

딥러닝을 이용하여 국내 노인인구의 COPD 사망 위험 추정

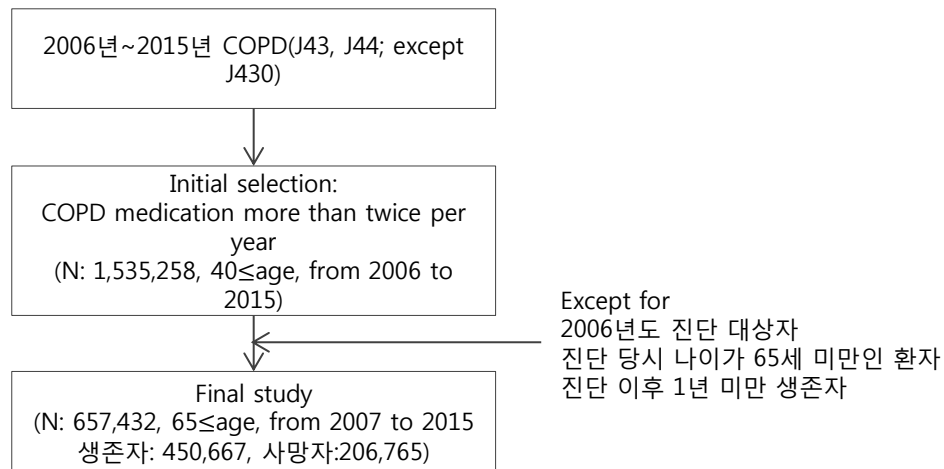
2018. 10. 04

한국환경정책평가연구원
사회환경연구부 강선아

딥러닝 이용 국내 노인인구 호흡기 질환 사망 위험 추정

- 목적: 딥러닝을 활용한 예측 의학 성과를 환경성 질환 분석에 적용
 - 예) 머신러닝 알고리즘으로 파악한 심혈관 발병 인자를 이용하여 추정한 발병 위험 추정치가 기존 학회 제공 발병 인자 이용 추정치보다 더 정확함을 확인(Stephen et al. 2017)
 - 실시간으로 갱신되는 데이터를 반영하여 결과를 update 할 수 있는 딥러닝의 장점 활용
- 내용: 만성폐쇄성 폐질환 사망 위험을 딥러닝을 이용하여 추정
 - 연구 대상 :65세 이상 만성폐쇄성폐질환(COPD) 환자
 - 2009년 현재 치료 중 환자 192,496명/ 2010년 전체 사망원인 중 7위에 해당
 - 자료 : 건강보험 맞춤형연구 DB , 2006-2015년 건강보험 코호트 DB version 2.0, 인구, 기후, 대기오염도 및 대기오염물질 배출량 자료를 연계
 - 맞춤형연구 DB: 만성폐쇄성 폐질환 질병에 영향을 끼치는 요인 분석
 - 건강보험 코호트 DB: 인구 특성 (성별, 연령) , 건강 관련 특성(병력, 식전혈당..), 진료기록
 - 기후: 기상청 제공 시군구별 기후 데이터
 - 환경자료: 대기오염물질 오염도, 대기오염물질 배출량
- 방법론: 딥러닝과 일반적인 호흡기 질환 사망위험 예측 모델링의 예측 정확도 비교
 - 머신러닝 방법론: Lag 변수를 변인(feature)으로 포함하는 ANN/시계열 분석이 가능한 RNN, LSTM 적용 점검
 - 일반적으로 알려진 위험인자: 대한결핵 및 호흡기 학회/WHO 제공

데이터 구축(일부 변경)



1차 연도

대기오염에 따른 COPD 질환자의
사망원인 분석

- 국내 COPD 질환자의 진료 데이터 수집
- 대기오염 및 기상기후요인 데이터 수집
- 로지스틱 회귀분석을 통한 사망 원인 분석

2차 연도

딥러닝을 이용한 국내 COPD 노인
환자의 사망위험 추정

- 딥러닝을 이용 사망예측 모델링 구축
- 1차연도 분석을 통해 밝혀진 사망원인을 이용한 예측 모델링과 딥러닝 이용 사망예측 모델링 정확도 비교

변수 정리

구분	변수	추출 방법 및 기준	상태
개인	나이		O
	성별	남/여	O
	거주지(도시/농촌)	대도시: 인구 50만 명 이상의 도시 중소 도시: 인구 10만 명~50만 명 소도시: 인구 10만 이하 (출처: 국토환경정보센터, http://www.neins.go.kr/ltr/intro/basic09.asp)	O
	소득	보험료 20분위	O
	직종	정무직/일반직/기능직/공안직/경찰, 소방직/ 1, 2종 고용직/ 경노무 고용직/ 교육직/ 법관, 검사/ 일용직/ 군인/ 미통보군인/연구직/공중보건의/전임전문직/청원경찰/지도직/원어민 영어교사/계약직/공익수의사	O
	장애등급	장애유형코드/ 장애등급코드(정상, 1~13등급, 기타)	O
진단	Charlson comorbidity index	다음 슬라이드 참고	코딩 완료
	입원 횟수(제거)	Copd 진단 기준 1년 동안의 입원 횟수	O
	수술여부	암/비암 구분(진단 후 1년)	X
	Copd 중증도(일부 변경)	진단 후 1년을 기준으로 copd 중증도 상태	O(재반영)
	Copd 의약품 일정기준 이상 처방	Histogram 참고 후 일정 기준 이상 사용자 구분	진행중
	외래경증악화	중등증 악화 : copd코드(J43, J44, expect for J430) + 항생제/steroid + 외래 중증 악화 : copd코드(J43, J44, expect for J430, J12~J17 (하위코드 모두 포함), I26, I26.0, I26.9, R06.0, J80) + 항생제/steroid + 입원/응급실 (진단 후 1년)	O
	산소처방전	가정산소치료, 기간/질병과 관계없이 한번이라도 처방받으면 추출	X
	진단 전 입원횟수(제거)	진단 1년 전 입원횟수	O

sub-analysis

구분	변수	추출 방법 및 기준
검진	체중	
	BMI	$G1E_WGHT / (G1E_HGHT * G1E_HGHT * 0.0001)$
	비만도	1저체중; 2정상체중; 3비만; 4고도비만;
	혈색소수치	
	혈당수치	
	콜레스테롤수치	
	심전도 검사	
	흉부방사선 검사	
	흡연 여부	<p>흡연여부 : ① 피우지 않는다 ② 과거에 피웠으나 지금은 끊었다 ③ 현재도 피운다</p> <p>흡연량 : ① 반갑 미만 ② 반갑 이상 ~ 한갑 미만 ③ 한갑 이상 ~ 두갑 미만 ④ 두갑 이상 (09년 이후 검진에서는 직접 기재하게 되어있음)</p> <p>흡연기간 : ① 5년미만 ② 5 ~ 9년 ③ 10 ~ 19년 ④ 20 ~ 29년 ⑤ 30년이상</p>

COPD 중증도

COPD 중증도 정의가 수정됨

- (1) were pre-scribed ICS&LABA&LAMA for 6 months
- (2) had at least two prescriptions for systemic steroids with ICSLABA,
- (3) had at least two prescriptions for systemic steroids + a LABA or a LAMA,
- (4) had at last two prescriptions for systemic steroids + a LABA and a LAMA prescribed together

수술 여부

수술 여부 자체는 T20_table에서 제공하나, 어떤 질병으로 수술을 받았는지 판단 불가

연도별 국강보험요양급여비용(국민보험심사평가원) 제공하는 코드를 분류하여 암과 비암 수술을 파악해야 함

- 일반처치, 수술후처치 등 제외 작업

분류번호	코 드	분 류	점 수
		[기관, 기관지 및 폐]	
		주 : 내시경 세척·소독료는 자-131-1, 자-131-2, 자-133, 자-134-1, 자-147-1을 실시한 경우에 제2장 제4절에 의하여 산정한다.	
자-130		기관절개술 Tracheostomy	
	O1300	가. 관혈적 기관절개술 Invasive Tracheostomy	2,072.29
	O1301*	나. 경피적 확장 기관절개술 Percutaneous Dilatational Tracheostomy	1,083.58
자-130-1		혼상갑상막절개술 Cricothyroidotomy	
	O1303	가. 투관침에 의한 경우 Trochar	927.92
	O1306	나. 피부절개에 의한 경우 Skin Incision	1,059.99
자-130-2	O1305	종격동 기관절개술 Mediastinal Tracheostomy	5,587.14
자-131		기관 또는 기관지 종양제거술	

Charlson comorbidity index(ICD 10-code 매칭)

질병	ICD_code
Myocardial infarction	I21.0-I21.9, I22.0, I22.1, I22.8, I22.9, I23.0-I23.8
congestive heart failure	I50, I50.0, I50.1, I50.9, I11.0, I13.0, I13.2
Peripheral vascular dis.	I70x, I71x, I73x, I74x, I77x, I78x, I79x, I80x, I81, I82x, I83x
Cerebrovascular dis.	I60-69
Dementia	F00x, F01x, F02x, F03, G30x, G31.1
Chronic pulmonary dis.	J41x-J47x, J60x-J66x, I27.8, I27.9, J68.4, J70.1, J70.3, J40.x, J67.x,
Connective tissue dis.	M30.x, L94x
Peptic ulcer dis.	K25.4-K25.7, K26.4-K26.7, K27.4-K27.7, K28.4-K28.7, K25.0-K25.3, K25.9, K26.0-K26.3, K26.9, K27.0-K27.3, K27.9, K28.0-K28.3, K28.9
Liver dis. (mild)	K70.3, K71.7, K73.x, K74.3-K74.6, B18.x, K70.0-K70.2, K70.9, K71.3-K71.5, K74.0-K74.2, K76.0, K76.2-K76.4, K76.8, K76.9, Z94.4
Liver dis. (moderate /severe)	K72.1, K72.9, K76.6, K76.7, I85.0, I85.9, I86.4, I98.2, K70.4, K71.1, K76.5
Diabetes (without complications)	E10.1, E10.9, E11.1 E11.9, E13.1, E13.9, E14.1, E14.9, E10.0, E10.6, E10.8, E11.0, E11.5, E11.6, E11.8, E12.0, E12.1, E12.6, E12.8, E12.9, E13.0, E13.5, E13.6, E13.8, E14.0, E14.5, E14.6, E14.8
Diabetes (with complications)	E10.2-E10.4, E11.2-E11.4, E13.2-E13.4, E14.2-E14.4, E10.5-E10.7, E11.5-E11.7, E12.2-E12.7, E13.5-E13.7, E14.5-E14.7
Paraplegia / hemiplegia	G81.x, G82.0-G82.2, G82.3-G82.5, G04.1, G11.4, G80.1, G80.2, G83.0-G83.4, G83.9
Renal dis.	I12.0, I13.1, N18.x, N03.x, N05.2-N05.7, N07.2-N07.4, N01.x, N19.x, N25.x, Z49.0-Z49.2, Z99.2,
Cancer	C00-D48
AIDS/HIV	B20.x-B22.x, B23.x, B24.x

출처: clinicalcodes.org; 김경훈(2010, 건강보험심사평가원)

기후 및 대기오염 데이터

	i	sigungu_c	sigungu_n	pred
2	20030101	31370	가평군	90.41764
3	20030101	11230	강남구	88.62311
4	20030101	11250	강동구	89.13578
5	20030101	32030	강릉시	81.37052
6	20030101	11090	강북구	90.09519
7	20030101	11160	강서구	89.27938
8	20030101	21120	강서구	80.8714
9	20030101	36390	강진군	55.96265
10	20030101	23310	강화군	90.30905
11	20030101	38090	거제시	77.495
12	20030101	38390	거창군	73.30078
13	20030101	37100	경산시	80.74238
14	20030101	37020	경주시	82.95442
15	20030101	34070	계룡시	74.57845
16	20030101	23070	계양구	89.1111
17	20030101	37370	고령군	76.71183
18	20030101	32400	고성군	84.83608
19	20030101	38340	고성군	73.51062
20	20030101	31101	고양시 덕	90.18133
21	20030101	31102	고양시 율	90.42557
22	20030101	35370	고창군	61.46316
23	20030101	36350	고흥군	60.35404
24	20030101	36320	곡성군	63.46482
25	20030101	34020	공주시	76.4561
26	20030101	31110	과천시	88.00147
27	20030101	11210	관악구	88.30338
28	20030101	31060	광명시	88.10199

오존, PM10, PM2.5 데이터 보간(크리깅 이용)

- 시간해상도: 2003년 1월 1일~2015년 12월 31일 일단위
- 공간해상도: 시군구 단위

평균기온, 평균습도 데이터 보간(크리깅 이용)

- 시간해상도: 2003년 1월 1일~2015년 12월 31일 일단위
- 공간해상도: 시군구 단위

향후 작업

- 미완료된 변수 추출
- 기후 및 대기오염 데이터와 결합
- GLM(logistic regression), GAM 분석