvate reaction	measurement	38080	255.9 (174.5)	229.1 (113.9)	240.4 (153.4)	293.7 (230.6)	263.5 (166.9)	307.2 (248.4)	<0.001

3007220	mean (SD)	Creatine kinase in Serum or Plasmas	measurement	38124	124.7 (215.7)	107.6 (160.2)	118.5 (147.7)	170.7 (343.8)	130.5 (216.7)	130.9 (268.9)	<0.001
Diabetes medication	n (%)		medication		23515 (58.1)	12197 (56.6)	8268 (59.1)	1864 (63.7)	707 (58.7)	479 (55.3)	<0.002
Hyperlipidemia medication	n (%)		medication		10282 (25.4)	6095 (28.3)	3268 (23.4)	431 (14.7)	339 (28.2)	149 (17.2)	<0.004
Anticoagulant	n (%)		medication		1740 (4.3)	835 (3.9)	330 (2.4)	379 (13.0)	140 (11.6)	56 (6.5)	<0.006
Antihypertension	n (%)		medication		11701 (28.9)	7009 (32.6)	2653 (19.0)	1119 (38.3)	682 (56.6)	238 (27.5)	<0.008
Antiplatelet agent	n (%)		medication		2204 (5.4)	1611 (7.5)	369 (2.6)	97 (3.3)	105 (8.7)	22 (2.5)	<0.010
Targets	n (%)		medication		3478 (8.6)	2517 (11.7)	649 (4.6)	145 (5.0)	116 (9.6)	51 (5.9)	<0.012
316866	n (%)	Hypertensive disorder	condition		5816 (14.4)	3832 (17.8)	1344 (9.6)	306 (10.5)	220 (18.3)	114 (13.2)	<0.001
4159131	n (%)	Dyslipidemia	condition		834 (2.1)	409 (1.9)	354 (2.5)	28 (1.0)	9 (0.7)	34 (3.9)	<0.001
320128	n (%)	Essential hypertension	condition		3115 (7.7)	2131 (9.9)	768 (5.5)	94 (3.2)	63 (5.2)	59 (6.8)	<0.001
4029305	n (%)	Hypercholesterolemia	condition		3586 (8.9)	2216 (10.3)	1193 (8.5)	63 (2.2)	48 (4.0)	66 (7.6)	<0.001
432867	n (%)	Hyperlipidemia	condition		1298 (3.2)	720 (3.3)	516 (3.7)	19 (0.6)	17 (1.4)	26 (3.0)	<0.001
4029263	n (%)	Primary combined hyperlipidemia	condition		66 (0.2)	55 (0.3)	10 (0.1)	0 (0.0)	1 (0.1)	0 (0.0)	<0.001
4292079	n (%)	Hyperlipoproteinemia	condition		38 (0.1)	29 (0.1)	9 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0.044
4120314	n (%)	Hypertriglyceridemia	condition		351 (0.9)	163 (0.8)	170 (1.2)	6 (0.2)	4 (0.3)	8 (0.9)	<0.001
4028741	n (%)	Benign hypertension	condition		376 (0.9)	253 (1.2)	85 (0.6)	16 (0.5)	15 (1.2)	7 (0.8)	<0.001

4289933	n (%)	Malignant hypertension	condition	37 (0.1)	21 (0.1)	13 (0.1)	0 (0.0)	1 (0.1)	2 (0.2)	0.328
4220010	n (%)	Fredrickson type IIa hyperproteinemia	condition	14 (0.0)	11 (0.1)	3 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.39
4029261	n (%)	Fredrickson type IV hyperlipoproteinemia	condition	3 (0.0)	3 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.619
4162681	n (%)	Familial type 3 hyperlipoproteinemia	condition	2 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.779
4134862	n (%)	Familial hypercholesterolemia	condition	4 (0.0)	4 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.474
4079876	n (%)	Familial combined hyperlipidemia	condition	5 (0.0)	3 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.951
40479758	n (%)	Pseudoaneurysm	condition	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	<0.001
201340	n (%)	Gastritis	condition	1061 (2.6)	661 (3.1)	263 (1.9)	81 (2.8)	28 (2.3)	28 (3.2)	<0.001
193249	n (%)	Acute hemorrhagic gastritis	condition	188 (0.5)	101 (0.5)	76 (0.5)	2 (0.1)	3 (0.2)	6 (0.7)	0.007
4232623	n (%)	Helicobacter-associated gastritis	condition	88 (0.2)	55 (0.3)	32 (0.2)	1 (0.0)	0 (0.0)	0 (0.0)	0.032
4225273	n (%)	Chronic gastritis	condition	216 (0.5)	140 (0.7)	59 (0.4)	8 (0.3)	5 (0.4)	4 (0.5)	0.012
192667	n (%)	Atrophic gastritis	condition	252 (0.6)	149 (0.7)	93 (0.7)	2 (0.1)	4 (0.3)	4 (0.5)	0.001
4057513	n (%)	Chronic superficial gastritis	condition	105 (0.3)	55 (0.3)	49 (0.4)	0 (0.0)	1 (0.1)	0 (0.0)	0.003
199866	n (%)	Acute gastritis	condition	134 (0.3)	64 (0.3)	40 (0.3)	18 (0.6)	9 (0.7)	3 (0.3)	0.004
195306	n (%)	Gastroduodenitis	condition	82 (0.2)	53 (0.2)	21 (0.2)	3 (0.1)	3 (0.2)	2 (0.2)	0.237
433516	n (%)	Duodenitis	condition	27 (0.1)	18 (0.1)	7 (0.1)	1 (0.0)	1 (0.1)	0 (0.0)	0.628
4087594	n (%)	Acute gastric mucosal erosion	condition	4 (0.0)	2 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.941
201059	n (%)	Hypertrophic gastritis	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755

198249	n (%)	Congenital anomaly of digestive system	condition	4 (0.0)	3 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.071
4247651	n (%)	Bile-induced gastritis	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
4057514	n (%)	Corrosive gastritis	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
194574	n (%)	Chronic type B viral hepatitis	condition	993 (2.5)	311 (1.4)	422 (3.0)	149 (5.1)	31 (2.6)	80 (9.2)	<0.001
4059290	n (%)	Steatosis of liver	condition	1822 (4.5)	782 (3.6)	853 (6.1)	35 (1.2)	9 (0.7)	143 (16.5)	<0.001
198964	n (%)	Chronic hepatitis C	condition	234 (0.6)	114 (0.5)	49 (0.4)	23 (0.8)	7 (0.6)	41 (4.7)	<0.001
4098583	n (%)	Posthepatic cirrhosis	condition	941 (2.3)	262 (1.2)	285 (2.0)	242 (8.3)	39 (3.2)	113 (13.0)	<0.001
4064161	n (%)	Cirrhosis of liver	condition	518 (1.3)	182 (0.8)	131 (0.9)	143 (4.9)	24 (2.0)	38 (4.4)	<0.001
4001171	n (%)	Liver cell carcinoma	condition	1185 (2.9)	391 (1.8)	252 (1.8)	363 (12.4)	49 (4.1)	130 (15.0)	<0.001
192680	n (%)	Portal hypertension	condition	4 (0.0)	1 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0.026
194984	n (%)	Disease of liver	condition	316 (0.8)	140 (0.7)	102 (0.7)	38 (1.3)	7 (0.6)	29 (3.3)	<0.001
4225905	n (%)	Liver cyst	condition	92 (0.2)	63 (0.3)	18 (0.1)	6 (0.2)	1 (0.1)	4 (0.5)	0.01
4098954	n (%)	Klatskin's tumor	condition	25 (0.1)	8 (0.0)	2 (0.0)	11 (0.4)	0 (0.0)	4 (0.5)	<0.001
4313567	n (%)	Postnecrotic cirrhosis	condition	34 (0.1)	16 (0.1)	9 (0.1)	7 (0.2)	0 (0.0)	2 (0.2)	0.014
4001664	n (%)	Intrahepatic bile duct carcinoma	condition	67 (0.2)	32 (0.1)	6 (0.0)	19 (0.6)	7 (0.6)	3 (0.3)	<0.001
4340394	n (%)	Hepatic sclerosis	condition	5 (0.0)	0 (0.0)	5 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.05
4135822	n (%)	Primary biliary cholangitis	condition	58 (0.1)	26 (0.1)	14 (0.1)	5 (0.2)	2 (0.2)	11 (1.3)	<0.001
4252074	n (%)	Cardiac cirrhosis	condition	10 (0.0)	4 (0.0)	1 (0.0)	4 (0.1)	1 (0.1)	0 (0.0)	0.001

4340379	n (%)	Hepatitis B carrier	condition	32 (0.1)	16 (0.1)	13 (0.1)	2 (0.1)	1 (0.1)	0 (0.0)	0.889
40484532	n (%)	Nonalcoholic steatohepatitis	condition	21 (0.1)	10 (0.0)	9 (0.1)	0 (0.0)	0 (0.0)	2 (0.2)	0.09
4133325	n (%)	Focal nodular hyperplasia of liver	condition	10 (0.0)	5 (0.0)	5 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.746
196455	n (%)	Hepatorenal syndrome	condition	21 (0.1)	2 (0.0)	3 (0.0)	9 (0.3)	4 (0.3)	3 (0.3)	<0.001
4046123	n (%)	Secondary biliary cirrhosis	condition	6 (0.0)	1 (0.0)	1 (0.0)	3 (0.1)	0 (0.0)	1 (0.1)	<0.001
4026131	n (%)	Non-alcoholic fatty liver	condition	53 (0.1)	9 (0.0)	29 (0.2)	1 (0.0)	0 (0.0)	14 (1.6)	<0.001
4208660	n (%)	Cholangiocarcinoma of biliary tract	condition	59 (0.1)	25 (0.1)	10 (0.1)	17 (0.6)	2 (0.2)	5 (0.6)	<0.001
4146459	n (%)	Esophageal varices in cirrhosis of the liver	condition	8 (0.0)	2 (0.0)	4 (0.0)	1 (0.0)	1 (0.1)	0 (0.0)	0.322
4246127	n (%)	Malignant neoplasm of liver	condition	20 (0.0)	4 (0.0)	5 (0.0)	7 (0.2)	0 (0.0)	4 (0.5)	<0.001
4130518	n (%)	Neoplasm of liver	condition	9 (0.0)	2 (0.0)	5 (0.0)	2 (0.1)	0 (0.0)	0 (0.0)	0.196
4089663	n (%)	Malignant neoplasm of intrahepatic canaliculi	condition	1 (0.0)	0 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.012
4159144	n (%)	Hepatopulmonary syndrome	condition	3 (0.0)	1 (0.0)	0 (0.0)	1 (0.0)	1 (0.1)	0 (0.0)	0.009
4277276	n (%)	Veno-occlusive disease of the liver	condition	3 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	3 (0.3)	<0.001
201519	n (%)	Primary malignant neoplasm of liver	condition	23 (0.1)	6 (0.0)	5 (0.0)	7 (0.2)	1 (0.1)	4 (0.5)	<0.001
192675	n (%)	Biliary cirrhosis	condition	1 (0.0)	0 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.012
4340384	n (%)	Ischemic hepatitis	condition	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	<0.001
4240725	n (%)	Peliosis hepatis	condition	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	<0.001
4044708	n (%)	Langerhans cell histiocytosis, unifocal	condition	1 (0.0)	0 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.012
4342772	n (%)	Chronic non-A non-B hepatitis	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755
4111998	n (%)	Esophageal varices associated with another disorder	condition	2 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.983

4166154	n (%)	Combined hepatocellular carcinoma and cholangiocarcinoma	condition	4 (0.0)	1 (0.0)	2 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.1
4267417	n (%)	Hepatic fibrosis	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755
4116347	n (%)	Portal hypertensive gastropathy	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
4151243	n (%)	Accidental organ perforation during a procedure	condition	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	<0.001
4088114	n (%)	Postoperative intra-abdominal abscess	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
432851	n (%)	Secondary malignant neoplastic disease	condition	2 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.983
443597	n (%)	Chronic kidney disease stage 3	condition	83 (0.2)	33 (0.2)	13 (0.1)	4 (0.1)	32 (2.7)	1 (0.1)	<0.001
198185	n (%)	Chronic renal failure	condition	203 (0.5)	97 (0.5)	51 (0.4)	5 (0.2)	49 (4.1)	1 (0.1)	<0.001
193782	n (%)	End-stage renal disease	condition	405 (1.0)	110 (0.5)	110 (0.8)	29 (1.0)	149 (12.4)	7 (0.8)	<0.001
443611	n (%)	Chronic kidney disease stage 5	condition	121 (0.3)	32 (0.1)	24 (0.2)	9 (0.3)	55 (4.6)	1 (0.1)	<0.001
46271022	n (%)	Chronic kidney disease	condition	486 (1.2)	177 (0.8)	66 (0.5)	28 (1.0)	211 (17.5)	4 (0.5)	<0.001
443601	n (%)	Chronic kidney disease stage 2	condition	12 (0.0)	5 (0.0)	4 (0.0)	0 (0.0)	2 (0.2)	1 (0.1)	0.028
443612	n (%)	Chronic kidney disease stage 4	condition	63 (0.2)	22 (0.1)	6 (0.0)	0 (0.0)	35 (2.9)	0 (0.0)	<0.001
4030520	n (%)	End stage renal failure on dialysis	condition	7 (0.0)	1 (0.0)	1 (0.0)	1 (0.0)	4 (0.3)	0 (0.0)	<0.001
443614	n (%)	Chronic kidney disease stage 1	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
316437	n (%)	Cerebral atherosclerosis	condition	39 (0.1)	32 (0.1)	2 (0.0)	1 (0.0)	2 (0.2)	2 (0.2)	0.001
312938	n (%)	Hypertensive encephalopathy	condition	3 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.044
4180026	n (%)	Vertebral artery aneurysm	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927

381591	n (%)	Cerebrovascular disease	condition	27 (0.1)	22 (0.1)	1 (0.0)	1 (0.0)	3 (0.2)	0 (0.0)	0.001
374384	n (%)	Cerebral ischemia	condition	36 (0.1)	29 (0.1)	3 (0.0)	1 (0.0)	2 (0.2)	1 (0.1)	0.007
4201411	n (%)	Carotid cavernous fistula	condition	3 (0.0)	2 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.389
4225722	n (%)	Dural arteriovenous malformation	condition	3 (0.0)	1 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.523
4278243	n (%)	Cerebral arteriosclerosis	condition	5 (0.0)	3 (0.0)	0 (0.0)	1 (0.0)	1 (0.1)	0 (0.0)	0.096
377001	n (%)	Nonruptured cerebral aneurysm	condition	3 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.03
4077957	n (%)	Unruptured aneurysm of middle cerebral artery	condition	2 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.779
4078443	n (%)	Unruptured aneurysm of posterior communicating artery	condition	2 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.003
4077829	n (%)	Unruptured aneurysm of carotid artery	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
4102124	n (%)	Carotid atherosclerosis	condition	4 (0.0)	3 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.916
4077199	n (%)	Unruptured aneurysm of anterior communicating artery	condition	3 (0.0)	1 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.828
4029497	n (%)	Cerebral arterial aneurysm	condition	3 (0.0)	1 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.828
432346	n (%)	Dissection of vertebral artery	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
4029966	n (%)	Intracranial aneurysm	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755
4078016	n (%)	Aneurysm of internal carotid artery	condition	4 (0.0)	4 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.474
378774	n (%)	Moyamoya disease	condition	3 (0.0)	2 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.389
4062391	n (%)	Injury of muscle and tendon of long flexor muscle of toe at ankle and foot level	condition	166 (0.4)	63 (0.3)	36 (0.3)	8 (0.3)	57 (4.7)	2 (0.2)	<0.001
4318545	n (%)	Cellulitis of hip	condition	139 (0.3)	48 (0.2)	19 (0.1)	9 (0.3)	60 (5.0)	3 (0.3)	<0.001
4320944	n (%)	Cellulitis of toe	condition	9 (0.0)	1 (0.0)	0 (0.0)	1 (0.0)	7 (0.6)	0 (0.0)	<0.001

4028237	n (%)	Cellulitis of foot	condition	37 (0.1)	11 (0.1)	6 (0.0)	2 (0.1)	17 (1.4)	1 (0.1)	<0.001
4100662	n (%)	Cellulitis of finger	condition	16 (0.0)	3 (0.0)	3 (0.0)	2 (0.1)	8 (0.7)	0 (0.0)	<0.001
4267714	n (%)	Cellulitis of hand	condition	5 (0.0)	2 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0.085
444129	n (%)	Injury of wrist	condition	3 (0.0)	2 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.976
375545	n (%)	Cataract	condition	2705 (6.7)	2013 (9.3)	436 (3.1)	116 (4.0)	110 (9.1)	30 (3.5)	<0.001
374028	n (%)	Age related macular degeneration	condition	109 (0.3)	83 (0.4)	18 (0.1)	5 (0.2)	3 (0.2)	0 (0.0)	<0.001
4290333	n (%)	Macular retinal edema	condition	300 (0.7)	182 (0.8)	95 (0.7)	13 (0.4)	8 (0.7)	2 (0.2)	0.032
378416	n (%)	Retinal disorder	condition	209 (0.5)	119 (0.6)	74 (0.5)	6 (0.2)	9 (0.7)	1 (0.1)	0.039
4146103	n (%)	Degeneration of macula and posterior pole	condition	96 (0.2)	83 (0.4)	8 (0.1)	4 (0.1)	1 (0.1)	0 (0.0)	<0.001
374025	n (%)	Retinal hemorrhage	condition	50 (0.1)	21 (0.1)	23 (0.2)	2 (0.1)	2 (0.2)	2 (0.2)	0.309
379010	n (%)	Epiretinal membrane	condition	332 (0.8)	266 (1.2)	46 (0.3)	4 (0.1)	15 (1.2)	1 (0.1)	<0.001
4224741	n (%)	Degenerative disorder of macula	condition	22 (0.1)	15 (0.1)	5 (0.0)	1 (0.0)	0 (0.0)	1 (0.1)	0.511
4338894	n (%)	Macular hole	condition	82 (0.2)	61 (0.3)	18 (0.1)	1 (0.0)	2 (0.2)	0 (0.0)	0.002
4090255	n (%)	Retinal drusen	condition	33 (0.1)	24 (0.1)	8 (0.1)	1 (0.0)	0 (0.0)	0 (0.0)	0.208
376965	n (%)	Hypertensive retinopathy	condition	52 (0.1)	25 (0.1)	20 (0.1)	3 (0.1)	4 (0.3)	0 (0.0)	0.226
4318985	n (%)	Degeneration of retina	condition	29 (0.1)	19 (0.1)	10 (0.1)	0 (0.0)	0 (0.0)	0 (0.0)	0.353
434337	n (%)	Retinal vascular disorder	condition	16 (0.0)	8 (0.0)	3 (0.0)	1 (0.0)	1 (0.1)	3 (0.3)	<0.001
4105178	n (%)	Cystoid macular edema	condition	43 (0.1)	28 (0.1)	13 (0.1)	2 (0.1)	0 (0.0)	0 (0.0)	0.409

372629	n (%)	Nonexudative age-related macular degeneration	condition	64 (0.2)	54 (0.3)	6 (0.0)	1 (0.0)	3 (0.2)	0 (0.0)	<0.001
376966	n (%)	Exudative age-related macular degeneration	condition	55 (0.1)	46 (0.2)	4 (0.0)	2 (0.1)	2 (0.2)	1 (0.1)	<0.001
432890	n (%)	Microaneurysm of retinal artery	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755
372894	n (%)	Central serous chorioretinopathy	condition	25 (0.1)	9 (0.0)	11 (0.1)	2 (0.1)	1 (0.1)	2 (0.2)	0.187
376973	n (%)	Cataract secondary to ocular disease	condition	4 (0.0)	3 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.916
4335998	n (%)	Proliferative vitreoretinopathy	condition	15 (0.0)	9 (0.0)	6 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.715
4195988	n (%)	Myopic macular degeneration	condition	10 (0.0)	5 (0.0)	4 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.609
4195046	n (%)	Clinically significant macular edema	condition	84 (0.2)	56 (0.3)	27 (0.2)	0 (0.0)	1 (0.1)	0 (0.0)	0.019
4103381	n (%)	Retinitis pigmentosa	condition	14 (0.0)	8 (0.0)	6 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.726
432619	n (%)	Angioid streaks of choroid	condition	2 (0.0)	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.209
435810	n (%)	After-cataract	condition	39 (0.1)	30 (0.1)	7 (0.1)	0 (0.0)	1 (0.1)	1 (0.1)	0.038
4232976	n (%)	Coats' disease	condition	2 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.983
378142	n (%)	Retinal ischemia	condition	2 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.779
4081146	n (%)	Lens capsule opacity in pseudophakia	condition	5 (0.0)	3 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.951
4338889	n (%)	Acute retinal necrosis	condition	4 (0.0)	2 (0.0)	1 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.135
436403	n (%)	Retinal vasculitis	condition	9 (0.0)	5 (0.0)	0 (0.0)	1 (0.0)	1 (0.1)	2 (0.2)	<0.001
4195500	n (%)	Idiopathic juxtafoveal telangiectasia	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755
4083487	n (%)	Macular drusen	condition	28 (0.1)	17 (0.1)	8 (0.1)	2 (0.1)	0 (0.0)	1 (0.1)	0.792
378414	n (%)	Retinal detachment	condition	6 (0.0)	1 (0.0)	3 (0.0)	2 (0.1)	0 (0.0)	0 (0.0)	0.095
4029156	n (%)	Subretinal hemorrhage	condition	4 (0.0)	3 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.916

377270	n (%)	Hereditary retinal dystrophy	condition	4 (0.0)	2 (0.0)	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.723
319016	n (%)	Retinal lattice degeneration	condition	4 (0.0)	1 (0.0)	3 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.56
4090244	n (%)	Retinal macroaneurysm	condition	3 (0.0)	0 (0.0)	2 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	0.017
433197	n (%)	Pseudoexfoliation of lens capsule	condition	5 (0.0)	4 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.809
4334885	n (%)	Parafoveal telangiectasia	condition	4 (0.0)	3 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.916
373499	n (%)	Disorder of eye region	condition	2 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.779
4197160	n (%)	Ocular ischemic syndrome	condition	5 (0.0)	4 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.809
4317952	n (%)	Radiation retinopathy	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
376401	n (%)	Traumatic cataract	condition	4 (0.0)	3 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.916
4109548	n (%)	Drug-induced cataract	condition	2 (0.0)	2 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.779
437267	n (%)	Peripheral retinal degeneration	condition	6 (0.0)	3 (0.0)	2 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0.9
441555	n (%)	Retinal telangiectasia	condition	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	0 (0.0)	<0.001
4204368	n (%)	Eales' disease	condition	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	<0.001
372903	n (%)	Retinal edema	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755
373766	n (%)	Retinopathy of prematurity	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
40479994	n (%)	Presenile cataract	condition	1 (0.0)	0 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.755
4164175	n (%)	Diabetic intraretinal microvascular anomaly	condition	1 (0.0)	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0.927
45770919	n (%)	Infantile and/or juvenile cataract	condition	1 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	0 (0.0)	1 (0.1)	<0.001
198803	n (%)	Benign prostatic hyperplasia	condition	156 (0.4)	133 (0.6)	6 (0.0)	6 (0.2)	8 (0.7)	3 (0.3)	<0.001
4093346	n (%)	Large prostate	condition	22 (0.1)	17 (0.1)	3 (0.0)	1 (0.0)	1 (0.1)	0 (0.0)	0.195

제 4 절 혼합 기관 환경 하에서 TIL 의 성능

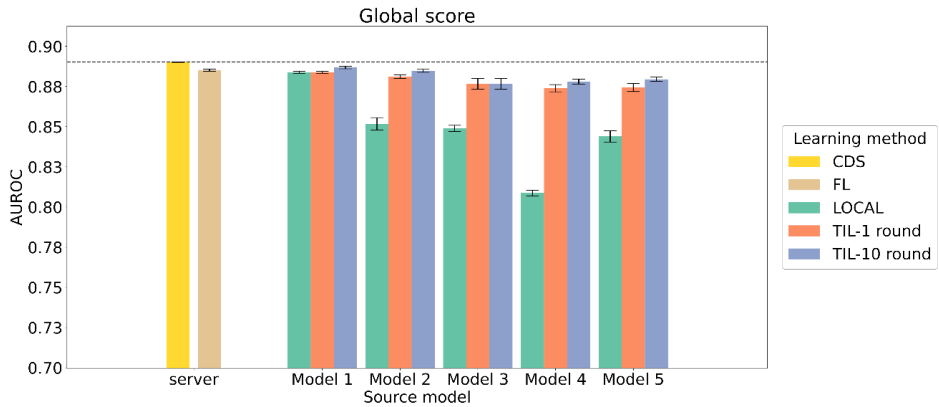


그림 3-3 혼합된 분포의 가상 기관 환경에서 훈련 알고리즘의 전역 성능.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

점선의 수평선은 CDS의 성능이며, 모형이 개인정보보호 환경에서 성취할 수 있는 상한을 의미한다. 표현된 TIL 모형은 매 라운드마다 데이터 크기의 내림차순으로 기관 1 부터 5 까지 방문하여 훈련된다.

막대 그래프는 모형 초기화를 무작위로 5 번 수행하여 생성된 AUROC 의 평균이며, 오차 막대는 95% 신뢰구간이다.

다양한 가상 기관 상태 하에서 TIL 의 효율성을 보이기 위해 다른 두 가지 PPDL 환경으로 훈련 알고리즘을 평가하였다. 혼합된 분포의 가상 기관 환경은 가장 큰 데이터부터 가장 작은 데이터까지 다양한 데이터 크기를 가진 가상 기관 환경이다.

성능 척도와 CDS 에 대한 상대적 효율성 (relative efficiency; RE)¹¹⁶ 은 다른 훈련 알고리즘의 학습 능력을 비교하기 위해 측정되었다. 모든 평가 척도는 그림 3-2 에서 보았듯이 지역 성능 점수와 전역 성능 점수로 측정되었다 (표 3-3 ~ 표 3-6).

표 3-3 (전역 성능) 전이반복학습 모형의 여러 평가 척도 결과.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

TIL-10 과 TIL-1 은 각각 10 라운드와 단일 라운드를 나타내며, asc 는 데이터가 적은 기관에서 많은 기관의 순서로 (오름차순), desc 는 데이터가 많은 기관에서 적은 기관의 순서로 (내림차순) 훈련한 것을 의미한다.

Specificity 를 0.750 으로 고정하여 계산하였다.

AUROC 는 Area Under the Receiver Operating Characteristics curve 를 의미하며, \pm SD 는 Standard Deviation 를 의미한다.

	Metric	AUROC	\pm SD	Accuracy	\pm SD	f1	\pm SD	Sensitivity	\pm SD
CDS	server	0.890	0.000	0.760	0.000	0.383	0.001	0.867	0.004
FL	server	0.885	0.001	0.760	0.001	0.381	0.003	0.862	0.008
LOCAL	model 1	0.884	0.001	0.759	0.000	0.378	0.002	0.853	0.006
	model 2	0.852	0.005	0.753	0.001	0.351	0.006	0.780	0.016
	model 3	0.849	0.003	0.753	0.001	0.355	0.004	0.790	0.010
	model 4	0.809	0.002	0.747	0.002	0.326	0.007	0.713	0.018
	model 5	0.844	0.005	0.754	0.001	0.359	0.002	0.800	0.007
TIL-10-asc	model 1	0.887	0.003	0.759	0.001	0.380	0.004	0.859	0.011
	model 2	0.874	0.005	0.758	0.002	0.373	0.009	0.839	0.024
	model 3	0.869	0.011	0.757	0.002	0.369	0.007	0.829	0.020
	model 4	0.880	0.004	0.758	0.000	0.373	0.002	0.841	0.005
	model 5	0.878	0.006	0.758	0.001	0.374	0.005	0.843	0.013
TIL-10-desc	model 1	0.887	0.001	0.760	0.000	0.384	0.002	0.869	0.006
	model 2	0.885	0.001	0.759	0.000	0.378	0.001	0.853	0.004
	model 3	0.877	0.005	0.758	0.001	0.373	0.003	0.839	0.009
	model 4	0.878	0.002	0.758	0.001	0.375	0.004	0.844	0.010
	model 5	0.880	0.002	0.759	0.001	0.377	0.003	0.851	0.007
TIL-1-asc	model 1	0.887	0.003	0.759	0.001	0.380	0.004	0.859	0.011
	model 2	0.873	0.005	0.758	0.002	0.373	0.009	0.838	0.024

	model 3	0.861	0.002	0.755	0.001	0.363	0.004	0.811	0.012
	model 4	0.852	0.007	0.753	0.001	0.352	0.006	0.782	0.015
	model 5	0.844	0.005	0.754	0.001	0.359	0.002	0.800	0.007
	model 1	0.884	0.001	0.759	0.000	0.378	0.002	0.853	0.006
	model 2	0.881	0.001	0.758	0.000	0.375	0.002	0.844	0.006
TIL-1-desc	model 3	0.877	0.005	0.758	0.001	0.373	0.003	0.839	0.009
	model 4	0.874	0.003	0.757	0.001	0.372	0.004	0.836	0.012
	model 5	0.874	0.003	0.758	0.001	0.373	0.003	0.840	0.007

이 연구를 통해, 개별 가상 기관에서 훈련된 모형과 개별 가상 기관의 테스트 세트는 모형 i 와 테스트 세트 j ,로 표기하였으며 i 와 j 는 기관의 번호를 의미한다.

첫째로, AUROC 의 전역 성능 점수는 전역 데이터 분포에 대한 모형 예측 성능을 의미한다 (그림 3-3).

CDS 모형은 가장 높은 전역 AUROC 점수를 달성하여 0.890 ± 0.000 이었으며 모든 다른 훈련 알고리즘보다 월등할 것으로 기대되었다.

지역 학습을 통해 각 가상 기관에서 훈련된 모형 간에 가장 높은 성능과 낮은 성능은 지역 학습 모형 1과 지역 학습 모형 4였으며 AUROC는 각각 0.884 ± 0.001 과 0.809 ± 0.002 였다. 데이터를 가장 많이 소유한 가상 기관으로 훈련한 지역 학습 모형 1 이 더 높은 전역 성능 점수를 보이는 경향을 보였다.

FL 모형은 CDS 모형에 비견할 정도인 0.885 ± 0.001 이었다.

TIL 모형은 전역 성능 점수에서 모든 지역 학습 모형을 능가하며 CDS와 FL 모형과 유사한 전역 점수를 보였다.

TIL-10-desc 모형에서 가장 높은 AUROC 와 가장 낮은 전역 AUROC 는 모형 1 의 0.887 ± 0.001 과 모형 4 의 0.877 ± 0.005 였다.

TIL-1-asc 와 TIL-1-desc (모형 1 과 5)에서 마지막 가상 기관에서 생성된 모형의 전역 AUROC 는 각각 0.887 ± 0.003 과 0.874 ± 0.003 로 상응하는 지역 학습 모형의 전역 성능 점수보다 높았다.

TIL-1-asc 과 TIL-1-desc (모형 5 와 1)에서 첫 가상 기관에서 생성된 모형의 전역 AUROC 는 각각 0.844 ± 0.004 와 0.884 ± 0.001 로 상응하는 지역 학습 모형의 전역 AUROC 와 유사하였다.

이는 TIL-1-asc 와 TIL-1-desc 로부터 생성된 첫 모형은 오로지 첫 가상 기관 데이터를 사용하여 훈련하기 때문으로 보인다. 다중 TIL 라운드 (TIL-10-asc 와 desc)는 전역 AUROC 에서 약간의 성능 향상을 보여주었다.

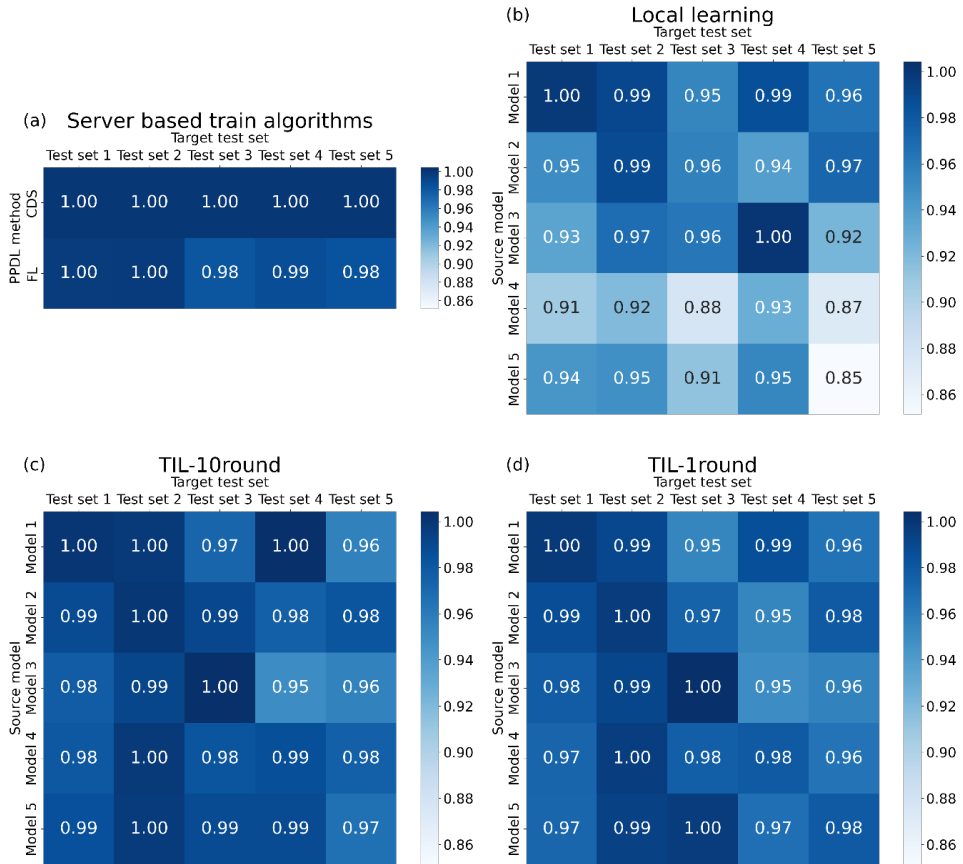


그림 3-4 혼합된 분포의 가상 기관 환경에서 훈련 알고리즘의 지역 성능. CDS 모형의 점수에 대해 AUROC 의 RE 가 측정되었다. 히트 맵 (heat map)의 각 행은 각 PPDL 훈련 알고리즘을 통해 훈련된 다른 모형의 의미한다. 각 열은 기관의 테스트 셋을 의미한다. (a) 서버 기반 PPDL 훈련 알고리즘의 성능. CDS 와 FL 을 통해 훈련된 모형의 RE, (b) 지역 학습 모형의 성능. (c) 가상 기관의 데이터 크기의 내림차순으로 10 라운드를 훈련한 TIL 모형의 성능. (d) 가상 기관의 데이터 크기의 내림차순으로 1 라운드를 훈련한 TIL 모형의 성능.

다음으로 지역 AUROC 는 개별 가상 기관의 데이터에 대한 개별 모형의 예측 성능을 비교하기 위해 측정되었다 (그림 3-4). CDS 를 통해 이를 수 있는 최고 점수는 개별 가상 기관에 따라 다르기 때문에 PPDL 훈련 알고리즘의 학습 능력을 평가하기 위해 CDS 모형의 점수에 대해 RE 로 지역 점수를 표현하였다. CDS 를 통해 학습된 모형은 모든 기관의 테스트 세트에 대하여 다른 모형과 비교했을 때 가장 높은 점수를 보였다 (그림 3-4 a 의 CDS 행).

데이터 수가 큰 기관에서 학습된 지역 학습 모형은 자신의 가상 기관 내의 테스트 데이터에 대해 높은 점수를 보였고 (그림 3-4 b 의 대각선 RE), 데이터 수가 가장 작은 지역 학습 모형 5 가 가장 낮은 RE 를 보였으며, 0.851 ± 0.027 의 AUROC 를 보였다 (그림 3-4 b 의 모형 5, 테스트 세트 5).

자신의 가상 기관 외의 외부 테스트 세트에 대한 지역 학습 모형의 점수 중 (그림 3-4, 대각선 외 RE), 테스트 세트 5 에 대한 모형 4 의 RE 가 가장 낮았으며 0.872 ± 0.031 의 AUROC 를 보였다 (그림 3-4 b 의 모형 4, 테스트 세트 5).

FL 모형은 CDS 모형과 유사한 RE 를 보였다.

표 3-4 (지역 성능) 전이반복학습 모형의 여러 평가 척도 결과.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

TIL-10 과 TIL-1 은 각각 10 라운드와 단일 라운드를 나타내며, asc 는 데이터가 적은 기관에서 많은 기관의 순서로 (오름차순), desc 는 데이터가 많은 기관에서 적은 기관의 순서로 (내림차순) 훈련한 것을 의미한다.

Specificity 를 0.750 으로 고정하여 계산하였다. AUROC 는 Area Under the Receiver Operating Characteristics curve 를, \pm SD 는 Standard Deviation 를 의미한다.

● 각 가상 기관에서 예측된 AUROC

	Metric	AUROC									
	Target test set	Test set 1	\pm SD	Test set 2	\pm SD	Test set 3	\pm SD	Test set 4	\pm SD	Test set 5	\pm SD
CDS	server	0.852	0.001	0.928	0.001	0.944	0.004	0.922	0.005	0.951	0.008
FL	server	0.849	0.002	0.926	0.001	0.927	0.009	0.910	0.005	0.936	0.011
LOCAL	model 1	0.850	0.001	0.920	0.001	0.901	0.016	0.909	0.007	0.918	0.014
	model 2	0.809	0.004	0.918	0.004	0.904	0.014	0.865	0.010	0.925	0.014
	model 3	0.796	0.003	0.899	0.006	0.909	0.009	0.922	0.013	0.878	0.013
	model 4	0.773	0.001	0.854	0.007	0.828	0.015	0.855	0.006	0.830	0.030
	model 5	0.804	0.005	0.880	0.008	0.863	0.007	0.879	0.019	0.810	0.026
TIL-10-asc	model 1	0.850	0.002	0.927	0.005	0.926	0.013	0.895	0.021	0.955	0.008
	model 2	0.827	0.005	0.926	0.004	0.947	0.010	0.890	0.022	0.946	0.011
	model 3	0.822	0.013	0.913	0.008	0.939	0.010	0.902	0.008	0.909	0.020
	model 4	0.837	0.006	0.920	0.004	0.932	0.007	0.912	0.012	0.947	0.011
	model 5	0.837	0.008	0.915	0.003	0.927	0.013	0.909	0.002	0.916	0.006
TIL-10-desc	model 1	0.852	0.001	0.926	0.002	0.919	0.010	0.925	0.005	0.910	0.011
	model 2	0.842	0.002	0.928	0.002	0.936	0.007	0.899	0.008	0.935	0.018
	model 3	0.833	0.007	0.920	0.003	0.948	0.007	0.875	0.019	0.910	0.009
	model 4	0.835	0.004	0.927	0.003	0.928	0.012	0.909	0.010	0.928	0.015
	model 5	0.839	0.004	0.926	0.001	0.934	0.016	0.911	0.014	0.919	0.017

TIL-1-asc	model 1	0.850	0.002	0.927	0.005	0.926	0.013	0.895	0.021	0.955	0.008
	model 2	0.825	0.004	0.925	0.003	0.947	0.009	0.881	0.026	0.950	0.005
	model 3	0.814	0.003	0.908	0.001	0.933	0.007	0.894	0.016	0.900	0.008
	model 4	0.809	0.004	0.894	0.006	0.909	0.021	0.902	0.004	0.891	0.030
	model 5	0.804	0.005	0.880	0.008	0.863	0.007	0.879	0.019	0.810	0.026
TIL-1-desc	model 1	0.850	0.001	0.920	0.001	0.901	0.016	0.909	0.007	0.918	0.014
	model 2	0.841	0.003	0.925	0.002	0.920	0.012	0.879	0.019	0.932	0.016
	model 3	0.833	0.007	0.920	0.003	0.948	0.007	0.875	0.019	0.910	0.009
	model 4	0.829	0.005	0.925	0.002	0.922	0.014	0.903	0.013	0.918	0.009
	model 5	0.828	0.006	0.923	0.002	0.941	0.012	0.891	0.017	0.932	0.016

● 각 가상 기관에서 예측된 Accuracy

	Metric	Accuracy									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.748	0.001	0.761	0.000	0.762	0.001	0.769	0.002	0.769	0.000
FL	server	0.748	0.002	0.760	0.000	0.759	0.001	0.766	0.005	0.765	0.004
LOCAL	model 1	0.749	0.001	0.761	0.000	0.758	0.001	0.764	0.002	0.762	0.000
	model 2	0.740	0.002	0.761	0.000	0.761	0.002	0.759	0.003	0.769	0.000
	model 3	0.739	0.001	0.760	0.000	0.759	0.002	0.768	0.000	0.768	0.003
	model 4	0.730	0.000	0.757	0.001	0.751	0.003	0.757	0.007	0.754	0.005
	model 5	0.743	0.001	0.756	0.003	0.755	0.001	0.761	0.009	0.749	0.004
TIL-10-asc	model 1	0.751	0.001	0.761	0.000	0.759	0.003	0.765	0.003	0.769	0.000
	model 2	0.745	0.001	0.761	0.000	0.761	0.002	0.765	0.006	0.769	0.000
	model 3	0.745	0.003	0.760	0.001	0.762	0.001	0.768	0.006	0.765	0.004
	model 4	0.747	0.002	0.760	0.000	0.761	0.002	0.769	0.007	0.769	0.000
	model 5	0.748	0.002	0.760	0.000	0.759	0.001	0.772	0.002	0.768	0.003
TIL-10-desc	model 1	0.750	0.001	0.761	0.000	0.758	0.003	0.769	0.005	0.760	0.003
	model 2	0.748	0.001	0.760	0.000	0.760	0.001	0.762	0.000	0.766	0.007
	model 3	0.748	0.002	0.760	0.000	0.762	0.001	0.762	0.000	0.766	0.004
	model 4	0.745	0.001	0.761	0.000	0.761	0.002	0.762	0.000	0.768	0.003
	model 5	0.747	0.002	0.761	0.000	0.761	0.003	0.766	0.003	0.765	0.004

TIL-1-asc	model 1	0.751	0.001	0.761	0.000	0.759	0.003	0.765	0.003	0.769	0.000
	model 2	0.744	0.001	0.761	0.000	0.760	0.001	0.761	0.005	0.769	0.000
	model 3	0.743	0.001	0.759	0.001	0.762	0.001	0.761	0.005	0.762	0.000
	model 4	0.741	0.001	0.759	0.000	0.760	0.002	0.769	0.002	0.760	0.008
	model 5	0.743	0.001	0.756	0.003	0.755	0.001	0.761	0.009	0.749	0.004
TIL-1-desc	model 1	0.749	0.001	0.761	0.000	0.758	0.001	0.764	0.002	0.762	0.000
	model 2	0.748	0.001	0.760	0.000	0.759	0.001	0.762	0.000	0.768	0.003
	model 3	0.748	0.002	0.760	0.000	0.762	0.001	0.762	0.000	0.766	0.004
	model 4	0.744	0.002	0.761	0.000	0.761	0.003	0.762	0.000	0.768	0.003
	model 5	0.744	0.002	0.761	0.000	0.763	0.001	0.765	0.003	0.766	0.004

● 각 가상 기관에서 예측된 f1

	Metric	f1									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.407	0.003	0.275	0.000	0.290	0.006	0.437	0.009	0.348	0.000
FL	server	0.406	0.006	0.273	0.001	0.279	0.006	0.424	0.019	0.326	0.020
LOCAL	model 1	0.409	0.003	0.275	0.000	0.269	0.007	0.415	0.010	0.311	0.000
	model 2	0.375	0.006	0.276	0.002	0.286	0.012	0.398	0.012	0.348	0.000
	model 3	0.369	0.004	0.269	0.002	0.274	0.008	0.432	0.000	0.340	0.016
	model 4	0.333	0.001	0.257	0.004	0.233	0.014	0.389	0.028	0.272	0.028
	model 5	0.386	0.004	0.251	0.013	0.257	0.007	0.406	0.036	0.249	0.022
TIL-10-asc	model 1	0.416	0.003	0.276	0.001	0.274	0.015	0.420	0.012	0.348	0.000
	model 2	0.395	0.005	0.275	0.002	0.286	0.008	0.419	0.024	0.348	0.000
	model 3	0.394	0.013	0.272	0.004	0.290	0.006	0.432	0.021	0.326	0.020
	model 4	0.403	0.006	0.272	0.001	0.288	0.010	0.436	0.028	0.348	0.000
	model 5	0.406	0.007	0.273	0.001	0.276	0.005	0.449	0.009	0.340	0.016
TIL-10-desc	model 1	0.413	0.004	0.275	0.000	0.271	0.016	0.436	0.018	0.303	0.017
	model 2	0.405	0.003	0.273	0.001	0.283	0.005	0.411	0.000	0.333	0.034
	model 3	0.404	0.008	0.272	0.000	0.293	0.006	0.411	0.000	0.333	0.020
	model 4	0.395	0.004	0.275	0.002	0.288	0.010	0.411	0.000	0.340	0.016
	model 5	0.400	0.007	0.274	0.002	0.286	0.014	0.424	0.012	0.326	0.020

TIL-1-asc	model 1	0.416	0.003	0.276	0.001	0.274	0.015	0.420	0.012	0.348	0.000
	model 2	0.391	0.005	0.276	0.001	0.283	0.005	0.406	0.018	0.348	0.000
	model 3	0.385	0.003	0.268	0.003	0.293	0.006	0.406	0.018	0.311	0.000
	model 4	0.376	0.002	0.267	0.003	0.283	0.013	0.437	0.009	0.303	0.042
	model 5	0.386	0.004	0.251	0.013	0.257	0.007	0.406	0.036	0.249	0.022
TIL-1-desc	model 1	0.409	0.003	0.275	0.000	0.269	0.007	0.415	0.010	0.311	0.000
	model 2	0.406	0.005	0.272	0.002	0.279	0.006	0.411	0.000	0.340	0.016
	model 3	0.404	0.008	0.272	0.000	0.293	0.006	0.411	0.000	0.333	0.020
	model 4	0.390	0.007	0.274	0.001	0.286	0.014	0.411	0.000	0.340	0.016
	model 5	0.388	0.008	0.275	0.002	0.295	0.005	0.420	0.012	0.333	0.020

● 각 가상 기관에서 예측된 Sensitivity

	Metric	Sensitivity									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.737	0.007	0.979	0.000	0.973	0.025	0.953	0.026	1.000	0.000
FL	server	0.736	0.013	0.973	0.006	0.927	0.025	0.918	0.053	0.925	0.068
LOCAL	model 1	0.741	0.006	0.979	0.000	0.891	0.025	0.894	0.026	0.875	0.000
	model 2	0.666	0.013	0.984	0.009	0.955	0.045	0.847	0.032	1.000	0.000
	model 3	0.653	0.008	0.955	0.009	0.909	0.032	0.941	0.000	0.975	0.056
	model 4	0.577	0.001	0.909	0.015	0.755	0.052	0.824	0.072	0.750	0.088
	model 5	0.690	0.010	0.882	0.054	0.845	0.025	0.871	0.097	0.675	0.068
TIL-10-asc	model 1	0.758	0.006	0.984	0.006	0.909	0.056	0.906	0.032	1.000	0.000
	model 2	0.710	0.012	0.981	0.009	0.955	0.032	0.906	0.067	1.000	0.000
	model 3	0.707	0.029	0.969	0.015	0.973	0.025	0.941	0.059	0.925	0.068
	model 4	0.728	0.013	0.967	0.005	0.964	0.038	0.953	0.077	1.000	0.000
	model 5	0.735	0.016	0.973	0.006	0.918	0.020	0.988	0.026	0.975	0.056
TIL-10-desc	model 1	0.752	0.009	0.979	0.000	0.900	0.059	0.953	0.049	0.850	0.056
	model 2	0.734	0.006	0.973	0.006	0.945	0.020	0.882	0.000	0.950	0.112
	model 3	0.731	0.017	0.969	0.000	0.982	0.025	0.882	0.000	0.950	0.068
	model 4	0.710	0.010	0.979	0.010	0.964	0.038	0.882	0.000	0.975	0.056
	model 5	0.722	0.016	0.975	0.009	0.955	0.056	0.918	0.032	0.925	0.068

TIL-1-asc	model 1	0.758	0.006	0.984	0.006	0.909	0.056	0.906	0.032	1.000	0.000
	model 2	0.702	0.011	0.984	0.006	0.945	0.020	0.871	0.049	1.000	0.000
	model 3	0.688	0.007	0.951	0.013	0.982	0.025	0.871	0.049	0.875	0.000
	model 4	0.669	0.005	0.948	0.010	0.945	0.050	0.953	0.026	0.850	0.137
	model 5	0.690	0.010	0.882	0.054	0.845	0.025	0.871	0.097	0.675	0.068
TIL-1-desc	model 1	0.741	0.006	0.979	0.000	0.891	0.025	0.894	0.026	0.875	0.000
	model 2	0.735	0.010	0.969	0.007	0.927	0.025	0.882	0.000	0.975	0.056
	model 3	0.731	0.017	0.969	0.000	0.982	0.025	0.882	0.000	0.950	0.068
	model 4	0.699	0.016	0.975	0.006	0.955	0.056	0.882	0.000	0.975	0.056
	model 5	0.696	0.019	0.981	0.009	0.991	0.020	0.906	0.032	0.950	0.068

TIL 모형은 지역 학습 모형에 비해 향상된 지역 성능 점수를 보였으며 CDS와 FL 모형과 유사한 지역 성능 점수를 보였다.

상응하는 지역 학습 모형에 대비하여, TIL-1-desc의 모형 5는 자신의 가상 기관 내의 테스트 세트 5에 대해 0.980 ± 0.017 (그림 3-4 d의 모형 5, 테스트 세트 5)을 보였다. 게다가, 테스트 세트 5에 대해 TIL-1-desc에서 모형 4는 0.965 ± 0.001 (그림 3-4 d의 모형 4, 테스트 세트 5)였으며, 상응하는 지역 학습 모형에 비교하여 외부 예측 점수의 향상을 보였다.

앞서 언급된 지역 점수 향상은 단일 TIL 라운드에서 비롯되었으며 앞에서 훈련된 다른 가상 기관에서 온 지식 전이의 효과를 보여준다.

TIL-10-desc 모형 중, 가장 낮은 지역 성능 점수는 0.950 ± 0.021 (그림 3-4 c의 모형 2, 테스트 세트 4)였으며 이는 지역 학습 모형의 가장 낮은 지역 성능 점수보다 훨씬 높다. TIL-10-desc는 TIL-1-desc에 비해 약간 더 나은 지역 점수를 보였으며 이는 모형 1, 4, 5 (그림 3-4 c, d)에서 보였다.

모형 1의 지역 성능 점수에 대하여 TIL-10-desc는 TIL-1-desc보다 나은 지역 성능 점수를 보였는데, TIL-10-desc에서 모형 1은 10라운드를 모두 돈 모든 가상 기관에서 온 데이터를 사용하여 훈련되었던 반면, TIL-1-desc에서 온 모형 1은 가상 기관 1로 훈련을 시작하여 가상 기관 1의 데이터만 사용하였기 때문일 것으로 사료 된다.

그림 3-4 d에서 지역 성능 점수 히트 맵의 대각선 RE의 위쪽 RE와 대각선 RE의 아래쪽 RE가 각각 TIL-1-asc과 TIL-1-desc로 해석되어야 하는데, 이는 각 행렬의 RE는 순차적 기관 방문을 통해 모형이 훈련된 테스트 세트에 대한 모형 성능이기 때문이다. 그러므로 TIL-1-asc와 TIL-

1-desc (모형 5와 1)의 첫 모형은 모든 테스트 세트에 걸쳐 상응하는 지역 학습 모형과 유사한 지역 성능 점수를 보인다.

제 5 절 혼합 동질 환경 하에서 TIL 의 성능

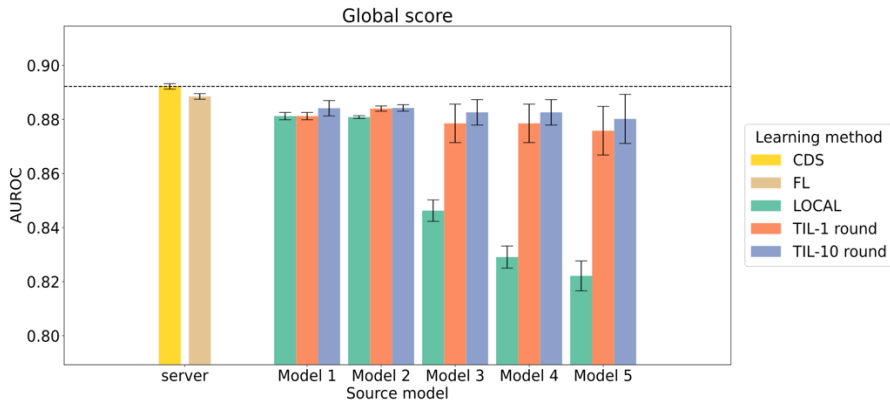


그림 3-5 혼합 동질 분포 환경에서 훈련 알고리즘의 전역 성능.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

점선의 수평선은 CDS 의 성능을 나타내며 이는 모형이 개인정보보호 환경에서 한 모형이 최대 성능을 낼 수 있는 상한이다.

제시된 TIL 모형은 각 라운드마다 기관 1에서 5 까지 방문하며 훈련된다.

막대 그래프는 다운 샘플된 데이터와 모형 초기화의 무작위 반복으로 생성된 평균 AUROC를 의미한다. 오차 막대는 95% 신뢰 구간을 의미한다.

앞 절의 혼합 분포의 가상 기관의 데이터의 수는 동일하지만, 분포를 동일하게 각 가상 기관을 분할하여 homogeneous 한 가상 기관의 상황을 실험하였다.

전역 성능의 AUROC 는 CDS 에서 0.892 ± 0.001 을 보여 이 실험에서 상한이었으며, 앞의 혼합 분포의 가상 기관 환경과 유사한 결과를 보였다.

지역 학습을 통해 각 가상 기관에서 훈련된 모형 간에 가장 높은 성능과 낮은 성능은 지역 학습 모형 1, 2 와 지역 학습 모형 5 였으며 AUROC 는 각각 0.881 ± 0.002 과 0.809 ± 0.002 였으며 앞의 결과와 큰 차이가 없었다. 데이터를 가장 많이 소유한 가상 기관으로 훈련한 지역 학습 모형 1, 2 가 더 높은 전역 성능 점수를 보이는 경향은 앞의 결과와 같았다.

FL 모형은 CDS 모형에 비견할 정도인 0.888 ± 0.001 이었다.

TIL 모형은 전역 성능 점수에서 모든 지역 학습 모형을 능가하며 CDS 와 FL 모형과 유사한 전역 점수를 보였다.

FL 과 TIL 중 혼합 분포의 가상 기관 환경과 동질 분포의 가상 기관 환경에서 AUROC 를 비롯한 전역 성능 점수의 부침을 비교해 보면, FL 의 경우 동질 분포의 가상 기관에서 0.888 ± 0.001 , 혼합 분포의 가상 기관 환경에서 0.885 ± 0.001 로 0.003 의 AUROC 의 하강이 관찰 되었지만 큰 차이를 두기는 어려울 것으로 사료된다.

또한, TIL 의 경우 동질 분포의 가상 기관에서 가장 성능이 좋았던 TIL-10-asc 의 모델 1 의 0.885 ± 0.002 와 혼합 분포의 가상 기관에서 TIL-10-asc 모델 1 을 비교해 보면 0.887 ± 0.003 로 혼합 분포에서 성능이 다소 좋았지만 역시 큰 차이를 두기 어려울 것으로 사료된다.

표 3-5 (전역 성능) 전이반복학습 모형의 여러 평가 척도 결과.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

TIL-10 과 TIL-1 은 각각 10 라운드와 단일 라운드를 의미한다.

Specificity 를 0.750 으로 고정하여 계산하였다. AUROC 는 Area Under the Receiver Operating Characteristics curve 를, \pm SD 는 Standard Deviation 를 의미한다.

	Metric	AUROC	\pm SD	Accuracy	\pm SD	f1	\pm SD	Sensitivity	\pm SD
CDS	server	0.892	0.001	0.761	0.000	0.387	0.002	0.877	0.006
FL	server	0.888	0.001	0.760	0.000	0.383	0.002	0.867	0.005
LOCAL	model 1	0.881	0.002	0.759	0.001	0.377	0.002	0.852	0.006
	model 2	0.881	0.001	0.758	0.001	0.373	0.002	0.841	0.006
	model 3	0.846	0.005	0.751	0.001	0.346	0.004	0.766	0.012
	model 4	0.829	0.006	0.750	0.002	0.341	0.011	0.751	0.029
	model 5	0.822	0.008	0.750	0.001	0.338	0.003	0.746	0.008
TIL-10-asc	model 1	0.885	0.002	0.759	0.001	0.380	0.004	0.857	0.010
	model 2	0.883	0.002	0.759	0.001	0.379	0.003	0.857	0.008
	model 3	0.871	0.005	0.757	0.001	0.369	0.003	0.828	0.009
	model 4	0.882	0.005	0.759	0.001	0.377	0.003	0.850	0.007
	model 5	0.879	0.007	0.758	0.001	0.376	0.002	0.846	0.006
TIL-10-desc	model 1	0.884	0.004	0.759	0.001	0.379	0.003	0.856	0.008
	model 2	0.884	0.002	0.759	0.001	0.379	0.003	0.855	0.007
	model 3	0.883	0.006	0.759	0.001	0.378	0.004	0.854	0.012
	model 4	0.883	0.006	0.759	0.001	0.378	0.004	0.854	0.012
	model 5	0.880	0.012	0.758	0.003	0.374	0.012	0.843	0.033
TIL-1-asc	model 1	0.885	0.002	0.759	0.001	0.380	0.004	0.857	0.010
	model 2	0.881	0.004	0.759	0.001	0.378	0.005	0.852	0.014
	model 3	0.855	0.004	0.753	0.001	0.353	0.004	0.786	0.010
	model 4	0.840	0.003	0.754	0.000	0.356	0.002	0.792	0.006
	model 5	0.822	0.008	0.750	0.001	0.338	0.003	0.746	0.008

	model 1	0.881	0.002	0.759	0.001	0.377	0.002	0.852	0.006
	model 2	0.884	0.001	0.759	0.001	0.379	0.003	0.856	0.009
TIL-1-desc	model 3	0.879	0.010	0.758	0.002	0.375	0.010	0.844	0.028
	model 4	0.879	0.010	0.758	0.002	0.375	0.010	0.844	0.028
	model 5	0.876	0.012	0.757	0.003	0.372	0.013	0.836	0.035

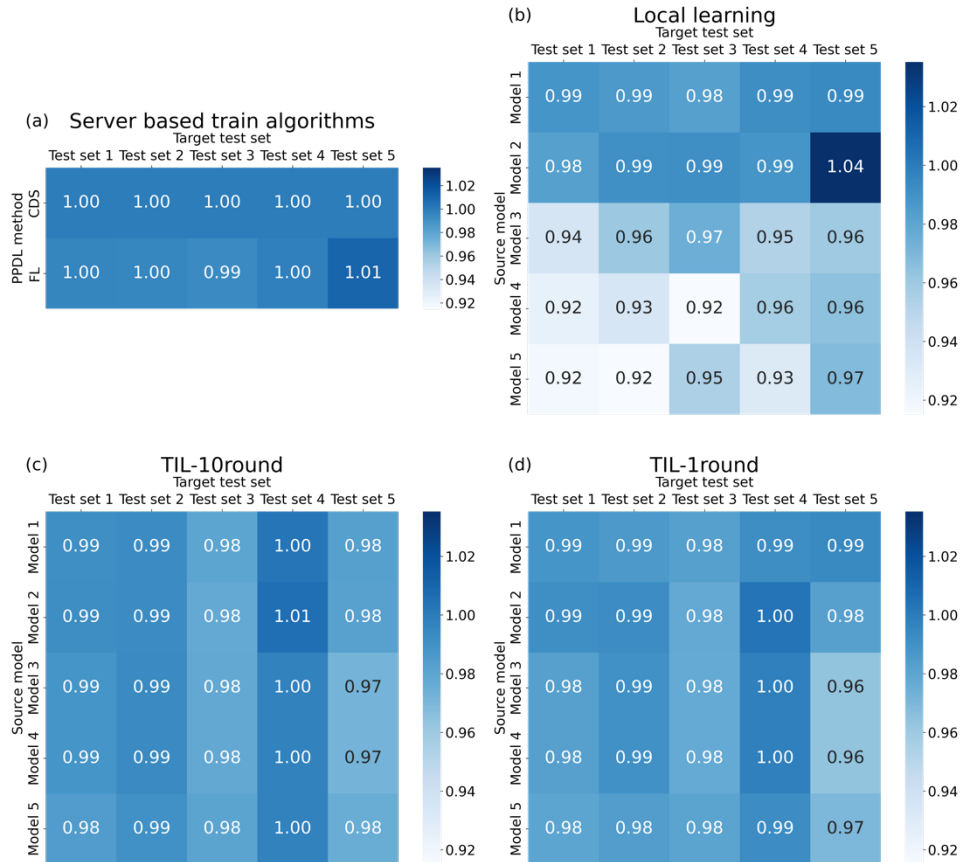


그림 3-6 혼합 동질 분포 환경에서 훈련 알고리즘의 지역 성능.

CDS 모형의 점수에 대해 AUROC 의 RE 가 측정되었다. 히트 맵 (heat map)의 각 행은 각 PPDL 훈련 알고리즘을 통해 훈련된 다른 모형을 의미한다. 각 열은 기관의 테스트 셋을 의미한다. (a) 서버 기반 PPDL 훈련 알고리즘의 성능. CDS 와 FL 을 통해 훈련된 모형의 RE, (b) 지역 학습 모형의 성능. (c) 가상 기관의 데이터 크기의 내림차순으로 10 라운드를 훈련한 TIL 모형의 성능. (d) 가상 기관의 데이터 크기의 내림차순으로 1 라운드를 훈련한 TIL 모형의 성능.

다음으로 지역 AUROC 는 개별 가상 기관의 데이터에 대한 개별 모형의 예측 성능을 비교하기 위해 측정되었다 (그림 3-6). CDS 를 통해 이를 수 있는 최고 점수는 개별 가상 기관에 따라 다르기 때문에 PPDL 훈련 알고리즘의 학습 능력을 평가하기 위해 CDS 모형의 점수에 대해 RE 로 지역 점수를 표현하였다 (그림 3-6 a 의 CDS 행).

FL 모형은 CDS 모형과 유사한 RE 를 보였다.

분포가 동질한 특성으로 각 가상 기관의 자신의 테스트 셋을 검정한 결과와 타 기관을 예측한 결과의 성능 차가 크게 없었으며, 이는 TIL 에서도 유사하였다.

표 3-6 (지역 성능) 전이반복학습 모형의 여러 평가 척도 결과.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

TIL-10 과 TIL-1 은 각각 10 라운드와 단일 라운드를 의미한다.

Specificity 를 0.750 으로 고정하여 계산하였다. AUROC 는 Area Under the Receiver Operating Characteristics curve 를, \pm SD 는 Standard Deviation 를 의미한다.

● 각 가상 기관에서 예측된 AUROC

	Metric	AUROC									
	Target test set	Test set 1	\pm SD	Test set 2	\pm SD	Test set 3	\pm SD	Test set 4	\pm SD	Test set 5	\pm SD
CDS	server	0.890	0.002	0.886	0.001	0.915	0.004	0.916	0.002	0.911	0.003
FL	server	0.887	0.001	0.882	0.001	0.908	0.006	0.914	0.003	0.921	0.012
LOCAL	model 1	0.880	0.003	0.874	0.000	0.899	0.006	0.908	0.003	0.905	0.011
	model 2	0.875	0.002	0.878	0.001	0.907	0.002	0.905	0.003	0.943	0.006
	model 3	0.833	0.007	0.852	0.004	0.891	0.003	0.873	0.005	0.874	0.006
	model 4	0.823	0.006	0.829	0.006	0.837	0.016	0.877	0.010	0.879	0.018
	model 5	0.817	0.006	0.811	0.010	0.873	0.005	0.853	0.014	0.882	0.017
TIL-10-asc	model 1	0.880	0.004	0.882	0.003	0.921	0.004	0.903	0.009	0.893	0.004
	model 2	0.876	0.004	0.883	0.003	0.918	0.005	0.913	0.005	0.879	0.016
	model 3	0.864	0.005	0.873	0.006	0.896	0.008	0.904	0.005	0.861	0.008
	model 4	0.878	0.006	0.877	0.007	0.915	0.012	0.904	0.010	0.891	0.008
	model 5	0.875	0.005	0.875	0.010	0.916	0.011	0.905	0.008	0.885	0.017
TIL-10-desc	model 1	0.882	0.004	0.880	0.005	0.896	0.005	0.918	0.005	0.895	0.005
	model 2	0.883	0.002	0.880	0.003	0.894	0.006	0.921	0.002	0.895	0.006
	model 3	0.880	0.008	0.880	0.004	0.895	0.008	0.914	0.010	0.885	0.023
	model 4	0.880	0.008	0.880	0.004	0.895	0.008	0.914	0.010	0.885	0.023
	model 5	0.877	0.016	0.878	0.010	0.897	0.005	0.914	0.015	0.889	0.015

TIL-1-asc	model 1	0.880	0.004	0.882	0.003	0.921	0.004	0.903	0.009	0.893	0.004
	model 2	0.874	0.005	0.882	0.004	0.919	0.005	0.901	0.012	0.893	0.016
	model 3	0.845	0.005	0.858	0.004	0.897	0.003	0.880	0.019	0.878	0.016
	model 4	0.837	0.003	0.831	0.004	0.874	0.008	0.863	0.018	0.896	0.013
	model 5	0.817	0.006	0.811	0.010	0.873	0.005	0.853	0.014	0.882	0.017
TIL-1-desc	model 1	0.880	0.003	0.874	0.000	0.899	0.006	0.908	0.003	0.905	0.011
	model 2	0.883	0.002	0.879	0.002	0.894	0.006	0.919	0.005	0.895	0.006
	model 3	0.874	0.014	0.878	0.004	0.894	0.007	0.914	0.008	0.878	0.024
	model 4	0.874	0.014	0.878	0.004	0.894	0.007	0.914	0.008	0.878	0.024
	model 5	0.872	0.016	0.873	0.010	0.897	0.005	0.909	0.014	0.883	0.016

● 각 가상 기관에서 예측된 Accuracy

	Metric	Accuracy									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.759	0.001	0.759	0.001	0.771	0.001	0.771	0.005	0.773	0.000
FL	server	0.759	0.001	0.757	0.001	0.770	0.001	0.769	0.003	0.773	0.000
LOCAL	model 1	0.758	0.001	0.757	0.001	0.769	0.002	0.769	0.005	0.773	0.000
	model 2	0.756	0.001	0.757	0.000	0.770	0.001	0.764	0.003	0.773	0.000
	model 3	0.748	0.001	0.753	0.001	0.765	0.003	0.756	0.004	0.762	0.004
	model 4	0.749	0.002	0.749	0.003	0.751	0.003	0.763	0.007	0.765	0.005
	model 5	0.749	0.001	0.746	0.001	0.757	0.002	0.758	0.003	0.761	0.004
TIL-10-asc	model 1	0.757	0.001	0.758	0.001	0.771	0.001	0.767	0.000	0.773	0.000
	model 2	0.758	0.002	0.758	0.001	0.772	0.001	0.767	0.000	0.770	0.004
	model 3	0.754	0.002	0.758	0.001	0.766	0.002	0.763	0.003	0.756	0.010
	model 4	0.758	0.001	0.757	0.001	0.769	0.004	0.767	0.000	0.771	0.003
	model 5	0.757	0.001	0.757	0.001	0.769	0.005	0.766	0.002	0.770	0.004
TIL-10-desc	model 1	0.758	0.001	0.758	0.001	0.767	0.002	0.774	0.003	0.773	0.000
	model 2	0.757	0.001	0.758	0.001	0.767	0.002	0.773	0.005	0.773	0.000
	model 3	0.757	0.002	0.758	0.001	0.767	0.002	0.770	0.006	0.768	0.010
	model 4	0.757	0.002	0.758	0.001	0.767	0.002	0.770	0.006	0.768	0.010
	model 5	0.756	0.004	0.757	0.002	0.767	0.002	0.772	0.004	0.770	0.007

TIL-1-asc	model 1	0.757	0.001	0.758	0.001	0.771	0.001	0.767	0.000	0.773	0.000
	model 2	0.757	0.001	0.757	0.001	0.771	0.001	0.764	0.003	0.773	0.000
	model 3	0.751	0.002	0.753	0.001	0.766	0.003	0.760	0.006	0.762	0.004
	model 4	0.754	0.001	0.751	0.000	0.762	0.002	0.763	0.005	0.767	0.003
	model 5	0.749	0.001	0.746	0.001	0.757	0.002	0.758	0.003	0.761	0.004
TIL-1-desc	model 1	0.758	0.001	0.757	0.001	0.769	0.002	0.769	0.005	0.773	0.000
	model 2	0.758	0.002	0.758	0.001	0.767	0.002	0.772	0.006	0.773	0.000
	model 3	0.756	0.004	0.757	0.001	0.766	0.004	0.771	0.005	0.767	0.014
	model 4	0.756	0.004	0.757	0.001	0.766	0.004	0.771	0.005	0.767	0.014
	model 5	0.755	0.004	0.757	0.002	0.766	0.003	0.769	0.006	0.767	0.008

● 각 가상 기관에서 예측된 f1

	Metric	f1									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.377	0.004	0.372	0.003	0.434	0.004	0.463	0.017	0.444	0.000
FL	server	0.379	0.003	0.365	0.002	0.432	0.005	0.455	0.011	0.444	0.000
LOCAL	model 1	0.373	0.003	0.364	0.003	0.427	0.006	0.455	0.018	0.444	0.000
	model 2	0.364	0.005	0.365	0.001	0.432	0.005	0.439	0.011	0.444	0.000
	model 3	0.331	0.004	0.345	0.005	0.411	0.012	0.405	0.015	0.403	0.017
	model 4	0.335	0.009	0.329	0.015	0.357	0.011	0.435	0.028	0.415	0.021
	model 5	0.335	0.004	0.316	0.003	0.381	0.007	0.414	0.012	0.397	0.017
TIL-10-asc	model 1	0.371	0.004	0.367	0.003	0.437	0.004	0.447	0.000	0.444	0.000
	model 2	0.371	0.007	0.367	0.002	0.439	0.005	0.447	0.000	0.433	0.016
	model 3	0.358	0.007	0.366	0.006	0.418	0.009	0.435	0.011	0.378	0.040
	model 4	0.371	0.006	0.362	0.006	0.429	0.015	0.447	0.000	0.439	0.013
	model 5	0.370	0.003	0.364	0.005	0.429	0.019	0.443	0.009	0.433	0.016
TIL-10-desc	model 1	0.372	0.005	0.367	0.006	0.422	0.008	0.475	0.011	0.444	0.000
	model 2	0.371	0.006	0.367	0.005	0.420	0.010	0.471	0.017	0.444	0.000
	model 3	0.369	0.010	0.369	0.005	0.420	0.010	0.459	0.023	0.426	0.041
	model 4	0.369	0.010	0.369	0.005	0.420	0.010	0.459	0.023	0.426	0.041
	model 5	0.364	0.018	0.365	0.009	0.420	0.010	0.467	0.014	0.432	0.027

TIL-1-asc	model 1	0.371	0.004	0.367	0.003	0.437	0.004	0.447	0.000	0.444	0.000
	model 2	0.368	0.006	0.365	0.004	0.437	0.004	0.439	0.011	0.444	0.000
	model 3	0.345	0.007	0.347	0.005	0.418	0.011	0.422	0.023	0.403	0.017
	model 4	0.354	0.006	0.336	0.002	0.400	0.006	0.435	0.019	0.421	0.013
	model 5	0.335	0.004	0.316	0.003	0.381	0.007	0.414	0.012	0.397	0.017
TIL-1-desc	model 1	0.373	0.003	0.364	0.003	0.427	0.006	0.455	0.018	0.444	0.000
	model 2	0.372	0.007	0.366	0.004	0.420	0.010	0.467	0.020	0.444	0.000
	model 3	0.365	0.015	0.363	0.006	0.416	0.016	0.463	0.017	0.420	0.056
	model 4	0.365	0.015	0.363	0.006	0.416	0.016	0.463	0.017	0.420	0.056
	model 5	0.361	0.019	0.362	0.007	0.416	0.013	0.455	0.023	0.421	0.033

● 각 가상 기관에서 예측된 Sensitivity

	Metric	Sensitivity									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.853	0.010	0.857	0.007	0.970	0.011	0.937	0.044	1.000	0.000
FL	server	0.857	0.007	0.838	0.006	0.965	0.014	0.916	0.029	1.000	0.000
LOCAL	model 1	0.840	0.009	0.834	0.009	0.950	0.018	0.916	0.047	1.000	0.000
	model 2	0.817	0.013	0.839	0.004	0.965	0.014	0.874	0.029	1.000	0.000
	model 3	0.726	0.012	0.783	0.012	0.905	0.033	0.789	0.037	0.883	0.046
	model 4	0.737	0.024	0.739	0.039	0.760	0.029	0.863	0.071	0.917	0.059
	model 5	0.739	0.010	0.705	0.009	0.825	0.018	0.811	0.029	0.867	0.046
TIL-10-asc	model 1	0.835	0.012	0.843	0.009	0.980	0.011	0.895	0.000	1.000	0.000
	model 2	0.836	0.018	0.844	0.006	0.985	0.014	0.895	0.000	0.967	0.046
	model 3	0.799	0.018	0.840	0.017	0.925	0.025	0.863	0.029	0.817	0.109
	model 4	0.836	0.017	0.830	0.016	0.955	0.041	0.895	0.000	0.983	0.037
	model 5	0.832	0.007	0.836	0.013	0.955	0.054	0.884	0.024	0.967	0.046
TIL-10-desc	model 1	0.838	0.014	0.845	0.016	0.935	0.022	0.968	0.029	1.000	0.000
	model 2	0.835	0.016	0.843	0.013	0.930	0.027	0.958	0.044	1.000	0.000
	model 3	0.830	0.027	0.849	0.015	0.930	0.027	0.926	0.060	0.950	0.112
	model 4	0.830	0.027	0.849	0.015	0.930	0.027	0.926	0.060	0.950	0.112
	model 5	0.817	0.049	0.839	0.024	0.930	0.027	0.947	0.037	0.967	0.075

TIL-1-asc	model 1	0.835	0.012	0.843	0.009	0.980	0.011	0.895	0.000	1.000	0.000
	model 2	0.827	0.017	0.838	0.011	0.980	0.011	0.874	0.029	1.000	0.000
	model 3	0.765	0.019	0.789	0.014	0.925	0.031	0.832	0.058	0.883	0.046
	model 4	0.790	0.016	0.757	0.005	0.875	0.018	0.863	0.047	0.933	0.037
	model 5	0.739	0.010	0.705	0.009	0.825	0.018	0.811	0.029	0.867	0.046
TIL-1-desc	model 1	0.840	0.009	0.834	0.009	0.950	0.018	0.916	0.047	1.000	0.000
	model 2	0.837	0.018	0.841	0.011	0.930	0.027	0.947	0.053	1.000	0.000
	model 3	0.819	0.042	0.833	0.016	0.920	0.045	0.937	0.044	0.933	0.149
	model 4	0.819	0.042	0.833	0.016	0.920	0.045	0.937	0.044	0.933	0.149
	model 5	0.808	0.050	0.829	0.020	0.920	0.037	0.916	0.060	0.933	0.091

제 6 절 작은 기관 환경 하에서 TIL 의 성능

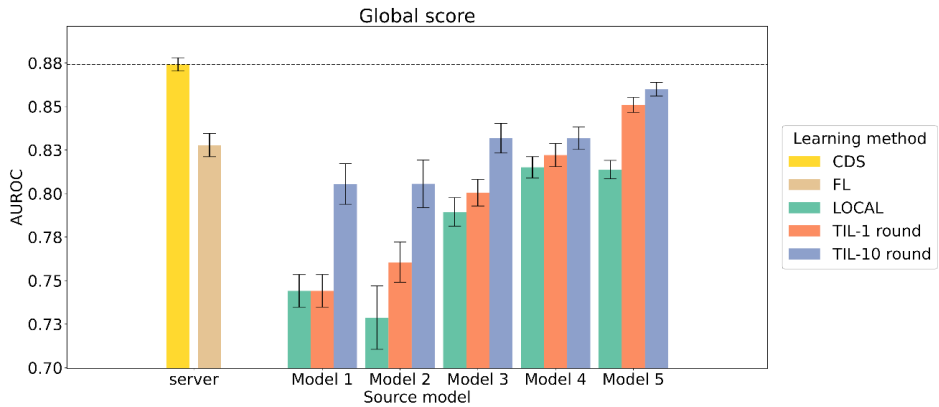


그림 3-7 작은 기관 환경에서 훈련 알고리즘의 전역 성능.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

점선의 수평선은 CDS 의 성능을 나타내며 이는 모형이 개인정보보호 환경에서 한 모형이 최대 성능을 낼 수 있는 상한이다.

제시된 TIL 모형은 각 라운드마다 기관 1 에서 5 까지 방문하며 훈련된다.

막대 그래프는 다운 샘플된 데이터와 모형 초기화의 무작위 반복으로 생성된 평균 AUROC 를 의미한다. 오차 막대는 95% 신뢰 구간을 의미한다.

데이터가 작은 기관 사이에서 TIL 의 성능을 평가하기 위해 작은 기관의 환경에서 동일한 PPDL 실험이 수행되었다. 개별 가상 기관 데이터는 가장 작은 기관 5 의 데이터에 맞춰 다운 샘플 되었고 여기에는 866 명의 환자가 포함되었다 (표 3-1). 작은 기관의 환경에서도 혼합 분포의 가상 기관 환경의 PPDL 모형과 유사한 성능 패턴이 역시 측정되었고 작은 기관들의 협동으로 TIL 의 효율성이 높아지는 것을 볼 수 있었다.

CDS 모형의 전역 AUROC 는 0.874 ± 0.017 이었다. 지역 학습을 통해 훈련된 개별 가상 기관의 모형 중, 가장 높은 전역 성능 점수와 가장 낮은 전역 성능 점수는 모형 4 와 모형 2 에서 보였고, 각각 AUROC 가 0.815 ± 0.026 과 0.729 ± 0.079 였다.

지역 학습 모형은 혼합 분포의 가상 기관 환경에 비해 작은 가상 기관 데이터 크기 때문에 혼합 분포의 CDS 모형보다 훨씬 낮은 성능을 보였다.

FL 의 전역 AUROC 는 0.828 ± 0.029 로 CDS 의 전역 성능에 못 미치는 결과를 보였다.

표 3-7 (전역 성능) 전이반복학습 모형의 여러 평가 척도 결과.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

TIL-10 과 TIL-1 은 각각 10 라운드와 단일 라운드를 의미한다.

Specificity를 0.750으로 고정하여 계산하였다. AUROC는 Area Under the Receiver Operating Characteristics curve 를, \pm SD 는 Standard Deviation 를 의미한다.

	Metric	AUROC	\pm SD	Accuracy	\pm SD	f1	\pm SD	Sensitivity	\pm SD
CDS	server	0.874	0.017	0.757	0.004	0.329	0.019	0.826	0.056
FL	server	0.828	0.029	0.753	0.005	0.311	0.023	0.773	0.067
LOCAL	model 1	0.744	0.040	0.745	0.006	0.275	0.029	0.669	0.081
	model 2	0.729	0.079	0.741	0.010	0.253	0.051	0.610	0.138
	model 3	0.789	0.035	0.749	0.005	0.291	0.025	0.714	0.070
	model 4	0.815	0.026	0.752	0.005	0.309	0.023	0.767	0.068
	model 5	0.814	0.023	0.751	0.004	0.305	0.018	0.754	0.051
TIL-10	model 1	0.805	0.050	0.752	0.006	0.306	0.030	0.759	0.086
	model 2	0.806	0.059	0.751	0.007	0.301	0.034	0.745	0.097
	model 3	0.832	0.037	0.754	0.005	0.315	0.025	0.785	0.073
	model 4	0.832	0.028	0.753	0.004	0.314	0.018	0.781	0.053
	model 5	0.860	0.017	0.757	0.003	0.331	0.015	0.831	0.044
TIL-1	model 1	0.744	0.040	0.745	0.006	0.275	0.029	0.669	0.081
	model 2	0.761	0.050	0.745	0.005	0.274	0.025	0.666	0.070
	model 3	0.801	0.033	0.750	0.005	0.298	0.024	0.736	0.069
	model 4	0.822	0.029	0.753	0.004	0.311	0.018	0.773	0.052
	model 5	0.851	0.019	0.756	0.003	0.328	0.016	0.822	0.047

TIL-10 모형은 모든 지역 학습 모형에서 일등했고 전역 성능 점수에 대해 FL 과 CDS 에 필적할 점수를 보였다. 가장 높은 AUROC 와 가장 낮은 AUROC 는 TIL-10 모형 (모형 5 와 1)에서 각각 0.860 ± 0.017 , 0.805 ± 0.050 이었고, CDS 와 비견할만 하였고 FL 을 능가하는 AUROC 를 보였다.

TIL-1 에서 모형 1 의 AUROC 는 0.744 ± 0.040 이었고 상응하는 지역 학습 모형의 AUROC 와 유사하였다. 그러나 TIL 의 단일 라운드 후의 마지막 모형 (모형 5)의 AUROC 는 0.851 ± 0.019 였고, 대응하는 지역 학습 모형 보다 0.814 ± 0.023 보다 나아졌다.

TIL-10 모형은 TIL-1 모형에 비해 여러 라운드를 돌아 훈련되어 더 나은 성능을 보여 준다고 할 수 있다.

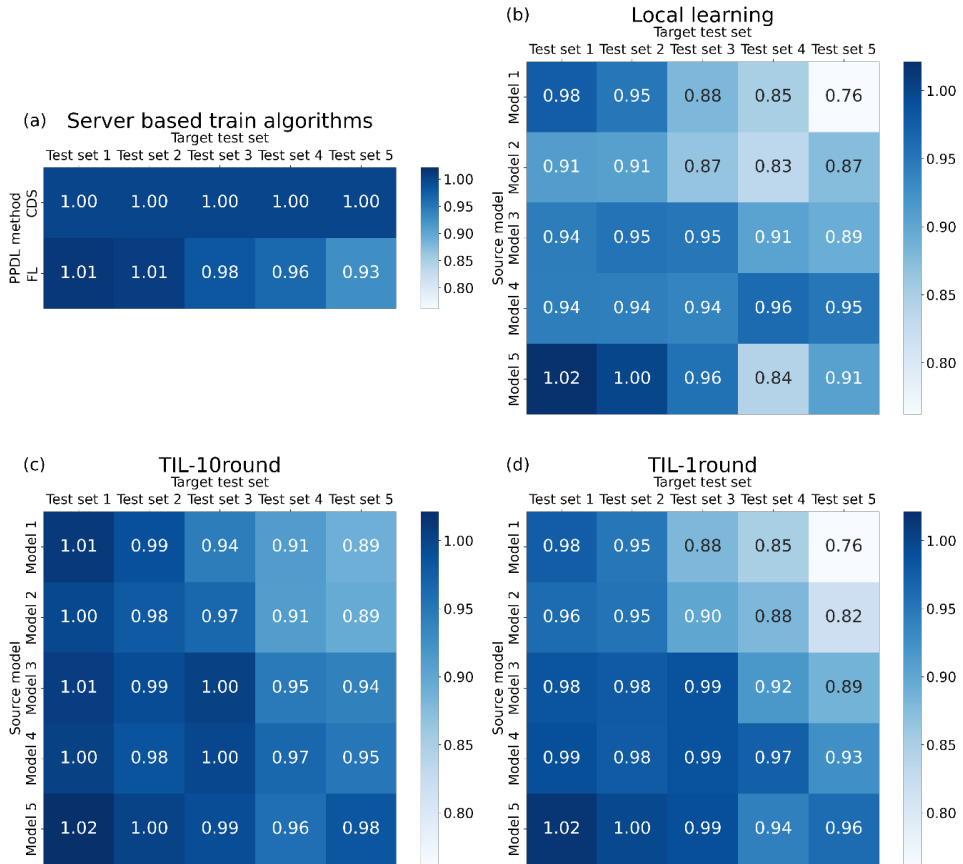


그림 3-8 작은 기관 환경에서 훈련 알고리즘의 지역 성능.

CDS 모형의 점수에 대해 AUROC 의 RE 가 측정되었다. 히트 맵 (heat map)의 각 행은 각 PPDL 훈련 알고리즘을 통해 훈련된 다른 모형을 의미한다. 각 열은 기관의 테스트 셋을 의미한다. (a) 서버 기반 PPDL 훈련 알고리즘의 성능. CDS 와 FL 을 통해 훈련된 모형의 RE, (b) 지역 학습 모형의 성능. (c) 가상 기관의 데이터 크기의 내림차순으로 10 라운드를 훈련한 TIL 모형의 성능. (d) 가상 기관의 데이터 크기의 내림차순으로 1 라운드를 훈련한 TIL 모형의 성능.

지역 AUROC 점수는 작은 가상 기관 환경에서 측정되었다 (그림 3-8). 지역 성능 점수는 CDS 모형의 AUROC 에 대해 RE 로 표시되었다. 가장 낮은 지역 학습 모형의 RE 는 자신의 가상 기관의 테스트 세트로 측정되었으며 0.907 ± 0.066 (그림 3-8 b 의 모형 5, 테스트셋 5)였고 외부 기관 테스트셋으로 시험된 가장 낮은 지역 학습 모형은 0.762 ± 0.026 (그림 3-8 b 의 모형 1, 테스트셋 5)이었다.

FL 모형은 CDS 모형과 유사한 지역 RE 를 보였으나 혼합 분포의 가상 기관 환경의 FL 의 RE 보다 약간 낮았다.

표 3-8 (지역 성능) 전이반복학습 모형의 여러 평가 척도 결과.

CDS, FL, TIL, LOCAL 은 각각 중앙 딥러닝 모형인 collaborative data sharing, 연합 학습 (federated learning), 전이반복학습 (transfer iterative learning), 지역 학습 (local learning)을 의미한다.

TIL-10 과 TIL-1 은 각각 10 라운드와 단일 라운드를 의미한다.

Specificity 를 0.750 으로 고정하여 계산하였다. AUROC 는 Area Under the Receiver Operating Characteristics curve 를, \pm SD 는 Standard Deviation 를 의미한다.

● 각 가상 기관에서 예측된 AUROC

	Metric	AUROC									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.807	0.044	0.890	0.043	0.897	0.053	0.883	0.038	0.893	0.023
FL	server	0.817	0.037	0.896	0.038	0.882	0.050	0.852	0.043	0.828	0.055
LOCAL	model 1	0.788	0.064	0.846	0.066	0.790	0.072	0.751	0.089	0.681	0.059
	model 2	0.736	0.088	0.808	0.074	0.777	0.087	0.737	0.123	0.780	0.093
	model 3	0.763	0.048	0.849	0.049	0.852	0.075	0.799	0.066	0.797	0.073
	model 4	0.762	0.046	0.839	0.064	0.841	0.060	0.849	0.048	0.848	0.040
	model 5	0.822	0.038	0.889	0.049	0.858	0.071	0.743	0.060	0.810	0.023
TIL_10round	model 1	0.816	0.047	0.880	0.049	0.848	0.081	0.805	0.083	0.795	0.081
	model 2	0.805	0.058	0.871	0.060	0.867	0.065	0.804	0.082	0.799	0.088
	model 3	0.813	0.037	0.884	0.035	0.900	0.049	0.837	0.058	0.840	0.058
	model 4	0.811	0.040	0.876	0.043	0.895	0.044	0.853	0.058	0.849	0.044
	model 5	0.824	0.038	0.892	0.044	0.890	0.066	0.845	0.057	0.878	0.028
TIL_1round	model 1	0.788	0.064	0.846	0.066	0.790	0.072	0.751	0.089	0.681	0.059
	model 2	0.776	0.060	0.849	0.049	0.809	0.072	0.777	0.070	0.731	0.081
	model 3	0.794	0.038	0.872	0.042	0.886	0.054	0.810	0.062	0.791	0.055
	model 4	0.800	0.033	0.871	0.040	0.886	0.041	0.859	0.045	0.826	0.038
	model 5	0.820	0.036	0.888	0.044	0.889	0.058	0.831	0.057	0.858	0.024

● 각 가상 기관에서 예측된 Accuracy

	Metric	Accuracy									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.743	0.016	0.757	0.005	0.758	0.005	0.764	0.012	0.762	0.005
FL	server	0.745	0.014	0.759	0.005	0.758	0.005	0.757	0.010	0.752	0.007
LOCAL	model 1	0.746	0.017	0.754	0.006	0.748	0.006	0.746	0.014	0.740	0.006
	model 2	0.730	0.020	0.752	0.008	0.748	0.008	0.743	0.020	0.748	0.013
	model 3	0.738	0.017	0.756	0.006	0.753	0.007	0.749	0.013	0.750	0.010
	model 4	0.736	0.012	0.755	0.005	0.754	0.006	0.759	0.011	0.755	0.009
	model 5	0.750	0.017	0.758	0.004	0.754	0.006	0.745	0.009	0.749	0.004
TIL_10round	model 1	0.750	0.015	0.757	0.005	0.753	0.007	0.753	0.012	0.748	0.009
	model 2	0.745	0.015	0.756	0.006	0.755	0.005	0.752	0.012	0.752	0.010
	model 3	0.748	0.015	0.757	0.005	0.758	0.004	0.755	0.012	0.755	0.008
	model 4	0.747	0.016	0.757	0.006	0.759	0.004	0.758	0.011	0.755	0.008
	model 5	0.750	0.015	0.758	0.004	0.758	0.005	0.757	0.011	0.758	0.005
TIL_1round	model 1	0.746	0.017	0.754	0.006	0.748	0.006	0.746	0.014	0.740	0.006
	model 2	0.743	0.016	0.754	0.005	0.751	0.005	0.749	0.012	0.746	0.010
	model 3	0.743	0.015	0.756	0.006	0.756	0.006	0.752	0.012	0.748	0.009
	model 4	0.745	0.014	0.756	0.007	0.758	0.005	0.762	0.008	0.753	0.008
	model 5	0.749	0.016	0.758	0.004	0.759	0.004	0.756	0.011	0.755	0.005

● 각 가상 기관에서 예측된 f1

	Metric	f1									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.361	0.066	0.253	0.028	0.259	0.028	0.402	0.049	0.312	0.024
FL	server	0.371	0.056	0.264	0.025	0.258	0.027	0.371	0.042	0.265	0.037
LOCAL	model 1	0.375	0.069	0.241	0.034	0.204	0.037	0.323	0.063	0.199	0.035
	model 2	0.307	0.088	0.224	0.045	0.205	0.044	0.307	0.089	0.241	0.069
	model 3	0.342	0.073	0.249	0.036	0.231	0.037	0.338	0.057	0.250	0.055
	model 4	0.335	0.049	0.245	0.029	0.238	0.033	0.380	0.046	0.279	0.047
	model 5	0.388	0.069	0.258	0.022	0.240	0.033	0.320	0.040	0.249	0.020
TIL_10round	model 1	0.392	0.060	0.254	0.028	0.232	0.039	0.353	0.055	0.242	0.048
	model 2	0.371	0.062	0.247	0.035	0.244	0.027	0.349	0.051	0.262	0.054
	model 3	0.382	0.060	0.254	0.028	0.259	0.022	0.365	0.050	0.280	0.041
	model 4	0.379	0.062	0.254	0.032	0.263	0.022	0.378	0.045	0.278	0.042
	model 5	0.390	0.057	0.262	0.022	0.259	0.027	0.373	0.048	0.291	0.027
TIL_1round	model 1	0.375	0.069	0.241	0.034	0.204	0.037	0.323	0.063	0.199	0.035
	model 2	0.362	0.066	0.236	0.025	0.222	0.028	0.339	0.054	0.231	0.053
	model 3	0.363	0.060	0.248	0.032	0.250	0.033	0.353	0.051	0.244	0.046
	model 4	0.369	0.056	0.248	0.036	0.258	0.025	0.392	0.033	0.267	0.043
	model 5	0.385	0.061	0.259	0.024	0.264	0.023	0.366	0.047	0.278	0.025

● 각 가상 기관에서 예측된 Sensitivity

	Metric	Sensitivity									
	Target test set	Test set 1	±SD	Test set 2	±SD	Test set 3	±SD	Test set 4	±SD	Test set 5	±SD
CDS	server	0.636	0.139	0.897	0.111	0.920	0.113	0.863	0.128	0.880	0.080
FL	server	0.656	0.120	0.940	0.100	0.913	0.108	0.780	0.110	0.728	0.118
LOCAL	model 1	0.668	0.145	0.847	0.134	0.703	0.140	0.663	0.150	0.528	0.102
	model 2	0.529	0.175	0.783	0.173	0.707	0.167	0.630	0.212	0.660	0.210
	model 3	0.597	0.148	0.880	0.139	0.807	0.144	0.700	0.144	0.683	0.170
	model 4	0.579	0.101	0.863	0.115	0.837	0.128	0.805	0.119	0.773	0.149
	model 5	0.696	0.145	0.913	0.091	0.843	0.132	0.653	0.097	0.675	0.062
TIL_10round	model 1	0.703	0.129	0.900	0.112	0.813	0.149	0.737	0.132	0.658	0.151
	model 2	0.659	0.132	0.873	0.137	0.860	0.108	0.727	0.126	0.723	0.168
	model 3	0.681	0.130	0.900	0.112	0.917	0.091	0.767	0.128	0.775	0.129
	model 4	0.676	0.136	0.900	0.126	0.937	0.088	0.800	0.114	0.770	0.134
	model 5	0.699	0.127	0.930	0.090	0.920	0.108	0.787	0.122	0.810	0.088
TIL_1round	model 1	0.668	0.145	0.847	0.134	0.703	0.140	0.663	0.150	0.528	0.102
	model 2	0.639	0.139	0.827	0.101	0.773	0.110	0.702	0.134	0.625	0.162
	model 3	0.640	0.127	0.877	0.125	0.883	0.131	0.735	0.126	0.663	0.141
	model 4	0.653	0.121	0.877	0.142	0.913	0.102	0.835	0.087	0.735	0.137
	model 5	0.688	0.136	0.920	0.097	0.940	0.094	0.768	0.119	0.768	0.080

TIL 모형은 지역 학습 모형과 비교해 반복 전이를 통한 지역 학습보다 향상 되었다. 상응하는 지역 학습 모형과 대조적으로 TIL-1 에서 모형 5 는 단일 라운드에서 전이학습의 효과를 재 반복한 자신의 가상 기관 테스트 셋 5 에 대해 0.961 ± 0.027 이었다. TIL-1 에서 모형 1 은 기관 1 의 데이터에서만 훈련되므로 지역 성능 점수는 지역 학습 모형 1 과 유사하다.

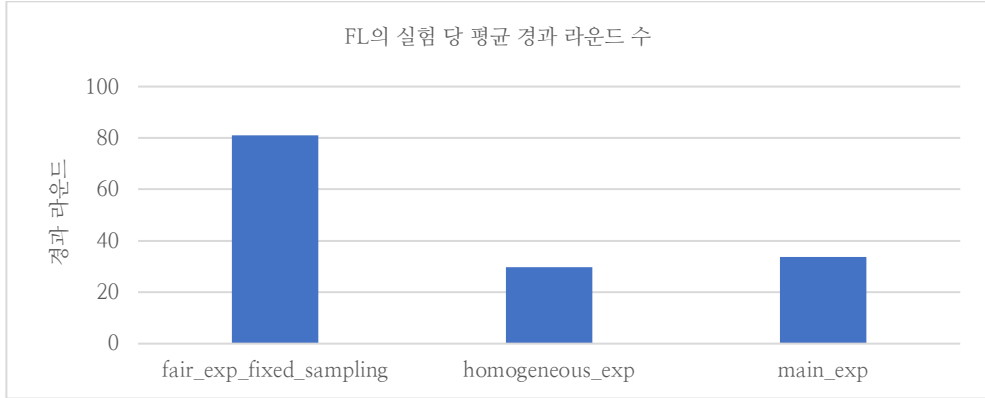
TIL-10 에서 모형 1 은 테스트셋 5 에서 0.890 ± 0.091 이었고 이는 지역 학습의 모형 1 에 비해 향상된 것이다.

TIL-1 의 앞 부분의 모형은 TIL-10 모형보다 외부 테스트 세트에 대해 더 낮은 지역 성능 점수를 보였고, 이는 다른 가상 기관의 데이터로 학습하지 않았기 때문일 것으로 사료된다 (그림 5-6 c, d 의 모형 1-4 와 테스트셋 2-5, 대각선 RE 위의 RE). 이 현상은 혼합 분포의 가상 기관 환경에서 보다 작은 기관 환경에서 더 잘 보여진다. TIL-10 는 다중 라운드를 통해 더 많은 지식을 얻었고 TIL-1 모형보다 더 높은 지역 성능 점수를 보여준 것으로 생각할 수 있다.

다른 모든 모형에 비해 CDS 모형의 성능이 월등해 보이지만 몇몇 모형의 경우 (FL 모형, 지역 학습의 모형 5, TIL-1 라운드의 모형 5, TIL-10 라운드의 모형 1, 3, 5) CDS 보다 높은 지역 성능 점수를 보였다. CDS 모형은 전역 데이터 분포를 예측하기 위해 훈련되므로 몇몇 모형은 지역 테스트 데이터를 더 정교하게 예측하기 때문으로 보인다. 추가적으로 이런 부침은 작은 훈련 세트와 작은 테스트 세트로 인해 일어난 것으로 보인다.

제 7 절 FL 과 TIL 의 통신 비용의 비교

표 3-9 FL 의 실험 당 평균 경과 라운드 수 및 평균 통신 비용



실험	Label	평균	표준편차	최소 round	최대 round	평균통신 비용
작은 기관 환경	fair_exp_fixed_sampling	80.98	22.60	39	100	810
혼합 동질 환경	homogeneous_exp	29.80	0.84	29	31	298
혼합 기관 환경	main_exp	33.60	2.61	31	37	336
	TIL 1round					5
	TIL 10rounds					50

FL 의 통신 비용과 TIL 의 통신 비용을 비교함으로써 RE 를 이용한 성능의 비교 뿐 아니라 통신의 효율의 측면에서 TIL 의 효율을 설명하고자 한다.

단일 가상 기관에서 1 epoch 의 훈련을 마치고 생성된 모형을 중앙 서버에 업로드하였다. 이를 전 가상 기관으로 확장하여 5 개 가상 기관 모두에서 중앙 서버로 업로드 된 모형을 통합하여 중앙 서버에서 업데이트 후, 그 업데이트한 통합된 모형을 각 가상 기관으로 다운로드하는 과정을 1 라운드로 하였다.

작은 기관 환경은 (3 장 6 절 참조) 각 가상 기관을 가장 데이터가 작은 가상 기관 5 에 맞추었으며, 평균 경과 라운드는 80.98 라운드였으며, 수렴에 이르지 못한 경우도 총 50 회의 실험 (2 장 6 절, 7 절 참조)에서 20 회가 있었다.

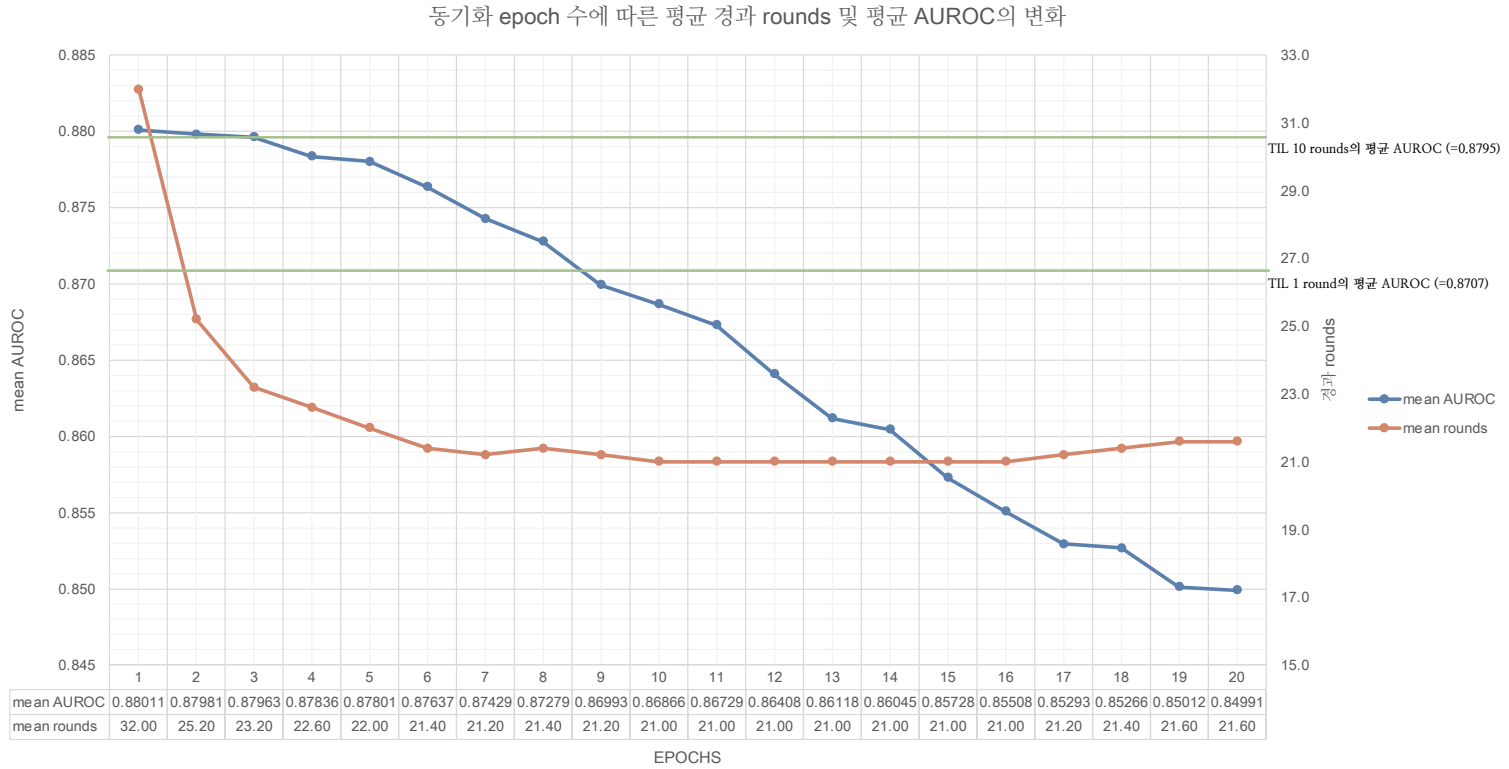
혼합 기관 환경의 샘플 수를 유지하였지만, 무작위 샘플링으로 동질한 분포를 실험한 혼합 동질 환경 (3 장 5 절 참조)의 평균 경과 라운드는 29.8 라운드였으며, 이질적인 분포와 샘플 수로 실험한 혼합 기관 환경 (3 장 4 절)의 평균 경과 라운드는 33.6 라운드였다.

1 라운드가 가상 기관 5 개에서 업로드와 다운로드 각 1 회씩 2 회 통신을 하므로 1라운드 당 왕복에 든 FL 시스템 전체의 통신 비용은 10 이라 할 수 있다. 이를 각 환경의 계산에 적용하면, 작은 기관 환경에서 평균 통신 비용은 810, 혼합 동질 환경에서 298, 혼합 기관 환경에서 336 이라 할 수 있다.

이는 TIL 1 라운드에서 같은 숫자의 가상 기관으로 훈련할 경우 통신 비용은 5, TIL 10 라운드에서 통신 비용이 50 인 것에 비하면 통신 비용이 높다고 할 수 있다.

또한 FL 에서 매 1 epoch 마다 업로드-업데이트-다운로드함으로써 발생하는 통신 비용을 절감할 수 있는 지를 보고자, 각 가상 기관 내에서 훈련에 동원되는 epoch 의 수를 늘려 훈련한 후 중앙 서버로 업데이트를 하는 방식으로 추가 실험을 해 본 결과는 다음과 같다.

표 3-10 동기화 epoch 수에 따른 평균 경과 라운드 및 평균 AUROC 의 변화



전체 데이터를 가장 이질적으로 5 개의 가상 기관으로 나누는 혼합 기관 환경의 데이터로 FL 의 epoch 수를 변화시켜 가며 조기 종료에 이른 라운드 수와 그 때의 AUROC 로 평가를 하였다.

1 epoch 마다 모델을 업로드하는 것이 본래의 FL 의 실험 세팅이었으며, 2 epoch 부터 20 epoch 를 순환 후 업로드하는 것으로 실험을 확장하였다. 그 결과 업로드 주기를 늘릴 수록 처음 4~5 epoch 마다 업로드를 하는 경우 점차 수렴에 이르는 라운드 수가 약 22 회에서 21 회까지 감소하였지만 21 회를 유지하는 모습을 보였다.

또한, AUROC 는 0.88 에서 0.87 을 유지하였으나 업로드 주기를 지연시키면 시킬수록 점차 성능이 하락하였다. TIL 10rounds 의 경우 AUROC 의 평균은 0.88 로 FL 의 경우 3 epoch 마다 업로드를 한 경우와 유사하였으며 이 이후로 TIL 10rounds 의 성능보다 하회하는 것을 관찰할 수 있었다. 또한 TIL 1round 의 평균 AUROC는 0.87이었으며 이는 FL에서 8 epoch 마다 업로드를 하는 것과 유사하였으며 이보다 업로드를 지연시킬수록 성능이 하락하였다.

그러므로, 본 실험의 제반 환경에서 FL 을 사용하고자 한다면 업로드를 3 epoch 마다 하는 것이 성능의 하락을 막으면서 통신 비용을 절감할 수 있는 방안이라 할 수 있다. 하지만, 앞의 계산 방식과 같이 단순화시킨 통신 비용의 계산을 적용한다면, 수렴을 할 때까지 23.2 회의 라운드가 필요하므로 각 가상 기관 개별의 업로드와 다운로드의 수를 고려하면 232 의 통신 비용이 발생하므로 여전히 TIL 10rounds 의 50 에 비해 비용이 높다고 할 수 있다.

각 가상 기관에서 epoch 수를 늘려 업로드의 주기를 지연시킴에 따라 AUROC 의 성능이 점차 하락하는 것은 각 가상 기관의 지역 데이터에 모형이 최적화됨에 따라 각 가상 기관 내의 데이터에 맞추어 가상 기관 별로 모형이 과적합 된 것으로 사료된다.

또한, epoch 수를 늘려도 조기 종료까지 경과한 라운드 수가 21 회에서 횡보하는 것은 각자 기관에서 과적합 된 모형이더라도 본 실험의 데이터셋이 원천적으로 단일 기관에서 분할된 데이터이므로 다수의 공통된 특성을 지니고 있어서 일정 라운드의 경과에서 수렴이 되지 않을까 하는 추측을 할 수 있다. 이는 차후 실세계의 데이터셋으로 더 실험을 해야 하는 주제라 할 수 있다.

제 4 장 프라이버시 보전 딥러닝 예측 모형의 고찰

충분한 데이터에 대한 필요와 환자 개인정보의 보호 사이의 딜레마는 이로운 AI 모형의 발전을 제한하고 이런 상황은 종종 의료 영역에서 마주쳐왔던 문제이다. PPDL 훈련 알고리즘을 통해 AI 모형을 훈련하는 것은 하나의 가능한 접근 방법으로 두 필요를 가장 잘 균형 잡아 줄 수 있는 방법이라 할 수 있다^{85,90}. EHR 데이터를 사용하여 이 연구에서 나온 전역과 지역 점수를 보면 PPDL 훈련 알고리즘이 지역 학습 모형보다 성능이 좋음을 보였고, CDS 모형과 비견할만한 성능을 보임을 알 수 있었다 (그림 5-3 ~ 그림 5-6). 그리고 그것은 이전 PPDL 연구와 일관성이 있다¹¹²⁻¹¹⁴. FL 은 성능이 좋지만 빈번한 모형 동기화를 요하며 이는 더 높은 통신 비용으로 이끈다¹⁰². 그에 대한 해결책으로, TIL 모형은 FL 모형에 비교할만한 경쟁적인 성능을 보였으며 통신 비용도 낮고 라운드의 숫자도 줄여 FL 과 경쟁할만 하다 할 수 있다. 이 연구의 결과로 TIL 모형은 FL 에 유사한 성능 점수를 보였으며 기관 방문의 단일 라운드만으로도 성취되었다. TIL 의 성능은 라운드의 횟수를 높일수록 좋아지며 그럼에도 불구하고 TIL 의 통신 비용은 FL 의 비용보다 여전히 낮다. 그러므로 TIL 은 효율적인 PPDL 훈련 알고리즘이며 PPDL 의 환경 에서 고려될 수 있는 FL 보다도 효율적이다. 모형 전이의 횟수가 FL 보다 적기 때문에 TIL 은 자동화된 프레임워크 없이 수동으로 개발하는 것도 가능하다. 이는 엄격한

보안 프로토콜이나 자동화된 프레임워크를 지지해 줄 수 있는 하부구조가 없는 작은 의료기관에서도 사용할 수 있다.

데이터의 특성 때문에 EHR 을 응용하여 의료 기관 간 TIL 을 적용하기 위해 추가적 보정이 필요하다. 사진이나 신호를 포함한 의료 데이터에서 데이터 구조의 차원은 시공간에 상응한다. 그런 응용에서 기관 사이의 데이터 구조의 이질성은 reshaping, interpolation, resampling 등의 방법으로 통일할 수 있다. 그러나 EHR 의 활용 과정에서 입력 벡터의 각 변수는 측정 (measurement), 상태 (condition), 약제 복용력 (medication) 등 형태가 다른 특성 변수이기 때문에 기관 사이 EHR 내 변수가 이질성이 심할 때 통일하기 어렵다. 그러므로 CDM 과 같은 EHR 데이터 구조에서 EHR 을 응용한 PDDL 을 적용하는 것에 대하여 데이터를 보유한 기관 사이 상호 consensus 가 필수적이다.

전이학습에 기반을 둔 TIL 이지만 서로 몇가지 차이가 있다. 전이학습은 다르지만 비슷한 목적에 대해 훈련된 모형을 전이하는 것으로 비슷한 영역으로부터 학습된 지식을 기반으로 현재 목적에 대해 모형을 조정하는 것이다 ¹²⁰. 그리고, 전이 학습에 대해 전이는 거의 항상 한 번만 수행 된다. 대조적으로 TIL 은 같은 목적에 관련된 분산된 데이터의 전역과 지역 특성에 초점을 맞춘다. 모형 전이는 한 라운드당 각 기관에 걸쳐 여러번 수행되고 추가 라운드는 모형 성능을 높이기 위해 수행될 수 있다. 그 결과로 TIL 연구는 같은 문제에 대해 분산된 데이터에 걸쳐 반복적으로 모형을 전이할 때 모형 성능을 제시할 필요가 있다.

TIL 의 목적은 전이 학습과 목적이 다르기 때문에 TIL 의 제한점은 전이 학습의 요소로부터 나온다. 훈련된 모형이 이질적인 특성 분포로 기관에 전이될 때 ⁹⁰, 모형은 이전 데이터로부터 배운 지식을 잊을 수 있고 이는 catastrophic forgetting problem¹²¹ 으로 알려져 있다. 이 현상은 작은 기관 환경에서 TIL-10-desc 모형의 전역 점수에서 볼 수 있는데, 전역 점수가 다중 훈련 라운드 후 기관의 모형에 따라 부침하는 것을 볼 수 있다. 게다가, TIL-10-desc 모형 1 에서 5 까지의 전체 지역 점수는 forgetting problem 을 보여 주는 것으로 보인다. 비록 forgetting problem 는 이 연구 결과에서 성능을 나쁘게 만드는 데 큰 역할을 하는 것으로 보이지 않지만 각 기관에서 온 테스트 데이터에 대해 각 기관의 TIL 모형의 지역과 전역 점수를 분석하는데 catastrophic forgetting 을 인식하는 것은 중요하다. 그러므로 단일 기관에서 단일 대표 모형을 선택하는 것보다 모든 기관의 모든 모형을 TIL 의 평가에 고려해야 한다 ^{101,108,109,112-114}.

TIL 은 몇 가지 방법으로 더 연구될 수 있다. forgetting problem 효과를 상쇄하기 위해 continual learning¹²²⁻¹²⁵ 기법이 적용될 수 있다. Continual learning 을 적용하는 것은 AI 모형이 이전 기관에서 얻은 지식을 기억하도록 하여 전역 데이터 분포를 더 잘 예측한다. 모형 성능에 대한 기관 간 방문 순서의 효과는 역시 더 연구되어야 한다. 다양한 TIL 기관 방문 순서가 데이터 크기보다 답안 분류의 숫자나 응용에 특화된 계층 방법 등 기준에 기반해 고안될 수 있으나 가장 좋은 전이 순서는 알려져 있지 않다. 추가적으로 각 기관에서 나온 optimizer 의 상태는 초기화, 복사 또는 다른 방법으로 집적될 수 있고 모형 성능에 영향을 미칠 수 있다. 이런

앞으로 기대되는 연구로 TIL 은 성능 좋은 모형을 더 낮은 통신 비용으로 만들 수 있는 강력한 PPDL 훈련 알고리즘이 될 수 있다.

결 론

기준으로 삼을 통계 모형을 위해 건강보험공단의 주관으로 건강검진을 시행 받은 21,581,796명의 건강검진 데이터와 명세서 데이터를 이용하여 생존분석을 수행하고 한국인의 검진 이후 8년 이내의 CVD 발병을 예측하는 Cox 비례 위험 모형을 작성하였으며 이를 현재 임상에서 널리 이용되고 있는 CVD의 예측 모형인 FRS, PCE와 비교하였다.

그 결과로 AUROC가 남성에서 0.762 ± 0.002 , 여성에서 0.811 ± 0.002 로 측정 되었으며 이는 서구에서 개발된 모형 보다 나은 성능을 보였다. 이 결과로 한국인 뿐만 아니라 아시아인에게 맞춘 정교한 CVD 모형의 필요성을 제고할 수 있다.

하지만, 이와 같은 연구와 자원을 모두 활용하여 환자 진료에 응용하기에는 현실적인 어려움이 많다. 또한, 개인정보보호에 대한 제한과 법률이 엄격해 짐에 따라 데이터 상호 교류와 외부 반출은 더욱더 어려워지고 있다.

분산된 데이터로 통합된 모형을 제작하고자 하는 실험으로 EHR 데이터를 사용하여 CVD 예측 모형에 전이반복합습 (TIL)을 포함한 PDDL의 환경 하에 훈련 알고리즘의 성능을 비교하였다. 서울대학교병원의 common data model (CDM) 을 이용하여 40,507명의 당뇨 환자 코호트를 구성하였고 이를 k-means clustering을 이용하여 가상의 5개 기관으로 나누었다. 다양한 기관들의 조건 아래 TIL의 성능을 평가하기 위해 상이한

데이터 분포를 지닌 두 가지 다른 기관 실험 환경이 사용되었다. 5개 기관의 데이터 크기에 따라 오름차순과 내림차순으로 훈련하였고, 단일 라운드와 10라운드를 시행하였다. 또한 가장 작은 기관의 데이터에 맞춰 5개 기관의 데이터 개수를 통일하고 동일하게 단일 라운드와 10라운드로 실험을 진행하였다.

FL과 TIL 중 어떤 PPDL 알고리즘이 실세계 데이터셋에서 얼마나 강건할 수 있을지를 비교해 보기 위해 동질적인 분포의 가상 기관 환경을 추가로 실험하였으나, FL, TIL 모두 성능의 하강이 동질적인 분포에 비해 혼합 분포의 성능이 크게 저하되지는 않았다.

비교와 평가를 위해 시뮬레이션 연구이기 때문에 가능한 모든 데이터를 모아서 테스트 세트의 결과를 평가할 수 있는 CDS 모형과, 닫힌 네트워크로 인해 실세계에서 사용하기 힘든 FL의 성능을 TIL과 비교하였다. 또한 각 기관의 협력이 없는 자 기관의 모형과 이를 테스트 세트에 적용한 모형과 TIL을 비교하였다.

그 결과로 혼합 기관 환경에서 CDS의 AUROC 대비 TIL의 AUROC는 전역 성능은 97.6%에서 99.6%였으며 FL은 99.4%를 보였다. 자 기관만을 예측한 경우 85.1%에서 99.3%로 다양했다. 전체 데이터를 모은 CDS에 비해 TIL은 성능이 많이 떨어지지 않았다. 또한, 협력이 없는 자 기관 예측의 성능은 다양한 반면 TIL의 경우 안정적인 예측 성능을 보였다.

또한, 이질적인 분포를 그대로 두고 가장 작은 기관의 데이터에 맞춰 수행한 실험에서 FL 보다 TIL의 성능이 더 좋았으며 이는 작은 기관들의 협력으로 충분히 데이터를 모두 모은 효과를 가질 수 있다는 가능성을 보여

주었다.

이로서 FL보다 통신 비용을 줄이며 안정적인 성능을 담보할 수 있는 TIL을 이용하여 데이터가 작은 기관도 큰 병원과 협력하여 데이터를 외부와 공유 없이 개인정보를 보호하며 모형을 제작하여 이를 의료 분야에 활용할 것을 제안한다.

학술지 사전 게재 정보

본 학위 논문은 저자가 사전에 1저자로 게재했던 학술지 논문의 내용을 바탕으로 함

- Choi, S.H., Lee, S.M., Kim, S.H. *et al.* Prediction of 8-year risk of cardiovascular diseases in Korean adult population. *Sci Rep* 11, 14339 (2021). <https://doi.org/10.1038/s41598-021-93840-2>

참 고 문 헌

- 1 Collaborators, G. B. D. R. F. Global, regional, and national comparative risk assessment of 79 behavioural, environmental and occupational, and metabolic risks or clusters of risks, 1990-2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1659-1724, doi:10.1016/S0140-6736(16)31679-8 (2016).
- 2 Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association. *Circulation* **137**, e67-e492, doi:10.1161/CIR.0000000000000558 (2018).
- 3 질병관리청. 만성질환예방관리 (심뇌혈관질환홍보사업), <<https://www.kdca.go.kr/contents.es?mid=a20303020300>> (2022).
- 4 Dahlof, B. Cardiovascular disease risk factors: epidemiology and risk assessment. *Am J Cardiol* **105**, 3A-9A, doi:10.1016/j.amjcard.2009.10.007 (2010).
- 5 Damen, J. A. *et al.* Prediction models for cardiovascular disease risk in the general population: systematic review. *BMJ* **353**, i2416, doi:10.1136/bmj.i2416 (2016).
- 6 Zhao, D., Liu, J., Xie, W. & Qi, Y. Cardiovascular risk assessment: a global perspective. *Nat Rev Cardiol* **12**, 301-311, doi:10.1038/nrcardio.2015.28 (2015).
- 7 Seong, S. C. *et al.* Data Resource Profile: The National Health Information Database of the National Health Insurance Service in South Korea. *International Journal of Epidemiology*, dyw253 (2016).
- 8 Seong, S. C. *et al.* Cohort profile: the National Health Insurance Service-National Health Screening Cohort (NHIS-HEALS) in Korea. *BMJ Open* **7**, e016640, doi:10.1136/bmjopen-2017-016640 (2017).
- 9 국민건강보험. 표본코호트DB 안내, <<https://nhiss.nhis.or.kr/bd/ab/bdaba002cv.do#>> (2022).
- 10 D'Agostino, R. B., Sr. *et al.* General cardiovascular risk profile for use in primary care: the Framingham Heart Study. *Circulation* **117**, 743-753, doi:10.1161/CIRCULATIONAHA.107.699579 (2008).
- 11 Kannel, W. B., McGee, D. & Gordon, T. A general cardiovascular risk profile: the Framingham Study. *Am J Cardiol* **38**, 46-51, doi:10.1016/0002-9149(76)90061-8 (1976).

- 12 Fox, C. S., Evans, J. C., Larson, M. G., Kannel, W. B. & Levy, D. Temporal trends in coronary heart disease mortality and sudden cardiac death from 1950 to 1999: the Framingham Heart Study. *Circulation* **110**, 522-527, doi:10.1161/01.CIR.0000136993.34344.41 (2004).
- 13 Brindle, P. *et al.* Predictive accuracy of the Framingham coronary risk score in British men: prospective cohort study. *BMJ* **327**, 1267, doi:10.1136/bmj.327.7426.1267 (2003).
- 14 Goff, D. C., Jr. *et al.* 2013 ACC/AHA guideline on the assessment of cardiovascular risk: a report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *J Am Coll Cardiol* **63**, 2935-2959, doi:10.1016/j.jacc.2013.11.005 (2014).
- 15 Investigators, A. R. i. C. The Atherosclerosis Risk in Communities (ARIC) study: design and objectives. *Am J Epidemiol* **129**, 687-702 (1989).
- 16 Fried, L. P. *et al.* The Cardiovascular Health Study: design and rationale. *Ann Epidemiol* **1**, 263-276, doi:10.1016/1047-2797(91)90005-w (1991).
- 17 Friedman, G. D. *et al.* Cardia - Study Design, Recruitment, and Some Characteristics of the Examined Subjects. *Journal of Clinical Epidemiology* **41**, 1105-1116, doi:Doi 10.1016/0895-4356(88)90080-7 (1988).
- 18 TR, D., WB, K. & LP, L. An approach to longitudinal studies in a community: the Framingham Study. *Annals of the New York Academy of Sciences* **107**, 539-556 (1963).
- 19 Kannel, W. B., Feinleib, M., McNamara, P. M., Garrison, R. J. & Castelli, W. P. An investigation of coronary heart disease in families. The Framingham offspring study. *Am J Epidemiol* **110**, 281-290, doi:10.1093/oxfordjournals.aje.a112813 (1979).
- 20 Go, A. S. *et al.* Heart disease and stroke statistics--2013 update: a report from the American Heart Association. *Circulation* **127**, e6-e245, doi:10.1161/CIR.0b013e31828124ad (2013).
- 21 Muntner, P. *et al.* Validation of the atherosclerotic cardiovascular disease Pooled Cohort risk equations. *JAMA* **311**, 1406-1415, doi:10.1001/jama.2014.2630 (2014).
- 22 Kavousi, M. *et al.* Comparison of application of the ACC/AHA guidelines, Adult Treatment Panel III guidelines, and European Society of Cardiology guidelines for cardiovascular disease prevention in a European cohort. *JAMA* **311**, 1416-1423, doi:10.1001/jama.2014.2632 (2014).
- 23 Mortensen, M. B., Afzal, S., Nordestgaard, B. G. & Falk, E. Primary Prevention With Statins: ACC/AHA Risk-Based Approach Versus Trial-Based Approaches to Guide Statin Therapy. *J Am Coll Cardiol* **66**, 2699-2709, doi:10.1016/j.jacc.2015.09.089 (2015).

- 24 Yadlowsky, S. *et al.* Clinical Implications of Revised Pooled Cohort Equations for Estimating Atherosclerotic Cardiovascular Disease Risk. *Ann Intern Med* **169**, 20-29, doi:10.7326/M17-3011 (2018).
- 25 OECD. *Elderly population*, <<https://data.oecd.org/pop/elderly-population.htm>> (2018).
- 26 Publications, U. N. *World Population Ageing, 2015*. (UN, 2017).
- 27 Jee, S. H. *et al.* A coronary heart disease prediction model: the Korean Heart Study. *BMJ Open* **4**, e005025, doi:10.1136/bmjopen-2014-005025 (2014).
- 28 Jung, K. J. *et al.* The ACC/AHA 2013 pooled cohort equations compared to a Korean Risk Prediction Model for atherosclerotic cardiovascular disease. *Atherosclerosis* **242**, 367-375, doi:10.1016/j.atherosclerosis.2015.07.033 (2015).
- 29 Cook, N. R. & Ridker, P. M. Calibration of the Pooled Cohort Equations for Atherosclerotic Cardiovascular Disease: An Update. *Ann Intern Med* **165**, 786-794, doi:10.7326/M16-1739 (2016).
- 30 Lee, C. H. *et al.* Validation of the Pooled Cohort equations in a long-term cohort study of Hong Kong Chinese. *J Clin Lipidol* **9**, 640-646 e642, doi:10.1016/j.jacl.2015.06.005 (2015).
- 31 Fawcett, T. An introduction to ROC analysis. *Pattern Recognition Letters* **27**, 861-874, doi:10.1016/j.patrec.2005.10.010 (2006).
- 32 Brier, G. W. Verification of Forecasts Expressed in Terms of Probability. *Monthly Weather Review* **78**, 1-3, doi:10.1175/1520-0493(1950)078<0001:Vofeit>2.0.Co;2 (1950).
- 33 Harrell, F. E., Jr., Califf, R. M., Pryor, D. B., Lee, K. L. & Rosati, R. A. Evaluating the yield of medical tests. *JAMA* **247**, 2543-2546, doi:10.1001/jama.1982.03320430047030 (1982).
- 34 Schmid, M., Wright, M. N. & Ziegler, A. On the use of Harrell's C for clinical risk prediction via random survival forests. *Expert Systems with Applications* **63**, 450-459, doi:10.1016/j.eswa.2016.07.018 (2016).
- 35 Korea, S. *2017 Population and Housing Census*, <<http://kostat.go.kr/portal/eng/index.action>> (2018).
- 36 Witkiewitz, K. *et al.* Clinical Validation of Reduced Alcohol Consumption After Treatment for Alcohol Dependence Using the World Health Organization Risk Drinking Levels. *Alcohol Clin Exp Res* **41**, 179-186, doi:10.1111/acer.13272 (2017).
- 37 World Health, O. (World Health Organization, Geneva, 2000).
- 38 Craig, C. L. *et al.* International physical activity questionnaire: 12-country reliability and validity. *Med Sci Sports Exerc* **35**, 1381-1395, doi:10.1249/01.MSS.0000078924.61453.FB (2003).
- 39 Arija, V. *et al.* Effectiveness of a physical activity program on cardiovascular disease risk in adult primary health-care users: the "Pas-a-Pas" community intervention trial. *BMC Public Health* **17**, 576, doi:10.1186/s12889-017-4485-3 (2017).

- 40 Physical status: the use and interpretation of anthropometry. Report
of a WHO Expert Committee. *World Health Organ Tech Rep Ser*
854, 1-452 (1995).
- 41 Levey, A. S. *et al.* A new equation to estimate glomerular filtration
rate. *Ann Intern Med* **150**, 604-612, doi:10.7326/0003-4819-150-9-
200905050-00006 (2009).
- 42 World Health, O. *ICD-10 : international statistical classification of*
diseases and related health problems / World Health Organization.
(World Health Organization, 2004).
- 43 Kleinbaum, D. G. & Klein, M. *Survival analysis: a self-learning*
text. (Springer, 2012).
- 44 김양진. *생존분석.* (자유아카데미, 2013).
- 45 Veeranna, V. *et al.* Homocysteine and reclassification of
cardiovascular disease risk. *J Am Coll Cardiol* **58**, 1025-1033,
doi:10.1016/j.jacc.2011.05.028 (2011).
- 46 Rokey, R., Rolak, L. A., Harati, Y., Kutka, N. & Verani, M. S.
Coronary artery disease in patients with cerebrovascular disease: a
prospective study. *Ann Neurol* **16**, 50-53,
doi:10.1002/ana.410160110 (1984).
- 47 Berenson, G. S. *et al.* Association between multiple cardiovascular
risk factors and atherosclerosis in children and young adults. The
Bogalusa Heart Study. *N Engl J Med* **338**, 1650-1656,
doi:10.1056/NEJM199806043382302 (1998).
- 48 Huxley, R. R. & Woodward, M. Cigarette smoking as a risk factor
for coronary heart disease in women compared with men: a
systematic review and meta-analysis of prospective cohort studies.
Lancet **378**, 1297-1305, doi:10.1016/S0140-6736(11)60781-2
(2011).
- 49 Thun, M. J. *et al.* 50-year trends in smoking-related mortality in the
United States. *N Engl J Med* **368**, 351-364,
doi:10.1056/NEJMsa1211127 (2013).
- 50 Stefan, N., Haring, H. U., Hu, F. B. & Schulze, M. B. Metabolically
healthy obesity: epidemiology, mechanisms, and clinical
implications. *Lancet Diabetes Endocrinol* **1**, 152-162,
doi:10.1016/S2213-8587(13)70062-7 (2013).
- 51 Phillips, C. M. Metabolically healthy obesity: definitions,
determinants and clinical implications. *Rev Endocr Metab Disord*
14, 219-227, doi:10.1007/s11154-013-9252-x (2013).
- 52 Erol, A. & Karpyak, V. M. Sex and gender-related differences in
alcohol use and its consequences: Contemporary knowledge and
future research considerations. *Drug Alcohol Depend* **156**, 1-13,
doi:10.1016/j.drugalcdep.2015.08.023 (2015).
- 53 Hippisley-Cox, J., Coupland, C. & Brindle, P. Development and
validation of QRISK3 risk prediction algorithms to estimate future

- risk of cardiovascular disease: prospective cohort study. *BMJ* **357**, j2099, doi:10.1136/bmj.j2099 (2017).
- 54 Hense, H. W., Schulte, H., Lowel, H., Assmann, G. & Keil, U. Framingham risk function overestimates risk of coronary heart disease in men and women from Germany--results from the MONICA Augsburg and the PROCAM cohorts. *Eur Heart J* **24**, 937-945, doi:10.1016/s0195-668x(03)00081-2 (2003).
- 55 Kunte, S. & Gore, A. P. The Paradox of Large Samples. *Current Science* **62**, 393-395 (1992).
- 56 Klein, J. P. & Moeschberger, M. L. *Survival analysis: techniques for censored and truncated data*. Vol. 1230 (Springer, 2003).
- 57 Esteva, A. *et al.* A guide to deep learning in healthcare. *Nat Med* **25**, 24-29, doi:10.1038/s41591-018-0316-z (2019).
- 58 Shah, P. *et al.* Artificial intelligence and machine learning in clinical development: a translational perspective. *NPJ Digit Med* **2**, 69, doi:10.1038/s41746-019-0148-3 (2019).
- 59 Ching, T. *et al.* Opportunities and obstacles for deep learning in biology and medicine. *J R Soc Interface* **15**, 20170387, doi:10.1098/rsif.2017.0387 (2018).
- 60 LeCun, Y., Bengio, Y. & Hinton, G. Deep learning review. *Weekly journal of science in nature international. Nature* **521** (2015).
- 61 Goodfellow, I., Bengio, Y. & Courville, A. *Deep learning*. (MIT press, 2016).
- 62 Zhang, J. X. *et al.* A deep learning model for predicting next-generation sequencing depth from DNA sequence. *Nat Commun* **12**, 4387, doi:10.1038/s41467-021-24497-8 (2021).
- 63 Kiyasseh, D., Zhu, T. & Clifton, D. A clinical deep learning framework for continually learning from cardiac signals across diseases, time, modalities, and institutions. *Nat Commun* **12**, 4221, doi:10.1038/s41467-021-24483-0 (2021).
- 64 Luo, Y. *et al.* ECNet is an evolutionary context-integrated deep learning framework for protein engineering. *Nature communications* **12**, 1-14 (2021).
- 65 Skreta, M. *et al.* Automatically disambiguating medical acronyms with ontology-aware deep learning. *Nat Commun* **12**, 5319, doi:10.1038/s41467-021-25578-4 (2021).
- 66 Hannun, A. Y. *et al.* Cardiologist-level arrhythmia detection and classification in ambulatory electrocardiograms using a deep neural network. *Nat Med* **25**, 65-69, doi:10.1038/s41591-018-0268-3 (2019).
- 67 Rajpurkar, P. *et al.* Deep learning for chest radiograph diagnosis: A retrospective comparison of the CheXNeXt algorithm to practicing radiologists. *PLoS Med* **15**, e1002686, doi:10.1371/journal.pmed.1002686 (2018).

- 68 Zhang, K. *et al.* Clinically Applicable AI System for Accurate Diagnosis, Quantitative Measurements, and Prognosis of COVID-19 Pneumonia Using Computed Tomography. *Cell* **181**, 1423-1433 e1411, doi:10.1016/j.cell.2020.04.045 (2020).
- 69 De Fauw, J. *et al.* Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nat Med* **24**, 1342-1350, doi:10.1038/s41591-018-0107-6 (2018).
- 70 Wang, F., Casalino, L. P. & Khullar, D. Deep Learning in Medicine-Promise, Progress, and Challenges. *JAMA Intern Med* **179**, 293-294, doi:10.1001/jamainternmed.2018.7117 (2019).
- 71 Cahan, E. M., Hernandez-Boussard, T., Thadaney-Israni, S. & Rubin, D. L. Putting the data before the algorithm in big data addressing personalized healthcare. *NPJ Digit Med* **2**, 78, doi:10.1038/s41746-019-0157-2 (2019).
- 72 Hu, Y. P. *et al.* The challenges of deploying artificial intelligence models in a rapidly evolving pandemic. *Nature Machine Intelligence* **2**, 298-300, doi:10.1038/s42256-020-0185-2 (2020).
- 73 Peiffer-Smadja, N. *et al.* Machine Learning for COVID-19 needs global collaboration and data-sharing. *Nature Machine Intelligence* **2**, 293-294, doi:10.1038/s42256-020-0181-6 (2020).
- 74 Luengo-Oroz, M. *et al.* Artificial intelligence cooperation to support the global response to COVID-19. *Nature Machine Intelligence* **2**, 295-297, doi:10.1038/s42256-020-0184-3 (2020).
- 75 Moorthy, V., Henao Restrepo, A. M., Preziosi, M. P. & Swaminathan, S. Data sharing for novel coronavirus (COVID-19). *Bull World Health Organ* **98**, 150, doi:10.2471/BLT.20.251561 (2020).
- 76 Price, W. N., 2nd & Cohen, I. G. Privacy in the age of medical big data. *Nat Med* **25**, 37-43, doi:10.1038/s41591-018-0272-7 (2019).
- 77 Thapa, C. & Camtepe, S. Precision health data: Requirements, challenges and existing techniques for data security and privacy. *Comput Biol Med* **129**, 104130, doi:10.1016/j.combiomed.2020.104130 (2021).
- 78 Kels, C. G. HIPAA in the Era of Data Sharing. *JAMA* **323**, 476-477, doi:10.1001/jama.2019.19645 (2020).
- 79 HIPAA. US Department of Health and Human Services. *Health information privacy*, <<https://www.hhs.gov/hipaa/index.html>> (2013).
- 80 European Commission. The EU General Data Protection Regulation. *Complete guide to GDPR compliance*, <<https://gdpr.eu/>> (2020).
- 81 van Panhuis, W. G. *et al.* A systematic review of barriers to data sharing in public health. *BMC Public Health* **14**, 1144, doi:10.1186/1471-2458-14-1144 (2014).

- 82 Cath, C. Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philos Trans A Math Phys Eng Sci* **376**, 20180080, doi:10.1098/rsta.2018.0080 (2018).
- 83 Jobin, A., Ienca, M. & Vayena, E. The global landscape of AI ethics guidelines. *Nature Machine Intelligence* **1**, 389-399, doi:10.1038/s42256-019-0088-2 (2019).
- 84 Shokri, R. & Shmatikov, V. in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security* 1310-1321 (Association for Computing Machinery, Denver, Colorado, USA, 2015).
- 85 Chang, K. *et al.* in *Artificial Intelligence in Medicine* (eds Lei Xing, Maryellen L. Giger, & James K. Min) 101-112 (Academic Press, 2021).
- 86 Kaissis, G. A., Makowski, M. R., Ruckert, D. & Braren, R. F. Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence* **2**, 305-311, doi:10.1038/s42256-020-0186-1 (2020).
- 87 Froelicher, D. *et al.* Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nat Commun* **12**, 5910, doi:10.1038/s41467-021-25972-y (2021).
- 88 Vizitiu, A., Niță, C. I., Puiu, A., Suciu, C. & Itu, L. M. in *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*. 1-6.
- 89 Tanuwidjaja, H. C., Choi, R., Baek, S. & Kim, K. Privacy-Preserving Deep Learning on Machine Learning as a Service-a Comprehensive Survey. *Ieee Access* **8**, 167425-167447, doi:10.1109/Access.2020.3023084 (2020).
- 90 Kairouz, P. *et al.* Advances and open problems in federated learning. *arXiv preprint arXiv:1912.04977* (2019).
- 91 Li, W. *et al.* in *International workshop on machine learning in medical imaging*. 133-141 (Springer).
- 92 Ziller, A. *et al.* Medical imaging deep learning with differential privacy. *Sci Rep* **11**, 13524, doi:10.1038/s41598-021-93030-0 (2021).
- 93 Sadilek, A. *et al.* Privacy-first health research with federated learning. *NPJ Digit Med* **4**, 132, doi:10.1038/s41746-021-00489-2 (2021).
- 94 Zhang, X. *et al.* in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 1169-1178.
- 95 McMahan, B., Moore, E., Ramage, D., Hampson, S. & Arcas, B. A. y. in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* Vol. 54 (eds Singh Aarti & Zhu Jerry) 1273--1282 (PMLR, Proceedings of Machine Learning Research, 2017).

- 96 Yang, Q., Liu, Y., Chen, T. J. & Tong, Y. X. Federated Machine Learning: Concept and Applications. *Acm Transactions on Intelligent Systems and Technology* **10**, Article 12, doi:Artn 12 10.1145/3298981 (2019).
- 97 Collaborative learning without sharing data. *Nature Machine Intelligence* **3**, 459-459, doi:10.1038/s42256-021-00364-5 (2021).
- 98 Rieke, N. *et al.* The future of digital health with federated learning. *NPJ Digit Med* **3**, 119, doi:10.1038/s41746-020-00323-1 (2020).
- 99 Yang, D. *et al.* Federated semi-supervised learning for COVID region segmentation in chest CT using multi-national data from China, Italy, Japan. *Med Image Anal* **70**, 101992, doi:10.1016/j.media.2021.101992 (2021).
- 100 Roth, H. R. *et al.* in *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning* 181-191 (Springer, 2020).
- 101 Sheller, M. J., Reina, G. A., Edwards, B., Martin, J. & Bakas, S. in *International MICCAI Brainlesion Workshop*. 92-104 (Springer).
- 102 Bai, X. *et al.* Advancing COVID-19 Diagnosis with Privacy-Preserving Collaboration in Artificial Intelligence. *ArXiv* **3**, 1081-1089, doi:10.1038/s42256-021-00421-z (2021).
- 103 Lu, M. Y. *et al.* Federated learning for computational pathology on gigapixel whole slide images. *Med Image Anal* **76**, 102298, doi:10.1016/j.media.2021.102298 (2022).
- 104 Dayan, I. *et al.* Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med* **27**, 1735-1743, doi:10.1038/s41591-021-01506-3 (2021).
- 105 Kaissis, G. *et al.* End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nature Machine Intelligence* **3**, 473-484, doi:10.1038/s42256-021-00337-8 (2021).
- 106 Cavalli, E., Mattasoglio, A., Pinciroli, F. & Spaggiari, P. Information security concepts and practices: the case of a provincial multi-specialty hospital. *Int J Med Inform* **73**, 297-303, doi:10.1016/j.ijmedinf.2003.12.008 (2004).
- 107 Reddy, M. C., Purao, S. & Kelly, M. Developing IT Infrastructure for Rural Hospitals: A Case Study of Benefits and Challenges of Hospital-to-Hospital Partnerships. *J Am Med Inform Assoc* **15**, 554-558, doi:10.1197/jamia.M2676 (2008).
- 108 Chang, K. *et al.* Distributed deep learning networks among institutions for medical imaging. *J Am Med Inform Assoc* **25**, 945-954, doi:10.1093/jamia/ocy017 (2018).
- 109 Festag, S. & Spreckelsen, C. Privacy-Preserving Deep Learning for the Detection of Protected Health Information in Real-World Data: Comparative Evaluation. *JMIR Form Res* **4**, e14064, doi:10.2196/14064 (2020).

- 110 Pan, S. J. & Yang, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering* **22**, 1345-1359 (2009).
- 111 Torrey, L. & Shavlik, J. in *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques* 242-264 (IGI global, 2010).
- 112 Balachandar, N., Chang, K., Kalpathy-Cramer, J. & Rubin, D. L. Accounting for data variability in multi-institutional distributed deep learning for medical imaging. *Journal of the American Medical Informatics Association* **27**, 700-708 (2020).
- 113 Sheller, M. J. *et al.* Federated learning in medicine: facilitating multi-institutional collaborations without sharing patient data. *Sci Rep* **10**, 12598, doi:10.1038/s41598-020-69250-1 (2020).
- 114 Alawad, M. *et al.* Privacy-Preserving Deep Learning NLP Models for Cancer Registries. *Ieee Transactions on Emerging Topics in Computing* **9**, 1219-1230, doi:10.1109/Tetc.2020.2983404 (2021).
- 115 Reich, C., Ryan, P., Belenkaya, R., Natarajan, K. & Blacketer, C. (2019).
- 116 Everitt, B. S. & Skrondal, A. *The Cambridge dictionary of statistics.* (2010).
- 117 Wannamethee, S. G., Shaper, A. G., Whincup, P. H., Lennon, L. & Sattar, N. Impact of Diabetes on Cardiovascular Disease Risk and All-Cause Mortality in Older Men: Influence of Age at Onset, Diabetes Duration, and Established and Novel Risk Factors. *Archives of Internal Medicine* **171**, 404-410, doi:10.1001/archinternmed.2011.2 (2011).
- 118 Nelson, S. J., Zeng, K., Kilbourne, J., Powell, T. & Moore, R. Normalized names for clinical drugs: RxNorm at 6 years. *J Am Med Inform Assoc* **18**, 441-448, doi:10.1136/amiajnl-2011-000116 (2011).
- 119 McMahan, B., Moore, E., Ramage, D., Hampson, S. & y Arcas, B. A. in *Artificial Intelligence and Statistics.* 1273-1282 (PMLR).
- 120 Gao, Y. & Cui, Y. Deep transfer learning for reducing health care disparities arising from biomedical data inequality. *Nature communications* **11**, 1-8 (2020).
- 121 French, R. M. Catastrophic forgetting in connectionist networks. *Trends Cogn Sci* **3**, 128-135, doi:10.1016/s1364-6613(99)01294-2 (1999).
- 122 Parisi, G. I., Kemker, R., Part, J. L., Kanan, C. & Wermter, S. Continual lifelong learning with neural networks: A review. *Neural Netw* **113**, 54-71, doi:10.1016/j.neunet.2019.01.012 (2019).
- 123 Kirkpatrick, J. *et al.* K., Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, and D. Hassabis. 2017. *Overcoming catastrophic forgetting in neural networks.* In *Proceedings of the National Academy of Sciences* **114**, 3521-3526 (2017).

- 124 Zenke, F., Poole, B. & Ganguli, S. in *International Conference on Machine Learning*. 3987-3995 (PMLR).
- 125 van de Ven, G. M., Siegelmann, H. T. & Tolias, A. S. Brain-inspired replay for continual learning with artificial neural networks. *Nat Commun* **11**, 4069, doi:10.1038/s41467-020-17866-2 (2020).

Abstract

Prediction of Risk of Cardiovascular Diseases in Korean Adult Population : Cox proportional hazards model and privacy preserving deep learning

Sung Hyouk Choi

Graduate School, College of Medicine, Seoul National University

Biomedical Engineering Major

While artificial intelligence (AI) has demonstrated extensive potential in various domains, developing AI models in healthcare is limited because of medical data privacy restrictions. Privacy preserving deep learning (PPDL) attempts to train AI models while protecting data privacy, and one type of PPDL training algorithm named federated learning (FL) has been gaining attention as a solution. However, FL is based on model synchronizations that require automated FL platforms, which may be difficult to implement in some medical institutions with closed networks. We present transfer iterative learning (TIL), a decentralized PPDL training algorithm based on transfer learning which can achieve model performances comparable to FL at lower

communication cost.

Therefore, the main purpose of this study is to preliminarily analyze predictive modeling applying the conventional statistical method using data from the National Health Insurance Service, and to analyze the patient data implemented in the Common Data Model (CDM) of a single institution. It is to implement FL and TIL by simulating with multicenter data and compare the performance.

For predictive modeling with clinical significance, although many prediction models for cardiovascular diseases (CVDs) have been developed and validated for Western populations, the development of CVD prediction models for Asians has been slow. Our cohort study retrospectively analyzed the incidence of CVD that occurred between January 1, 2009, and December 31, 2016, in all Koreans who underwent national health screening. This dataset included 21,581,796 adults between the ages of 40 and 79 years (10,412,947 men, 11,168,849 women) without CVD at baseline. The primary outcome, CVD, was defined as the development of any of the following: acute coronary syndrome, cerebral infarction, and cerebral hemorrhage, as defined with health insurance claims data. The prediction model was constructed by Cox proportional hazard regression and validated with tenfold cross-validation. The performance of the models was evaluated through Harrell's C-index and Brier score. The discrimination of the models was assessed by the area under the receiver operating characteristic curve (AUROC). Our model

showed an AUROC of 0.762 in men and 0.811 in women. The Brier score of our model was 0.018 in men and 0.010 in women, which was better than the pooled cohort equation (PCE). Our novel model performed better than the FRS and PCE for Koreans.

To test the performance of TIL as a follow-up experiment, a predictive model of CVDs was created using data composed of 40,507 diabetic patients implemented in the CDM of Seoul National University Hospital. The worst and best TIL models presented relative efficiency (RE) of 97.6% and 99.6 %, respectively, compared to the model trained on aggregated data. FL demonstrated 99.4% RE, and the worst and best local learning models showed 85.1% and 99.3% RE, respectively. The results suggest that TIL is a communication efficient PPDL training algorithm which could be considered as an alternative to FL for medical institutions that could not support an FL framework.

Keywords: Transfer Iterative Learning, Privacy Preserving Deep Learning, Prediction Model, Cardiovascular Diseases, Federated Learning, Common Data Model

감사의 말

외과 전문의가 되고 펠로우를 하던 아들이 갑자기 빅데이터와 인공지능을 공부하고 싶다고 대학원에 가겠다고 하였을 때, 적잖이 당황하셨을 터임에도 불구하고 자식의 결정을 존중해 주신 부모님께 이 논문을 바칩니다. 부모님의 가르침과 돌봄이 없었다면 이 글은 세상에 나올 수 없었을 것입니다.

학위 과정 중 만나 결혼하게 되어 물심양면으로 옆에서 저를 지켜 봐주고 보살피며 응원하였던 아내 이슬과 그 안의 작은 생명, 아직 정식 이름이 없는 튼튼이에게 이 논문을 선물해주고 싶습니다. 매주 저희 부부와 튼튼이를 잘 챙겨 들라고 고생하시는 장인 어른과 장모님께도 이 자리를 빌어 감사 드리고 싶습니다.

연구실에서 과제를 수행하며 엉뚱한 곳으로 빠지게 되거나 미흡한 점을 예리하고 논리적으로 보며, 이 논문이 제대로 된 결론을 도출할 수 있도록 항상 지도와 조언을 아끼지 않으셨던 저의 지도교수 윤형진 교수님께 깊은 감사의 말씀을 드립니다. 또한, 박사 학위 논문의 심사위원이셨던 서울대학교병원 의공학교실의 김희찬 교수님과 최진욱 교수님, 서울대학교병원 내과학 교실의 이해영 교수님, 그리고 고려대학교 전기전자공학부의 김중헌 교수님의 가르침이 없었다면 이 논문이 나올 수 없었을 것입니다.

어쩌다 보니 형제가 모두 조금 늦게까지 공부를 하게 되었는데, 학위 과정한다고 자주 못 보던 동생 승준이에게도 감사합니다.

저는 이웃 복이 많았던 것 같습니다. 연구실에서 지내며, 처음 봤던 이왕진과 김정모, 그리고 박희원과 소리나, 박주영, 김효태, 그리고 이충호 지금도 소식이 닿고 자주 보기는 왕진이와 경모 둘이지만, 이들에게도 깊은 감사 말씀 드립니다.

그리고, 이승민, 광진슬 이 둘 덕분에 연구실 생활이 외롭고 힘들지만은 않았습니다. 오히려 같이 이야기하고 연구도 같이하고, 스튜디오도 하고, 많은 활동을 같이 한 덕분에 즐거웠습니다. 승민이도 빨리 학위를 받았으면 좋겠습니다. 또한 이제는 여기를 떠났지만 항시 저와 승민이를 격려해 주고 계신 박민수 교수님께 심심한 감사의 말씀 드립니다. 이 논문의 전반부의 데이터 분석과 통계적 방법에 승민이와 박교수의 노력과 세심한 분석이 없었다면 제대로 수행될 수 없었을 것입니다.

현재 같은 연구실을 쓰며 같이 학위를 받은 박유진, 그리고 박사 후 과정으로 있는 김수환 박사, 석사 학위를 하고 있는 김성표와 안건 모두 제가 여러 방법론과 코딩의 문제로 고통 받고 있을 때, 통계적 방법론과 유사한 연구 결과에 대한 참고 자료를 건네 주고, 코딩 방법에 대한 힌트를 주며 저의 짐을 기꺼이 같이 들어 주는 고마운 분들입니다.

이 논문의 후반부는 고려대학교 전기전자공학부의 유재성의 난이도 있는 코딩과 꼼꼼하고 다양한 모델링이 없었다면 실험을 성공적으로 수행할 수 없었을 것입니다. 이 자리를 빌어 유재성의 노력과 수고에 깊은 감사의 말씀 드립니다.

저를 외과 의사로 만들어 주신 한양대학교의 외과학 교실의 교수님을 빼놓을 수 없습니다. 석사 지도 교수이셨던 남영수 교수님, 그리고 정민성

교수님 감사합니다. 그리고 저의 결혼식에서 기꺼이 주례를 맡아 주셨던 권성준 교수님 감사합니다. 권오정 교수님, 이경근 교수님과 최동호 교수님, 이강홍 교수님과 안병규 교수님, 하태경 교수님께도 감사의 말씀 드리고 싶습니다.

외과 수련 동기들이고 지금도 연락하며 생업에 바쁜 김신혁, 이정희, 김은진 세 분에게도 감사 드립니다.

이 외에도 이루 말할 수 없이 감사 드릴만한 분들이 많습니다. 이 모든 분들의 관심과 배려, 격려와 조언으로 이 논문이 세상에 나오게 되었습니다. 감사합니다.