

# 일산병원과 건강보험공단 자료를 결합한 인공지능 기반 뇌졸중 및 뇌손상 환자의 발병 후 예후 예측 모델

김형섭 · 김영은 · 권남우 · 김종현 · 손강주 · 차지욱 · 김민영  
조한열 · 박덕호 · 신은경 · 손세호 · 김수영 · 배세진

*h·well*  
국민건강보험



국민건강보험

National Health  
Insurance Service

일산병원

Ilsan Hospital

공동연구보고서

NHIMC-2021-PR-021

# 일산병원과 건강보험공단 자료를 결합한 인공지능 기반 뇌졸중 및 뇌손상 환자의 발병 후 예후 예측 모델

김형섭 · 김영은 · 권남우 · 김종헌 · 손강주 · 차지욱 · 김민영  
조한얼 · 박덕호 · 신은경 · 손세호 · 김수영 · 배세진

*h·well*  
국민건강보험



[저 자]

책임 연구자:	국민건강보험 일산병원 재활의학과	김형섭
	국민건강보험공단	김영은
공동 연구원:	국민건강보험 일산병원 재활의학과	권남우
	국민건강보험 일산병원 신경과	김종헌
	국민건강보험 일산병원 연구소 연구분석부	손강주
	서울대학교 사회과학대학 심리학과	차지욱
	차의과학대학 분당차병원 재활의학과	김민영
	연세대학교 강남세브란스병원 재활의학과	조한얼
	에스포항병원 재활의학과	박덕호
	고려대학교 사회과학대학 사회학과	신은경
	서울대학교 사회과학대학 경제학과	손세호
	서울대학교 사회과학대학 심리학과	김수영
	국민건강보험공단	배세진

연구관리번호	IRB 번호
NHIS-2022-1-062	NHIMC 2021-07-038

본 연구보고서에 실린 내용은 국민건강보험 일산병원의 공식적인 견해와 다를 수 있음을 밝혀둡니다.

## 머리말

의학은 늘 변화하는 과학이다(Medicine is ever changing field of science). 경험의학에서 근거의학(evidence based medicine)에서 이제는 빅데이터 기반 정밀의학(precision medicine)이 대세가 되어 이는 재활의학 분야에서도 그 파도가 거세게 몰아치고 있다. 뇌졸중과 뇌손상은 손상의 정도에 따라 환자와 가족의 가정복귀와 사회 참여에 심각한 제한을 가져오는 질병으로, 이에 대한 재활치료와 적절한 사회 복귀 프로그램이 필요하다. 급성기 기능 회복의 재활치료 이후 만성기 유지기 치료까지, 사회 복귀 이후에 지역사회재활 (community-based rehabilitation)이 연속적으로 이어져 있음에도 이에 대한 분명한 이정표(milestone)가 명확하지 않았다. 한편, 병원 자료와 공단 자료를 연속적으로 모으게 된다면, 병전의 상태, 급성기 신경학적 상태 및 내과적인 상태, 가정 복귀 이후의 의료이용 및 건강형태를 안다면, 이에 대한 구체적인 재활치료의 근거와 목표를 제시할 수 있게 된다. 이번 연구는 국민건강보험 일산병원에 입원한 뇌졸중과 뇌손상 환자의 병원자료와 공단 결합 자료를 이용한 뇌졸중 및 뇌손상 환자의 기계학습 모델을 이용한 예후 예측 모델로, 뇌졸중 환자와 뇌손상 환자의 생존 및 사망, 사망 위험 인자 분석을 통해, 내과적 합병증 예방 및 뇌손상 이후 기능 향상 방법을 확인하여, 적정 재활 치료의 근거를 제공하고자 하였다. 비록 단일 병원 자료이므로 대표성을 지적받을 수 있으나, 여기에 공개한 coding 자료를 통하여, 누구나 자신의 병원 자료를 결합할 수 있으므로 추후 국내 여러 병원과 공단 자료 결합 연구를 통해서 재현성과 대표성을 구현할 수 있으리라 생각한다. 끝으로 본 보고서에서 저술한 내용은 저자들의 의견이며, 보고서 내용상의 하자가 있는 경우 저자들의 책임으로 국민건강보험 일산병원 연구서의 공식적인 견해가 아님을 밝혀둔다.

2023년 5월

국민건강보험공단 이사장

형재룡

국민건강보험 일산병원장

김성우

국민건강보험공단 건강보험연구원장

이해중

국민건강보험 일산병원 연구소장

이천준



# 목차

요약 .....	1
제1장 서론 .....	11
제1절 통계 분석과 기계 학습(Machine learning)과의 차이 .....	16
제2절 선행 연구 분석 .....	25
제2장 연구방법 .....	29
제1절 병원 자료 .....	31
제2절 국민건강보험 맞춤형자료 .....	32
제3절 대상자 추출 과정 .....	33
제4절 분석 방법 .....	34
제3장 결과 .....	35
제1절 노인 장기요양보험 서비스 이용자군의 생존 분석의 기계학습(machine learning) .....	59
제2절 노인 장기요양보험 서비스 미이용자군의 기계학습(Machine Learning) 분석 결과 .....	70
제4장 결론 .....	81
참고문헌 .....	87
부록 .....	93

## 표목차

〈표 3-1〉 노인 장기요양보험 서비스 여부에 따른 인구학적 특성 및 병원 자료의 비교 ..	38
〈표 3-2〉 노인 장기요양보험 서비스 이용자군의 생존 여부에 따른 인구학적 특성 및 병원 자료의 비교 .....	44
〈표 3-3〉 노인 장기요양보험 서비스 이용자군의 생존군과 사망군의 비교 .....	49
〈표 3-4〉 노인 장기요양보험 서비스 미이용자군의 생존 여부에 따른 인구학적 특성 및 병원 자료의 비교 .....	53
〈표 3-5〉 장기요양보험 이용자군의 생존 분석의 NN 기계 학습 주요 인자 .....	59
〈표 3-6〉 NN의 실험 세트 검증 .....	60
〈표 3-7〉 SVM의 실험 세트 검증 .....	61
〈표 3-8〉 KNN의 실험 세트 검증 .....	61
〈표 3-9〉 노인 장기요양보험 서비스 이용자군의 생존 분석의 RF기계 학습 주요 인자 ..	61
〈표 3-10〉 RF의 실험 세트 검증 .....	62
〈표 3-11〉 노인 장기요양보험 서비스 이용자군의 생존 분석의 GBM기계 학습 주요 인자 .....	63
〈표 3-12〉 GBM의 실험 세트 검증 .....	64
〈표 3-13〉 노인 장기요양보험 서비스 이용자군의 생존 분석의 r_part 기계 학습 주요 인자 ..	65
〈표 3-14〉 r-Part의 실험 세트 검증 .....	66
〈표 3-15〉 노인 장기요양보험 서비스 이용자군의 생존 분석의 Xgboost .....	66
〈표 3-16〉 Xgboost의 실험 세트 검증 .....	67
〈표 3-17〉 노인 장기요양보험 서비스 이용자군의 인공지능 알고리즘에 따른 정확도와 AUC value .....	67

〈표 3-18〉 노인 장기요양 서비스 이용군의 생존 회귀 분석에서 상위 20개 변수 .....	68
〈표 3-19〉 노인 장기요양 서비스 이용군의 기계학습 알고리즘에 따른 주요 인자의 비교 .....	69
〈표 3-20〉 장기요양보험 미이용자군의 사망에 관련한 NN 기계 학습 주요 인자 .....	70
〈표 3-21〉 NN의 실험 세트 검증 .....	71
〈표 3-22〉 SVM의 실험 세트 검증 .....	71
〈표 3-23〉 KNN의 실험 세트 검증 .....	72
〈표 3-24〉 노인 장기요양보험 서비스 미이용자군의 생존 분석의 RF기계 학습 주요 인자 .....	72
〈표 3-25〉 RF의 실험 세트 검증 .....	73
〈표 3-26〉 노인 장기요양보험 서비스 미이용자군의 생존 분석의 GBM 기계 학습 주요 인자 .....	74
〈표 3-27〉 GBM의 실험 세트 검증 .....	75
〈표 3-28〉 노인 장기요양보험 서비스 미이용자군의 생존 분석의 r_part 기계 학습 주요 인자 .....	75
〈표 3-29〉 r-Part의 실험 세트 검증 .....	76
〈표 3-30〉 노인 장기요양보험 서비스 미이용자군의 생존 분석의 Xgboost 기계 학습 주요 인자 .....	77
〈표 3-31〉 Xgboost의 실험 세트 검증 .....	78
〈표 3-32〉 노인 장기요양보험 서비스 미이용자군의 생존 분석에서 기계학습 방법에 따른 정확도와 AUC value .....	78
〈표 3-33〉 전체 노인 장기요양 서비스 미이용군의 생존 회귀 분석에서 상위 20개 항목 .....	79
〈표 3-34〉 전체 노인 장기요양 서비스 미이용자군의 기계학습 알고리즘 분석에 따른 주요 인자의 비교 .....	80

## 그림목차

[그림 1-1] 기계학습 분석 모식도 .....	18
[그림 1-2] 주요 지도 기계학습 알고리즘 모식도 .....	23
[그림 1-3] 주요 비지도 기계학습 알고리즘 모식도 .....	24
[그림 2-1] 자료 결합 형태 .....	33
[그림 2-2] 자료 결합 모식도 .....	34
[그림 3-3] 장기요양보험 이용자군의 생존 분석의 NN 기계 학습 주요 인자 .....	60
[그림 3-4] 노인 장기요양보험 서비스 이용자군의 생존 분석의 RF기계 학습 주요 인자 .....	62
[그림 3-5] 노인 장기요양보험 서비스 이용자군의 생존 분석의 GBM기계 학습 주요 인자 .....	64
[그림 3-6] 노인 장기요양보험 서비스 이용자군의 생존 분석의 r-part 기계 학습 주요 인자 .....	65
[그림 3-7] 노인 장기요양보험 서비스 이용자군의 생존 분석의 Xgboost .....	67
[그림 3-8] 노인 장기요양보험 서비스 미이용자군의 사망에 관련한 NN 기계 학습 주요 인자 .....	71
[그림 3-9] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 RF기계 학습 주요 인자 .....	73
[그림 3-10] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 GBM기계 학습 주요 인자 .....	74
[그림 3-11] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 r-part 기계 학습 주요 인자 .....	76
[그림 3-12] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 Xgboost 기계 학습 주요 인자 .....	77



# 요약



## 요약

### 1. 연구 배경과 목적

뇌졸중과 외상성 뇌손상의 후유증을 최소화하고, 퇴원 후 환자의 일상생활 복귀를 적절하게 준비하기 위해서는 예후를 예측하여 적절한 재활치료를 수립하는 것이 중요하다. 기존의 예후 예측 모델은 예측에 대한 민감도 및 특이도가 높지 않고, 각 변수들을 현실에서 적용하기 어렵다는 제한점이 있다. 따라서, 새로운 신기술을 통한 예후 예측에 대한 필요성이 대두되었다. 특히 기계 학습과 딥러닝 기술이 단순 점수 합산의 고식적인 방법 보다 예측도가 더욱 높다는 것이 밝혀지면서 임상 진단과 예후 예측에 많이 사용되기 시작하였다.

뇌졸중과 외상성 뇌손상 환자에서 임상양상 양상은 원인에 따라 다르지 않고 침범한 부위에 따라 언어, 편측 무시가 나타나기도 하며, 실행증(apraxia)과, 정동 장애(mood disorder)가 나타난다. 따라서 국내에서는 뇌병변 장애의 범주에 뇌졸중 환자와 외상성 뇌손상을 함께 포함하고 있다.

한편, 우리의 선행 연구 “국민건강보험 빅데이터를 이용한 퇴행성 뇌질환(뇌졸중, 치매, 파킨슨) 진단 이후의 장기 요양으로 전달되는 의료체계 개선 연구 (2009)”에 의하면 2009년 초발 뇌졸중 환자 10,549명 환자 중에 33.69%가 2년 이내 사망하고, 중증 뇌졸중 환자는 초기의 충분한 재활치료를 받았을 경우에는 그렇지 않은 군에 비해 생존율을 높음을 알 수 있었으나, 경증 환자일 경우에는 재활치료의 빈도와 생존율과 큰 차이가 없음을 확인하였다. 따라서 중증 환자의 경우 합병증 방지를 통한 사망률을 낮추는 것이 치료 전략이 되어야 할 것이고, 경증 환자의 경우 초기의 기능회복을 유도한 후 적극적인 가정이나 사회 복귀가 치료 전략이 되어야 할 것이다.

이번 연구를 통해 일산병원 자료와 공단 자료의 결합을 통해 인공지능 분석이 가능한지 확인하는 것이 일차적인 목표이며, 이차적으로는 병원 자료를 통해서 뇌손상 이후에 생존과 사망에 영향을 미치는 인자를 분석하고자 하였으며, 뇌손상 환자의 사망과 관련된 인자 분석을 통해서 환자들과 가족들에게 합리적인 재활치료 전략 및 재활의료 전달



체계를 마련하는 근거를 제공하는 것이 이 연구의 최종 목표라 하겠다.

## 2. 연구방법

2009년부터 2018년 10년간 20세 이상 성인을 대상으로 본원에서 뇌졸중 및 외상성 뇌손상, 편마비, 및 사지 마비로 진단된 환자를 포함 (KCD I6\*, G81\*, G82\*, S06\*) 하고, 뇌영상의학 검사를 본원에서 시행하고 뇌영상 판독 결과가 있는 자로 제한하였다. 우리 병원 입원 이전에 장애 등록된 환자는 제외하였으며, 사지 마비 중에도 척수 손상을 시사하는 사지마비(Tetraplegia)와 양하지마비(Paraplegia) 환자는 제외하였다.

병원 자료로는 인구사회학적 변수: 성별, 나이, 출생 및 사망 일자, 질병 코드 및 진단명, 혈액검사, 진료 경과, 영상의학검사 등을 수집하였으며, 국민건강보험 맞춤형 자료에서는 사망 및 생존기간, 장애유형 및 등급, 등록일자, 거주 지역, 가입자 유형, 보험료 분위를 조사하였다.

대상자 추출 과정은 다음과 같다. 대상자는 2009-2018 사이에 사람은 진단명이 뇌손상 및 마비로 (KCD I6\*, G81\*, G82\*, S06\*) 입원한 환자(n=9,741) 중 20세 미만 (n=340), 사회 경제적 자료가 없는 사람(n=758) 을 제거하여, 총 8,643 명이 자료 결합에 성공하였다. 이중 척수 손상을 의미하는 (Tetraplegia, Paraplegia) 환자 n=569명을 제거하였다. 또한 뇌영상 (CT, MRI) 판독지가 없는 경우 (n=2821), 입원 일수가 1일로 응급실에만 있다 퇴원한 환자 464명을 제외하여, 총 4,789명이 이 연구에 포함되었다. 분석은 R version 3.5.2 (2018-12-20)을 이용하여 인공지능 및 통계분석을 시행하였다. 결측치 처리는 각항목의 평균치를 삽입하여 계산하였다. 분석한 기계학습 알고리즘은 Neural Network(NN), Random Forest(RF), Gradient boosting machine(GBM), K-nearest neighbor method (K-NN), Support vector machine(SVM), r-part, Extreme Gradient boost(Xgboost), Caret Ensemble을 이용하였다. 또한 생존 사망에 대해서 linear regression 분석을 하여, 인공지능 분석과 비교하였다.

## 3. 결과

노인 장기요양보험 서비스 이용자군은 1,839명으로 노인 장기요양 변수를 포함한 237개 변수였으며, 반면에 미이용자군은 2,950명으로 147개 변수이며 각군의 인공지능 분석에서는 훈련세트를 70%, 검증세트로 30%를 나누었으며, 총 변수 중에 주요 20개만 추출하였다.

노인 장기요양보험 서비스를 이용자군이 전체 사망자가 962명(52.3%)으로, 미이용자군 695명(23.6%) 보다 많았으며, 발병 후 생존기간도 1,668.1±1,135.4(일)로 서비스를 받지 않은군의 1,831.7±1,114.6 보다 짧았다. 성별로는 노인 장기요양보험 서비스를 이용자군에서 여성이 1031(56.1%)명으로 이는 평균 연령이 여자가 높았다. 장기요양보험 서비스를 이용자군은 입원중 총입원기간, 일반병동 입원기간, 중환자실 입원기간, 총 협진횟수, 호흡기 및 감염내과 협진, 위루술, L-tube 삽입, 입원중 폐렴 여부, 기관절개 관삽입(T-tube), 욕창발생, Albumin 및 혈색소(Hb) 저하, 공복혈당 및 당화혈색소, BUN, Cr의 상승이 서비스를 받지 않는 군보다 통계적으로 차이가 있게 관찰되었다.

노인 장기요양 서비스 이용자군에서 사망자군이 생존기간 1234.4±1090.9일로 생존군(2143.8±982.9)보다 짧았다. 성별은 양군간의 차이가 없었으며 사망군은 생존군에 비해 입원중 총입원기간, 중환자실 입원기간, 호흡기 내과 협진, 위루술, L-tube 삽입, 입원중 폐렴 여부, 기관절개 관삽입(T-tube), 욕창발생, Albumin 저하, 공복혈당 및 당화혈색소, BUN/Cr의 상승이 생존군 보다 통계적으로 차이가 있게 관찰되었다.

노인 장기요양보험 서비스 이용자군에서 사망자군에서 인증조사 점수가 높았으며, 사망자 군에서 판정 인정등급이 높았다. 일상생활 환산 점수, 인지 환산점수, 행동 환산 점수, 간호 환산 점수, 재활환산 점수, 흡인, 산소요법, 욕창간호, 경관 영양, 도뇨, 간호 투석, 약챙겨먹기, 완전 와상상태, 치매 완전 의존 등 모두 사망군에서 높았다.

노인 장기요양보험 서비스 미이용자의 생존군과 사망자 군간의 비교에서, 사망자군이 생존기간 1,234.4±1,090.9일로 생존군(2,175.0±921.7) 보다 짧았다. 또한 연령 또한 사망자군이 71.3±13.0세로 생존군의 59.7±12.5세로 볼 때 사망자군이 노인 장기요양보험 서비스를 받지 못하고 발병후 조기에 사망하여 받지 못한 것으로 여겨진다. 성별은 양군간의 차이가 없었으며 사망군은 생존군에 비해 입원중 총입원기간, 일반병동 입원기간, 중환자실 입원기간, 총 협진횟수, 호흡기 및 감염내과 협진, 위루술, L-tube 삽입, 입원중 폐렴 여부, 기관절개 관삽입(T-tube), 욕창발생, Albumin 및 혈색소(Hb) 저하, 공복혈당 및 당화혈색소, BUN/Cr의 상승이 통계적으로 차이가 있었다.

노인 장기요양보험 이용자군의 기계학습은 NN, SVM, KNN, RF, GBM, r-part Xgboost를 이용하였으며, 정확도(Accuracy)와 AUC(Area under curve)는 다음과 같다 <표 1>.

<표 1> 노인 장기요양보험 서비스 이용자군의 인공지능 알고리즘에 따른 정확도와 AUC value

	NN	SVM	KNN	RF	GBM	r-part	Xgboost
정확도	0.8161	0.8105	0.7073	0.8368	0.8199	0.7041	0.8368
AUC	0.8157	0.8082	0.7104	0.8335	0.8183	0.7041	0.8337

정확도와 AUC는 Xgboost에서 가장 좋은 결과를 나타내었으나 여러 인공지능 분석 tool 사이에서 Ensemble model 분석에서 가장 최적의 모델은 Random Forest로 밝혀졌으며, 정확도는 0.8643, AUC는 0.8639로 측정되었다. (appendix page 197p)

상위 20개 변수는 각 알고리즘에 따라 순위가 다르지만, 환자의 기능과 내과적 지표에 따라 많이 다른 것을 확인할 수 있었다. 노란색 부위는 각 알고리즘에서 20개 상위 변수에서 공통된 것만 표시한 것으로 2차 장애 1등급 및 2 등급, 장기요양등급구분 저소득으로, 뇌손상 이후 기능과 경제력 상태가 생존에 지대한 영향을 미침을 알 수 있다 <표 2>.

〈표 2〉 노인 장기요양보험 서비스 이용자군의 기계학습 알고리즘에 따른 주요 인자의 비교

순위	NN	RF	GBM	Xgboost	Logistic Regression
1	2차 장애 1등급	장기요양경감구분_저소득	장기요양경감구분_저소득	장기요양경감구분_저소득	2차 장애 1등급
2	주수발자_자녀	주수발자_머느리,사위	ALBUMIN	인증조사점수	연령
3	대구광역시	장기요양경감구분_건강보험	인증조사인정점수	ALBUMIN	2차 장애 2등급
4	2차 장애 2등급	ALBUMIN	연령	연령	ALBUMIN
5	장기요양경감구분_저소득	연령	주수발자_머느리, 사위	주수발자_머느리,사위	수발자(머느리, 사위포함)
6	동거인_부모	인증조사점수	장기요양경감구분_건강보험	CR	인증조사인정점수
7	장기요양경감구분_값없음	2차 장애 2등급	CR	HB	장기요양경감구분_건강보험
8	2차 장애 3등급	2차 장애 5등급	2차 장애 1등급	장기요양경감구분_건강보험	뇌출혈(+)
9	1차 청각장애인	2차 장애 3등급	주수발자_자녀	2차 장애 1등급	L_tube_feeding(+)
10	1차 지적장애인	주수발자_자녀	공복 혈당 (GLUCOSE_AC)	BUN	성별_여자
11	경상남도	2차 장애 1등급	2차 장애 2등급	공복 혈당 (GLUCOSE_AC)	약행겨머기_부분의존
12	의료급여	CR	BUN	2차 장애 2등급	장기요양경감구분_저소득
13	노인요양공동생활가정	HB	2차 장애 5등급	HBA1C	치매_완전의존
14	2차 청각장애인	간호환산점수	총입원기간	총입원기간	multifocal_large(+)
15	2차 장애 5등급	등급판정인정등급코드_5등급	2차 장애 3등급	주수발자_자녀	약행겨머기_완전의존
16	장기요양경감구분_저소득	2차 장애 4등급	HB	총합진횟수	장기요양대상자구분코드
17	장기요양경감구분_만성질환	주수발자_시설관계자	간호환산점수	일반병동 입원기간	장기요양경감구분_값없음
18	2차 장애 6등급	2차 장애 6등급	L_tube_feeding,1	간호환산점수	페렴(+)
19	2차 장애 4등급	등급판정인정등급코드_4등급	당화혈색소(HBA1C)	2차 장애 5등급	등급판정인정등급코드_값없음
20	주수발자_머느리, 사위	장기요양경감구분_만성질환	주수발자_시설관계자	인지환산점수	Temporal lobe(+)

노인 장기요양보험 미이용자군의 기계학습도 이용자군과 동일하게 NN, SVM, K-nearest method, random forest, gredient boosting machine, r-part, Xgboost를 이용하였으며, 정확도 AUC는 다음과 같다 <표 3>.

<표 3> 노인 장기요양보험 서비스 미이용자군의 인공지능 알고리즘에 따른 정확도와 AUC value

	NN	SVM	KNN	RF	GBM	r-part	Xgboost
정확도	0.8711	0.9518	0.8354	0.9533	0.9471	0.8124	0.9611
AUC	0.8110	0.8299	0.7138	0.7798	0.8054	0.6652	0.8102

노인 장기요양보험 서비스 미이용자군의 정확도와 AUC는 노인 장기요양보험 서비스 이용자군의 결과처럼, Xgboost에서 가장 좋은 결과를 나타내었으나 여러 인공지능 분석 tool 사이에서 Ensemble model 분석에서 가장 최적의 모델은 Random Forest로 밝혀졌으며, 정확도는 0.8711, AUC는 0.8010로 측정되었다. (appendix page 268p)

노인 장기요양보험 이용자군과 다른 점은 Albumin, multifocal\_large(+), L\_tube\_feeding(+), 폐렴(+), 공복혈당 (AC\_glucose)과 같이 병원 입원중의 환자의 내과적 및 신경과적 상태가 상위 중요 인자로 관찰되었다. 상위 20개 변수는 각 알고리즘에 따라 순위가 다르지만, 상위 20개 알고리즘은 유사한 양상을 보인다. 다만 NN에서는 장애등급과 장애 종류가 주요 상위 20개 항목이었다 <표 4>.

<표 4> 노인 장기요양보험 서비스 미이용자군의 기계학습 알고리즘에 따른 주요 인자의 비교

순위	NN		RF		GBM		Xgboost		Logistic Regression	
	연령	연령	연령	연령	연령	연령	연령	연령	연령	연령
1	2차 장애 1등급	연령	연령	연령	연령	연령	연령	연령	연령	연령
2	2차 지적장애인	ALBUMIN	ALBUMIN	ALBUMIN	ALBUMIN	ALBUMIN	ALBUMIN	ALBUMIN	2차 장애 1등급	
3	2차 장애 5등급	2차 장애 3등급	L_tube_feeding(+)	L_tube_feeding(+)	L_tube_feeding(+)	L_tube_feeding(-)	L_tube_feeding(-)	L_tube_feeding(+)	L_tube_feeding(+)	
4	1차 간질장애인	L_tube_feeding(+)	CR	CR	BUN	BUN	BUN	폐렴(+)	폐렴(+)	
5	1차 장애 6등급	2차 장애 5등급	BUN	BUN	공복 혈당 (GLUCOSE_AC)	공복 혈당 (GLUCOSE_AC)	공복 혈당 (GLUCOSE_AC)	multifocal_large(+)	multifocal_large(+)	
6	1차 장애 1등급	L_tube_feeding(-)	공복 혈당 (GLUCOSE_AC)	공복 혈당 (GLUCOSE_AC)	CR	CR	ALBUMIN	ALBUMIN	ALBUMIN	
7	2차 간질장애인	BUN	폐렴(+)	폐렴(-)	폐렴(-)	폐렴(-)	2차 장애 2등급	2차 장애 2등급	2차 장애 2등급	
8	2차 지적장애인	폐렴(-)	총임원기간	총임원기간	총임원기간	총임원기간	군지역-음	군지역-음	군지역-음	
9	2차 자폐성장애인	폐렴(+)	HB	HB	HB	HB	IVH(+)	IVH(+)	IVH(+)	
10	2차 언어장애인	CR	총협진횟수	총협진횟수	HBA1C	HBA1C	HB	HB	HB	
11	1차 언어장애인	multifocal_large(+)	multifocal_large(-)	multifocal_large(-)	multifocal_large(-)	multifocal_large(-)	Coil(+)	Coil(+)	Coil(+)	
12	2차 장애 6등급	multifocal_large(-)	폐렴(-)	폐렴(-)	총협진횟수	총협진횟수	성별_여자	성별_여자	성별_여자	
13	2차 장애 2등급	총협진횟수	HBA1C	HBA1C	일반병동 입원기간	일반병동 입원기간	공복 혈당 (GLUCOSE_AC)	공복 혈당 (GLUCOSE_AC)	공복 혈당 (GLUCOSE_AC)	
14	간장애인	의료급여	L_tube_feeding (-)	L_tube_feeding (-)	폐렴(+)	폐렴(+)	1차 장애 1등급	1차 장애 1등급	1차 장애 1등급	
15	세종특별자치시	군지역-음	일반병동 입원기간	일반병동 입원기간	L_tube_feeding(+)	L_tube_feeding(+)	Aneurysm(+)	Aneurysm(+)	Aneurysm(+)	
16	1차 장애 5등급	총임원기간	2차 장애 5등급	2차 장애 5등급	2차 장애 5등급	2차 장애 5등급	호흡기 협진	호흡기 협진	호흡기 협진	
17	1차 장애 3등급	HB	군지역-음	군지역-음	2차 장애 1등급	2차 장애 1등급	1차 장애 3등급	1차 장애 3등급	1차 장애 3등급	
18	2차 뇌병변장애인	공복 혈당 (GLUCOSE_AC)	2차 장애 1등급	2차 장애 1등급	multifocal_large(+)	multifocal_large(+)	1차 장애 2등급	1차 장애 2등급	1차 장애 2등급	
19	1차 뇌병변장애인	일반병동 입원기간	2차 장애 3등급	2차 장애 3등급	2차 장애 2등급	2차 장애 2등급	brain_stem(+)	brain_stem(+)	brain_stem(+)	
20	1차 지적장애인	2차 장애 1등급	multifocal_large(+)	multifocal_large(+)	2차 장애 3등급	2차 장애 3등급	광역시-음	광역시-음	광역시-음	

## 4. 결론

이제까지 병원 자료와 국민건강 보험 자료를 결합한 자료로 뇌손상 환자의 예후에 대한 기계학습 분석을 한 것이 최초 시도이며, 이를 통해 병원의 입원 자료가 뇌손상 환자의 예후 인자와 어떤 관련이 있는지 파악 할 수 있었다. 노인 장기요양보험 서비스 이용군 및 미이용군, 모두 생존 예측 모델은 Xgboost가 가장 높은 정확도와 AUC 값을 보였으나, RF가 ensemble 분석에서 최적의 모델로 선정되었다. 인공지능 분석 알고리즘에 따라서 주요 변수가 순위에서만 차이가 나고 거의 유사한 양상이 관찰되었으며, 기계 학습 알고리즘 분석 결과와, 통계분석 방법인 logistic regression 결과 값의 비교에서도 20개 상위 주요 항목은 순위의 차이가 있지만 비슷한 결과 값을 보였다. 노인 장기요양보험 서비스 이용군의 RF분석에서 주요 변수를 살펴보면, 장기요양경감 구분, 두번째 제출한 장애등급, 장기요양 판정 등급, 내과적 상태를 반영하는 Albumin, HB, 연령이 주요 변수였으며, 노인 장기요양 미이용자군의 경우도 비슷하나, 이는 노인 장기요양보험 서비스 이용자군과는 달리, L-tube feeding, 폐렴여부, 뇌병변의 상태, 총입원기간, 일반 병동 입원기간, 협진 횟수등, 병원의 자료가 주요 변수였다. 특히 주목하여야 할 것은 연령과 albumin, HB, CR이 노인 장기요양보험 서비스 이용자군 및 미이용자군 모두에 내과적 질환 및 영양상태가 뇌손상 이후 사망을 예측하는 인자임을 확인할 수 있었다.

# 제 1 장

## 서론

제1절 통계 분석과 기계 학습(Machine learning)과의 차이 16

---

제2절 선행 연구 분석 25

---





# 제 1 장

## 서론

최근 인공지능 분석이 산업계, 의학계를 포함한 다양한 학문 분야에 도입되고 있다. 이는 자료 확보가 이전과 일일이 연구자에 의해서 수집한 것과 달리, 컴퓨터 하드웨어의 발전, 인공지능 알고리즘의 개발, 인터넷의 발전과 전 세계적인 보급으로 인해 수집된 빅데이터, 다양한 공공 빅데이터 개방이 배경으로 지적할 수 있다.

인공지능은 인간의 지능을 모방하여 사람이 하는 일을 기계가 할 수 있도록 하는 기술이다. 인공지능을 구현하는 방법에는 기계 학습과 딥러닝이 있는데, 딥러닝이 기계 학습에 포함되는 개념이다. 기계 학습과 딥러닝 둘 다 학습 모델을 바탕으로 데이터를 분류하는 기술이나, 동작 원리에 차이가 존재한다. 기계 학습은 인간이 먼저 전처리한 데이터를 바탕으로 판단하고 예측하는 기술로, 데이터의 특징을 스스로 추출하지 못한다. 반면 딥러닝은 대량의 데이터에 신경망을 적용하여 컴퓨터 스스로가 데이터의 특성을 추출하고 분류한 뒤 판단하는 기술이다.

대표적인 기계 학습의 학습 알고리즘에는 지도학습, 비지도학습, 강화학습이 있다. 지도학습은 컴퓨터에게 정답을 알려주고 학습시키는 방법으로, 대표 알고리즘으로는 서포트 벡터 머신(Support Vector Machine), 결정 트리(Decision tree), 로지스틱 회귀(Logistic regression), 선형 회귀(Linear regression)가 있다. 비지도 학습은 정답을 알려주지 않고, 특징이 비슷한 데이터를 범주화하여 예측하는 학습 방법으로, 대표 알고리즘으로는 K-평균 군집화(K-means clustering), 주성분 분석(Principal component analysis)가 있으며, 강화 학습은 행동에 대한 보상을 받으며 학습하는 것으로, 알고리즘으로는 마르코프 결정과정 (Markov decision process)가 있다.

딥러닝은 인간의 신경망 원리를 모방하여 심층 신경망 이론을 기반으로 고안된 학습방법을 말한다. 다층 구조의 신경망과 이를 학습시키는 역전파 알고리즘은 오래 전부터 알려져 있었으나, 최근 데이터의 증가, 그래픽카드와 같은 병렬처리 하드웨어가 발달하면서 매우 깊은 신경망의 학습이 가능하게 되었다. 하지만, 아직은 사람이 결정해 줘야 하는 변수인 하이퍼파라미터가 많기 때문에, 이 변수들을 변경해가며 학습하는 과정에

많은 노력이 필요하다. 딥러닝의 지도학습 알고리즘에는 이미지 분류, 인식, 분할 등이 가능한 합성곱 신경망(Convolutional neural network, CNN)과 시계열 데이터를 분류할 때 사용하는 순환 신경망(Recurrent neural network, RNN), LSTM(Long short term memory)가 있다. 이외에 워드 임베딩, 군집과 같은 비지도 학습이 있고, 이미 훈련된 모델을 활용하여 학습하는 기법인 전이학습이 있다.

기계 학습과 딥러닝 알고리즘의 발전으로 이를 이용하여 전통적인 예후 예측 인자의 한계를 극복하려는 시도들이 많이 이루어지고 있다. 뇌졸중과 뇌손상의 예후 예측을 위해서 다양한 기계 학습과 딥러닝 알고리즘이 사용되고 있다. 기계 학습 알고리즘 중에서는 랜덤포레스트, 서포트 벡터 머신, 선형 회귀, 결정 트리와 같은 알고리즘이, 딥러닝에서는 심층신경망(Deep neural network, DNN)과 같은 알고리즘이 사용되고 있다. 기계 학습은 데이터의 양에 비례하여 성능이 향상되는 경향성을 보이므로 많은 양의 데이터가 필요하고, 더하여 가공된 양질의 데이터가 필요하다. 이런 점에서 공단의 빅데이터는 기계 학습과 딥러닝을 적용하기에 적절한 것으로 사료된다.

뇌졸중은 뇌혈관의 폐쇄 또는 파열로 인하여 갑작스럽게 국소적인 신경학적 장애가 발생하여 24시간 이상 지속되는 경우로 정의한다. 2019년 World Stroke Organization의 보고에 의하면, 전세계적으로 초발 뇌졸중 환자는 매년 1,367만명에 달하며, 유병률은 8,000만명으로 추정되고, 매년 550만명이 뇌졸중으로 사망한다. 2018년 뇌졸중 역학보고서에 따르면 국내에서는 매년 10만여명의 뇌졸중 환자가 발생하며, 유병률은 약 80만명으로 추정된다.

뇌졸중은 성인 장애 원인 질환 중 가장 많은 부분을 차지하며, 다양한 기능 장애를 초래한다. 병변 위치에 따라 뇌의 특정 영역이나 신경계가 손상되면 특정 신경학적 증상이 나타난다. 손상 영역에 따라 상하지 근력 저하, 보행 장애, 균형 장애, 인지기능 장애, 무시 증후군, 실어증, 우울증, 통증 등의 장애가 나타날 수 있으며, 후유증이 남기도 한다. 후유증으로 인해 보행 장애, 일상생활 동작 수행 능력의 저하 시에는 주변의 도움이 지속적으로 필요할 수 있으며 이로 인해 사회 경제적 부담 및 의료 비용이 증가하게 된다.

외상성 뇌손상이란 외부로부터의 물리적인 충격에 의해 뇌가 손상 받는 것으로, 원인은 교통사고, 산업재해, 추락, 폭력 및 스포츠 손상 등이다. 2016년의 보고에 따르면 외상성 뇌손상은 전세계적으로 매년 2,700만명정도가 발생하는 것으로 추정된다. 외상성 뇌손상은 일차적 뇌손상과 이차적 뇌손상으로 나눌 수 있는데, 일차적 뇌손상이란

충격 당시 직접적인 물리적 힘에 의한 손상이며, 그 종류에는 미만성 축삭 손상, 뇌좌상, 기능해리, 두개내 출혈이 있다. 이차적 뇌손상은 일차적 손상에 의한 이차적 병리생리적 변화에 의한 손상이다. 두개내 출혈, 부종 등에 의해 두개강 내압이 증가하면 뇌혈류 공급이 제한되어 허혈성 뇌손상이 발생하기도 한다. 외상성 뇌손상 환자는 신체적 기능의 저하, 인지 및 행동의 장애를 보이는 경우가 많아 인지 기능 재교육과 교정에 둔 포괄적인 재활치료가 필요하다. 특히 뇌손상 환자는 젊은 사람에게 발생하는 경우가 많기에 노동력 상실이라는 문제를 야기하므로, 더욱 적극적인 재활프로그램이 필요하다.

따라서, 뇌졸중과 외상성 뇌손상의 후유증을 최소화하고, 퇴원 후 환자의 일상생활 복귀를 적절하게 준비하기 위해서는 예후를 예측하여 적절한 재활치료를 수립하는 것이 중요하다. 뇌졸중의 예후 예측 위해서는 전통적으로 여러 점수 체계가 사용되었다. 먼저, NIHSS(National Institutes of Health Stroke Scale)는 의식, 시야, 안면마비, 근력, 감각, 실어증 등으로 점수를 매기며, 점수가 높을수록 뇌졸중 이후 중증 후유증과 사망률 확률이 높았다. THRIVE(Total Health Risk In Vascular Events) 스코어는 NIHSS 점수, 나이, 만성 질환 등으로 점수를 매겨 뇌졸중 환자의 신경학적 예후와 사망률을 예측한다. ABCD2 스코어는 나이, 혈압, 임상양상과 지속시간, 당뇨 등으로 점수를 매기며 일과성 허혈성 발작 이후에 뇌졸중 발생의 위험도를 평가하고 예측한다. 뇌손상의 경우는 눈뜨기, 운동반응, 언어 반응을 종합적으로 평가한 GCS(Glasgow coma scale) 점수의 총점을 구하여 점수가 높을수록 예후가 좋다고 예측한다.

하지만, 위에서 살펴보았던 점수를 단순 합산하여 예후를 예측하는 모델은 예측에 대한 민감도 및 특이도가 높지 않고, 각 변수들을 현실에서 적용하기 어렵다는 제한점이 있다. 따라서, 새로운 신기술을 통한 예후 예측에 대한 필요성이 대두되었다. 특히 기계 학습과 딥러닝 기술이 단순 점수 합산의 고식적인 방법 보다 예측도가 더욱 높다는 것이 밝혀지면서 임상 진단과 예후 예측에 많이 사용되기 시작하였다.

뇌졸중과 외상성 뇌손상은 원인이 다르지만, 나타나는 양상은 뇌의 국소부위의 침범되는 것에 따라 다르게 나타난다. 뇌손상이 있을 경우, 인지 영역의 경우 침범한 부위에 따라 언어, 편측 무시가 나타나기도 하며, 실행증(apraxia)과, 정동 장애 (mood disorder)가 나타난다. 따라서 국내에서는 뇌병변 장애의 범주에 뇌졸중 환자와 외상성 뇌손상을 함께 포함하고 있다.

한편 국민건강보험 자료 (국민건강보험자료 공유서비스 <https://nhiss.nhis.or.kr/bd/ay/bdaya001iv.do>). 자와 병원 자료를 결합을 통해 각 자료의 가진 단점을 극복할

수 있다. 병원 자료는 시계열적 자료를 확보하기 어려운 반면에, 건강보험 자료에는 없는 진단 및 병원내 실제 처방 및 수술적 치료, 환자의 질병의 심각정도(severity)등을 알 수가 있으며, 영상 판독지가 있어, 동반된 부진단을 파악할 수가 있다. 반면 국민건강 보험자료는 코호트 형태의 자료가 제공되므로, 장기 추적 연구가 가능하며, 사망일자 및 원인도 가능한 장점이 있다.

한편, 우리의 선행 연구 “국민건강보험 빅데이터를 이용한 퇴행성 뇌질환(뇌졸중, 치매, 파킨슨) 진단 이후의 장기요양으로 전달되는 의료체계 개선 연구 (2009)”에 의하면 2009년 초발 뇌졸중 환자 10,549명 환자 중에 33.69%가 2년 이내 사망하고, 중증 뇌졸중 환자는 초기의 충분한 재활치료를 받았을 경우에는 그렇지 않은 군에 비해 생존율을 높음을 알 수 있었으나, 경증 환자일 경우에는 재활치료의 빈도와 생존율과 큰 차이가 없음을 확인하였다. 따라서 중증 환자의 경우 합병증 방지를 통한 사망률을 낮추는 것이 치료 전략이 되어야 할 것이고, 경증 환자의 경우 초기의 기능회복을 유도한 후 적극적인 가정이나 사회 복귀가 재활 치료 전략이 되어야 할 것이다. 뇌졸중과 외손상 환자의 재활은 환자의 예후에 근거한 재활치료 전략을 수립해야 함에도 불구하고, 환자 가족의 의지나, 환자의 경제적 상태 혹은 자동차 보험 혹은 산재보험의 특성에 맞추어 재활을 받고 있는 실정이다. 그리하여, 상당수의 환자들이 보험급여의 기준에 맞추어, 여러 병원에 입퇴원을 반복하면서, 치료를 받는 “재활 난민”이 발생하는 것이 현실이다.

이제까지 병원 자료와 공단 자료의 결합이 법적인 제한되어, 연구가 활발하게 진행되지 않은 않았다. 따라서 병원 자료와 공단 자료를 결합을 통해서, 뇌손상 이후 사망까지 시계열 자료를 구축할 수 있었다. 이번 연구를 통해 일산병원 자료와 공단 자료의 결합을 통해 인공지능 분석이 가능한지 확인하는 것이 일차적인 목표가 되겠으며, 이차적으로는 병원자료를 통해서 뇌손상 이후에 사망에 영향을 미치는 인자를 분석하고자 하였으며, 뇌손상 환자의 사망과 관련된 인자 분석을 통해서 환자들과 가족들에게 합리적인 재활치료 전략 및 재활의료 전달 체계를 마련하는 근거를 제공하는 것이 이 연구의 최종 목표라 하겠다.

## 제1절 통계 분석과 기계 학습(Machine learning)과의 차이

의료분야는 각종 임상 및 검진의료기기(device) 데이터의 통합, 병원 간 검사 기록과 연구데이터 공유를 통해 질병의 징후를 조기에 발견하고 질병을 예측하는 데 빅데이터

도입과 활용이 확대되는 추세에 있다. 영국의 UK biobank는 유전자 자료뿐만 아니라, 의학 자료 및 검사 결과 자료 및 사망 자료까지 결합한 자료를 연구자에게 공개하고 있으며, 미국의 Beth Israel Deaconess Medical Center에서는 2001년부터 2012년까지, 중환자실에 입원한 약 4만 명 이상의 비식별 보건의료 데이터를 (Medical Information Mart for Intensive Care, MIMIC)의 이름으로 공개하고 있으며, 우리나라도 공공부문 데이터를 공개하고 있으며, 국민건강보험공단과 심사평가원에서 보건 의료 자료를 연구자에게 공개하고 있다.

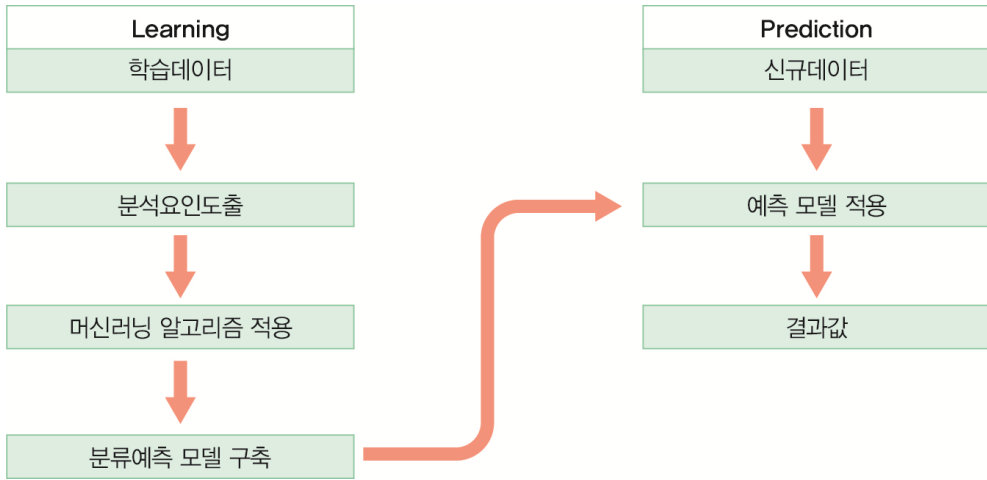
기존의 연구가 자료 분석을 통해서 결과값에 대한 원인 변수에 대한 관계 규명, 표본 추출을 통해 실제 다수 집단의 통계 추정이라면 인공 지능은 빅데이터 기반으로 훈련을 한 다음, 검증 데이터를 통해 예측하는 것이 통계 분석과의 차이점이라고 할 수 있다. 데이터 마이닝 기법(Data Mining Technics)을 통해서 빅데이터를 통계기법 및 인공지능 기법을 이용하여 분석하고 이에 대한 평가를 거쳐 일반화시킴으로써 새로운 자료에 대한 예측 및 추측을 할 수 있는 의사결정을 지원하게 된다. 데이터 마이닝은 대규모로 저장된 데이터 안에서 주로 인공지능기법을 활용하여 전통적인 통계학 이론으로는 설명이 힘든 패턴(pattern)과 규칙(rule)을 발견하게 되어, 기계 학습을 통해 분류(classification), 추정(estimation), 예측(prediction), 유사집단화(affinity grouping), 군집화(clustering)을 수행하게 된다.

의료자료는 방대하지만 정리가 되지 않아 이에 대한 정리와 추출 작업이 필요하다. 자연어(natural language) 형태로 구성된 비정형 또는 반정형데이터에서 패턴 또는 관계를 추출하여, 정형데이터로 구조화하는 작업이 필요하다. 불필요한 정보를 제거하고, 의미없는 단어나 연구 주제와 일치하지 않는 단어를 제거해주는 불용어(stopword: 관사, 전치사, 조사, 접속사)나 불필요한 단어나 기호를 제거하는 토큰화(tokenization)이 필요하다. 이를 텍스트마이닝(Text mining)이라고 하며, 자연어 처리(natural language processing, NLP)가 핵심기술이다.

빅데이터 분석은 (1) 데이터 수집, (2) 데이터 전처리와 정제, (3) 데이터 분석과 정리 및 처리결과의 수용, (4) 해석과 결과제시의 순서로 진행된다. 먼저, 데이터 수집은 선정된 변수에 의해 구성된 분석모형이나 과제를 해결하기 위해 관련 데이터를 수집하고 이를 분석하는 단계를 말하며, 데이터 전처리(data preprocessing)와 정제(refinement)는 데이터 공급원으로부터 획득한 데이터 중 분석하기에 부적합하거나 수정이 필요한 경우 데이터를 전처리, 정제하는 과정을 말하며, 빅데이터 과제분석단계에서 데이터

전처리와 정제 과정은 프로젝트의 90% 이상을 차지한다.

데이터 분석과 정리 및 처리결과 수용은 분석 대상과 관련된 변수의 데이터가 수집되면 이를 분석하여 변수들 간의 관련성을 파악하는 과정을 말한다. 마지막으로, 훈련세트와 검증세트를 나누어 인공지능 분석을 통해, 예측력과 분석결과 의미 제시하게 된다.



[그림 1-1] 기계학습 분석 모식도

기계학습은 크게 지도 학습(supervised learning)과 비지도 학습(unsupervised learning)으로 구분할 수 있다. 지도 학습 기법을 활용하면(Classification), 회귀(Regression), 예측(prediction)의 목적으로 주로 시행하며 주요 알고리즘으로는, 일반화 선형 모델(Generalized linear models, GLM), 결정나무(Decision Tree), random Forrest, K-최근접 이웃법(K-nearest neighbor method), 점진적 부스팅 머신(Gradient boosting machine, GBM), 지지 벡터 머신(Support Vector Machine(SVM)), 인공신경망(neural network, NN), Ensemble model 등이 있다 [그림 1-2].

일반화 선형 모델은 선형 회귀의 발전된 형태로, 다양한 기능성의 분석과 연결함수를 위해서 분석자가 더 효과적으로 데이터를 모델링할 수 있으며, 그리드 탐색(grid search)으로 강화된 GLM은 전통적 통계와 가장 발전된 ML의 조합이다.

의사결정 트리(Decision trees)는 모 집단을 대상(target) 변수에 대해 동질적인 더 작은 조각으로 점진적으로 분할하는 규칙 집합을 학습하는 비지도 학습 방법이다. 랜덤 포레스트(Random forests) 널리 사용되는 총체적 학습 방법으로, 여러 개의 의사결

정 트리를 학습한 다음, 트리 전반에 걸친 평균을 구해 예측을 산출한다. 평균 프로세스는 일반화가 가능한 솔루션을 제공하며 데이터의 불규칙 잡음(random noise)을 걸러내는 효과가 있다.

점진적 부스팅 머신(Gradient boosting machine, GBM) 의사결정 트리의 시퀀스 교육을 통해 예측 모델을 생성하는 방법으로, 연속되는 트리가 이전 트리의 예측 오류를 수정하여 진행한다.

Boosting은 여러 개의 약한 Decision Tree를 조합해서 사용하는 Ensemble 기법 중 하나로 약한 예측 모형들의 학습 에러에 가중치를 두고, 순차적으로 다음 학습 모델에 반영하여 강한 예측모형을 만드는 것이다. XGBoost(Extreme Gradient Boosting)는 Boosting 기법을 이용하여 구현한 알고리즘은 Gradient Boost가 대표적인데, 이 알고리즘을 병렬 학습이 지원되도록 구현한 라이브러리가 XGBoost이다. Regression, Classification 문제를 모두 지원하며, 성능과 자원 효율이 좋아서, 인기있게 사용되는 알고리즘이다. SVM은 결정 경계(Decision Boundary), 즉 분류를 위한 기준 선을 정의하는 모델이다. 그래서 분류되지 않은 새로운 점이 나타나면 경계의 어느 쪽에 속하는지 확인해서 분류 과제를 수행할 수 있게 된다.

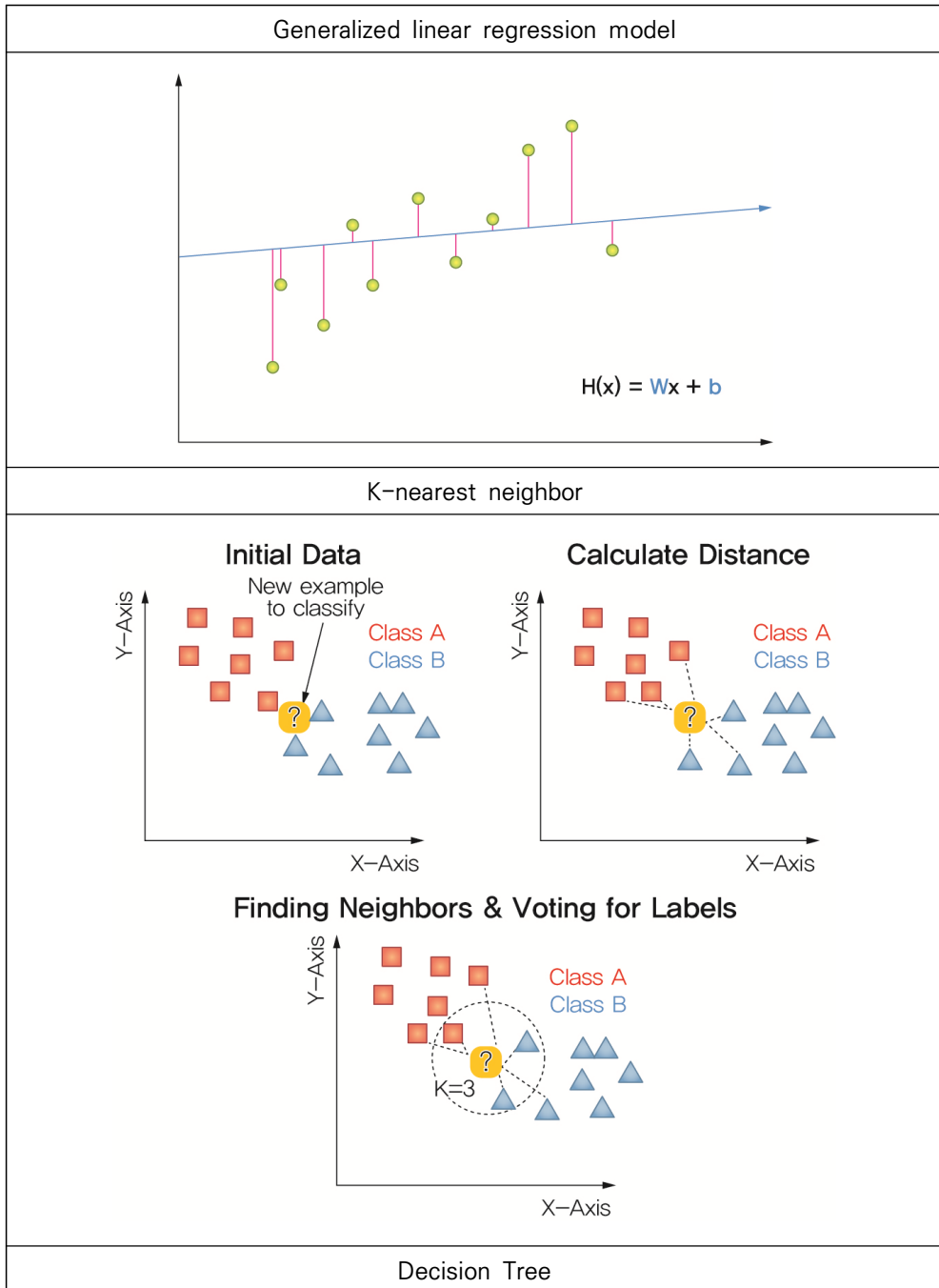
K-최근접 이웃(K-NN, K-Nearest Neighbor) 알고리즘은 가장 간단한 머신러닝 알고리즘으로, 분류(Classification) 알고리즘이다. 비슷한 특성을 가진 데이터는 비슷한 범주에 속하는 경향이 있다는 가정하에 사용한다. 특정 데이터가 주어졌을 초록색 삼각형 그룹과 빨간색 별 모양 그룹 중에 어디에 가까운지를 파악하는 것을 말한다. 이처럼, 주변의 가장 가까운 K개의 데이터를 보고 데이터가 속할 그룹을 판단하는 알고리즘이 K-NN 알고리즘이다.

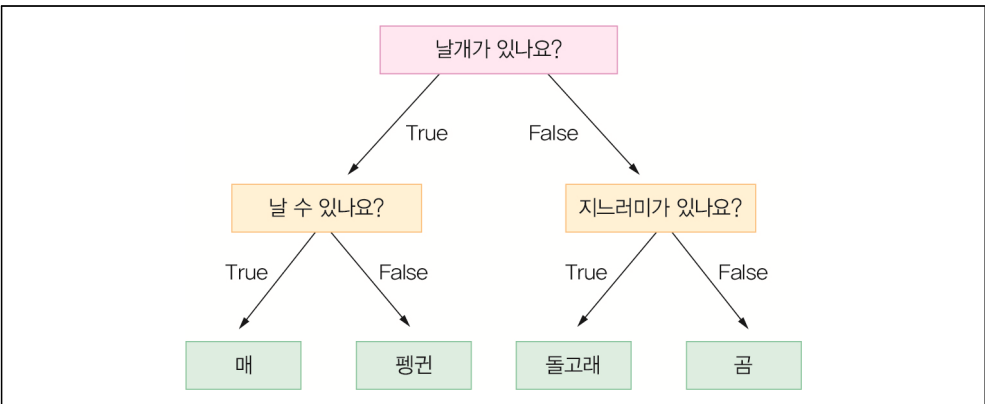
Deep neural network (DNN)은 입력층(input layer)과 출력층(output layer) 사이에 여러 개의 은닉층(hidden layer)들로 이뤄진 인공신경망(Artificial Neural Network, ANN)이다. input에 대해서 output까지 Layer를 거치게 되어, input Layer 다음에 W랑 b가 있어서  $Wx+b$ 를 거치게 되고 hidden layer에서는 활성화 함수가 있다고 보면 된다. 그 후 다시  $Wx+b$ 를 거치고 활성화 함수를 거친 후, output layer가 만약 분류였다면 마지막에는 softmax를 activation function(활성화 함수)을 활용한다.

Ensemble model은 여러 개의 단일 모델들의 평균치를 내거나, 투표를 해서 다수결에 의한 결정을 하는 등 여러 모델들의 집단 지성을 활용하여 더 나은 결과를 도출해 내는 것에 주 목적이 있다. 앙상블 기법에는 투표(Voting) - 투표를 통해 결과 도출, 샘플을

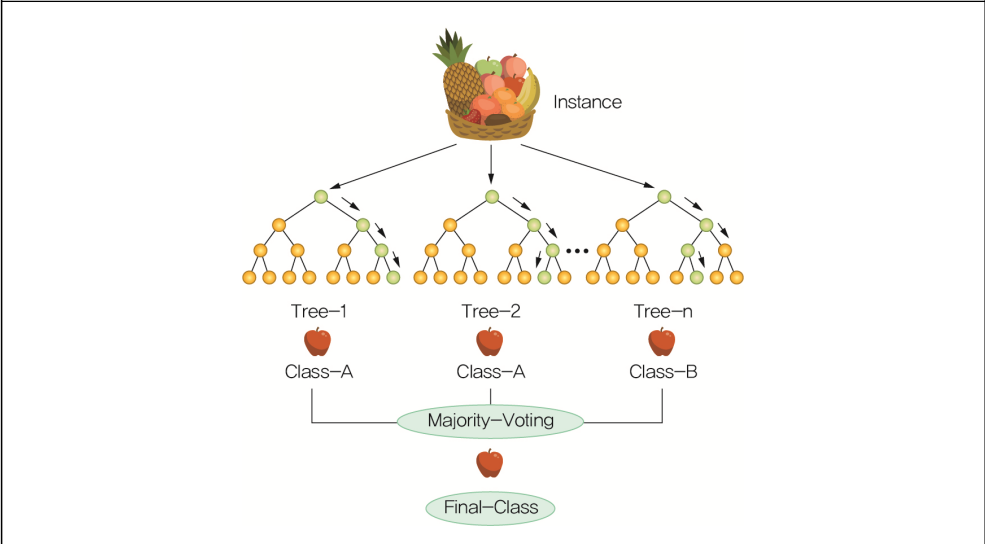


다양하게 중복 생성(Bagging - Bootstrap Aggregating), 이전 오차를 보완하며 가중치 부여 (Boosting), 여러 모델을 기반으로 meta 모델(Stacking) 다양한 방법이 있다.

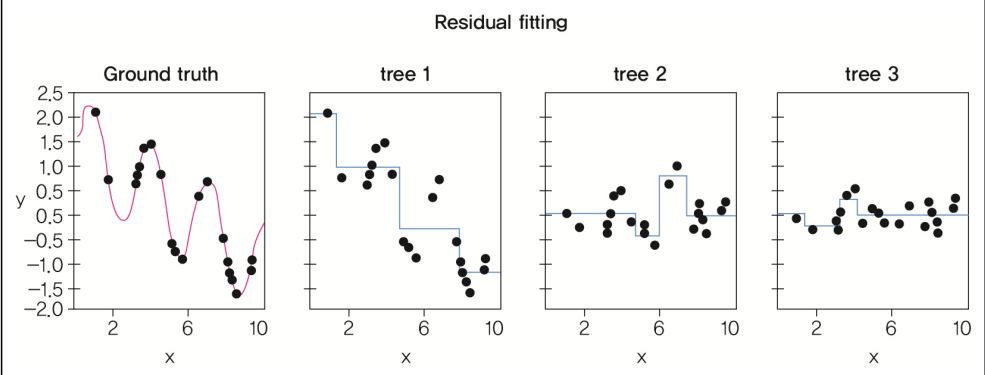




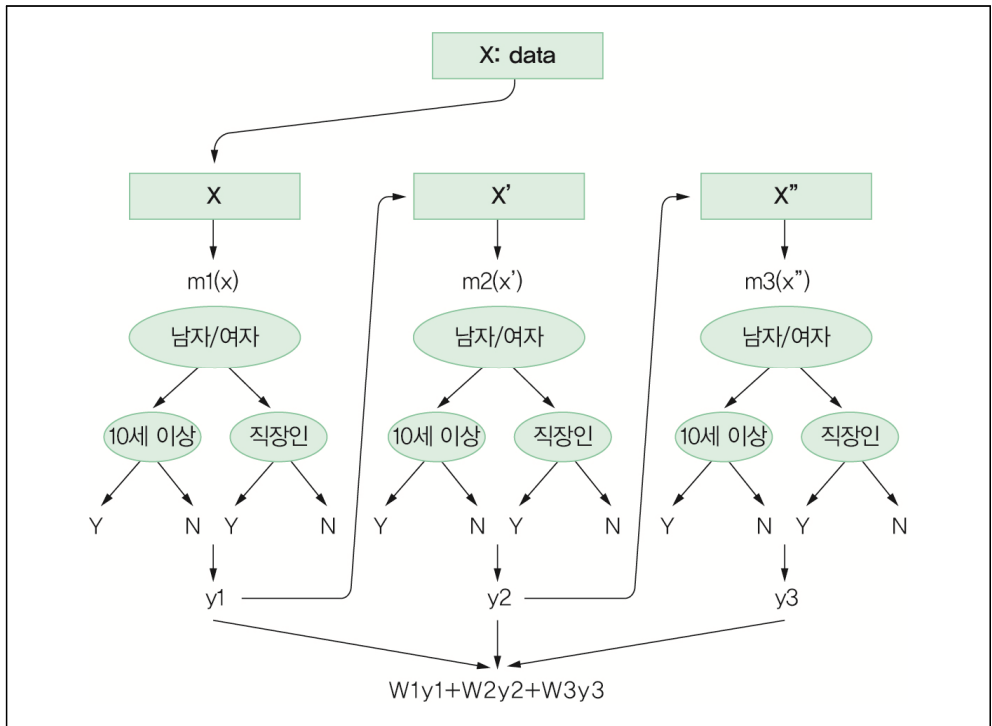
Random Forrest



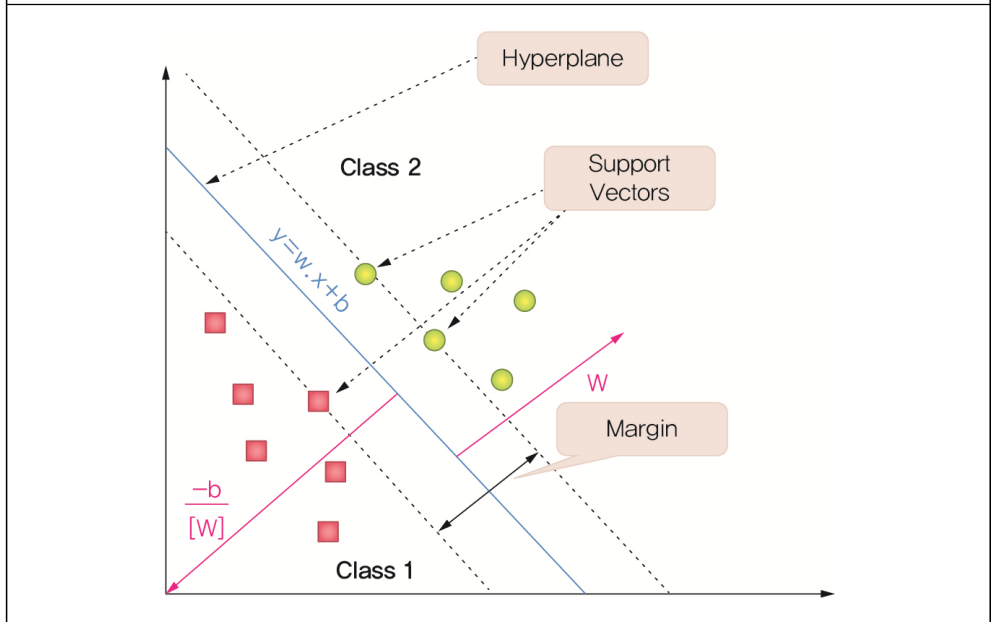
Gradient Boosting Machine



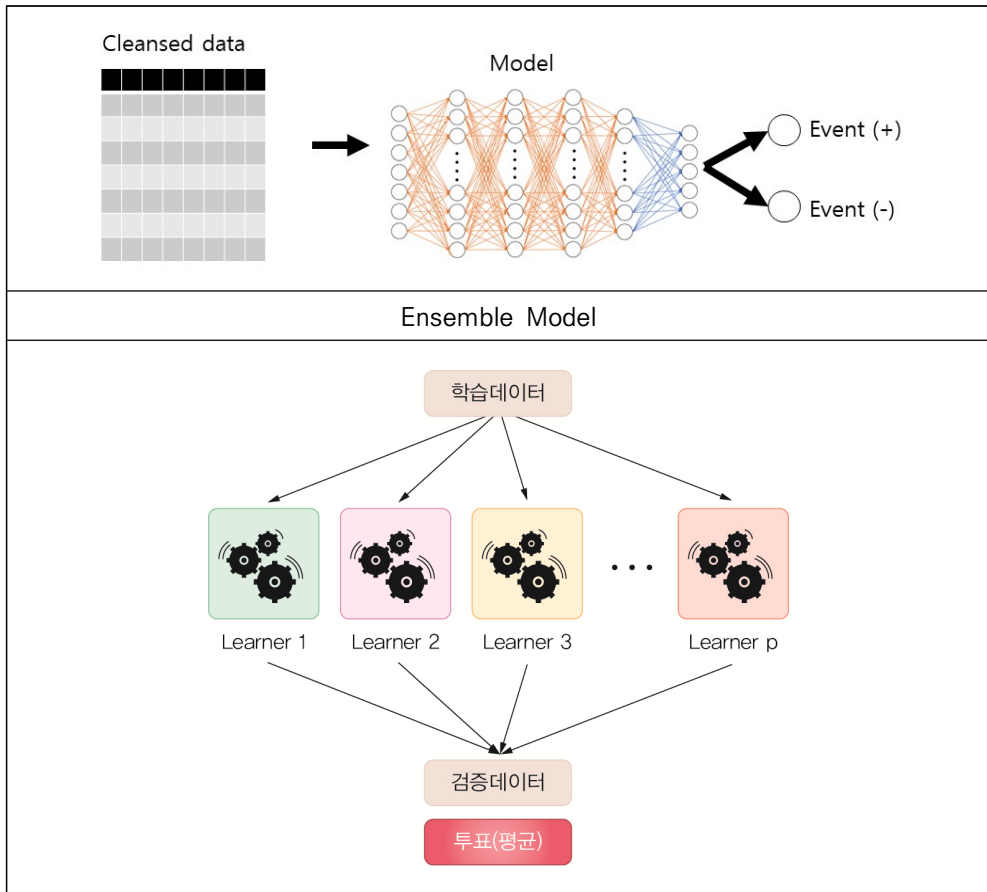
Xgboost



Support Vector Machine(SVM)



Deep neural network

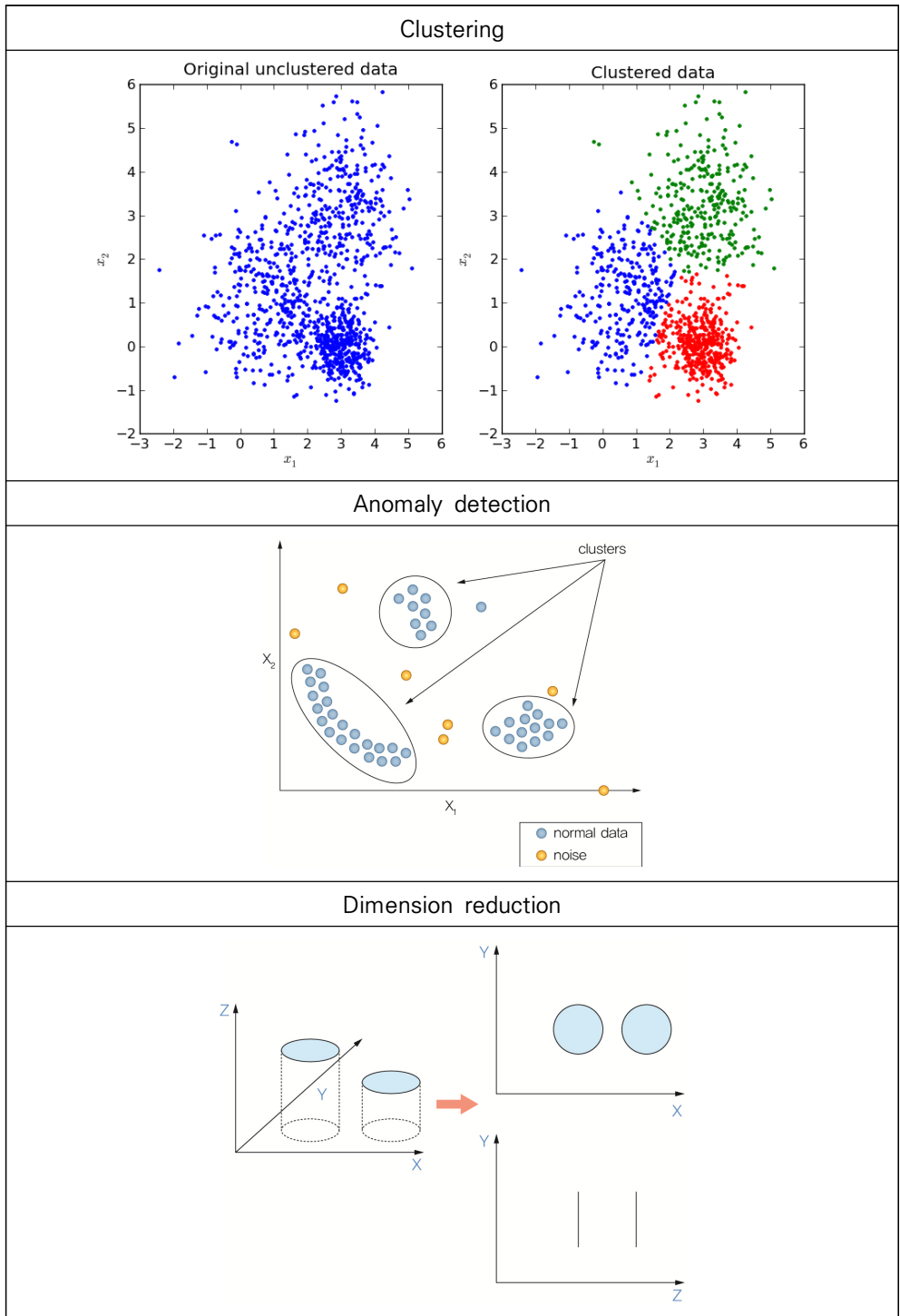


[그림 1-2] 주요 지도 기계학습 알고리즘 모식도

비지도 학습의 주요 기술에는 클러스터링(Clustering), 비정상 탐지(Anomaly detection), 차원 축소(Dimension reduction)이 있다 [그림 1-3]. 클러스터링(Clustering)은 개체를 다수의 매트릭스에서 상호 유사한 세그먼트 또는 클러스터로 그룹화하는 기법으로, 가장 널리 사용되는 것이 k-평균(k-means) 이다.

비정상 탐지(Anomaly detection)는 예상치 못한 이벤트 또는 결과를 식별하는 프로세스이다.

차원 축소(Dimension reduction)는 고려 대상 변수의 수를 줄이는 프로세스로 더 많은 데이터를 수집할수록 예측에 사용 가능한 예측 변수(또는 특성)의 수도 급격히 증가하게 된다. 주성분 요소 분석 (Principal components analysis, PCA)은 일련의 원시 특성을 평가해 이를 상호 독립적인 인덱스로 축소하는 기법이다.



[그림 1-3] 주요 비지도 기계학습 알고리즘 모식도

기계학습 분석은 이제까지의 통계학적 모델과는 다른 측면이 있다. 가령, 통계학적 모델은 여러 선행인자로 인한 결과에 대한 분석이며 소수 sample을 통해 다수를 추정하는 목적이다. 즉 결과값, 가령 질병의 발생 혹은 사망 여부에 대해서 원인 인자를 분석을 통하게 되면, 수집된 자료를 전체 인구 집단에 적용하지만, 기계학습은 과거의 자료 혹은 훈련 자료를 모델링을 통해서 미래를 예측하는 것이 목적이 된다. 또한 통계 분석에서는 대상이나 표본자료 간의 관계가 성립되거나 유사한 대상을 기준으로 하며, 정교한 분석 기법을 실행해 그 가치의 정확성을 높이는 데 치중하지만, 빅데이터 분석은 전혀 인과 관계로 여겨지고 있는 인자 간에도 관계를 도출하고 연관성을 찾을 수 있다.

## 제2절 선행 연구 분석

앞서 언급한 대로 뇌졸중이나 외상성 뇌손상은 신경학적 손상에 따른 기능적 제한을 동반하는 경우가 많다. 따라서 뇌졸중이나 외상성 뇌손상 이후의 기능적 예후를 예측하는 것은 항상 관심도가 높은 연구 주제였다. 뇌병변 환자의 예후를 조기에 예측하는 것은 환자나 보호자가 매우 궁금해하는 정보이기도 하고, 의료인에게도 임상적 의사결정에 중요한 정보로서 활용될 수 있기 때문에 그 중요성이 더 높다고 할 수 있다.

특히 최근에는 인공지능 기계학습에 기반한 뇌졸중 후 예후를 예측하고자 하는 시도가 활발히 이루어지고 있다. 기계학습은 한 번에 많은 데이터를 처리할 수 있으며, 입원 초기 정보를 통해 빠른 예후 예측을 가능하게 해준다는 장점이 있다. Heo 등은 급성기 뇌졸중 이후 이후 3개월 시점의 기능적 예후를 예측하는 연구를 제시하였는데, 그들은 modified Rankin scale 0-2를 좋은 예후로 정의하고 Acute Stroke Registry and Analysis of Lausanne (ASTRAL) score의 예측력과 기계학습의 예측력을 비교하였다. 그들의 연구에서 DNN은 ASTRAL score 보다 유의하게 높은 예측력을 보여주었으나 random forest(RF)과 logistic regression에서는 모델들의 예측력이 ASTRAL score와 유의한 차이를 보이지 않았다. 한편 Jang 등은 한국의 뇌졸중 등록 사업 데이터를 활용하여 급성기 뇌졸중 환자의 3개월 예후 예측 모델을 발표하였다. 이 연구에서는 DNN, SVM, random forest, and extreme gradient boosting와 같은 기계학습 알고리즘이 사용되었으며, 모든 기계학습 모델들이 전통적인 logistic regression 방법에 비해 더 나은 예측력을 보여주었다. Park 등은 한국의 단일 뇌혈관 전문 병원 환경에서 급성기 뇌졸중 환자의 3개월 예후 예측 모델을 발표하였다. 이 연구에서는 3개월 후 기능적

독립성(modified Rankin scale 0-1)을 좋은 예후로 정의하였으며, 정규화(regularization) 방법을 사용한 logistic regression model이 앙상블(ensemble) 모델인 RF나 extreme gradient boosting 알고리즘 보다 더 나은 예측력을 보여주었다. 또한 그들의 연구는 예측에 필요한 인자 중요도를 밝혔는데, 입원 시 NIHSS 점수와 나이가 단기 예후 예측에 가장 중요한 인자임을 밝혔다. 호주에서 시행한 관련 연구에서는 급성기 뇌경색 환자들을 대상으로 3개월 기능적 예후 예측을 시행하였으며, 이 연구에서는 AUC 범위가 0.66-0.71로 다른 관련 연구들에 비해 다소 낮은 예측력을 보여주었는데 이는 적은 샘플 수 및 변수 종류에 기인한 것으로 사료된다. 한편 Fang 등은 뇌경색 이후 6개월 시점의 기능적 예후를 예측하는 모델에서 DNN 알고리즘과 다른 기계학습 알고리즘 (random forest, Support Vector Machine etc.)간에는 예측력에 유의한 차이가 없음을 밝힌 바 있다. 반면 뇌졸중 발병 후 6개월 후 운동기능을 예측하는 연구에서 상지는 modified Brunnstorm classification, 하지는 functional ambulatory category를 결과 변수로 예측하였고, DNN, logistic regression, RF 3가지 기계 학습 중 DNN의 예측도가 가장 높다는 결과를 발표하기도 하였다. 이처럼 데이터 세트의 종류, 연구가 이루어진 기관 및 지역 등에 따라서 같은 기계학습 알고리즘을 적용하더라도 서로 일관되지 않은 결과를 보이기도 한다.

한편 뇌경색 이후 예후와 관련하여 세부 항목의 예측 모델 또한 연구되어져 왔는데, 뇌졸중 발생 이후 7일과 14일에 폐렴 발생률을 예측하는 모델에서는 logistic regression, SVM은 attention-augmented gated recurrent unit (GRU)등의 알고리즘을 비교하였고, attention-augmented GRU 모델이 예측도가 가장 높았다. 뇌졸중 이후 만성 실어증 환자에서 재활치료에 대한 반응을 예측하는 시범 연구에서는 random forest가 SVM이나 gradient boosting 모델에 예측도가 다소 높았으며, 개별화된 재활치료의 결과를 예측할 수 있는 적용가능한 모델을 제시하였다는 점에서 연구의 의의가 있었다.

외상성 뇌손상 분야에서도 관련 연구는 활발히 이루어지고 있다. Pease 등은 초기 임상 정보 및 CT scan을 기반으로 DL 모델을 수립하였다. 이 연구에서는 발병 후 6개월 시점의 사망 및 Glasgow Outcomes Score score 1-3을 결과값으로 설정하였다. 결과적으로 제시된 결합 모델은 신경외과 의사들의 예측과 비교하여 더 우수한 성능을 보여주었다. 한편 중증도 이상의 뇌손상 환자의 사망률을 예측하는 모델에서는 logistic regression, SVM, decision tree, artificial neural network (ANN) 알고리즘 중 ANN이 사망에 대한 예측도가 가장 높았다. Brushetta 등은 외상성 뇌손상 환자군의 예후 예측에

있어 전통적인 linear regression model이 기계학습 알고리즘과 비교하여 예측력에 유의한 차이가 없다는 결과를 제시한 바 있어 뇌졸중과 마찬가지로 외상성 뇌손상 분야에서 개별 연구들에서 일관된 결과를 보여주지는 못하였다. 한편 최근의 활발한 연구에도 불구하고 외상성 뇌손상 분야는 뇌졸중에 비해 인공지능 기계학습을 적용한 연구가 아직 부족한 실정이라고 할 수 있다.





# 제2장

## 연구방법

제1절 병원 자료	31
제2절 국민건강보험 맞춤형자료	32
제3절 대상자 추출 과정	33
제4절 분석 방법	34



## 제2장 연구방법

2009년부터 2018년 10년간 20세 이상 성인을 대상으로 본원에서 뇌졸중 및 외상성 뇌손상, 편마비, 및 사지 마비로 진단된 환자를 포함 (KCD I6\*, G81\*, G82\*, S06\*) 하고, 뇌영상의학 검사를 본원에서 시행하고 뇌영상 판독 결과가 있는 자로 제한하였다. 우리 병원 입원 이전에 장애 등록된 환자는 제외하였으며, 사지 마비 중에도 척수 손상을 시사하는 사지마비(Tetraplegia)와 양하지마비(Paraplegia) 환자는 제외하였다. 수집한 자료와 정의는 다음과 같다.

### 제1절 병원 자료

- 인구사회학적 변수: 성별, 나이, 출생 및 사망 일자
- 건강행태 및 검사: 흡연력, 음주력, 질병과거력, 신체계측 (키, 체중), 흡연력과 음주력은 총량은 계산하지 않고, 현재도 흡연 및 음주 중인 경우, 과거에 흡연 및 음주를 하였으나, 지금은 하지 않는 경우, 과거나 현재도 하지 않은 경우로 나누어 조사하였다.
- 질병 코드 및 진단명: 뇌경색, 뇌출혈, 외상성 뇌손상, 경막하 출혈, 경막외 출혈, 지주막하 출혈, 뇌실질 출혈, 뇌실내 출혈, 편마비, 사지마비 이 중 소아마비와 척수 손상으로 인한 사지마비(tetraplegia) 제외하였다.

진단명은 주·부상병에 상기 병명이 있거나 영상의학 판독지에 위의 항목이 있을 경우로 하였다.

- 혈액검사: Glucose AC, Albumin, Hb, HbA1c, BUN/, Cr를 조사하였다. 여러 번 검사할 경우에는 최초 시행 검사 결과만 적용하였다.
- 진료 경과: 중환자실 입원 기간, 일반 병동 입원 기간, 전체 입원 기간, 기관 절개술

시행 여부, 경관 급식, 폐렴을 비롯한 합병증 여부(폐렴, 욕창, 심부정맥혈전증), 수술 및 시술 여부(craniectomy, craniotomy, coil embolization, clipping, gastrostomy), 입원 기간내 전체 협진 횟수, 감염 및 호흡기 내과 협진 횟수를 조사하였다.

- 영상의학검사: chest PA or AP, brain CT, brain MRI 판독 결과지에 대해서 R package이용 자연어 처리를 통하여 폐렴과 뇌손상의 원인과 뇌병변의 해부학적 위치, 범위를 지정하였다. 예를 들어, traumatic intracranial hemorrhage가 있을 경우, 텍스트 마이닝에서는 외상(trauma)과 뇌내출혈(intracranial hemorrhage)로, traumatic subdural hemorrhage와 non-traumatic subdural hemorrhage로 판독지나 진단명이 있다고 하면, 각각 trauma와 non-trauma, subdural hemorrhage을 나누어 각각 인자로 구분하도록 하였다. 따라서, 급성 뇌졸중으로 진단받고 입원하였다고, 뇌영상에서 현재 임상을 일으킬 수 있는 직접적인 원인이 되는 뇌병변 뿐만 아니라, 퇴행성 소견이 있을 수도 있을 경우, 모두 병변 영역으로 포함하였다.

폐렴(Pneumonia)은 Chest X-ray에서 판독지에 pneumonia, atelectasis, consolidation 있을 경우 이거나, 부상명에 pneumonia가 있을 경우, 입원중 폐렴으로 진단하였다.

뇌손상 원인에 따른 분류는 infarction, hemorrhage(subdural, epidural, subarachnoidal, intraventricular), traumatic brain injury, non-traumatic brain injury로 지정하였다.

뇌손상의 해부학적 부위는 Frontal lobe, Temporal lobe, parietal lobe, occipital lobe, Cerebellum, basal ganglia, thalamus, subcortex, brain stem, ACA, MCA, PCA, basilar artery, Aneurysm을 지정하였다.

뇌병변 특성을 focal\_small lesion, multifocal\_large\_lesion, Acute lesion, Chronic and old lesion로 구분하였다.

## 제2절 국민건강보험 맞춤형자료

이번 연구의 환자의 병원 자료와 국민건강보험의 자료를 결합하여, 사망 및 생존기간, 장애유형 및 등급, 등록일자, 거주 지역, 가입자 유형, 보험료 분위를 조사하였다. 1차 장애 유형 및 등급은 최초 장애등록을 말하며, 2차 장애 유형 및 등급은 1차 장애등록 이후 2년 지난 시점의 장애등록을 지칭한다. 장기요양보험 변수를 신청 자료(성별, 나이,

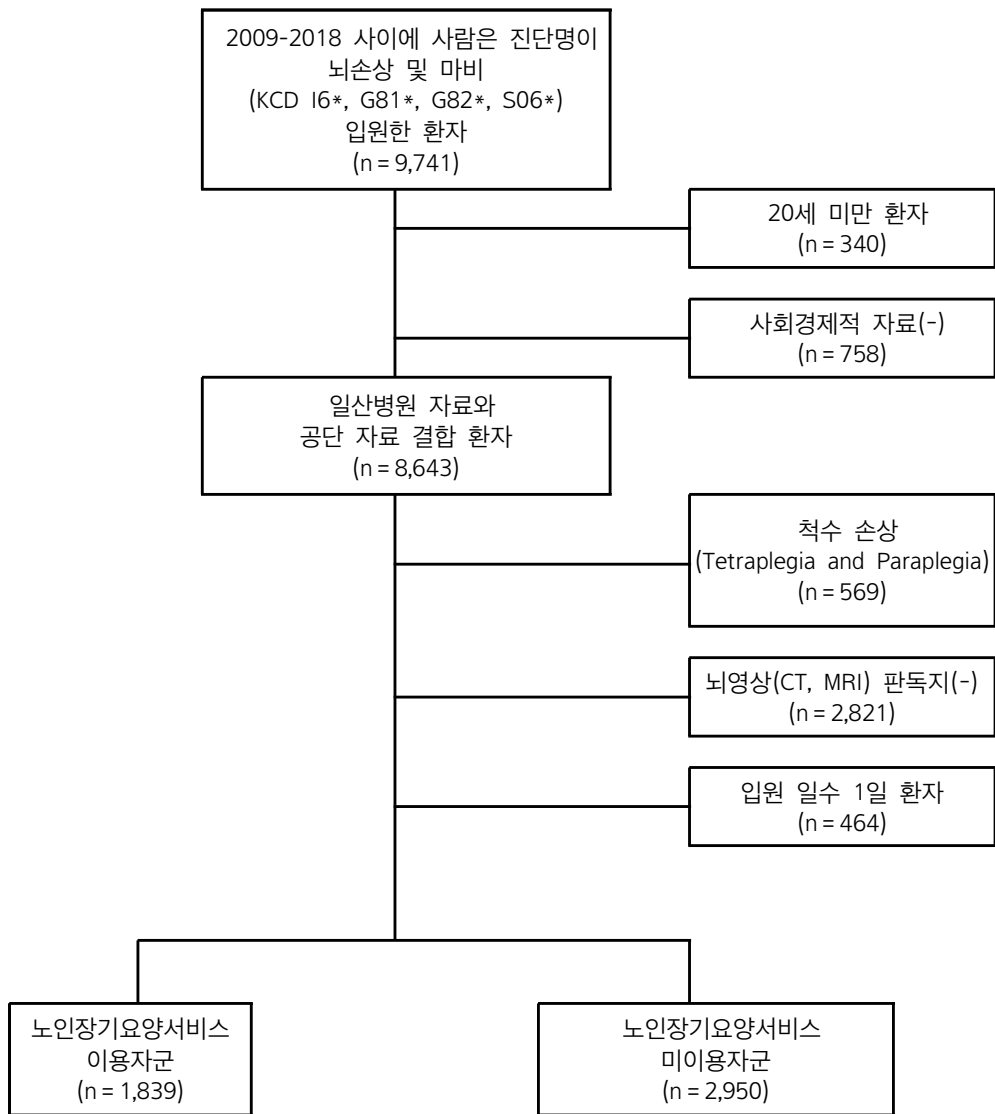
신청 일자) 및 인정 조사 자료, 등급판정 자료를 조사하였다. 결측치 처리는 항목의 평균치를 삽입하여 계산하였다.



[그림 2-1] 자료 결합 형태

### 제3절 대상자 추출 과정

대상자는 2009-2018 사이에 사립은 진단명이 뇌손상 및 마비로 (KCD I6\*, G81\*, G82\*, S06\*) 입원한 환자(n=9,741) 중 20세 미만 (n=340), 사회 경제적 자료가 없는 사람(n=758) 을 제거하여, 총 8,643 명이 자료 결합에 성공하였다. 이중 척수 손상을 의미하는 (Tetraplegia, Paraplegia) 환자 n=569명을 제거하였다. 또한 뇌영상 (CT, MRI) 판독지가 없는 경우 (n=2821), 입원 일수가 1일로 응급실에만 있다 퇴원한 환자 464명을 제외하여, 총 4,789명이 이 연구에 포함되었다. 자료 결합 코드는 Appendix 98page부터 152page에서 확인할 수 있다.



[그림 2-2] 자료 결합 모식도

## 제4절 분석 방법

R을 이용하여 인공지능 및 통계분석을 시행하였다. R version 3.5.2 (2018-12-20)이었으며 설치된 통계와 인공지능 알고리즘 package는 Appendix에서 확인할 수 있다. 분석한 기계학습 알고리즘은 NN, RF, GBM, Xgboost, KNN, SVM을 이용하였다. 또한 생존 사망에 대해서 linear regression 분석을 하여, 인공지능 분석과 비교하였다.

# 제3장

## 결과

제1절 노인 장기요양보험 서비스 이용자군  
의 생존 분석의 기계학습(machine learning) 59

---

제2절 노인 장기요양보험 서비스 미이용자  
군의 기계학습(Machine Learning) 70  
분석 결과

---





## 제3장 결과

노인 장기요양보험 서비스 이용자군 (1,839명)은 총 노인 장기요양 변수를 포함한 237개 변수였으며, 반면에 미이용자군 (2,950명)은 147개 변수이며 각 군의 인공지능 분석에서는 훈련세트를 70%, 검증세트로 30%를 나누었으며, 총 변수 중에 주요 20개만 추출하였으며, ensemble model 분석을 통해서 최적의 분석 모듈을 확인하였다.

병원 자료와 공단 자료 노인 장기요양보험 자료 결합 코드 및 노인 장기요양보험 여부에 따른 생존 여부 분석에 대한 기계학습의 coding 자료는 appendix에 찾을 수 있다.

국민건강보험 일산병원의 의무기록 자료와 국민건강보험 결합 자료를 이용하여, 노인 장기요양 서비스 여부에 따른 인구학적 특성비교는 표를 통해 알 수 있다. 노인 장기요양 보험 서비스는 65세 이상 노인 또는 치매, 중풍, 파킨슨병 등 노인성 질병을 앓고 있는 65세 미만인 자 중 6개월 이상의 기간 동안 일상생활을 수행하기 어려워 장기요양보험 서비스가 필요하다고 인정되는 자에 대해서 급여가 가능하다. 노인 장기요양보험 서비스를 신청한 노인의 연령 평균이  $76.1 \pm 9.4$ 세로 서비스를 받지 않은 군( $62.4 \pm 13.6$ )보다 많았다. 이는 노인 장기요양보험 서비스가 65세 이상 노인에게 주로 서비스가 제공되고 65세 미만 장애인에게는 지역자치단체에서 장애인활동보조서비스가 제공되어 이런 차이가 발생한 것으로 보인다.

노인 장기요양보험 서비스를 이용자군이 전체 사망자가 962명(52.3%)으로, 서비스를 받지않은 군 695명(23.6%) 보다 많았으며, 발병 후 생존기간도  $1,668.1 \pm 1,135.4$ (일)로 서비스를 받지 않은 군의  $1,831.7 \pm 1,114.6$  보다 짧았다. 성별로는 노인 장기요양보험 서비스를 이용자군에서 여성이 1,031(56.1%)명으로 이는 평균 연령이 여자가 높으면서 기인하는 것으로 보인다. 거주지역은 일산병원이 경기도 고양시에 소재하여, 고양시뿐만 아니라 서울 및 경기, 인천지역에서 거주하는 환자에서 많이 입원하여 이런 결과가 나타난 것으로 추측된다. 또한 도시 규모에서 광역시 읍, 면은 인천광역시의 읍, 면 거주 환자들이 일산병원을 많이 진료를 받은 것에 기인한 것으로 여겨진다.

장기요양보험 서비스를 이용자군은 입원중 총입원기간, 일반병동 입원기간, 중환자실 입원기간, 총 협진횟수, 호흡기 및 감염내과 협진, 위루술, L-tube 삽입, 입원중 폐렴 여부, 기관절개 관삽입(T-tube), 욕창발생, Albumin 및 혈색소(Hb) 저하, 공복혈당 및 당화혈색소, BUN, Cr의 상승이 서비스를 받지 않는 군보다 통계적으로 차이가 있게 관찰되어 이는 장기요양보험 서비스를 이용자군이 받지 않은 군보다 연령이 높은 것에 기인한 것으로 여겨진다 <표 3-1>.

<표 3-1> 노인 장기요양보험 서비스 여부에 따른 인구학적 특성 및 병원 자료의 비교

	노인 장기요양보험 서비스 이용군	노인 장기요양보험 서비스 미이용군	P-value
<b>생존여부</b>			0.000
(-)	877 (47.7%)	2255 (76.4%)	
(+)	962 (52.3%)	695 (23.6%)	
<b>발병후 생존기간 (일)</b>	1668.1 ± 1135.4	1831.7 ± 1114.6	0.000
<b>연령 (연도)</b>	76.1 ± 9.4	62.4 ± 13.6	0.000
<b>성별</b>			0.000
남	808 (43.9%)	1813 (61.5%)	
여	1031 (56.1%)	1137 (38.5%)	
<b>거주지역</b>			0.000
서울	241 (13.1%)	383 (13.0%)	
부산	5 (0.3%)	9 (0.3%)	
대구	3 (0.2%)	7 (0.2%)	
인천	38 (2.1%)	67 (2.3%)	
광주	4 (0.2%)	7 (0.2%)	
대전	3 (0.2%)	6 (0.2%)	
울산	1 (0.1%)	1 (0.0%)	
세종	0 (0.0%)	1 (0.0%)	
경기도	1406 (76.5%)	2349 (79.6%)	
강원도	17 (0.9%)	22 (0.7%)	
충청북도	15 (0.8%)	13 (0.4%)	
충청남도	41 (2.2%)	18 (0.6%)	
전라북도	16 (0.9%)	14 (0.5%)	
전라남도	26 (1.4%)	17 (0.6%)	
경상북도	13 (0.7%)	25 (0.8%)	
경상남도	8 (0.4%)	8 (0.3%)	
제주특별자치도	2 (0.1%)	3 (0.1%)	

	노인 장기요양보험 서비스 이용군	노인 장기요양보험 서비스 미이용군	P-value
<b>도시규모</b>			0.000
광역시	276 (15.0%)	613 (20.8%)	
광역시-읍	242 (13.2%)	379 (12.8%)	
광역시-면	61 (3.3%)	53 (1.8%)	
시지역-읍	58 (3.2%)	42 (1.4%)	
시지역-면	115 (6.3%)	658 (22.3%)	
시지역-면	930 (50.6%)	1082 (36.7%)	
군지역-읍	133 (7.2%)	110 (3.7%)	
군지역-면	24 (1.3%)	13 (0.4%)	
<b>1차 장애 등급</b>			0.000
(-)	954 (51.9%)	2321 (78.7%)	
1	143 (7.8%)	64 (2.2%)	
2	166 (9.0%)	131 (4.4%)	
3	126 (6.9%)	56 (1.9%)	
4	151 (8.2%)	86 (2.9%)	
5	182 (9.9%)	135 (4.6%)	
6	117 (6.4%)	157 (5.3%)	
<b>1차 장애 종류</b>			0.000
비장애인	954 (51.9%)	2321 (78.7%)	
지체장애인	236 (12.8%)	204 (6.9%)	
뇌병변장애인	415 (22.6%)	168 (5.7%)	
시각장애인	51 (2.8%)	57 (1.9%)	
청각장애인	85 (4.6%)	83 (2.8%)	
지적장애인	15 (0.8%)	23 (0.8%)	
정신장애인	3 (0.2%)	12 (0.4%)	
신장장애인	7 (0.4%)	3 (0.1%)	
언어장애인	56 (3.0%)	60 (2.0%)	
자폐성장애인	9 (0.5%)	5 (0.2%)	
심장장애인	2 (0.1%)	4 (0.1%)	
호흡기장애인	1 (0.1%)	4 (0.1%)	
간장애인	5 (0.3%)	4 (0.1%)	
간질장애인	0 (0.0%)	2 (0.1%)	
<b>2차 장애등급</b>			0.000
(-)	954 (51.9%)	2321 (78.7%)	
1	355 (19.3%)	194 (6.6%)	
2	342 (18.6%)	323 (10.9%)	

	노인 장기요양보험 서비스 이용군	노인 장기요양보험 서비스 미이용군	P-value
3	54 (2.9%)	27 (0.9%)	
4	47 (2.6%)	24 (0.8%)	
5	56 (3.0%)	32 (1.1%)	
6	31 (1.7%)	29 (1.0%)	
<b>2차 장애 종류</b>			0.000
비장애인	954 (51.9%)	2321 (78.7%)	
지체장애인	196 (10.7%)	192 (6.5%)	
뇌병변장애인	443 (24.1%)	173 (5.9%)	
시각장애인	47 (2.6%)	55 (1.9%)	
청각장애인	83 (4.5%)	88 (3.0%)	
지적장애인	21 (1.1%)	21 (0.7%)	
정신장애인	7 (0.4%)	15 (0.5%)	
신장장애인	7 (0.4%)	3 (0.1%)	
언어장애인	72 (3.9%)	69 (2.3%)	
자폐성장애인	4 (0.2%)	4 (0.1%)	
심장장애인	2 (0.1%)	2 (0.1%)	
호흡기장애인	0 (0.0%)	2 (0.1%)	
간장애인	3 (0.2%)	4 (0.1%)	
간질장애인	0 (0.0%)	1 (0.0%)	
<b>총입원기간 (일)</b>	25.0 ± 40.6	16.1 ± 24.6	0.000
<b>일반병동 입원기간 (일)</b>	15.3 ± 18.1	10.2 ± 15.3	0.000
<b>중환자실 입원기간 (일)</b>	9.7 ± 35.8	5.9 ± 16.9	0.000
<b>총 협진 횟수</b>	9.8 ± 9.9	7.1 ± 8.3	0.000
<b>호흡기내과 협진 횟수</b>	0.3 ± 1.0	0.2 ± 0.8	0.000
<b>감염내과 협진 횟수</b>	0.4 ± 1.4	0.3 ± 1.2	0.001
<b>위루술시행</b>			0.000
(-)	1808 (98.3%)	2934 (99.5%)	
(+)	31 (1.7%)	16 (0.5%)	
<b>L_tube삽입</b>			0.000
(-)	1171 (63.7%)	2545 (86.3%)	
(+)	668 (36.3%)	405 (13.7%)	
<b>폐렴</b>			0.000
(-)	732 (39.8%)	1774 (60.1%)	
(+)	1107 (60.2%)	1176 (39.9%)	
<b>T_tube 삽입</b>			0.001
(-)	1778 (96.7%)	2898 (98.2%)	

	노인 장기요양보험 서비스 이용군	노인 장기요양보험 서비스 미이용군	P-value
(+)	61 (3.3%)	52 (1.8%)	
<b>외상성 뇌손상</b>			0.052
(-)	1754 (95.4%)	2848 (96.5%)	
(+)	85 (4.6%)	102 (3.5%)	
<b>비외상성 뇌손상</b>			0.043
(-)	15 (0.8%)	10 (0.3%)	
(+)	1824 (99.2%)	2940 (99.7%)	
<b>뇌경색</b>			0.000
(-)	434 (23.6%)	1379 (46.7%)	
(+)	1405 (76.4%)	1571 (53.3%)	
<b>지주막하출혈 (SAH)</b>			0.000
(-)	1705 (92.7%)	2613 (88.6%)	
(+)	134 (7.3%)	337 (11.4%)	
<b>경막하출혈 (SDH)</b>			0.009
(-)	1657 (90.1%)	2723 (92.3%)	
(+)	182 (9.9%)	227 (7.7%)	
<b>뇌실질내출혈 (ICH)</b>			0.001
(-)	1365 (74.2%)	2313 (78.4%)	
(+)	474 (25.8%)	637 (21.6%)	
<b>뇌실내출혈 (IVH)</b>			0.008
(-)	1695 (92.2%)	2778 (94.2%)	
(+)	144 (7.8%)	172 (5.8%)	
<b>Focal_small</b>			0.000
(-)	672 (36.5%)	1831 (62.1%)	
(+)	1167 (63.5%)	1119 (37.9%)	
<b>multifocal_large</b>			0.000
(-)	1081 (58.8%)	2334 (79.1%)	
(+)	758 (41.2%)	616 (20.9%)	
<b>Acute_recent</b>			0.000
(-)	431 (23.4%)	1131 (38.3%)	
(+)	1408 (76.6%)	1819 (61.7%)	
<b>Chronic_old</b>			0.000
(-)	524 (28.5%)	1452 (49.2%)	
(+)	1315 (71.5%)	1498 (50.8%)	
<b>Frontal</b>			
(-)	1839 (100.0%)	2950 (100.0%)	

	노인 장기요양보험 서비스 이용군	노인 장기요양보험 서비스 미이용군	P-value
<b>Temporal</b>			0.001
(-)	1668 (90.7%)	2755 (93.4%)	
(+)	171 (9.3%)	195 (6.6%)	
<b>Parietal</b>			0.536
(-)	1832 (99.6%)	2943 (99.8%)	
(+)	7 (0.4%)	7 (0.2%)	
<b>Occipital</b>			0.000
(-)	1694 (92.1%)	2812 (95.3%)	
(+)	145 (7.9%)	138 (4.7%)	
<b>Cerebellum</b>			0.000
(-)	1465 (79.7%)	2613 (88.6%)	
(+)	374 (20.3%)	337 (11.4%)	
<b>BG</b>			0.000
(-)	948 (51.5%)	2039 (69.1%)	
(+)	891 (48.5%)	911 (30.9%)	
<b>subcortex</b>			0.772
(-)	1833 (99.7%)	2943 (99.8%)	
(+)	6 (0.3%)	7 (0.2%)	
<b>brain_stem</b>			0.000
(-)	1612 (87.7%)	2736 (92.7%)	
(+)	227 (12.3%)	214 (7.3%)	
<b>Thalamus</b>			0.000
(-)	1519 (82.6%)	2676 (90.7%)	
(+)	320 (17.4%)	274 (9.3%)	
<b>ACA</b>			0.000
(-)	1706 (92.8%)	2843 (96.4%)	
(+)	133 (7.2%)	107 (3.6%)	
<b>MCA</b>			0.000
(-)	1312 (71.3%)	2279 (77.3%)	
(+)	527 (28.7%)	671 (22.7%)	
<b>PCA</b>			0.000
(-)	1672 (90.9%)	2825 (95.8%)	
(+)	167 (9.1%)	125 (4.2%)	
<b>basilar_artery</b>			0.000
(-)	1753 (95.3%)	2875 (97.5%)	
(+)	86 (4.7%)	75 (2.5%)	

	노인 장기요양보험 서비스 이용군	노인 장기요양보험 서비스 미이용군	P-value
<b>Aneurysm</b>			0.000
(-)	1628 (88.5%)	2270 (76.9%)	
(+)	211 (11.5%)	680 (23.1%)	
<b>Coil</b>			0.000
(-)	1808 (98.3%)	2790 (94.6%)	
(+)	31 (1.7%)	160 (5.4%)	
<b>craniectomy</b>			0.004
(-)	1793 (97.5%)	2911 (98.7%)	
(+)	46 (2.5%)	39 (1.3%)	
<b>craniotomy</b>			0.000
(-)	1774 (96.5%)	2733 (92.6%)	
(+)	65 (3.5%)	217 (7.4%)	
<b>흡연여부 (smoke_yn)</b>			0.000
현재도 흡연	954 (51.9%)	1249 (42.3%)	
흡연하다 금연	718 (39.0%)	1048 (35.5%)	
금연	167 (9.1%)	653 (22.1%)	
<b>음주여부 (drink_yn)</b>			0.000
현재도 음주	1322 (71.9%)	1604 (54.4%)	
음주하다 금주	297 (16.2%)	409 (13.9%)	
금주	220 (12.0%)	937 (31.8%)	
<b>ALBUMIN/ (g/dL)</b>	3.7 ± 0.5	4.0 ± 0.5	0.000
<b>BUN (mg/dL)</b>	19.3 ± 12.4	16.7 ± 10.3	0.000
<b>CR (mg/dL)</b>	1.2 ± 1.3	1.1 ± 1.4	0.014
<b>GLUCOSE_AC (mg/dL)</b>	133.8 ± 54.3	129.4 ± 49.3	0.004
<b>HB (g/dL)</b>	13.0 ± 2.0	13.8 ± 2.0	0.000
<b>HBA1C (%)</b>	6.6 ± 1.4	6.5 ± 1.3	0.032
<b>욕창</b>			0.000
(-)	1726 (93.9%)	2888 (97.9%)	
(+)	113 (6.1%)	62 (2.1%)	

노인 장기요양보험 서비스를 이용자군에서 사망자군이 생존기간 1234.4 ± 1090.9일로 생존군(2143.8 ± 982.9보다 짧았다. 성별은 양군간의 차이가 없었으며 사망군은 생존군에 비해 입원중 총입원기간, 중환자실 입원기간, 호흡기 내과 협진, 위루술, L-tube 삽입, 입원중 폐렴 여부, 기관절개 관삽입(T-tube), 욕창발생, Albumin 저하, 공복혈당 및 당화혈색소, BUN/Cr의 상승이 생존군 보다 통계적으로 차이가 있게 관찰되어, 사망



군이 연령뿐만 아니라 동반된 내과적 질환이 많고, 초기 뇌손상이 생존군보다 심한 것을 의미한다 <표 3-2>.

<표 3-2> 노인 장기요양보험 서비스 이용자군의 생존 여부에 따른 인구학적 특성 및 병원 자료의 비교

	생존 (n=877)	사망 (n=962)	P-value
<b>발병후 생존기간 (일)</b>	2143.8 ± 982.9	1234.4 ± 1090.9	0.000
<b>연령 (연도)</b>	73.4 ± 9.9	78.6 ± 8.2	0.000
<b>성별</b>			0.193
남	371 (42.3%)	437 (45.4%)	
여	506 (57.7%)	525 (54.6%)	
<b>거주지역</b>			0.368
서울	110 (12.5%)	131 (13.6%)	
부산	2 (0.2%)	3 (0.3%)	
대구	0 (0.0%)	3 (0.3%)	
인천	16 (1.8%)	22 (2.3%)	
광주	2 (0.2%)	2 (0.2%)	
대전	0 (0.0%)	3 (0.3%)	
울산	0 (0.0%)	1 (0.1%)	
세종	685 (78.1%)	721 (74.9%)	
경기도	7 (0.8%)	10 (1.0%)	
강원도	8 (0.9%)	7 (0.7%)	
충청북도	16 (1.8%)	25 (2.6%)	
충청남도	7 (0.8%)	9 (0.9%)	
전라북도	15 (1.7%)	11 (1.1%)	
전라남도	3 (0.3%)	10 (1.0%)	
경상북도	4 (0.5%)	4 (0.4%)	
경상남도	2 (0.2%)	0 (0.0%)	
제주특별자치도			0.368
<b>도시규모</b>			0.000
광역시	149 (17.0%)	127 (13.2%)	
광역시-읍	99 (11.3%)	143 (14.9%)	
광역시-면	23 (2.6%)	38 (4.0%)	
시지역-읍	30 (3.4%)	28 (2.9%)	
시지역-면	83 (9.5%)	32 (3.3%)	
시지역-면	432 (49.3%)	498 (51.8%)	
군지역-읍	53 (6.0%)	80 (8.3%)	

	생존 (n = 877)	사망 (n = 962)	P-value
군지역-면	8 (0.9%)	16 (1.7%)	
<b>1차 장애 등급</b>			0.000
(-)	420 (47.9%)	534 (55.5%)	
1	68 (7.7%)	75 (7.8%)	
2	75 (8.6%)	91 (9.5%)	
3	69 (7.9%)	57 (5.9%)	
4	82 (9.4%)	69 (7.2%)	
5	98 (11.2%)	84 (8.7%)	
6	65 (7.4%)	52 (5.4%)	
<b>1차 장애 종류</b>			0.000
비장애인	420 (47.9%)	534 (55.5%)	
지체장애인	116 (13.2%)	120 (12.5%)	
뇌병변장애인	239 (27.3%)	176 (18.3%)	
시각장애인	26 (3.0%)	25 (2.6%)	
청각장애인	41 (4.7%)	44 (4.6%)	
지적장애인	10 (1.1%)	5 (0.5%)	
정신장애인	2 (0.2%)	1 (0.1%)	
언어장애인	4 (0.5%)	3 (0.3%)	
자폐성장래인	16 (1.8%)	40 (4.2%)	
심장장애인	2 (0.2%)	7 (0.7%)	
호흡기장애인	0 (0.0%)	2 (0.2%)	
간장애인	0 (0.0%)	1 (0.1%)	
장루요루장애인	1 (0.1%)	4 (0.4%)	
<b>2차 장애등급</b>			0.000
(-)			
1	420 (47.9%)	534 (55.5%)	
2	228 (26.0%)	127 (13.2%)	
3	226 (25.8%)	116 (12.1%)	
4	0 (0.0%)	54 (5.6%)	
5	0 (0.0%)	47 (4.9%)	
6	2 (0.2%)	54 (5.6%)	
<b>2차 장애 종류</b>			0.000
비장애인	420 (47.9%)	534 (55.5%)	
지체장애인	101 (11.5%)	95 (9.9%)	
뇌병변장애인	248 (28.3%)	195 (20.3%)	
시각장애인	22 (2.5%)	25 (2.6%)	

	생존 (n = 877)	사망 (n = 962)	P-value
청각장애인	43 (4.9%)	40 (4.2%)	
지적장애인	17 (1.9%)	4 (0.4%)	
정신장애인	4 (0.5%)	3 (0.3%)	
언어장애인	4 (0.5%)	3 (0.3%)	
자폐성장애인	18 (2.1%)	54 (5.6%)	
심장장애인	0 (0.0%)	4 (0.4%)	
호흡기장애인	0 (0.0%)	2 (0.2%)	
장루요루장애인	0 (0.0%)	3 (0.3%)	
<b>총입원기간 (일)</b>	25.9 ± 50.5	24.2 ± 28.6	0.381
<b>일반병동 입원기간 (일)</b>	14.8 ± 17.3	15.8 ± 18.8	0.211
<b>중환자실 입원기간 (일)</b>	11.2 ± 46.5	8.4 ± 21.8	0.103
<b>총 협진 횟수</b>	9.3 ± 9.7	10.2 ± 10.1	0.068
<b>호흡기내과 협진 횟수</b>	0.3 ± 0.8	0.4 ± 1.2	0.000
<b>감염내과 협진 횟수</b>	0.4 ± 1.3	0.4 ± 1.4	0.352
<b>위루술시행</b>			0.000
(-)	2934 (99.5%)	1808 (98.3%)	
(+)	16 (0.5%)	31 (1.7%)	
<b>L_tube삽입</b>			0.000
(-)	647 (73.8%)	524 (54.5%)	
(+)	230 (26.2%)	438 (45.5%)	
<b>폐렴</b>			0.000
(-)	437 (49.8%)	295 (30.7%)	
(+)	440 (50.2%)	667 (69.3%)	
<b>T_tube 삽입</b>			0.499
(-)	851 (97.0%)	927 (96.4%)	
(+)	26 (3.0%)	35 (3.6%)	
<b>외상성 뇌손상</b>			0.994
(-)	837 (95.4%)	917 (95.3%)	
(+)	40 (4.6%)	45 (4.7%)	
<b>비외상성 뇌손상</b>			1.000
(-)	7 (0.8%)	8 (0.8%)	
(+)	870 (99.2%)	954 (99.2%)	
<b>뇌경색</b>			0.069
(-)	224 (25.5%)	210 (21.8%)	
(+)	653 (74.5%)	752 (78.2%)	
<b>지주막하출혈 (SAH)</b>			0.801

	생존 (n = 877)	사망 (n = 962)	P-value
(-)	815 (92.9%)	890 (92.5%)	
(+)	62 (7.1%)	72 (7.5%)	
<b>경막하출혈 (SDH)</b>			0.325
(-)	797 (90.9%)	860 (89.4%)	
(+)	80 (9.1%)	102 (10.6%)	
<b>뇌실질내출혈 (ICH)</b>			0.000
(-)	601 (68.5%)	764 (79.4%)	
(+)	276 (31.5%)	198 (20.6%)	
<b>뇌실내출혈 (IVH)</b>			0.283
(-)	815 (92.9%)	880 (91.5%)	
(+)	62 (7.1%)	82 (8.5%)	
<b>Focal_small</b>			0.206
(-)	334 (38.1%)	338 (35.1%)	
(+)	543 (61.9%)	624 (64.9%)	
<b>multifocal_large</b>			0.000
(-)	566 (64.5%)	515 (53.5%)	
(+)	311 (35.5%)	447 (46.5%)	
<b>Acute_recent</b>			0.000
(-)	929 (41.2%)	202 (29.1%)	
(+)	1326 (58.8%)	493 (70.9%)	
<b>Chronic_old</b>			1.000
(-)	206 (23.5%)	225 (23.4%)	
(+)	671 (76.5%)	737 (76.6%)	
<b>Temporal</b>			0.230
(-)	262 (29.9%)	262 (27.2%)	
(+)	615 (70.1%)	700 (72.8%)	
<b>Parietal</b>			1.000
(-)	874 (99.7%)	958 (99.6%)	
(+)	3 (0.3%)	4 (0.4%)	
<b>Occipital</b>			0.527
(-)	812 (92.6%)	882 (91.7%)	
(+)	65 (7.4%)	80 (8.3%)	
<b>Cerebellum</b>			0.021
(-)	719 (82.0%)	746 (77.5%)	
(+)	158 (18.0%)	216 (22.5%)	
<b>BG</b>			0.489

	생존 (n = 877)	사망 (n = 962)	P-value
(-)	460 (52.5%)	488 (50.7%)	
(+)	417 (47.5%)	474 (49.3%)	
<b>subcortex</b>			0.767
(-)	875 (99.8%)	958 (99.6%)	
(+)	2 (0.2%)	4 (0.4%)	
<b>brain_stem</b>			0.594
(-)	773 (88.1%)	839 (87.2%)	
(+)	104 (11.9%)	123 (12.8%)	
<b>Thalamus</b>			0.026
(-)	743 (84.7%)	776 (80.7%)	
(+)	134 (15.3%)	186 (19.3%)	
<b>ACA</b>			0.285
(-)	820 (93.5%)	886 (92.1%)	
(+)	57 (6.5%)	76 (7.9%)	
<b>MCA</b>			0.102
(-)	642 (73.2%)	670 (69.6%)	
(+)	235 (26.8%)	292 (30.4%)	
<b>PCA</b>			0.889
(-)	796 (90.8%)	876 (91.1%)	
(+)	81 (9.2%)	86 (8.9%)	
<b>basilar_artery</b>			0.738
(-)	838 (95.6%)	915 (95.1%)	
(+)	39 (4.4%)	47 (4.9%)	
<b>Aneurysm</b>			0.020
(-)	760 (86.7%)	868 (90.2%)	
(+)	117 (13.3%)	94 (9.8%)	
<b>Coil</b>			0.325
(-)	859 (97.9%)	949 (98.6%)	
(+)	18 (2.1%)	13 (1.4%)	
<b>craniectomy</b>			0.024
(-)	847 (96.6%)	946 (98.3%)	
(+)	30 (3.4%)	16 (1.7%)	
<b>craniotomy</b>			0.008
(-)	835 (95.2%)	939 (97.6%)	
(+)	42 (4.8%)	23 (2.4%)	
<b>흡연여부 (smoke_yn)</b>			0.040

	생존 (n = 877)	사망 (n = 962)	P-value
현재도 흡연	451 (51.4%)	503 (52.3%)	
흡연하다 금연	331 (37.7%)	387 (40.2%)	
금연	95 (10.8%)	72 (7.5%)	
<b>음주여부 (drink_yn)</b>			0.000
현재도 음주	605 (69.0%)	717 (74.5%)	
음주하다 금주	137 (15.6%)	160 (16.6%)	
금주	135 (15.4%)	85 (8.8%)	
<b>ALBUMIN/ (g/dL)</b>	3.9 ±0.4	3.6 ±0.6	0.000
<b>BUN (mg/dL)</b>	17.4 ± 10.1	20.9 ± 13.9	0.000
<b>CR (mg/dL)</b>	1.0 ±1.1	1.3 ±1.5	0.000
<b>GLUCOSE_AC (mg/dL)</b>	129.3 ± 45.9	137.9 ± 60.6	0.001
<b>HB (g/dL)</b>	13.4 ±2.0	12.7 ±2.0	0.000
<b>HBA1C (%)</b>	6.5 ±1.3	6.6 ±1.5	0.036
<b>육창</b>			0.000
(-)	846 (96.5%)	880 (91.5%)	
(+)	31 (3.5%)	82 (8.5%)	

노인 장기요양보험 서비스를 이용자군에서 사망자군에서 인정 조사 점수가 높았으며, 사망자 군에서 등급판정인정등급이 높았다. 일상생활 환산 점수, 인지 환산점수, 행동 환산 점수, 간호 환산 점수, 재활환산 점수, 흡인, 산소요법, 육창간호, 경관 영양, 도뇨, 간호 투석, 약챙겨먹기, 완전 외상상태, 치매 완전 의존 모두 사망군에서 높았다 <표 3-3>.

<표 3-3> 노인 장기요양보험 서비스 이용자군의 생존군과 사망군의 비교

	생존 (n = 877)	사망 (n = 962)	P-value
<b>장기요양대상자구분코드</b>			0.000
국민기초수급권자	117 (13.3%)	108 (11.2%)	
건강보험감경대상자	216 (24.6%)	76 (7.9%)	
건강보험가입자	541 (61.7%)	768 (79.8%)	
의료급여수급권자	3 (0.3%)	10 (1.0%)	
<b>장기요양경감구분</b>			0.000
기초	119 (13.6%)	111 (11.5%)	
의료급여	5 (0.6%)	11 (1.1%)	
건강보험	479 (54.6%)	826 (85.9%)	

	생존 (n = 877)	사망 (n = 962)	P-value
저소득	238 (27.1%)	1 (0.1%)	
희귀난치성	3 (0.3%)	1 (0.1%)	
만성질환	19 (2.2%)	0 (0.0%)	
값없음	14 (1.6%)	12 (1.2%)	
<b>인정조사인정점수</b>	60.8 ± 17.3	70.1 ± 22.4	0.000
<b>등급판정인정등급코드</b>			0.000
1등급	48 (5.5%)	121 (12.6%)	
2등급	64 (7.3%)	146 (15.2%)	
3등급	232 (26.5%)	320 (33.3%)	
4등급	247 (28.2%)	144 (15.0%)	
5등급	66 (7.5%)	21 (2.2%)	
<b>인지지원등급</b>	18 (2.1%)	2 (0.2%)	
등급외A	50 (5.7%)	61 (6.3%)	
등급외B	45 (5.1%)	21 (2.2%)	
등급외C	14 (1.6%)	5 (0.5%)	
기각	29 (3.3%)	40 (4.2%)	
각하	64 (7.3%)	81 (8.4%)	
<b>주거형태코드</b>			0.000
노인요양시설	629 (71.7%)	629 (65.4%)	
노인전문요양시설	82 (9.4%)	94 (9.8%)	
노인요양공동생활가정	0 (0.0%)	4 (0.4%)	
단기보호시설	1 (0.1%)	13 (1.4%)	
노인요양시설	1 (0.1%)	1 (0.1%)	
요양병원	128 (14.6%)	199 (20.7%)	
기타	35 (4.0%)	21 (2.2%)	
양로시설	1 (0.1%)	1 (0.1%)	
<b>동거인코드</b>			0.000
독거	158 (18.0%)	89 (9.3%)	
배우자	212 (24.2%)	223 (23.2%)	
부모	4 (0.5%)	4 (0.4%)	
자녀(며느리,사위포함)	168 (19.2%)	220 (22.9%)	
손자녀	2 (0.2%)	3 (0.3%)	
친척	4 (0.5%)	4 (0.4%)	
친구·이웃	2 (0.2%)	2 (0.2%)	
입소시설관계자	108 (12.3%)	134 (13.9%)	
기타	115 (13.1%)	151 (15.7%)	
배우자+자녀	59 (6.7%)	57 (5.9%)	

	생존 (n = 877)	사망 (n = 962)	P-value
배우자+손자녀	2 (0.2%)	0 (0.0%)	
배우자+자녀+손자녀	6 (0.7%)	14 (1.5%)	
부모+배우자	3 (0.3%)	0 (0.0%)	
부모+자녀	0 (0.0%)	1 (0.1%)	
부모+손자녀	1 (0.1%)	0 (0.0%)	
부모+배우자+자녀	3 (0.3%)	0 (0.0%)	
부모+배우자+손자녀	30 (3.4%)	49 (5.1%)	
자녀+손자녀	0 (0.0%)	11 (1.1%)	
<b>수발자코드</b>			0.000
없음	63 (7.2%)	44 (4.6%)	
배우자	243 (27.7%)	254 (26.4%)	
시설관계자	5 (0.6%)	0 (0.0%)	
자녀	56 (6.4%)	247 (25.7%)	
며느리	4 (0.5%)	5 (0.5%)	
사위	12 (1.4%)	11 (1.1%)	
형제·자매	14 (1.6%)	15 (1.6%)	
부모	118 (13.5%)	197 (20.5%)	
자녀(며느리,사위포함)	2 (0.2%)	3 (0.3%)	
손자녀	24 (2.7%)	49 (5.1%)	
친척	61 (7.0%)	34 (3.5%)	
친구·이웃	229 (26.1%)	89 (9.3%)	
간병인	33 (3.8%)	13 (1.4%)	
자원봉사자	2 (0.2%)	0 (0.0%)	
기타	11 (1.3%)	1 (0.1%)	
<b>일상생활환산점수</b>	36.6 ± 20.2	47.1 ± 23.6	0.000
<b>인지환산점수</b>	44.3 ± 23.0	52.9 ± 24.2	0.000
<b>행동환산점수</b>	5.1 ± 9.9	6.9 ± 12.0	0.000
<b>간호환산점수</b>	2.8 ± 9.0	7.1 ± 14.0	0.000
<b>재활환산점수</b>	15.9 ± 16.5	19.5 ± 18.4	0.000
<b>기관지절개관간호</b>			0.190
(-)	868 (99.0%)	944 (98.1%)	
(+)	9 (1.0%)	18 (1.9%)	
<b>흡인</b>			0.035
(-)	861 (98.2%)	928 (96.5%)	
(+)	16 (1.8%)	34 (3.5%)	
<b>산소요법</b>			0.004
(-)	872 (99.4%)	940 (97.7%)	



	생존 (n = 877)	사망 (n = 962)	P-value
(+)	5 (0.6%)	22 (2.3%)	
<b>육창간호</b>			0.000
(-)	862 (98.3%)	900 (93.6%)	
(+)	15 (1.7%)	62 (6.4%)	
<b>경관영양(튜브급식)</b>			0.000
(-)	842 (96.0%)	870 (90.4%)	
(+)	35 (4.0%)	92 (9.6%)	
<b>암성통증간호</b>			0.804
(-)	862 (98.3%)	943 (98.0%)	
(+)	15 (1.7%)	19 (2.0%)	
<b>도뇨관리</b>			0.000
(-)	847 (96.6%)	870 (90.4%)	
(+)	30 (3.4%)	92 (9.6%)	
<b>장루(인공항문)간호</b>			0.767
(-)	875 (99.8%)	958 (99.6%)	
(+)	2 (0.2%)	4 (0.4%)	
<b>간호투석</b>			0.000
(-)	860 (98.1%)	911 (94.7%)	
(+)	17 (1.9%)	51 (5.3%)	
<b>당뇨발간호</b>			0.163
(-)	876 (99.9%)	956 (99.4%)	
(+)	1 (0.1%)	6 (0.6%)	
<b>약행겨먹기</b>			0.000
완전자립	188 (21.4%)	102 (10.6%)	
부분도움	529 (60.3%)	549 (57.1%)	
완전도움	160 (18.2%)	311 (32.3%)	
<b>장애노인(와상도)</b>			0.000
정상	115 (13.1%)	86 (8.9%)	
생활자립	463 (52.8%)	369 (38.4%)	
준 와상상태	226 (25.8%)	343 (35.7%)	
완전 와상 상태	73 (8.3%)	164 (17.0%)	
<b>치매(인지증)노인</b>			0.000
자립	270 (30.8%)	281 (29.2%)	
불완전 자립	508 (57.9%)	497 (51.7%)	
부분 의존	79 (9.0%)	151 (15.7%)	
완전 의존	20 (2.3%)	33 (3.4%)	

노인 장기요양 서비스 미이용자군의 생존군과 사망자 군간의 비교에서, 사망자군이 생존기간  $1,234.4 \pm 1,090.9$ 일로 생존군 ( $2,175.0 \pm 921.7$ ) 보다 짧았다. 또한 연령 또한 사망자군이  $71.3 \pm 13.0$ 세로 생존군의  $59.7 \pm 12.5$ 세로 볼 때 사망자군이, 노인 장기요양보험 서비스를 받지 못하고 발병후 조기에 사망하여 받지 못한 것으로 여겨진다. 성별은 양군간의 차이가 없었으며 사망군은 생존군에 비해 입원중 총입원기간, 일반 병동 입원기간, 중환자실 입원기간, 총 협진횟수, 호흡기 및 감염내과 협진, 위루술, L-tube 삽입, 입원중 폐렴 여부, 기관절개 관삽입(T-tube), 욕창발생, Albumin 및 혈색소(Hb) 저하, 공복혈당 및 당화혈색소, BUN/Cr의 상승이 통계적으로 차이가 있었다 <표 3-4>.

<표 3-4> 노인 장기요양보험 서비스 미이용자군의 생존 여부에 따른 인구학적 특성 및 병원 자료의 비교

	생존 (n=2255)	사망 (n=695)	P-value
<b>발병후 생존기간(일)</b>	2175.0 ± 921.7	717.8 ± 946	0.000
<b>연령(연도)</b>	59.7 ± 12.5	71.3 ± 13.0	0.000
<b>성별</b>			0.764
남	1382 (61.3%)	431 (62.0%)	
여	873 (38.7%)	264 (38.0%)	
<b>거주지역</b>			0.564
서울	294 (13.0%)	89 (12.8%)	
부산	6 (0.3%)	3 (0.4%)	
대구	5 (0.2%)	2 (0.3%)	
인천	55 (2.4%)	12 (1.7%)	
광주	4 (0.2%)	3 (0.4%)	
대전	5 (0.2%)	1 (0.1%)	
울산	1 (0.0%)	0 (0.0%)	
세종	0 (0.0%)	1 (0.1%)	
경기도	1800 (79.8%)	549 (79.0%)	
강원도	15 (0.7%)	7 (1.0%)	
충청북도	9 (0.4%)	4 (0.6%)	
충청남도	11 (0.5%)	7 (1.0%)	
전라북도	11 (0.5%)	3 (0.4%)	
전라남도	10 (0.4%)	7 (1.0%)	
경상북도	21 (0.9%)	4 (0.6%)	
경상남도	6 (0.3%)	2 (0.3%)	

	생존 (n=2255)	사망 (n=695)	P-value
제주특별자치도	2 (0.1%)	1 (0.1%)	
<b>도시규모</b>			0.000
광역시	513 (22.7%)	100 (14.4%)	
광역시-읍	271 (12.0%)	108 (15.5%)	
광역시-면	41 (1.8%)	12 (1.7%)	
시지역-읍	27 (1.2%)	15 (2.2%)	
시지역-면	579 (25.7%)	79 (11.4%)	
시지역-면	765 (33.9%)	317 (45.6%)	
군지역-읍	49 (2.2%)	61 (8.8%)	
군지역-면	10 (0.4%)	3 (0.4%)	
<b>1차 장애 등급</b>			0.000
(-)	1813 (80.4%)	508 (73.1%)	
1	32 (1.4%)	32 (4.6%)	
2	87 (3.9%)	44 (6.3%)	
3	36 (1.6%)	20 (2.9%)	
4	59 (2.6%)	27 (3.9%)	
5	100 (4.4%)	35 (5.0%)	
6	128 (5.7%)	29 (4.2%)	
<b>1차 장애 종류</b>			0.000
비장애인	1813 (80.4%)	508 (73.1%)	
지체장애인	149 (6.6%)	55 (7.9%)	
뇌병변장애인	134 (5.9%)	34 (4.9%)	
시각장애인	34 (1.5%)	23 (3.3%)	
청각장애인	59 (2.6%)	24 (3.5%)	
지적장애인	20 (0.9%)	3 (0.4%)	
정신장애인	9 (0.4%)	3 (0.4%)	
신장장애인	3 (0.1%)	0 (0.0%)	
언어장애인	26 (1.2%)	34 (4.9%)	
자폐성장장애인	3 (0.1%)	2 (0.3%)	
심장장애인	2 (0.1%)	2 (0.3%)	
호흡기장애인	0 (0.0%)	4 (0.6%)	
간장애인	2 (0.1%)	2 (0.3%)	
간질장애인	1 (0.0%)	1 (0.1%)	
<b>2차 장애등급</b>			0.000
(-)	1813 (80.4%)	508 (73.1%)	
1	149 (6.6%)	45 (6.5%)	

	생존 (n=2255)	사망 (n=695)	P-value
2	279 (12.4%)	44 (6.3%)	
3	4 (0.2%)	23 (3.3%)	
4	2 (0.1%)	22 (3.2%)	
5	1 (0.0%)	31 (4.5%)	
6	7 (0.3%)	22 (3.2%)	
<b>2차 장애 종류</b>			0.000
비장애인	1813 (80.4%)	508 (73.1%)	
지체장애인	141 (6.3%)	51 (7.3%)	
뇌병변장애인	130 (5.8%)	43 (6.2%)	
시각장애인	36 (1.6%)	19 (2.7%)	
청각장애인	63 (2.8%)	25 (3.6%)	
지적장애인	20 (0.9%)	1 (0.1%)	
정신장애인	13 (0.6%)	2 (0.3%)	
신장장애인	3 (0.1%)	0 (0.0%)	
언어장애인	30 (1.3%)	39 (5.6%)	
자폐성장장애인	2 (0.1%)	2 (0.3%)	
심장장애인	1 (0.0%)	1 (0.1%)	
호흡기장애인	0 (0.0%)	2 (0.3%)	
간장애인	2 (0.1%)	2 (0.3%)	
간질장애인	1 (0.0%)	0 (0.0%)	
<b>총입원기간 (일)</b>	13.7 ± 18.0	23.9 ± 37.9	0.000
<b>일반병동 입원기간 (일)</b>	8.7 ± 10.6	15.1 ± 24.6	0
<b>중환자실 입원기간 (일)</b>	5.0 ± 12.6	8.8 ± 26.1	0.000
<b>총 협진 횟수</b>	5.8 ± 5.5	11.3 ± 13.0	0.000
<b>호흡기내과 협진 횟수</b>	0.1 ± 0.5	0.5 ± 1.3	0.000
<b>감염내과 협진 횟수</b>	0.1 ± 0.8	0.7 ± 1.9	0.000
<b>위루술시행</b>			0.000
(-)	2934 (99.5%)	1808 (98.3%)	
(+)	16 (0.5%)	31 (1.7%)	
<b>L_tube삽입</b>			0.000
(-)	2116 (93.8%)	429 (61.7%)	
(+)	139 (6.2%)	266 (38.3%)	
<b>폐렴</b>			0.000
(-)	1567 (69.5%)	207 (29.8%)	
(+)	688 (30.5%)	488 (70.2%)	
<b>T_tube 삽입</b>			0.000

	생존 (n=2255)	사망 (n=695)	P-value
(-)	2242 (99.4%)	656 (94.4%)	
(+)	13 (0.6%)	39 (5.6%)	
<b>외상성 뇌손상</b>			0.558
(-)	2180 (96.7%)	668 (96.1%)	
(+)	75 (3.3%)	27 (3.9%)	
<b>비외상성 뇌손상</b>			0.914
(-)	7 (0.3%)	3 (0.4%)	
(+)	2248 (99.7%)	692 (99.6%)	
<b>뇌경색</b>			0.000
(-)	1134 (50.3%)	245 (35.3%)	
(+)	1121 (49.7%)	450 (64.7%)	
<b>지주막하출혈 (SAH)</b>			0.042
(-)	1982 (87.9%)	631 (90.8%)	
(+)	273 (12.1%)	64 (9.2%)	
<b>경막하출혈 (SDH)</b>			0.050
(-)	2094 (92.9%)	629 (90.5%)	
(+)	161 (7.1%)	66 (9.5%)	
<b>뇌실질내출혈 (ICH)</b>			0.083
(-)	1785 (79.2%)	528 (76.0%)	
(+)	470 (20.8%)	167 (24.0%)	
<b>뇌실내출혈(IVH)</b>			0.000
(-)	2169 (96.2%)	609 (87.6%)	
(+)	86 (3.8%)	86 (12.4%)	
<b>Focal_small</b>			0.000
(-)	1458 (64.7%)	373 (53.7%)	
(+)	797 (35.3%)	322 (46.3%)	
<b>multifocal_large</b>			0.000
(-)	1932 (85.7%)	402 (57.8%)	
(+)	323 (14.3%)	293 (42.2%)	
<b>Acute_recent</b>			0.000
(-)	929 (41.2%)	202 (29.1%)	
(+)	1326 (58.8%)	493 (70.9%)	
<b>Chronic_old</b>			0.000
(-)	1174 (52.1%)	278 (40.0%)	
(+)	1081 (47.9%)	417 (60.0%)	
<b>Frontal</b>			

	생존 (n=2255)	사망 (n=695)	P-value
(-)	2950 (100.0%)	1839 (100.0%)	
<b>Temporal</b>			0.065
(-)	2117 (93.9%)	638 (91.8%)	
(+)	138 (6.1%)	57 (8.2%)	
<b>Parietal</b>			0.894
(-)	2249 (99.7%)	694 (99.9%)	
(+)	6 (0.3%)	1 (0.1%)	
<b>Occipital</b>			0.000
(-)	2177 (96.5%)	635 (91.4%)	
(+)	78 (3.5%)	60 (8.6%)	
<b>Cerebellum</b>			0.000
(-)	2049 (90.9%)	564 (81.2%)	
(+)	206 (9.1%)	131 (18.8%)	
<b>BG</b>			0.000
(-)	1602 (71.0%)	437 (62.9%)	
(+)	653 (29.0%)	258 (37.1%)	
<b>subcortex</b>			1.000
(-)	2250 (99.8%)	693 (99.7%)	
(+)	5 (0.2%)	2 (0.3%)	
<b>brain_stem</b>			0.000
(-)	2118 (93.9%)	618 (88.9%)	
(+)	137 (6.1%)	77 (11.1%)	
<b>Thalamus</b>			0.000
(-)	2065 (91.6%)	611 (87.9%)	
(+)	190 (8.4%)	84 (12.1%)	
<b>ACA</b>			0.004
(-)	2186 (96.9%)	657 (94.5%)	
(+)	69 (3.1%)	38 (5.5%)	
<b>MCA</b>			0.000
(-)	1777 (78.8%)	502 (72.2%)	
(+)	478 (21.2%)	193 (27.8%)	
<b>PCA</b>			0.000
(-)	2188 (97.0%)	637 (91.7%)	
(+)	67 (3.0%)	58 (8.3%)	
<b>basilar_artery</b>			0.000
(-)	2212 (98.1%)	663 (95.4%)	

	생존 (n=2255)	사망 (n=695)	P-value
(+)	43 (1.9%)	32 (4.6%)	
<b>Aneurysm</b>			0.000
(-)	1675 (74.3%)	595 (85.6%)	
(+)	580 (25.7%)	100 (14.4%)	
<b>Coil</b>			0.000
(-)	2108 (93.5%)	682 (98.1%)	
(+)	147 (6.5%)	13 (1.9%)	
<b>craniectomy</b>			0.006
(-)	2072 (91.9%)	661 (95.1%)	
(+)	183 (8.1%)	34 (4.9%)	
<b>craniotomy</b>			0.006
(-)	2072 (91.9%)	661 (95.1%)	
(+)	183 (8.1%)	34 (4.9%)	
<b>흡연여부 (smoke_yn)</b>			0.000
현재도 흡연	1154 (51.2%)	450 (64.7%)	
흡연하다 금연	306 (13.6%)	103 (14.8%)	
금연	795 (35.3%)	142 (20.4%)	
<b>음주여부 (drink_yn)</b>			0.000
현재도 음주	1604 (54.4%)	1322 (71.9%)	
음주하다 금주	409 (13.9%)	297 (16.2%)	
금주	937 (31.8%)	220 (12.0%)	
<b>ALBUMIN/ (g/dL)</b>	4.1 ±0.4	3.6 ±0.6	0.000
<b>BUN (mg/dL)</b>	15.0 ±6.9	22.3 ± 16.1	0.000
<b>CR (mg/dL)</b>	0.9 ±0.9	1.6 ±2.2	0.000
<b>GLUCOSE_AC (mg/dL)</b>	125.0 ± 45.0	143.4 ± 59.	0.034
<b>HB (g/dL)</b>	14.2 ±1.8	12.8 ±2.4	0.000
<b>HBA1C (%)</b>	6.4 ±1.3	6.6 ±1.2	0.032
<b>육창</b>			0.000
(-)	2233 (99.0%)	655 (94.2%)	
(+)	22 (1.0%)	40 (5.8%)	

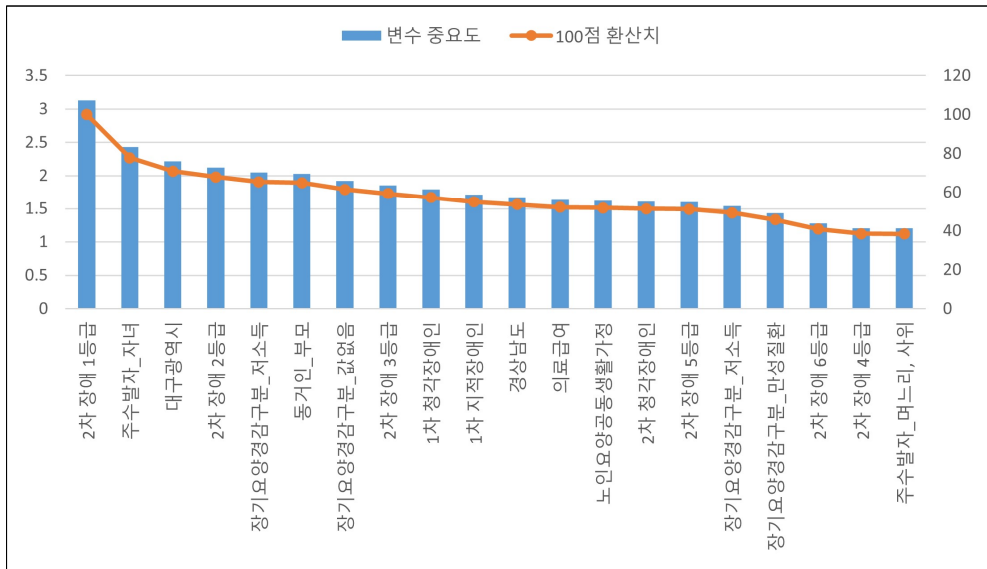
## 제1절 노인 장기요양보험 서비스 이용자군의 생존 분석의 기계 학습(machine learning)

NN 분석에 관련된 코드와 결과 값은 appendix 166p에 있으며, 237개 변수 중 상위 20개 주요 항목은 다음과 같다.

<표 3-5> 장기요양보험 이용자군의 생존 분석의 NN 기계 학습 주요 인자

상위 20개 주요 항목	주요 인자	100점 환산치
2차 장애 1등급	3.128	100
주수발자_자녀	2.434	77.79
대구광역시	2.22	70.95
2차 장애 2등급	2.127	67.98
장기요양경감구분_저소득	2.048	65.45
동거인_노인	2.032	64.96
장기요양경감구분_값없음	1.926	61.55
2차 장애 3등급	1.857	59.36
1차 청각장애인	1.793	57.3
1차 지적장애인	1.718	54.91
경상남도	1.677	53.59
의료급여	1.633	52.19
노인요양공동생활가정	1.622	51.84
2차 청각장애인	1.607	51.34
2차 장애 5등급	1.602	51.2
장기요양경감구분_저소득	1.542	49.26
장기요양경감구분_만성질환	1.431	45.72
2차 장애 6등급	1.282	40.96
2차 장애 4등급	1.206	38.53
주수발자_며느리, 사위	1.202	38.4





[그림 3-3] 장기요양보험 이용자군의 생존 분석의 NN 기계 학습 주요 인자

정확도(Accuracy)는 0.8161, 95% 신뢰구간(CI)은 (0.7806, 0.8481), No Information Rate는 0.5235로 P-Value [Acc > NIR]: <math>2e-16</math>로 관찰되었다.

민감도(Sensitivity)는 0.8071, Specificity는 0.8244, 양성 예측도(Positive Predictive Value)는 0.8071, 음성 예측도 (Negative Predictive Value)는 0.8244, Balanced Accuracy인 AUC 값은 0.8157이었다 <표 3-6>.

<표 3-6> NN의 실험 세트 검증

Prediction	생존	사망
생존	205	49
사망	49	230

SVM 분석에 관련된 코드와 결과 값은 appendix 166p에 있으며 정확도는 0.8105, 95% 신뢰구간(CI)은 (0.7746, 0.8429), No Information Rate는 0.52350 로 P-Value [Acc > NIR]: <math>2e-16</math>로 관찰되었다.

민감도는 0.7598, 특이도는 0.8566, 양성 예측도는 0.8283, 음성 예측도는 0.7967, Balanced Accuracy인 AUC 값은 0.8082이었다 <표 3-7>.

<표 3-7> SVM의 실험 세트 검증

Prediction	생존	사망
생존	193	40
사망	61	239

KNN 분석에 관련된 코드와 결과 값은 appendix 174p에 있으며 정확도는 0.7073, 95% 신뢰구간(CI)은 (0.6667, 0.7456), No Information Rate는 0.52350로 P-Value [Acc > NIR]: 2.2e-16 로 관찰되었다.

민감도는 0.7756, 특이도는 0.642, 양성 예측도는 0.6452, 음성 예측도는 0.7595, Balanced Accuracy인 AUC 값은 0.714이었다 <표 3-8>.

<표 3-8> KNN의 실험 세트 검증

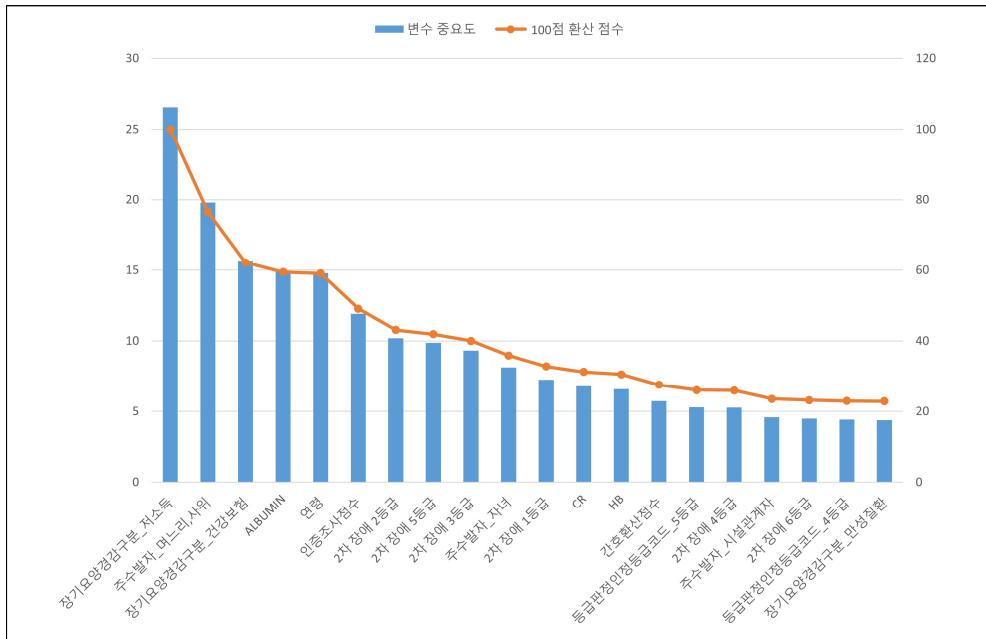
Prediction	생존	사망
생존	197	99
사망	57	180

RF분석에 관련된 코드와 결과 값은 appendix 178p에 있으며, 237개 변수 중 상위 20개 주요 항목은 다음과 같다 <표 3-9>, [그림 3-4].

<표 3-9> 노인 장기요양보험 서비스 이용자군의 생존 분석의 RF기계 학습 주요 인자

상위 20개 주요 항목	변수 중요도	100점 환산 점수
장기요양경감구분_저소득	26.559	100.000
주수발자_머느리,사위	19.805	76.510
장기요양경감구분_건강보험	15.69	62.190
ALBUMIN	14.907	59.470
연령	14.794	59.080
인증조사점수	11.929	49.110
2차 장애 2등급	10.201	43.100
2차 장애 5등급	9.861	41.920
2차 장애 3등급	9.315	40.020
주수발자_자녀	8.133	35.900
2차 장애 1등급	7.245	32.810
CR	6.797	31.260
HB	6.596	30.560

상위 20개 주요 항목	변수 중요도	100점 환산 점수
간호환산점수	5.74	27.580
등급판정인정등급코드_5등급	5.302	26.060
2차 장애 4등급	5.277	25.970
주수발자_시설관계자	4.583	23.560
2차 장애 6등급	4.483	23.210
등급판정인정등급코드_4등급	4.417	22.980
장기요양경감구분_만성질환	4.386	22.870



[그림 3-4] 노인 장기요양보험 서비스 이용자군의 생존 분석의 RF기계 학습 주요 인자

정확도는 0.8368, 95% 신뢰구간(CI)은 (0.8026, 0.8671), No Information Rate는 0.5235로 P-Value [Acc > NIR]는 <math>2.2e-16</math>로 관찰되었다.

민감도는 0.7638, 특이도는 0.9032, 양성 예측도는 0.8778, 음성 예측도는 0.8077, Balanced Accuracy인 AUC 값은 0.8335이었다 <표 3-10>.

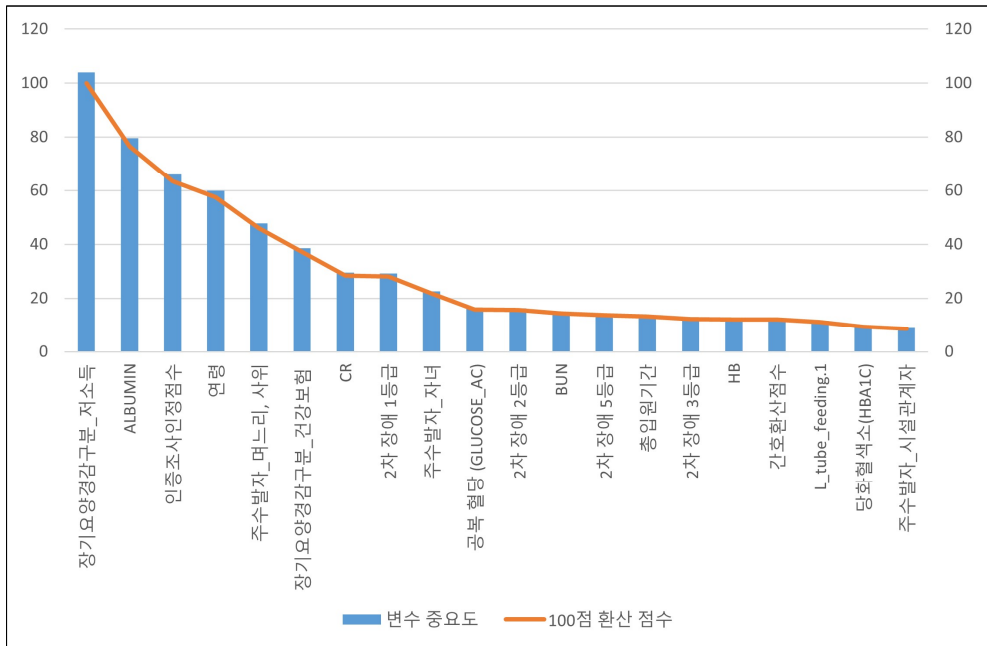
<표 3-10> RF의 실험 세트 검증

Prediction	생존	사망
생존	194	27
사망	60	252

GBM분석에 관련된 코드와 결과 값은 appendix 183p에 있으며, 237개 변수중 상위 20개 주요 항목은 다음과 같다 <표 3-11>, [그림 3-5].

<표 3-11> 노인 장기요양보험 서비스 이용자군의 생존 분석의 GBM기계 학습 주요 인자

상위 20개 주요 항목	변수 중요도	100점 환산 점수
장기요양경감구분_저소득	103.965	100
ALBUMIN	79.561	76.527
인증조사인정점수	65.986	63.469
연령	59.838	57.556
주수발자_머느리, 사위	47.804	45.981
장기요양경감구분_건강보험	38.65	37.176
CR	29.547	28.42
2차 장애 1등급	29.168	28.056
주수발자_자녀	22.634	21.771
공복 혈당 (GLUCOSE_AC)	16.527	15.897
2차 장애 2등급	16.326	15.703
BUN	14.984	14.413
2차 장애 5등급	14.306	13.761
총입원기간	13.677	13.155
2차 장애 3등급	12.707	12.222
HB	12.591	12.111
간호환산점수	12.573	12.094
L_tube_feeding.1	11.574	11.133
당화혈색소(HBA1C)	9.654	9.285
주수발자_시설관계자	8.931	8.59



[그림 3-5] 노인 장기요양보험 서비스 이용자군의 생존 분석의 GBM기계 학습 주요 인자

정확도는 0.8199, 95% 신뢰구간(CI)은 (0.7846, 0.8516), No Information Rate는 0.5235로 P-Value [Acc > NIR]는 <math>2e-16</math>로 관찰되었다.

민감도는 0.7835, 특이도는 0.9032, 양성 예측도는 0.8778, 음성 예측도는 0.8077, Balanced Accuracy인 AUC 값은 0.8335이었다 <표 3-12>.

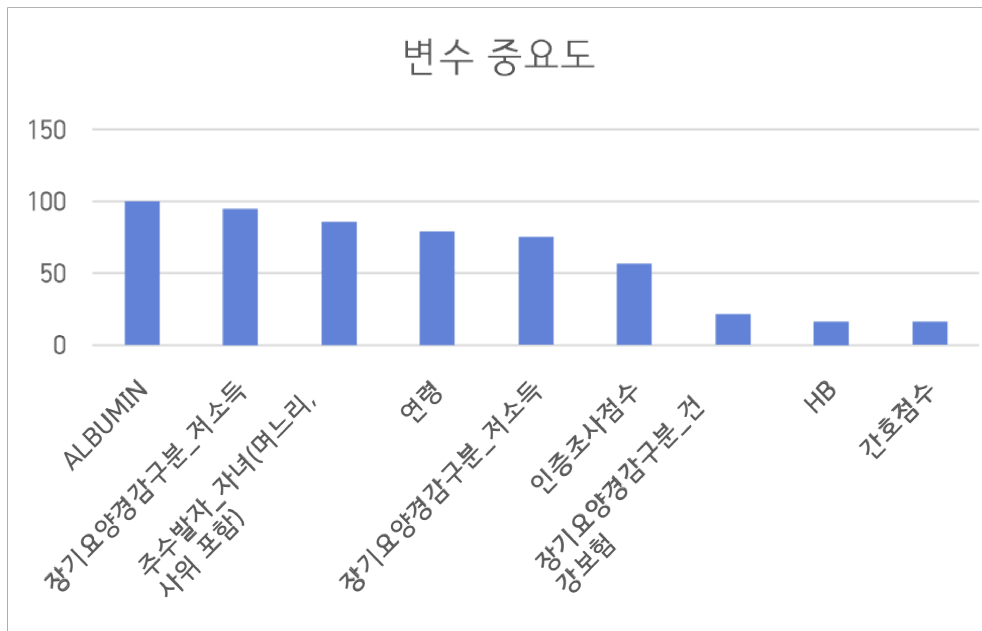
<표 3-12> GBM의 실험 세트 검증

Prediction	생존	사망
생존	199	41
사망	55	238

r\_part 분석에 관련된 코드와 결과 값은 appendix 187p에 있으며, 147개 변수중 상위 20개 주요 항목은 다음과 같다 <표 3-13>, [그림 3-6].

<표 3-13> 노인 장기요양보험 서비스 이용자군의 생존 분석의 r\_part 기계 학습 주요 인자

상위 20개 주요 항목	변수 중요도
ALBUMIN	100
장기요양경감구분_저소득	95.08
주수발자_자녀(머느리, 사위 포함)	85.96
연령	79.3
장기요양경감구분_저소득	75.14
인증조사점수	56.59
장기요양경감구분_건강보험	21.17
HB	16.54
간호점수	15.99



[그림 3-6] 노인 장기요양보험 서비스 이용자군의 생존 분석의 r-part 기계 학습 주요 인자

r-part의 실험 세트 검증의 정확도는 0.7041, 95% 신뢰구간(CI)은 (0.6634, 0.7425), No Information Rate는 0.5로 P-Value [Acc > NIR]는 < 2.2e-16 로 관찰되었다. 민감도는 0.6404, 특이도는 0.7678, 양성 예측도는 0.7339, 음성 예측도는 0.6811, Balanced Accuracy인 AUC 값은 0.7041이었다 <표 3-14>.

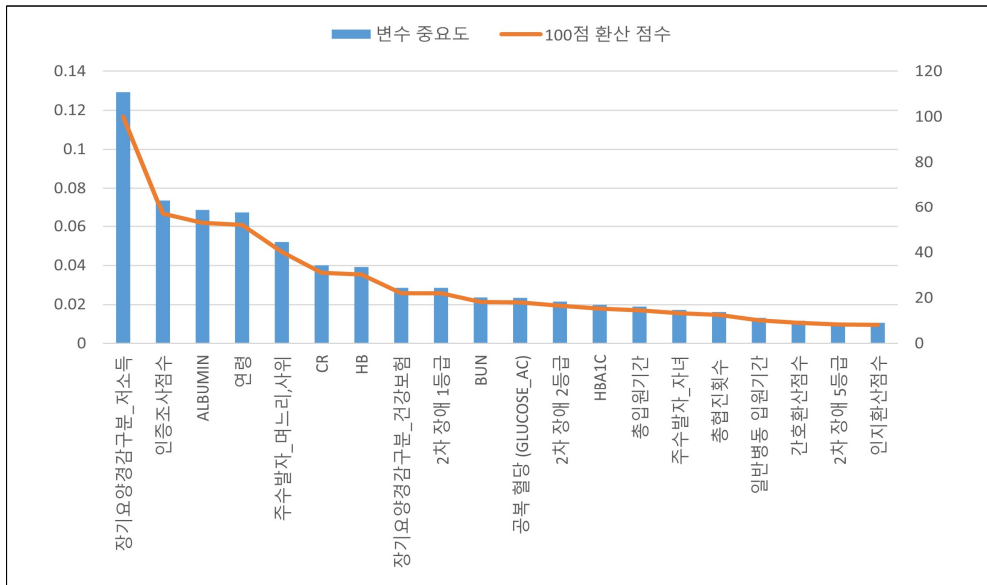
<표 3-14> r-Part의 실험 세트 검증

Prediction	생존	사망
생존	171	62
사망	96	205

Xgboost분석에 관련된 코드와 결과 값은 appendix 191p에 있으며, 237개 변수중 상위 20개 주요 항목은 다음과 같다 <표 3-15>, [그림 3-7].

<표 3-15> 노인 장기요양보험 서비스 이용자군의 생존 분석의 Xgboost

상위 20개 주요 항목	변수 중요도	100점 환산 점수
장기요양경감구분_저소득	0.12928	100
인증조사점수	0.07385	57.119
ALBUMIN	0.06847	52.957
연령	0.06726	52.027
주수발자_며느리,사위	0.05207	40.273
CR	0.04009	31.013
HB	0.03919	30.312
장기요양경감구분_건강보험	0.02851	22.051
2차 장애 1등급	0.02845	22.009
BUN	0.02354	18.205
공복 혈당 (GLUCOSE_AC)	0.02339	18.091
2차 장애 2등급	0.02155	16.666
HBA1C	0.01981	15.319
총입원기간	0.01894	14.647
주수발자_자녀	0.01731	13.388
총협진횟수	0.01614	12.485
일반병동 입원기간	0.0131	10.132
간호환산점수	0.01178	9.113
2차 장애 5등급	0.01071	8.287
인지환산점수	0.01067	8.256



[그림 3-7] 노인 장기요양보험 서비스 이용자군의 생존 분석의 Xgboost

Xgboost의 실험 세트 검증의 정확도는 0.8368, 95% 신뢰구간(CI)은 (0.8026, 0.8671), No Information Rate는 0.5235로 P-Value [Acc > NIR]는 <math>2e-16</math>로 관찰되었다. 민감도는 0.7677, 특이도는 0.8996, 양성 예측도는 0.8097, 음성 예측도는 0.8337, Balanced Accuracy인 AUC 값은 0.8335이었다 <표 3-16>.

<표 3-16> Xgboost의 실험 세트 검증

Prediction	생존	사망
생존	195	28
사망	59	251

여러 인공지능 분석 tool 사이에서 Ensemble model 분석에서 가장 최적의 모델은 Random Forest로 밝혀졌으며 정확도는 0.8643, AUC는 0.8639로 측정되었다 (appendix 195p).

<표 3-17> 노인 장기요양보험 서비스 이용자군의 인공지능 알고리즘에 따른 정확도와 AUC value

	NN	SVM	KNN	RF	gbm	Xgboost
정확도	0.8161	0.8105	0.7073	0.8368	0.8199	0.8368
AUC	0.8157	0.8082	0.7104	0.8335	0.8183	0.8337



노인장기 영양보험 이용자군의 생존 예측의 회귀 분석에서도 기계학습 분석과 유사한 결과가 관찰되었다. 뇌손상 이후의 기능 상태를 반영하는 장애등급, 인증조사점수, 약책 가져먹기 항목, 치매 완전의존, 연령과 성별, 내과적 상태와 영양 상태를 반영하는 albumin, L-tube feeding, 폐렴, multifocal\_large(+) 뇌병변 부위, 기타 장기요양비용\_감액 사유, 시설 입소 등 사회 경제적 이용이 생존 예측 회귀 분석에서 주요 인자로 나타났다 <표 3-18>.

<표 3-18> 노인 장기요양 서비스 이용자군의 생존 회귀 분석에서 상위 20개 변수

항목	Estimate	standard error	t value	r(> t )
2차 장애 1등급	-0.6766041	0.0821004	-8.241	3.31E-16***
연령	0.0091567	0.0011352	8.066	1.33E-15***
2차 장애 2등급	-0.4844421	0.0752091	-6.441	1.53E-10***
ALBUMIN	-0.0980809	0.019042	-5.151	2.89E-07***
수발자(머느리, 사위포함)	0.1888185	0.0425434	4.438	9.64E-06***
인증조사인정점수	0.0050641	0.0011486	4.409	1.10E-05***
장기요양경감구분_건강보험	0.2434287	0.0595411	4.088	4.54E-05***
뇌출혈(+)	-0.08484	0.0218715	-3.879	0.000109***
L_tube_feeding(+)	0.0756416	0.0205017	3.69	0.000232***
성별_여자	-0.0749876	0.0204767	-3.662	0.000258***
약책가져먹기_부분의존	0.0968358	0.0269063	3.599	0.000328***
장기요양경감구분_저소득	-0.1824759	0.0515289	-3.541	0.000409***
치매_완전의존	-0.195052	0.0591424	-3.298	0.000993***
multifocal_large(+)	0.058627	0.0191864	3.054	0.002289**
약책가져먹기_완전의존	0.1057237	0.0346817	3.048	0.002335**
장기요양대상자구분코드	-0.0791161	0.0267261	-2.96	0.003115**
장기요양경감구분_값없음	0.2290171	0.083419	2.745	0.006106**
폐렴(+)	0.0510431	0.0187599	2.721	0.006576**
등급판정인정등급코드_값없음	0.2076688	0.0817281	2.541	0.01114*
Temporal lobe(+)	0.0748856	0.0303653	2.466	0.013753*

여러 기계 분석 알고리즘과 회귀 분석간에 상위 20개 주요 인자간의 인자의 비교는 다음과 같다. 여기서 노란색은 각 알고리즘에서 공통으로 나타난 주요 인자로 2차 장애 1등급과 2등급, 장기요양등급경감 구분에서 저소득층으로 밝혀져, 뇌손상 이후 생존이 기능적 수준과 경제적 수준이 영향을 미침을 알 수 있다 <표 3-19>.

<표 3-19> 노인 장기요양 서비스 이용군의 기계학습 알고리즘에 따른 주요 인자의 비교

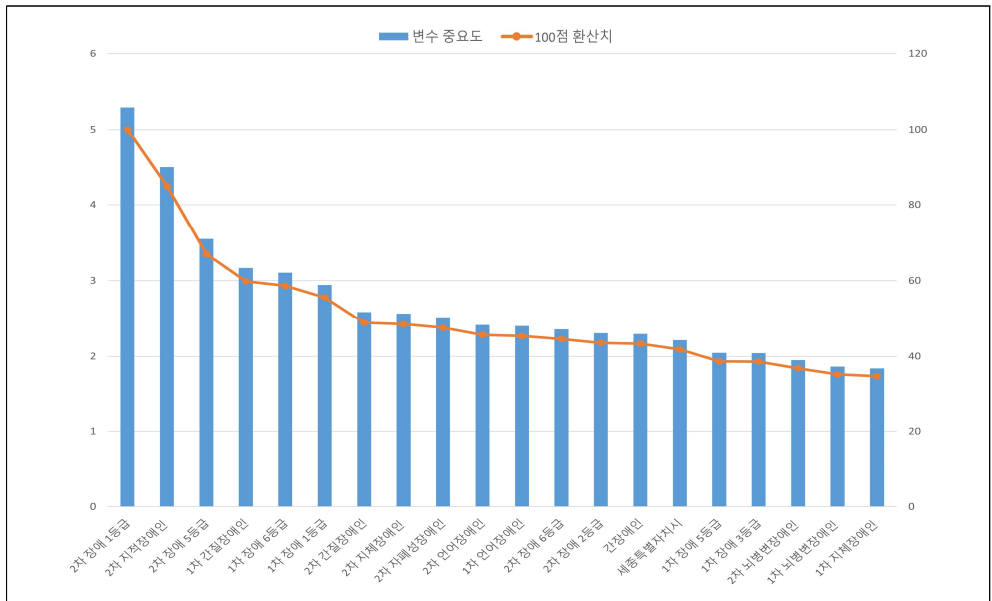
순위	NN	RF	GBM	Xgboost	Logistic Regression
1	2차 장애 1등급	장기요양경감구분_저소득	장기요양경감구분_저소득	장기요양경감구분_저소득	2차 장애 1등급
2	주수발자_저녀	주수발자_머느리,사위	ALBUMIN	인종조사점수	연령
3	대구광역시	장기요양경감구분_건강보험	인종조사인정점수	ALBUMIN	2차 장애 2등급
4	2차 장애 2등급	ALBUMIN	연령	연령	ALBUMIN
5	장기요양경감구분_저소득	연령	주수발자_머느리, 사위	주수발자_머느리,사위	수발자(머느리, 사위포함)
6	동거인_부모	인종조사점수	장기요양경감구분_건강보험	CR	인종조사인정점수
7	장기요양경감구분_값없음	2차 장애 2등급	CR	HB	장기요양경감구분_건강보험
8	2차 장애 3등급	2차 장애 5등급	2차 장애 1등급	장기요양경감구분_건강보험	뇌출혈(+)
9	1차 청각장애인	2차 장애 3등급	주수발자_저녀	2차 장애 1등급	L_tube_feeding(+)
10	1차 지적장애인	주수발자_저녀	공복 혈당 (GLUCOSE_AC)	BUN	성별_여자
11	경상남도	2차 장애 1등급	2차 장애 2등급	공복 혈당 (GLUCOSE_AC)	약행겨머기_부분의존
12	의료급여	CR	BUN	2차 장애 2등급	장기요양경감구분_저소득
13	노인요양동생활가정	HB	2차 장애 5등급	HBA1C	치매_완전의존
14	2차 청각장애인	간호환산점수	총입원기간	총입원기간	multifocal_large(+)
15	2차 장애 5등급	등급판정인정등급코드_5등급	2차 장애 3등급	주수발자_저녀	약행겨머기_완전의존
16	장기요양경감구분_저소득	2차 장애 4등급	HB	총합진횟수	장기요양대상자구분코드
17	장기요양경감구분_만성질환	주수발자_시설관계자	간호환산점수	일반병동 입원기간	장기요양경감구분_값없음
18	2차 장애 6등급	2차 장애 6등급	L_tube_feeding,1	간호환산점수	폐렴(+)
19	2차 장애 4등급	등급판정인정등급코드_4등급	당화혈색소(HBA1C)	2차 장애 5등급	등급판정인정등급코드_값없음
20	주수발자_머느리, 사위	장기요양경감구분_만성질환	주수발자_시설관계자	인지환산점수	Temporal lobe(+)

## 제2절 노인 장기요양보험 서비스 미이용자군의 기계학습(Machine Learning) 분석 결과

NN 분석에 관련된 코드와 결과 값은 appendix 246p에 있으며, 147개 변수 중 상위 20개 주요 항목은 다음과 같다 <표 3-20>, [그림 3-8].

<표 3-20> 장기요양보험 미이용자군의 사망에 관련한 NN 기계 학습 주요 인자

상위 20개 주요 항목	변수 중요도	100점 환산 점수
2차 장애 1등급	5.292	100
2차 지적장애인	4.497	84.97
2차 장애 5등급	3.545	66.99
1차 간질장애인	3.168	59.86
1차 장애 6등급	3.104	58.65
1차 장애 1등급	2.943	55.6
2차 간질장애인	2.588	48.9
2차 지체장애인	2.569	48.53
2차 자폐성장애인	2.518	47.58
2차 언어장애인	2.417	45.67
1차 언어장애인	2.402	45.37
2차 장애 6등급	2.36	44.59
2차 장애 2등급	2.306	43.57
간장애인	2.296	43.37
세종특별자치시	2.214	41.83
1차 장애 5등급	2.047	38.67
1차 장애 3등급	2.045	38.63
2차 뇌병변장애인	1.951	36.86
1차 뇌병변장애인	1.865	35.24
1차 지체장애인	1.84	34.76



[그림 3-8] 노인 장기요양보험 서비스 미이용자군의 사망에 관련한 NN 기계 학습 주요 인자

정확도는 0.8711, 95% 신뢰구간(CI)은 (0.8470, 0.8927), No Information Rate는 0.7399로 P-Value [Acc > NIR]: <2.2e-16 로 관찰되었다.

민감도는 0.9362, 특이도는 0.6858, 양성 예측도는 0.8945, 음성 예측도는 0.7908, Balanced Accuracy인 AUC 값은 0.8110이었다 <표 3-21>.

<표 3-21> NN의 실험 세트 검증

Prediction	생존	사망
생존	602	71
사망	41	155

SVM 분석에 관련된 코드와 결과 값은 appendix 241p에 있으며 정확도는 0.8884, 95% 신뢰구간(CI)은 (0.8665, 0.9085), No Information Rate는 0.7339로 P-Value [Acc > NIR]: <2.2e-16 로 관찰되었다 <표 3-22>.

<표 3-22> SVM의 실험 세트 검증

Prediction	생존	사망
생존	612	66
사망	31	150

KNN 분석에 관련된 코드와 결과 값은 appendix 243p에 있으며 정확도는 0.8354, 95% 신뢰구간(CI)은 (0.8091, 0.8595), No Information Rate는 0.5235로 P-Value [Acc > NIR]: 1.144e-11로 관찰되었다.

민감도는 0.9673, 특이도는 0.4602, 양성 예측도는 0.8360, 음성 예측도는 0.8320, Balanced Accuracy인 AUC 값은 0.7138이었다 <표 3-23>.

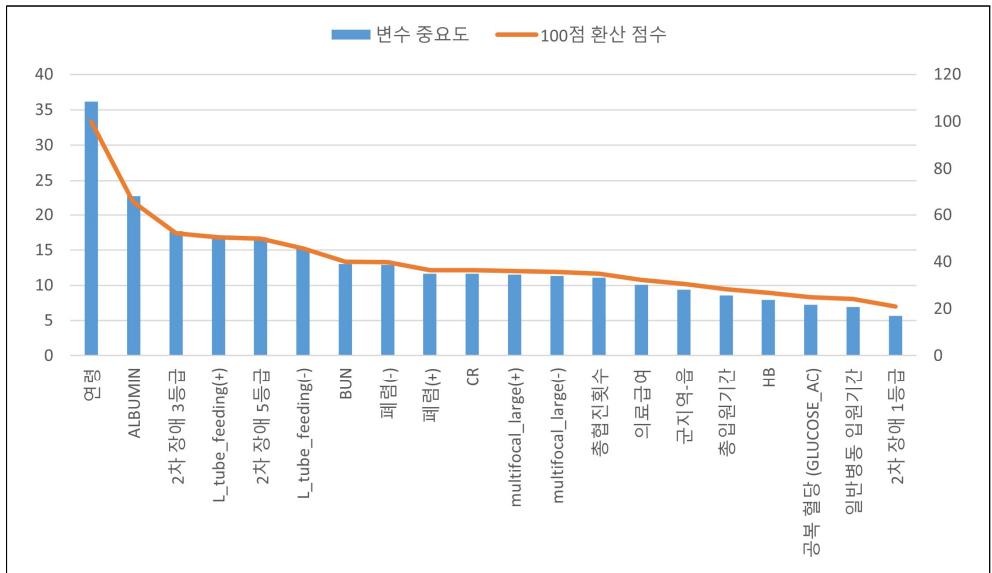
<표 3-23> KNN의 실험 세트 검증

Prediction	생존	사망
생존	622	122
사망	21	104

RF분석에 관련된 코드와 결과 값은 appendix 247p에 있으며, 237개 변수 중 상위 20개 주요 항목은 다음과 같다 <표 3-24>, [그림 3-9].

<표 3-24> 노인 장기요양보험 서비스 미이용자군의 생존 분석의 RF기계 학습 주요 인자

주요 인자	변수 중요도	100점 환산 점수
연령	36.164	100
ALBUMIN	22.784	65.32
2차 장애 3등급	17.727	52.21
L_tube_feeding(+)	17.064	50.49
2차 장애 5등급	16.814	49.84
L_tube_feeding(-)	15.204	45.67
BUN	13.06	40.12
폐렴(-)	12.935	39.79
폐렴(+)	11.672	36.52
CR	11.674	36.52
multifocal_large(+)	11.52	36.12
multifocal_large(-)	11.352	35.69
총혈진횟수	11.087	35
의료급여	10.067	32.36
군지역-읍	9.378	30.57
총입원기간	8.536	28.39
HB	7.912	26.77
공복 혈당 (GLUCOSE_AC)	7.218	24.97
일반병동 입원기간	6.912	24.18
2차 장애 1등급	5.677	20.98



[그림 3-9] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 RF기계 학습 주요 인자

정확도는 0.8631, 95% 신뢰구간(CI)은 (0.8384, 0.88852), No Information Rate는 0.7399로 P-Value [Acc > NIR]는 <math>2.2e-16</math>로 관찰되었다.

민감도는 0.9533, 특이도는 0.6062, 양성 예측도는 0.8732, 음성 예측도는 0.8204, Balanced Accuracy인 AUC 값은 0.7798이었다 <표 3-25>.

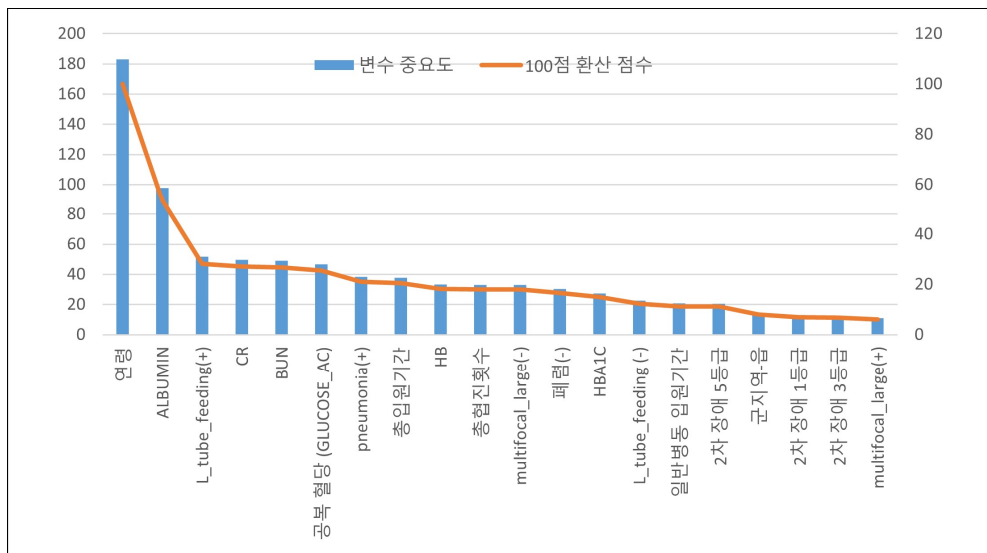
<표 3-25> RF의 실험 세트 검증

Prediction	생존	사망
생존	613	89
사망	30	137

GBM 분석에 관련된 코드와 결과 값은 appendix 251p에 있으며, 147개 변수중 상위 20개 주요 항목은 다음과 같다 <표 3-26>, [그림 3-10].

<표 3-26> 노인 장기요양보험 서비스 미이용자군의 생존 분석의 GBM 기계 학습 주요 인자

주요인자	변수 중요도	100점 환산 점수
연령	183.1	100
ALBUMIN	97.81	53.422
L_tube_feeding(+)	51.72	28.245
CR	49.63	27.108
BUN	48.91	26.715
공복 혈당 (GLUCOSE_AC)	46.53	25.415
폐렴(+)	38.41	20.98
총입원기간	37.67	20.571
HB	33.21	18.141
총흡진횟수	32.96	18.003
multifocal_large(-)	32.86	17.944
폐렴(-)	30.21	16.501
HBA1C	27.45	14.995
L_tube_feeding (-)	22.59	12.336
일반병동 입원기간	20.66	11.281
2차 장애 5등급	20.49	11.19
군지역-읍	14.54	7.941
2차 장애 1등급	12.65	6.907
2차 장애 3등급	12.44	6.793
multifocal_large(+)	11.03	6.024



[그림 3-10] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 GBM기계 학습 주요 인자

정확도는 0.8734, 95% 신뢰구간(CI)은 (0.8495, 0.8948), No Information Rate는 0.7399로 P-Value [Acc > NIR]는 <2.2e-16로 관찰되었다.

민감도는 0.9471, 특이도는 0.6637, 양성 예측도는 0.8891, 음성 예측도는 0.6637, Balanced Accuracy인 AUC 값은 0.8054이었다 <표 3-27>.

<표 3-27> GBM의 실험 세트 검증

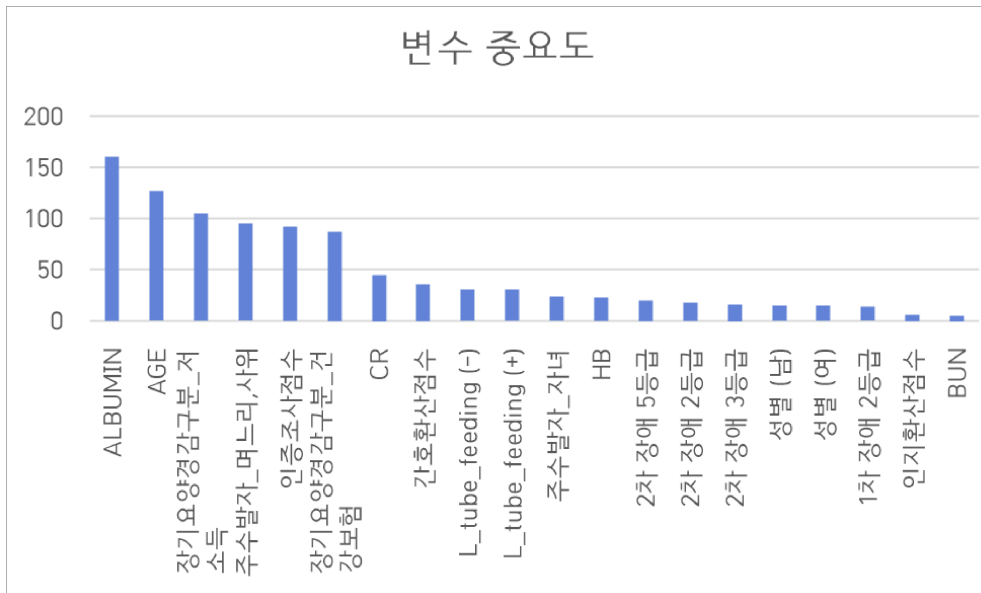
Prediction	생존	사망
생존	609	76
사망	34	150

r\_part 분석에 관련된 코드와 결과 값은 appendix 259p에 있으며, 147개 변수중 상위 20개 주요 항목은 다음과 같다 <표 3-28>, [그림 3-11].

<표 3-28> 노인 장기요양보험 서비스 미이용자군의 생존 분석의 r\_part 기계 학습 주요 인자

상위 20개 주요 항목	변수 중요도
ALBUMIN	160.554024
AGE	126.330119
장기요양경감구분_저소득	104.719361
주수발자_머느리,사위	94.677169
인증조사점수	92.16592
장기요양경감구분_건강보험	87.480856
CR	44.932035
간호환산점수	35.955768
L_tube_feeding (-)	30.430029
L_tube_feeding (+)	30.430029
주수발자_자녀	23.318628
HB	22.574412
2차장애5등급	19.850712
2차장애2등급	17.761503
2차장애3등급	16.162466
성별 (남)	14.417745
성별 (여)	14.417745
1차 장애 2등급	13.31
인지환산점수	6.022037
BUN	5.29





[그림 3-11] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 r-part 기계 학습 주요 인자

r-part의 실험 세트 검증의 정확도는 0.8124, 95% 신뢰구간(CI)은 (0.7849, 0.8379), No Information Rate는 0.7399로 P-Value [Acc > NIR]는 2.925e-07로 관찰되었다.

민감도는 0.9720, 특이도는 0.3584, 양성 예측도는 0.8117, 음성 예측도는 0.8182, Balanced Accuracy인 AUC 값은 0.6652이었다 <표 3-29>.

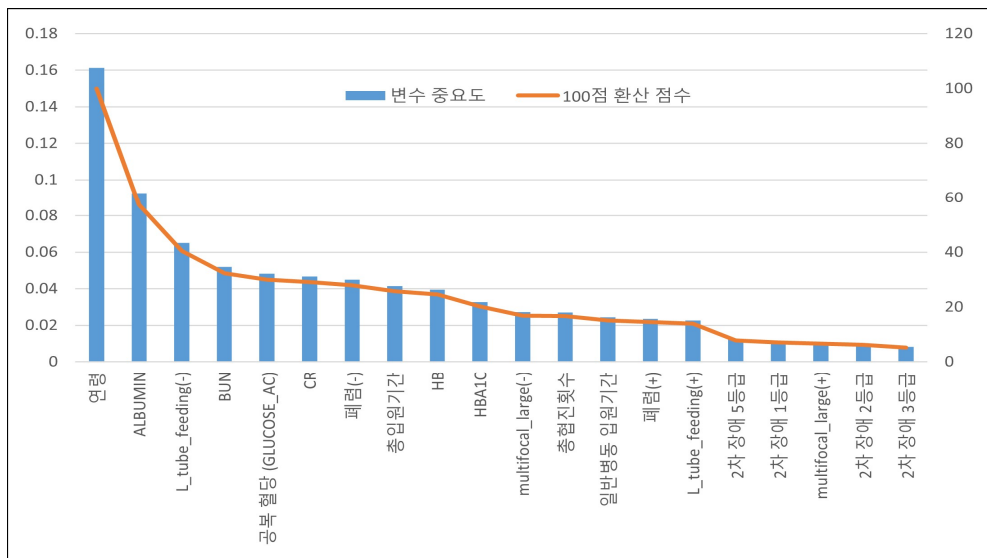
<표 3-29> r-Part의 실험 세트 검증

Prediction	생존	사망
생존	625	145
사망	18	81

Xgboost 분석에 관련된 코드와 결과 값은 appendix 263p에 있으며, 147개 변수중 상위 20개 주요 항목은 다음과 같다 <표 3-30>, [그림 3-12].

<표 3-30> 노인 장기요양보험 서비스 미이용자군의 생존 분석의 Xgboost 기계 학습 주요 인자

상위 20개 주요 항목	변수 중요도	100점 환산 점수
연령	0.161187	100
ALBUMIN	0.092533	57.407
L_tube_feeding(-)	0.065175	40.434
BUN	0.052007	32.265
공복 혈당 (GLUCOSE_AC)	0.048241	29.929
CR	0.046796	29.032
폐렴(-)	0.044998	27.916
총입원기간	0.041538	25.77
HB	0.039556	24.541
HBA1C	0.032675	20.272
multifocal_large(-)	0.027154	16.846
총흡진횟수	0.026928	16.706
일반병동 입원기간	0.024364	15.115
폐렴(+)	0.023461	14.555
L_tube_feeding(+)	0.022578	14.007
2차 장애 5등급	0.0126	7.817
2차 장애 1등급	0.011366	7.052
multifocal_large(+)	0.010809	6.706
2차 장애 2등급	0.009946	6.17
2차 장애 3등급	0.008289	5.143



[그림 3-12] 노인 장기요양보험 서비스 미이용자군의 생존 분석의 Xgboost 기계 학습 주요 인자

Xgboost의 실험 세트 검증의 정확도는 0.8826, 95% 신뢰구간(CI)은 (0.8593, 0.6593), No Information Rate는 0.7399로 P-Value [Acc > NIR]는 <math>2.2e-16</math>로 관찰되었다.

민감도는 0.9611, 특이도는 0.6593, 양성 예측도는 0.8892, 음성 예측도는 0.8563, Balanced Accuracy인 AUC 값은 0.8102이었다 <표 3-31>.

<표 3-31> Xgboost의 실험 세트 검증

Prediction	생존	사망
생존	678	77
사망	25	149

노인 장기요양 미이용자군의 사망 예측 기계학습 분석 알고리즘 사이에서도, 노인 장기요양보험 서비스 이용자군의 사망 예측 모델처럼, Ensemble model 분석에서 가장 최적의 모델은 Random Forest로 밝혀졌다 <표 3-32>, (Appendix 267p).

<표 3-32> 노인 장기요양보험 서비스 미이용자군의 생존 분석에서 기계학습 방법에 따른 정확도와 AUC value

	Neural Network (NN)	Support Vector Machine (SVM)	KNN	RF	gbm	Xgboost
Accuracy	0.8711	0.9518	0.8354	0.9533	0.9471	0.9611
AUC	0.8110	0.8299	0.7138	0.7798	0.8054	0.8102

노인장기 요양보험 미이용자군의 생존 회귀 분석에서도 기계학습 분석과 유사한 결과가 관찰되었다. 노인 장기요양보험 서비스 이용자군과는 달리 병원내 상태의 반영하는 항목인 폐렴, 뇌손상 부위 multifocal\_large, IVH(+), Coil (+), Aneurysm(+), brain stem (+)이 주요한 항목으로 관찰되었다. 또한 생존, 1차 장애등급이 많이 관찰되었으며, 보통 1차 장애 등급을 신청후 2년이 지나면 갱신하게 되는데, 장애등록 이후 추가 갱신을 하지 못해서 이런 결과가 나온 것으로 보인다. 기타 연령과 성별과 같은 인구학적인 인자와 albumin, L-tube feeding과 같은 영양학적 지표는 노인 장기요양보험 서비스 이용자군 및 다른 기계학습 알고리즘의 분석과 유사하게 관찰되었다 <표 3-33>.

<표 3-33> 전체 노인 장기요양 서비스 미이용군의 생존 회귀 분석에서 상위 20개 항목

항목	Estimate	standard error	t value	r(> t )
연령	0.00618	0.00052	11.896	2.00E-16***
2차 장애 1등급	-0.64161	0.074159	-8.652	2.00E-16***
L_tube_feeding(+)	0.220334	0.019885	11.08	2.00E-16***
폐렴(+)	0.123612	0.013092	9.442	2.00E-16***
multifocal_large(+)	0.126946	0.016021	7.924	3.25E-15***
ALBUMIN	-0.11116	0.01434	-7.752	1.24E-14***
2차 장애 2등급	-0.42523	0.065071	-6.535	7.50E-11***
군지역-읍	0.190125	0.033843	5.618	2.12E-08***
IVH(+)	0.152174	0.027127	5.61	2.22E-08***
HB	-0.01974	0.003718	-5.311	1.18E-07***
Coil(+)	-0.14613	0.028835	-5.068	4.28E-07***
성별_여자	-0.06128	0.013963	-4.389	1.18E-05***
공복 혈당 (GLUCOSE_AC)	0.000457	0.000122	3.749	0.000181***
1차 장애 1등급	0.853372	0.242499	3.519	0.00044***
Aneurysm(+)	-0.052	0.015683	-3.316	0.000926***
호흡기 협진	0.026637	0.008101	3.288	0.001021**
1차 장애 3등급	0.697349	0.241303	2.89	0.003882**
1차 장애 2등급	0.644658	0.237908	2.71	0.006774**
brain_stem(+)	0.059964	0.022802	2.63	0.008588**
광역시_읍	0.052078	0.021034	2.476	1.33E-02*

여러 기계 분석 알고리즘과 회귀 분석 간에 상위 20개 변수 중요도 간의 인자의 비교는 다음과 같다. 모든 분석에서 공통으로 나타나는 항목은 2차 장애 1등급, 4개 이상 알고리즘에서 공통으로 관찰되는 항목은 연령, Albumin, multifocal\_large(+), 2차 장애 5등급, L\_tube\_feeding(+), 폐렴(+), 공복혈당 (AC\_glucose)가 있었다. 각 알고리즘 간의 순위의 차이는 관찰되지만, 주요 인자는 비슷하게 관찰되었지만 NN에서는 장애등급과 장애종류가 주요 상위 20개 항목이었다 <표 3-34>.

<표 3-34> 전체 노인 장기요양 서비스 미이용자군의 기계학습 알고리즘 분석에 따른 주요 인자의 비교

순위	NN	RF	GBM	Xgboost	Logistic Regression
1	2차 장애 1등급	연령	연령	연령	연령
2	2차 지적장애인	ALBUMIN	ALBUMIN	ALBUMIN	2차 장애 1등급
3	2차 장애 5등급	2차 장애 3등급	L_tube_feeding(+)	L_tube_feeding(-)	L_tube_feeding(+)
4	1차 간질장애인	L_tube_feeding(+)	CR	BUN	폐렴(+)
5	1차 장애 6등급	2차 장애 5등급	BUN	공복 혈당 (GLUCOSE_AC)	multifocal_large(+)
6	1차 장애 1등급	L_tube_feeding(-)	공복 혈당 (GLUCOSE_AC)	CR	ALBUMIN
7	2차 간질장애인	BUN	폐렴(+)	폐렴(-)	2차 장애 2등급
8	2차 지적장애인	폐렴(-)	총임원기간	총임원기간	군지역-음
9	2차 자폐성장애인	폐렴(+)	HB	HB	IVH(+)
10	2차 언어장애인	CR	총합진횟수	HBA1C	HB
11	1차 언어장애인	multifocal_large(+)	multifocal_large(-)	multifocal_large(-)	Coil(+)
12	2차 장애 6등급	multifocal_large(-)	폐렴(-)	총합진횟수	성별_여자
13	2차 장애 2등급	총합진횟수	HBA1C	일반병동 입원기간	공복 혈당 (GLUCOSE_AC)
14	간장애인	의료급여	L_tube_feeding (-)	폐렴(+)	1차 장애 1등급
15	세종특별자치시	군지역-음	일반병동 입원기간	L_tube_feeding(+)	Aneurysm(+)
16	1차 장애 5등급	총임원기간	2차 장애 5등급	2차 장애 5등급	호흡기 협진
17	1차 장애 3등급	HB	군지역-음	2차 장애 1등급	1차 장애 3등급
18	2차 뇌병변장애인	공복 혈당 (GLUCOSE_AC)	2차 장애 1등급	multifocal_large(+)	1차 장애 2등급
19	1차 뇌병변장애인	일반병동 입원기간	2차 장애 3등급	2차 장애 2등급	brain_stem(+)
20	1차 지적장애인	2차 장애 1등급	multifocal_large(+)	2차 장애 3등급	광역시-음

# 제4장

## 결론



## 제4장 결론

이번 연구는 병원 자료와 공단자료를 결합을 통한 연구로 그 의의를 둘 수 있다. 그러나, 일산병원 환자 중 상당수는 건강 검진을 받지 않아, 건강검진 자료는 결측 치가 너무 많아서 결합할 수가 없었다.

기계학습 분석관련 연구는 데이터 수집과 전처리 및 정제에 대부분의 시간의 소요된다. 대부분의 자연어 처리 과정이, 각 언어의 구조에 맞게 구조화되어 있으나, 우리나라의 의무기록은 한국어와 의학 영어와 중복된 형태로 되어 있어, 의학자료에 맞는 자연어 처리 모듈이 개발되지 않아, R-package를 이용, 영어 단어와 한글 단어를 하나 하나를 구조화하는데 애로 사항이 있었다. 또한 다양한 변수, 오타자 처리가 문제가 될 수 있다. 가령 예를 들어 판독지에 오타자가 “MCA infarction” 오자가 있을 경우(r과 c가 바뀌게 될 경우), 주요 단어를 “MCA”와 “infarction”으로 설정하면 “MCA”영역만 설정이 되고, “infarction”은 빠지게 된다. 다음으로, 혈액 검사 및 흉부 X-선 검사와 같이 반복되는 검사 결과의 정제, 결측치 처리, 부정확한 자료(흡연력, 음주력)에 대한 정의에 대해서는 문제점으로 지적될 수 있다. 예를 들어 혈액 검사는 입원 기간 중에 여러 번 시행하게 되는데, 어느 시기의 검사를 한 것을 대표값으로 취할 것에 대해서 논란의 소지가 있다.

이 연구가 공단 서버에서 운용을 하다 보니, R-version이 2018년도 version으로 update 되지 않아, 새로운 R-package 설치를 하지 못하였다. 이제까지는 각 알고리즘에 맞는 package를 설치하여, 각 package에 맞추어 코딩하고 분석하였지만, 최근에 개발된, H<sub>2</sub>O를 위시한 machine learning automatic analysis package를 이용하면, 다양한 기계학습 알고리즘을 한 번에 처리가 가능하여, 시간을 단축시킬 수 있으나, 이를 이용할 수가 없었다. 실례로, H<sub>2</sub>O를 package를 설치는 되었으나, 동반 package는 2018년도 R-version에는 설치가 되지 않아 분석을 할 수 없었다. 또한 여러 명의 연구자가 동시에 서버에 접속을 할 때는 Xgboost 분석에서는 무려 2주나 소모되어, 자료 오류를 수정해서 다시 돌릴 때 또한 2주 이상 소모되어, 추가 분석하는데 시간적 여유를 확보하기가 어려웠다.



기존의 기계 학습 방식이 black box model로 데이터 간의 원인과 결과를 설명할 수가 없었지만, 최근 개발된 causal modeling learning (Athey, S., 2019)과 causal interpretation methods (Schwab, 2019)가 개발되어 이러한 기계학습의 발전은 정확한 원인 인자를 밝혀내어 진보된 예측 모델링이 가능할 것으로 생각된다. 이에 맞추어 최근 개발된 r-package로는 generalized random forest, light gbm package가 있으나, 이를 설치하지 못하여 추가 분석은 이루어지지 않았다.

노인 장기요양보험 서비스 이용군 및 미이용군, 모두 Xgboost의 값이 정확도 및 AUC 값이 높았으나, ensemble 분석에서는 random forest가 최적의 예측 모델로 선정되었다. 하지만 각 인공지능 알고리즘 모델에 따라서 상위 20개 주요 인자가 순위에서만 차이나고 거의 유사한 양상이 관찰되었으며, 기존의 통계분석 방법인 logistic regression에서도 비슷한 결과가 관찰되었다. RF분석에서 주요 인자를 살펴보면, 장기요양경감 구분, 두번째 제출한 장애등급, 장기요양 판정 등급, 내과적 상태를 반영하는 Albumin, HB, 연령이 주요 변수였으며, 노인 장기요양보험 서비스 미이용자군의 경우도 비슷하나, 이는 이용자군과는 달리, L-tube feeding, 폐렴여부, 뇌병변의 상태, 총입원기간, 일반병동 입원기간, 협진 횟수등, 병원의 자료가 주요 변수였다. 특히 주목하여야 할 것은 연령과 albumin, HB, CR이 노인 장기요양보험 이용자군 및 미이용자군 모두에 내과적 질환 및 영양상태가 뇌손상 이후 사망을 예측하는 인자임을 확인할 수 있었다.

65세 이상 퇴행성 뇌질환 환자라도 병원에 입원한 상태에서는 서비스를 제공받지 못한다. 따라서 노인 장기요양보험 서비스를 받지 못한다는 것은 병원에 입원할 정도로 상태가 좋지 못하다는 것을 의미한다. 즉 노인 장기요양보험 서비스 미이용자군들은 인구학적 구성에서, 65세 미만이고, 퇴행성 뇌손상이 아닌 외상성 뇌손상일 경우, 혹은 장애정도가 경미한 경우나 혹은 뇌손상 이후 조기에 사망하거나, 내과적, 신경학적 상태가 좋지 않아 병원에 장기간 입원, 요양병원에 장기간 입원한 경우로 여겨진다. 한편 장기요양소견서 비용 경감 구분이 사망 예측의 주요 인자로 지정되는 것으로 볼 때, 사회경제적 상태 또한 뇌손상 이후 사망 예측에 중요한 인자임을 확인할 수가 있었다.

추가 연구에서는 65세를 기준으로 하여, 65세 이상 노인 장기요양보험 서비스를 받지 못한 군과 65세 미만 노인 장기요양보험 서비스 받지 못한군에 대한 추가 분석과 예측 모델을 구할 필요가 있다. 또한 병원내 입원 당시, 자료 분석에 대한 합의도 필요하며, 무엇보다도 한국어와 영어가 혼재된 한국의 의무기록에서 자연어 처리 모듈 개발이 환자의 기능적 수준에 대한 평가를 포함하지 않아, 의무기록의 자연어 처리 모듈이 필요하며,

병원 및 건강보험 자료 결합일 경우에는 병원내 자료를 좀 더 정밀하게 추출한 다음, 공단 자료를 이용하여 시계열적 자료를 구축한 다음, 예후 예측 모델에 관한 연구를 진행하는 것이 필요하다.

국민건강보험 일산병원의 단일 자료로 예후 예측한 것이므로, 사회 경제적 조건이 제한되어 대표성에 관하여 단점으로 지적할 수 있다. 따라서 추후에는 전체 지역을 대표할 여러 병원의 자료를 동일한 기준으로 자료를 획득하여, 공단 자료를 결합한다면, 이러한 단점을 극복할 수 있을 것으로 생각된다.



참고문헌



## 참고문헌

1. 김준엽, 배희준, 박종무. “뇌졸중 역학보고서, 2018.” 주간 건강과 질병 12.43 (2019): 1845-1860.
2. 대한재활의학회. “재활의학.” 군자출판사(2020).
3. 서지영. “딥러닝 파이토치 교과서.” 길벗(2022).
4. 의사를 위한 실전 인공지능. 군자출판사.
5. <https://davincilabs.ai/blog/?q=YToxOntzOjEyOjEjZl3b3JkX3R5cGUiO3M6MzoiYWxsIjt9&bmode=view&idx=10608836&t=board>.
6. 국민건강보험 빅데이터를 이용한 퇴행성 뇌질환(뇌졸중, 치매, 파킨슨) 진단 이후의 장기요양으로 전달되는 의료체계 개선 연구 (2009).
7. Lindsay, M. Patrice, et al. “World Stroke Organization (WSO): global stroke fact sheet 2019.” (2019): 806-817.
8. Kim, Jeoung Kun, Yoo Jin Choo, and Min Cheol Chang. “Prediction of motor function in stroke patients using machine learning algorithm: Development of practical models.” Journal of Stroke and Cerebrovascular Diseases 30,8 (2021): 105856.
9. Rau, Cheng-Shyuan, et al. “Mortality prediction in patients with isolated moderate and severe traumatic brain injury using machine learning models.” PloS one 13.11 (2018): e0207192.
10. Ge, Yanqiu, et al. “Predicting post-stroke pneumonia using deep neural network approaches.” International journal of medical informatics 132 (2019): 103986.
11. Gu, Yiwen, et al. “A machine learning approach for predicting post-stroke aphasia recovery: A pilot study.” Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments. 2020.
12. Moraffah, R., Karami, M., Guo, R., Raglin, A., & Liu, H. (2020). Causal interpretability for machine learning-problems, methods and evaluation. ACM SIGKDD Explorations Newsletter, 22(1), 18-33.
13. Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. The Annals of Statistics, 47(2), 1148-1178.

14. Schwab, P., & Karlen, W. (2019). Explain: Causal explanations for model interpretation under uncertainty. *Advances in Neural Information Processing Systems*, 32.
15. Shameer, K., Johnson, K. W., Glicksberg, B. S., Dudley, J. T., & Sengupta, P. P. (2018). Machine learning in cardiovascular medicine: are we there yet?. *Heart*, 104(14), 1156-1164.
16. Ntaios, G.; Faouzi, M.; Ferrari, J.; et al. An integer-based score to predict functional outcome in acute ischemic stroke: the ASTRAL score. *Neurology* 2012, 78, 1916-1922.
17. Heo, J.; Yoon, J.G.; Park, H.; et al. Machine Learning-Based Model for Prediction of Outcomes in Acute Stroke. *Stroke* 2019, 50, 1263-1265.
18. Jang, S.K.; Chang, J.Y.; Lee, J.S.; et al. Reliability and Clinical Utility of Machine Learning to Predict Stroke Prognosis: Comparison with Logistic Regression. *J Stroke* 2020, 22, 403-406.
19. Park, D.; Jeong, E.; Kim, H.; et al. Machine Learning-Based Three-Month Outcome Prediction in Acute Ischemic Stroke: A Single Cerebrovascular-Specialty Hospital Study in South Korea. *Diagnostics* 2021, 11, 1909.
20. Alaka, S.A.; Menon, B.K.; Brobbey, A.; et al. Functional Outcome Prediction in Ischemic Stroke: A Comparison of Machine Learning Algorithms and Regression Models. *Front Neurol* 2020, 11, 889.
21. Fang, G.; Huang, Z.; Wang, Z. Predicting ischemic stroke outcome using deep learning approaches. *Front Genet* 2022, 12, 827522.
22. Ge, Y.; Wang, Q.; Wang, L. Predicting post-stroke pneumonias using deep neural network approaches. *Int J Med Inform.* 2019, 132, 103986.
23. Gu, Y.; Bahrani, M.; Billot, A.; et al. A machine learning approach for predicting post-stroke aphasia recovery: A pilot study. *Petra*. June 30-July 3, 2020. Corfu, Greece.
24. Nourlahi, M.; Dadboud, F.; Khalili, H.; et al. A machine learning model for predicting favorable outcome in severe traumatic brain injury patients after 9 months. *Acute Crit Care*. 2022, 37, 45-52.
25. Pease, M.; Arefan, D.; Barber, J; et al. Outcome prediction in patients with severe traumatic brain injury using deep learning from head CT scans. *Radiology*. 2022, 304, 385-394.

26. Brushetta, R.; Tartarisco, G.; Lucca, L.F.; et al. Predicting outcome of traumatic brain injury: Is machine learning the best way? *Biomedicines*, 2022, 10, 686.
27. Athey, S., Tibshirani, J., & Wager, S. (2019). Generalized random forests. *The Annals of Statistics*, 47(2), 1148-1178.
28. Schwab, P., & Karlen, W. (2019). Cxplain: Causal explanations for model interpretation under uncertainty. *Advances in Neural Information Processing Systems*, 32.





# 부 록





```
Sessioninfo()
R version 3.5.2 (2018-12-20)
Platform: powerpc64le-unknown-linux-gnu (64-bit)
Running under: Ubuntu 16.04.4 LTS

Matrix products: default
BLAS: /usr/local/lib/R/lib/libRblas.so
LAPACK: /usr/local/lib/R/lib/libRlapack.so

locale:
[1] LC_CTYPE=en_US.UTF-8      LC_NUMERIC=C
     LC_TIME=en_US.UTF-8     LC_COLLATE=en_US.UTF-8
     LC_MONETARY=en_US.UTF-8
[6] LC_MESSAGES=C            LC_PAPER=en_US.UTF-8
     LC_NAME=C              LC_ADDRESS=C
     LC_TELEPHONE=C
[11] LC_MEASUREMENT=en_US.UTF-8 LC_IDENTIFICATION=C

attached base packages:
[1] parallel stats graphics grDevices utils datasets
     methods base

other attached packages:
[1] h2o_3.24.0.1      xgboost_0.71.2    forcats_0.4.0
     stringr_1.4.0    purrr_0.3.1       readr_1.3.1
     tidyr_0.8.3
[8] tibble_2.0.1      tidyverse_1.2.1   randomForest_4.6-14
     kernlab_0.9-27   gbm_2.1.5         e1071_1.7-0.1
     dplyr_0.8.0.1
[15] data.table_1.12.0 doParallel_1.0.14 iterators_1.0.10
     foreach_1.4.4   caretEnsemble_2.0.0 caret_6.0-81
     ggplot2_3.1.0
[22] lattice_0.20-38   moonBook_0.2.3    lubridate_1.7.4
     ipred_0.9-8      nnet_7.3-12       rpart_4.1-13
     class_7.3-15
```

```

[29] lava_1.6.4          survival_2.43-3     haven_2.1.0
loaded via a namespace (and not attached):
[1] httr_1.4.0          jsonlite_1.6        splines_3.5.2
     prodlim_2018.04.18 modelr_0.1.4         assertthat_0.2.0
     stats4_3.5.2
[8] cellranger_1.1.0    yaml_2.2.0          pillar_1.3.1
     backports_1.1.3    glue_1.3.0          digest_0.6.18
     rvest_0.3.2
[15] snakecase_0.9.2     colorspace_1.4-0    recipes_0.1.4
     Matrix_1.2-16     plyr_1.8.4          timeDate_3043.102
     pkgconfig_2.0.2
[22] broom_0.5.1         scales_1.0.0        stringdist_0.9.5.1
     gower_0.1.2        generics_0.0.2      sjlabelled_1.0.16
     withr_2.1.2
[29] pbapply_1.3-4       lazyeval_0.2.1      cli_1.0.1
     magrittr_1.5        crayon_1.3.4        readxl_1.3.0
     nlme_3.1-137
[36] MASS_7.3-51.1       xml2_1.2.0          tools_3.5.2
     hms_0.4.2          munsell_0.5.0       prediction_0.3.6.1
     compiler_3.5.2
[43] rlang_0.3.3         RCurl_1.95-4.12     grid_3.5.2
     rstudioapi_0.9.0  bitops_1.0-6        gtable_0.2.0
     ModelMetrics_1.2.2
[50] codetools_0.2-15    reshape2_1.4.3      sjmisc_2.7.7
     R6_2.4.0          gridExtra_2.3       stringi_1.3.1
     Rcpp_1.0.1

```

Data preprocessing code

```

# prerequisite
library(haven)
library(survival)
library(lava)
library(class)
library(rpart)
library(nnet)

```

```

library(ipred)
library(lubridate)
library(dplyr)
library(moonBook)
# machine_Learning_library
library(caret)
library(caretEnsemble)
library(doParallel)
library(data.table)
library(dplyr)
library(e1071)
library(gbm)
library(kernlab)
library(randomForest)
library(tidyverse)
library(xgboost)
library(h2o)

path="~/data/data_source/user_data/"
spath=paste0(path, "handled_data/")

# data_fusion (자격변수)

final_data=read.csv(paste0(spath,"final_data.csv"))

...

mode(final_data$MDCARE_STRT_DT)
final_data$MDCARE_STRT_DT=as.numeric(final_data$MDCARE_STRT_DT)

mode(final_data$DTH_ASSMD_DT)

x=data.frame(final_data$DTH_ASSMD_DT)
ncol(x)
sum(!is.na(x))

```

```

final_data$Live_days=ymd(final_data$DTH_ASSMD_DT)-ymd(final_data$
MDCARE_STRT_DT)

data_2$data_2_min=ymd(data_2$MDCARE_STRT_DT)-ymd(data_2$HME_DT)

#final_data_1=final_data %>% filter(final_data$Ipwon==0)
final_data_2=final_data %>% filter(final_data$Ipwon==1)
#final_data_3=final_data %>% filter(final_data$Ipwon==1 &
final_data$LTC_YN==1)
#final_data_4=final_data %>% filter(final_data$Ipwon==1 &
final_data$LTC_YN==0)
#final_data_5=final_data %>% filter(final_data$LTC_YN==0)
#final_data_6=final_data %>% filter(final_data$LTC_YN==1)

#dim(final_data_1)
dim(final_data_2)
#dim(final_data_3)
#dim(final_data_4)
#dim(final_data_5)
#dim(final_data_6)

#table(is.na(final_data_1))
table(is.na(final_data_2))
#table(is.na(final_data_3))
#table(is.na(final_data_4))
#table(is.na(final_data_5))
#table(is.na(final_data_6))

final_data_2$ALBUMIN=as.numeric(final_data_2$ALBUMIN)
final_data_2$BUN=as.numeric(final_data_2$BUN)

#mean(final_data_2$ALBUMIN, na.rm = T)

mean(final_data_2$BUN, na.rm = T)
final_data_2$BUN=ifelse(is.na(final_data_2$BUN),17.7131,final_dat

```

```

a_2$BUN)
table(is.na(final_data_2$BUN))

mean(final_data_2$CR, na.rm = T)
final_data_2$CR=ifelse(is.na(final_data_2$CR),1.101941,final_data_2$CR)
table(is.na(final_data_2$CR))

mean(final_data_2$GLUCOSE.AC., na.rm = T)
final_data_2$GLUCOSE.AC.=ifelse(is.na(final_data_2$GLUCOSE.AC.),131.1165,final_data_2$GLUCOSE.AC.)
table(is.na(final_data_2$GLUCOSE.AC.))

mean(final_data_2$HB, na.rm = T)
final_data_2$HB=ifelse(is.na(final_data_2$HB),13.53255,final_data_2$HB)
table(is.na(final_data_2$HB))

mean(final_data_2$HBA1C, na.rm = T)
final_data_2$HBA1C=ifelse(is.na(final_data_2$HBA1C),6.510671,final_data_2$HBA1C)
table(is.na(final_data_2$HBA1C))

write.csv(final_data_2,paste0(spath, "final_data_2.csv"))

write.csv(paste0(spath,data=final_data_2,"final_data_2.csv"))

final_data$Ipwon=NA
final_data$Ipwon[(final_data$ipwon_period==0)]=0
final_data$Ipwon[(!final_data$ipwon_period==0)]=1

final_data$DEATH_YN=NA
final_data$DEATH_YN[(is.na(final_data$DTH_YN))]=0
final_data$DEATH_YN[(!is.na(final_data$DTH_YN))]=1

```



```

final_data$L_tube_feeding=NA
final_data$L_tube_feeding[(final_data$L_tubefeeding==0)]=0
final_data$L_tube_feeding[(!final_data$L_tubefeeding==0)]=1

final_data$LTC_YN=NA
final_data$LTC_YN[(final_data$APFRM_TYPE_CD==0)]=0
final_data$LTC_YN[(!final_data$APFRM_TYPE_CD==0)]=1

final_data$T_tube=NA
final_data$T_tube[final_data$tracheostom1==0 &
final_data$jeolgae==0]=0
final_data$T_tube[final_data$tracheostom1==1 |
final_data$jeolgae==1]=1
final_data$T_tube[is.na(final_data$tracheostom1) &
is.na(final_data$jeolgae)]=NA
final_data$T_tube[is.na(final_data$tracheostom1)
|final_data$jeolgae==0]=0
final_data$T_tube[final_data$tracheostom1==0
|is.na(final_data$jeolgae)]=0
final_data$T_tube[final_data$tracheostom1==1
|is.na(final_data$jeolgae)]=1

write.csv(final_data,paste0(spath, "final_data_1.csv"))

data_2$data_2_min=ymd(data_2$MDCARE_STRT_DT)-ymd(data_2$HME_DT)

length(data_2$data_2_min)

##test
data_2$interval = as.duration(ymd(data_2$MDCARE_STRT_DT) %--%
ymd(data_2$HME_DT))
data_2$interval_year =
as.duration(interval(ymd(data_2$HME_DT),ymd(data_2$MDCARE_STRT_DT
))) %/% as.duration(years(1))

```

```

x=data.frame(data_2$interval)
sum(!is.na(x))

data_2$interval = duration(data_2$interval,units="seconds")
data_3=data_2[!is.na(data_2$HME_DT), c("interval",
"MDCARE_STRT_DT", "HME_DT")]
data_3
sum(!is.na(data_3))

colnames(data_2)
g1eq_t_max =setDT(g1eq_t)[order(-STD_YYYY, INDI_DSCM_NO),
head(.SD,1), by=INDI_DSCM_NO]
g1eq_t_min =setDT(g1eq_t)[order(STD_YYYY, INDI_DSCM_NO),
head(.SD,1), by=INDI_DSCM_NO]
length(g1eq_t_max$INDI_DSCM_NO)
length(g1eq_t_min$INDI_DSCM_NO)
setnames(g1eq_t_max, c("STD_YYYY", "CMPR_DSB_GRADE",
"MAIN_DSB_TYPE"), c("STD_YYYY_L", "CMPR_DSB_GRADE_L",
"MAIN_DSB_TYPE_L"))

g1eq_t_uniq=merge(g1eq_t_min[,c("INDI_DSCM_NO","STD_YYYY",
"CMPR_DSB_GRADE", "MAIN_DSB_TYPE")],
g1eq_t_max[,c("INDI_DSCM_NO","STD_YYYY_L", "CMPR_DSB_GRADE_L",
"MAIN_DSB_TYPE_L")],by="INDI_DSCM_NO", all.x=T)

data_1=merge(ilsan_hospital_brain_injury,g1eq_t_uniq,
by="INDI_DSCM_NO", all.x=T)

#length(g1eq_t$INDI_DSCM_NO)
g1eq_t_data1=merge(g1eq_t,data_1[,c("INDI_DSCM_NO", "STD_YYYY")],
by="INDI_DSCM_NO", all.x=T)
#length(g1eq_t_data1$INDI_DSCM_NO)

g1eq_t_data1$YR_diff=g1eq_t_data1$EXMD_BZ_YYYY -
g1eq_t_data1$STD_YYYY

```

```

#i=2002
#file =paste0(path,"bfc_",i,".sas7bat")

#bfc_=read_sas(paste0("bfc_",i,".sas7bdat"))
# 장기요양 변수

rvsn_hfvt_tbnibase <-
read_sas("~/data/data_source/user_data/rvsn_hfvt_tbnibase.sas7bdat", NULL) # long term care 기본(신청, 인정조사, 등급 판정)

rvsn_hfvt_tbnyinst <-
read_sas("~/data/data_source/user_data/rvsn_hfvt_tbnyinst.sas7bdat", NULL) # long term care 시설 변수

rvsn_hfvt_tbnypay <-
read_sas("~/data/data_source/user_data/rvsn_hfvt_tbnypay.sas7bdat", NULL) # long term care 지불 변수

rvsn_hfvt_tbnyuc34 <-
read_sas("~/data/data_source/user_data/rvsn_hfvt_tbnyuc34.sas7bdat", NULL)

# 장기요양보험 인정욕구 조사

colnames(rvsn_hfvt_tbnyuc34)

write.csv(rvsn_hfvt_tbnibase,paste0(spath, "LTC_data.csv"))

LTC_data_1=rvsn_hfvt_tbnibase[,c(1:21)]

colnames(LTC_data_1)

LTC_data_2=rvsn_hfvt_tbnyuc34[,c(1:4,72,74,76,78,80,82,84,86,88,90,158,160)]
colnames(LTC_data_2)

```

```

#장기 요양 변수 최종 final 자료
LTC_data_3=merge(LTC_data_1,LTC_data_2,by="INDI_DSCM_NO",
all.x=T)
LTC_data_3$APLY_YM=as.numeric(LTC_data_3$APLY_YM)
LTC_data_3$GRADE_JUDG_YM=as.numeric(LTC_data_3$GRADE_JUDG_YM)

LTC_data_final=LTC_data_3[!duplicated(LTC_data_3$INDI_DSCM_NO),]
length(LTC_data_final$INDI_DSCM_NO)

colnames(LTC_data_3)

#LTC_data_3_last_day= LTC_data_3 %>%
#group_by (INDI_DSCM_NO) %>%
#filter(APLY_YM == max(APLY_YM))

mod_data_brain_ct_k_trauma_Final<-mod_data_brain_ct_k_trauma[!dup
licated(mod_data_brain_ct_k_trauma$INDI_DSCM_NO_R),]

LTC_data_2=LTC_data_1[!duplicated(LTC_data_1$INDI_DSCM_NO),]
length(LTC_data_1$INDI_DSCM_NO)
length(LTC_data_2$INDI_DSCM_NO)

colnames(rvsn_hfvt_tbnabase)
colnames(rvsn_hfvt_tbnabase)
colnames(rvsn_hfvt_tbnyst)
colnames(rvsn_hfvt_tbnypay)
colnames(rvsn_hfvt_tbnyc34)

#-----

# 일산병원 자료 (brain CT, trauma and non-trauma)

mod_data_brain_ct_k <-

```

```

read_sas("~/data/data_source/user_data/mod_data_brain_ct_k.sas7bd
at", NULL)
# 진단에 trauma_non_trauma 포함
mod_data_brain_ct_k_trauma <-
mod_data_brain_ct_k[grepl("Traumatic|traumatic|Trauma|trauma|Conc
ussion|concussion|Contusion|contusion",mod_data_brain_ct_k$Result
s, perl=TRUE),]
mod_data_brain_ct_k_trauma_Final<-mod_data_brain_ct_k_trauma[!dup
licated(mod_data_brain_ct_k_trauma$INDI_DSCM_NO_R),]

# 진단에 trauma_trauma 포함
mod_data_brain_ct_k_non_trauma<-
mod_data_brain_ct_k[grepl("Nontrauma|nontrauma",mod_data_brain_ct
_k$Diagnosis, perl=TRUE),]
mod_data_brain_ct_k_non_trauma_Final <-
mod_data_brain_ct_k[!grepl("Traumatic|traumatic|Trauma|trauma",mo
d_data_brain_ct_k$Results, perl=TRUE),]

# 진단에 infarction 포함
mod_data_brain_ct_k_infarction
<-mod_data_brain_ct_k[grepl("infarction|Infarction",mod_data_brai
n_ct_k$Results, mod_data_brain_ct_k$Diagnosis, perl=TRUE),]
brain_ct_infarction_Final=mod_data_brain_ct_k_infarction
[!duplicated(mod_data_brain_ct_k_infarction$INDI_DSCM_NO_R),]

# 병변에 lesion_focal_small 포함
mod_data_brain_ct_k_Focal_small<-mod_data_brain_ct_k[grepl("Focal
|focal|Small|small|minimal",mod_data_brain_ct_k$Results,
perl=TRUE),]
brain_ct_Focal_small_Final<-mod_data_brain_ct_k_Focal_small[!dupl
icated(mod_data_brain_ct_k_Focal_small$INDI_DSCM_NO_R),]

# 병변에 lesion_multifocal_large 포함
mod_data_brain_ct_k_multifocal_large<-mod_data_brain_ct_k[grepl("
Multifocal|multifocal|mutiple|Multiple|Large|large|Diffuse|diffus

```

```

e",mod_data_brain_ct_k$Results, perl=TRUE),]
brain_ct_multifocal_large_Final<-mod_data_brain_ct_k_multifocal_la
rge[!duplicated(mod_data_brain_ct_k_multifocal_large$INDI_DSCM_NO_
R),]
# 병변에 lesion_Acute 포함
mod_data_brain_ct_k_Acute_recent<-mod_data_brain_ct_k[grepl("Acut
e|acute|Recent|recent",mod_data_brain_ct_k$Results,mod_data_brain
_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Acute_recent_Final<-mod_data_brain_ct_k_Acute_recent[!du
plicated(mod_data_brain_ct_k_Acute_recent$INDI_DSCM_NO_R),]

# 병변에 lesion_Chronic and old 포함
mod_data_brain_ct_k_Chronic_old<-mod_data_brain_ct_k[grepl("Chron
ic|chronic|Recent|recent|Encephalomal|encephalomal|Old|old",mod_d
ata_brain_ct_k$Results,mod_data_brain_ct_k$Diagnosis,
perl=TRUE),]
brain_ct_Chronic_old_Final<-mod_data_brain_ct_k_Chronic_old[!dupl
icated(mod_data_brain_ct_k_Chronic_old$INDI_DSCM_NO_R),]

# 병변에 SAH 포함
mod_data_brain_ct_k_SAH<-mod_data_brain_ct_k[grepl("subarachnoida
l|SAH",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_SAH_Final<-mod_data_brain_ct_k_SAH[!duplicated(mod_data_
brain_ct_k_SAH$INDI_DSCM_NO_R),]

# 병변에 SDH 포함
mod_data_brain_ct_k_SDH<-mod_data_brain_ct_k[grepl("subdural|Subd
ural|SDH",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_SDH_Final<-mod_data_brain_ct_k_SDH[!duplicated(mod_data_
brain_ct_k_SDH$INDI_DSCM_NO_R),]

# 병변에 ICH 포함
mod_data_brain_ct_k_ICH<-mod_data_brain_ct_k[grepl("Intracranial

```

```

hemorrhage|intracranial hemorrhage|intracerebral
hemorrhage|ICH|hematoma|hemorrhagic",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_ICH_Final<-mod_data_brain_ct_k_ICH[!duplicated(mod_data_
brain_ct_k_ICH$INDI_DSCM_NO_R),]

# 병변에 IVH 포함
mod_data_brain_ct_k_IVH<-mod_data_brain_ct_k[grepl("Intraventricu
lar hemorrhage|intrventricular
hemorrhage|IVH",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_IVH_Final<-mod_data_brain_ct_k_IVH[!duplicated(mod_data_
brain_ct_k_IVH$INDI_DSCM_NO_R),]

# 병변에 Frotallobe 포함
mod_data_brain_ct_k_Frontal<-mod_data_brain_ct_k[grepl("Frontal|f
ronatal",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Frontal_Final<-mod_data_brain_ct_k_Frontal[!duplicated(m
od_data_brain_ct_k_Frontal$INDI_DSCM_NO_R),]

# 병변에 Temporal lobe 포함
mod_data_brain_ct_k_Temporal
<-mod_data_brain_ct_k[grepl("Temporal|temporal",mod_data_brain_ct
_k$Results, mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Temporal_Final<-mod_data_brain_ct_k_Temporal[!duplicated
(mod_data_brain_ct_k_Temporal$INDI_DSCM_NO_R),]

# 병변에 parietal lobe 포함
mod_data_brain_ct_k_Parietal<-mod_data_brain_ct_k[grepl("Parietal
|pareital",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Parietal_Final<-mod_data_brain_ct_k_Parietal[!duplicated
(mod_data_brain_ct_k_Parietal$INDI_DSCM_NO_R),]
# 병변에 Occipital lobe 포함

```

```

mod_data_brain_ct_k_Occipital<-mod_data_brain_ct_k[grepl("Occipital|occipital",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Occipital_Final<-mod_data_brain_ct_k_Occipital[!duplicated(mod_data_brain_ct_k_Occipital$INDI_DSCM_NO_R),]

# 병변에 Cerebellum 포함
mod_data_brain_ct_k_Cerebellum<-mod_data_brain_ct_k[grepl("Cerebellum|cerebellum|Cerebellar|cerebellar",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Cerebellum_Final<-mod_data_brain_ct_k_Cerebellum[!duplicated(mod_data_brain_ct_k_Cerebellum$INDI_DSCM_NO_R),]

# 병변에 BG 포함
mod_data_brain_ct_k_BG<-mod_data_brain_ct_k[grepl("BG|BG|Basal ganglia|basal ganglia",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_BG_Final<-mod_data_brain_ct_k_BG[!duplicated(mod_data_brain_ct_k_BG$INDI_DSCM_NO_R),]

# 병변에 thalamus 포함
mod_data_brain_ct_k_Thalamus
<-mod_data_brain_ct_k[grepl("Thalamus|thalamus|Thalamic|thalamic",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Thalamus_Final<-mod_data_brain_ct_k_Thalamus[!duplicated(mod_data_brain_ct_k_Thalamus$INDI_DSCM_NO_R),]

# 병변에 subcortex 포함
mod_data_brain_ct_k_subcortex<-mod_data_brain_ct_k[grepl("Subcortex|subcortex|Subcortical|subcortical",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_subcortex_Final<-mod_data_brain_ct_k_subcortex[!duplicated(mod_data_brain_ct_k_subcortex$INDI_DSCM_NO_R),]

```



```

# 병변에 brain_stem 포함
mod_data_brain_ct_k_brain_stem<-mod_data_brain_ct_k[grepl("Midbra
in|midbrain|Pons|pons|Lateral medullary|lateral medullary|Medial
medullary|medial medullary",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_brain_stem_Final<-mod_data_brain_ct_k_brain_stem[!duplic
ated(mod_data_brain_ct_k_brain_stem$INDI_DSCM_NO_R),]

# 병변에 ACA 포함
mod_data_brain_ct_k_ACA<-mod_data_brain_ct_k[grepl("ACA|Anterior
cerebral artery|arterial cerebral
artery",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_ACA_Final<-mod_data_brain_ct_k_ACA[!duplicated(mod_data_
brain_ct_k_ACA$INDI_DSCM_NO_R),]

# 병변에 MCA 포함
mod_data_brain_ct_k_MCA<-mod_data_brain_ct_k[grepl("MCA|middle
cerebral artery|middle cerebral
artery",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_MCA_Final<-mod_data_brain_ct_k_MCA[!duplicated(mod_data_
brain_ct_k_MCA$INDI_DSCM_NO_R),]

# 병변에 PCA 포함
mod_data_brain_ct_k_PCA<-mod_data_brain_ct_k[grepl("PCA|Posterior
cerebral artery|Posterior cerebral
artery",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_PCA_Final<-mod_data_brain_ct_k_PCA[!duplicated(mod_data_
brain_ct_k_PCA$INDI_DSCM_NO_R),]

# 병변에 basilar artery 포함
mod_data_brain_ct_k_basilar_artery<-mod_data_brain_ct_k[grepl("ba

```

```

silar|Basilar",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_basilar_artery_Final<-mod_data_brain_ct_k_basilar_artery
[!duplicated(mod_data_brain_ct_k_basilar_artery$INDI_DSCM_NO_R),]

# 병변에 Aneurysm 포함
mod_data_brain_ct_k_Aneurysm
<-mod_data_brain_ct_k[grepl("Aneurysm|aneurysm",mod_data_brain_ct
_k$Results, mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Aneurysm_Final<-mod_data_brain_ct_k_Aneurysm
[!duplicated(mod_data_brain_ct_k_Aneurysm$INDI_DSCM_NO_R),]

# 시술 및 수술 포함 Clipping
mod_data_brain_ct_k_Clipping<-mod_data_brain_ct_k[grepl("Clipping
|clipping|Clip|clip",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Clipping_Final<-mod_data_brain_ct_k_Clipping[!duplicated
(mod_data_brain_ct_k_Clipping$INDI_DSCM_NO_R),]

# 시술 및 수술 포함 Coil
mod_data_brain_ct_k_Coil<-mod_data_brain_ct_k[grepl("Coil|coil|Em
bolization|embolization",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_Coil_Final<-mod_data_brain_ct_k_Coil[!duplicated(mod_dat
a_brain_ct_k_Coil$INDI_DSCM_NO_R),]

# 병변에 craniectomy 포함
mod_data_brain_ct_k_craniectomy
<-mod_data_brain_ct_k[grepl("Craniectomy|craniectomy ",mod_data_
brain_ct_k$Results, mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_craniectomy_Final<-mod_data_brain_ct_k_craniectomy
[!duplicated(mod_data_brain_ct_k_craniectomy $INDI_DSCM_NO_R),]

# 병변에 craniotomy 포함
mod_data_brain_ct_k_craniotomy<-mod_data_brain_ct_k[grepl("Cranio

```

```

tomy|craniotomy",mod_data_brain_ct_k$Results,
mod_data_brain_ct_k$Diagnosis,perl=TRUE),]
brain_ct_craniotomy_Final<-mod_data_brain_ct_k_craniotomy[!duplic
ated(mod_data_brain_ct_k_craniotomy$INDI_DSCM_NO_R),]
# 일산병원 자료 (brain MRI)-----

mod_data_brain_mri_k <-
read_sas("~/data/data_source/user_data/mod_data_brain_mri_k.sas7b
dat", NULL)
write.csv (mod_data_brain_mri_k ,paste0(spath,
"mod_data_brain_mri_k .csv"))

# 특이 소견 없음 제거 (정상 소견 제거)
mod_data_brain_mri_k<- mod_data_brain_mri_k[!grepl("No
remarkable|No signal abnormality|No evidence|No definate|No
Significant",mod_data_brain_mri_k$Results, perl=TRUE),]

# brain lesion이 있는 군중 중복 인원 제거
brain_mri_lesion=mod_data_brain_mri_k[!duplicated(mod_data_brain_
mri_k$INDI_DSCM_NO_R),]

# 진단에 Nontrauma|nontrauma 중에 중복 인원 제거
mod_data_brain_mri_k_non_trauma<-
mod_data_brain_mri_k[grepl("Nontrauma|nontrauma",mod_data_brain_m
ri_k$Diagnosis, perl=TRUE),]
brain_MRI_non_trauma=mod_data_brain_mri_k_non_trauma[!duplicated(
mod_data_brain_mri_k_non_trauma$INDI_DSCM_NO_R),]
mod_data_brain_mri_k_non_trauma_1<-
mod_data_brain_mri_k_non_trauma[!grepl("Traumatic|traumatic|Traum
a|trauma",mod_data_brain_mri_k_non_trauma$Results, perl=TRUE),]
brain_MRI_non_trauma_Final=mod_data_brain_mri_k_non_trauma_1[!dup
licated(mod_data_brain_mri_k_non_trauma_1$INDI_DSCM_NO_R),]

# 진단에 traumatic brain injury 포함
mod_data_brain_mri_k_trauma<-

```

```

mod_data_brain_mri_k[grepl("Traumatic|traumatic|Trauma|trauma|Con
cussion|concussion|Contusion|contusion|Diffuse axonal|diffuse
axonal|DAI ",mod_data_brain_mri_k$Diagnosis,mod_data_brain_
mri_k$Results, perl=TRUE),]
brain_MRI_trauma_Final=mod_data_brain_mri_k_trauma[!duplicated(mod
_data_brain_mri_k_trauma$INDI_DSCM_NO_R),]

# 진단에 infarction 포함
mod_data_brain_mri_k_infarction
<-mod_data_brain_mri_k[grepl("infarction|Infarction",mod_data_bra
in_mri_k$Results, mod_data_brain_mri_k$Diagnosis, perl=TRUE),]
brain_mri_infarction_Final=mod_data_brain_mri_k_infarction
[!duplicated(mod_data_brain_mri_k_infarction$INDI_DSCM_NO_R),]

# 병변에 lesion_focal_small 포함
mod_data_brain_mri_k_Focal_small<-mod_data_brain_mri_k[grepl("Foc
al|focal|Small|small|minimal",mod_data_brain_mri_k$Results,
perl=TRUE),]
brain_mri_Focal_small_Final<-mod_data_brain_mri_k_Focal_small[!du
plicated(mod_data_brain_mri_k_Focal_small$INDI_DSCM_NO_R),]

# 병변에 lesion_multifocal_large 포함
mod_data_brain_mri_k_multifocal_large<-mod_data_brain_mri_k[grepl
("Multifocal|multifocal|mutiple|Multiple|Large|large|Diffuse|diff
use",mod_data_brain_mri_k$Results, perl=TRUE),]
brain_mri_multifocal_large_Final<-mod_data_brain_mri_k_multifocal
_large[!duplicated(mod_data_brain_mri_k_multifocal_large$INDI_DSC
M_NO_R),]

# 병변에 lesion_Acute포함
mod_data_brain_mri_k_Acute_recent<-mod_data_brain_mri_k[grepl("Ac
ute|acute|Recent|recent",mod_data_brain_mri_k$Results,mod_data_br
ain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Acute_recent_Final<-mod_data_brain_mri_k_Acute_recent[!
duplicated(mod_data_brain_mri_k_Acute_recent$INDI_DSCM_NO_R),]

```

```

# 병변에 lesion_Chronic and old 포함
mod_data_brain_mri_k_Chronic_old<-mod_data_brain_mri_k[grepl("Chronic|chronic|Recent|recent|Encephalomal|encephalomal|Old|old",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Chronic_old_Final<-mod_data_brain_mri_k_Chronic_old[!duplicated(mod_data_brain_mri_k_Chronic_old$INDI_DSCM_NO_R),]

# 병변에 SAH 포함
mod_data_brain_mri_k_SAH<-mod_data_brain_mri_k[grepl("subarachnoidal|SAH",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_SAH_Final<-mod_data_brain_mri_k_SAH[!duplicated(mod_data_brain_mri_k_SAH$INDI_DSCM_NO_R),]

# 병변에 SDH 포함
mod_data_brain_mri_k_SDH<-mod_data_brain_mri_k[grepl("subdural|Subdural|SDH",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_SDH_Final<-mod_data_brain_mri_k_SDH[!duplicated(mod_data_brain_mri_k_SDH$INDI_DSCM_NO_R),]

# 병변에 ICH 포함
mod_data_brain_mri_k_ICH<-mod_data_brain_mri_k[grepl("Intracranial hemorrhage|intracranial hemorrhage|intracerebral hemorrhage|ICH|hematoma|horrhagic",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_ICH_Final<-mod_data_brain_mri_k_ICH[!duplicated(mod_data_brain_mri_k_ICH$INDI_DSCM_NO_R),]

# 병변에 IVH 포함
mod_data_brain_mri_k_IVH<-mod_data_brain_mri_k[grepl("Intraventricular hemorrhage|intrventricular hemorrhage|IVH",mod_data_brain_mri_k$Results,

```

```

mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_IVH_Final<-mod_data_brain_mri_k_IVH[!duplicated(mod_data_brain_mri_k_IVH$INDI_DSCM_NO_R),]

# 병변에 Frontal lobe 포함
mod_data_brain_mri_k_Frontal<-mod_data_brain_mri_k[grepl("Frontal|fronatal",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Frontal_Final<-mod_data_brain_mri_k_Frontal[!duplicated(mod_data_brain_mri_k_Frontal$INDI_DSCM_NO_R),]

# 병변에 Temporal lobe 포함
mod_data_brain_mri_k_Temporal<-mod_data_brain_mri_k[grepl("Temporal|temporal",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Temporal_Final<-mod_data_brain_mri_k_Temporal[!duplicated(mod_data_brain_mri_k_Temporal$INDI_DSCM_NO_R),]

# 병변에 parietallobe 포함
mod_data_brain_mri_k_Parietal<-mod_data_brain_mri_k[grepl("Parietal|pareital",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Parietal_Final<-mod_data_brain_mri_k_Parietal[!duplicated(mod_data_brain_mri_k_Parietal$INDI_DSCM_NO_R),]

# 병변에 Occipital lobe 포함
mod_data_brain_mri_k_Occipital<-mod_data_brain_mri_k[grepl("Occipital|occipital",mod_data_brain_mri_k$Results,mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Occipital_Final<-mod_data_brain_mri_k_Occipital[!duplicated(mod_data_brain_mri_k_Occipital$INDI_DSCM_NO_R),]

# 병변에 Cerebellum 포함
mod_data_brain_mri_k_Cerebellum<-mod_data_brain_mri_k[grepl("Cerebellum|cerebellum|Cerebellar|cerebellar",mod_data_brain_mri_k$Res

```

```

ults, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Cerebellum_Final<-mod_data_brain_mri_k_Cerebellum[!duplicated(mod_data_brain_mri_k_Cerebellum$INDI_DSCM_NO_R),]

# 병변에 BG 포함
mod_data_brain_mri_k_BG<-mod_data_brain_mri_k[grepl("BG|BG|Basal ganglia|basal ganglia",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_BG_Final<-mod_data_brain_mri_k_BG[!duplicated(mod_data_brain_mri_k_BG$INDI_DSCM_NO_R),]

# 병변에 thalamus 포함
mod_data_brain_mri_k_Thalamus
<-mod_data_brain_mri_k[grepl("Thalamus|thalamus|Thalamic|thalamic",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Thalamus_Final<-mod_data_brain_mri_k_Thalamus[!duplicated(mod_data_brain_mri_k_Thalamus$INDI_DSCM_NO_R),]

# 병변에 subcortex 포함
mod_data_brain_mri_k_subcortex<-mod_data_brain_mri_k[grepl("Subcortex|subcortex|Subcortical|subcortical",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_subcortex_Final<-mod_data_brain_mri_k_subcortex[!duplicated(mod_data_brain_mri_k_subcortex$INDI_DSCM_NO_R),]

# 병변에 brain_stem 포함
mod_data_brain_mri_k_brain_stem<-mod_data_brain_mri_k[grepl("Midbrain|midbrain|Pons|pons|Lateral medullary|lateral medullary|Medial medullary|medial medullary",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_brain_stem_Final<-mod_data_brain_mri_k_brain_stem[!duplicated(mod_data_brain_mri_k_brain_stem$INDI_DSCM_NO_R),]

```

```

# 병변에 ACA 포함
mod_data_brain_mri_k_ACA<-mod_data_brain_mri_k[grepl("ACA|Anterior cerebral artery|arterial cerebral artery",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_ACA_Final<-mod_data_brain_mri_k_ACA[!duplicated(mod_data_brain_mri_k_ACA$INDI_DSCM_NO_R),]

# 병변에 MCA 포함
mod_data_brain_mri_k_MCA<-mod_data_brain_mri_k[grepl("MCA|middle cerebral artery|middle cerebral artery",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_MCA_Final<-mod_data_brain_mri_k_MCA[!duplicated(mod_data_brain_mri_k_MCA$INDI_DSCM_NO_R),]

# 병변에 PCA 포함
mod_data_brain_mri_k_PCA<-mod_data_brain_mri_k[grepl("PCA|Posterior cerebral artery|Posterior cerebral artery",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_PCA_Final<-mod_data_brain_mri_k_PCA[!duplicated(mod_data_brain_mri_k_PCA$INDI_DSCM_NO_R),]

# 병변에 basilar artery 포함
mod_data_brain_mri_k_basilar_artery<-mod_data_brain_mri_k[grepl("basilar|Basilar",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_basilar_artery_Final<-mod_data_brain_mri_k_basilar_artery[!duplicated(mod_data_brain_mri_k_basilar_artery$INDI_DSCM_NO_R),]

# 병변에 Aneurysm 포함
mod_data_brain_mri_k_Aneurysm
<-mod_data_brain_mri_k[grepl("Aneurysm|aneurysm",mod_data_brain_mri_k$Results, mod_data_brain_mri_k$Diagnosis,perl=TRUE),]

```



```

brain_mri_Aneurysm_Final<-mod_data_brain_mri_k_Aneurysm
[!duplicated(mod_data_brain_mri_k_Aneurysm$INDI_DSCM_NO_R),]

# 시술 및 수술 포함 Clipping
mod_data_brain_mri_k_Clipping<-mod_data_brain_mri_k[grepl("Clippi
ng|clipping|Clip|clip",mod_data_brain_mri_k$Results,
mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Clipping_Final<-mod_data_brain_mri_k_Clipping[!duplicat
ed(mod_data_brain_mri_k_Clipping$INDI_DSCM_NO_R),]

# 시술 및 수술 포함 Coil
mod_data_brain_mri_k_Coil<-mod_data_brain_mri_k[grepl("Coil|coil"
,mod_data_brain_mri_k$Results,
mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_Coil_Final<-mod_data_brain_mri_k_Coil[!duplicated(mod_d
ata_brain_mri_k_Coil$INDI_DSCM_NO_R),]

# 병변에 craniectomy 포함
mod_data_brain_mri_k_craniectomy
<-mod_data_brain_mri_k[grepl("Craniectomy|craniectomy
",mod_data_brain_mri_k$Results,
mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_craniectomy_Final<-mod_data_brain_mri_k_craniectomy
[!duplicated(mod_data_brain_mri_k_craniectomy $INDI_DSCM_NO_R),]

# 병변에 craniotomy 포함
mod_data_brain_mri_k_craniotomy<-mod_data_brain_mri_k[grepl("Cran
iotomy|craniotomy",mod_data_brain_mri_k$Results,
mod_data_brain_mri_k$Diagnosis,perl=TRUE),]
brain_mri_craniotomy_Final<-mod_data_brain_mri_k_craniotomy[!dupl
icated(mod_data_brain_mri_k_craniotomy$INDI_DSCM_NO_R),]

# 일산병원 Ct_brain_MRI 자료 결합
# traumatic_brain_injury_CT_MRI_최종결합자료
brain_ct_mri_trauma_Final=merge(mod_data_brain_ct_k_trauma_Final,

```

```

brain_MRI_trauma_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_trauma_Final=brain_ct_mri_trauma_Final[!duplicated(brain_ct
_mri_trauma_Final$INDI_DSCM_NO_R),]
brain_trauma_Final$TBI=NA
brain_trauma_Final$TBI=brain_trauma_Final$INDI_DSCM_NO_R
brain_trauma_Final$TBI<-"Y"
TBI<-brain_trauma_Final[,c("INDI_DSCM_NO_R", "TBI")]

# non_traumatic_brain_injury_CT_MRI_최종결합자료
brain_ct_mri_non_trauma_Final=merge(mod_data_brain_ct_k_non_traum
a_Final,brain_MRI_non_trauma_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_non_trauma_Final=brain_ct_mri_non_trauma_Final[!duplicated(
brain_ct_mri_non_trauma_Final$INDI_DSCM_NO_R),]
colnames(mod_data_brain_ct_k_non_trauma_Final)
colnames(brain_MRI_non_trauma_Final)
brain_non_trauma_Final$non_trauma=NA
brain_non_trauma_Final$non_trauma=brain_non_trauma_Final$INDI_DSC
M_NO_R
brain_non_trauma_Final$non_trauma<-"Y"
Non_trauma<-brain_non_trauma_Final[,c("INDI_DSCM_NO_R", "non_trauma")]

# Infarction_최종결합자료
brain_ct_mri_infarction_Final=merge(brain_ct_infarction_Final,bra
in_mri_infarction_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_infarction_Final=brain_ct_mri_infarction_Final[!duplicated(
brain_ct_mri_infarction_Final$INDI_DSCM_NO_R),]
brain_infarction_Final$Infarction=NA
brain_infarction_Final$Infarction=NA
brain_infarction_Final$Infarction=brain_infarction_Final$INDI_DSC
M_NO_R
brain_infarction_Final$Infarction<-"Y"
Infarction<-brain_infarction_Final[,c("INDI_DSCM_NO_R", "Infarction")]

# SAH_최종결합자료
brain_ct_mri_SAH_Final=merge(brain_ct_SAH_Final,brain_MRI_SAH_Fin

```

```

al,by="INDI_DSCM_NO_R", all.x=T)
brain_SAH_Final=brain_ct_mri_SAH_Final[!duplicated(brain_ct_mri_S
AH_Final$INDI_DSCM_NO_R),]
brain_SAH_Final$SAH=NA
brain_SAH_Final$SAH=brain_SAH_Final$INDI_DSCM_NO_R
brain_SAH_Final$SAH<-"Y"
SAH<-brain_SAH_Final[,c("INDI_DSCM_NO_R","SAH")]

#SDH_최종결합자료
brain_ct_mri_SDH_Final=merge(brain_ct_SDH_Final,brain_MRI_SDH_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_SDH_Final=brain_ct_mri_SDH_Final[!duplicated(brain_ct_mri_S
DH_Final$INDI_DSCM_NO_R),]
brain_SDH_Final$SDH=NA
brain_SDH_Final$SDH=brain_SDH_Final$INDI_DSCM_NO_R
brain_SDH_Final$SDH<-"Y"
SDH<-brain_SDH_Final[,c("INDI_DSCM_NO_R","SDH")]

#ICH_최종결합자료
brain_ct_mri_ICH_Final=merge(brain_ct_ICH_Final,brain_mri_ICH_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_ICH_Final=brain_ct_mri_ICH_Final[!duplicated(brain_ct_mri_I
CH_Final$INDI_DSCM_NO_R),]
brain_ICH_Final$ICH=NA
brain_ICH_Final$ICH=brain_ICH_Final$INDI_DSCM_NO_R
brain_ICH_Final$ICH<-"Y"
ICH<-brain_ICH_Final[,c("INDI_DSCM_NO_R","ICH")]

#IVH_최종결합자료
brain_ct_mri_IVH_Final=merge(brain_ct_IVH_Final,brain_mri_IVH_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_IVH_Final=brain_ct_mri_IVH_Final[!duplicated(brain_ct_mri_I
VH_Final$INDI_DSCM_NO_R),]
brain_IVH_Final$IVH=NA
brain_IVH_Final$IVH=brain_IVH_Final$INDI_DSCM_NO_R
brain_IVH_Final$IVH<-"Y"

```

```

IVH<-brain_IVH_Final[,c("INDI_DSCM_NO_R","IVH")]
#Focal_small_최종결합자료
brain_ct_mri_Focal_small_Final=merge(brain_ct_Focal_small_Final,b
rain_MRI_Focal_small_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Focal_small_Final=brain_ct_mri_Focal_small_Final[!duplicate
d(brain_ct_mri_Focal_small_Final$INDI_DSCM_NO_R),]
brain_Focal_small_Final$Focal_small=NA
brain_Focal_small_Final$Focal_small=brain_Focal_small_Final$INDI_
DSCM_NO_R
brain_Focal_small_Final$Focal_small<-"Y"
Focal_small<-brain_Focal_small_Final[,c("INDI_DSCM_NO_R","Focal_s
mall")]

#lesion_multifocal_large_Diffuse_최종결합자료
brain_ct_mri_multifocal_large_Final=merge(brain_ct_multifocal_lar
ge_Final,brain_mri_multifocal_large_Final,by="INDI_DSCM_NO_R",
all.x=T)
brain_multifocal_large_Final=brain_ct_mri_multifocal_large_Final[
!duplicated(brain_ct_mri_multifocal_large_Final$INDI_DSCM_NO_R),]
brain_multifocal_large_Final$multifocal_large=NA
brain_multifocal_large_Final$multifocal_large=brain_multifocal_la
rge_Final$INDI_DSCM_NO_R
brain_multifocal_large_Final$multifocal_large<-"Y"
multifocal_large<-brain_multifocal_large_Final[,c("INDI_DSCM_NO_R
","multifocal_large")]

#Acute_recent_최종결합자료
brain_ct_mri_Acute_recent_Final=merge(brain_ct_Acute_recent_Final
,brain_mri_Acute_recent_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Acute_recent_Final=brain_ct_mri_Acute_recent_Final[!duplica
ted(brain_ct_mri_Acute_recent_Final$INDI_DSCM_NO_R),]
brain_Acute_recent_Final$Acute_recent=NA
brain_Acute_recent_Final$Acute_recent=brain_Acute_recent_Final$IN
DI_DSCM_NO_R
brain_Acute_recent_Final$Acute_recent<-"Y"

```

```

Acute_recent<-brain_Acute_recent_Final[,c("INDI_DSCM_NO_R", "Acute
_recent")]

#lesion_Chronic and old_최종결합자료
brain_ct_mri_Chronic_old_Final=merge(brain_ct_Chronic_old_Final,b
rain_mri_Chronic_old_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Chronic_old_Final=brain_ct_mri_Chronic_old_Final[!duplicate
d(brain_ct_mri_Chronic_old_Final$INDI_DSCM_NO_R),]
brain_Chronic_old_Final$Chronic_old=NA
brain_Chronic_old_Final$Chronic_old=brain_Chronic_old_Final$INDI_
DSCM_NO_R
brain_Chronic_old_Final$Chronic_old<-"Y"
Chronic_old<-brain_Chronic_old_Final[,c("INDI_DSCM_NO_R", "Chronic
_old")]

#Frontal_lobe_최종결합자료
brain_ct_mri_Frontal_Final=merge(brain_ct_Frontal_Final,brain_mri
_Frontal_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Frontal_Final=brain_ct_mri_Frontal_Final[!duplicated(brain_
ct_mri_Frontal_Final$INDI_DSCM_NO_R),]
brain_Frontal_Final$Frontal_Final=NA
brain_Frontal_Final$Frontal_Final=brain_Frontal_Final$INDI_DSCM_NO_R
brain_Frontal_Final$Frontal_Final<-"Y"
Frontal_Final<-brain_Frontal_Final[,c("INDI_DSCM_NO_R", "Frontal_F
inal")]

#Temporal_lobe_최종결합자료
brain_ct_mri_Temporal_Final=merge(brain_ct_Temporal_Final,brain_m
ri_Temporal_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Temporal_Final=brain_ct_mri_Temporal_Final[!duplicated(brai
n_ct_mri_Temporal_Final$INDI_DSCM_NO_R),]
brain_Temporal_Final$Temporal_Final=NA
brain_Temporal_Final$Temporal_Final=brain_Temporal_Final$INDI_DSC
M_NO_R
brain_Temporal_Final$Temporal_Final<-"Y"

```

```
Temporal_Final<-brain_Temporal_Final[,c("INDI_DSCM_NO_R", "Tempora  
l_Final")]
```

```
#Parietal_lobe_최종결합자료
```

```
brain_ct_mri_Parietal_Final=merge(brain_ct_Parietal_Final,brain_m  
ri_Parietal_Final,by="INDI_DSCM_NO_R", all.x=T)
```

```
brain_Parietal_Final=brain_ct_mri_Parietal_Final[!duplicated(brai  
n_ct_mri_Parietal_Final$INDI_DSCM_NO_R),]
```

```
brain_Parietal_Final$Parietal_Final=NA
```

```
brain_Parietal_Final$Parietal_Final=brain_Parietal_Final$INDI_DSC  
M_NO_R
```

```
brain_Parietal_Final$Parietal_Final<-"Y"
```

```
Parietal_Final<-brain_Parietal_Final[,c("INDI_DSCM_NO_R", "Parieta  
l_Final")]
```

```
#Occipital_lobe_최종결합자료
```

```
brain_ct_mri_Occipital_Final=merge(brain_ct_Occipital_Final,brain  
_mri_Occipital_Final,by="INDI_DSCM_NO_R", all.x=T)
```

```
brain_Occipital_Final=brain_ct_mri_Occipital_Final[!duplicated(bra  
in_ct_mri_Occipital_Final$INDI_DSCM_NO_R),]
```

```
brain_Occipital_Final$Occipital_Final=NA
```

```
brain_Occipital_Final$Occipital_Final=brain_Occipital_Final$INDI_  
DSCM_NO_R
```

```
brain_Occipital_Final$Occipital_Final<-"Y"
```

```
Occipital_Final<-brain_Occipital_Final[,c("INDI_DSCM_NO_R", "Occip  
ital_Final")]
```

```
#Cerebellum_최종자료
```

```
brain_ct_mri_Cerebellum_Final=merge(brain_ct_Cerebellum_Final,bra  
in_mri_Cerebellum_Final,by="INDI_DSCM_NO_R", all.x=T)
```

```
brain_Cerebellum_Final=brain_ct_mri_Cerebellum_Final[!duplicated(  
brain_ct_mri_Cerebellum_Final$INDI_DSCM_NO_R),]
```

```
brain_Cerebellum_Final$Cerebellum_Final=NA
```

```
brain_Cerebellum_Final$Cerebellum_Final=brain_Cerebellum_Final$IN  
DI_DSCM_NO_R
```

```

brain_Cerebellum_Final$Cerebellum_Final<-"Y"
Cerebellum_Final<-brain_Cerebellum_Final[,c("INDI_DSCM_NO_R", "Cer
ebellum_Final")]

#BG_최종자료
brain_ct_mri_BG_Final=merge(brain_ct_BG_Final,brain_mri_BG_Final,
by="INDI_DSCM_NO_R", all.x=T)
brain_BG_Final=brain_ct_mri_BG_Final[!duplicated(brain_ct_mri_BG_
Final$INDI_DSCM_NO_R),]
brain_BG_Final$BG_Final=NA
brain_BG_Final$BG_Final=brain_BG_Final$INDI_DSCM_NO_R
brain_BG_Final$BG_Final<-"Y"
BG_Final<-brain_BG_Final[,c("INDI_DSCM_NO_R", "BG_Final")]

#subcortex_최종자료
brain_ct_mri_subcortex_Final=merge(brain_ct_subcortex_Final,brain
_mri_subcortex_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_subcortex_Final=brain_ct_mri_subcortex_Final[!duplicated(br
ain_ct_mri_subcortex_Final$INDI_DSCM_NO_R),]
brain_subcortex_Final$subcortex_Final=NA
brain_subcortex_Final$subcortex_Final=brain_subcortex_Final$INDI
DSCM_NO_R
brain_subcortex_Final$subcortex_Final<-"Y"
subcortex_Final<-brain_subcortex_Final[,c("INDI_DSCM_NO_R", "subco
rtex_Final")]

#brain_stem_최종자료
brain_ct_mri_brain_stem_Final=merge(brain_ct_brain_stem_Final,bra
in_mri_brain_stem_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_brain_stem_Final=brain_ct_mri_brain_stem_Final[!duplicated(
brain_ct_mri_brain_stem_Final$INDI_DSCM_NO_R),]
brain_brain_stem_Final$brain_stem_Final=NA
brain_brain_stem_Final$brain_stem_Final=brain_brain_stem_Final$IN
DI_DSCM_NO_R
brain_brain_stem_Final$brain_stem_Final<-"Y"

```

```

brain_stem_Final<-brain_brain_stem_Final[,c("INDI_DSCM_NO_R", "bra
in_stem_Final")]

#Thalamus_최종자료
brain_ct_mri_Thalamus_Final=merge(brain_ct_Thalamus_Final,brain_m
ri_Thalamus_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Thalamus_Final=brain_ct_mri_Thalamus_Final[!duplicated(brai
n_ct_mri_Thalamus_Final$INDI_DSCM_NO_R),]
brain_Thalamus_Final$Thalamus_Final=NA
brain_Thalamus_Final$Thalamus_Final=brain_Thalamus_Final$INDI_DSC
M_NO_R
brain_Thalamus_Final$Thalamus_Final<-"Y"
Thalamus_Final<-brain_Thalamus_Final[,c("INDI_DSCM_NO_R", "Thalamu
s_Final")]

#ACA_최종자료
brain_ct_mri_ACA_Final=merge(brain_ct_ACA_Final,brain_mri_ACA_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_ACA_Final=brain_ct_mri_ACA_Final[!duplicated(brain_ct_mri_A
CA_Final$INDI_DSCM_NO_R),]
brain_ACA_Final$ACA_Final=NA
brain_ACA_Final$ACA_Final=brain_ACA_Final$INDI_DSCM_NO_R
brain_ACA_Final$ACA_Final<-"Y"
ACA_Final<-brain_ACA_Final[,c("INDI_DSCM_NO_R", "ACA_Final")]

#MCA_최종자료
brain_ct_mri_MCA_Final=merge(brain_ct_MCA_Final,brain_mri_MCA_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_MCA_Final=brain_ct_mri_MCA_Final[!duplicated(brain_ct_mri_M
CA_Final$INDI_DSCM_NO_R),]
brain_MCA_Final$MCA_Final=NA
brain_MCA_Final$MCA_Final=brain_MCA_Final$INDI_DSCM_NO_R
brain_MCA_Final$MCA_Final<-"Y"
MCA_Final<-brain_MCA_Final[,c("INDI_DSCM_NO_R", "MCA_Final")]

```



```

#PCA_최종자료
brain_ct_mri_PCA_Final=merge(brain_ct_PCA_Final,brain_mri_PCA_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_PCA_Final=brain_ct_mri_PCA_Final[!duplicated(brain_ct_mri_PCA_Final$INDI_DSCM_NO_R),]
brain_PCA_Final$PCA_Final=NA
brain_PCA_Final$PCA_Final=brain_PCA_Final$INDI_DSCM_NO_R
brain_PCA_Final$PCA_Final<-"Y"
PCA_Final<-brain_PCA_Final[,c("INDI_DSCM_NO_R","PCA_Final")]

#basilar_artery_최종자료
brain_ct_mri_basilar_artery_Final=merge(brain_ct_basilar_artery_Final,brain_mri_basilar_artery_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_basilar_artery_Final=brain_ct_mri_basilar_artery_Final[!duplicated(brain_ct_mri_basilar_artery_Final$INDI_DSCM_NO_R),]
brain_basilar_artery_Final$ basilar_artery_Final=NA
brain_basilar_artery_Final$
basilar_artery_Final=brain_basilar_artery_Final$INDI_DSCM_NO_R
brain_basilar_artery_Final$ basilar_artery_Final<-"Y"
basilar_artery_Final<-brain_basilar_artery_Final[,c("INDI_DSCM_NO_R","basilar_artery_Final")]

#Aneurysm_최종자료
brain_ct_mri_Aneurysm_Final=merge(brain_ct_Aneurysm_Final,brain_mri_Aneurysm_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Aneurysm_Final=brain_ct_mri_Aneurysm_Final[!duplicated(brain_ct_mri_Aneurysm_Final$INDI_DSCM_NO_R),]
brain_Aneurysm_Final$Aneurysm_Final=NA
brain_Aneurysm_Final$Aneurysm_Final=brain_Aneurysm_Final$INDI_DSCM_NO_R
brain_Aneurysm_Final$Aneurysm_Final<-"Y"
Aneurysm_Final<-brain_Aneurysm_Final[,c("INDI_DSCM_NO_R","Aneurysm_Final")]

#Clipping_최종자료

```

```

brain_ct_mri_Clipping_Final=merge(brain_ct_Clipping_Final,brain_mri_Clipping_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Clipping_Final=brain_ct_mri_Clipping_Final[!duplicated(brain_ct_mri_Clipping_Final$INDI_DSCM_NO_R),]
brain_Clipping_Final$Clipping_Final=NA
brain_Clipping_Final$Clipping_Final=brain_Clipping_Final$INDI_DSCM_NO_R
brain_Clipping_Final$Clipping_Final<-"Y"
Clipping_Final<-brain_Clipping_Final[,c("INDI_DSCM_NO_R","Clipping_Final")]

#Coil_최종자료
brain_ct_mri_Coil_Final=merge(brain_ct_Coil_Final,brain_mri_Coil_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Coil_Final=brain_ct_mri_Coil_Final[!duplicated(brain_ct_mri_Coil_Final$INDI_DSCM_NO_R),]
brain_Coil_Final$Coil_Final=NA
brain_Coil_Final$Coil_Final=brain_Coil_Final$INDI_DSCM_NO_R
brain_Coil_Final$Coil_Final<-"Y"
Coil_Final<-brain_Coil_Final[,c("INDI_DSCM_NO_R","Coil_Final")]

#Craniectomy_최종자료
brain_ct_mri_craniectomy_Final=merge(brain_ct_craniectomy_Final,brain_mri_craniectomy_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_craniectomy_Final=brain_ct_mri_craniectomy_Final[!duplicated(brain_ct_mri_craniectomy_Final$INDI_DSCM_NO_R),]
brain_craniectomy_Final$craniectomy_Final=NA
brain_craniectomy_Final$craniectomy_Final=brain_craniectomy_Final$INDI_DSCM_NO_R
brain_craniectomy_Final$craniectomy_Final<-"Y"
craniectomy_Final<-brain_craniectomy_Final[,c("INDI_DSCM_NO_R","craniectomy_Final")]

#Craniotomy_최종자료
brain_ct_mri_craniotomy_Final=merge(brain_ct_craniotomy_Final,bra

```

```

in_mri_craniotomy_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_craniotomy_Final=brain_ct_mri_craniotomy_Final[!duplicated(
brain_ct_mri_craniotomy_Final$INDI_DSCM_NO_R),]
brain_craniotomy_Final$craniotomy_Final=NA
brain_craniotomy_Final$craniotomy_Final=brain_craniotomy_Final$IN
DI_DSCM_NO_R
brain_craniotomy_Final$craniotomy_Final<-"Y"
craniotomy_Final<-brain_craniotomy_Final[,c("INDI_DSCM_NO_R","cra
niotomy_Final")]

# column name 변경
setnames(data_2, c("INDI_DSCM_NO"), c("INDI_DSCM_NO_R"))

#gastrostomy patients
mod_tubefeed_result_1108_k_2=mod_tubefeed_result_1108_k[(!is.na(m
od_tubefeed_result_1108_k$SUGA_CD) &
mod_tubefeed_result_1108_k$SUGA_CD==82661),]
mod_tubefeed_result_1108_k_2_gastrostomy=mod_tubefeed_result_1108
_k_2[!duplicated(mod_tubefeed_result_1108_k_2$INDI_DSCM_NO_R),]
mod_tubefeed_result_1108_k_2_gastrostomy$gastrostomy<-mod_tubefee
d_result_1108_k_2_gastrostomy$INDI_DSCM_NO_R
mod_tubefeed_result_1108_k_2_gastrostomy$gastrostomy<-"Y"

# 최종 gastrostomy feeding patients

gastrostomy<-mod_tubefeed_result_1108_k_2_gastrostomy[,c("INDI_DS
CM_NO_R","gastrostomy")]

#일산병원자료+장애등급+사망자료+gastrostomy

data_2=merge(data_1,ilsan_hospital_brain_injury_dth,
by="INDI_DSCM_NO_R", all.x=T)

data_3=merge(data_2,gastrostomy,by="INDI_DSCM_NO_R", all.x=T)

```

```

# 일산병원 자료 (chest, pneumonia 유무)

mod_data_brain_cxr_k <-
read_sas("~/data/data_source/user_data/mod_data_brain_cxr_k.sas7b
dat", NULL)
write.csv (mod_data_brain_cxr_k ,paste0(spath,
"mod_data_brain_cxr_k .csv"))

mod_data_brain_cxr_k_pn <-
mod_data_brain_cxr_k[grepl("pneumonia|atelectasis|Pneumonia|Atele
ctasis|Consolidation|consolidation",mod_data_brain_cxr_k$Results,
perl=TRUE),]
mod_data_brain_cxr_k_pn_Final
<-mod_data_brain_cxr_k_pn[!duplicated(mod_data_brain_cxr_k_pn$IND
I_DSCM_NO_R),]
colnames(mod_data_brain_cxr_k_pn_Final)

INDI_DSCM_NO_R=as.factor("INDI_DSCM_NO_R")

mod_data_brain_cxr_k_pn_Final$pneumonia<-mod_data_brain_cxr_k_pn_
Final$INDI_DSCM_NO_R
mod_data_brain_cxr_k_pn_Final$pneumonia<-"Y"

# 일산병원 자료 (tube_feeding 여부)

mod_tubefeed_result_1108_k <-
read_sas("~/data/data_source/user_data/mod_tubefeed_result_1108_k
.sas7bdat", NULL)

# L-tube patients
mod_tubefeed_result_1108_k_3=mod_tubefeed_result_1108_k[!is.na(mo
d_tubefeed_result_1108_k$SUGA_CD) &
!is.na(mod_tubefeed_result_1108_k$SUGA_CD==82661) &
((mod_tubefeed_result_1108_k$SUGA_CD==82660)) ,]
mod_tubefeed_result_1108_k_3_L_tubefeeding=mod_tubefeed_result_11

```

```

08_k_3[!duplicated(mod_tubefeed_result_1108_k_3$INDI_DSCM_NO_R),]
mod_tubefeed_result_1108_k_3_L_tubefeeding$L_tubefeeding=mod_tube
feed_result_1108_k_3_L_tubefeeding$INDI_DSCM_NO_R
mod_tubefeed_result_1108_k_3_L_tubefeeding$L_tubefeeding<-"Y"

# 최종 L-tube feeding patients
L_tube_feeding<-mod_tubefeed_result_1108_k_3_L_tubefeeding[,c("IN
DI_DSCM_NO_R", "L_tubefeeding")]

#일산병원자료+장애등급+사망자료+gastrostomy+L_tube_feeding

data_4=merge(data_3,L_tube_feeding,by="INDI_DSCM_NO_R", all.x=T)

# chest-Xray에서 pneumonia 최종 환자 명단
pneumonia<-mod_data_brain_cxr_k_pn_Final[,c("INDI_DSCM_NO_R", "pne
umonia")]

#일산병원자료+장애등급+사망자료+gastrostomy+L_tube_feeding+Pneumonia

data_5=merge(data_4,pneumonia,by="INDI_DSCM_NO_R", all.x=T)

# Tracheostomy 최종 명단
mod_data_brain_cxr_k_tracheostomy <-
mod_data_brain_cxr_k[grepl("Tracheo|tracheo",mod_data_brain_cxr_k
$Results, perl=TRUE),]
mod_data_brain_cxr_k_tracheostomy_Final
<-mod_data_brain_cxr_k_tracheostomy[!duplicated(mod_data_brain_cx
r_k_tracheostomy$INDI_DSCM_NO_R),]
mod_data_brain_cxr_k_tracheostomy_Final$tracheostomy<-mod_data_br
ain_cxr_k_tracheostomy_Final$INDI_DSCM_NO_R
mod_data_brain_cxr_k_tracheostomy_Final$tracheostomy<-"Y"
tracheostomy<-mod_data_brain_cxr_k_tracheostomy_Final[,c("INDI_DS
CM_NO_R", "tracheostomy")]

```

```

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy

data_6=merge(data_5,tracheostomy,by="INDI_DSCM_NO_R", all.x=T)

# 일산병원 Ct_brain_MRI 자료 결합
# traumatic_brain_injury_CT_MRI_최종 결합자료
brain_ct_mri_trauma_Final=merge(mod_data_brain_ct_k_trauma_Final,
brain_MRI_trauma_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_trauma_Final=brain_ct_mri_trauma_Final[!duplicated(brain_ct
_mri_trauma_Final$INDI_DSCM_NO_R),]
brain_trauma_Final$TBI=NA
brain_trauma_Final$TBI=brain_trauma_Final$INDI_DSCM_NO_R
brain_trauma_Final$TBI<-"Y"
TBI<-brain_trauma_Final[,c("INDI_DSCM_NO_R","TBI")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI

data_7=merge(data_6,TBI,by="INDI_DSCM_NO_R", all.x=T)

# non_traumatic_brain_injury_CT_MRI_최종 결합자료
brain_ct_mri_non_trauma_Final=merge(mod_data_brain_ct_k_non_traum
a_Final,brain_MRI_non_trauma_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_non_trauma_Final=brain_ct_mri_non_trauma_Final[!duplicated(
brain_ct_mri_non_trauma_Final$INDI_DSCM_NO_R),]
colnames(mod_data_brain_ct_k_non_trauma_Final)
colnames(brain_MRI_non_trauma_Final)
brain_non_trauma_Final$non_trauma=NA
brain_non_trauma_Final$non_trauma=brain_non_trauma_Final$INDI_DSC
M_NO_R
brain_non_trauma_Final$non_trauma<-"Y"
Non_trauma<-brain_non_trauma_Final[,c("INDI_DSCM_NO_R","non_trauma")]

```

```

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma

data_8=merge(data_7,Non_trauma,by="INDI_DSCM_NO_R", all.x=T)

# Infarction_최종결합자료
brain_ct_mri_infarction_Final=merge(brain_ct_infarction_Final,brain_mri_infarction_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_infarction_Final=brain_ct_mri_infarction_Final[!duplicated(brain_ct_mri_infarction_Final$INDI_DSCM_NO_R),]
brain_infarction_Final$Infarction=NA
brain_infarction_Final$Infarction=NA
brain_infarction_Final$Infarction=brain_infarction_Final$INDI_DSCM_NO_R
brain_infarction_Final$Infarction<-"Y"
Infarction<-brain_infarction_Final[,c("INDI_DSCM_NO_R","Infarction")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction

data_9=merge(data_8,Infarction,by="INDI_DSCM_NO_R", all.x=T)

# SAH_최종결합자료
brain_ct_mri_SAH_Final=merge(brain_ct_SAH_Final,brain_MRI_SAH_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_SAH_Final=brain_ct_mri_SAH_Final[!duplicated(brain_ct_mri_SAH_Final$INDI_DSCM_NO_R),]
brain_SAH_Final$SAH=NA
brain_SAH_Final$SAH=brain_SAH_Final$INDI_DSCM_NO_R
brain_SAH_Final$SAH<-"Y"
SAH<-brain_SAH_Final[,c("INDI_DSCM_NO_R","SAH")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma

```

```

+Infarction+SAH

data_10=merge(data_9,SAH,by="INDI_DSCM_NO_R", all.x=T)

#SDH_최종결합자료
brain_ct_mri_SDH_Final=merge(brain_ct_SDH_Final,brain_MRI_SDH_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_SDH_Final=brain_ct_mri_SDH_Final[!duplicated(brain_ct_mri_SDH_Final$INDI_DSCM_NO_R),]
brain_SDH_Final$SDH=NA
brain_SDH_Final$SDH=brain_SDH_Final$INDI_DSCM_NO_R
brain_SDH_Final$SDH<-"Y"
SDH<-brain_SDH_Final[,c("INDI_DSCM_NO_R","SDH")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH

data_11=merge(data_10,SDH,by="INDI_DSCM_NO_R", all.x=T)

#ICH_최종결합자료
brain_ct_mri_ICH_Final=merge(brain_ct_ICH_Final,brain_mri_ICH_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_ICH_Final=brain_ct_mri_ICH_Final[!duplicated(brain_ct_mri_ICH_Final$INDI_DSCM_NO_R),]
brain_ICH_Final$ICH=NA
brain_ICH_Final$ICH=brain_ICH_Final$INDI_DSCM_NO_R
brain_ICH_Final$ICH<-"Y"
ICH<-brain_ICH_Final[,c("INDI_DSCM_NO_R","ICH")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH

data_12=merge(data_11,ICH,by="INDI_DSCM_NO_R", all.x=T)

```



```

#IVH_최종결합자료
brain_ct_mri_IVH_Final=merge(brain_ct_IVH_Final,brain_mri_IVH_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_IVH_Final=brain_ct_mri_IVH_Final[!duplicated(brain_ct_mri_IVH_Final$INDI_DSCM_NO_R),]
brain_IVH_Final$IVH=NA
brain_IVH_Final$IVH=brain_IVH_Final$INDI_DSCM_NO_R
brain_IVH_Final$IVH<-"Y"
IVH<-brain_IVH_Final[,c("INDI_DSCM_NO_R","IVH")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH

data_13=merge(data_12,IVH,by="INDI_DSCM_NO_R", all.x=T)

#Focal_small_최종결합자료
brain_ct_mri_Focal_small_Final=merge(brain_ct_Focal_small_Final,brain_MRI_Focal_small_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Focal_small_Final=brain_ct_mri_Focal_small_Final[!duplicated(brain_ct_mri_Focal_small_Final$INDI_DSCM_NO_R),]
brain_Focal_small_Final$Focal_small=NA
brain_Focal_small_Final$Focal_small=brain_Focal_small_Final$INDI_DSCM_NO_R
brain_Focal_small_Final$Focal_small<-"Y"
Focal_small<-brain_Focal_small_Final[,c("INDI_DSCM_NO_R","Focal_small")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small

data_14=merge(data_13,Focal_small,by="INDI_DSCM_NO_R", all.x=T)

#lesion_multifocal_large_Diffuse_최종결합자료

```

```

brain_ct_mri_multifocal_large_Final=merge(brain_ct_multifocal_lar
ge_Final,brain_mri_multifocal_large_Final,by="INDI_DSCM_NO_R",
all.x=T)
brain_multifocal_large_Final=brain_ct_mri_multifocal_large_Final[
!duplicated(brain_ct_mri_multifocal_large_Final$INDI_DSCM_NO_R),]
brain_multifocal_large_Final$multifocal_large=NA
brain_multifocal_large_Final$multifocal_large=brain_multifocal_la
rge_Final$INDI_DSCM_NO_R
brain_multifocal_large_Final$multifocal_large<-"Y"
multifocal_large<-brain_multifocal_large_Final[,c("INDI_DSCM_NO_R
","multifocal_large")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large

data_15=merge(data_14,multifocal_large,by="INDI_DSCM_NO_R",
all.x=T)

#Acute_acute_최종결합자료
brain_ct_mri_Acute_recent_Final=merge(brain_ct_Acute_recent_Final
,brain_mri_Acute_recent_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Acute_recent_Final=brain_ct_mri_Acute_recent_Final[!duplica
ted(brain_ct_mri_Acute_recent_Final$INDI_DSCM_NO_R),]
brain_Acute_recent_Final$Acute_recent=NA
brain_Acute_recent_Final$Acute_recent=brain_Acute_recent_Final$IN
DI_DSCM_NO_R
brain_Acute_recent_Final$Acute_recent<-"Y"
Acute_recent<-brain_Acute_recent_Final[,c("INDI_DSCM_NO_R","Acute
_recent")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent

```

```

data_16=merge(data_15,Acute_recent,by="INDI_DSCM_NO_R", all.x=T)

#lesion_Chronic and old_최종결합자료
brain_ct_mri_Chronic_old_Final=merge(brain_ct_Chronic_old_Final,brain_mri_Chronic_old_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Chronic_old_Final=brain_ct_mri_Chronic_old_Final[!duplicated(brain_ct_mri_Chronic_old_Final$INDI_DSCM_NO_R),]
brain_Chronic_old_Final$Chronic_old=NA
brain_Chronic_old_Final$Chronic_old=brain_Chronic_old_Final$INDI_DSCM_NO_R
brain_Chronic_old_Final$Chronic_old<-"Y"
Chronic_old<-brain_Chronic_old_Final[,c("INDI_DSCM_NO_R","Chronic_old")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old

data_17=merge(data_16,Chronic_old,by="INDI_DSCM_NO_R", all.x=T)

#Frontal_lobe_최종결합자료
brain_ct_mri_Frontal_Final=merge(brain_ct_Frontal_Final,brain_mri_Frontal_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Frontal_Final=brain_ct_mri_Frontal_Final[!duplicated(brain_ct_mri_Frontal_Final$INDI_DSCM_NO_R),]
brain_Frontal_Final$Frontal_Final=NA
brain_Frontal_Final$Frontal_Final=brain_Frontal_Final$INDI_DSCM_NO_R
brain_Frontal_Final$Frontal_Final<-"Y"
Frontal_Final<-brain_Frontal_Final[,c("INDI_DSCM_NO_R","Frontal_Final")]

#일산병원자료+장애등급+사망
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent

```

```

+Chronic_old+Frontal_Final

data_18=merge(data_17,Frontal_Final,by="INDI_DSCM_NO_R", all.x=T)

#Temporal_lobe_최종결합자료
brain_ct_mri_Temporal_Final=merge(brain_ct_Temporal_Final,brain_m
ri_Temporal_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Temporal_Final=brain_ct_mri_Temporal_Final[!duplicated(brai
n_ct_mri_Temporal_Final$INDI_DSCM_NO_R),]
brain_Temporal_Final$Temporal_Final=NA
brain_Temporal_Final$Temporal_Final=brain_Temporal_Final$INDI_DSC
M_NO_R
brain_Temporal_Final$Temporal_Final<-"Y"
Temporal_Final<-brain_Temporal_Final[,c("INDI_DSCM_NO_R","Tempora
l_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final

data_19=merge(data_18,Temporal_Final,by="INDI_DSCM_NO_R",
all.x=T)

#Parietal_lobe_최종결합자료
brain_ct_mri_Parietal_Final=merge(brain_ct_Parietal_Final,brain_m
ri_Parietal_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Parietal_Final=brain_ct_mri_Parietal_Final[!duplicated(brai
n_ct_mri_Parietal_Final$INDI_DSCM_NO_R),]
brain_Parietal_Final$Parietal_Final=NA
brain_Parietal_Final$Parietal_Final=brain_Parietal_Final$INDI_DSC
M_NO_R
brain_Parietal_Final$Parietal_Final<-"Y"
Parietal_Final<-brain_Parietal_Final[,c("INDI_DSCM_NO_R","Parieta
l_Final")]

```

```

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final

data_20=merge(data_19,Parietal_Final,by="INDI_DSCM_NO_R",
all.x=T)
#Occipital_lobe_최종결합자료
brain_ct_mri_Occipital_Final=merge(brain_ct_Occipital_Final,brain
_mri_Occipital_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Occipital_Final=brain_ct_mri_Occipital_Final[!duplicated(br
ain_ct_mri_Occipital_Final$INDI_DSCM_NO_R),]
brain_Occipital_Final$Occipital_Final=NA
brain_Occipital_Final$Occipital_Final=brain_Occipital_Final$INDI_
DSCM_NO_R
brain_Occipital_Final$Occipital_Final<-"Y"
Occipital_Final<-brain_Occipital_Final[,c("INDI_DSCM_NO_R","Occip
ital_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipital
_Final

data_21=merge(data_20,Occipital_Final,by="INDI_DSCM_NO_R",
all.x=T)

#Cerebellum_최종자료
brain_ct_mri_Cerebellum_Final=merge(brain_ct_Cerebellum_Final,bra
in_mri_Cerebellum_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Cerebellum_Final=brain_ct_mri_Cerebellum_Final[!duplicated(
brain_ct_mri_Cerebellum_Final$INDI_DSCM_NO_R),]
brain_Cerebellum_Final$Cerebellum_Final=NA
brain_Cerebellum_Final$Cerebellum_Final=brain_Cerebellum_Final$IN

```

```

DI_DSCM_NO_R
brain_Cerebellum_Final$Cerebellum_Final<-"Y"
Cerebellum_Final<-brain_Cerebellum_Final[,c("INDI_DSCM_NO_R", "Cer
ebellum_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final

data_22=merge(data_21,Cerebellum_Final,by="INDI_DSCM_NO_R",
all.x=T)

#BG_최종자료
brain_ct_mri_BG_Final=merge(brain_ct_BG_Final,brain_mri_BG_Final,
by="INDI_DSCM_NO_R", all.x=T)
brain_BG_Final=brain_ct_mri_BG_Final[!duplicated(brain_ct_mri_BG_
_Final$INDI_DSCM_NO_R),]
brain_BG_Final$BG_Final=NA
brain_BG_Final$BG_Final=brain_BG_Final$INDI_DSCM_NO_R
brain_BG_Final$BG_Final<-"Y"
BG_Final<-brain_BG_Final[,c("INDI_DSCM_NO_R", "BG_Final")]
#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final

data_23=merge(data_22,BG_Final,by="INDI_DSCM_NO_R", all.x=T)

#subcortex_최종자료
brain_ct_mri_subcortex_Final=merge(brain_ct_subcortex_Final,brain
_mri_subcortex_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_subcortex_Final=brain_ct_mri_subcortex_Final[!duplicated(br

```

```

ain_ct_mri_subcortex_Final$INDI_DSCM_NO_R),]
brain_subcortex_Final$subcortex_Final=NA
brain_subcortex_Final$subcortex_Final=brain_subcortex_Final$INDI_
DSCM_NO_R
brain_subcortex_Final$subcortex_Final<-"Y"
subcortex_Final<-brain_subcortex_Final[,c("INDI_DSCM_NO_R", "subco
rtex_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final

data_24=merge(data_23,subcortex_Final,by="INDI_DSCM_NO_R",
all.x=T)

#brain_stem_최종자료
brain_ct_mri_brain_stem_Final=merge(brain_ct_brain_stem_Final,bra
in_mri_brain_stem_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_brain_stem_Final=brain_ct_mri_brain_stem_Final[!duplicated(
brain_ct_mri_brain_stem_Final$INDI_DSCM_NO_R),]
brain_brain_stem_Final$brain_stem_Final=NA
brain_brain_stem_Final$brain_stem_Final=brain_brain_stem_Final$IN
DI_DSCM_NO_R
brain_brain_stem_Final$brain_stem_Final<-"Y"
brain_stem_Final<-brain_brain_stem_Final[,c("INDI_DSCM_NO_R", "bra
in_stem_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Final

```

```

data_25=merge(data_24,brain_stem_Final,by="INDI_DSCM_NO_R",
all.x=T)

#Thalamus_최종자료
brain_ct_mri_Thalamus_Final=merge(brain_ct_Thalamus_Final,brain_m
ri_Thalamus_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Thalamus_Final=brain_ct_mri_Thalamus_Final[!duplicated(brai
n_ct_mri_Thalamus_Final$INDI_DSCM_NO_R),]
brain_Thalamus_Final$Thalamus_Final=NA
brain_Thalamus_Final$Thalamus_Final=brain_Thalamus_Final$INDI_DSC
M_NO_R
brain_Thalamus_Final$Thalamus_Final<-"Y"
Thalamus_Final<-brain_Thalamus_Final[,c("INDI_DSCM_NO_R","Thalamu
s_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final

data_26=merge(data_25,Thalamus_Final,by="INDI_DSCM_NO_R",
all.x=T)

#ACA_최종자료
brain_ct_mri_ACA_Final=merge(brain_ct_ACA_Final,brain_mri_ACA_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_ACA_Final=brain_ct_mri_ACA_Final[!duplicated(brain_ct_mri_A
CA_Final$INDI_DSCM_NO_R),]
brain_ACA_Final$ACA_Final=NA
brain_ACA_Final$ACA_Final=brain_ACA_Final$INDI_DSCM_NO_R
brain_ACA_Final$ACA_Final<-"Y"
ACA_Final<-brain_ACA_Final[,c("INDI_DSCM_NO_R","ACA_Final")]

```



```

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final+ACA_Final

data_27=merge(data_26,ACA_Final,by="INDI_DSCM_NO_R", all.x=T)

#MCA_최종자료
brain_ct_mri_MCA_Final=merge(brain_ct_MCA_Final,brain_mri_MCA_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_MCA_Final=brain_ct_mri_MCA_Final[!duplicated(brain_ct_mri_M
CA_Final$INDI_DSCM_NO_R),]
brain_MCA_Final$MCA_Final=NA
brain_MCA_Final$MCA_Final=brain_MCA_Final$INDI_DSCM_NO_R
brain_MCA_Final$MCA_Final<-"Y"
MCA_Final<-brain_MCA_Final[,c("INDI_DSCM_NO_R","MCA_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final+ACA_Final+MCA_Final

data_28=merge(data_27,MCA_Final,by="INDI_DSCM_NO_R", all.x=T)

#PCA_최종자료
brain_ct_mri_PCA_Final=merge(brain_ct_PCA_Final,brain_mri_PCA_Fin
al,by="INDI_DSCM_NO_R", all.x=T)
brain_PCA_Final=brain_ct_mri_PCA_Final[!duplicated(brain_ct_mri_P
CA_Final$INDI_DSCM_NO_R),]
brain_PCA_Final$PCA_Final=NA
brain_PCA_Final$PCA_Final=brain_PCA_Final$INDI_DSCM_NO_R

```

```

brain_PCA_Final$PCA_Final<-"Y"
PCA_Final<-brain_PCA_Final[,c("INDI_DSCM_NO_R", "PCA_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final+ACA_Final+MCA_Final+PCA_Final

data_29=merge(data_28,PCA_Final,by="INDI_DSCM_NO_R", all.x=T)
#basilar_artery_최종자료
brain_ct_mri_basilar_artery_Final=merge(brain_ct_basilar_artery_F
inal,brain_mri_basilar_artery_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_basilar_artery_Final=brain_ct_mri_basilar_artery_Final[!dup
licated(brain_ct_mri_basilar_artery_Final$INDI_DSCM_NO_R),]
brain_basilar_artery_Final$ basilar_artery_Final=NA
brain_basilar_artery_Final$
basilar_artery_Final=brain_basilar_artery_Final$INDI_DSCM_NO_R
brain_basilar_artery_Final$ basilar_artery_Final<-"Y"
basilar_artery_Final<-brain_basilar_artery_Final[,c("INDI_DSCM_NO
_R", "basilar_artery_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final+ACA_Final+MCA_Final+PCA_Final+basilar_artery_Fina

data_30=merge(data_29,basilar_artery_Final,by="INDI_DSCM_NO_R",
all.x=T)

#Aneurysm_최종자료
brain_ct_mri_Aneurysm_Final=merge(brain_ct_Aneurysm_Final,brain_m

```

```

ri_Aneurysm_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Aneurysm_Final=brain_ct_mri_Aneurysm_Final[!duplicated(brai
n_ct_mri_Aneurysm_Final$INDI_DSCM_NO_R),]
brain_Aneurysm_Final$Aneurysm_Final=NA
brain_Aneurysm_Final$Aneurysm_Final=brain_Aneurysm_Final$INDI_DSC
M_NO_R
brain_Aneurysm_Final$Aneurysm_Final<-"Y"
Aneurysm_Final<-brain_Aneurysm_Final[,c("INDI_DSCM_NO_R","Aneurys
m_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final+ACA_Final+MCA_Final+PCA_Final+basilar_artery_Fin
al+Aneurysm_Final

data_31=merge(data_30,Aneurysm_Final,by="INDI_DSCM_NO_R",
all.x=T)

#Clipping_최종자료
brain_ct_mri_Clipping_Final=merge(brain_ct_Clipping_Final,brain_m
ri_Clipping_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_Clipping_Final=brain_ct_mri_Clipping_Final[!duplicated(brai
n_ct_mri_Clipping_Final$INDI_DSCM_NO_R),]
brain_Clipping_Final$Clipping_Final=NA
brain_Clipping_Final$Clipping_Final=brain_Clipping_Final$INDI_DSC
M_NO_R
brain_Clipping_Final$Clipping_Final<-"Y"
Clipping_Final<-brain_Clipping_Final[,c("INDI_DSCM_NO_R","
Clipping_Final")]

#일산병원자료+장애등급+사망자료
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma

```

```
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent  
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita  
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina  
l+Thalamus_Final+ACA_Final+MCA_Final+PCA_Final+basilar_artery_Fin  
al+Aneurysm_Final+Clipping_Final
```

```
data_32=merge(data_31,Aneurysm_Final,by="INDI_DSCM_NO_R",  
all.x=T)
```

```
#Coil_최종자료
```

```
brain_ct_mri_Coil_Final=merge(brain_ct_Coil_Final,brain_mri_Coil_  
Final,by="INDI_DSCM_NO_R", all.x=T)
```

```
brain_Coil_Final=brain_ct_mri_Coil_Final[!duplicated(brain_ct_mri_  
_Coil_Final$INDI_DSCM_NO_R),]
```

```
brain_Coil_Final$Coil_Final=NA
```

```
brain_Coil_Final$Coil_Final=brain_Coil_Final$INDI_DSCM_NO_R
```

```
brain_Coil_Final$Coil_Final<-"Y"
```

```
Coil_Final<-brain_Coil_Final[,c("INDI_DSCM_NO_R","Coil_Final")]
```

```
#일산병원자료+장애등급+사망자료
```

```
+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma  
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent  
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita  
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina  
l+Thalamus_Final+ACA_Final+MCA_Final+PCA_Final+basilar_artery_Fin  
al+Aneurysm_Final+Clipping_Final+Coil_Final
```

```
data_33=merge(data_32,Coil_Final,by="INDI_DSCM_NO_R", all.x=T)
```

```
#Craniectomy_최종자료
```

```
brain_ct_mri_craniectomy_Final=merge(brain_ct_craniectomy_Final,b  
rain_mri_craniectomy_Final,by="INDI_DSCM_NO_R", all.x=T)
```

```
brain_craniectomy_Final=brain_ct_mri_craniectomy_Final[!duplicate  
d(brain_ct_mri_craniectomy_Final$INDI_DSCM_NO_R),]
```

```
brain_craniectomy_Final$craniectomy_Final=NA
```

```

brain_craniectomy_Final$craniectomy_Final=brain_craniectomy_Final
$INDI_DSCM_NO_R
brain_craniectomy_Final$craniectomy_Final<-"Y"
craniectomy_Final<-brain_craniectomy_Final[,c("INDI_DSCM_NO_R", "c
raniectomy_Final")]

```

#일산병원자료+장애등급+사망자료

```

+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final+ACA_Final+MCA_Final+PCA_Final+basilar_artery_Fin
al+Aneurysm_Final+Clipping_Final+Coil_Final+craniectomy_Final
data_34=merge(data_33,craniectomy_Final,by="INDI_DSCM_NO_R",
all.x=T)

```

#Craniotomy\_최종자료

```

brain_ct_mri_craniotomy_Final=merge(brain_ct_craniotomy_Final,bra
in_mri_craniotomy_Final,by="INDI_DSCM_NO_R", all.x=T)
brain_craniotomy_Final=brain_ct_mri_craniotomy_Final[!duplicated(
brain_ct_mri_craniotomy_Final$INDI_DSCM_NO_R),]
brain_craniotomy_Final$craniotomy_Final=NA
brain_craniotomy_Final$craniotomy_Final=brain_craniotomy_Final$IN
DI_DSCM_NO_R
brain_craniotomy_Final$craniotomy_Final<-"Y"
craniotomy_Final<-brain_craniotomy_Final[,c("INDI_DSCM_NO_R", "cra
niotomy_Final")]

```

#일산병원자료+장애등급+사망자료

```

+gastrostomy+L_tube_feeding+Pneumonia+Tracheostomy+TBI+non_trauma
+Infarction+SAH+ICH+IVH+Focal_small+multifocal_large+Acute_recent
+Chronic_old+Frontal_Final+Temporal_Final+Parietal_Final+Occipita
l_Final+Cerebellum_Final+BG_Final+subcortex_Final+brain_stem_Fina
l+Thalamus_Final+ACA_Final+MCA_Final+PCA_Final+basilar_artery_Fin
al+Aneurysm_Final+Clipping_Final+Coil_Final+craniectomy_Final+cran

```

```

iotomy_Final

data_35=merge(data_34,craniotomy_Final,by="INDI_DSCM_NO_R",
all.x=T)
# LTC+공단자료+병원자료 결합

LTC_data_3_last_day= LTC_data_3 %>%
  group_by (INDI_DSCM_NO) %>%
  slice(which.max(APLY_YM))

write.csv (LTC_data_3, paste0(spath,"LTC_temp.csv"))
write.csv (LTC_data_3_last_day,
paste0(spath,"LTC_last_temp.csv"))

write.csv(data_35,paste0(spath, "data_35.csv"))
LTC_data_final=setnames(LTC_data_final,"INDI_DSCM_NO","INDI_DSCM_
NO_R")

data=merge(data_35,LTC_data_final,by="INDI_DSCM_NO_R", all.x=T)
write.csv(data,paste0(spath, "data.csv"))

hospital_data=read.csv("~/data/data_source/user_data/khs_wonnae_d
ata.csv",header = T, stringsAsFactors = FALSE)

data_final=merge(data,hospital_data,by="INDI_DSCM_NO_R", all.x=T)
write.csv(data_final,paste0(spath, "data_final.csv"))

colnames(data_final)

data_36=data_final[,c(2:7,9:54,57:76,80,82,84,86,88,90,92:122)]

colnames(data_36)

final_data=read.csv(paste0(spath,"final_data.csv"))
colnames(final_data)

```

```

nrow(final_data)
#5512

final_data$Ipwon=as.factor(final_data$Ipwon)
final_data_admission=final_data%>% filter(Ipwon==1)
final_data$ALBUMIN_temp = as.numeric(final_data$ALBUMIN)

final_data_admission=read.csv(paste0(spath,"final_data_admission.
csv"))
colnames(final_data_admission)
nrow(final_data_admission)
#4789
# 결측치 처리

#
final_data_admission$ALBUMIN=as.numeric(final_data_admission$ALBUMIN)
# mean(final_data_admission$ALBUMIN, na.rm=T)
#
#final_data_admission$ALBUMIN=ifelse(is.na(final_data_admission$A
LBUMIN),3.9, final_data_admission$ALBUMIN)
# table(is.na(final_data_admission$ALBUMIN))
#
# final_data_admission$BUN=as.numeric(final_data_admission$BUN)
# mean(final_data_admission$BUN, na.rm=T)
#
#final_data_admission$BUN=ifelse(is.na(final_data_admission$BUN),
17.7, final_data_admission$BUN)
# table(is.na(final_data_admission$BUN))
#
# final_data_admission$CR=as.numeric(final_data_admission$CR)
# mean(final_data_admission$CR, na.rm=T)
#
#
final_data_admission$CR=ifelse(is.na(final_data_admission$CR),1.1
0, final_data_admission$CR)

```

```

# table(is.na(final_data_admission$CR))
#
#
final_data_admission$GLUCOSE.AC.=as.numeric(final_data_admission$
GLUCOSE.AC.)
# mean(final_data_admission$GLUCOSE.AC., na.rm=T)
#
#final_data_admission$GLUCOSE.AC.=ifelse(is.na(final_data_admissi
on$GLUCOSE.AC.),130, final_data_admission$GLUCOSE.AC.)
# table(is.na(final_data_admission$GLUCOSE.AC.))
# final_data_admission$HB=as.numeric(final_data_admission$HB)
# mean(final_data_admission$HB, na.rm=T)
#
#
final_data_admission$HB=ifelse(is.na(final_data_admission$HB),13.
7, final_data_admission$HB)
# table(is.na(final_data_admission$HB))
#
#
final_data_admission$HBA1C=as.numeric(final_data_admission$HBA1C)
# mean(final_data_admission$HBA1C, na.rm=T)
#
#final_data_admission$HBA1C=ifelse(is.na(final_data_admission$HBA
1C),6.5, final_data_admission$HBA1C)
# table(is.na(final_data_admission$HBA1C))

write.csv(final_data_admission,paste0(spath,
"final_data_admission.csv"))

final_data=read.csv(paste0(spath,"final_data_admission.csv"))

final_data_admission_no_LTC=final_data_admission%>%
filter(LTC_YN==0)

write.csv(final_data_admission_LTC,paste0(spath,

```



```

"final_data_admission_no_LTC.csv"))

final_data_admission=final_data_admission[,c(6:11,13:52,81:93)]
colnames(final_data_admission)

final_data_admission$SIDO_NM=as.factor(final_data_admission$SIDO_NM)
final_data_admission$CITY_SIZE=as.factor(final_data_admission$CITY_SIZE)
final_data_admission$GAIBJA_TYPE=as.factor(final_data_admission$GAIBJA_TYPE)
final_data_admission$CMPR_DSB_GRADE=as.factor(final_data_admission$CMPR_DSB_GRADE)

result=mytable(LTC_YN~ipwon_period+Live_days+AGE+SEX_TYPE+SIDO_NM+CITY_SIZE+CMPR_DSB_GRADE, data=final_data_admission)

#2. The Prognosis of Patients with Long Term Care Service Machine Learning Analysis

if (FALSE)
{
  install.packages("class")
  install.packages("rpart")
  install.packages("nnet")
  install.packages("ipred")
  install.packages("caret")
  install.packages("caretEnsemble")
  install.packages("dplyr")
  install.packages("xgboost")
  install.packages("survival")
  install.packages("lava")
  install.packages("h2o")
  install.packages("lattice")
  install.packages("data.table")
  install.packages("MLevel")
}

```

```

}

install.packages("MLeval")
install.packages("installr")

# prerequisite

library(haven)
library(survival)
library(lava)
library(class)
library(rpart)
library(nnet)
library(ipred)
library(lubridate)
library(dplyr)
library(moonBook)
# machine_Learning_library

library(caret)
library(caretEnsemble)
library(doParallel)
library(data.table)
library(dplyr)
library(e1071)
library(gbm)
library(kernlab)
library(randomForest)
library(tidyverse)
library(xgboost)
library(h2o)
library(data.table)

path=~ /data/data_source/user_data/"
spath=paste0(path, "handled_data/")

```

```

path=~ /data/data_source/user_data/"
spath=paste0(path, "handled_data/")
# data_fusion (자격변수)

data=read.csv(paste0(spath,"final_LTC.csv"))

colnames(data)

# [1] "INDI_DSCM_NO_R" "MDCARE_STRT_DT"      "DTH_ASSMD_DT"
"ipwon_period" "Live_days"
# [6] "AGE"                "SEX"                "RVSN_ADDR_CD"
      "SIDO_NM"          "CITY_SIZE"
# [11] "GAIBJA_TYPE"      "CMPR_DSB_GRADE"
"MAIN_DSB_TYPE"      "CMPR_DSB_GRADE_L"    "MAIN_DSB_TYPE_L"
# [16] "DEATH_YN"        "gastrostomy"
"L_tube_feeding"    "pneumonia"          "T_tube"
# [21] "TBI"              "non_trauma"         "Infarction"
      "SAH"              "SDH"
# [26] "ICH"              "IVH"                "Focal_small"
      "multifocal_large"  "Acute_recent"
# [31] "Chronic_old"     "Frontal"            "Temporal"
      "Parietal"         "Occipital"
# [36] "Cerebellum"     "BG"                 "subcortex"
      "brain_stem"       "Thalamus"
# [41] "ACA"             "MCA"                "PCA"
      "basilar_artery"   "Aneurysm"
# [46] "Coil"            "craniectomy"       "craniotomy"
      "APFRM_TYPE_CD"    "LTC_OBJTR_TYPE_CD"
# [51] "LTC_REDU_TYPE"   "RCG_RCGT_SCORE"
"GJU_RCGT_GRADE_CD"  "DISM_JUDG_RSN_CD"   "RSDN_TP_CD"
# [56] "LVTGT_PRSN_CD"   "CAREPSN_CD"
"EVRDY_LIFE_ARFF_SCORE" "PCT_ARFF_SCORE"     "BHVR_ARFF_SCORE"

# [61] "TNDNC_ARFF_SCORE"  "REHBT_ARFF_SCORE"  "N01"
      "N02"              "N03"

```

```

# [66] "N04"          "N05"          "N06"
        "N07"          "N08"
# [71] "N09"          "N10"          "S10"
        "PE1"          "PE2"
# [76] "smoke_yn"      "drink_yn"     "ALBUMIN"
        "BUN"          "CR"
# [81] "GLUCOSE_AC"    "HB"           "HBA1C"
        "jaewon_GW"    "jaewon_ICU"
# [86] "total_cowork"  "pum_cowork"   "idm_cowork"
        "SORE"

```

```
Data_final=data[,c(4,6:7,9:89)]
```

```
colnames(Data_final)
```

```
#3. Data_format_assignment
```

```

Data_final$ipwon_period=as.numeric(Data_final$ipwon_period)
Data_final$AGE=as.numeric(Data_final$AGE)
Data_final$SEX=as.factor(Data_final$SEX)
Data_final$SIDO_NM=as.factor(Data_final$SIDO_NM)
Data_final$CITY_SIZE=as.factor(Data_final$CITY_SIZE)
Data_final$GAIBJA_TYPE=as.factor(Data_final$GAIBJA_TYPE)
Data_final$CMPR_DSB_GRADE=as.factor(Data_final$CMPR_DSB_GRADE)
Data_final$MAIN_DSB_TYPE=as.factor(Data_final$MAIN_DSB_TYPE)
Data_final$CMPR_DSB_GRADE_L=as.factor(Data_final$CMPR_DSB_GRADE_L)
Data_final$MAIN_DSB_TYPE_L=as.factor(Data_final$MAIN_DSB_TYPE_L)
Data_final$DEATH_YN=as.factor(Data_final$DEATH_YN)
Data_final$gastrostomy=as.factor(Data_final$gastrostomy)
Data_final$L_tube_feeding=as.factor(Data_final$L_tube_feeding)
Data_final$pneumonia=as.factor(Data_final$pneumonia)
Data_final$T_tube=as.factor(Data_final$T_tube)
Data_final$TBI=as.factor(Data_final$TBI)
Data_final$non_trauma=as.factor(Data_final$non_trauma)
Data_final$Infarction=as.factor(Data_final$Infarction)
Data_final$SAH=as.factor(Data_final$SAH)

```

```

Data_final$SDH=as.factor(Data_final$SDH)
Data_final$ICH=as.factor(Data_final$ICH)
Data_final$IVH=as.factor(Data_final$IVH)
Data_final$Focal_small=as.factor(Data_final$Focal_small)
Data_final$multifocal_large=as.factor(Data_final$multifocal_large)
Data_final$Acute_recent=as.factor(Data_final$Acute_recent)
Data_final$Chronic_old=as.factor(Data_final$Chronic_old)
Data_final$Frontal=as.factor(Data_final$Frontal)
Data_final$Temporal=as.factor(Data_final$Temporal)
Data_final$Parietal=as.factor(Data_final$Parietal)
Data_final$Occipital=as.factor(Data_final$Occipital)
Data_final$Cerebellum=as.factor(Data_final$Cerebellum)
Data_final$BG=as.factor(Data_final$BG)
Data_final$subcortex=as.factor(Data_final$subcortex)
Data_final$brain_stem=as.factor(Data_final$brain_stem)
Data_final$Thalamus=as.factor(Data_final$Thalamus)
Data_final$ACA=as.factor(Data_final$ACA)
Data_final$MCA=as.factor(Data_final$MCA)
Data_final$PCA=as.factor(Data_final$PCA)
Data_final$basilar_artery=as.factor(Data_final$basilar_artery)
Data_final$Aneurysm=as.factor(Data_final$Aneurysm)
Data_final$Coil=as.factor(Data_final$Coil)
Data_final$craniectomy=as.factor(Data_final$craniectomy)
Data_final$craniotomy=as.factor(Data_final$craniotomy)
Data_final$APFRM_TYPE_CD=as.factor(Data_final$APFRM_TYPE_CD)
Data_final$LTC_OBJTR_TYPE_CD=as.factor(Data_final$LTC_OBJTR_TYPE_CD)
Data_final$LTC_REDU_TYPE=as.factor(Data_final$LTC_REDU_TYPE)
Data_final$RCG_RCGT_SCORE=as.numeric(Data_final$RCG_RCGT_SCORE)
Data_final$GJU_RCGT_GRADE_CD=as.factor(Data_final$GJU_RCGT_GRADE_CD)
Data_final$DISM_JUDG_RSN_CD=as.factor(Data_final$DISM_JUDG_RSN_CD)
Data_final$RSDN_TP_CD=as.factor(Data_final$RSDN_TP_CD)
Data_final$LVTGT_PRSN_CD=as.factor(Data_final$LVTGT_PRSN_CD)
Data_final$CAREPSN_CD=as.factor(Data_final$CAREPSN_CD)
Data_final$EVRDY_LIFE_ARFF_SCORE=as.numeric(Data_final$EVRDY_LIFE
_ARFF_SCORE)

```

```

Data_final$PCT_ARFF_SCORE=as.numeric(Data_final$PCT_ARFF_SCORE)
Data_final$BHVR_ARFF_SCORE=as.numeric(Data_final$BHVR_ARFF_SCORE)
Data_final$TNDNC_ARFF_SCORE=as.numeric(Data_final$TNDNC_ARFF_SCORE)
Data_final$REHBT_ARFF_SCORE=as.numeric(Data_final$REHBT_ARFF_SCORE)
Data_final$N01=as.factor(Data_final$N01)
Data_final$N02=as.factor(Data_final$N02)
Data_final$N03=as.factor(Data_final$N03)
Data_final$N04=as.factor(Data_final$N04)
Data_final$N05=as.factor(Data_final$N05)
Data_final$N06=as.factor(Data_final$N06)
Data_final$N07=as.factor(Data_final$N07)
Data_final$N08=as.factor(Data_final$N08)
Data_final$N09=as.factor(Data_final$N09)
Data_final$N10=as.factor(Data_final$N10)
Data_final$S1=as.factor(Data_final$S1)
Data_final$PE1=as.factor(Data_final$PE1)
Data_final$PE2=as.factor(Data_final$PE2)
Data_final$smoke_yn=as.factor(Data_final$smoke_yn)
Data_final$drink_yn=as.factor(Data_final$drink_yn)
Data_final$ALBUMIN=as.numeric(Data_final$ALBUMIN)
Data_final$BUN=as.numeric(Data_final$BUN)
Data_final$CR=as.numeric(Data_final$CR)
Data_final$GLUCOSE_AC=as.numeric(Data_final$GLUCOSE_AC)
Data_final$HB=as.numeric(Data_final$HB)
Data_final$HBA1C=as.numeric(Data_final$HBA1C)
Data_final$jaewon_GW=as.numeric(Data_final$jaewon_GW)
Data_final$jaewon_ICU=as.numeric(Data_final$jaewon_ICU)
Data_final$total_cowork=as.numeric(Data_final$total_cowork)
Data_final$pum_cowork=as.numeric(Data_final$pum_cowork)
Data_final$idm_cowork=as.numeric(Data_final$idm_cowork)
Data_final$SORE=as.factor(Data_final$SORE)

nearZeroVar(Data_final, saveMetrics = TRUE)

# freqRatio percentUnique zeroVar  nzv

```

# ipwon_period	1.167883	7.17781403	FALSE	FALSE
# AGE	1.166667	3.20826536	FALSE	FALSE
# SEX	1.275990	0.10875476	FALSE	FALSE
# SIDO_NM	5.834025	0.87003806	FALSE	FALSE
# CITY_SIZE	3.369565	0.43501903	FALSE	FALSE
# GAIBJA_TYPE	1.357143	1.14192496	FALSE	FALSE
# CMPR_DSB_GRADE	5.241758	0.48939641	FALSE	FALSE
# MAIN_DSB_TYPE	2.298795	0.70690593	FALSE	FALSE
# CMPR_DSB_GRADE_L	2.687324	0.38064165	FALSE	FALSE
# MAIN_DSB_TYPE_L	2.153499	0.65252855	FALSE	FALSE
# DEATH_YN	1.096921	0.10875476	FALSE	FALSE
# gastrostomy	58.322581	0.10875476	FALSE	TRUE
# L_tube_feeding	1.752994	0.10875476	FALSE	FALSE
# pneumonia	1.512295	0.10875476	FALSE	FALSE
# T_tube	29.147541	0.10875476	FALSE	TRUE
# TBI	20.635294	0.10875476	FALSE	TRUE
# non_trauma	121.600000	0.10875476	FALSE	TRUE
# Infarction	3.237327	0.10875476	FALSE	FALSE
# SAH	12.723881	0.10875476	FALSE	FALSE
# SDH	9.104396	0.10875476	FALSE	FALSE
# ICH	2.879747	0.10875476	FALSE	FALSE
# IVH	11.770833	0.10875476	FALSE	FALSE
# Focal_small	1.736607	0.10875476	FALSE	FALSE
# multifocal_large	1.426121	0.10875476	FALSE	FALSE
# Acute_recent	3.266821	0.10875476	FALSE	FALSE
# Chronic_old	2.509542	0.10875476	FALSE	FALSE
# Frontal	0.000000	0.05437738	TRUE	TRUE
# Temporal	9.754386	0.10875476	FALSE	FALSE
# Parietal	261.714286	0.10875476	FALSE	TRUE
# Occipital	11.682759	0.10875476	FALSE	FALSE
# Cerebellum	3.917112	0.10875476	FALSE	FALSE
# BG	1.063973	0.10875476	FALSE	FALSE
# subcortex	305.500000	0.10875476	FALSE	TRUE
# brain_stem	7.101322	0.10875476	FALSE	FALSE

# Thalamus	4.746875	0.10875476	FALSE	FALSE
# ACA	12.827068	0.10875476	FALSE	FALSE
# MCA	2.489564	0.10875476	FALSE	FALSE
# PCA	10.011976	0.10875476	FALSE	FALSE
# basilar_artery	20.383721	0.10875476	FALSE	TRUE
# Aneurysm	7.715640	0.10875476	FALSE	FALSE
# Coil	58.322581	0.10875476	FALSE	TRUE
# craniectomy	38.978261	0.10875476	FALSE	TRUE
# craniotomy	27.292308	0.10875476	FALSE	TRUE
# APFRM_TYPE_CD	1.741710	0.27188690	FALSE	FALSE
# LTC_OBJTR_TYPE_CD	4.482877	0.21750952	FALSE	FALSE
# LTC_REDU_TYPE	5.460251	0.38064165	FALSE	FALSE
# RCG_RCGT_SCORE	1.640000	26.10114192	FALSE	FALSE
# GJU_RCGT_GRADE_CD	1.411765	0.59815117	FALSE	FALSE
# DISM_JUDG_RSN_CD	17.197917	0.27188690	FALSE	FALSE
# RSDN_TP_CD	3.847095	0.43501903	FALSE	FALSE
# LVTGT_PRSN_CD	1.121134	0.97879282	FALSE	FALSE
# CAREPSN_CD	1.562893	0.81566069	FALSE	FALSE
# EVRDY_LIFE_ARFF_SCORE	1.395604	1.35943448	FALSE	FALSE
# PCT_ARFF_SCORE	1.071661	0.43501903	FALSE	FALSE
# BHVR_ARFF_SCORE	5.512195	0.59815117	FALSE	FALSE
# TNDNC_ARFF_SCORE	7.234450	0.38064165	FALSE	FALSE
# REHBT_ARFF_SCORE	2.130435	1.08754758	FALSE	FALSE
# N01	67.111111	0.10875476	FALSE	TRUE
# N02	35.780000	0.10875476	FALSE	TRUE
# N03	67.111111	0.10875476	FALSE	TRUE
# N04	22.883117	0.10875476	FALSE	TRUE
# N05	13.480315	0.10875476	FALSE	FALSE
# N06	53.088235	0.10875476	FALSE	TRUE
# N07	14.073770	0.10875476	FALSE	FALSE
# N08	305.500000	0.10875476	FALSE	TRUE
# N09	26.044118	0.10875476	FALSE	TRUE
# N1	261.714286	0.10875476	FALSE	TRUE



```

# S1          2.286624      0.16313214      FALSE FALSE
# PE1        1.462214      0.21750952      FALSE FALSE
# PE2        1.823956      0.21750952      FALSE FALSE
# smoke_yn   1.328691      0.16313214      FALSE FALSE
# drink_yn   4.451178      0.16313214      FALSE FALSE
# ALBUMIN    1.016949      2.01196302      FALSE FALSE
# BUN        1.250000      21.53344209     FALSE FALSE
# CR         1.114286      15.44317564     FALSE FALSE
# GLUCOSE_AC 2.666667      12.72430669     FALSE FALSE
# HB         1.000000      6.96030451     FALSE FALSE
# HBA1C      3.419048      4.94834149     FALSE FALSE
# jaewon_GW  1.125000      5.11147363     FALSE FALSE
# jaewon_ICU 35.756757     5.65524742     FALSE TRUE
# total_cowork 1.208791      3.58890701     FALSE FALSE
# pum_cowork  9.472050      0.65252855     FALSE FALSE
# idm_cowork 15.568627     0.81566069     FALSE FALSE
# SORE       15.274336     0.10875476     FALSE FALSE

predata=Data_final[,-nearZeroVar(Data_final)] # near Zero
Variance를 제거

colnames(predata)

# [1] "ipwon_period"          "AGE"              "SEX"
      "SIDO_NM"           "CITY_SIZE"
# [6] "GAIBJA_TYPE"         "CMPR_DSB_GRADE"
"MAIN_DSB_TYPE"        "CMPR_DSB_GRADE_L"  "MAIN_DSB_TYPE_L"

# [11] "DEATH_YN"            "L_tube_feeding"   "pneumonia"
      "Infarction"        "SAH"
# [16] "SDH"                "ICH"              "IVH"
      "Focal_small"        "multifocal_large"
# [21] "Acute_recent"        "Chronic_old"      "Temporal"
      "Occipital"         "Cerebellum"

```

```

# [26] "BG"                "brain_stem"          "Thalamus"
      "ACA"                "MCA"
# [31] "PCA"                "Aneurysm"
"APFRM_TYPE_CD"          "LTC_OBJTR_TYPE_CD"   "LTC_REDU_TYPE"

# [36] "RCG_RCGT_SCORE"      "GJU_RCGT_GRADE_CD"
"DISM_JUDG_RSN_CD"      "RSDN_TP_CD"          "LVTGT_PRSN_CD"

# [41] "CAREPSN_CD"          "EVRDY_LIFE_ARFF_SCORE"
"PCT_ARFF_SCORE"        "BHVR_ARFF_SCORE"     "TNDNC_ARFF_SCORE"

# [46] "REHBT_ARFF_SCORE"    "N05"                  "N07"
      "S1"                "PE1"
# [51] "PE2"                "smoke_yn"            "drink_yn"
      "ALBUMIN"          "BUN"
# [56] "CR"                "GLUCOSE_AC"          "HB"
      "HBA1C"            "jaewon_GW"
# [61] "total_cowork"       "pum_cowork"          "idm_cowork"
      "SORE"

findCorrelation(cor(predata[,c(1:2,36,42:46,54:63)]),cutoff=.90,
names=TRUE)

# [1] "EVRDY_LIFE_ARFF_SCORE"

predata=subset(predata,select=-c(EVRDY_LIFE_ARFF_SCORE))

colnames(predata)

# [1] "ipwon_period"        "AGE"                  "SEX"
"SIDO_NM"              "CITY_SIZE"           "GAIBJA_TYPE"
# [7] "CMPR_DSB_GRADE"     "MAIN_DSB_TYPE"       "CMPR_DSB_GRADE_L"
"MAIN_DSB_TYPE_L"     "DEATH_YN"            "L_tube_feeding"
# [13] "pneumonia"          "Infarction"          "SAH"
"SDH"                  "ICH"                  "IVH"

```

```

# [19] "Focal_small"      "multifocal_large" "Acute_recent"
"Chronic_old"      "Temporal"          "Occipital"
# [25] "Cerebellum"      "BG"                "brain_stem"
"Thalamus"         "ACA"               "MCA"
# [31] "PCA"              "Aneurysm"          "APFRM_TYPE_CD"
"LTC_OBJTR_TYPE_CD" "LTC_REDU_TYPE"     "RCG_RCGT_SCORE"
# [37] "GJU_RCGT_GRADE_CD" "DISM_JUDG_RSN_CD" "RSDN_TP_CD"
"LVTGT_PRSN_CD"    "CAREPSN_CD"        "PCT_ARFF_SCORE"
# [43] "BHVR_ARFF_SCORE"  "TNDNC_ARFF_SCORE" "REHBT_ARFF_SCORE"
"N05"              "N07"               "S1"
# [49] "PE1"              "PE2"               "smoke_yn"
"drink_yn"         "ALBUMIN"           "BUN"
# [55] "CR"               "GLUCOSE_AC"        "HB"
"HBA1C"            "jaewon_GW"         "total_cowork"
# [61] "pum_cowork"       "idm_cowork"        "SORE"

st_model=preProcess(predata[,c(1:63)],method=c("center","scale"))
data=predict(st_model, predata)

data=as.data.frame(data)

colnames(data)
data=na.omit(data)

ohe_feats=c('ipwon_period','AGE','SEX','SIDO_NM','CITY_SIZE','GAI
BJA_TYPE','CMPR_DSB_GRADE','MAIN_DSB_TYPE','CMPR_DSB_GRADE_L','MA
IN_DSB_TYPE_L','L_tube_feeding','pneumonia','Infarction','SAH','S
DH','ICH','IVH',
'Focal_small','multifocal_large','Acute_recent','Chronic_old','Te
mporal','Occipital','Cerebellum','BG','brain_stem','Thalamus','AC
A','MCA','PCA','Aneurysm','APFRM_TYPE_CD','LTC_OBJTR_TYPE_CD','LT
C_REDU_TYPE','RCG_RCGT_SCORE','GJU_RCGT_GRADE_CD','DISM_JUDG_RSN_
CD','RSDN_TP_CD','LVTGT_PRSN_CD','CAREPSN_CD','PCT_ARFF_SCORE','B
HVR_ARFF_SCORE','TNDNC_ARFF_SCORE','REHBT_ARFF_SCORE','N05','N07'
,'S1','PE1','PE2','smoke_yn','drink_yn','ALBUMIN','BUN','CR','GLU

```

```
COSE_AC', 'HB', 'HBA1C', 'jaewon_GW', 'total_cowork', 'pum_cowork', 'id
m_cowork')
```

```
dummies =
dummyVars(~ipwon_period+AGE+SEX+SIDO_NM+CITY_SIZE+GAIBJA_TYPE+CMP
R_DSB_GRADE+MAIN_DSB_TYPE+CMPR_DSB_GRADE_L+MAIN_DSB_TYPE_L+L_tube
_feeding+pneumonia+Infarction+SAH+SDH+ICH+IVH+Focal_small+multifo
cal_large+Acute_recent+Chronic_old+Temporal+Occipital+Cerebellum+
BG+brain_stem+Thalamus+ACA+MCA+PCA+Aneurysm+APFRM_TYPE_CD+LTC_OBJ
TR_TYPE_CD+LTC_REDU_TYPE+RCG_RCGT_SCORE+GJU_RCGT_GRADE_CD+DISM_JU
DG_RSN_CD
+RSDN_TP_CD+LVTGT_PRSN_CD+CAREPSN_CD+PCT_ARFF_SCORE
+BHVR_ARFF_SCORE+TNDNC_ARFF_SCORE+REHBT_ARFF_SCORE+N05
+N07+S1+PE1+PE2+smoke_yn+drink_yn
+ALBUMIN+BUN+CR+GLUCOSE_AC+HB+HBA1C+jaewon_GW+total_cowork+pum_co
work+idm_cowork, data=data)
```

```
df_ohe <- as.data.frame(predict(dummies, newdata = data))
df_combined <- cbind(data[, -c(which(colnames(data) %in%
ohe_feats))], df_ohe)
Data=as.data.table(df_combined)
```

```
colnames(Data)
# [1] "DEATH_YN"          "SORE"              "ipwon_period"
"AGE"              "SEX.1"            "SEX.2"
# [7] "SIDO_NM.1"        "SIDO_NM.2"        "SIDO_NM.3"
"SIDO_NM.4"        "SIDO_NM.5"        "SIDO_NM.6"
# [13] "SIDO_NM.7"        "SIDO_NM.9"        "SIDO_NM.10"
"SIDO_NM.11"       "SIDO_NM.12"       "SIDO_NM.13"
# [19] "SIDO_NM.14"       "SIDO_NM.15"       "SIDO_NM.16"
"SIDO_NM.17"       "CITY_SIZE.1"      "CITY_SIZE.2"

# [25] "CITY_SIZE.3"      "CITY_SIZE.4"      "CITY_SIZE.5"
"CITY_SIZE.6"      "CITY_SIZE.7"      "CITY_SIZE.8"
```

```

# [31] "GAIBJA_TYPE.0"      "GAIBJA_TYPE.1"      "GAIBJA_TYPE.2"
      "GAIBJA_TYPE.3"      "GAIBJA_TYPE.4"      "GAIBJA_TYPE.5"
# [37] "GAIBJA_TYPE.6"      "GAIBJA_TYPE.7"      "GAIBJA_TYPE.8"
      "GAIBJA_TYPE.9"      "GAIBJA_TYPE.10"     "GAIBJA_TYPE.11"
# [43] "GAIBJA_TYPE.12"     "GAIBJA_TYPE.13"
      "GAIBJA_TYPE.14"     "GAIBJA_TYPE.15"     "GAIBJA_TYPE.16"
      "GAIBJA_TYPE.17"
# [49] "GAIBJA_TYPE.18"     "GAIBJA_TYPE.19"
      "GAIBJA_TYPE.20"     "CMPR_DSB_GRADE.0"   "CMPR_DSB_GRADE.1"
      "CMPR_DSB_GRADE.2"
# [55] "CMPR_DSB_GRADE.3"     "CMPR_DSB_GRADE.4"
      "CMPR_DSB_GRADE.5"   "CMPR_DSB_GRADE.6"   "MAIN_DSB_TYPE.0"
      "MAIN_DSB_TYPE.1"
# [61] "MAIN_DSB_TYPE.2"     "MAIN_DSB_TYPE.3"
      "MAIN_DSB_TYPE.4"     "MAIN_DSB_TYPE.5"     "MAIN_DSB_TYPE.6"
      "MAIN_DSB_TYPE.8"
# [67] "MAIN_DSB_TYPE.9"     "MAIN_DSB_TYPE.10"
      "MAIN_DSB_TYPE.11"   "MAIN_DSB_TYPE.12"   "MAIN_DSB_TYPE.14"
      "CMPR_DSB_GRADE_L.0"
# [73] "CMPR_DSB_GRADE_L.1"   "CMPR_DSB_GRADE_L.2"
      "CMPR_DSB_GRADE_L.3" "CMPR_DSB_GRADE_L.4" "CMPR_DSB_GRADE_L.5"
      "CMPR_DSB_GRADE_L.6"
# [79] "MAIN_DSB_TYPE_L.0"    "MAIN_DSB_TYPE_L.1"
      "MAIN_DSB_TYPE_L.2"   "MAIN_DSB_TYPE_L.3"   "MAIN_DSB_TYPE_L.4"
      "MAIN_DSB_TYPE_L.5"
# [85] "MAIN_DSB_TYPE_L.6"    "MAIN_DSB_TYPE_L.8"
      "MAIN_DSB_TYPE_L.9"   "MAIN_DSB_TYPE_L.10"  "MAIN_DSB_TYPE_L.11"
      "MAIN_DSB_TYPE_L.14"
# [91] "L_tube_feeding.0"    "L_tube_feeding.1"   "pneumonia.0"
      "pneumonia.1"        "Infarction.0"        "Infarction.1"
# [97] "SAH.0"                "SAH.1"                "SDH.0"
      "SDH.1"                "ICH.0"                "ICH.1"
# [103] "IVH.0"                "IVH.1"                "Focal_small.0"
      "Focal_small.1"       "multifocal_large.0"
      "multifocal_large.1"

```

```

# [109] "Acute_recent.0"      "Acute_recent.1"
"Chronic_old.0"      "Chronic_old.1"      "Temporal.0"
"Temporal.1"
# [115] "Occipital.0"      "Occipital.1"      "Cerebellum.0"
      "Cerebellum.1"      "BG.0"      "BG.1"
# [121] "brain_stem.0"      "brain_stem.1"      "Thalamus.0"
      "Thalamus.1"      "ACA.0"      "ACA.1"
# [127] "MCA.0"      "MCA.1"      "PCA.0"
      "PCA.1"      "Aneurysm.0"      "Aneurysm.1"
# [133] "APFRM_TYPE_CD.1"      "APFRM_TYPE_CD.2"
"APFRM_TYPE_CD.3"      "APFRM_TYPE_CD.4"      "APFRM_TYPE_CD.5"
"LTC_OBJTR_TYPE_CD"
# [139] "LTC_REDU_TYPE.A"      "LTC_REDU_TYPE.B"
"LTC_REDU_TYPE.C"      "LTC_REDU_TYPE.D"      "LTC_REDU_TYPE.E"
"LTC_REDU_TYPE.F"
# [145] "LTC_REDU_TYPE.z"      "RCG_RCGT_SCORE"
"GJU_RCGT_GRADE_CD.1" "GJU_RCGT_GRADE_CD.2" "GJU_RCGT_GRADE_CD.3"
"GJU_RCGT_GRADE_CD.4"
# [151] "GJU_RCGT_GRADE_CD.5" "GJU_RCGT_GRADE_CD.6"
"GJU_RCGT_GRADE_CD.A" "GJU_RCGT_GRADE_CD.B" "GJU_RCGT_GRADE_CD.C"
"GJU_RCGT_GRADE_CD.Y"
# [157] "GJU_RCGT_GRADE_CD.Z" "DISM_JUDG_RSN_CD.0"
"DISM_JUDG_RSN_CD.1" "DISM_JUDG_RSN_CD.2" "DISM_JUDG_RSN_CD.4"
"DISM_JUDG_RSN_CD.5"
# [163] "RSDN_TP_CD.1"      "RSDN_TP_CD.2"      "RSDN_TP_CD.3"
      "RSDN_TP_CD.4"      "RSDN_TP_CD.5"      "RSDN_TP_CD.7"
# [169] "RSDN_TP_CD.9"      "RSDN_TP_CD.z"
"LVTGT_PRSN_CD.1"      "LVTGT_PRSN_CD.2"      "LVTGT_PRSN_CD.3"
"LVTGT_PRSN_CD.4"
# [175] "LVTGT_PRSN_CD.5"      "LVTGT_PRSN_CD.6"
"LVTGT_PRSN_CD.7"      "LVTGT_PRSN_CD.8"      "LVTGT_PRSN_CD.9"
"LVTGT_PRSN_CD.A"
# [181] "LVTGT_PRSN_CD.B"      "LVTGT_PRSN_CD.C"
"LVTGT_PRSN_CD.D"      "LVTGT_PRSN_CD.E"      "LVTGT_PRSN_CD.F"
"LVTGT_PRSN_CD.G"

```

```

# [187] "LVTGT_PRSN_CD.J"      "LVTGT_PRSN_CD.z"      "CAREPSN_CD.0"
      "CAREPSN_CD.1"      "CAREPSN_CD.2"      "CAREPSN_CD.3"
# [193] "CAREPSN_CD.4"      "CAREPSN_CD.5"      "CAREPSN_CD.6"
      "CAREPSN_CD.7"      "CAREPSN_CD.8"      "CAREPSN_CD.9"
# [199] "CAREPSN_CD.10"      "CAREPSN_CD.11"
"CAREPSN_CD.12"      "CAREPSN_CD.13"      "CAREPSN_CD.14"
"PCT_ARFF_SCORE"
# [205] "BHVR_ARFF_SCORE"      "TNDNC_ARFF_SCORE"
"REHBT_ARFF_SCORE"      "N05.0"      "N05.1"
"N07.0"
# [211] "N07.1"      "S1.1"      "S1.2"
      "S1.3"      "PE1.1"      "PE1.2"
# [217] "PE1.3"      "PE1.4"      "PE2.1"
      "PE2.2"      "PE2.3"      "PE2.4"
# [223] "smoke_yn.1"      "smoke_yn.2"      "smoke_yn.3"
      "drink_yn.1"      "drink_yn.2"      "drink_yn.3"

# [229] "ALBUMIN"      "BUN"      "CR"
      "GLUCOSE_AC"      "HB"      "HBA1C"
# [235] "jaewon_GW"      "total_cowork"      "pum_cowork"
      "idm_cowork"

# Data splitting

seed=0923
set.seed(seed)

ind=sample(2, nrow(Data), replace=T, prob=c(0.7,0.3))
trainmc<-Data[ind==1,]
testmc<-Data[ind==2,]

#Multi-class classification

ctrl <- trainControl(method="repeatedcv", number=10, repeats=5)
metric <- "Accuracy"

```

```

nn_fit <- train(DEATH_YN~., data = trainmc, method = "nnet",
trControl=ctrl, metric=metric, tuneLength = 5,
verbose=FALSE,na.action = na.omit)

nn_fit

# Neural Network(NN)
#
# 1305 samples
# 237 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1174, 1174, 1175, 1174, 1175, 1174, ...
# Resampling results across tuning parameters:
#
#   size decay Accuracy  Kappa
# 1    0e+00  0.7578960  0.5137118
# 1    1e-04  0.7701054  0.5393119
# 1    1e-03  0.7713090  0.5397254
# 1    1e-02  0.7811705  0.5608433
# 1    1e-01  0.7971106  0.5930881
# 3    0e+00  0.7698221  0.5383056
# 3    1e-04  0.7666067  0.5320510
# 3    1e-03  0.7670705  0.5329404
# 3    1e-02  0.7691844  0.5370984
# 3    1e-01  0.7673466  0.5334804
# 5    0e+00      NaN      NaN
# 5    1e-04      NaN      NaN
# 5    1e-03      NaN      NaN
# 5    1e-02      NaN      NaN
# 5    1e-01      NaN      NaN
# 7    0e+00      NaN      NaN
# 7    1e-04      NaN      NaN

```



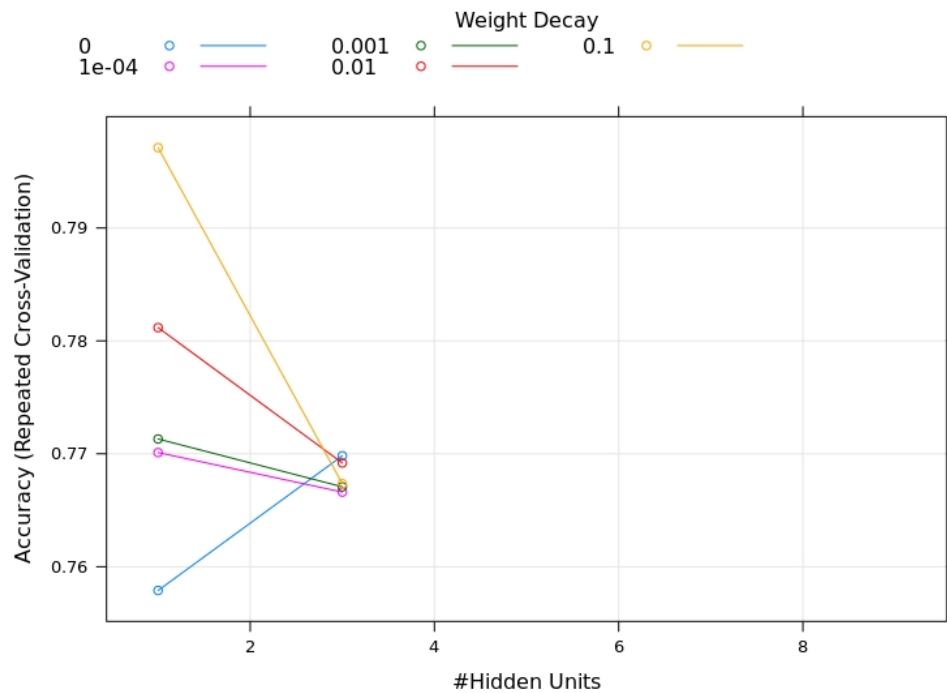
```
# 7 1e-03 NaN NaN
# 7 1e-02 NaN NaN
# 7 1e-01 NaN NaN
# 9 0e+00 NaN NaN
# 9 1e-04 NaN NaN
# 9 1e-03 NaN NaN
# 9 1e-02 NaN NaN
# 9 1e-01 NaN NaN
```

```
#
```

```
# Accuracy was used to select the optimal model using the largest value.
```

```
# The final values used for the model were size = 1 and decay = 0.1.
```

```
plot(nn_fit)
```



```
test_pred_nn <- predict(nn_fit, newdata = testmc)
confusionMatrix (test_pred_nn, testmc$DEATH_YN)
# Confusion Matrix and Statistics
```

```

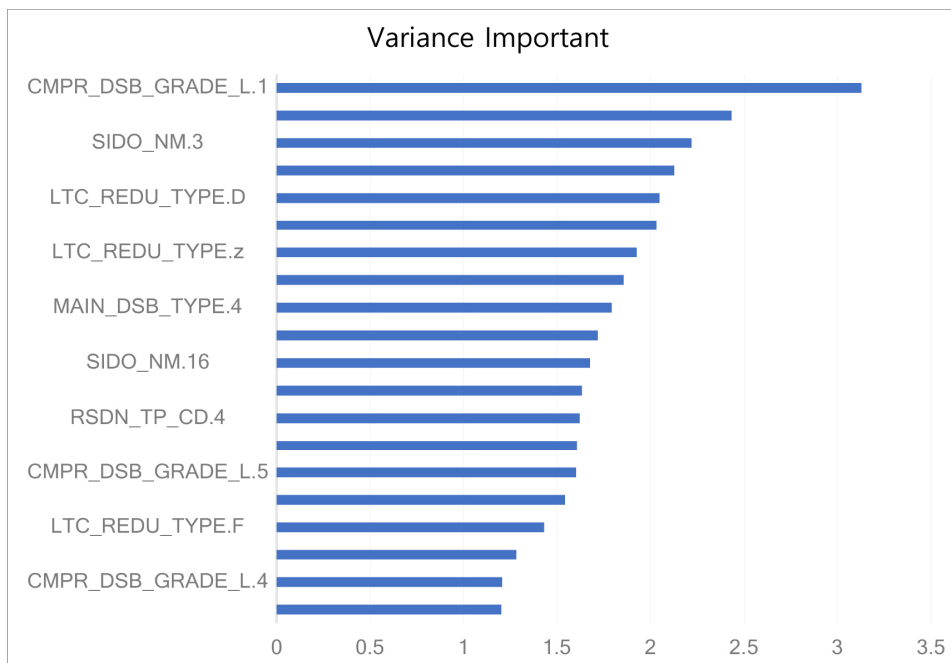
#
# Reference
# Prediction   N   Y
# N 205  49
# Y  49 230
#
# Accuracy : 0.8161
# 95% CI : (0.7806, 0.8481)
# No Information Rate : 0.5235
# P-Value [Acc > NIR] : <2e-16
#
# Kappa : 0.6315
# McNemar's Test P-Value : 1
#
#           Sensitivity : 0.8071
#           Specificity : 0.8244
#           Pos Pred Value : 0.8071
#           Neg Pred Value : 0.8244
#           Prevalence : 0.4765
#           Detection Rate : 0.3846
#           Detection Prevalence : 0.4765
#           Balanced Accuracy : 0.8157
#
# 'Positive' Class : N

imp<-varImp(nn_fit, scale =FALSE)
imp

# nnet variable importance
#
# only 20 most important variables shown (out of 237)
#
# Overall
# CMPR_DSB_GRADE_L.1   3.128
# CAREPSN_CD.2        2.434

```

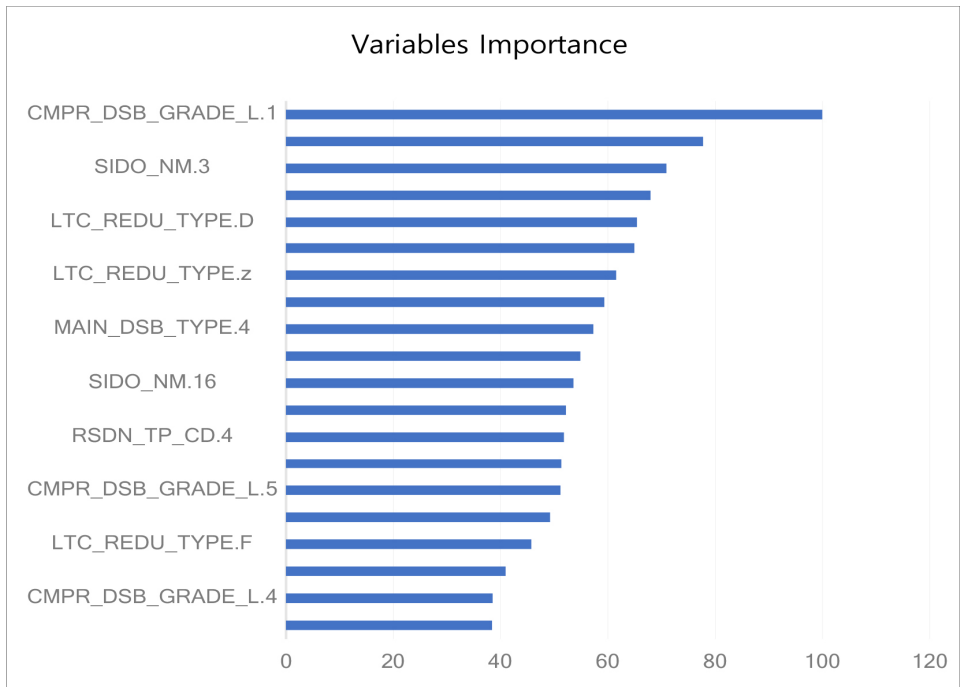
# SIDO_NM.3	2.220
# CMPR_DSB_GRADE_L.2	2.127
# LTC_REDU_TYPE.D	2.048
# LVTGT_PRSN_CD.3	2.032
# LTC_REDU_TYPE.z	1.926
# CMPR_DSB_GRADE_L.3	1.857
# MAIN_DSB_TYPE.4	1.793
# MAIN_DSB_TYPE.5	1.718
# SIDO_NM.16	1.677
# GAIBJA_TYPE.0	1.633
# RSDN_TP_CD.4	1.622
# MAIN_DSB_TYPE_L.4	1.607
# CMPR_DSB_GRADE_L.5	1.602
# LTC_REDU_TYPE.C	1.542
# LTC_REDU_TYPE.F	1.431
# MAIN_DSB_TYPE_L.6	1.282
# CMPR_DSB_GRADE_L.4	1.206
# CAREPSN_CD.3	1.202



```

imp<-varImp(nn_fit, scale =TRUE)
imp
# nnet variable importance
# only 20 most important variables shown (out of 237)
#
# Overall
# CMPR_DSB_GRADE_L.1  100.00
# CAREPSN_CD.2       77.79
# SIDO_NM.3          70.95
# CMPR_DSB_GRADE_L.2  67.98
# LTC_REDU_TYPE.D    65.45
# LVTGT_PRSN_CD.3    64.96
# LTC_REDU_TYPE.z    61.55
# CMPR_DSB_GRADE_L.3  59.36
# MAIN_DSB_TYPE.4    57.30
# MAIN_DSB_TYPE.5    54.91
# SIDO_NM.16         53.59
# GAIBJA_TYPE.0     52.19
# RSDN_TP_CD.4       51.84
# MAIN_DSB_TYPE_L.4  51.34
# CMPR_DSB_GRADE_L.5  51.20
# LTC_REDU_TYPE.C    49.26
# LTC_REDU_TYPE.F    45.72
# MAIN_DSB_TYPE_L.6  40.96
# CMPR_DSB_GRADE_L.4  38.53
# CAREPSN_CD.3       38.40

```



```

set.seed(seed)
svm_fit <- caret::train(DEATH_YN ~., data = trainmc, method =
"svmRadial",trControl=ctrl, metric=metric, tuneLength =
10,importance=TRUE,proximity=TRUE, na.action = na.omit)

svm_fit

# support vector machines with Radial Basis Function Kernel
#
# 1305 samples
# 237 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1174, 1174, 1174, 1173, 1175, 1175, ...
# Resampling results across tuning parameters:

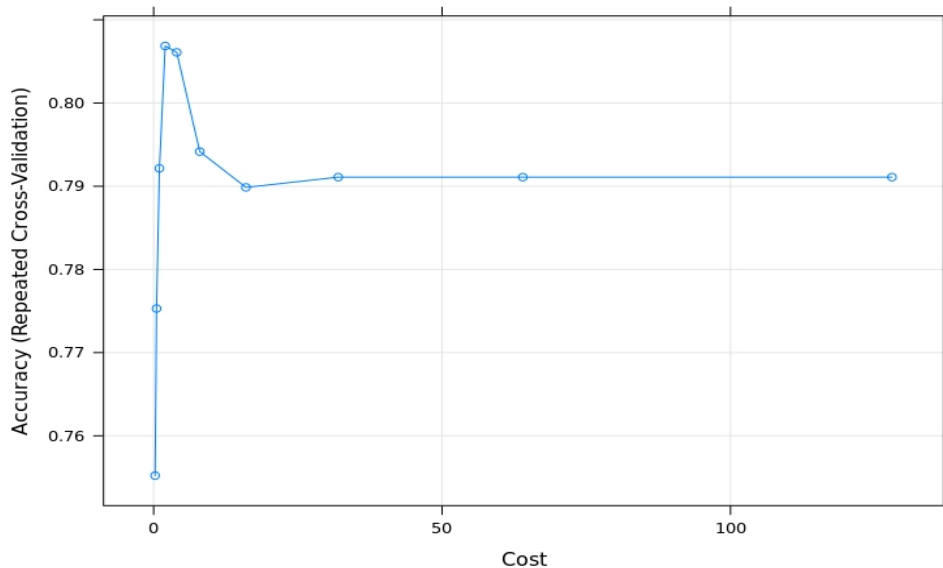
```

```

#
# C      Accuracy  Kappa
# 0.25  0.7558363  0.5080513
# 0.50  0.7762205  0.5495375
# 1.00  0.7888065  0.5751774
# 2.00  0.8071884  0.6126858
# 4.00  0.8032060  0.6048041
# 8.00  0.7924789  0.5835132
# 16.00 0.7857436  0.5699243
# 32.00 0.7862016  0.5708988
# 64.00 0.7862016  0.5708988
# 128.00 0.7862016  0.5708988
#
# Tuning parameter 'sigma' was held constant at a value of
0.01631908
# Accuracy was used to select the optimal model using the largest
value.
# The final values used for the model were sigma = 0.01631908 and
C = 2.

```

```
plot(svm_fit)
```



```

test_pred_svm <- predict(svm_fit, newdata = testmc)
confusionMatrix(test_pred_svm, testmc$DEATH_YN)

# Reference
# Prediction   N   Y
#           N 193  40
#           Y  61 239
#
# Accuracy : 0.8105
# 95% CI : (0.7746, 0.8429)
# No Information Rate : 0.5235
# P-Value [Acc > NIR] : < 2e-16
#
# Kappa : 0.6188
# Mcnemar's Test P-Value : 0.04658
#
#           Sensitivity : 0.7598
#           Specificity : 0.8566
#           Pos Pred Value : 0.8283
#           Neg Pred Value : 0.7967
#           Prevalence : 0.4765
#           Detection Rate : 0.3621
#           Detection Prevalence : 0.4371
#           Balanced Accuracy : 0.8082
#
#           'Positive' Class : N

knn_fit <- train(DEATH_YN~., data = trainmc, method = "knn",
trControl=ctrl, metric=metric, tuneLength=50, na.action =
na.omit)

knn_fit

# k-Nearest Neighbors
#

```

```

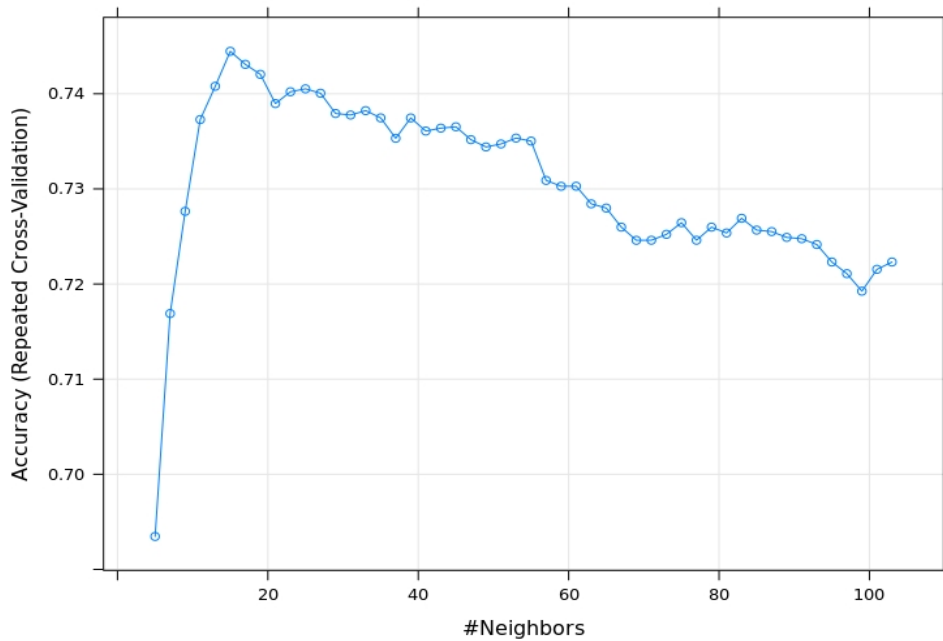
# 1305 samples
# 237 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1175, 1175, 1174, 1174, 1174, 1174, ...
# Resampling results across tuning parameters:
#
#  k    Accuracy  Kappa
#  5  0.6934584  0.3884296
#  7  0.7168961  0.4354293
#  9  0.7276394  0.4566965
# 11  0.7372895  0.4757003
# 13  0.7407905  0.4826947
# 15  0.7444676  0.4901690
# 17  0.7430934  0.4877344
# 19  0.7420423  0.4854501
# 21  0.7389723  0.4793790
# 23  0.7402135  0.4818175
# 25  0.7405213  0.4826955
# 27  0.7400598  0.4818284
# 29  0.7379293  0.4776709
# 31  0.7377685  0.4774076
# 33  0.7382288  0.4784292
# 35  0.7374677  0.4770597
# 37  0.7353233  0.4729011
# 39  0.7374607  0.4770964
# 41  0.7360832  0.4743449
# 43  0.7363897  0.4748597
# 45  0.7365295  0.4752215
# 47  0.7351753  0.4726036
# 49  0.7344107  0.4711849
# 51  0.7347243  0.4718457

```



```
# 53 0.7353419 0.4729313
# 55 0.7350296 0.4722679
# 57 0.7308933 0.4641897
# 59 0.7302767 0.4630918
# 61 0.7302826 0.4632545
# 63 0.7284295 0.4594930
# 65 0.7279867 0.4586925
# 67 0.7259808 0.4546327
# 69 0.7245962 0.4519591
# 71 0.7246044 0.4520265
# 73 0.7252221 0.4532958
# 75 0.7264471 0.4557363
# 77 0.7246044 0.4520041
# 79 0.7259797 0.4547039
# 81 0.7253655 0.4535308
# 83 0.7269109 0.4563846
# 85 0.7256685 0.4539878
# 87 0.7255123 0.4537544
# 89 0.7249016 0.4526608
# 91 0.7247606 0.4522964
# 93 0.7241522 0.4511357
# 95 0.7223143 0.4475869
# 97 0.7210881 0.4451730
# 99 0.7192479 0.4416902
# 101 0.7215439 0.4462002
# 103 0.7223131 0.4477554
# Accuracy was used to select the optimal model using the largest
value.
# The final value used for the model was k = 15.

plot(knn_fit)
```



```
test_pred_knn <- predict(knn_fit, newdata = testmc)
confusionMatrix (test_pred_knn, testmc$DEATH_YN)
```

```
# Confusion Matrix and Statistics
```

```
#
```

```
# Reference
```

```
# Prediction  N  Y
```

```
# N 197  99
```

```
# Y  57 180
```

```
#
```

```
# Accuracy : 0.7073
```

```
# 95% CI : (0.6667, 0.7456)
```

```
# No Information Rate : 0.5235
```

```
# P-Value [Acc > NIR] : < 2.2e-16
```

```
#
```

```
# Kappa : 0.4177
```

```
# Mcnemar's Test P-Value : 0.001028
```

```
#
```

```
#           Sensitivity : 0.7756
```

```

#           Specificity : 0.6452
#           Pos Pred Value : 0.6655
#           Neg Pred Value : 0.7595
#           Prevalence : 0.4765
#           Detection Rate : 0.3696
#           Detection Prevalence : 0.5553
#           Balanced Accuracy : 0.7104
#
#           'Positive' Class : N

set.seed(seed)
rf_fit <- train(DEATH_YN~., data = trainmc, method = "rf",
trControl=ctrl, metric=metric,
tuneLength=15,importance=TRUE,proximity=TRUE, na.action =
na.omit)

rf_fit

# 1305 samples
# 237 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1174, 1174, 1174, 1173, 1175, 1175, ...
# Resampling results across tuning parameters:
#
#  mtry Accuracy  Kappa
#  2   0.7674735  0.5307242
# 18   0.7920080  0.5812385
# 35   0.7927866  0.5826649
# 52   0.7901782  0.5774124
# 69   0.7918669  0.5808855
# 85   0.7878868  0.5728543
# 102  0.7877506  0.5726535

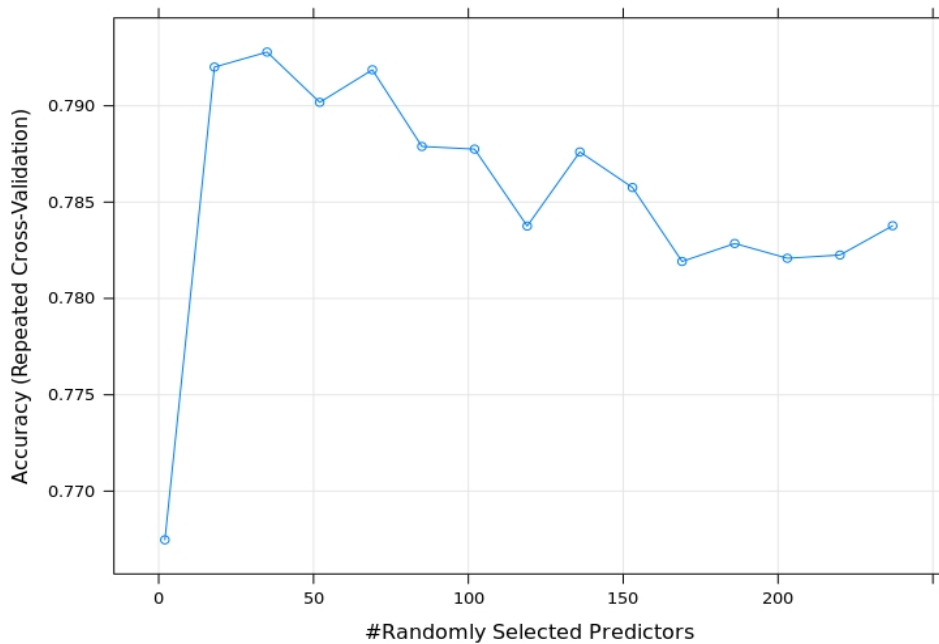
```

```

# 119  0.7837577  0.5645623
# 136  0.7876037  0.5724196
# 153  0.7857565  0.5686533
# 169  0.7819197  0.5609589
# 186  0.7828522  0.5628687
# 203  0.7820864  0.5614102
# 220  0.7822555  0.5616130
# 237  0.7837741  0.5648462
#
# Accuracy was used to select the optimal model using the largest
value.
# The final value used for the model was mtry = 35.

```

```
plot(rf_fit)
```



```

test_pred_rf <- predict(rf_fit, newdata = testmc)
confusionMatrix(test_pred_rf, testmc$DEATH_YN)

```

```
# Confusion Matrix and Statistics
```

```

#
# Reference
# Prediction   N   Y
# N 194  27
# Y  60 252
#
# Accuracy : 0.8368
# 95% CI : (0.8026, 0.8671)
# No Information Rate : 0.5235
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.6709
# Mcnemar's Test P-Value : 0.0006019
#
#           Sensitivity : 0.7638
#           Specificity : 0.9032
#           Pos Pred Value : 0.8778
#           Neg Pred Value : 0.8077
#           Prevalence : 0.4765
#           Detection Rate : 0.3640
#           Detection Prevalence : 0.4146
#           Balanced Accuracy : 0.8335
#
#           'Positive' Class : N

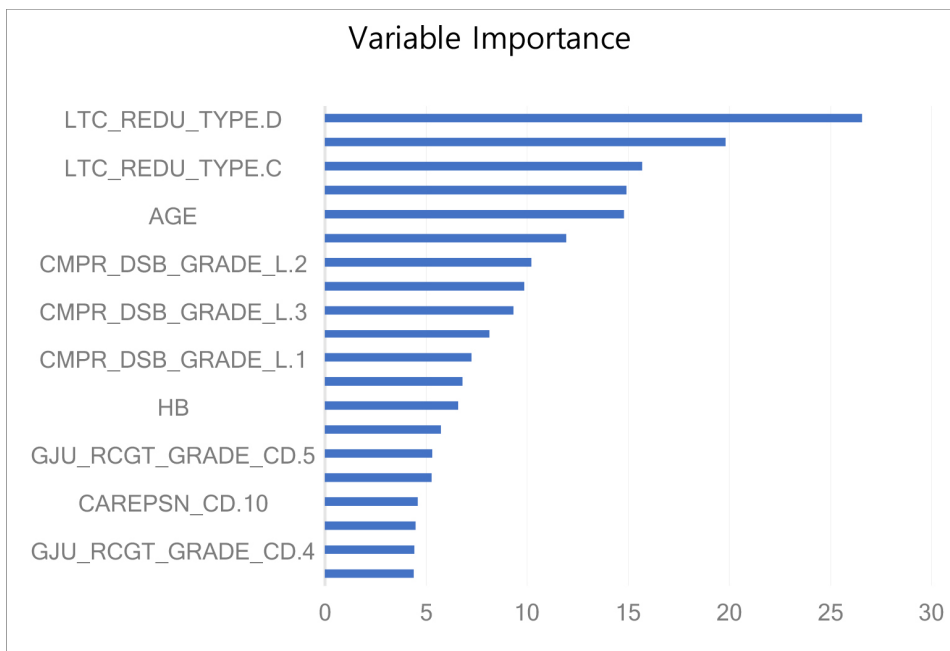
imp<-varImp(rf_fit, scale =FALSE)

imp

# rf variable importance
#
# only 20 most important variables shown (out of 237)
# Importance
# LTC_REDU_TYPE.D           26.559
# CAREPSN_CD.3             19.805

```

# LTC_REDU_TYPE.C	15.690
# ALBUMIN	14.907
# AGE	14.794
# RCG_RCGT_SCORE	11.929
# CMPR_DSB_GRADE_L.2	10.201
# CMPR_DSB_GRADE_L.5	9.861
# CMPR_DSB_GRADE_L.3	9.315
# CAREPSN_CD.11	8.133
# CMPR_DSB_GRADE_L.1	7.245
# CR	6.797
# HB	6.596
# TNDNC_ARFF_SCORE	5.740
# GJU_RCGT_GRADE_CD.5	5.302
# CMPR_DSB_GRADE_L.4	5.277
# CAREPSN_CD.10	4.583
# CMPR_DSB_GRADE_L.6	4.483
# GJU_RCGT_GRADE_CD.4	4.417
# LTC_REDU_TYPE.F	4.386

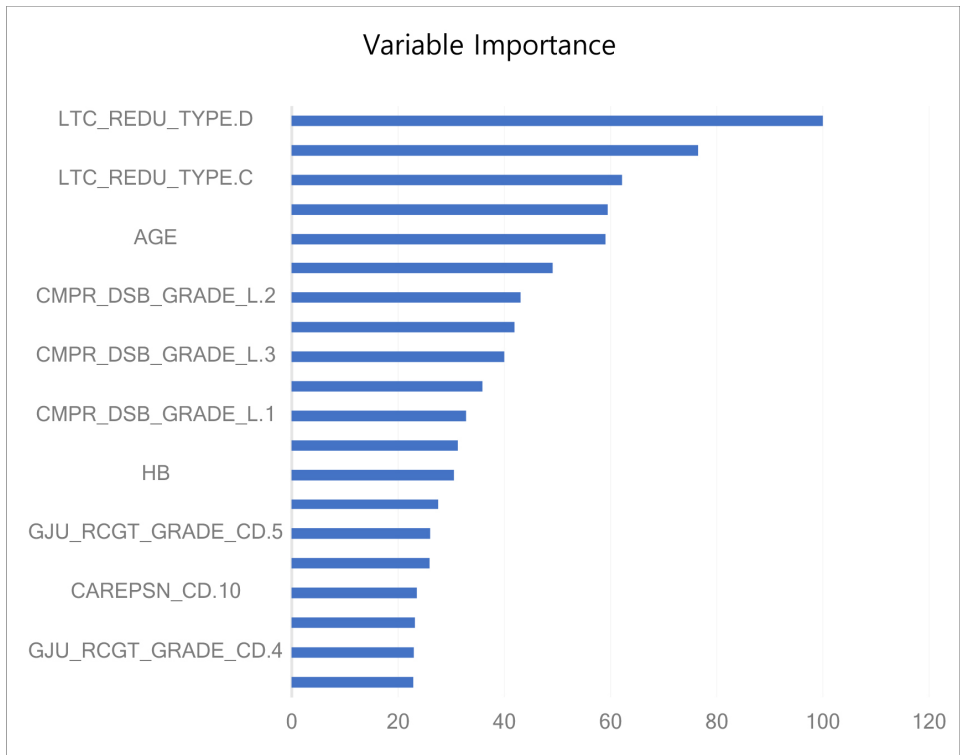


```

imp<-varImp(rf_fit, scale =TRUE)
imp

# rf variable importance
#
# only 20 most important variables shown (out of 237)
#
# Importance
# LTC_REDU_TYPE.D          100.00
# CAREPSN_CD.3             76.51
# LTC_REDU_TYPE.C          62.19
# ALBUMIN                   59.47
# AGE                       59.08
# RCG_RCGT_SCORE           49.11
# CMPR_DSB_GRADE_L.2       43.10
# CMPR_DSB_GRADE_L.5       41.92
# CMPR_DSB_GRADE_L.3       40.02
# CAREPSN_CD.11            35.90
# CMPR_DSB_GRADE_L.1       32.81
# CR                        31.26
# HB                        30.56
# TNDNC_ARFF_SCORE         27.58
# GJU_RCGT_GRADE_CD.5      26.06
# CMPR_DSB_GRADE_L.4       25.97
# CAREPSN_CD.10            23.56
# CMPR_DSB_GRADE_L.6       23.21
# GJU_RCGT_GRADE_CD.4      22.98
#LTC_REDU_TYPE. F         22.87

```



```
gbm_fit <- train(DEATH_YN~., data = trainmc, method = "gbm",
metric=metric, trControl=ctrl, tuneLength=10, verbose=FALSE,
na.action = na.omit)
gbm_fit

# Stochastic Gradient Boosting
#
# 1305 samples
# 237 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1173, 1175, 1175, 1174, 1175, 1174, ...
# Resampling results across tuning parameters:
#
```



#	interaction.depth	n.trees	Accuracy	Kappa
# 1	50	0.7641743	0.5244474	
# 1	100	0.7908300	0.5791742	
# 1	150	0.7979952	0.5940035	
# 1	200	0.8027632	0.6037499	
# 1	250	0.8056676	0.6097132	
# 1	300	0.8058214	0.6100837	
# 1	350	0.8073412	0.6131278	
# 1	400	0.8059648	0.6104459	
# 1	450	0.8048972	0.6083670	
# 1	500	0.8058180	0.6101728	
# 2	50	0.7903498	0.5780485	
# 2	100	0.7998438	0.5977115	
# 2	150	0.8024486	0.6031993	
# 2	200	0.8035115	0.6054265	
# 2	250	0.8050371	0.6085587	
# 2	300	0.8048891	0.6083300	
# 2	350	0.8012027	0.6009991	
# 2	400	0.7996665	0.5979320	
# 2	450	0.8005849	0.5998316	
# 2	500	0.7987423	0.5960098	
# 3	50	0.7967902	0.5914692	
# 3	100	0.8035161	0.6052922	
# 3	150	0.8053611	0.6091724	
# 3	200	0.8044381	0.6073194	
# 3	250	0.8036654	0.6058647	
# 3	300	0.8018286	0.6021565	
# 3	350	0.8002889	0.5991113	
# 3	400	0.7979661	0.5943087	
# 3	450	0.7973472	0.5931182	
# 3	500	0.7970640	0.5925700	
# 4	50	0.7951120	0.5879353	
# 4	100	0.8049078	0.6080830	
# 4	150	0.8041269	0.6066078	

# 4	200	0.8050370	0.6086015
# 4	250	0.8007341	0.5998200
# 4	300	0.7985884	0.5955515
# 4	350	0.7976654	0.5937296
# 4	400	0.7961375	0.5905703
# 4	450	0.7965978	0.5914522
# 4	500	0.7988973	0.5961262
# 5	50	0.7934020	0.5846365
# 5	100	0.8033798	0.6050002
# 5	150	0.8038412	0.6060770
# 5	200	0.8052083	0.6087394
# 5	250	0.8021327	0.6026299
# 5	300	0.7976805	0.5937368
# 5	350	0.7956817	0.5896908
# 5	400	0.7964637	0.5913539
# 5	450	0.7946293	0.5875517
# 5	500	0.7952400	0.5888648
# 6	50	0.7949476	0.5877767
# 6	100	0.7989254	0.5962274
# 6	150	0.7993694	0.5970606
# 6	200	0.7961538	0.5907821
# 6	250	0.7940106	0.5865805
# 6	300	0.7940164	0.5866015
# 6	350	0.7921726	0.5828527
# 6	400	0.7940000	0.5865758
# 6	450	0.7917040	0.5819421
# 6	500	0.7903300	0.5792208
# 7	50	0.7961817	0.5902686
# 7	100	0.8064556	0.6111311
# 7	150	0.8041222	0.6064938
# 7	200	0.8024381	0.6031650
# 7	250	0.7998367	0.5980713
# 7	300	0.8008915	0.6002038
# 7	350	0.7979871	0.5943640
# 7	400	0.7984510	0.5953607

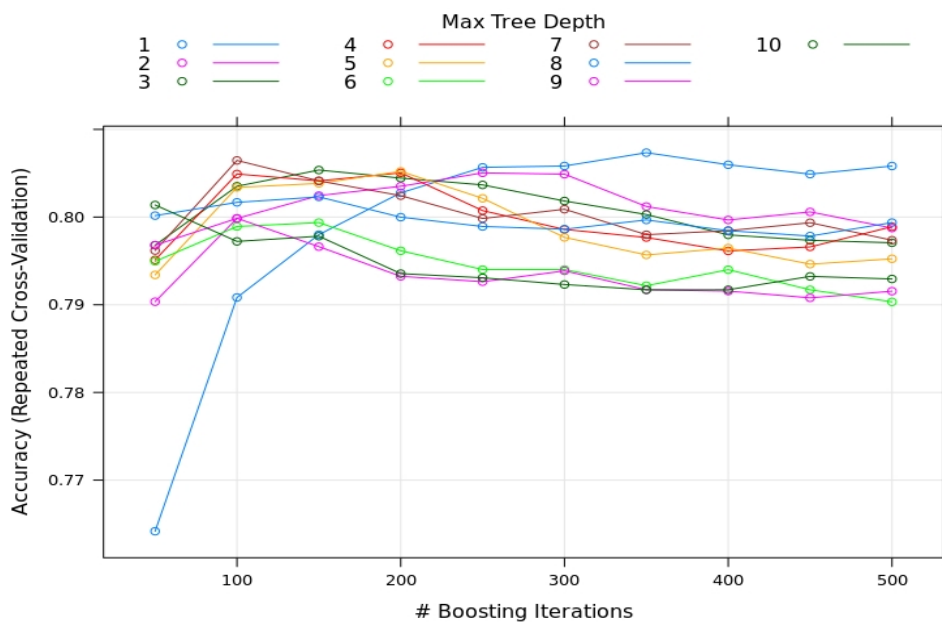
# 7	450	0.7993543	0.5971155
# 7	500	0.7973566	0.5931321
# 8	50	0.8001630	0.5984857
# 8	100	0.8016769	0.6017664
# 8	150	0.8022982	0.6030701
# 8	200	0.7999859	0.5984252
# 8	250	0.7989229	0.5962443
# 8	300	0.7986176	0.5957497
# 8	350	0.7996687	0.5978448
# 8	400	0.7984415	0.5954750
# 8	450	0.7978343	0.5942883
# 8	500	0.7993693	0.5973237
# 9	50	0.7967961	0.5917101
# 9	100	0.7998507	0.5980864
# 9	150	0.7966247	0.5917681
# 9	200	0.7932401	0.5848289
# 9	250	0.7926294	0.5836510
# 9	300	0.7938671	0.5861846
# 9	350	0.7917098	0.5818015
# 9	400	0.7915571	0.5815491
# 9	450	0.7907891	0.5799847
# 9	500	0.7915466	0.5813608
# 10	50	0.8013751	0.6007844
# 10	100	0.7972236	0.5926926
# 10	150	0.7978145	0.5939722
# 10	200	0.7935500	0.5855080
# 10	250	0.7930767	0.5845758
# 10	300	0.7923076	0.5830221
# 10	350	0.7917028	0.5818485
# 10	400	0.7917121	0.5817944
# 10	450	0.7932342	0.5849507
# 10	500	0.7929393	0.5842619
#			
# Tuning parameter 'shrinkage' was held constant at a value of 0.1			

```

# Tuning parameter 'n.minobsinnode' was held constant at a value
of 10
# Accuracy was used to select the optimal model using the largest
value.
# The final values used for the model were n.trees = 350,
interaction.depth = 1, shrinkage = 0.1 and n.minobsinnode = 10.

```

```
plot(gbm_fit)
```



```

test_pred_gbm <- predict(gbm_fit, newdata = testmc)
confusionMatrix(test_pred_gbm, testmc$DEATH_YN)
#
# Confusion Matrix and Statistics
#
# Reference
# Prediction  N  Y
# N 199  41
# Y  55 238
#
# Accuracy : 0.8199

```

```

# 95% CI : (0.7846, 0.8516)
# No Information Rate : 0.5235
# P-Value [Acc > NIR] : <2e-16
#
# Kappa : 0.6381
# McNemar's Test P-Value : 0.1846
#
# Sensitivity : 0.7835
# Specificity : 0.8530
# Pos Pred Value : 0.8292
# Neg Pred Value : 0.8123
# Prevalence : 0.4765
# Detection Rate : 0.3734
# Detection Prevalence : 0.4503
# Balanced Accuracy : 0.8183
#
# 'Positive' Class : N

imp<-varImp(gbm_fit, scale = FALSE)

imp

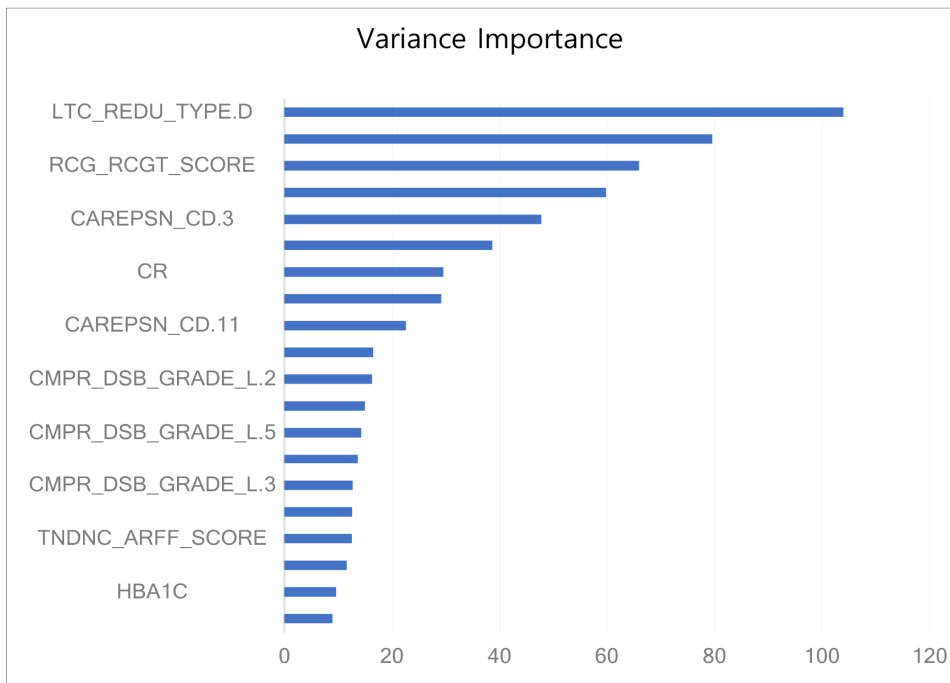
# gbm variable importance
#
# only 20 most important variables shown (out of 237)
#
# Overall
# LTC_REDU_TYPE.D 103.965
# ALBUMIN 79.561
# RCG_RCGT_SCORE 65.986
# AGE 59.838
# CAREPSN_CD.3 47.804
# LTC_REDU_TYPE.C 38.650
# CR 29.547
# CMPR_DSB_GRADE_L.1 29.168

```

```

# CAREPSN_CD.11      22.634
# GLUCOSE_AC         16.527
# CMPR_DSB_GRADE_L.2 16.326
# BUN                14.984
# CMPR_DSB_GRADE_L.5 14.306
# ipwon_period       13.677
# CMPR_DSB_GRADE_L.3 12.707
# HB                 12.591
# TNDNC_ARFF_SCORE   12.573
# L_tube_feeding.1   11.574
# HBA1C              9.654
# CAREPSN_CD.10      8.931

```



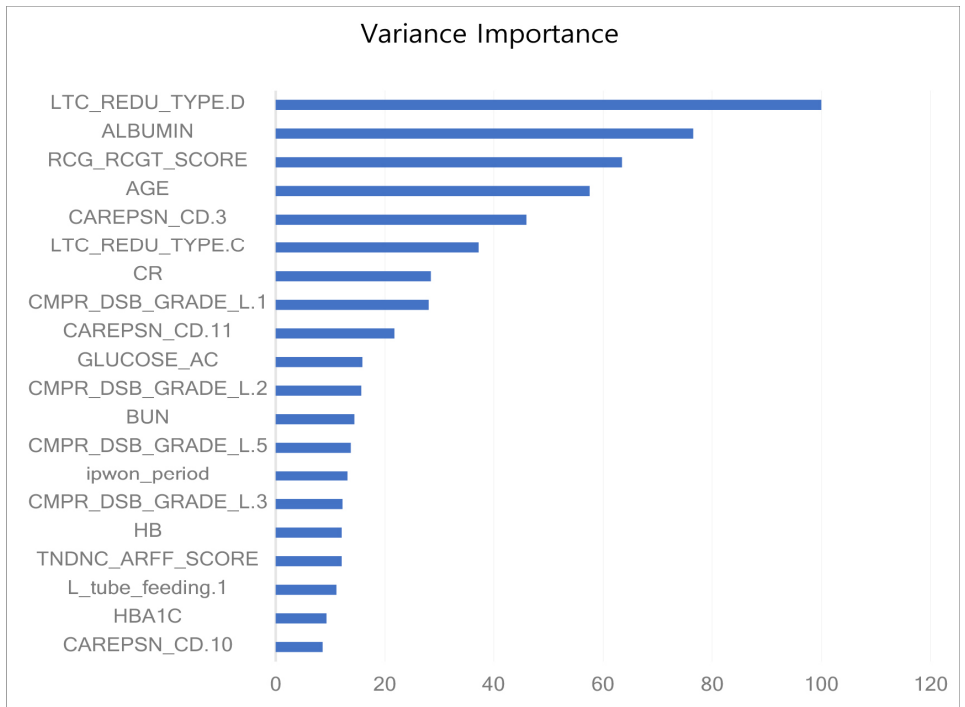
```
imp<-varImp(gbm_fit, scale = TRUE)
```

```
imp
```

```

# gbm variable importance
#
#   only 20 most important variables shown (out of 237)
#
#           Overall
# LTC_REDU_TYPE.D           100.000
# ALBUMIN                    76.527
# RCG_RCGT_SCORE            63.469
# AGE                        57.556
# CAREPSN_CD.3              45.981
# LTC_REDU_TYPE.C           37.176
# CR                         28.420
# CMPR_DSB_GRADE_L.1        28.056
# CAREPSN_CD.11             21.771
# GLUCOSE_AC                 15.897
# CMPR_DSB_GRADE_L.2        15.703
# BUN                        14.413
# CMPR_DSB_GRADE_L.5        13.761
# ipwon_period               13.155
# CMPR_DSB_GRADE_L.3        12.222
# HB                          12.111
# TNDNC_ARFF_SCORE          12.094
# L_tube_feeding.1          11.133
# HBA1C                      9.285
# CAREPSN_CD.10             8.590

```



```
rpart_fit <- train(DEATH_YN~., data = trainmc, method = "rpart")
rpart_fit
```

CART

```
# 1305 samples
# 237 predictor
# 2 classes: '0', '1'
#
# No pre-processing
# Resampling: Bootstrapped (25 reps)
# Summary of sample sizes: 1305, 1305, 1305, 1305, 1305, ...
# Resampling results across tuning parameters:
#
#   cp          Accuracy   Kappa
# 0.04590164  0.7147868  0.41997464
# 0.06967213  0.6760848  0.33326086
```



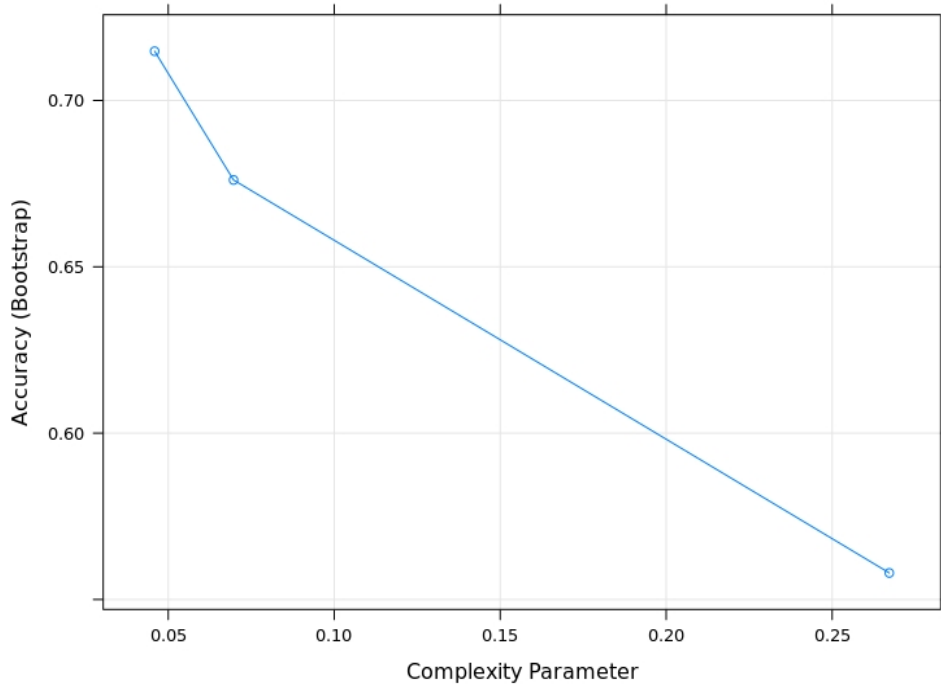
```

# 0.26721311 0.5579830 0.06341989
#
# Accuracy was used to select the optimal model using the largest
value.
# The final value used for the model was cp = 0.04590164.

test_pred_rpart <- predict(rpart_fit, newdata = testmc)
confusionMatrix (test_pred_rpart, testmc$DEATH_YN)

# Confusion Matrix and Statistics
#
# Reference
# Prediction  0  1
# 0 171  62
# 1  96 205
#
# Accuracy : 0.7041
# 95% CI : (0.6634, 0.7425)
# No Information Rate : 0.5
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.4082
# McNemar's Test P-Value : 0.008656
#
#           Sensitivity : 0.6404
#           Specificity : 0.7678
#           Pos Pred Value : 0.7339
#           Neg Pred Value : 0.6811
#           Prevalence : 0.5000
#           Detection Rate : 0.3202
#           Detection Prevalence : 0.4363
#           Balanced Accuracy : 0.7041
#
#           'Positive' Class : 0
Plot(r_part)

```



```
imp<-varImp(rpart_fit, scale =TRUE)
```

```
imp
```

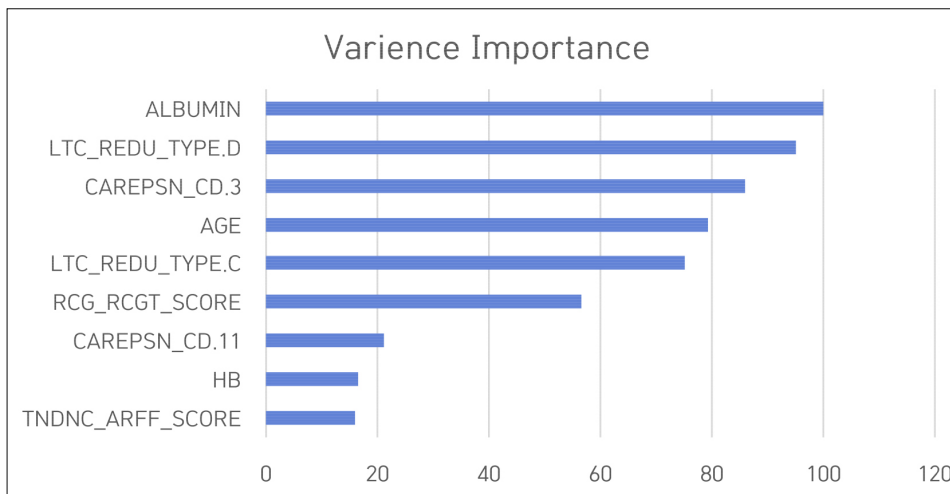
```
# Overall
# ALBUMIN          100.00
# LTC_REDU_TYPE.D  95.08
# CAREPSN_CD.3     85.96
# AGE              79.30
# LTC_REDU_TYPE.C  75.14
# RCG_RCGT_SCORE   56.59
# CAREPSN_CD.11    21.17
# HB               16.54
# TNDNC_ARFF_SCORE 15.99
# LVTGT_PRSN_CD.G  0.00
# LTC_REDU_TYPE.B  0.00
# MAIN_DSB_TYPE_L.10 0.00
```

```

# CMPR_DSB_GRADE.3      0.00
# S10.1                 0.00
# IVH.1                 0.00
# GAIBJA_TYPE.5        0.00
# MCA.0                 0.00
# PE2.2                 0.00
# SAH.1                 0.00
# CMPR_DSB_GRADE_L.3   0.00

```

```
plot(imp)
```



```

tunedxgbgrid <- expand.grid(nrounds=500, max_depth=7, eta=0.01,
gamma=0, colsample_bytree=0.6, min_child_weight=1, subsample=0.8)

```

```

ctrl <- trainControl(method="repeatedcv",
number=10, repeats=5, classProbs = TRUE)

```

```
set.seed(seed)
```

```

xgb_fit <- train(DEATH_YN~., data = trainmc, method = "xgbTree",
metric=metric, trControl=ctrl, tuneGrid=tunedxgbgrid, na.action =
na.omit)

```

```
xgb_fit
```

```
# eXtreme Gradient Boosting
```

```

# 1305 samples
# 237 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1174, 1174, 1174, 1173, 1175, 1175, ...
# Resampling results:
#
# Accuracy Kappa
# 0.801368 0.6006442
#
# Tuning parameter 'nrounds' was held constant at a value of 500
# Tuning parameter 'max_depth' was held constant at a value of 7
# parameter 'colsample_bytree' was held constant at a value of
0.6
# Tuning parameter 'min_child_weight' was held constant at a
value of
# 1
# Tuning parameter 'subsample' was held constant at a value of 0.8

test_pred_xgb <- predict(xgb_fit, newdata = testmc)
confusionMatrix(test_pred_xgb, testmc$DEATH_YN)

# Reference

# Prediction  N  Y
#           N 195 28
#           Y  59 251
#
#           Accuracy : 0.8368
#           95% CI : (0.8026, 0.8671)
#           No Information Rate : 0.5235
#           P-Value [Acc > NIR] : < 2.2e-16
#

```

```

#           Kappa : 0.671
# McNemar's Test P-Value : 0.001298
#
# Sensitivity : 0.7677
# Specificity : 0.8996
# Pos Pred Value : 0.8744
# Neg Pred Value : 0.8097
# Prevalence : 0.4765
# Detection Rate : 0.3659
# Detection Prevalence : 0.4184
# Balanced Accuracy : 0.8337
#
# 'Positive' Class : N

imp<-varImp(xgb_fit, scale = FALSE)

imp

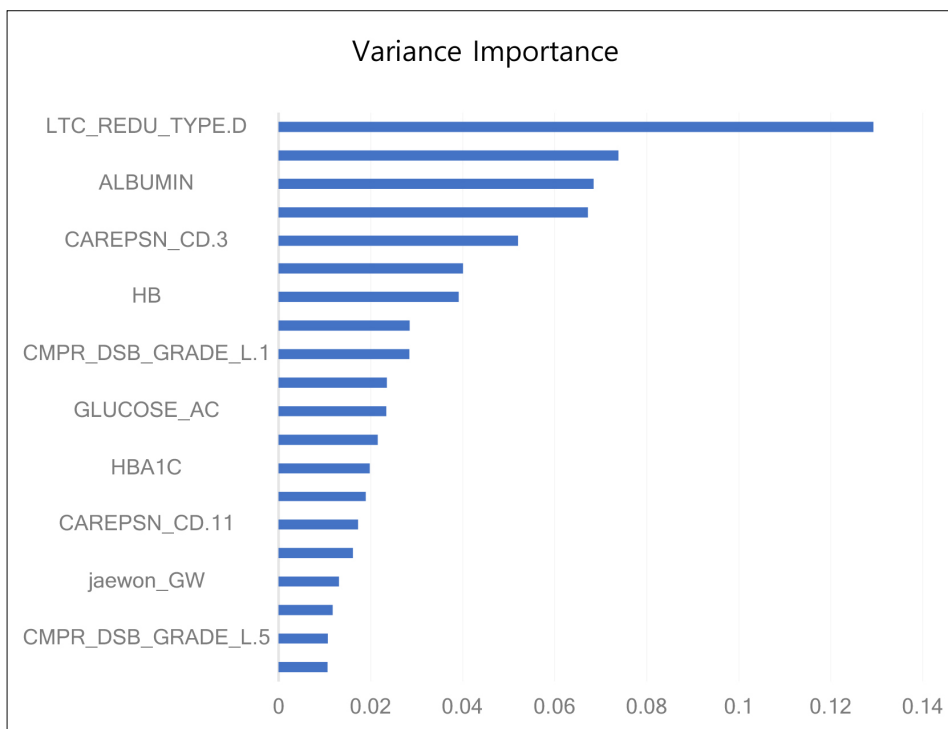
# xgbTree variable importance
#
# only 20 most important variables shown (out of 237)
#
# Overall
# LTC_REDU_TYPE.D      0.12928
# RCG_RCGT_SCORE      0.07385
# ALBUMIN              0.06847
# AGE                  0.06726
# CAREPSN_CD.3        0.05207
# CR                   0.04009
# HB                   0.03919
# LTC_REDU_TYPE.C     0.02851
# CMPR_DSB_GRADE_L.1 0.02845
# BUN                  0.02354
# GLUCOSE_AC          0.02339
# CMPR_DSB_GRADE_L.2 0.02155

```

```

# HBA1C          0.01981
# ipwon_period   0.01894
# CAREPSN_CD.11 0.01731
# total_cowork   0.01614
# jaewon_GW      0.01310
# TNDNC_ARFF_SCORE 0.01178
# CMPR_DSB_GRADE_L.5 0.01071
# PCT_ARFF_SCORE 0.01067

```



```

imp<-varImp(xgb_fit, scale = TRUE)
imp

# xgbTree variable importance
#
# only 20 most important variables shown (out of 237)
#

```

```

# Overall
# LTC_REDU_TYPE.D      0.12928
# RCG_RCGT_SCORE      0.07385
# ALBUMIN              0.06847
# AGE                  0.06726
# CAREPSN_CD.3        0.05207
# CR                   0.04009
# HB                   0.03919
# LTC_REDU_TYPE.C     0.02851
# CMPR_DSB_GRADE_L.1  0.02845
# BUN                  0.02354
# GLUCOSE_AC          0.02339
# CMPR_DSB_GRADE_L.2  0.02155
# HBA1C               0.01981
# ipwon_period        0.01894
# CAREPSN_CD.11       0.01731
# total_cowork        0.01614
# jaewon_GW           0.01310
# TNDNC_ARFF_SCORE    0.01178
# CMPR_DSB_GRADE_L.5  0.01071
# PCT_ARFF_SCORE      0.01067

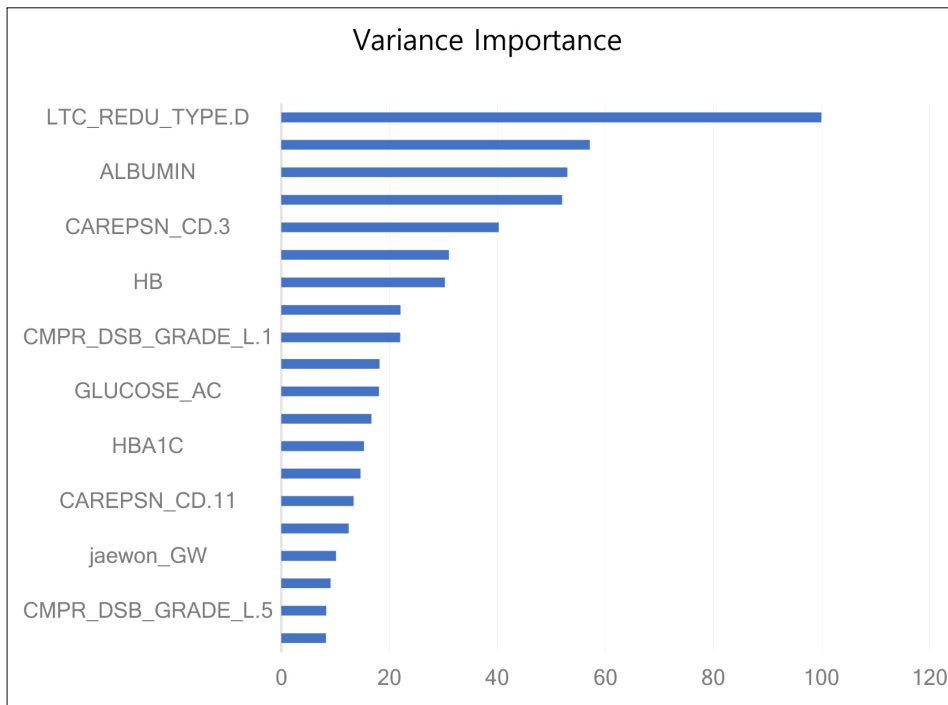
# xgbTree variable importance
#
# only 20 most important variables shown (out of 237)
#
# Overall
# LTC_REDU_TYPE.D      100.000
# RCG_RCGT_SCORE       57.119
# ALBUMIN              52.957
# AGE                  52.027
# CAREPSN_CD.3        40.273
# CR                   31.013
# HB                   30.312
# LTC_REDU_TYPE.C     22.051

```

```

# CMPR_DSB_GRADE_L.1  22.009
# BUN                  18.205
# GLUCOSE_AC          18.091
# CMPR_DSB_GRADE_L.2  16.666
# HBA1C                15.319
# ipwon_period        14.647
# CAREPSN_CD.11       13.388
# total_cowork         12.485
# jaewon_GW           10.132
# TNDNC_ARFF_SCORE    9.113
# CMPR_DSB_GRADE_L.5  8.287
# PCT_ARFF_SCORE      8.256

```



```

ovrctrl <- trainControl(method="repeatedcv", number=10,
repeats=5,
                        savePredictions='final',
                        classProbs=TRUE,

```



```

summaryFunction = twoClassSummary,
allowParallel=TRUE,
index = createFolds(trainmc$DEATH_YN, 10))
metric <- "ROC"
algorithmList <-
c('nnet','svmRadial','knn','rpart','rf','gbm','xgbTree')

model_DEATH <- caretList(DEATH_YN~., data=trainmc,
metric=metric,
trControl=ovrctrl,
methodList= algorithmList,
tuneList = NULL,#ClassProbs=F,
continue_on_fail = FALSE)# ,na.action =
na.omit) # na.action = na.roughfix

stctrl <- trainControl(method="repeatedcv", number=10, repeats=5,
savePredictions=TRUE, classProbs=TRUE,
summaryFunction = twoClassSummary,
allowParallel=TRUE,
index = createFolds(trainmc$DEATH_YN, 10))

set.seed(seed)
stack.rf_DEATH <- caretStack(model_DEATH, method="rf",
metric="ROC", trControl=stctrl)
print(stack.rf_DEATH)

# A rf ensemble of 2 base models: nnet, svmRadial, knn, rpart, rf,
gbm, xgbTree
#
# Ensemble results:
# Random Forest
#
# 11745 samples
# 7 predictor
# 2 classes: 'N', 'Y'

```

```

#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 130, 130, 132, 130, 131, 130, ...
# Resampling results across tuning parameters:
#
#  mtry  ROC  Sens  Spec
#  2     0.5  NaN   NaN
#  4     0.5  NaN   NaN
#  7     0.5  NaN   NaN
#
# ROC was used to select the optimal model using the largest
value.
# The final value used for the model was mtry = 2.

test_pred_DEATH <- predict(stack.rf_DEATH, newdata = testmc)
confusionMatrix(test_pred_DEATH, testmc$DEATH_YN)
Confusion Matrix and Statistics

# Reference
# Prediction   N   Y
#           N 223  24
#           Y  51 255
#
# Accuracy : 0.8643
# 95% CI : (0.8340, 0.8946)
# No Information Rate : 0.7469
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.6247
# Mcnemar's Test P-Value : 0.001189

#
#           Sensitivity : 0.9139
#           Specificity : 0.8138

```

```

#           Pos Pred Value : 0.8333
#           Neg Pred Value : 0.9028
#           Prevalence : 0.5045
#           Detection Rate : 0.4611
#   Detection Prevalence : 0.5533
#           Balanced Accuracy : 0.8639
#
#           'Positive' Class : N

```

장기요양 서비스 이용자군의 생존 기간과 생존에 대한 회귀 분석

```

result=lm(Live_days~ipwon_period+AGE+SEX+SIDO_NM+CITY_SIZE+GAIBJA
_TYPE+CMPR_DSB_GRADE+MAIN_DSB_TYPE+CMPR_DSB_GRADE_L+MAIN_DSB_TYPE
_L+L_tube_feeding+pneumonia+Infarction+SAH+SDH+ICH+IVH+Focal_small
l+multifocal_large+Acute_recent+Chronic_old+Temporal+Occipital+Cerebellum+BG+brain_stem+Thalamus+ACA+MCA+PCA+Aneurysm+APFRM_TYPE_CD+LTC_OBJSR_TYPE_CD+LTC_REDU_TYPE+RCG_RCGT_SCORE+GJU_RCGT_GRADE_CD+DISM_JUDG_RSN_CD+RSDN_TP_CD+LVTGT_PRSN_CD+CAREPSN_CD+PCT_ARFF_SCORE+BHVR_ARFF_SCORE+TNDNC_ARFF_SCORE+REHBT_ARFF_SCORE+N05+N07+S10+PE1+PE2+smoke_yn+drink_yn+ALBUMIN+BUN+CR+GLUCOSE_AC+HB+HBA1C+jaewon_GW+total_cowork+pum_cowork+idm_cowork, data=Data_final)

```

```
summary(result)
```

Call:

```

lm(formula = Live_days ~ ipwon_period + AGE + SEX + RVSN_ADDR_CD +
    SIDO_NM + CITY_SIZE + GAIBJA_TYPE + CMPR_DSB_GRADE +
    MAIN_DSB_TYPE +
    CMPR_DSB_GRADE_L + MAIN_DSB_TYPE_L + L_tube_feeding +
    pneumonia +
    Infarction + SAH + SDH + ICH + IVH + Focal_small +
    multifocal_large +
    Acute_recent + Chronic_old + Temporal + Occipital +
    Cerebellum +
    BG + brain_stem + Thalamus + ACA + MCA + PCA + Aneurysm +

```

```

APFRM_TYPE_CD + LTC_OBJTR_TYPE_CD + LTC_REDU_TYPE +
RCG_RCGT_SCORE +
    GJU_RCGT_GRADE_CD + DISM_JUDG_RSN_CD + RSDN_TP_CD +
LVTGT_PRSN_CD +
    CAREPSN_CD + PCT_ARFF_SCORE + BHVR_ARFF_SCORE +
TNDNC_ARFF_SCORE +
    REHBT_ARFF_SCORE + N05 + N07 + S10 + PE1 + PE2 + smoke_yn +
    drink_yn + ALBUMIN + BUN + CR + GLUCOSE_AC + HB + HBA1C +
    jaewon_GW + total_cowork + pum_cowork + idm_cowork, data =
Data_final)

```

Residuals:

Min	1Q	Median	3Q	Max
-2413.8	-656.1	-110.5	577.5	3513.8

Coefficients: (6 not defined because of singularities)

Estimate	Std. Error	t value	Pr(> t )
(Intercept)	5333.4689	675.8667	7.891 5.40e-15 ***
ipwon_period	1.7066	0.8445	2.021 0.043450 *
AGE	-50.0449	3.2901	-15.211 < 2e-16 ***
SEX2	118.8162	65.2906	1.820 0.068970 .
SIDO_NM2	-845.7870	450.4975	-1.877 0.060633 .
SIDO_NM3	-366.2705	590.4950	-0.620 0.535161
SIDO_NM4	-46.4090	174.9815	-0.265 0.790871
SIDO_NM5	105.3035	502.5971	0.210 0.834069
SIDO_NM6	-139.0638	589.3425	-0.236 0.813490
SIDO_NM7	-66.9296	1022.2104	-0.065 0.947803
SIDO_NM9	-6.3813	70.6030	-0.090 0.927994
SIDO_NM10	-72.4940	252.0833	-0.288 0.773705
SIDO_NM11	116.4156	297.9162	0.391 0.696020
SIDO_NM12	161.3116	216.6135	0.745 0.456560
SIDO_NM13	837.1002	289.2124	2.894 0.003849 **
SIDO_NM14	-78.4125	227.7332	-0.344 0.730652
SIDO_NM15	383.7834	325.1910	1.180 0.238099
SIDO_NM16	120.7903	401.4361	0.301 0.763532
SIDO_NM17	801.3396	721.6589	1.110 0.266983

CITY_SIZE2	-38.4451	93.6973	-0.410	0.681631
CITY_SIZE3	-672.4431	195.0187	-3.448	0.000579 ***
CITY_SIZE4	NA	NA	NA	NA
CITY_SIZE5	57.3954	116.8445	0.491	0.623342
CITY_SIZE6	-58.6329	74.4235	-0.788	0.430911
CITY_SIZE7	178.8583	342.3993	0.522	0.601485
CITY_SIZE8	201.3179	394.2090	0.511	0.609638
GAIBJA_TYPE1	240.5582	355.2429	0.677	0.498396
GAIBJA_TYPE2	52.3856	365.4914	0.143	0.886048
GAIBJA_TYPE3	77.7084	376.4348	0.206	0.836479
GAIBJA_TYPE4	-139.2108	370.8601	-0.375	0.707431
GAIBJA_TYPE5	370.9248	364.8448	1.017	0.309462
GAIBJA_TYPE6	95.6809	371.4899	0.258	0.796779
GAIBJA_TYPE7	150.0377	367.0664	0.409	0.682778
GAIBJA_TYPE8	382.3786	364.4494	1.049	0.294242
GAIBJA_TYPE9	369.9585	371.9922	0.995	0.320109
GAIBJA_TYPE10	319.9017	358.0578	0.893	0.371754
GAIBJA_TYPE11	273.2847	369.4119	0.740	0.459537
GAIBJA_TYPE12	456.2095	365.4316	1.248	0.212057
GAIBJA_TYPE13	257.6250	358.2977	0.719	0.472227
GAIBJA_TYPE14	340.5218	359.1780	0.948	0.343238
GAIBJA_TYPE15	266.6007	359.4443	0.742	0.458373
GAIBJA_TYPE16	278.8524	359.1637	0.776	0.437628
GAIBJA_TYPE17	482.3128	353.9426	1.363	0.173167
GAIBJA_TYPE18	342.6623	352.4645	0.972	0.331099
GAIBJA_TYPE19	365.1225	353.0142	1.034	0.301147
GAIBJA_TYPE20	344.2284	350.2840	0.983	0.325893
CMPR_DSB_GRADE1	-226.3612	672.6285	-0.337	0.736512
CMPR_DSB_GRADE2	53.6526	669.6213	0.080	0.936148
CMPR_DSB_GRADE3	195.0406	668.1235	0.292	0.770382
CMPR_DSB_GRADE4	216.1613	659.1511	0.328	0.742999
CMPR_DSB_GRADE5	157.9372	663.2573	0.238	0.811815
CMPR_DSB_GRADE6	169.3322	650.4184	0.260	0.794631
CMPR_DSB_GRADE10	-451.7739	743.4694	-0.608	0.543499
CMPR_DSB_GRADE20	-49.6201	711.5149	-0.070	0.944410

MAIN_DSB_TYPE1	94.3985	737.6408	0.128	0.898186
MAIN_DSB_TYPE2	182.3610	740.0740	0.246	0.805396
MAIN_DSB_TYPE3	410.3520	731.7700	0.561	0.575033
MAIN_DSB_TYPE4	586.5380	769.7291	0.762	0.446165
MAIN_DSB_TYPE5	693.1631	798.6029	0.868	0.385537
MAIN_DSB_TYPE6	-14.9331	996.6418	-0.015	0.988047
MAIN_DSB_TYPE8	-575.7469	742.5167	-0.775	0.438215
MAIN_DSB_TYPE9	316.6127	743.2301	0.426	0.670167
MAIN_DSB_TYPE10	-422.8136	828.9716	-0.510	0.610088
MAIN_DSB_TYPE11	164.9139	959.3455	0.172	0.863535
MAIN_DSB_TYPE12	2456.4082	1240.5308	1.980	0.047855 *
MAIN_DSB_TYPE14	NA	NA	NA	NA
CMPR_DSB_GRADE_L1	387.1182	228.8912	1.691	0.090973 .
CMPR_DSB_GRADE_L2	87.1832	209.4833	0.416	0.677331
CMPR_DSB_GRADE_L3	-315.8904	263.9954	-1.197	0.231644
CMPR_DSB_GRADE_L4	-586.8066	263.5622	-2.226	0.026119 *
CMPR_DSB_GRADE_L5	-280.7352	257.9897	-1.088	0.276681
CMPR_DSB_GRADE_L6	NA	NA	NA	NA
MAIN_DSB_TYPE_L1	-231.5973	962.5555	-0.241	0.809890
MAIN_DSB_TYPE_L2	-244.9567	960.3245	-0.255	0.798695
MAIN_DSB_TYPE_L3	-546.9172	947.2189	-0.577	0.563753
MAIN_DSB_TYPE_L4	-554.5488	985.2628	-0.563	0.573618
MAIN_DSB_TYPE_L5	-793.0781	992.6582	-0.799	0.424438
MAIN_DSB_TYPE_L6	-443.6615	1064.4392	-0.417	0.676877
MAIN_DSB_TYPE_L8	NA	NA	NA	NA
MAIN_DSB_TYPE_L9	-919.6885	949.5394	-0.969	0.332905
MAIN_DSB_TYPE_L10	777.0452	1120.0496	0.694	0.487931
MAIN_DSB_TYPE_L11	NA	NA	NA	NA
MAIN_DSB_TYPE_L14	NA	NA	NA	NA
L_tube_feeding1	-156.4359	59.8834	-2.612	0.009074 **
pneumonia1	39.1723	53.1386	0.737	0.461122
Infarction1	-57.4194	84.4407	-0.680	0.496602
SAH1	-137.8273	114.3212	-1.206	0.228139
SDH1	158.6798	91.0572	1.743	0.081583 .
ICH1	-115.0021	64.1201	-1.794	0.073069 .

IVH1	-53.3268	104.0841	-0.512	0.608479	
Focal_small1	179.6271	60.6629	2.961	0.003109	**
multifocal_large1	12.2824	55.0144	0.223	0.823362	
Acute_recent1	71.9676	70.3688	1.023	0.306591	
Chronic_old1	173.8747	73.1233	2.378	0.017528	*
Temporal1	-51.0088	87.7247	-0.581	0.561007	
Occipital1	-51.3655	91.9845	-0.558	0.576637	
Cerebellum1	44.0846	62.6958	0.703	0.482061	
BG1	-13.8634	57.4074	-0.241	0.809205	
brain_stem1	36.7576	76.4676	0.481	0.630797	
Thalamus1	-4.9276	65.9878	-0.075	0.940483	
ACA1	-205.6638	96.5617	-2.130	0.033330	*
MCA1	-80.2819	59.0210	-1.360	0.173944	
PCA1	38.2338	89.7221	0.426	0.670064	
Aneurysm1	-168.1799	81.6109	-2.061	0.039483	*
APFRM_TYPE_CD2	222.8999	136.0408	1.638	0.101513	
APFRM_TYPE_CD3	204.0876	61.9510	3.294	0.001007	**
APFRM_TYPE_CD4	119.4915	80.9018	1.477	0.139868	
APFRM_TYPE_CD5	-30.7344	852.0296	-0.036	0.971229	
LTC_OBJTR_TYPE_CD	47.0719	76.4815	0.615	0.538331	
LTC_REDU_TYPEB	571.1858	330.6480	1.727	0.084270	.
LTC_REDU_TYPEC	61.0214	170.0810	0.359	0.719807	
LTC_REDU_TYPED	340.9109	148.3721	2.298	0.021705	*
LTC_REDU_TYPEE	373.2679	524.7408	0.711	0.476975	
LTC_REDU_TYPEF	278.7597	259.7661	1.073	0.283375	
LTC_REDU_TYPEz	156.8671	240.9723	0.651	0.515153	
RCG_RCGT_SCORE	-4.6931	3.8931	-1.205	0.228184	
GJU_RCGT_GRADE_CD2	-124.1318	166.0273	-0.748	0.454772	
GJU_RCGT_GRADE_CD3	-161.9342	190.8048	-0.849	0.396177	
GJU_RCGT_GRADE_CD4	-122.8808	223.3182	-0.550	0.582222	
GJU_RCGT_GRADE_CD5	-80.6309	262.2683	-0.307	0.758550	
GJU_RCGT_GRADE_CD6	-60.0973	342.7225	-0.175	0.860824	
GJU_RCGT_GRADE_CDA	-136.1786	254.4386	-0.535	0.592575	
GJU_RCGT_GRADE_CDB	-3.5454	278.6527	-0.013	0.989850	
GJU_RCGT_GRADE_CDC	-542.4326	363.3673	-1.493	0.135682	

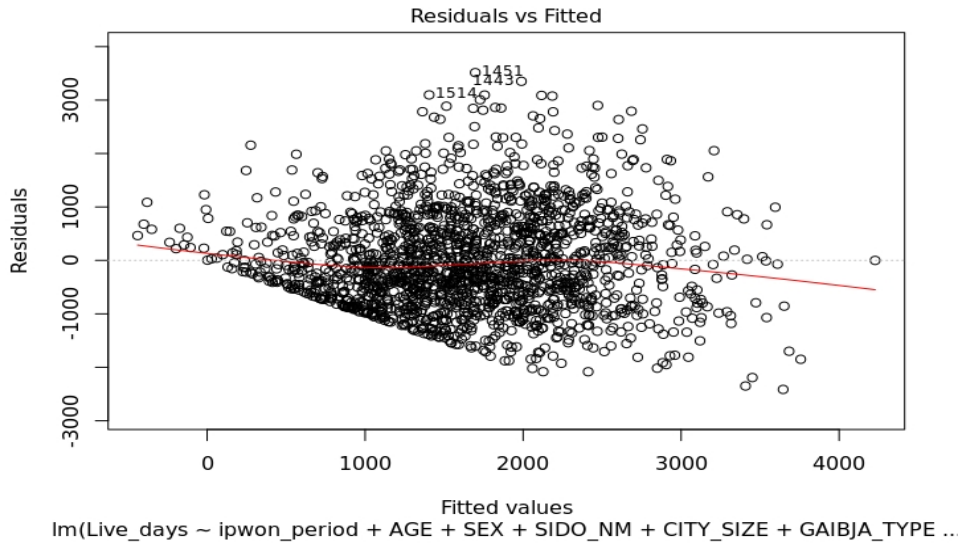
GJU_RCGT_GRADE_CDY	-331.5625	408.2347	-0.812	0.416802
GJU_RCGT_GRADE_CDZ	-236.9532	301.2813	-0.786	0.431696
DISM_JUDG_RSN_CD1	477.8714	420.9523	1.135	0.256450
DISM_JUDG_RSN_CD2	74.2796	334.1411	0.222	0.824108
DISM_JUDG_RSN_CD4	96.3132	221.3979	0.435	0.663603
DISM_JUDG_RSN_CD5	-335.2296	1006.0381	-0.333	0.739012
RSDN_TP_CD2	57.1229	149.0032	0.383	0.701497
RSDN_TP_CD3	221.9291	524.9678	0.423	0.672534
RSDN_TP_CD4	-230.2980	304.2392	-0.757	0.449180
RSDN_TP_CD5	-956.9544	724.6153	-1.321	0.186805
RSDN_TP_CD7	-126.6151	107.9206	-1.173	0.240875
RSDN_TP_CD9	-254.8127	152.6245	-1.670	0.095200
RSDN_TP_CDz	865.5284	875.6776	0.988	0.323097
LVTGT_PRSN_CD2	-27.4543	112.3930	-0.244	0.807051
LVTGT_PRSN_CD3	-571.6306	400.8112	-1.426	0.154004
LVTGT_PRSN_CD4	-45.4108	97.5176	-0.466	0.641515
LVTGT_PRSN_CD5	333.8198	513.7069	0.650	0.515895
LVTGT_PRSN_CD6	-751.4207	448.0793	-1.677	0.093735
LVTGT_PRSN_CD7	-672.3072	527.6158	-1.274	0.202759
LVTGT_PRSN_CD8	-79.2449	145.5618	-0.544	0.586235
LVTGT_PRSN_CD9	54.7061	122.8660	0.445	0.656198
LVTGT_PRSN_CDA	-34.0068	133.3668	-0.255	0.798765
LVTGT_PRSN_CDB	114.7500	725.2665	0.158	0.874305
LVTGT_PRSN_CDC	28.0817	246.4085	0.114	0.909280
LVTGT_PRSN_CDD	-83.9301	582.2026	-0.144	0.885392
LVTGT_PRSN_CDE	-1409.2819	984.9417	-1.431	0.152669
LVTGT_PRSN_CDF	-1412.4157	1164.8460	-1.213	0.225481
LVTGT_PRSN_CDG	-849.2058	587.8190	-1.445	0.148740
LVTGT_PRSN_CDJ	-60.6164	140.9676	-0.430	0.667250
LVTGT_PRSN_CDz	19.8907	326.7244	0.061	0.951463
CAREPSN_CD1	0.7979	134.7030	0.006	0.995274
CAREPSN_CD2	141.7183	592.6887	0.239	0.811049
CAREPSN_CD3	-63.1613	129.7049	-0.487	0.626350
CAREPSN_CD4	-373.6639	404.0843	-0.925	0.355248
CAREPSN_CD5	-29.8851	278.1229	-0.107	0.914443



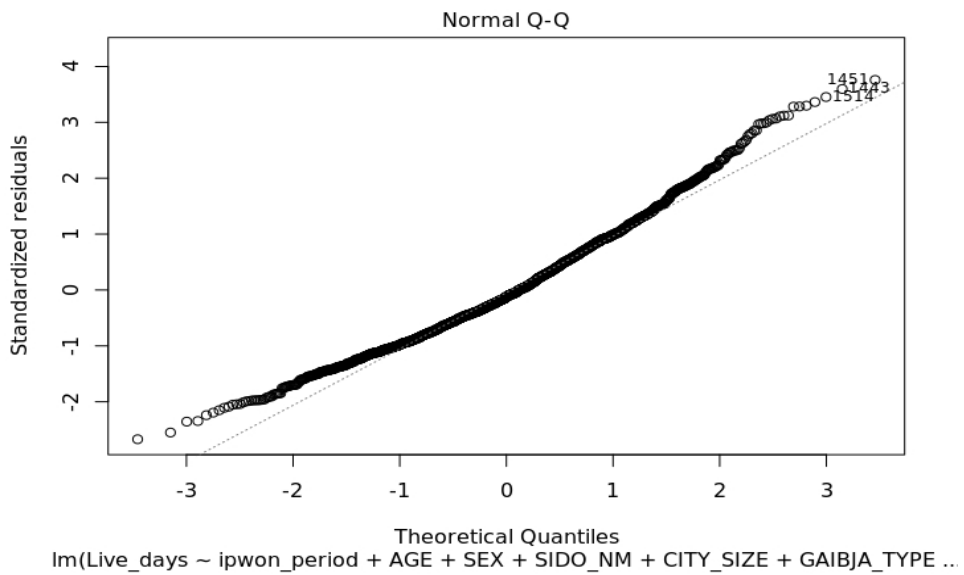
CAREPSN_CD6	4.0567	213.2952	0.019	0.984828
CAREPSN_CD7	32.9308	136.4878	0.241	0.809374
CAREPSN_CD8	437.1525	470.3137	0.929	0.352770
CAREPSN_CD9	-79.8920	161.4352	-0.495	0.620747
CAREPSN_CD10	183.9931	166.7400	1.103	0.269983
CAREPSN_CD11	208.3568	124.6582	1.671	0.094827 .
CAREPSN_CD12	101.4693	186.6556	0.544	0.586778
CAREPSN_CD13	-1140.1230	703.2460	-1.621	0.105159
CAREPSN_CD14	170.2883	316.3026	0.538	0.590393
PCT_ARFF_SCORE	0.4952	1.2649	0.391	0.695478
BHVR_ARFF_SCORE	-3.7985	2.4427	-1.555	0.120138
TDNC_ARFF_SCORE	-2.1934	3.3966	-0.646	0.518522
REHBT_ARFF_SCORE	0.9206	2.1165	0.435	0.663666
N051	75.5624	141.0611	0.536	0.592258
N071	-231.1378	120.0803	-1.925	0.054419 .
S102	-30.0416	78.4739	-0.383	0.701901
S103	-38.0932	105.4860	-0.361	0.718055
PE12	-62.0840	85.4171	-0.727	0.467431
PE13	-82.7342	103.4520	-0.800	0.423979
PE14	105.6002	139.0385	0.760	0.447660
PE22	117.4380	61.3054	1.916	0.055586 .
PE23	141.6781	96.3780	1.470	0.141745
PE24	94.9837	174.3252	0.545	0.585920
smoke_yn	-41.2864	39.8376	-1.036	0.300182
drink_yn	-104.1309	39.8645	-2.612	0.009080 **
ALBUMIN	-62.1594	53.8190	-1.155	0.248269
BUN	-2.0459	2.4735	-0.827	0.408281
CR	-3.1764	31.6714	-0.100	0.920124
GLUCOSE_AC	-2.0772	0.5241	-3.963	7.71e-05 ***
HB	40.5359	14.1124	2.872	0.004126 **
HBA1C	30.1110	20.0345	1.503	0.133041
jaewon_GW	-2.6196	1.5603	-1.679	0.093373 .
total_cowork	-17.7076	4.4740	-3.958	7.88e-05 ***
pum_cowork	6.9399	28.7328	0.242	0.809172
idm_cowork	9.4981	23.1025	0.411	0.681031

plot(result)

Hit <Return> to see next plot:



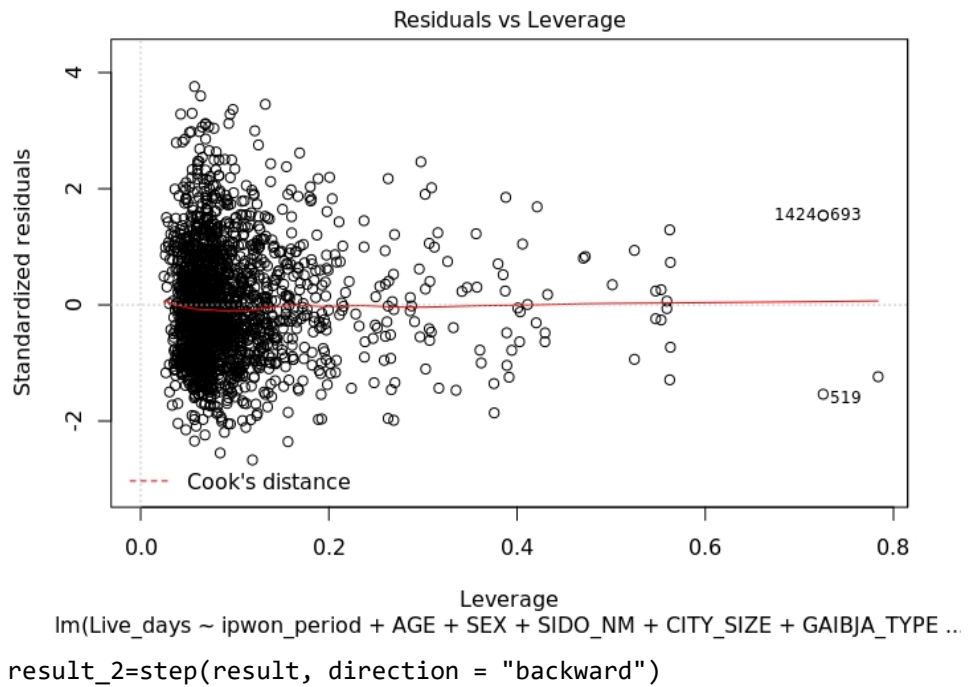
Hit <Return> to see next plot:



Hit <Return> to see next plot:



Hit <Return> to see next plot



```
summary(result_2)
```

```
Call:
```

```
lm(formula = Live_days ~ ipwon_period + AGE + SEX +  
CMPR_DSB_GRADE_L +  
    MAIN_DSB_TYPE_L + L_tube_feeding + SAH + SDH + ICH +  
Focal_small +  
    Chronic_old + ACA + MCA + Aneurysm + APFRM_TYPE_CD +  
LTC_REDU_TYPE +  
    RCG_RCGT_SCORE + BHVR_ARFF_SCORE + N07 + PE1 + PE2 +  
drink_yn +  
    GLUCOSE_AC + HB + HBA1C + jaewon_GW + total_cowork, data =  
Data_final)
```

```
Residuals:
```

```
    Min      1Q  Median      3Q      Max  
-2135.2 -689.9 -112.2  617.0  3426.4
```

```
Coefficients: (1 not defined because of singularities)
```

```
Estimate Std. Error t value Pr(>|t|)
```

```
(Intercept)      5077.6836   346.8428  14.640 < 2e-16 ***  
ipwon_period          1.4128    0.8081   1.748 0.080568 .  
AGE                  -49.5161    2.9412 -16.835 < 2e-16 ***  
SEX2                  129.4705    54.7291   2.366 0.018104 *  
CMPR_DSB_GRADE_L1    396.7928    587.4885   0.675 0.499506  
CMPR_DSB_GRADE_L2    219.6132    586.3539   0.375 0.708047  
CMPR_DSB_GRADE_L3   -169.1423    600.0424  -0.282 0.778065  
CMPR_DSB_GRADE_L4   -436.9543    565.0257  -0.773 0.439426  
CMPR_DSB_GRADE_L5   -178.6170    597.8864  -0.299 0.765168  
CMPR_DSB_GRADE_L6    141.2627    610.0269   0.232 0.816900  
MAIN_DSB_TYPE_L1     -46.1839    583.6284  -0.079 0.936936  
MAIN_DSB_TYPE_L2     -74.1247    584.4140  -0.127 0.899085  
MAIN_DSB_TYPE_L3    -185.1147    599.2492  -0.309 0.757425  
MAIN_DSB_TYPE_L4      92.7753    588.9388   0.158 0.874845  
MAIN_DSB_TYPE_L5    -305.7849    622.5590  -0.491 0.623364  
MAIN_DSB_TYPE_L6     -75.2090    690.3928  -0.109 0.913265
```

MAIN_DSB_TYPE_L8	-825.6231	692.5412	-1.192	0.233355
MAIN_DSB_TYPE_L9	-719.4562	594.5077	-1.210	0.226373
MAIN_DSB_TYPE_L10	484.7017	761.1522	0.637	0.524337
MAIN_DSB_TYPE_L11	-175.2220	900.3748	-0.195	0.845720
MAIN_DSB_TYPE_L14	NA	NA	NA	NA
L_tube_feeding1	-151.3631	54.9232	-2.756	0.005913 **
SAH1	-148.3017	105.1710	-1.410	0.158684
SDH1	167.1286	85.7826	1.948	0.051538 .
ICH1	-102.3826	56.2165	-1.821	0.068741 .
Focal_small1	172.2086	52.0031	3.312	0.000946 ***
Chronic_old1	194.3779	55.9647	3.473	0.000527 ***
ACA1	-177.5422	92.1138	-1.927	0.054085 .
MCA1	-105.3033	54.4726	-1.933	0.053377 .
Aneurysm1	-164.1995	78.1924	-2.100	0.035874 *
APFRM_TYPE_CD2	216.3502	127.4583	1.697	0.089792 .
APFRM_TYPE_CD3	281.7525	53.4401	5.272	1.51e-07 ***
APFRM_TYPE_CD4	187.5188	76.6419	2.447	0.014513 *
APFRM_TYPE_CD5	591.5579	705.5468	0.838	0.401897
LTC_REDU_TYPEB	560.7343	253.8289	2.209	0.027294 *
LTC_REDU_TYPEC	213.3411	71.9335	2.966	0.003059 **
LTC_REDU_TYPED	459.1180	91.7276	5.005	6.13e-07 ***
LTC_REDU_TYPEE	152.1132	494.5162	0.308	0.758423
LTC_REDU_TYPEF	309.4150	232.7832	1.329	0.183952
LTC_REDU_TYPEz	188.5359	203.6158	0.926	0.354603
RCG_RCGT_SCORE	-4.6690	1.6287	-2.867	0.004197 **
BHVR_ARFF_SCORE	-4.3551	2.2579	-1.929	0.053908 .
N071	-274.3883	102.2034	-2.685	0.007326 **
PE12	-55.9431	78.3701	-0.714	0.475424
PE13	-115.2429	89.0561	-1.294	0.195816
PE14	78.9930	123.1016	0.642	0.521157
PE22	148.8670	54.8403	2.715	0.006700 **
PE23	186.4577	86.6774	2.151	0.031597 *
PE24	89.8114	157.6710	0.570	0.569012
drink_yn	-106.6997	37.7562	-2.826	0.004766 **
GLUCOSE_AC	-2.2760	0.5015	-4.538	6.06e-06 ***

```

HB          38.4000    12.0596    3.184 0.001477 **
HBA1C       32.8301    19.0798    1.721 0.085484 .
jaewon_GW   -3.4102     1.5059   -2.265 0.023659 *
total_cowork -15.7902     3.4263   -4.609 4.34e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

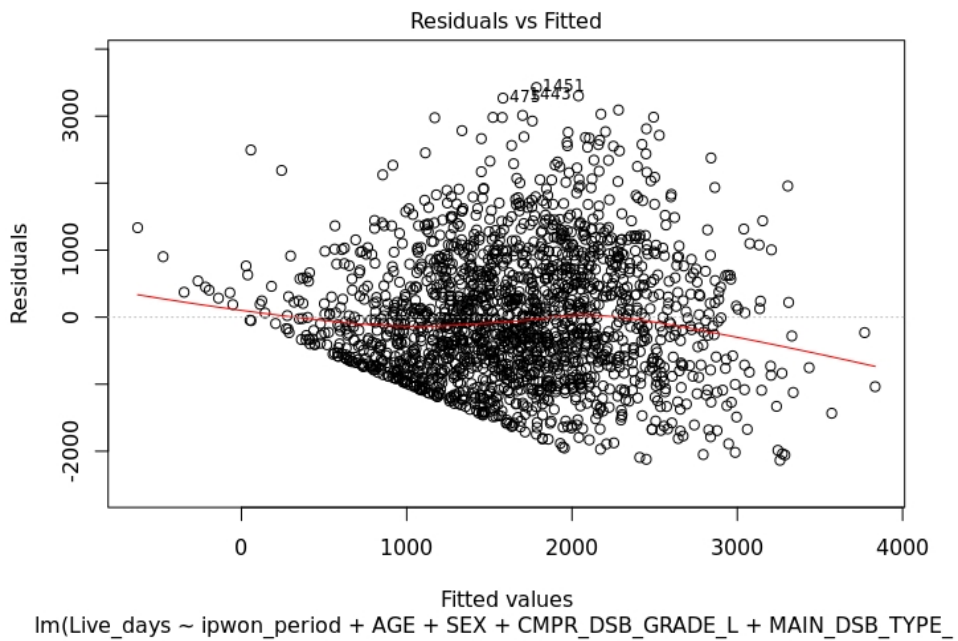
```

Residual standard error: 964.8 on 1785 degrees of freedom  
Multiple R-squared: 0.2988, Adjusted R-squared: 0.278  
F-statistic: 14.35 on 53 and 1785 DF, p-value: < 2.2e-16

```

plot(result_2)
Hit <Return> to see next plot:

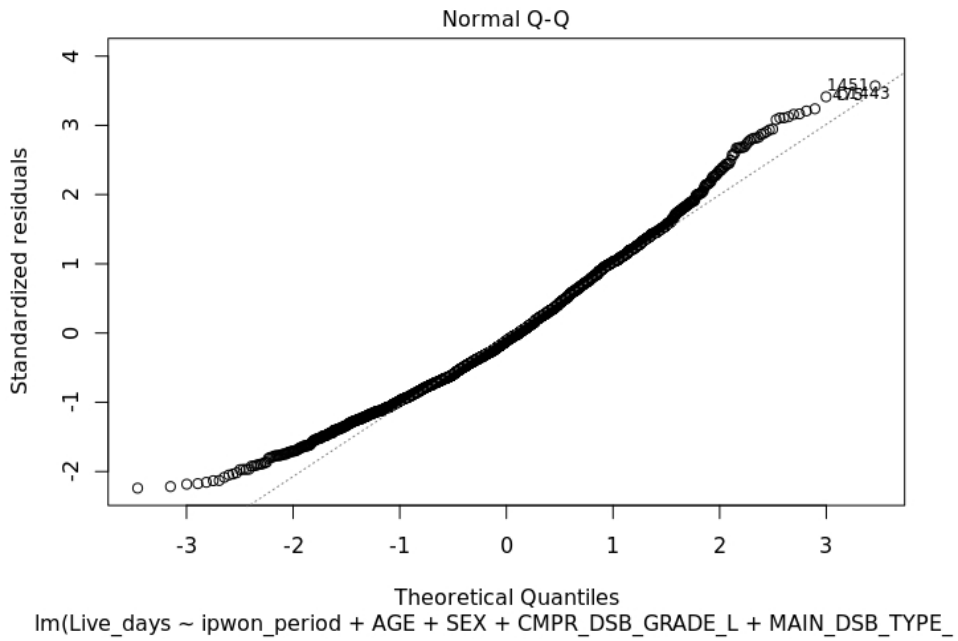
```



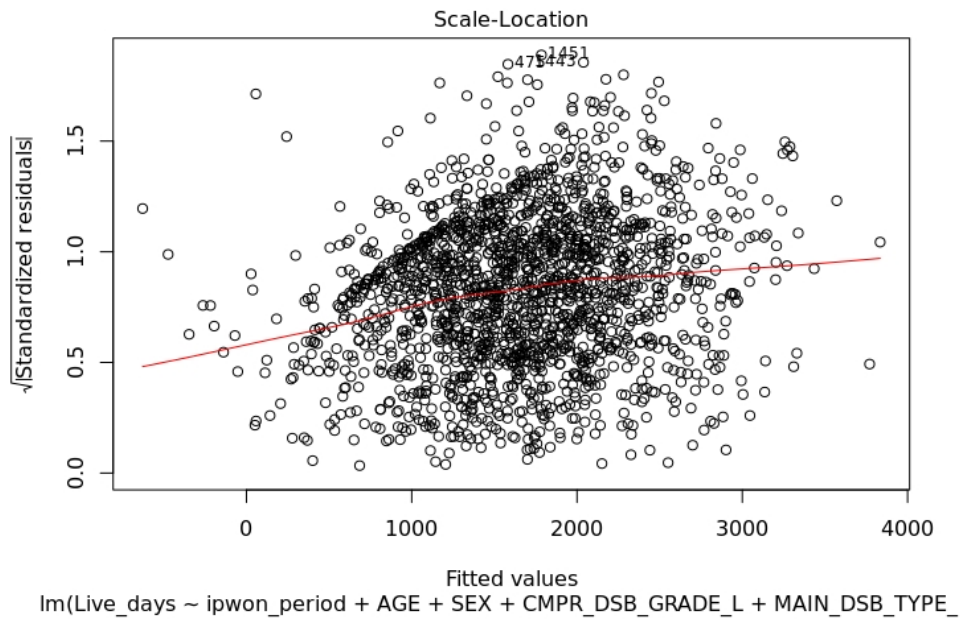
```

Hit <Return> to see next plot:

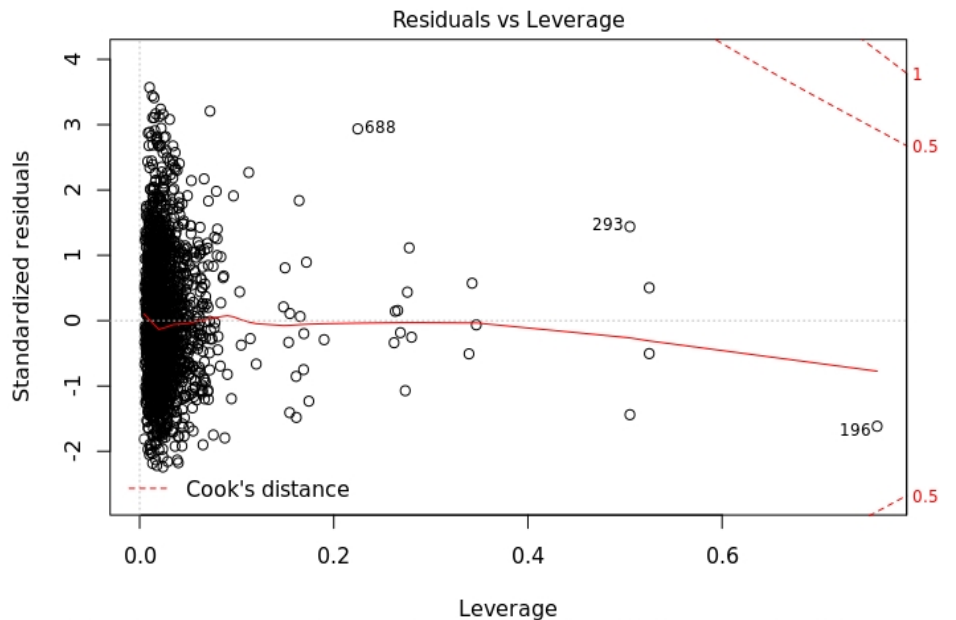
```



Hit <Return> to see next plot:



Hit <Return> to see next plot:



lm(Live\_days ~ ipwon\_period + AGE + SEX + CMPR\_DSB\_GRADE\_L + MAIN\_DSB\_TYPE\_

anova(result, result\_2)

Analysis of Variance Table

Model 1: Live\_days ~ ipwon\_period + AGE + SEX + SIDO\_NM +  
 CITY\_SIZE +  
 GAIBJA\_TYPE + CMPR\_DSB\_GRADE + MAIN\_DSB\_TYPE +  
 CMPR\_DSB\_GRADE\_L +  
 MAIN\_DSB\_TYPE\_L + L\_tube\_feeding + pneumonia + Infarction +  
 SAH + SDH + ICH + IVH + Focal\_small + multifocal\_large +  
 Acute\_recent + Chronic\_old + Temporal + Occipital + Cerebellum +  
 BG + brain\_stem + Thalamus + ACA + MCA + PCA + Aneurysm +  
 APFRM\_TYPE\_CD + LTC\_OBJTR\_TYPE\_CD + LTC\_REDU\_TYPE +  
 RCG\_RCGT\_SCORE +  
 GJU\_RCGT\_GRADE\_CD + DISM\_JUDG\_RSN\_CD + RSDN\_TP\_CD +  
 LVTGT\_PRSN\_CD +  
 CAREPSN\_CD + PCT\_ARFF\_SCORE + BHVR\_ARFF\_SCORE +  
 TNDNC\_ARFF\_SCORE +



```

REHBT_ARFF_SCORE + N05 + N07 + S10 + PE1 + PE2 + smoke_yn +
drink_yn + ALBUMIN + BUN + CR + GLUCOSE_AC + HB + HBA1C +
jaewon_GW + total_cowork + pum_cowork + idm_cowork
Model 2: Live_days ~ ipwon_period + AGE + SEX + CMPR_DSB_GRADE_L
+ MAIN_DSB_TYPE_L +
L_tube_feeding + SAH + SDH + ICH + Focal_small + Chronic_old +
ACA + MCA + Aneurysm + APFRM_TYPE_CD + LTC_REDU_TYPE +
RCG_RCGT_SCORE +
BHVR_ARFF_SCORE + N07 + PE1 + PE2 + drink_yn + GLUCOSE_AC +
HB + HBA1C + jaewon_GW + total_cowork
Res.Df      RSS    Df  Sum of Sq    F Pr(>F)
1   1651 1530183019
2   1785 1661452961 -134 -131269942 1.057 0.3173

```

DEATH\_YN

```

result=lm(DEATH_YN~ipwon_period+AGE+SEX+SIDO_NM+CITY_SIZE+GAIBJA_
TYPE+CMPR_DSB_GRADE+MAIN_DSB_TYPE+CMPR_DSB_GRADE_L+MAIN_DSB_TYPE_
L+L_tube_feeding+pneumonia+Infarction+SAH+SDH+ICH+IVH+Focal_small
+multifocal_large+Acute_recent+Chronic_old+Temporal+Occipital+Cer
ebellum+BG+brain_stem+Thalamus+ACA+MCA+PCA+Aneurysm+APFRM_TYPE_CD
+LTC_OBJTR_TYPE_CD+LTC_REDU_TYPE+RCG_RCGT_SCORE+GJU_RCGT_GRADE_CD
+DISM_JUDG_RSN_CD+RSDN_TP_CD+LVTGT_PRSN_CD+CAREPSN_CD+PCT_ARFF_SC
ORE+BHVR_ARFF_SCORE+TNDNC_ARFF_SCORE+REHBT_ARFF_SCORE+N05+N07+S10
+PE1+PE2+smoke_yn+drink_yn+ALBUMIN+BUN+CR+GLUCOSE_AC+HB+HBA1C+jae
won_GW+total_cowork+pum_cowork+idm_cowork, data=Data_final)

```

```
summary(result)
```

Call:

```

lm(formula = DEATH_YN ~ ipwon_period + AGE + SEX + SIDO_NM +
    CITY_SIZE + GAIBJA_TYPE + CMPR_DSB_GRADE + MAIN_DSB_TYPE +
    CMPR_DSB_GRADE_L + MAIN_DSB_TYPE_L + L_tube_feeding + pneumonia +
    Infarction + SAH + SDH + ICH + IVH + Focal_small +
    multifocal_large +

```

```

Acute_recent + Chronic_old + Temporal + Occipital + Cerebellum +
BG + brain_stem + Thalamus + ACA + MCA + PCA + Aneurysm +
APFRM_TYPE_CD + LTC_OBJTR_TYPE_CD + LTC_REDU_TYPE +
RCG_RCGT_SCORE +
GJU_RCGT_GRADE_CD + DISM_JUDG_RSN_CD + RSDN_TP_CD +
LVTGT_PRSN_CD +
CAREPSN_CD + PCT_ARFF_SCORE + BHVR_ARFF_SCORE +
TNDNC_ARFF_SCORE +
REHBT_ARFF_SCORE + N05 + N07 + S10 + PE1 + PE2 + smoke_yn +
drink_yn + ALBUMIN + BUN + CR + GLUCOSE_AC + HB + HBA1C +
jaewon_GW + total_cowork + pum_cowork + idm_cowork, data =
Data_final)

```

Residuals:

Min	1Q	Median	3Q	Max
-1.04275	-0.23403	0.01368	0.24931	0.85850

Coefficients: (6 not defined because of singularities)

Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-0.0749545	0.2472566	-0.303 0.761818
ipwon_period	-0.0001875	0.0003089	-0.607 0.543880
AGE	0.0087522	0.0012036	7.271 5.47e-13 ***
SEX2	-0.0905052	0.0238857	-3.789 0.000157 ***
SIDO_NM2	0.0136070	0.1648084	0.083 0.934210
SIDO_NM3	0.4076940	0.2160245	1.887 0.059301 .
SIDO_NM4	0.0891098	0.0640146	1.392 0.164103
SIDO_NM5	0.0969725	0.1838683	0.527 0.597985
SIDO_NM6	0.2152389	0.2156029	0.998 0.318275
SIDO_NM7	0.5907382	0.3739617	1.580 0.114373
SIDO_NM9	0.0158669	0.0258292	0.614 0.539100
SIDO_NM10	-0.0098744	0.0922213	-0.107 0.914744
SIDO_NM11	-0.1763550	0.1089886	-1.618 0.105831
SIDO_NM12	0.0516570	0.0792451	0.652 0.514580
SIDO_NM13	-0.0212152	0.1058044	-0.201 0.841104
SIDO_NM14	-0.0015012	0.0833131	-0.018 0.985626

SIDO_NM15	0.0851401	0.1189667	0.716	0.474300
SIDO_NM16	-0.2653334	0.1468599	-1.807	0.070989 .
SIDO_NM17	-0.2038732	0.2640091	-0.772	0.440094
CITY_SIZE2	0.0033205	0.0342779	0.097	0.922841
CITY_SIZE3	0.0198470	0.0713449	0.278	0.780906
CITY_SIZE4	NA	NA	NA	NA
CITY_SIZE5	-0.0985391	0.0427460	-2.305	0.021277 *
CITY_SIZE6	-0.0474230	0.0272268	-1.742	0.081734 .
CITY_SIZE7	-0.0853733	0.1252621	-0.682	0.495615
CITY_SIZE8	-0.1600742	0.1442160	-1.110	0.267177
GAIBJA_TYPE1	-0.2338511	0.1299608	-1.799	0.072138 .
GAIBJA_TYPE2	-0.1435128	0.1337100	-1.073	0.283287
GAIBJA_TYPE3	-0.1026572	0.1377136	-0.745	0.456112
GAIBJA_TYPE4	-0.1249214	0.1356741	-0.921	0.357317
GAIBJA_TYPE5	-0.2229138	0.1334735	-1.670	0.095090 .
GAIBJA_TYPE6	-0.1685780	0.1359045	-1.240	0.214998
GAIBJA_TYPE7	-0.2049709	0.1342862	-1.526	0.127109
GAIBJA_TYPE8	-0.2168006	0.1333288	-1.626	0.104128
GAIBJA_TYPE9	-0.2745645	0.1360883	-2.018	0.043800 *
GAIBJA_TYPE10	-0.1182784	0.1309906	-0.903	0.366682
GAIBJA_TYPE11	-0.2773665	0.1351443	-2.052	0.040291 *
GAIBJA_TYPE12	-0.2508546	0.1336882	-1.876	0.060774 .
GAIBJA_TYPE13	-0.2323957	0.1310783	-1.773	0.076421 .
GAIBJA_TYPE14	-0.2484243	0.1314004	-1.891	0.058854 .
GAIBJA_TYPE15	-0.2807196	0.1314978	-2.135	0.032925 *
GAIBJA_TYPE16	-0.2329651	0.1313951	-1.773	0.076411 .
GAIBJA_TYPE17	-0.2231500	0.1294851	-1.723	0.085010 .
GAIBJA_TYPE18	-0.2830526	0.1289443	-2.195	0.028291 *
GAIBJA_TYPE19	-0.2366349	0.1291454	-1.832	0.067085 .
GAIBJA_TYPE20	-0.2215181	0.1281466	-1.729	0.084062 .
CMPR_DSB_GRADE1	0.5121537	0.2460720	2.081	0.037559 *
CMPR_DSB_GRADE2	0.4495611	0.2449718	1.835	0.066663 .
CMPR_DSB_GRADE3	0.2850462	0.2444239	1.166	0.243703
CMPR_DSB_GRADE4	0.2735961	0.2411414	1.135	0.256713
CMPR_DSB_GRADE5	0.2263739	0.2426436	0.933	0.350983

CMPR_DSB_GRADE6	0.1896700	0.2379467	0.797	0.425501	
CMPR_DSB_GRADE10	0.2530444	0.2719881	0.930	0.352325	
CMPR_DSB_GRADE20	0.1621154	0.2602980	0.623	0.533498	
MAIN_DSB_TYPE1	-0.0448176	0.2698558	-0.166	0.868114	
MAIN_DSB_TYPE2	-0.1984578	0.2707460	-0.733	0.463660	
MAIN_DSB_TYPE3	-0.2638677	0.2677081	-0.986	0.324447	
MAIN_DSB_TYPE4	-0.0154897	0.2815949	-0.055	0.956140	
MAIN_DSB_TYPE5	-0.3154352	0.2921580	-1.080	0.280445	
MAIN_DSB_TYPE6	-0.7900320	0.3646078	-2.167	0.030393	*
MAIN_DSB_TYPE8	-0.1738002	0.2716396	-0.640	0.522379	
MAIN_DSB_TYPE9	-0.3476624	0.2719006	-1.279	0.201204	
MAIN_DSB_TYPE10	0.0236191	0.3032680	0.078	0.937932	
MAIN_DSB_TYPE11	0.5408936	0.3509635	1.541	0.123468	
MAIN_DSB_TYPE12	-0.0056106	0.4538313	-0.012	0.990138	
MAIN_DSB_TYPE14	NA	NA	NA	NA	
CMPR_DSB_GRADE_L1	-0.6793856	0.0837367	-8.113	9.51e-16	***
CMPR_DSB_GRADE_L2	-0.4849743	0.0766366	-6.328	3.19e-10	***
CMPR_DSB_GRADE_L3	-0.0718105	0.0965791	-0.744	0.457260	
CMPR_DSB_GRADE_L4	-0.0595474	0.0964206	-0.618	0.536937	
CMPR_DSB_GRADE_L5	0.0029351	0.0943820	0.031	0.975195	
CMPR_DSB_GRADE_L6	NA	NA	NA	NA	
MAIN_DSB_TYPE_L1	0.1423640	0.3521378	0.404	0.686056	
MAIN_DSB_TYPE_L2	0.3019030	0.3513216	0.859	0.390280	
MAIN_DSB_TYPE_L3	0.3315351	0.3465271	0.957	0.338840	
MAIN_DSB_TYPE_L4	0.0344330	0.3604450	0.096	0.923906	
MAIN_DSB_TYPE_L5	0.3519695	0.3631505	0.969	0.332582	
MAIN_DSB_TYPE_L6	0.6476993	0.3894106	1.663	0.096446	.
MAIN_DSB_TYPE_L8	NA	NA	NA	NA	
MAIN_DSB_TYPE_L9	0.4773166	0.3473760	1.374	0.169609	
MAIN_DSB_TYPE_L10	0.1507370	0.4097549	0.368	0.713016	
MAIN_DSB_TYPE_L11	NA	NA	NA	NA	
MAIN_DSB_TYPE_L14	NA	NA	NA	NA	
L_tube_feeding1	0.0778117	0.0219075	3.552	0.000393	***
pneumonia1	0.0513241	0.0194400	2.640	0.008366	**
Infarction1	0.0016152	0.0308915	0.052	0.958308	

SAH1	0.0022644	0.0418229	0.054	0.956829	
SDH1	-0.0311796	0.0333120	-0.936	0.349417	
ICH1	-0.0794574	0.0234575	-3.387	0.000722	***
IVH1	0.0615459	0.0380777	1.616	0.106216	
Focal_small1	-0.0272277	0.0221927	-1.227	0.220043	
multifocal_large1	0.0523067	0.0201263	2.599	0.009435	**
Acute_recent1	-0.0215569	0.0257435	-0.837	0.402504	
Chronic_old1	-0.0024657	0.0267512	-0.092	0.926573	
Temporal1	0.0837244	0.0320929	2.609	0.009168	**
Occipital1	-0.0097721	0.0336513	-0.290	0.771553	
Cerebellum1	0.0051770	0.0229364	0.226	0.821454	
BG1	0.0233426	0.0210017	1.111	0.266531	
brain_stem1	-0.0215494	0.0279746	-0.770	0.441220	
Thalamus1	0.0531042	0.0241407	2.200	0.027961	*
ACA1	0.0249802	0.0353258	0.707	0.479580	
MCA1	-0.0028905	0.0215920	-0.134	0.893524	
PCA1	-0.0248472	0.0328236	-0.757	0.449163	
Aneurysm1	-0.0182893	0.0298562	-0.613	0.540240	
APFRM_TYPE_CD2	-0.0712565	0.0497687	-1.432	0.152403	
APFRM_TYPE_CD3	-0.0086839	0.0226639	-0.383	0.701652	
APFRM_TYPE_CD4	-0.0199628	0.0295968	-0.674	0.500094	
APFRM_TYPE_CD5	0.1264495	0.3117034	0.406	0.685036	
LTC_OBJTR_TYPE_CD	-0.0756258	0.0279797	-2.703	0.006945	**
LTC_REDU_TYPEB	0.1342851	0.1209631	1.110	0.267103	
LTC_REDU_TYPEC	0.2539583	0.0622218	4.081	4.69e-05	***
LTC_REDU_TYPED	-0.1822789	0.0542799	-3.358	0.000803	***
LTC_REDU_TYPEE	-0.3382706	0.1919693	-1.762	0.078236	.
LTC_REDU_TYPEF	-0.2346241	0.0950319	-2.469	0.013654	*
LTC_REDU_TYPEz	0.2095439	0.0881564	2.377	0.017570	*
RCG_RCGT_SCORE	0.0038613	0.0014242	2.711	0.006775	**
GJU_RCGT_GRADE_CD2	0.1292056	0.0607388	2.127	0.033549	*
GJU_RCGT_GRADE_CD3	0.1288874	0.0698033	1.846	0.065008	.
GJU_RCGT_GRADE_CD4	0.1179952	0.0816979	1.444	0.148848	
GJU_RCGT_GRADE_CD5	0.0235150	0.0959473	0.245	0.806423	
GJU_RCGT_GRADE_CD6	0.0143478	0.1253804	0.114	0.908908	

GJU_RCGT_GRADE_CDA	0.1714592	0.0930829	1.842	0.065654	.
GJU_RCGT_GRADE_CDB	0.0850162	0.1019413	0.834	0.404418	
GJU_RCGT_GRADE_CDC	0.2100388	0.1329330	1.580	0.114290	
GJU_RCGT_GRADE_CDY	0.1410373	0.1493471	0.944	0.345124	
GJU_RCGT_GRADE_CDZ	0.2684481	0.1102196	2.436	0.014973	*
DISM_JUDG_RSN_CD1	-0.2280696	0.1539997	-1.481	0.138804	
DISM_JUDG_RSN_CD2	-0.0252365	0.1222410	-0.206	0.836466	
DISM_JUDG_RSN_CD4	-0.0755957	0.0809954	-0.933	0.350785	
DISM_JUDG_RSN_CD5	0.1831616	0.3680453	0.498	0.618790	
RSDN_TP_CD2	0.0992113	0.0545108	1.820	0.068935	.
RSDN_TP_CD3	0.4324958	0.1920523	2.252	0.024455	*
RSDN_TP_CD4	0.2828712	0.1113018	2.541	0.011129	*
RSDN_TP_CD5	0.0934481	0.2650906	0.353	0.724498	
RSDN_TP_CD7	0.0640846	0.0394813	1.623	0.104745	
RSDN_TP_CD9	-0.0723847	0.0558356	-1.296	0.195022	
RSDN_TP_CDz	-0.1513204	0.3203547	-0.472	0.636737	
LVTGT_PRSN_CD2	0.0514703	0.0411174	1.252	0.210824	
LVTGT_PRSN_CD3	0.2906142	0.1466313	1.982	0.047652	*
LVTGT_PRSN_CD4	0.0468337	0.0356755	1.313	0.189443	
LVTGT_PRSN_CD5	-0.0232006	0.1879327	-0.123	0.901764	
LVTGT_PRSN_CD6	-0.0215312	0.1639237	-0.131	0.895516	
LVTGT_PRSN_CD7	0.2034848	0.1930210	1.054	0.291941	
LVTGT_PRSN_CD8	-0.0617296	0.0532518	-1.159	0.246542	
LVTGT_PRSN_CD9	-0.0504810	0.0449489	-1.123	0.261568	
LVTGT_PRSN_CDA	0.0322822	0.0487904	0.662	0.508288	
LVTGT_PRSN_CDB	-0.2995948	0.2653289	-1.129	0.259001	
LVTGT_PRSN_CDC	0.0499107	0.0901452	0.554	0.579880	
LVTGT_PRSN_CDD	-0.2219543	0.2129909	-1.042	0.297526	
LVTGT_PRSN_CDE	0.2713867	0.3603275	0.753	0.451457	
LVTGT_PRSN_CDF	0.5451015	0.4261430	1.279	0.201023	
LVTGT_PRSN_CDG	0.0681104	0.2150456	0.317	0.751492	
LVTGT_PRSN_CDJ	0.0855034	0.0515711	1.658	0.097513	.
LVTGT_PRSN_CDz	0.1765078	0.1195277	1.477	0.139944	
CAREPSN_CD1	0.0127394	0.0492793	0.259	0.796042	
CAREPSN_CD2	-0.4211991	0.2168271	-1.943	0.052239	.

CAREPSN_CD3	0.1824456	0.0474508	3.845	0.000125	***
CAREPSN_CD4	0.0707573	0.1478287	0.479	0.632255	
CAREPSN_CD5	0.0351509	0.1017475	0.345	0.729784	
CAREPSN_CD6	0.0623479	0.0780311	0.799	0.424398	
CAREPSN_CD7	0.0505225	0.0499322	1.012	0.311771	
CAREPSN_CD8	0.0480338	0.1720579	0.279	0.780147	
CAREPSN_CD9	0.1188410	0.0590589	2.012	0.044356	*
CAREPSN_CD10	-0.1272392	0.0609996	-2.086	0.037141	*
CAREPSN_CD11	-0.0697760	0.0456045	-1.530	0.126203	
CAREPSN_CD12	-0.1315232	0.0682854	-1.926	0.054266	.
CAREPSN_CD13	-0.0391878	0.2572730	-0.152	0.878953	
CAREPSN_CD14	-0.1385628	0.1157150	-1.197	0.231304	
PCT_ARFF_SCORE	0.0003319	0.0004628	0.717	0.473334	
BHVR_ARFF_SCORE	0.0014258	0.0008936	1.595	0.110793	
TNDNC_ARFF_SCORE	0.0019907	0.0012426	1.602	0.109332	
REHBT_ARFF_SCORE	0.0003066	0.0007743	0.396	0.692225	
N051	-0.0028535	0.0516053	-0.055	0.955911	
N071	-0.0355150	0.0439298	-0.808	0.418948	
S102	0.0775235	0.0287086	2.700	0.006997	**
S103	0.0784439	0.0385906	2.033	0.042240	*
PE12	0.0139346	0.0312487	0.446	0.655709	
PE13	0.0374481	0.0378465	0.989	0.322577	
PE14	0.0480970	0.0508653	0.946	0.344503	
PE22	-0.0221059	0.0224277	-0.986	0.324450	
PE23	-0.0335229	0.0352586	-0.951	0.341859	
PE24	-0.1661552	0.0637745	-2.605	0.009260	**
smoke_yn	0.0017357	0.0145740	0.119	0.905213	
drink_yn	-0.0210068	0.0145839	-1.440	0.149941	
ALBUMIN	-0.0885391	0.0196889	-4.497	7.38e-06	***
BUN	-0.0002553	0.0009049	-0.282	0.777887	
CR	0.0113366	0.0115865	0.978	0.328008	
GLUCOSE_AC	0.0003510	0.0001918	1.831	0.067350	.
HB	-0.0079499	0.0051628	-1.540	0.123794	
HBA1C	0.0168605	0.0073293	2.300	0.021549	*
jaewon_GW	-0.0005975	0.0005708	-1.047	0.295395	

```

total_cowork      -0.0009599  0.0016368  -0.586  0.557636
pum_cowork        0.0073177  0.0105115   0.696  0.486426
idm_cowork        0.0065220  0.0084518   0.772  0.440417

```

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

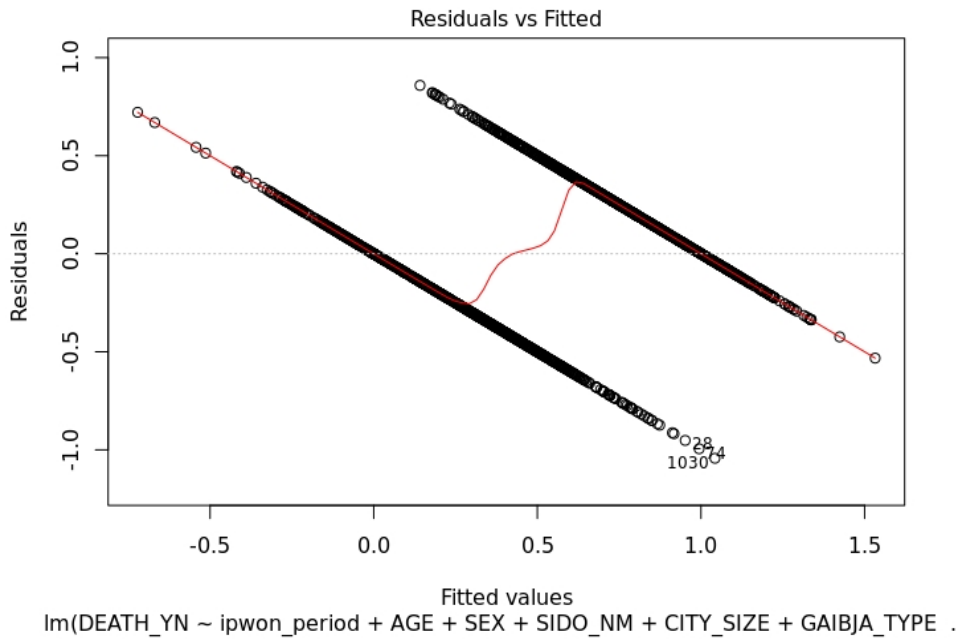
Residual standard error: 0.3522 on 1651 degrees of freedom

Multiple R-squared: 0.5536, Adjusted R-squared: 0.503

F-statistic: 10.95 on 187 and 1651 DF, p-value: < 2.2e-16

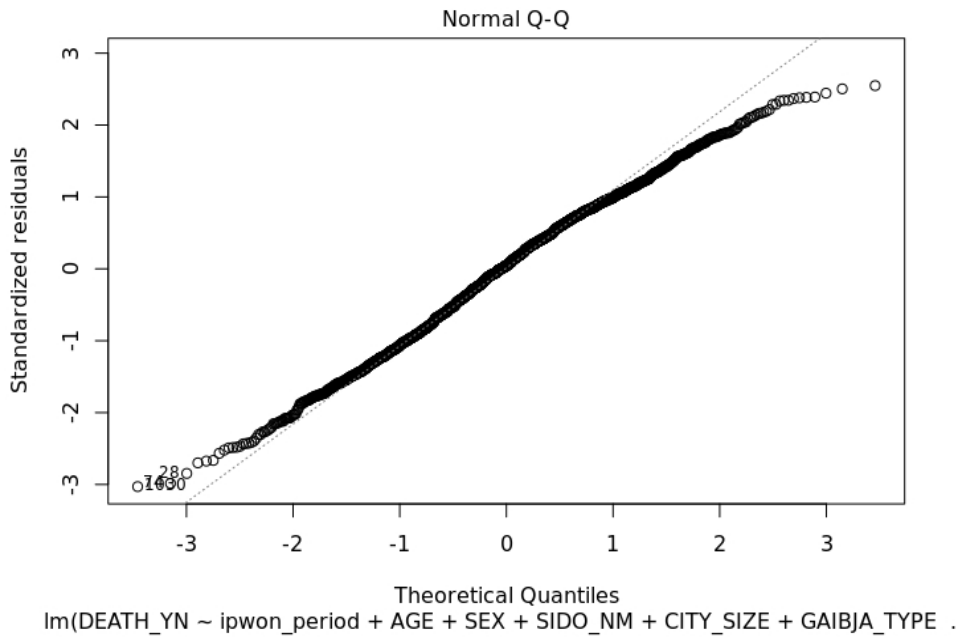
plot(result)

Hit <Return> to see next plot:

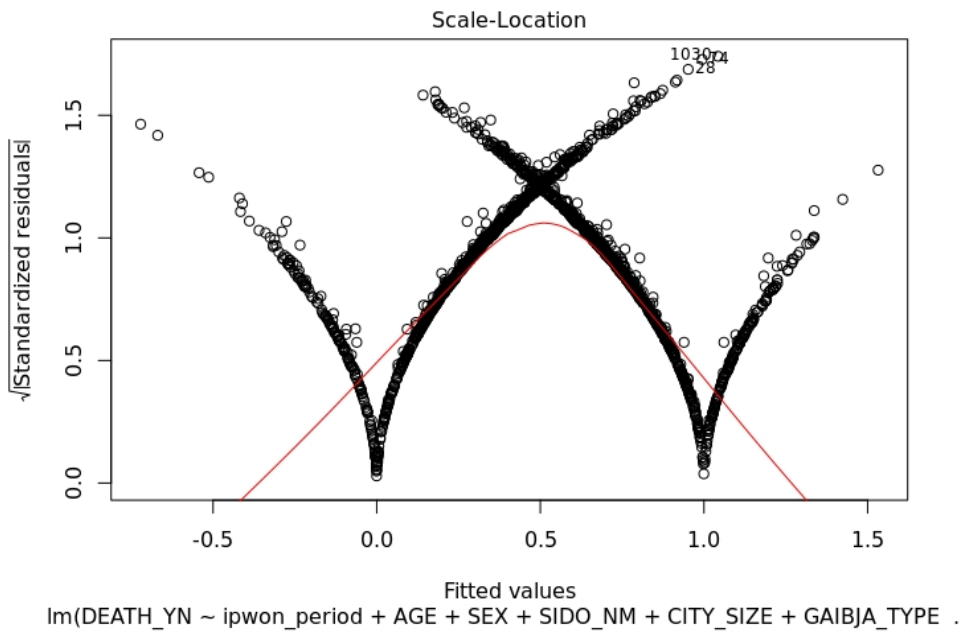


Hit <Return> to see next plot:

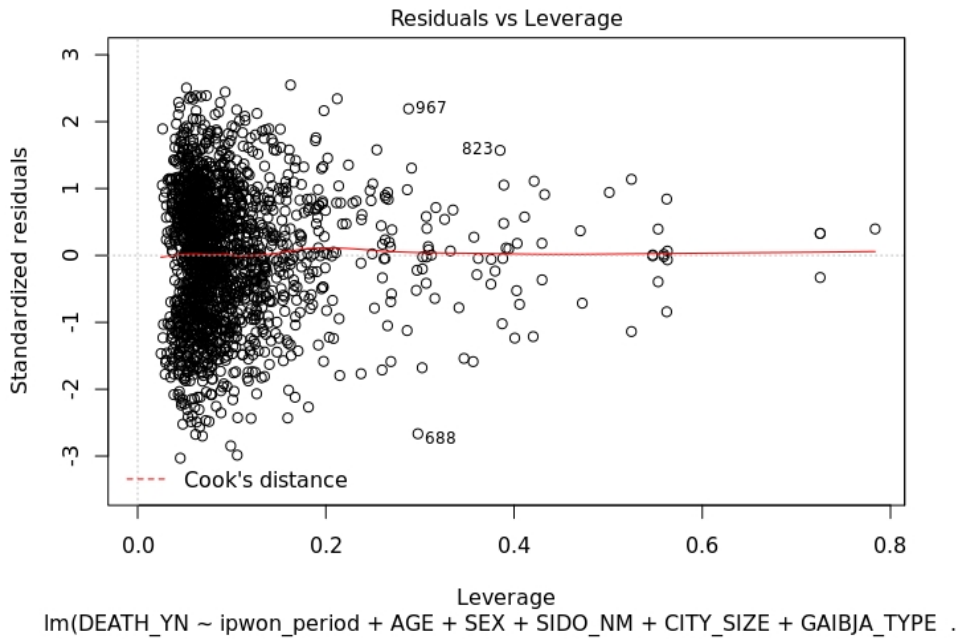




Hit <Return> to see next plot:



Hit <Return> to see next plot:



```
result_2=step(result, direction="backward")
```

```
summary(result_2)
```

Call:

```
lm(formula = DEATH_YN ~ ipwon_period + AGE + SEX + CITY_SIZE +
    CMPR_DSB_GRADE + MAIN_DSB_TYPE + CMPR_DSB_GRADE_L +
    MAIN_DSB_TYPE_L +
    L_tube_feeding + pneumonia + ICH + IVH + Focal_small +
    multifocal_large +
    Temporal + BG + Thalamus + LTC_OBJTR_TYPE_CD +
    LTC_REDU_TYPE +
    RCG_RCGT_SCORE + GJU_RCGT_GRADE_CD + CAREPSN_CD +
    BHVR_ARFF_SCORE +
    S10 + PE2 + ALBUMIN + CR + GLUCOSE_AC + HB + HBA1C, data =
    Data_final)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.07292	-0.25557	0.02202	0.25282	1.05814

Coefficients: (5 not defined because of singularities)

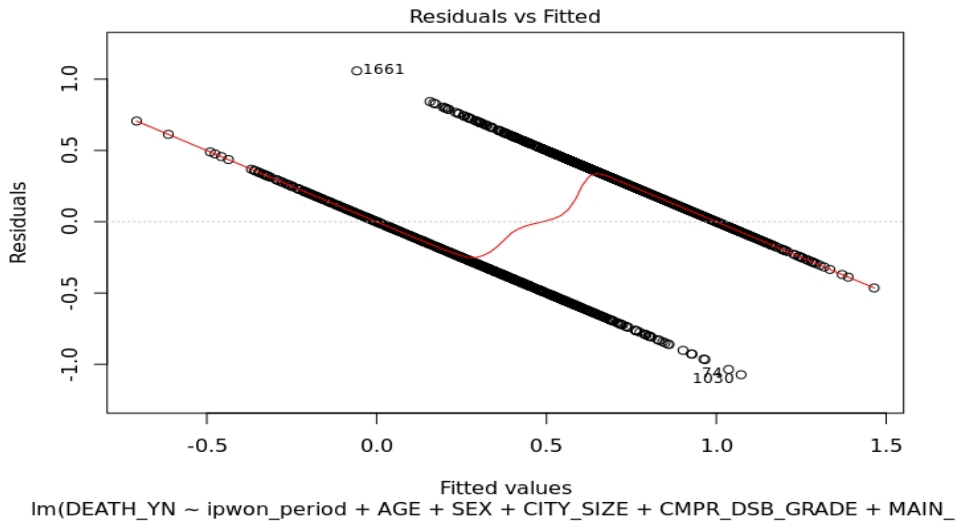
	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-0.3147760	0.1965498	-1.602	0.109446	
ipwon_period	-0.0003945	0.0002275	-1.734	0.083118	.
AGE	0.0091567	0.0011352	8.066	1.33e-15	***
SEX2	-0.0749876	0.0204767	-3.662	0.000258	***
CITY_SIZE2	0.0134672	0.0328752	0.410	0.682116	
CITY_SIZE3	-0.0139358	0.0518676	-0.269	0.788207	
CITY_SIZE4	-0.0016978	0.0530670	-0.032	0.974480	
CITY_SIZE5	-0.0814907	0.0410722	-1.984	0.047403	*
CITY_SIZE6	-0.0483349	0.0262042	-1.845	0.065273	.
CITY_SIZE7	0.1072830	0.0483078	2.221	0.026491	*
CITY_SIZE8	0.0328880	0.0846899	0.388	0.697816	
CMPR_DSB_GRADE1	0.4819307	0.2361308	2.041	0.041406	*
CMPR_DSB_GRADE2	0.4153001	0.2349716	1.767	0.077328	.
CMPR_DSB_GRADE3	0.2519223	0.2348312	1.073	0.283518	
CMPR_DSB_GRADE4	0.2474675	0.2315355	1.069	0.285303	
CMPR_DSB_GRADE5	0.1934901	0.2331530	0.830	0.406718	
CMPR_DSB_GRADE6	0.1705678	0.2279697	0.748	0.454438	
CMPR_DSB_GRADE10	0.1994725	0.2605356	0.766	0.444003	
CMPR_DSB_GRADE20	0.1176343	0.2498062	0.471	0.637769	
MAIN_DSB_TYPE1	-0.0403745	0.2644705	-0.153	0.878683	
MAIN_DSB_TYPE2	-0.2065638	0.2649183	-0.780	0.435658	
MAIN_DSB_TYPE3	-0.2316257	0.2620721	-0.884	0.376913	
MAIN_DSB_TYPE4	-0.0162854	0.2756459	-0.059	0.952894	
MAIN_DSB_TYPE5	-0.2927050	0.2862167	-1.023	0.306606	
MAIN_DSB_TYPE6	-0.8042236	0.3541857	-2.271	0.023291	*
MAIN_DSB_TYPE8	-0.0830973	0.2601620	-0.319	0.749457	
MAIN_DSB_TYPE9	-0.3174050	0.2671728	-1.188	0.234990	
MAIN_DSB_TYPE10	0.0239487	0.2934101	0.082	0.934957	
MAIN_DSB_TYPE11	0.5973335	0.3393503	1.760	0.078545	.
MAIN_DSB_TYPE12	-0.0663710	0.4446717	-0.149	0.881367	
MAIN_DSB_TYPE14	NA	NA	NA	NA	
CMPR_DSB_GRADE_L1	-0.6766041	0.0821004	-8.241	3.31e-16	***
CMPR_DSB_GRADE_L2	-0.4844421	0.0752091	-6.441	1.53e-10	***

CMPR_DSB_GRADE_L3	-0.0724680	0.0947003	-0.765	0.444235
CMPR_DSB_GRADE_L4	-0.0580720	0.0946230	-0.614	0.539481
CMPR_DSB_GRADE_L5	0.0160859	0.0921570	0.175	0.861455
CMPR_DSB_GRADE_L6	NA	NA	NA	NA
MAIN_DSB_TYPE_L1	0.1541684	0.3432588	0.449	0.653393
MAIN_DSB_TYPE_L2	0.3282562	0.3428280	0.957	0.338450
MAIN_DSB_TYPE_L3	0.3082825	0.3379202	0.912	0.361740
MAIN_DSB_TYPE_L4	0.0529488	0.3510902	0.151	0.880141
MAIN_DSB_TYPE_L5	0.3552985	0.3546925	1.002	0.316623
MAIN_DSB_TYPE_L6	0.5888333	0.3759972	1.566	0.117517
MAIN_DSB_TYPE_L8	NA	NA	NA	NA
MAIN_DSB_TYPE_L9	0.4973989	0.3394960	1.465	0.143071
MAIN_DSB_TYPE_L10	0.1832579	0.3967737	0.462	0.644232
MAIN_DSB_TYPE_L11	NA	NA	NA	NA
MAIN_DSB_TYPE_L14	NA	NA	NA	NA
L_tube_feeding1	0.0756416	0.0205017	3.690	0.000232 ***
pneumonia1	0.0510431	0.0187599	2.721	0.006576 **
ICH1	-0.0848400	0.0218715	-3.879	0.000109 ***
IVH1	0.0585970	0.0345453	1.696	0.090019 .
Focal_small1	-0.0443700	0.0201509	-2.202	0.027804 *
multifocal_large1	0.0586027	0.0191864	3.054	0.002289 **
Temporal1	0.0748856	0.0303653	2.466	0.013753 *
BG1	0.0282695	0.0189267	1.494	0.135453
Thalamus1	0.0537600	0.0232094	2.316	0.020657 *
LTC_OBJTR_TYPE_CD	-0.0791161	0.0267261	-2.960	0.003115 **
LTC_REDU_TYPEB	0.1559483	0.1167736	1.335	0.181895
LTC_REDU_TYPEC	0.2434287	0.0595411	4.088	4.54e-05 ***
LTC_REDU_TYPED	-0.1824759	0.0515289	-3.541	0.000409 ***
LTC_REDU_TYPEE	-0.3049841	0.1874312	-1.627	0.103880
LTC_REDU_TYPEF	-0.2037565	0.0924446	-2.204	0.027648 *
LTC_REDU_TYPEz	0.2290171	0.0834190	2.745	0.006106 **
RCG_RCGT_SCORE	0.0050641	0.0011486	4.409	1.10e-05 ***
GJU_RCGT_GRADE_CD2	0.1219250	0.0553458	2.203	0.027727 *
GJU_RCGT_GRADE_CD3	0.1203832	0.0644303	1.868	0.061870 .
GJU_RCGT_GRADE_CD4	0.1025401	0.0763192	1.344	0.179263

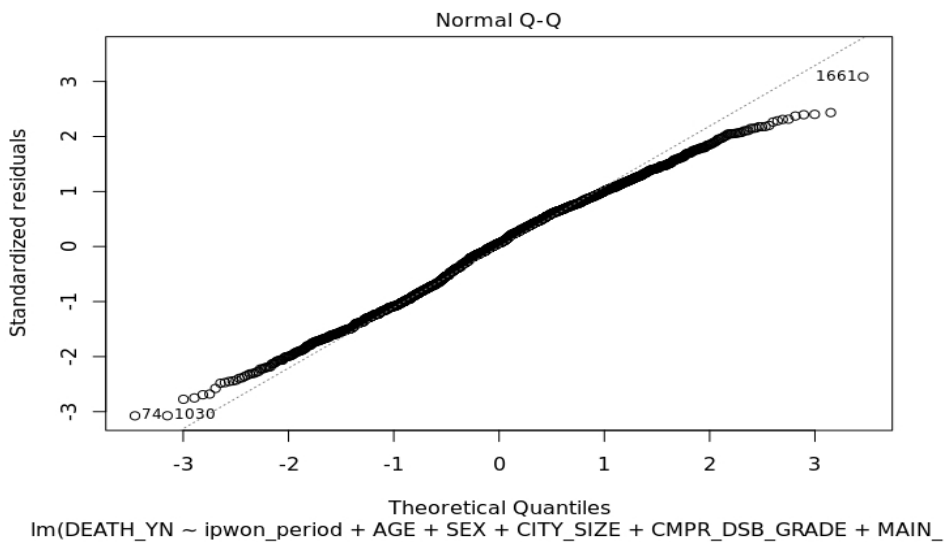
GJU_RCGT_GRADE_CD5	0.0191814	0.0900579	0.213	0.831360							
GJU_RCGT_GRADE_CD6	-0.0012552	0.1189164	-0.011	0.991579							
GJU_RCGT_GRADE_CDA	0.1697843	0.0869912	1.952	0.051129	.						
GJU_RCGT_GRADE_CDB	0.0684803	0.0956594	0.716	0.474164							
GJU_RCGT_GRADE_CDC	0.1994618	0.1255836	1.588	0.112405							
GJU_RCGT_GRADE_CDY	0.1050517	0.0680407	1.544	0.122781							
GJU_RCGT_GRADE_CDZ	0.2076688	0.0817281	2.541	0.011140	*						
CAREPSN_CD1	0.0215898	0.0406662	0.531	0.595554							
CAREPSN_CD2	-0.2415346	0.1738544	-1.389	0.164921							
CAREPSN_CD3	0.1888185	0.0425434	4.438	9.64e-06	***						
CAREPSN_CD4	-0.0183136	0.1275957	-0.144	0.885889							
CAREPSN_CD5	0.0279742	0.0837894	0.334	0.738523							
CAREPSN_CD6	0.0614804	0.0753039	0.816	0.414365							
CAREPSN_CD7	0.0396334	0.0431890	0.918	0.358917							
CAREPSN_CD8	0.0963470	0.1642932	0.586	0.557661							
CAREPSN_CD9	0.1156157	0.0558392	2.071	0.038551	*						
CAREPSN_CD10	-0.1282243	0.0526640	-2.435	0.015001	*						
CAREPSN_CD11	-0.0707369	0.0414980	-1.705	0.088449	.						
CAREPSN_CD12	-0.1328220	0.0648786	-2.047	0.040784	*						
CAREPSN_CD13	-0.1034857	0.2546306	-0.406	0.684488							
CAREPSN_CD14	-0.1454283	0.1118389	-1.300	0.193657							
BHVR_ARFF_SCORE	0.0013529	0.0008264	1.637	0.101792							
S102	0.0968358	0.0269063	3.599	0.000328	***						
S103	0.1057237	0.0346817	3.048	0.002335	**						
PE22	-0.0315448	0.0212441	-1.485	0.137758							
PE23	-0.0457754	0.0332040	-1.379	0.168192							
PE24	-0.1950520	0.0591424	-3.298	0.000993	***						
ALBUMIN	-0.0980809	0.0190420	-5.151	2.89e-07	***						
CR	0.0140348	0.0095721	1.466	0.142769							
GLUCOSE_AC	0.0004014	0.0001844	2.177	0.029631	*						
HB	-0.0083785	0.0049885	-1.680	0.093223	.						
HBA1C	0.0152096	0.0070595	2.154	0.031338	*						
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

Residual standard error: 0.3528 on 1744 degrees of freedom  
Multiple R-squared: 0.5268, Adjusted R-squared: 0.5013  
F-statistic: 20.66 on 94 and 1744 DF, p-value: < 2.2e-16  
plot(result\_2)

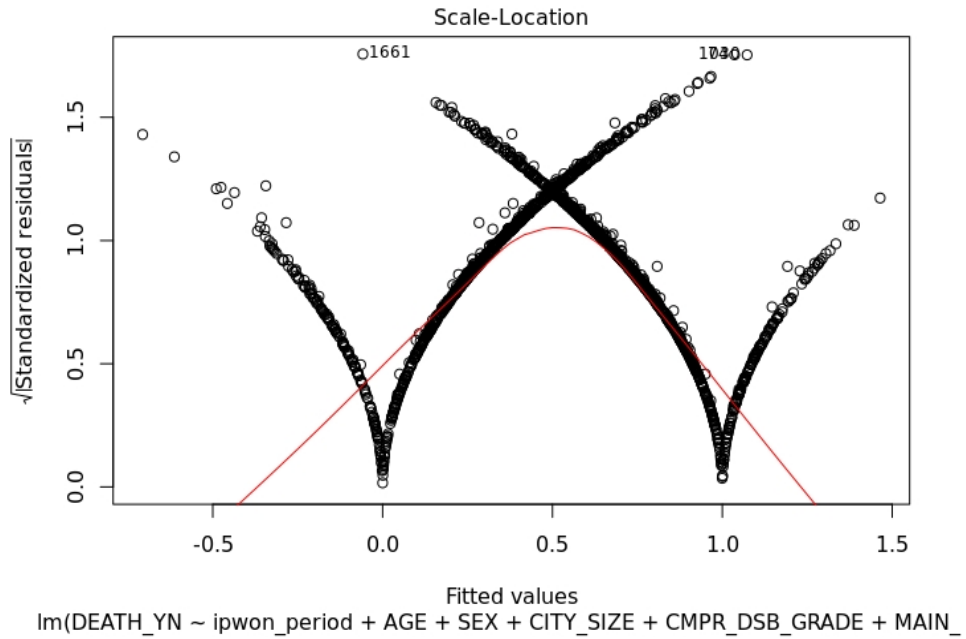
Hit <Return> to see next plot:



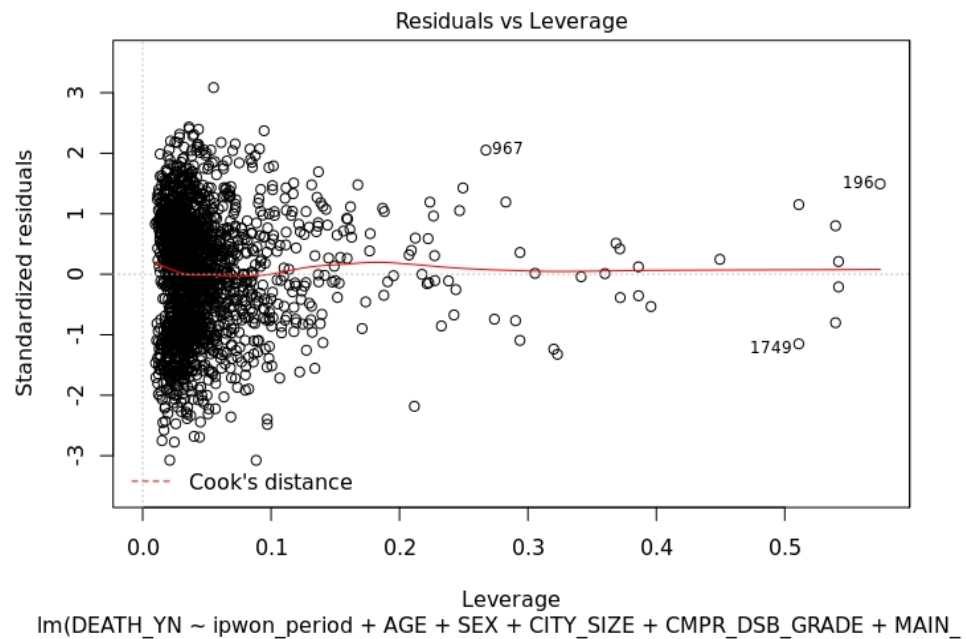
Hit <Return> to see next plot:



Hit <Return> to see next plot:



Hit <Return> to see next plot:



```
anova(result, result_2)
```

```
Analysis of Variance Table
```

```
Model 1: DEATH_YN ~ ipwon_period + AGE + SEX + SIDO_NM +  
CITY_SIZE + GAIBJA_TYPE +  
  CMPR_DSB_GRADE + MAIN_DSB_TYPE + CMPR_DSB_GRADE_L +  
MAIN_DSB_TYPE_L +  
  L_tube_feeding + pneumonia + Infarction + SAH + SDH + ICH +  
  IVH + Focal_small + multifocal_large + Acute_recent +  
Chronic_old +  
  Temporal + Occipital + Cerebellum + BG + brain_stem + Thalamus +  
  ACA + MCA + PCA + Aneurysm + APFRM_TYPE_CD + LTC_OBJTR_TYPE_CD +  
  LTC_REDU_TYPE + RCG_RCGT_SCORE + GJU_RCGT_GRADE_CD +  
DISM_JUDG_RSN_CD +  
  RSDN_TP_CD + LVTGT_PRSN_CD + CAREPSN_CD + PCT_ARFF_SCORE +  
  BHVR_ARFF_SCORE + TNDNC_ARFF_SCORE + REHBT_ARFF_SCORE + N05 +  
  N07 + S10 + PE1 + PE2 + smoke_yn + drink_yn + ALBUMIN + BUN +  
  CR + GLUCOSE_AC + HB + HBA1C + jaewon_GW + total_cowork +  
  pum_cowork + idm_cowork
```

```
Model 2: DEATH_YN ~ ipwon_period + AGE + SEX + CITY_SIZE +  
CMPR_DSB_GRADE +  
  MAIN_DSB_TYPE + CMPR_DSB_GRADE_L + MAIN_DSB_TYPE_L +  
L_tube_feeding +  
  pneumonia + ICH + IVH + Focal_small + multifocal_large +  
  Temporal + BG + Thalamus + LTC_OBJTR_TYPE_CD + LTC_REDU_TYPE +  
  RCG_RCGT_SCORE + GJU_RCGT_GRADE_CD + CAREPSN_CD +  
BHVR_ARFF_SCORE +  
  S10 + PE2 + ALBUMIN + CR + GLUCOSE_AC + HB + HBA1C
```

```
Res.Df    RSS  Df Sum of Sq    F Pr(>F)  
1    1651 204.79  
2    1744 217.07 -93   -12.278 1.0644 0.3219
```

```
#2. The Prognosis of Patients without Long Term Care Service  
Machine Learning Analysis
```

```
path="~/data/data_source/user_data/"
```



```

spath=paste0(path, "handled_data/")
# data_fusion (자격변수)

data=read.csv(paste0(spath,"final_data_admission_LTC_N.csv"))

colnames(data)

# [1] "X"                "INDI_DSCM_NO_R"  "MDCARE_STRT_DT"
"DTH_ASSMD_DT"    "ipwon_period"    "Live_days"       "AGE"
# [8] "SEX"              "RVSN_ADDR_CD"   "SIDO_NM"
"CITY_SIZE"       "GAIBJA_TYPE"    "CMPR_DSB_GRADE"
"MAIN_DSB_TYPE"
# [15] "CMPR_DSB_GRADE_L" "MAIN_DSB_TYPE_L" "DEATH_YN"
"gastrostomy"     "L_tube_feeding"  "pneumonia"       "T_tube"
# [22] "TBI"              "non_trauma"      "Infarction"
"SAH"              "SDH"              "ICH"              "IVH"
# [29] "Focal_small"      "multifocal_large" "Acute_recent"
"Chronic_old"     "Frontal"          "Temporal"         "Parietal"
# [36] "Occipital"        "Cerebellum"      "BG"
"subcortex"       "brain_stem"      "Thalamus"         "ACA"
# [43] "MCA"              "PCA"              "basilar_artery"
"Aneurysm"        "Coil"             "craniectomy"
"craniotomy"
# [50] "X.1"              "smoke_yn"        "drink_yn"
"ALBUMIN"         "BUN"              "CR"
"GLUCOSE.AC."
# [57] "HB"              "HBA1C"           "jaewon_GW"
"jaewon_ICU"      "total_cowork"    "pum_cowork"
"idm_cowork"
# [64] "SORE

Data_final=data[,c(5,7,8,10:49,51:64)]

colnames(Data_final)

#3. Data_format_assignment

```

```

Data_final$ipwon_period=as.numeric(Data_final$ipwon_period)
Data_final$AGE=as.numeric(Data_final$AGE)
Data_final$SEX=as.factor(Data_final$SEX)
Data_final$SIDO_NM=as.factor(Data_final$SIDO_NM)
Data_final$CITY_SIZE=as.factor(Data_final$CITY_SIZE)
Data_final$GAIBJA_TYPE=as.factor(Data_final$GAIBJA_TYPE)
Data_final$CMPR_DSB_GRADE=as.factor(Data_final$CMPR_DSB_GRADE)
Data_final$MAIN_DSB_TYPE=as.factor(Data_final$MAIN_DSB_TYPE)
Data_final$CMPR_DSB_GRADE_L=as.factor(Data_final$CMPR_DSB_GRADE_L)
Data_final$MAIN_DSB_TYPE_L=as.factor(Data_final$MAIN_DSB_TYPE_L)
Data_final$DEATH_YN=as.factor(Data_final$DEATH_YN)
Data_final$gastrostomy=as.factor(Data_final$gastrostomy)
Data_final$L_tube_feeding=as.factor(Data_final$L_tube_feeding)
Data_final$pneumonia=as.factor(Data_final$pneumonia)
Data_final$T_tube=as.factor(Data_final$T_tube)
Data_final$TBI=as.factor(Data_final$TBI)
Data_final$non_trauma=as.factor(Data_final$non_trauma)
Data_final$Infarction=as.factor(Data_final$Infarction)
Data_final$SAH=as.factor(Data_final$SAH)
Data_final$SDH=as.factor(Data_final$SDH)
Data_final$ICH=as.factor(Data_final$ICH)
Data_final$IVH=as.factor(Data_final$IVH)
Data_final$Focal_small=as.factor(Data_final$Focal_small)
Data_final$multifocal_large=as.factor(Data_final$multifocal_large)
Data_final$Acute_recent=as.factor(Data_final$Acute_recent)
Data_final$Chronic_old=as.factor(Data_final$Chronic_old)
Data_final$Frontal=as.factor(Data_final$Frontal)
Data_final$Temporal=as.factor(Data_final$Temporal)
Data_final$Parietal=as.factor(Data_final$Parietal)
Data_final$Occipital=as.factor(Data_final$Occipital)
Data_final$Cerebellum=as.factor(Data_final$Cerebellum)
Data_final$BG=as.factor(Data_final$BG)
Data_final$subcortex=as.factor(Data_final$subcortex)
Data_final$brain_stem=as.factor(Data_final$brain_stem)
Data_final$Thalamus=as.factor(Data_final$Thalamus)

```

```

Data_final$ACA=as.factor(Data_final$ACA)
Data_final$MCA=as.factor(Data_final$MCA)
Data_final$PCA=as.factor(Data_final$PCA)
Data_final$basilar_artery=as.factor(Data_final$basilar_artery)
Data_final$Aneurysm=as.factor(Data_final$Aneurysm)
Data_final$Coil=as.factor(Data_final$Coil)
Data_final$craniectomy=as.factor(Data_final$craniectomy)
Data_final$craniotomy=as.factor(Data_final$craniotomy)
Data_final$smoke_yn=as.factor(Data_final$smoke_yn)
Data_final$drink_yn=as.factor(Data_final$drink_yn)
Data_final$ALBUMIN=as.numeric(Data_final$ALBUMIN)
Data_final$BUN=as.numeric(Data_final$BUN)
Data_final$CR=as.numeric(Data_final$CR)
Data_final$GLUCOSE_AC=as.numeric(Data_final$GLUCOSE_AC)
Data_final$HB=as.numeric(Data_final$HB)
Data_final$HBA1C=as.numeric(Data_final$HBA1C)
Data_final$jaewon_GW=as.numeric(Data_final$jaewon_GW)
Data_final$jaewon_ICU=as.numeric(Data_final$jaewon_ICU)
Data_final$total_cowork=as.numeric(Data_final$total_cowork)
Data_final$pum_cowork=as.numeric(Data_final$pum_cowork)
Data_final$idm_cowork=as.numeric(Data_final$idm_cowork)
Data_final$SORE=as.factor(Data_final$SORE)

nearZeroVar(Data_final, saveMetrics = TRUE)

# freqRatio percentUnique zeroVar   nzv
# ipwon_period      1.227129    4.33898305  FALSE FALSE
# AGE                1.021505    2.64406780  FALSE FALSE
# SEX                1.594547    0.06779661  FALSE FALSE
# SIDO_NM           6.133159    0.57627119  FALSE FALSE
# CITY_SIZE         1.644377    0.27118644  FALSE FALSE
# GAIBJA_TYPE       1.208696    0.71186441  FALSE FALSE
# CMPR_DSB_GRADE   14.783439    0.23728814  FALSE FALSE
# MAIN_DSB_TYPE    11.377451    0.47457627  FALSE FALSE
# CMPR_DSB_GRADE_L  7.185759    0.23728814  FALSE FALSE

```

# MAIN_DSB_TYPE_L	12.088542	0.47457627	FALSE	FALSE
# DEATH_YN	3.244604	0.06779661	FALSE	FALSE
# gastrostomy	183.375000	0.06779661	FALSE	TRUE
# L_tube_feeding	6.283951	0.06779661	FALSE	FALSE
# pneumonia	1.508503	0.06779661	FALSE	FALSE
# T_tube	55.730769	0.06779661	FALSE	TRUE
# TBI	27.921569	0.06779661	FALSE	TRUE
# non_trauma	294.000000	0.06779661	FALSE	TRUE
# Infarction	1.139231	0.06779661	FALSE	FALSE
# SAH	7.753709	0.06779661	FALSE	FALSE
# SDH	11.995595	0.06779661	FALSE	FALSE
# ICH	3.631083	0.06779661	FALSE	FALSE
# IVH	16.151163	0.06779661	FALSE	FALSE
# Focal_small	1.636282	0.06779661	FALSE	FALSE
# multifocal_large	3.788961	0.06779661	FALSE	FALSE
# Acute_recent	1.608311	0.06779661	FALSE	FALSE
# Chronic_old	1.031680	0.06779661	FALSE	FALSE
# Frontal	0.000000	0.03389831	TRUE	TRUE
# Temporal	14.128205	0.06779661	FALSE	FALSE
# Parietal	420.428571	0.06779661	FALSE	TRUE
# Occipital	20.376812	0.06779661	FALSE	TRUE
# Cerebellum	7.753709	0.06779661	FALSE	FALSE
# BG	2.238200	0.06779661	FALSE	FALSE
# subcortex	420.428571	0.06779661	FALSE	TRUE
# brain_stem	12.785047	0.06779661	FALSE	FALSE
# Thalamus	9.766423	0.06779661	FALSE	FALSE
# ACA	26.570093	0.06779661	FALSE	TRUE
# MCA	3.396423	0.06779661	FALSE	FALSE
# PCA	22.600000	0.06779661	FALSE	TRUE
# basilar_artery	38.333333	0.06779661	FALSE	TRUE
# Aneurysm	3.338235	0.06779661	FALSE	FALSE
# Coil	17.437500	0.06779661	FALSE	FALSE
# craniectomy	74.641026	0.06779661	FALSE	TRUE
# craniotomy	12.594470	0.06779661	FALSE	FALSE
# smoke_yn	1.191794	0.10169492	FALSE	FALSE

```

# drink_yn      1.711846    0.10169492  FALSE FALSE
# ALBUMIN      1.077181     1.38983051  FALSE FALSE
# BUN          1.026316    13.22033898  FALSE FALSE
# CR           1.016129     9.38983051  FALSE FALSE
# GLUCOSE_AC   2.295082     8.74576271  FALSE FALSE
# HB           1.000000     4.50847458  FALSE FALSE
# HBA1C        4.542373     3.25423729  FALSE FALSE
# jaewon_GW    1.290735     2.88135593  FALSE FALSE
# jaewon_ICU   24.819277    3.22033898  FALSE TRUE
# total_cowork 1.097990     2.13559322  FALSE FALSE
# pum_cowork   14.870056    0.40677966  FALSE FALSE
# idm_cowork   22.897436    0.47457627  FALSE TRUE
# SORE         46.580645    0.06779661  FALSE TRUE
predata=Data_final[,-nearZeroVar(Data_final)] # near Zero
Variance를 제거

colnames(predata)

# [1] "ipwon_period"      "AGE"           "SEX"
"SIDO_NM"           "CITY_SIZE"     "GAIBJA_TYPE"
"CMPR_DSB_GRADE"
# [8] "MAIN_DSB_TYPE"    "CMPR_DSB_GRADE_L" "MAIN_DSB_TYPE_L"
"DEATH_YN"          "L_tube_feeding" "pneumonia"
"Infarction"
# [15] "SAH"              "SDH"           "ICH"           "IVH"
"Focal_small"      "multifocal_large" "Acute_recent"
# [22] "Chronic_old"      "Temporal"      "Cerebellum"
"BG"               "brain_stem"    "Thalamus"      "MCA"
# [29] "Aneurysm"         "Coil"          "craniotomy"
"smoke_yn"         "drink_yn"      "ALBUMIN"       "BUN"
# [36] "CR"              "GLUCOSE_AC"    "HB"
"HBA1C"            "jaewon_GW"     "total_cowork"
"pum_cowork"

findCorrelation(cor(predata[,c(1:2,34:42)]),cutoff=.90,

```

```

names=FALSE)

# integer(0)

st_model=preProcess(predata,method=c("center","scale")) #
Live_days 제거
data=predict(st_model, predata)

data=as.data.frame(data)
data=na.omit(data)

ohe_feats=c('ipwon_period','AGE','SEX','SIDO_NM','CITY_SIZE','GAI
BJA_TYPE','CMPR_DSB_GRADE','MAIN_DSB_TYPE','CMPR_DSB_GRADE_L','MA
IN_DSB_TYPE_L','L_tube_feeding','pneumonia','Infarction','SAH','S
DH','ICH','IVH',
'Focal_small','multifocal_large','Acute_recent','Chronic_old','Te
mporal','Cerebellum','BG','brain_stem','Thalamus','MCA','Aneurysm
','Coil','craniotomy','smoke_yn','drink_yn',
'ALBUMIN','BUN','CR','GLUCOSE_AC','HB','HBA1C','jaewon_GW','total
_cowork','pum_cowork')

dummies =
dummyVars(~ipwon_period+AGE+SEX+SIDO_NM+CITY_SIZE+GAIBJA_TYPE+CMP
R_DSB_GRADE+MAIN_DSB_TYPE+CMPR_DSB_GRADE_L+MAIN_DSB_TYPE_L+L_tube
_feeding+pneumonia+Infarction+SAH+SDH+ICH+IVH+Focal_small+multifo
cal_large+Acute_recent+Chronic_old
+Temporal+Cerebellum+BG+brain_stem+Thalamus+MCA+Aneurysm+Coil+cra
niotomy+smoke_yn+drink_yn
+ALBUMIN+BUN+CR+GLUCOSE_AC+HB+HBA1C+jaewon_GW+total_cowork+pum_co
work, data=data)

df_ohe <- as.data.frame(predict(dummies, newdata = data))
colnames(df_ohe)
df_combined <- cbind(data[c("DEATH_YN")],df_ohe) # Live_days제거
Data=as.data.table(df_combined)

```

```

colnames(Data)

# [1] "ipwon_period"      "AGE"              "SEX.1"
"SEX.2"              "SIDO_NM.1"        "SIDO_NM.2"
# [7] "SIDO_NM.3"         "SIDO_NM.4"        "SIDO_NM.5"
"SIDO_NM.6"         "SIDO_NM.7"        "SIDO_NM.8"
# [13] "SIDO_NM.9"         "SIDO_NM.10"       "SIDO_NM.11"
"SIDO_NM.12"        "SIDO_NM.13"       "SIDO_NM.14"
# [19] "SIDO_NM.15"        "SIDO_NM.16"       "SIDO_NM.17"
"CITY_SIZE.1"       "CITY_SIZE.2"      "CITY_SIZE.3"
# [25] "CITY_SIZE.4"       "CITY_SIZE.5"      "CITY_SIZE.6"
"CITY_SIZE.7"       "CITY_SIZE.8"      "GAIBJA_TYPE.0"
# [31] "GAIBJA_TYPE.1"    "GAIBJA_TYPE.2"    "GAIBJA_TYPE.3"
"GAIBJA_TYPE.4"    "GAIBJA_TYPE.5"    "GAIBJA_TYPE.6"
# [37] "GAIBJA_TYPE.7"    "GAIBJA_TYPE.8"    "GAIBJA_TYPE.9"
"GAIBJA_TYPE.10"   "GAIBJA_TYPE.11"   "GAIBJA_TYPE.12"
# [43] "GAIBJA_TYPE.13"   "GAIBJA_TYPE.14"   "GAIBJA_TYPE.15"
"GAIBJA_TYPE.16"   "GAIBJA_TYPE.17"   "GAIBJA_TYPE.18"
# [49] "GAIBJA_TYPE.19"   "GAIBJA_TYPE.20"
"CMPR_DSB_GRADE.0" "CMPR_DSB_GRADE.1" "CMPR_DSB_GRADE.2"
"CMPR_DSB_GRADE.3"
# [55] "CMPR_DSB_GRADE.4" "CMPR_DSB_GRADE.5"
"CMPR_DSB_GRADE.6" "MAIN_DSB_TYPE.0"  "MAIN_DSB_TYPE.1"
"MAIN_DSB_TYPE.2"
# [61] "MAIN_DSB_TYPE.3"  "MAIN_DSB_TYPE.4"  "MAIN_DSB_TYPE.5"
"MAIN_DSB_TYPE.6"  "MAIN_DSB_TYPE.8"  "MAIN_DSB_TYPE.9"
# [67] "MAIN_DSB_TYPE.10" "MAIN_DSB_TYPE.11"
"MAIN_DSB_TYPE.12" "MAIN_DSB_TYPE.14" "MAIN_DSB_TYPE.15"
"CMPR_DSB_GRADE_L.0"
# [73] "CMPR_DSB_GRADE_L.1" "CMPR_DSB_GRADE_L.2"
"CMPR_DSB_GRADE_L.3" "CMPR_DSB_GRADE_L.4" "CMPR_DSB_GRADE_L.5"
"CMPR_DSB_GRADE_L.6"
# [79] "MAIN_DSB_TYPE_L.0" "MAIN_DSB_TYPE_L.1"
"MAIN_DSB_TYPE_L.2" "MAIN_DSB_TYPE_L.3" "MAIN_DSB_TYPE_L.4"
"MAIN_DSB_TYPE_L.5"

```

```

# [85] "MAIN_DSB_TYPE_L.6" "MAIN_DSB_TYPE_L.8"
"MAIN_DSB_TYPE_L.9" "MAIN_DSB_TYPE_L.10" "MAIN_DSB_TYPE_L.11"
"MAIN_DSB_TYPE_L.12"
# [91] "MAIN_DSB_TYPE_L.14" "MAIN_DSB_TYPE_L.15"
"L_tube_feeding.0" "L_tube_feeding.1" "pneumonia.0"
"pneumonia.1"
# [97] "Infarction.0" "Infarction.1" "SAH.0"
"SAH.1" "SDH.0" "SDH.1"
# [103] "ICH.0" "ICH.1" "IVH.0"
"IVH.1" "Focal_small.0" "Focal_small.1"
# [109] "multifocal_large.0" "multifocal_large.1"
"Acute_recent.0" "Acute_recent.1" "Chronic_old.0"
"Chronic_old.1"
# [115] "Temporal.0" "Temporal.1" "Cerebellum.0"
"Cerebellum.1" "BG.0" "BG.1"
# [121] "brain_stem.0" "brain_stem.1" "Thalamus.0"
"Thalamus.1" "MCA.0" "MCA.1"
# [127] "Aneurysm.0" "Aneurysm.1" "Coil.0"
"Coil.1" "craniotomy.0" "craniotomy.1"
# [133] "smoke_yn.1" "smoke_yn.2" "smoke_yn.3"
"drink_yn.1" "drink_yn.2" "drink_yn.3"
# [139] "ALBUMIN" "BUN" "CR"
"GLUCOSE_AC" "HB" "HBA1C"
# [145] "jaewon_GW" "total_cowork" "pum_cowork"

# Data splitting

seed=0923
set.seed(seed)

ind=sample(2, nrow(Data), replace=T, prob=c(0.7,0.3))
trainmc<-Data[ind==1,]
testmc<-Data[ind==2,]

#Multi-class classification

```



```

ctrl <- trainControl(method="repeatedcv", number=10, repeats=5)
metric <- "Accuracy"

nn_fit <- train(DEATH_YN~., data = trainmc, method = "nnet",
trControl=ctrl, metric=metric, tuneLength = 5,
verbose=FALSE, na.action = na.omit)

nn_fit

# Neural Network(NN)
#
# 2081 samples
# 147 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1873, 1873, 1873, 1872, 1873, 1873, ...
# Resampling results across tuning parameters:

#  size decay Accuracy  Kappa
# 1    0e+00 0.8379610 0.5368343
# 1    1e-04 0.8384353 0.5341560
# 1    1e-03 0.8383350 0.5547432
# 1    1e-02 0.8493969 0.5749626
# 1    1e-01 0.8667902 0.6012551
# 3    0e+00 0.8385379 0.5184132
# 3    1e-04 0.8431561 0.5305945
# 3    1e-03 0.8494949 0.5587845
# 3    1e-02 0.8450782 0.5458841
# 3    1e-01 0.8536259 0.5634418
# 5    0e+00 0.8308503 0.4795180
# 5    1e-04 0.8480489 0.5532172
# 5    1e-03 0.8432393 0.5421152
# 5    1e-02 0.8449844 0.5518638

```

```

# 5    1e-01  0.8417055  0.5387075
# 7    0e+00      NaN      NaN
# 7    1e-04      NaN      NaN
# 7    1e-03      NaN      NaN
# 7    1e-02      NaN      NaN
# 7    1e-01      NaN      NaN
# 9    0e+00      NaN      NaN
# 9    1e-04      NaN      NaN
# 9    1e-03      NaN      NaN
# 9    1e-02      NaN      NaN
# 9    1e-01      NaN      NaN

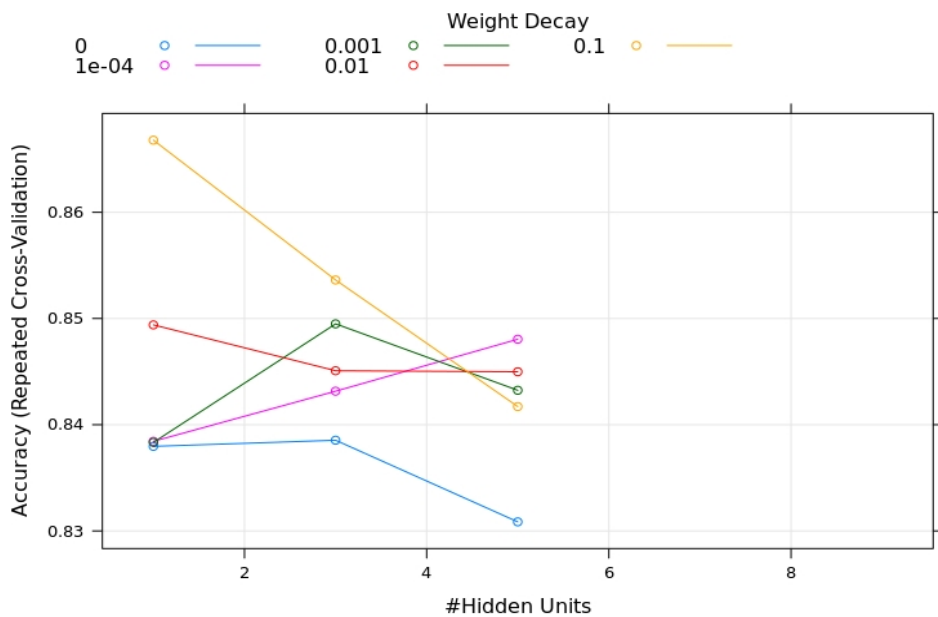
```

```
#
```

```
# Accuracy was used to select the optimal model using the largest value.
```

```
# The final values used for the model were size = 1 and decay = 0.1.
```

```
plot(nn_fit)
```



```
test_pred_nn <- predict(nn_fit, newdata = testmc)
```

```

confusionMatrix (test_pred_nn, testmc$DEATH_YN)

# Confusion Matrix and Statistics
#
# Reference
# Prediction   N   Y
#           N 602  71
#           Y  41 155
#
# Accuracy : 0.8711
# 95% CI : (0.847, 0.8927)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.6501
# Mcnemar's Test P-Value : 0.006139
#
#           Sensitivity : 0.9362
#           Specificity : 0.6858
#           Pos Pred Value : 0.8945
#           Neg Pred Value : 0.7908
#           Prevalence : 0.7399
#           Detection Rate : 0.6928
#           Detection Prevalence : 0.7745
#           Balanced Accuracy : 0.8110
#
#           'Positive' Class : N

imp<-varImp(nn_fit, scale =FALSE)
imp

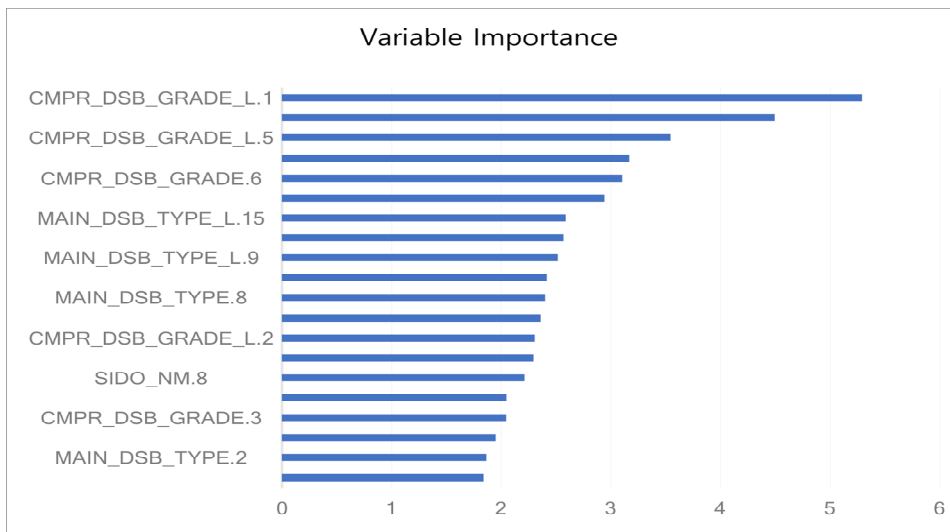
# nnet variable importance
#
# only 20 most important variables shown (out of 147)
#

```

```

# Overall
# CMPR_DSB_GRADE_L.1  5.292
# MAIN_DSB_TYPE_L.5   4.497
# CMPR_DSB_GRADE_L.5  3.545
# MAIN_DSB_TYPE.15    3.168
# CMPR_DSB_GRADE.6    3.104
# CMPR_DSB_GRADE.1    2.943
# MAIN_DSB_TYPE_L.15  2.588
# MAIN_DSB_TYPE_L.1   2.569
# MAIN_DSB_TYPE_L.9   2.518
# MAIN_DSB_TYPE_L.8   2.417
# MAIN_DSB_TYPE.8     2.402
# CMPR_DSB_GRADE_L.6  2.360
# CMPR_DSB_GRADE_L.2  2.306
# MAIN_DSB_TYPE.12    2.296
# SIDO_NM.8           2.214
# CMPR_DSB_GRADE.5    2.047
# CMPR_DSB_GRADE.3    2.045
# MAIN_DSB_TYPE_L.2   1.951
# MAIN_DSB_TYPE.2     1.865
# MAIN_DSB_TYPE.1     1.840

```

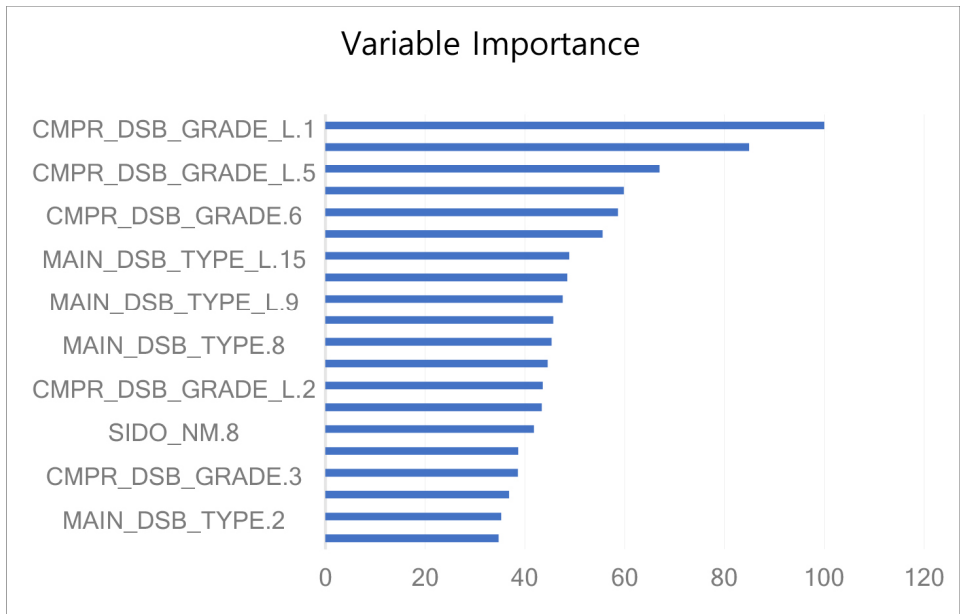


```

imp<-varImp(nn_fit, scale =TRUE)
imp

# nnet variable importance
#
# only 20 most important variables shown (out of 147)
#
# Overall
# CMPR_DSB_GRADE_L.1  100.00
# MAIN_DSB_TYPE_L.5   84.97
# CMPR_DSB_GRADE_L.5  66.99
# MAIN_DSB_TYPE.15    59.86
# CMPR_DSB_GRADE.6    58.65
# CMPR_DSB_GRADE.1    55.60
# MAIN_DSB_TYPE_L.15  48.90
# MAIN_DSB_TYPE_L.1   48.53
# MAIN_DSB_TYPE_L.9   47.58
# MAIN_DSB_TYPE_L.8   45.67
# MAIN_DSB_TYPE.8     45.37
# CMPR_DSB_GRADE_L.6  44.59
# CMPR_DSB_GRADE_L.2  43.57
# MAIN_DSB_TYPE.12    43.37
# SIDO_NM.8           41.83
# CMPR_DSB_GRADE.5    38.67
# CMPR_DSB_GRADE.3    38.63
# MAIN_DSB_TYPE_L.2   36.86
# MAIN_DSB_TYPE.2     35.24
# MAIN_DSB_TYPE.1     34.76

```



```

set.seed(seed)
svm_fit <- train(DEATH_YN ~., data = trainmc, method =
"svmRadial",trControl=ctrl, metric=metric, tuneLength =
10,importance=TRUE,proximity=TRUE, na.action = na.omit)

svm_fit

# support vector machines with Radial Basis Function Kernel
#
# 2081 samples
# 147 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1873, 1873, 1872, 1872, 1874, 1873, ...
# Resampling results across tuning parameters:
#
# C      Accuracy  Kappa

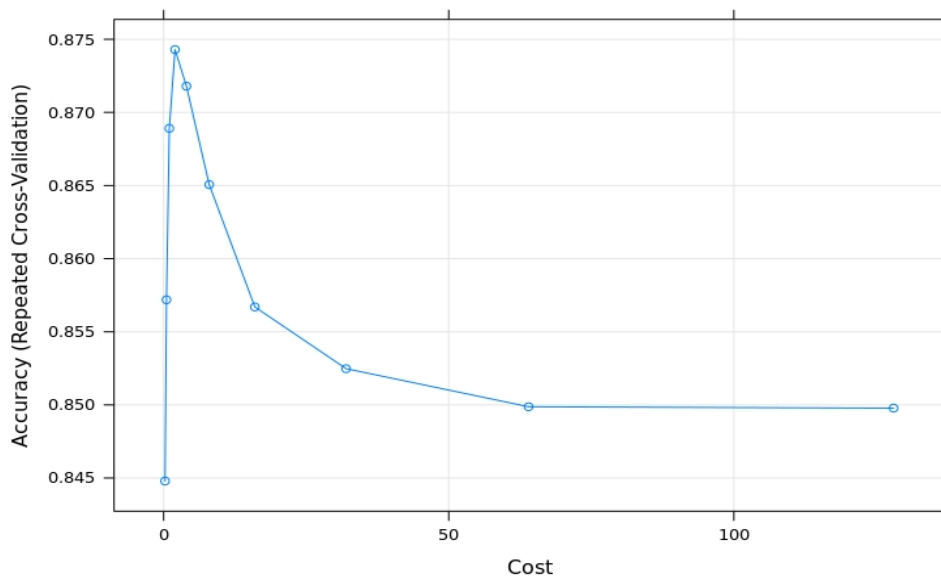
```

```

# 0.25 0.8447933 0.4718684
# 0.50 0.8571880 0.5273312
# 1.00 0.8689087 0.5746594
# 2.00 0.8742933 0.6028709
# 4.00 0.8717919 0.6013766
# 8.00 0.8650588 0.5879635
# 16.00 0.8566962 0.5690810
# 32.00 0.8524741 0.5633407
# 64.00 0.8498729 0.5574791
# 128.00 0.8497786 0.5567284
#
# Tuning parameter 'sigma' was held constant at a value of
0.03274065
# Accuracy was used to select the optimal model using the largest
value.
# The final values used for the model were sigma = 0.03274065 and
C = 2.

```

```
plot(svm_fit)
```



Support Vector Machines with Radial Basis Function Kernel

```
2081 samples
147 predictor
  2 classes: 'N', 'Y'
```

No pre-processing

Resampling: Cross-Validated (10 fold, repeated 5 times)

Summary of sample sizes: 1873, 1873, 1872, 1872, 1874, 1873, ...

Resampling results across tuning parameters:

C	Accuracy	Kappa
0.25	0.8564243	0.5456500
0.50	0.8619000	0.5657169
1.00	0.8712159	0.5990942
2.00	0.8733313	0.6080527
4.00	0.8724650	0.6035754
8.00	0.8672695	0.5856389
16.00	0.8598697	0.5588344
32.00	0.8547773	0.5441073
64.00	0.8535282	0.5398720
128.00	0.8525671	0.5368546

Tuning parameter 'sigma' was held constant at a value of 0.03274065

Accuracy was used to select the optimal model using the largest value.

The final values used for the model were sigma = 0.03274065 and C = 2.

```
> test_pred_svm<- predict(svm_fit, newdata = testmc)
```

```
> confusionMatrix (test_pred_svm, testmc$DEATH_YN)
```

Confusion Matrix and Statistics

```
# Reference
```

```
# Prediction   N   Y
```

```
# N  612  66
```

```
# Y   31 160
```

```
#
```

```
# Accuracy : 0.8884
```



```

# 95% CI : (0.8655, 0.9085)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.6946
# McNemar's Test P-Value : 0.0005561
#
#           Sensitivity : 0.9518
#           Specificity : 0.7080
#           Pos Pred Value : 0.9027
#           Neg Pred Value : 0.8377
#           Prevalence : 0.7399
#           Detection Rate : 0.7043
#           Detection Prevalence : 0.7802
#           Balanced Accuracy : 0.8299
#
#           'Positive' Class : N

knn_fit <- train(DEATH_YN~., data = trainmc, method = "knn",
trControl=ctrl, metric=metric, tuneLength=50, na.action =
na.omit)

knn_fit

# 2081 samples
# 147 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1873, 1873, 1873, 1874, 1873, 1873, ...
# Resampling results across tuning parameters:
#
# k   Accuracy   Kappa
# 5  0.8381545  0.4460898

```

# 7 0.8397956 0.4379334  
# 9 0.8367215 0.4156099  
# 11 0.8353716 0.4043210  
# 13 0.8331614 0.3963178  
# 15 0.8319105 0.3892877  
# 17 0.8333542 0.3939323  
# 19 0.8322905 0.3899462  
# 21 0.8322923 0.3916406  
# 23 0.8324874 0.3895104  
# 25 0.8281600 0.3690639  
# 27 0.8279709 0.3673524  
# 29 0.8277795 0.3659751  
# 31 0.8273963 0.3644648  
# 33 0.8265291 0.3600203  
# 35 0.8261435 0.3569290  
# 37 0.8254709 0.3524536  
# 39 0.8252768 0.3497496  
# 41 0.8247003 0.3456859  
# 43 0.8234517 0.3394299  
# 45 0.8230675 0.3377818  
# 47 0.8228734 0.3368207  
# 49 0.8229704 0.3359871  
# 51 0.8232594 0.3370053  
# 53 0.8226843 0.3340739  
# 55 0.8222021 0.3300980  
# 57 0.8226801 0.3320089  
# 59 0.8224878 0.3305308  
# 61 0.8223912 0.3287093  
# 63 0.8215267 0.3245650  
# 65 0.8217195 0.3243895  
# 67 0.8212392 0.3205457  
# 69 0.8202781 0.3151646  
# 71 0.8196069 0.3122309  
# 73 0.8197049 0.3115709  
# 75 0.8195112 0.3088503

```

# 77 0.8198001 0.3086818
# 79 0.8194146 0.3055134
# 81 0.8197040 0.3050729
# 83 0.8190318 0.3015146
# 85 0.8190313 0.3006471
# 87 0.8184553 0.2972991
# 89 0.8185510 0.2966746
# 91 0.8184540 0.2950272
# 93 0.8180698 0.2926793
# 95 0.8186458 0.2943961
# 97 0.8178780 0.2911948
# 99 0.8171083 0.2870252
# 101 0.8166280 0.2841169
# 103 0.8167232 0.2842185

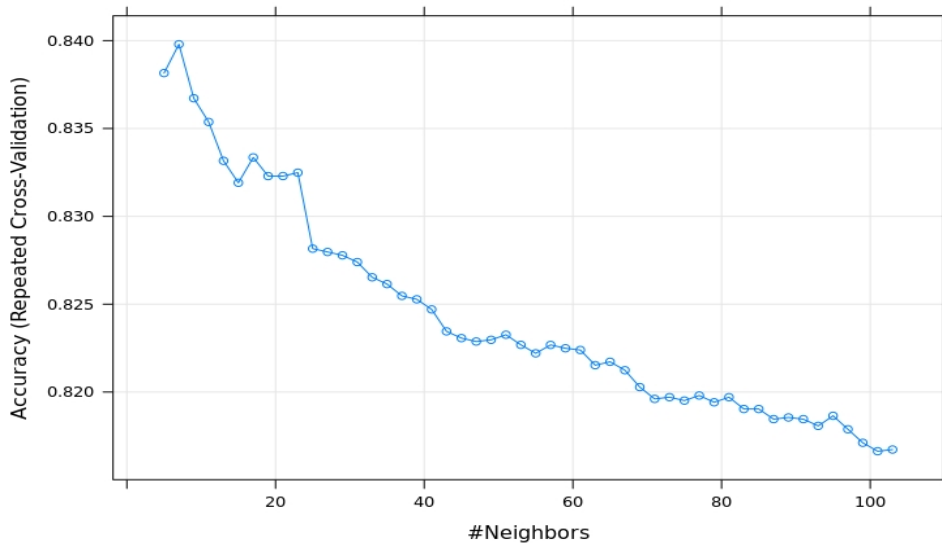
```

```
#
```

```
# Accuracy was used to select the optimal model using the largest value.
```

```
# The final value used for the model was k = 7.
```

```
plot(knn_fit)
```



```
knn_fit <- train(DEATH_YN~., data = trainmc, method = "knn",
```

```

trControl=ctrl, metric=metric, tuneLength=50, na.action =
na.omit)
> test_pred_knn<- predict(knn_fit, newdata = testmc)
> confusionMatrix (test_pred_knn, testmc$DEATH_YN)
Confusion Matrix and Statistics
# Reference
# Prediction   N   Y
# N  622 122
# Y   21 104
#
# Accuracy : 0.8354
# 95% CI   : (0.8091, 0.8595)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : 1.144e-11
#
# Kappa : 0.5
# McNemar's Test P-Value : < 2.2e-16
#
#           Sensitivity : 0.9673
#           Specificity : 0.4602
#           Pos Pred Value : 0.8360
#           Neg Pred Value : 0.8320
#           Prevalence : 0.7399
#           Detection Rate : 0.7158
#           Detection Prevalence : 0.8562
#           Balanced Accuracy : 0.7138
#
#           'Positive' Class : N

set.seed(seed)
rf_fit <- train(DEATH_YN~., data = trainmc, method = "rf",
trControl=ctrl, metric=metric,
tuneLength=15,importance=TRUE,proximity=TRUE, na.action =
na.omit)
rf_fit

```

```

# Random Forest

# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1174, 1174, 1174, 1173, 1175, 1175, ...
# Resampling results across tuning parameters:
#
#  mtry Accuracy  Kappa
#  2   0.7820538  0.5601806
# 18   0.8491887  0.6972969
# 35   0.8484136  0.6958825
# 52   0.8491980  0.6974626
# 69   0.8493530  0.6978030
# 86   0.8505721  0.7002963
# 103  0.8490254  0.6972652
# 120  0.8470360  0.6932114
# 136  0.8487353  0.6966280
# 153  0.8498052  0.6989137
# 170  0.8493402  0.6978969
# 187  0.8487353  0.6966961
# 204  0.8488798  0.6970128
# 221  0.8476654  0.6945459
# 238  0.8465873  0.6923908
#
# Accuracy was used to select the optimal model using the largest
value.
# The final value used for the model was mtry = 86.

plot(rf_fit)

test_pred_rf <- predict(rf_fit, newdata = testmc)
confusionMatrix(test_pred_rf, testmc$DEATH_YN)

# Confusion Matrix and Statistics
#

```

```

# Reference
# Prediction   N   Y
# N 613  89
# Y  30 137
#
# Accuracy : 0.8631
# 95% CI : (0.8384, 0.8852)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.6113
# McNemar's Test P-Value : 1.056e-07
#
#           Sensitivity : 0.9533
#           Specificity : 0.6062
#           Pos Pred Value : 0.8732
#           Neg Pred Value : 0.8204
#           Prevalence : 0.7399
#           Detection Rate : 0.7054
#           Detection Prevalence : 0.8078
#           Balanced Accuracy : 0.7798
#
#           'Positive' Class : N

imp<-varImp(rf_fit, scale =FALSE)

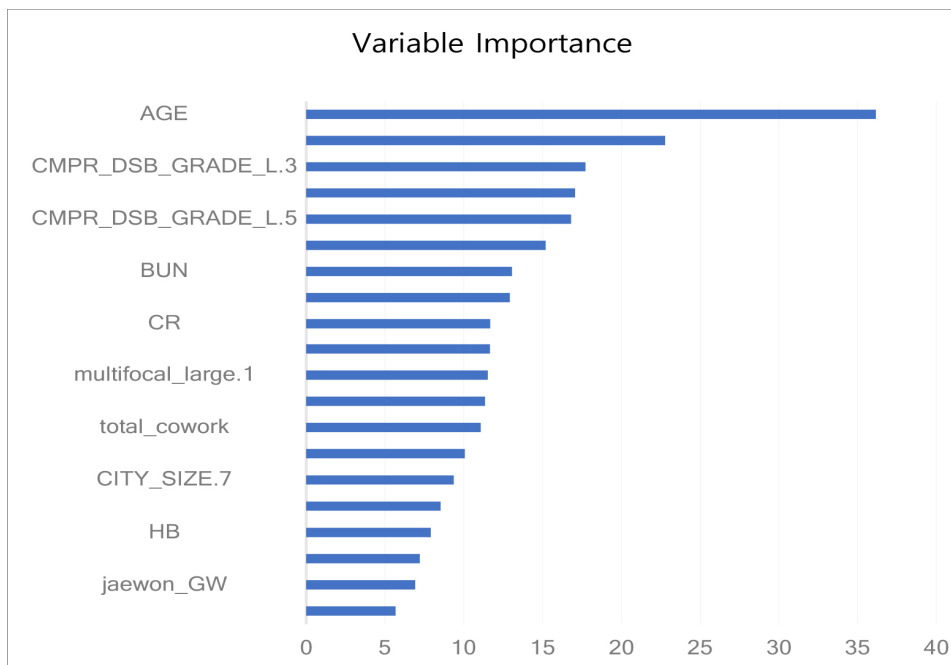
imp

#rf variable importance

# only 20 most important variables shown (out of 147)
#
# Importance
# AGE                36.164
# ALBUMIN             22.784

```

# CMPR_DSB_GRADE_L.3	17.727
# L_tube_feeding.1	17.064
# CMPR_DSB_GRADE_L.5	16.814
# L_tube_feeding.0	15.204
# BUN	13.060
# pneumonia.0	12.935
# CR	11.674
# pneumonia.1	11.672
# multifocal_large.1	11.520
# multifocal_large.0	11.352
# total_cowork	11.087
# GAIBJA_TYPE.0	10.067
# CITY_SIZE.7	9.378
# ipwon_period	8.536
# HB	7.912
# GLUCOSE_AC	7.218
# jaewon_GW	6.912
# CMPR_DSB_GRADE_L.1	5.677



```

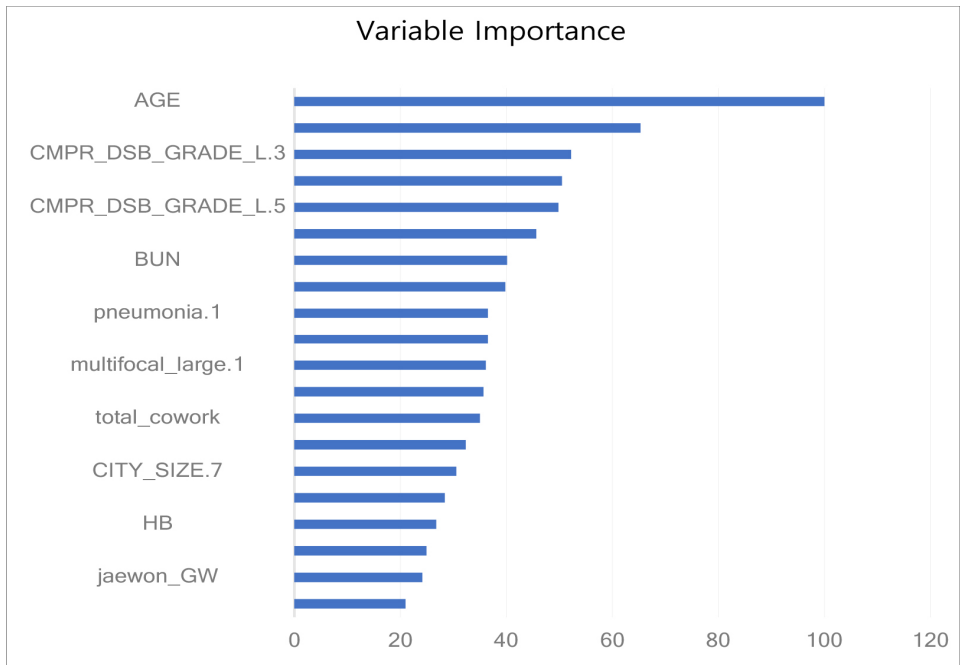
imp<-varImp(rf_fit, scale =TRUE)

imp

# rf variable importance
#
# only 20 most important variables shown (out of 147)
#
# Importance
# AGE                100.00
# ALBUMIN            65.32
# CMPR_DSB_GRADE_L.3  52.21
# L_tube_feeding.1   50.49
# CMPR_DSB_GRADE_L.5  49.84
# L_tube_feeding.0   45.67
# BUN                40.12
# pneumonia.0        39.79
# CR                 36.52
# pneumonia.1        36.52
# multifocal_large.1  36.12
# multifocal_large.0  35.69
# total_cowork        35.00
# GAIBJA_TYPE.0      32.36
# CITY_SIZE.7        30.57
# ipwon_period        28.39
# HB                 26.77
# GLUCOSE_AC          24.97
# jaewon_GW           24.18
# CMPR_DSB_GRADE_L.1  20.98

```





```
gbm_fit <- train(DEATH_YN~., data = trainmc, method = "gbm",
metric=metric, trControl=ctrl, tuneLength=10, verbose=FALSE,
na.action = na.omit)
```

```
# gbm_fit
# Stochastic Gradient Boosting
#
# 2081 samples
# 147 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1873, 1873, 1873, 1873, 1874, 1873, ...
# Resampling results across tuning parameters:
#
# interaction.depth n.trees Accuracy Kappa
# 1 50 0.8458406 0.4603123
```

# 1	100	0.8576620	0.5305717
# 1	150	0.8631418	0.5595801
# 1	200	0.8641011	0.5664360
# 1	250	0.8667000	0.5780322
# 1	300	0.8679500	0.5840866
# 1	350	0.8667966	0.5829369
# 1	400	0.8675677	0.5857158
# 1	450	0.8660343	0.5808889
# 1	500	0.8660338	0.5817486
# 2	50	0.8544917	0.5152425
# 2	100	0.8634289	0.5619001
# 2	150	0.8667975	0.5791580
# 2	200	0.8681423	0.5860851
# 2	250	0.8679519	0.5875709
# 2	300	0.8683323	0.5907175
# 2	350	0.8694881	0.5954579
# 2	400	0.8677568	0.5922936
# 2	450	0.8676648	0.5920622
# 2	500	0.8678571	0.5929116
# 3	50	0.8596821	0.5427980
# 3	100	0.8660274	0.5761898
# 3	150	0.8691002	0.5903053
# 3	200	0.8686185	0.5914178
# 3	250	0.8696771	0.5967496
# 3	300	0.8696813	0.5971423
# 3	350	0.8691011	0.5958089
# 3	400	0.8690068	0.5958587
# 3	450	0.8678539	0.5917934
# 3	500	0.8657417	0.5864193
# 4	50	0.8631372	0.5580862
# 4	100	0.8708296	0.5936428
# 4	150	0.8708346	0.5979725
# 4	200	0.8717033	0.6036415
# 4	250	0.8687225	0.5958440
# 4	300	0.8677591	0.5933430

# 4	350	0.8677600	0.5948977
# 4	400	0.8660251	0.5896208
# 4	450	0.8651588	0.5876040
# 4	500	0.8641959	0.5841348
# 5	50	0.8605448	0.5535377
# 5	100	0.8669908	0.5835743
# 5	150	0.8655526	0.5815498
# 5	200	0.8657486	0.5850922
# 5	250	0.8668951	0.5894964
# 5	300	0.8649766	0.5838793
# 5	350	0.8625723	0.5774817
# 5	400	0.8599738	0.5708066
# 5	450	0.8593987	0.5688951
# 5	500	0.8596886	0.5697274
# 6	50	0.8633342	0.5641117
# 6	100	0.8683333	0.5890496
# 6	150	0.8675631	0.5894680
# 6	200	0.8660265	0.5864107
# 6	250	0.8646808	0.5828923
# 6	300	0.8636259	0.5808139
# 6	350	0.8623781	0.5774556
# 6	400	0.8617051	0.5755023
# 6	450	0.8616071	0.5754769
# 6	500	0.8608378	0.5739950
# 7	50	0.8657367	0.5715139
# 7	100	0.8669875	0.5840844
# 7	150	0.8672746	0.5892537
# 7	200	0.8665105	0.5879612
# 7	250	0.8656446	0.5864022
# 7	300	0.8632426	0.5791666
# 7	350	0.8637211	0.5804668
# 7	400	0.8643914	0.5826361
# 7	450	0.8634303	0.5800163
# 7	500	0.8629505	0.5780440
# 8	50	0.8666982	0.5773269

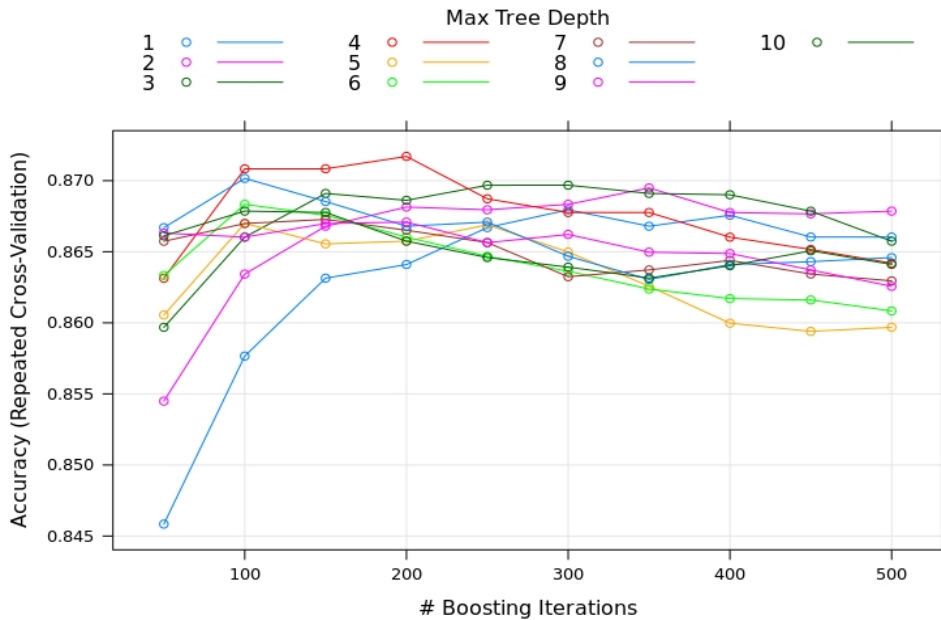
```

# 8          100      0.8701602  0.5957020
# 8          150      0.8685302  0.5953597
# 8          200      0.8667966  0.5897222
# 8          250      0.8670846  0.5906228
# 8          300      0.8646817  0.5838983
# 8          350      0.8630489  0.5794627
# 8          400      0.8641025  0.5820819
# 8          450      0.8643008  0.5840213
# 8          500      0.8645856  0.5839953
# 9           50      0.8663163  0.5788130
# 9          100      0.8660283  0.5846319
# 9          150      0.8669866  0.5906435
# 9          200      0.8670860  0.5919099
# 9          250      0.8656396  0.5871260
# 9          300      0.8662151  0.5896058
# 9          350      0.8649679  0.5861512
# 9          400      0.8648736  0.5870093
# 9          450      0.8637188  0.5829436
# 9          500      0.8625682  0.5793860
# 10         50      0.8661236  0.5772415
# 10         100      0.8678585  0.5920147
# 10         150      0.8677600  0.5934132
# 10         200      0.8657445  0.5884063
# 10         250      0.8645911  0.5834116
# 10         300      0.8639185  0.5823690
# 10         350      0.8631497  0.5802777
# 10         400      0.8640160  0.5817082
# 10         450      0.8650673  0.5852786
# 10         500      0.8641089  0.5829041
#
# Tuning parameter 'shrinkage' was held constant at a value of 0.1
# Tuning parameter 'n.minobsinnode' was held constant at a value
of 10
# Accuracy was used to select the optimal model using the largest
value.

```

```
# The final values used for the model were n.trees = 200,
interaction.depth = 4, shrinkage = 0.1 and n.minobsinnode = 10.
```

```
plot(gbm_fit)
```



```
test_pred_gbm <- predict(gbm_fit, newdata = testmc)
confusionMatrix(test_pred_gbm, testmc$DEATH_YN)
```

```
#           Reference
# Prediction  N   Y
#           N 609 76
#           Y  34 150
#
# Accuracy : 0.8734
# 95% CI : (0.8495, 0.8948)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.65
```

```

# McNemar's Test P-Value : 9.26e-05
#
#           Sensitivity : 0.9471
#           Specificity : 0.6637
#           Pos Pred Value : 0.8891
#           Neg Pred Value : 0.8152
#           Prevalence : 0.7399
#           Detection Rate : 0.7008
#           Detection Prevalence : 0.7883
#           Balanced Accuracy : 0.8054
#
#           'Positive' Class : N

imp<-varImp(gbm_fit, scale = FALSE)

imp

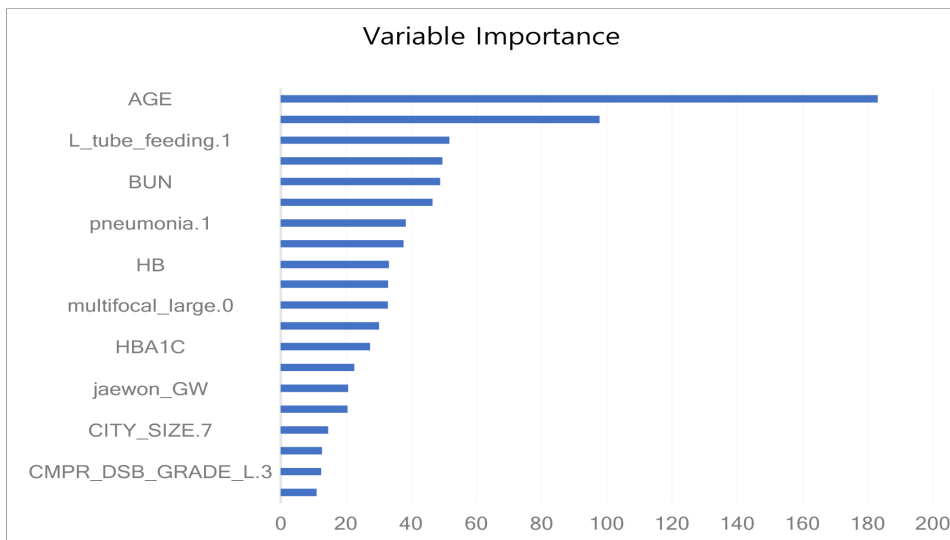
# gbm variable importance
#
# only 20 most important variables shown (out of 147)
#
# Overall
# AGE                183.10
# ALBUMIN             97.81
# L_tube_feeding.1   51.72
# CR                  49.63
# BUN                 48.91
# GLUCOSE_AC         46.53
# pneumonia.1        38.41
# ipwon_period        37.67
# HB                  33.21
# total_cowork        32.96
# multifocal_large.0 32.86
# pneumonia.0         30.21
# HBA1C              27.45

```

```

# L_tube_feeding.0      22.59
# jaewon_GW            20.66
# CMPR_DSB_GRADE_L.5  20.49
# CITY_SIZE.7         14.54
# CMPR_DSB_GRADE_L.1  12.65
# CMPR_DSB_GRADE_L.3  12.44
# multifocal_large.1  11.03

```



```
imp<-varImp(gbm_fit, scale = TRUE)
```

```
imp
```

```
# gbm variable importance
```

```
#
```

```
# only 20 most important variables shown (out of 147)
```

```
#
```

```
# Overall
```

```
# AGE                100
```

```
# ALBUMIN            57.407
```

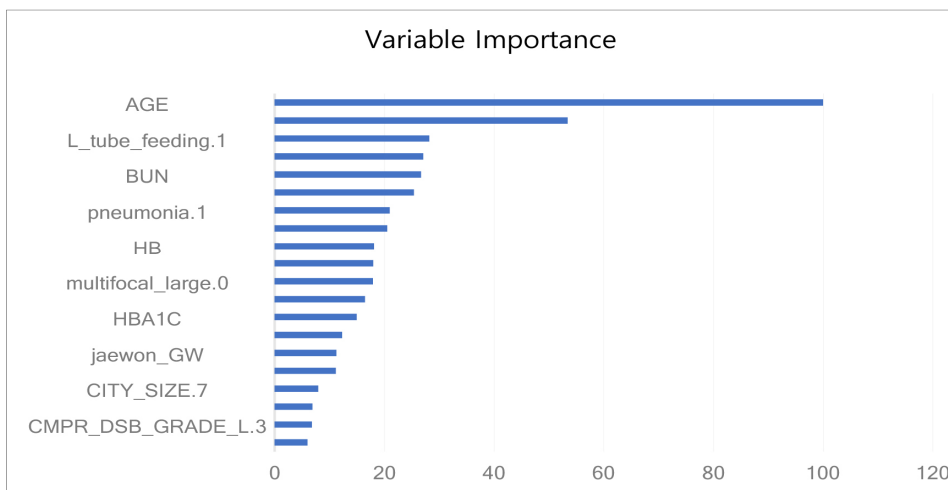
```
# L_tube_feeding.0  40.434
```

```
# BUN                32.265
```

```

# GLUCOSE_AC      29.929
# CR              29.032
# pneumonia.0    27.916
# ipwon_period   25.77
# HB              24.541
# HBA1C           20.272
# multifocal_large.0 16.846
# total_cowork    16.706
# jaewon_GW       15.115
# pneumonia.1    14.555
# L_tube_feeding.1 14.007
# CMPR_DSB_GRADE_L.5 7.817
# CMPR_DSB_GRADE_L.1 7.052
# multifocal_large.1 6.706
# CMPR_DSB_GRADE_L.2 6.17
# CMPR_DSB_GRADE_L.3 5.143

```



```

> rpart_fit <- train(DEATH_YN~., data = trainmc, method =
"rpart")
> rpart_fit

```

CART

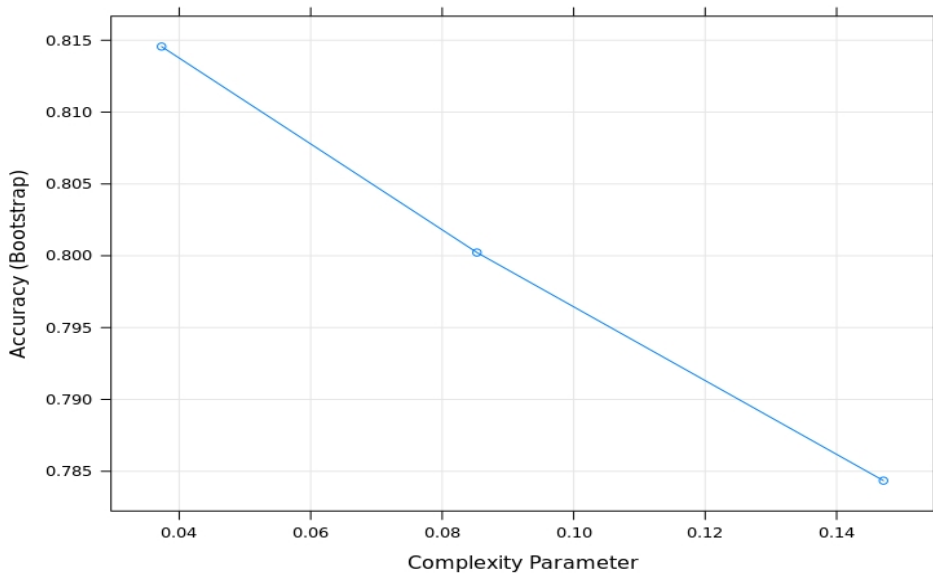


```

# 2081 samples
# 147 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Bootstrapped (25 reps)
# Summary of sample sizes: 2081, 2081, 2081, 2081, 2081, 2081, ...
# Resampling results across tuning parameters:
#
#   cp          Accuracy   Kappa
# 0.03731343  0.8145619  0.3664190
# 0.08528785  0.8002250  0.3227488
# 0.14712154  0.7843474  0.1772301
#
# Accuracy was used to select the optimal model using the largest
value.
# The final value used for the model was cp = 0.03731343.

```

```
plot(rpart_fit)
```



```
test_pred_rpart <- predict(rpart_fit, newdata = testmc)
```

```
confusionMatrix (test_pred_rpart, testmc$DEATH_YN)
```

```
# Reference
# Prediction   N   Y
# N  625  145
# Y   18   81
#
# Accuracy : 0.8124
# 95% CI : (0.7849, 0.8379)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : 2.925e-07
#
# Kappa : 0.404
# Mcnemar's Test P-Value : < 2.2e-16
#
#           Sensitivity : 0.9720
#           Specificity : 0.3584
#           Pos Pred Value : 0.8117
#           Neg Pred Value : 0.8182
#           Prevalence : 0.7399
#           Detection Rate : 0.7192
#           Detection Prevalence : 0.8861
#           Balanced Accuracy : 0.6652
#
#           'Positive' Class : N
```

```
imp<-varImp(rpart_fit, scale =TRUE)
```

```
imp
```

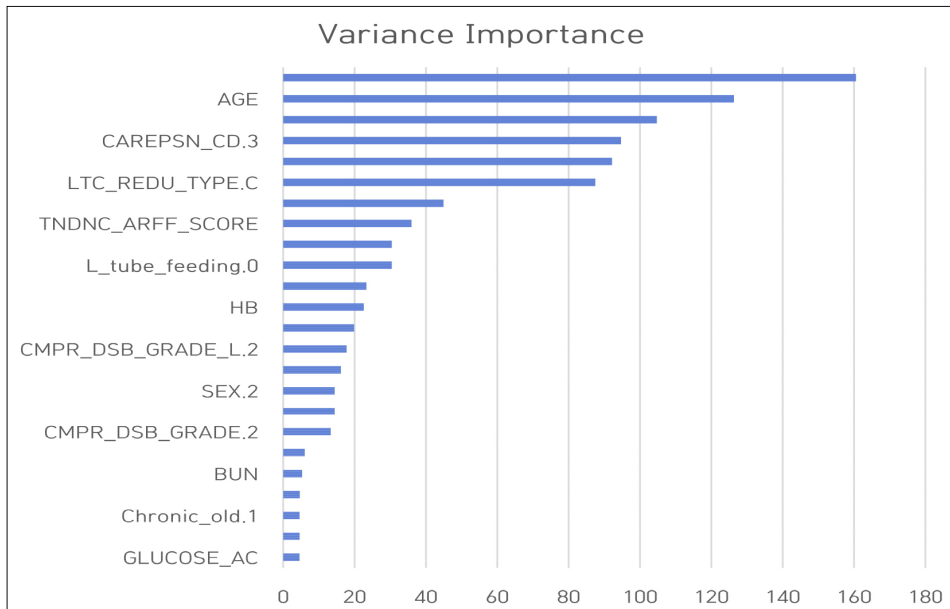
```
# AGE                126.330119
# ALBUMIN            160.554024
# BUN                 5.290000
# CAREPSN_CD.11      23.318628
# CAREPSN_CD.3       94.677169
# Chronic_old.0      4.576784
```

```

# Chronic_old.1      4.576784
# CMPR_DSB_GRADE_L.1 16.162466
# CMPR_DSB_GRADE_L.2 17.761503
# CMPR_DSB_GRADE_L.5 19.850712
# CMPR_DSB_GRADE.1  4.641653
# CMPR_DSB_GRADE.2  13.310000
# CR                 44.932035
# GLUCOSE_AC         4.558651
# HB                 22.574412
# L_tube_feeding.0  30.430029
# L_tube_feeding.1  30.430029
# LTC_REDU_TYPE.C   87.480856
# LTC_REDU_TYPE.D   104.719361
# PCT_ARFF_SCORE    6.022037
# RCG_RCGT_SCORE    92.165920
# SEX.1              14.417745
# SEX.2              14.417745
# TNDNC_ARFF_SCORE  35.955768

```

```
plot(rpart_fit)
```



```

tunedxgbgrid <- expand.grid(nrounds=500, max_depth=7, eta=0.01,
gamma=0, colsample_bytree=0.6, min_child_weight=1, subsample=0.8)

ctrl <- trainControl(method="repeatedcv",
number=10,repeats=5,classProbs = TRUE)

set.seed(seed)
xgb_fit <- train(DEATH_YN~., data = trainmc, method = "xgbTree",
metric=metric, trControl=ctrl, tuneGrid=tunedxgbgrid,na.action =
na.omit)
xgb_fit

# eXtreme Gradient Boosting
#
# 2081 samples
# 147 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 1873, 1873, 1872, 1872, 1874, 1873, ...
# Resampling results:
#
# Accuracy   Kappa
# 0.8694889  0.584074
#
# Tuning parameter 'nrounds' was held constant at a value of 500
# Tuning parameter 'max_depth' was held constant at a value of 7
# parameter 'colsample_bytree' was held constant at a value of 0.6
# Tuning parameter 'min_child_weight' was held constant at a
value of
# 1
# Tuning parameter 'subsample' was held constant at a value of 0.8

test_pred_xgb <- predict(xgb_fit, newdata = testmc)

```

```

confusionMatrix(test_pred_xgb, testmc$DEATH_YN)

# Reference
# Prediction   N   Y
#           N 618  77
#           Y  25 149
#
# Accuracy : 0.8826
# 95% CI : (0.8593, 0.9033)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.6704
# McNemar's Test P-Value : 4.424e-07
#
#           Sensitivity : 0.9611
#           Specificity : 0.6593
#           Pos Pred Value : 0.8892
#           Neg Pred Value : 0.8563
#           Prevalence : 0.7399
#           Detection Rate : 0.7112
#           Detection Prevalence : 0.7998
#           Balanced Accuracy : 0.8102
#
#           'Positive' Class : N

imp<-varImp(xgb_fit, scale = FALSE)

imp

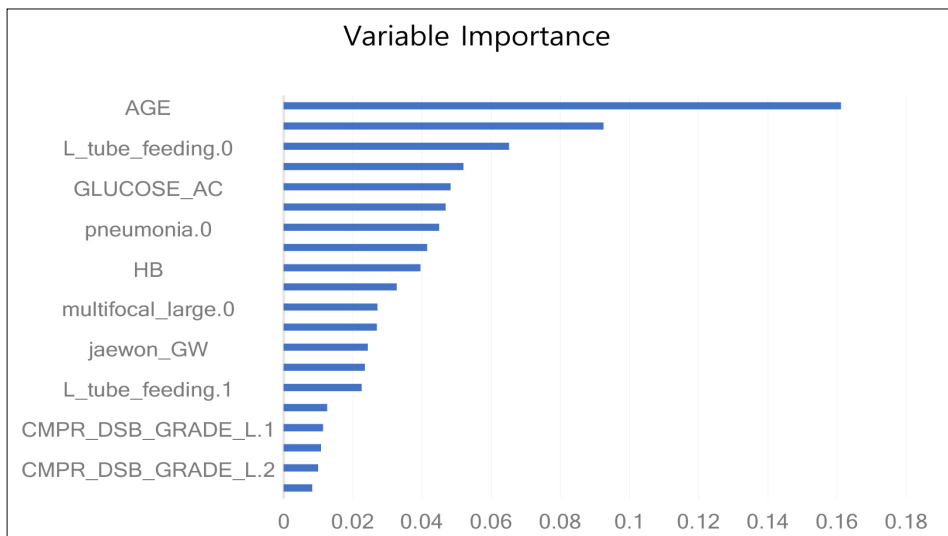
# only 20 most important variables shown (out of 147)
#
# Overall
# AGE                0.161187
# ALBUMIN            0.092533

```

```

# L_tube_feeding.0  0.065175
# BUN                0.052007
# GLUCOSE_AC        0.048241
# CR                0.046796
# pneumonia.0       0.044998
# ipwon_period      0.041538
# HB                0.039556
# HBA1C             0.032675
# multifocal_large.0 0.027154
# total_cowork      0.026928
# jaewon_GW         0.024364
# pneumonia.1       0.023461
# L_tube_feeding.1  0.022578
# CMPR_DSB_GRADE_L.5 0.012600
# CMPR_DSB_GRADE_L.1 0.011366
# multifocal_large.1 0.010809
# CMPR_DSB_GRADE_L.2 0.009946
# CMPR_DSB_GRADE_L.3 0.008289
#

```



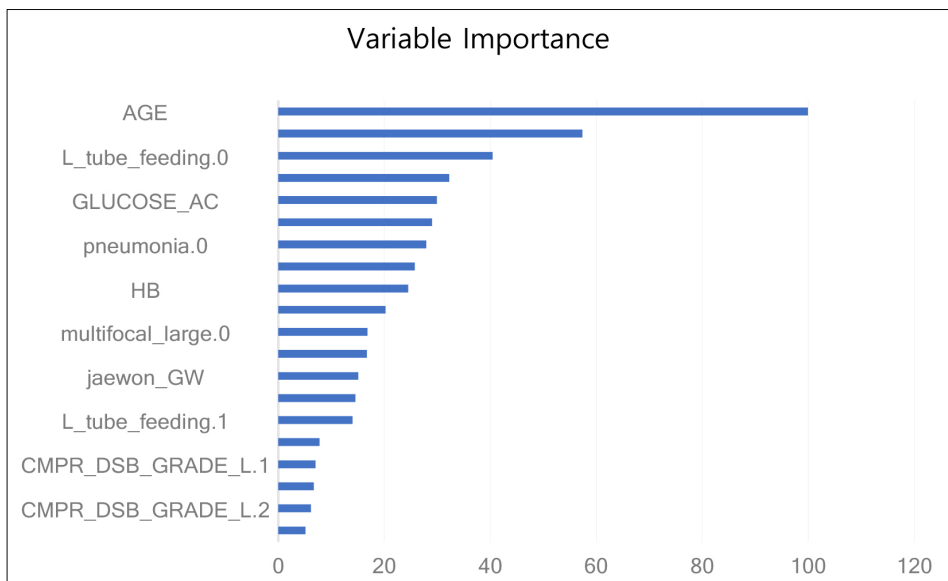
```
imp<-varImp(xgb_fit, scale = TRUE)
```

```

imp

# xgbTree variable importance
#
# only 20 most important variables shown (out of 147)
#
# Overall
# AGE                100.000
# ALBUMIN             57.407
# L_tube_feeding.0   40.434
# BUN                 32.265
# GLUCOSE_AC         29.929
# CR                  29.032
# pneumonia.0        27.916
# ipwon_period        25.770
# HB                  24.541
# HBA1C               20.272
# multifocal_large.0 16.846
# total_cowork        16.706
# jaewon_GW           15.115
# pneumonia.1         14.555
# L_tube_feeding.1   14.007
# CMPR_DSB_GRADE_L.5  7.817
# CMPR_DSB_GRADE_L.1  7.052
# multifocal_large.1  6.706
# CMPR_DSB_GRADE_L.2  6.170
# CMPR_DSB_GRADE_L.3  5.143

```



```

ovrctrl <- trainControl(method="repeatedcv", number=10,
repeats=5,
                        savePredictions='final',
                        classProbs=TRUE,
                        summaryFunction = twoClassSummary,
                        allowParallel=TRUE,
                        index = createFolds(trainmc$DEATH_YN, 10))

metric <- "ROC"
algorithmList <-
c('nnet','svmRadial','knn','rpart','rf','gbm','xgbTree')

model_DEATH <- caretList(DEATH_YN~., data=trainmc,
                        metric=metric,
                        trControl=ovrctrl,
                        methodList= algorithmList,
                        tuneList = NULL,#ClassProbs=F,
                        continue_on_fail = FALSE)# ,na.action =
na.omit) # na.action = na.roughfix

```



```

stctrl <- trainControl(method="repeatedcv", number=10, repeats=5,
                      savePredictions=TRUE, classProbs=TRUE,
summaryFunction = twoClassSummary,
                      allowParallel=TRUE,
                      index = createFolds(trainmc$DEATH_YN, 10))

set.seed(seed)
stack.rf_DEATH <- caretStack(model_DEATH, method="rf",
metric="ROC", trControl=stctrl)
print(stack.rf_DEATH)

# A rf ensemble of 2 base models: nnet, svmRadial, knn, rpart, rf,
gbm, xgbTree
#
# Ensemble results:
#   Random Forest
#
# 18729 samples
# 7 predictor
# 2 classes: 'N', 'Y'
#
# No pre-processing
# Resampling: Cross-Validated (10 fold, repeated 5 times)
# Summary of sample sizes: 208, 209, 208, 208, 208, 208, ...
# Resampling results across tuning parameters:
#
#  mtry  ROC  Sens  Spec
#  2     0.5  NaN   NaN
#  4     0.5  NaN   NaN
#  7     0.5  NaN   NaN
#
# ROC was used to select the optimal model using the largest
value.

```

```

# The final value used for the model was mtry = 2.

test_pred_DEATH <- predict(stack.rf_DEATH, newdata = testmc)
confusionMatrix(test_pred_DEATH, testmc$DEATH_YN)
Confusion Matrix and Statistics

# Reference
# Prediction   N   Y
#           N 609  78
#           Y  34 148
#
# Accuracy : 0.8711
# 95% CI : (0.8470, 0.8927)
# No Information Rate : 0.7399
# P-Value [Acc > NIR] : < 2.2e-16
#
# Kappa : 0.6426
# McNemar's Test P-Value : 4.842e-05

#
#           Sensitivity : 0.9471
#           Specificity : 0.6549
#           Pos Pred Value : 0.8865
#           Neg Pred Value : 0.8132
#           Prevalence : 0.7399
#           Detection Rate : 0.7008
#           Detection Prevalence : 0.7906
#           Balanced Accuracy : 0.8010
#
#           'Positive' Class : N
>

```

장기요양 서비스 미이용자군의 생존 기간과 생존에 대한 회귀 분석

```
result=lm(Live_days~ipwon_period+AGE+SEX+SIDO_NM+CITY_SIZE+GAIBJA_T  
YPE+CMPR_DSB_GRADE+MAIN_DSB_TYPE+CMPR_DSB_GRADE_L+MAIN_DSB_TYPE_L  
+  
+L_tube_feeding+pneumonia+Infarction+SAH+SDH+ICH+IVH+Focal_small+  
multifocal_large+Acute_recent+Chronic_old  
+  
+Temporal+Cerebellum+BG+brain_stem+Thalamus+MCA+Aneurysm+Coil+cra  
niotomy+smoke_yn+drink_yn  
+  
+ALBUMIN+BUN+CR+GLUCOSE_AC+HB+HBA1C+jaewon_GW+total_cowork+pum_co  
work, data=Data_final)  
> summary(result)
```

Call:

```
lm(formula = Live_days ~ ipwon_period + AGE + SEX + SIDO_NM +  
CITY_SIZE + GAIBJA_TYPE + CMPR_DSB_GRADE + MAIN_DSB_TYPE +  
CMPR_DSB_GRADE_L + MAIN_DSB_TYPE_L + L_tube_feeding +  
pneumonia +  
Infarction + SAH + SDH + ICH + IVH + Focal_small +  
multifocal_large +  
Acute_recent + Chronic_old + Temporal + Cerebellum + BG +  
brain_stem + Thalamus + MCA + Aneurysm + Coil + craniotomy +  
smoke_yn + drink_yn + ALBUMIN + BUN + CR + GLUCOSE_AC + HB +  
HBA1C + jaewon_GW + total_cowork + pum_cowork, data =  
Data_final)
```

Residuals:

Min	1Q	Median	3Q	Max
-2782.9	-675.2	-47.2	624.9	3796.8

Coefficients: (6 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept)		3661.9860	526.1608	6.960	4.21e-12 ***
-------------	--	-----------	----------	-------	--------------

ipwon_period	8.0676	1.5247	5.291	1.31e-07	***
AGE	-25.3162	1.6769	-15.097	< 2e-16	***
SEX2	57.3669	48.4744	1.183	0.236731	
SIDO_NM2	-495.7921	330.1862	-1.502	0.133324	
SIDO_NM3	-862.3053	366.4514	-2.353	0.018684	*
SIDO_NM4	-31.6609	127.0812	-0.249	0.803271	
SIDO_NM5	-1049.8113	362.8829	-2.893	0.003845	**
SIDO_NM6	67.5984	394.4934	0.171	0.863957	
SIDO_NM7	464.3649	951.2194	0.488	0.625461	
SIDO_NM8	-1077.6408	951.4700	-1.133	0.257475	
SIDO_NM9	-119.1778	52.9912	-2.249	0.024588	*
SIDO_NM10	253.1752	211.0683	1.199	0.230436	
SIDO_NM11	-71.2884	321.2695	-0.222	0.824411	
SIDO_NM12	-427.4683	291.7657	-1.465	0.143002	
SIDO_NM13	-685.1712	287.0798	-2.387	0.017066	*
SIDO_NM14	-412.0281	244.7039	-1.684	0.092334	.
SIDO_NM15	-256.9915	248.8241	-1.033	0.301774	
SIDO_NM16	-177.7513	351.1764	-0.506	0.612784	
SIDO_NM17	24.2253	615.1701	0.039	0.968590	
CITY_SIZE2	-232.9987	66.0181	-3.529	0.000423	***
CITY_SIZE3	180.1639	219.5369	0.821	0.411912	
CITY_SIZE4	NA	NA	NA	NA	
CITY_SIZE5	-212.4848	55.1872	-3.850	0.000121	***
CITY_SIZE6	-196.7662	50.4775	-3.898	9.92e-05	***
CITY_SIZE7	-65.9322	464.0357	-0.142	0.887023	
CITY_SIZE8	907.2581	528.3715	1.717	0.086073	.
GAIBJA_TYPE1	530.7412	460.7933	1.152	0.249501	
GAIBJA_TYPE2	396.1655	460.6508	0.860	0.389855	
GAIBJA_TYPE3	537.2362	461.8499	1.163	0.244835	
GAIBJA_TYPE4	482.9498	462.0310	1.045	0.295985	
GAIBJA_TYPE5	476.8966	460.3052	1.036	0.300270	
GAIBJA_TYPE6	449.2077	460.2818	0.976	0.329177	
GAIBJA_TYPE7	333.7787	463.2251	0.721	0.471243	
GAIBJA_TYPE8	431.2357	459.9593	0.938	0.348555	
GAIBJA_TYPE9	316.1747	461.7987	0.685	0.493615	

GAIBJA_TYPE10	496.4381	462.0768	1.074	0.282751
GAIBJA_TYPE11	341.1432	461.9360	0.739	0.460267
GAIBJA_TYPE12	550.2502	458.7726	1.199	0.230474
GAIBJA_TYPE13	371.0136	460.5619	0.806	0.420560
GAIBJA_TYPE14	519.6929	460.4198	1.129	0.259104
GAIBJA_TYPE15	442.0987	459.6393	0.962	0.336213
GAIBJA_TYPE16	445.3898	458.1642	0.972	0.331075
GAIBJA_TYPE17	504.5604	457.8524	1.102	0.270548
GAIBJA_TYPE18	613.2314	456.6654	1.343	0.179429
GAIBJA_TYPE19	541.1780	456.8959	1.184	0.236328
GAIBJA_TYPE20	529.4426	455.6396	1.162	0.245343
CMPR_DSB_GRADE1	71.7746	1003.9107	0.071	0.943009
CMPR_DSB_GRADE2	153.2879	996.5705	0.154	0.877766
CMPR_DSB_GRADE3	269.9762	1001.2992	0.270	0.787468
CMPR_DSB_GRADE4	158.2106	982.5480	0.161	0.872088
CMPR_DSB_GRADE5	446.9712	994.2575	0.450	0.653067
CMPR_DSB_GRADE6	348.8729	987.1999	0.353	0.723818
MAIN_DSB_TYPE1	-1097.5059	995.3574	-1.103	0.270284
MAIN_DSB_TYPE2	-1034.8903	971.0699	-1.066	0.286640
MAIN_DSB_TYPE3	-823.7301	1031.3264	-0.799	0.424526
MAIN_DSB_TYPE4	-838.0259	1060.4886	-0.790	0.429462
MAIN_DSB_TYPE5	-1057.6849	1025.5055	-1.031	0.302451
MAIN_DSB_TYPE6	-1177.5909	1072.1880	-1.098	0.272164
MAIN_DSB_TYPE8	-467.9184	1118.5556	-0.418	0.675742
MAIN_DSB_TYPE9	-742.5103	1067.0786	-0.696	0.486589
MAIN_DSB_TYPE10	-1669.0839	1435.6323	-1.163	0.245085
MAIN_DSB_TYPE11	-1997.0914	1241.4349	-1.609	0.107794
MAIN_DSB_TYPE12	-2888.1543	1180.7335	-2.446	0.014503 *
MAIN_DSB_TYPE14	-899.8982	1087.0354	-0.828	0.407827
MAIN_DSB_TYPE15	NA	NA	NA	NA
CMPR_DSB_GRADE_L1	687.8252	230.9749	2.978	0.002927 **
CMPR_DSB_GRADE_L2	417.7359	199.5194	2.094	0.036374 *
CMPR_DSB_GRADE_L3	-201.3624	302.1895	-0.666	0.505245
CMPR_DSB_GRADE_L4	234.8369	294.0385	0.799	0.424554
CMPR_DSB_GRADE_L5	-452.1088	273.0600	-1.656	0.097891 .

CMPR_DSB_GRADE_L6	NA	NA	NA	NA
MAIN_DSB_TYPE_L1	602.3318	1390.4918	0.433	0.664918
MAIN_DSB_TYPE_L2	709.8949	1366.6266	0.519	0.603487
MAIN_DSB_TYPE_L3	457.1146	1412.1184	0.324	0.746183
MAIN_DSB_TYPE_L4	482.2526	1431.7642	0.337	0.736274
MAIN_DSB_TYPE_L5	911.0374	1409.2196	0.646	0.518018
MAIN_DSB_TYPE_L6	336.1281	1438.3763	0.234	0.815246
MAIN_DSB_TYPE_L8	NA	NA	NA	NA
MAIN_DSB_TYPE_L9	50.6959	1430.2146	0.035	0.971726
MAIN_DSB_TYPE_L10	1083.1795	1797.3251	0.603	0.546782
MAIN_DSB_TYPE_L11	1065.0757	1718.7563	0.620	0.535519
MAIN_DSB_TYPE_L12	1613.3950	1676.6894	0.962	0.336006
MAIN_DSB_TYPE_L14	NA	NA	NA	NA
MAIN_DSB_TYPE_L15	NA	NA	NA	NA
L_tube_feeding1	-252.5080	62.4545	-4.043	5.42e-05 ***
pneumonia1	-162.0117	40.4771	-4.003	6.43e-05 ***
Infarction1	88.5182	59.5152	1.487	0.137041
SAH1	42.2602	73.6616	0.574	0.566211
SDH1	204.5093	73.9073	2.767	0.005692 **
ICH1	-244.9451	51.4472	-4.761	2.02e-06 ***
IVH1	-236.5221	85.6272	-2.762	0.005778 **
Focal_small1	54.7965	43.2576	1.267	0.205349
multifocal_large1	-152.3883	49.6842	-3.067	0.002182 **
Acute_recent1	-9.4431	46.8892	-0.201	0.840407
Chronic_old1	128.1326	51.8740	2.470	0.013567 *
Temporal1	135.5661	76.2464	1.778	0.075511 .
Cerebellum1	-37.8516	59.2162	-0.639	0.522738
BG1	66.5906	46.3552	1.437	0.150962
brain_stem1	-101.2757	71.5656	-1.415	0.157136
Thalamus1	48.0326	64.4978	0.745	0.456504
MCA1	-65.0828	46.4504	-1.401	0.161286
Aneurysm1	11.5391	53.2607	0.217	0.828494
Coil1	253.0840	97.7750	2.588	0.009691 **
craniotomy1	176.2629	78.6039	2.242	0.025011 *
smoke_yn2	-92.1750	42.7600	-2.156	0.031196 *

```

smoke_yn3      -161.9347    54.7615   -2.957 0.003131 **
drink_yn2      -168.5373    57.1147   -2.951 0.003195 **
drink_yn3      -236.6559    47.5343   -4.979 6.79e-07 ***
ALBUMIN        -239.4982    44.3137   -5.405 7.03e-08 ***
BUN            -13.8243     2.6246   -5.267 1.49e-07 ***
CR             48.5361    24.9131    1.948 0.051487 .
GLUCOSE_AC     -1.5502     0.4599   -3.370 0.000760 ***
HB            55.2838    11.5601    4.782 1.82e-06 ***
HBA1C         64.3654    16.8222    3.826 0.000133 ***
jaewon_GW     -7.4694     1.7313   -4.314 1.66e-05 ***
total_cowork   -36.8833     4.3156   -8.547 < 2e-16 ***
pum_cowork     68.2425    28.8323    2.367 0.018006 *

```

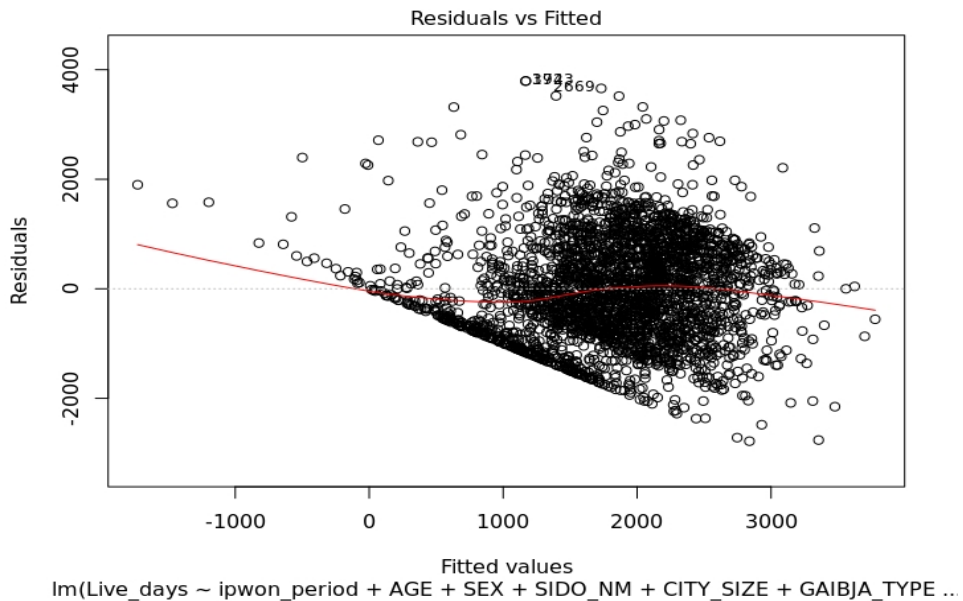
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

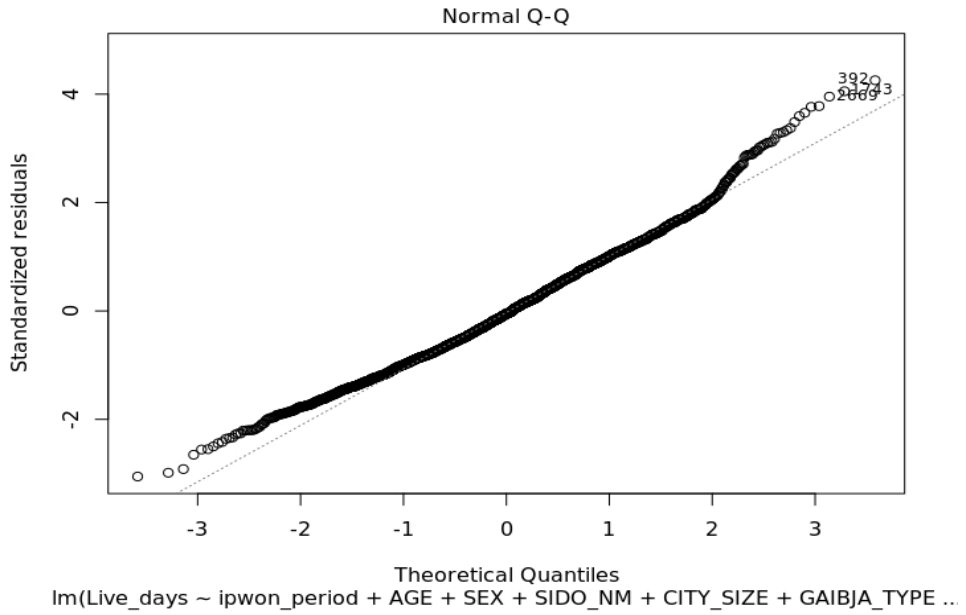
Residual standard error: 942.8 on 2838 degrees of freedom  
Multiple R-squared: 0.3115, Adjusted R-squared: 0.2846  
F-statistic: 11.57 on 111 and 2838 DF, p-value: < 2.2e-16

plot(result)

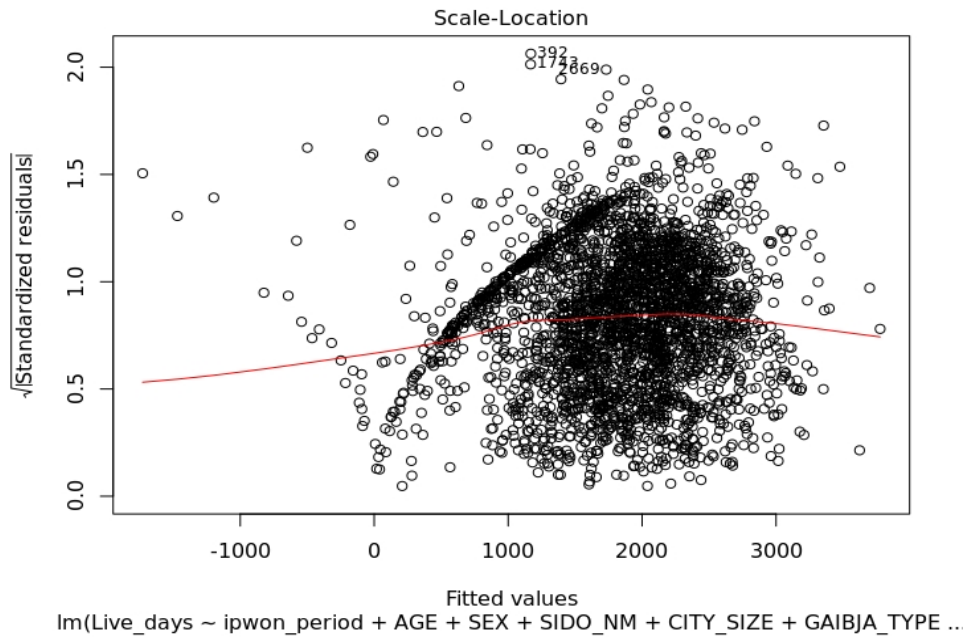
Hit <Return> to see next plot:



Hit <Return> to see next plot:

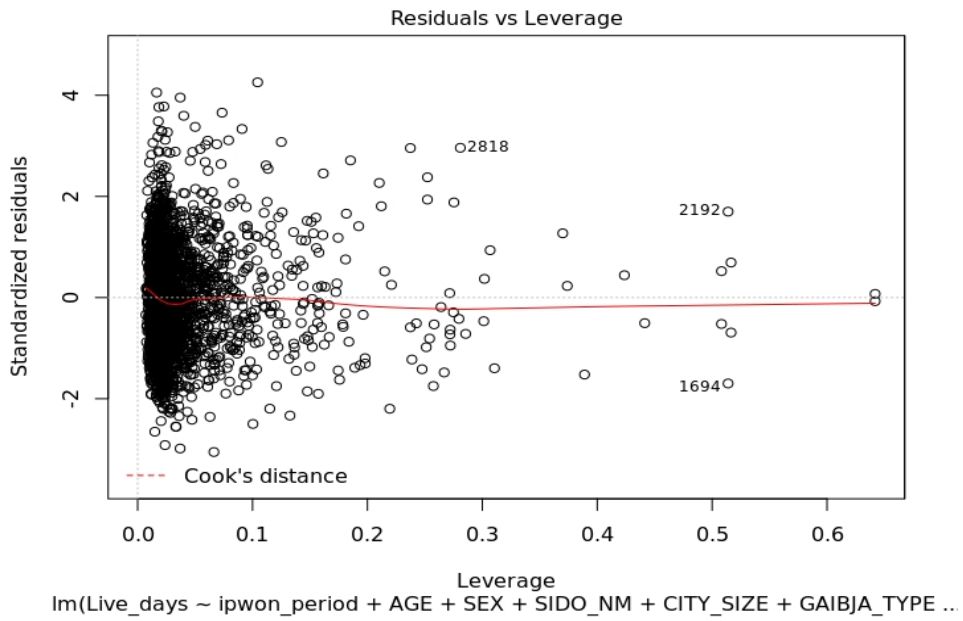


Hit <Return> to see next plot:





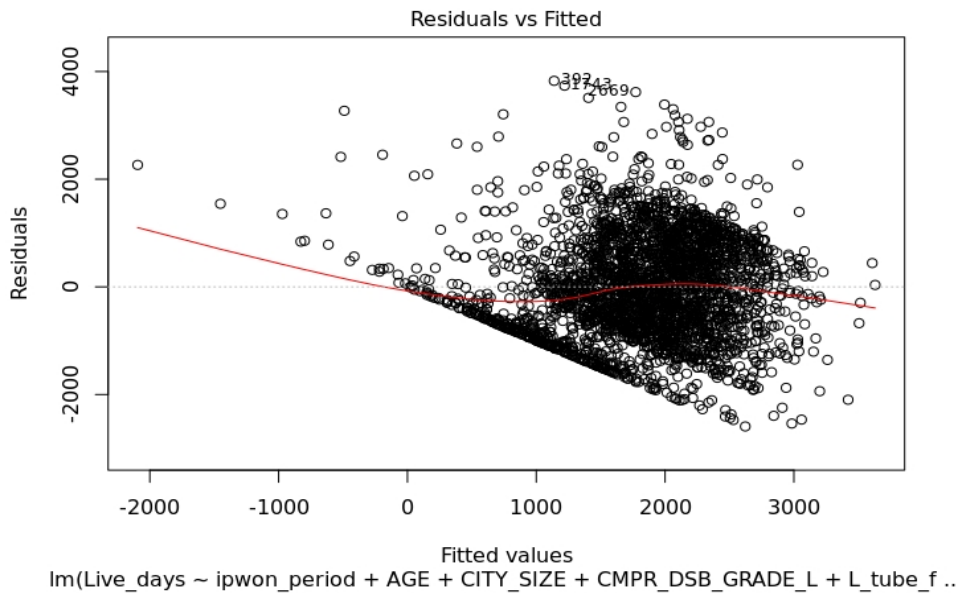
Hit <Return> to see next plot:



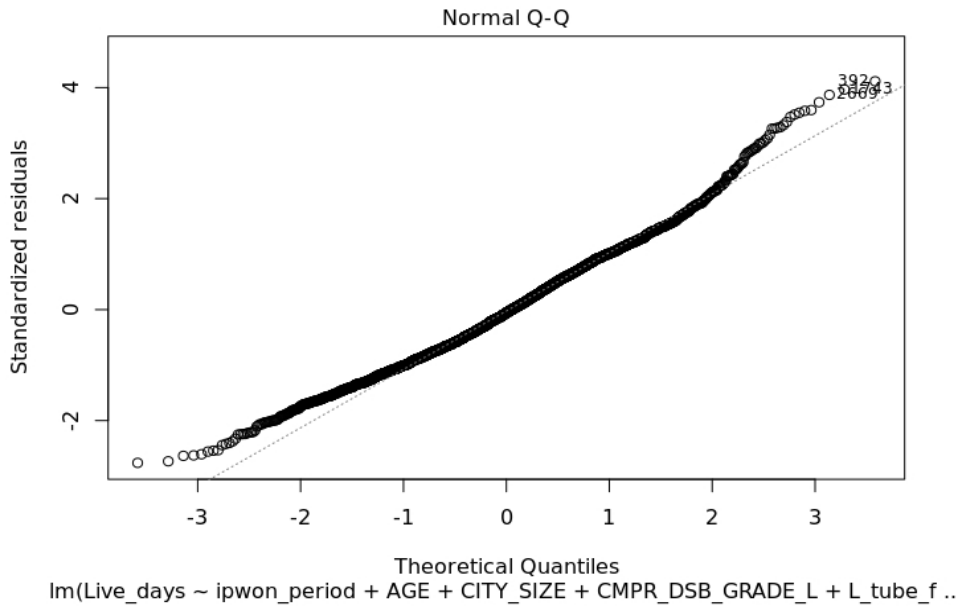
```
> result_2=step(result, direction="backward")
```

```
plot(result_2)
```

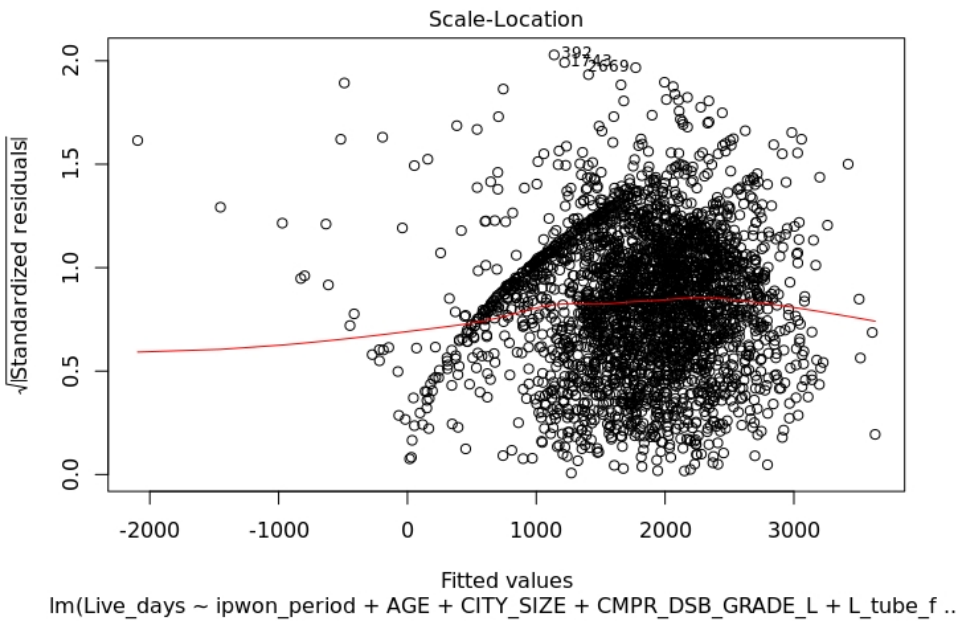
Hit <Return> to see next plot:



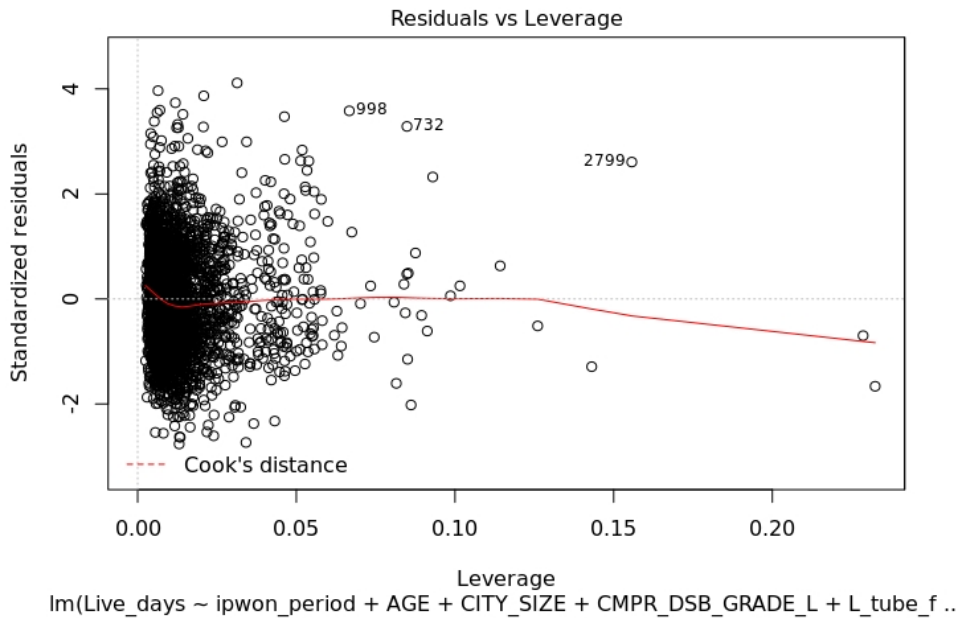
Hit <Return> to see next plot:



Hit <Return> to see next plot:



Hit <Return> to see next plot:



```

>
> summary(result_2)
Call:
lm(formula = DEATH_YN ~ AGE + SEX + CITY_SIZE + CMPR_DSB_GRADE
+
      MAIN_DSB_TYPE + CMPR_DSB_GRADE_L + L_tube_feeding +
      pneumonia +
      SDH + ICH + IVH + Focal_small + multifocal_large +
      Acute_recent +
      brain_stem + Aneurysm + Coil + ALBUMIN + BUN + CR +
      GLUCOSE_AC +
      HB + jaewon_GW + pum_cowork, data = Data_final)

Residuals:
    Min     1Q   Median     3Q      Max
-1.0743 -0.1928 -0.0541  0.1123  1.1418

Coefficients: (2 not defined because of singularities)
Estimate Std. Error t value Pr(>|t|)

```

(Intercept)	1.3680880	0.0782912	17.474	< 2e-16	***
AGE	0.0061796	0.0005195	11.896	< 2e-16	***
SEX2	-0.0612800	0.0139626	-4.389	1.18e-05	***
CITY_SIZE2	0.0520778	0.0210338	2.476	0.013346	*
CITY_SIZE3	0.0634164	0.0449182	1.412	0.158110	
CITY_SIZE4	0.1005160	0.0502460	2.000	0.045542	*
CITY_SIZE5	0.0131021	0.0177450	0.738	0.460360	
CITY_SIZE6	0.0236277	0.0164407	1.437	0.150786	
CITY_SIZE7	0.1901250	0.0338433	5.618	2.12e-08	***
CITY_SIZE8	0.0082392	0.0898882	0.092	0.926974	
CMPR_DSB_GRADE1	0.8533718	0.2424986	3.519	0.000440	***
CMPR_DSB_GRADE2	0.6446583	0.2379081	2.710	0.006774	**
CMPR_DSB_GRADE3	0.6973487	0.2413028	2.890	0.003882	**
CMPR_DSB_GRADE4	0.5675386	0.2340676	2.425	0.015383	*
CMPR_DSB_GRADE5	0.5070147	0.2334651	2.172	0.029960	*
CMPR_DSB_GRADE6	0.4893608	0.2322049	2.107	0.035164	*
MAIN_DSB_TYPE1	-0.2325934	0.2247912	-1.035	0.300891	
MAIN_DSB_TYPE2	-0.2617657	0.2254521	-1.161	0.245709	
MAIN_DSB_TYPE3	-0.1967296	0.2282152	-0.862	0.388739	
MAIN_DSB_TYPE4	-0.2847167	0.2268039	-1.255	0.209455	
MAIN_DSB_TYPE5	-0.3022788	0.2328855	-1.298	0.194400	
MAIN_DSB_TYPE6	-0.1352572	0.2419078	-0.559	0.576118	
MAIN_DSB_TYPE8	-0.5492589	0.2908209	-1.889	0.059039	.
MAIN_DSB_TYPE9	-0.2093353	0.2339812	-0.895	0.371039	
MAIN_DSB_TYPE10	-0.3424211	0.2687771	-1.274	0.202767	
MAIN_DSB_TYPE11	-0.0223481	0.2744162	-0.081	0.935099	
MAIN_DSB_TYPE12	0.3098591	0.2742072	1.130	0.258562	
MAIN_DSB_TYPE14	-0.1718252	0.2750153	-0.625	0.532162	
MAIN_DSB_TYPE15	NA	NA	NA	NA	
CMPR_DSB_GRADE_L1	-0.6416134	0.0741586	-8.652	< 2e-16	***
CMPR_DSB_GRADE_L2	-0.4252258	0.0650711	-6.535	7.50e-11	***
CMPR_DSB_GRADE_L3	-0.0494609	0.0971815	-0.509	0.610824	
CMPR_DSB_GRADE_L4	0.0881195	0.0951354	0.926	0.354391	
CMPR_DSB_GRADE_L5	0.2163423	0.0888736	2.434	0.014982	*
CMPR_DSB_GRADE_L6	NA	NA	NA	NA	

L_tube_feeding1	0.2203344	0.0198854	11.080	< 2e-16	***
pneumonia1	0.1236117	0.0130919	9.442	< 2e-16	***
SDH1	-0.0511272	0.0225680	-2.265	0.023557	*
ICH1	-0.0262236	0.0157092	-1.669	0.095162	.
IVH1	0.1521735	0.0271270	5.610	2.22e-08	***
Focal_small1	-0.0305576	0.0130281	-2.346	0.019068	*
multifocal_large1	0.1269455	0.0160205	7.924	3.25e-15	***
Acute_recent1	0.0198919	0.0131377	1.514	0.130106	
brain_stem1	0.0599641	0.0228015	2.630	0.008588	**
Aneurysm1	-0.0519983	0.0156830	-3.316	0.000926	***
Coil1	-0.1461262	0.0288353	-5.068	4.28e-07	***
ALBUMIN	-0.1111647	0.0143398	-7.752	1.24e-14	***
BUN	0.0012650	0.0008482	1.491	0.135978	
CR	0.0155330	0.0075404	2.060	0.039492	*
GLUCOSE_AC	0.0004571	0.0001219	3.749	0.000181	***
HB	-0.0197426	0.0037176	-5.311	1.18e-07	***
jaewon_GW	0.0008715	0.0004291	2.031	0.042364	*
pum_cowork	0.0266370	0.0081010	3.288	0.001021	**

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3111 on 2899 degrees of freedom

Multiple R-squared: 0.4719, Adjusted R-squared: 0.4628

F-statistic: 51.8 on 50 and 2899 DF, p-value: < 2.2e-16

anova(result, result\_2)

Analysis of Variance Table

Model 1: Live\_days ~ ipwon\_period + AGE + SEX + SIDO\_NM + CITY\_SIZE +

GAIBJA\_TYPE + CMPR\_DSB\_GRADE + MAIN\_DSB\_TYPE + CMPR\_DSB\_GRADE\_L +

MAIN\_DSB\_TYPE\_L + L\_tube\_feeding + pneumonia + Infarction + SAH + SDH + ICH + IVH + Focal\_small + multifocal\_large + Acute\_recent + Chronic\_old + Temporal + Cerebellum + BG + brain\_stem + Thalamus + MCA + Aneurysm + Coil + craniotomy +

```

smoke_yn + drink_yn + ALBUMIN + BUN + CR + GLUCOSE_AC + HB +
HBA1C + jaewon_GW + total_cowork + pum_cowork
Model 2: Live_days ~ ipwon_period + AGE + CITY_SIZE +
CMPR_DSB_GRADE_L +
L_tube_feeding + pneumonia + SDH + ICH + IVH + Focal_small +
multifocal_large + Chronic_old + Temporal + BG + Coil +
craniotomy +
smoke_yn + drink_yn + ALBUMIN + BUN + GLUCOSE_AC + HB + HBA1C +
jaewon_GW + total_cowork + pum_cowork
Res.Df      RSS  Df Sum of Sq      F Pr(>F)
1   2838 2522457340
2   2910 2600201291 -72 -77743951 1.2149 0.1071

# DEATH_YN

result=lm(DEATH_YN~ipwon_period+AGE+SEX+SIDO_NM+CITY_SIZE+GAIBJA_
TYPE+CMPR_DSB_GRADE+MAIN_DSB_TYPE+CMPR_DSB_GRADE_L+MAIN_DSB_TYPE_
L+L_tube_feeding+pneumonia+Infarction+SAH+SDH+ICH+IVH+Focal_small
+multifocal_large+Acute_recent+Chronic_old+Temporal+Cerebellum+BG
+brain_stem+Thalamus+MCA+Aneurysm+Coil+craniotomy+smoke_yn+drink_
yn+ALBUMIN+BUN+CR+GLUCOSE_AC+HB+HBA1C+jaewon_GW+total_cowork+pum_
cowork, data=Data_final)

Call:
lm(formula = DEATH_YN ~ ipwon_period + AGE + SEX + SIDO_NM +
    CITY_SIZE + GAIBJA_TYPE + CMPR_DSB_GRADE + MAIN_DSB_TYPE +
    CMPR_DSB_GRADE_L + MAIN_DSB_TYPE_L + L_tube_feeding +
pneumonia +
    Infarction + SAH + SDH + ICH + IVH + Focal_small +
multifocal_large +
    Acute_recent + Chronic_old + Temporal + Cerebellum + BG +
    brain_stem + Thalamus + MCA + Aneurysm + Coil + craniotomy +
    smoke_yn + drink_yn + ALBUMIN + BUN + CR + GLUCOSE_AC + HB +
    HBA1C + jaewon_GW + total_cowork + pum_cowork, data =
Data_final)

```

Residuals:

Min	1Q	Median	3Q	Max
-1.0634	-0.1918	-0.0543	0.1166	1.1271

Coefficients: (6 not defined because of singularities)

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	1.5352458	0.1737819	8.834	< 2e-16 ***
ipwon_period	-0.0005202	0.0005036	-1.033	0.30169
AGE	0.0060581	0.0005538	10.938	< 2e-16 ***
SEX2	-0.0661767	0.0160103	-4.133	3.68e-05 ***
SIDO_NM2	0.0160869	0.1090548	0.148	0.88274
SIDO_NM3	0.0164214	0.1210326	0.136	0.89209
SIDO_NM4	-0.0290219	0.0419727	-0.691	0.48934
SIDO_NM5	0.1332981	0.1198540	1.112	0.26616
SIDO_NM6	-0.0770637	0.1302944	-0.591	0.55426
SIDO_NM7	-0.0559158	0.3141714	-0.178	0.85875
SIDO_NM8	0.8130156	0.3142541	2.587	0.00973 **
SIDO_NM9	-0.0111717	0.0175021	-0.638	0.52333
SIDO_NM10	-0.0142540	0.0697122	-0.204	0.83800
SIDO_NM11	0.1008507	0.1061098	0.950	0.34197
SIDO_NM12	0.1341820	0.0963652	1.392	0.16390
SIDO_NM13	0.0799460	0.0948175	0.843	0.39921
SIDO_NM14	0.1458284	0.0808215	1.804	0.07129 .
SIDO_NM15	0.0076005	0.0821823	0.092	0.92632
SIDO_NM16	-0.0039015	0.1159875	-0.034	0.97317
SIDO_NM17	-0.2567636	0.2031801	-1.264	0.20643
CITY_SIZE2	0.0551849	0.0218046	2.531	0.01143 *
CITY_SIZE3	-0.0205324	0.0725093	-0.283	0.77707
CITY_SIZE4	NA	NA	NA	NA
CITY_SIZE5	0.0111069	0.0182274	0.609	0.54234
CITY_SIZE6	0.0253896	0.0166719	1.523	0.12790
CITY_SIZE7	0.0495696	0.1532630	0.323	0.74639
CITY_SIZE8	-0.1159594	0.1745120	-0.664	0.50644
GAIBJA_TYPE1	-0.1097810	0.1521921	-0.721	0.47076
GAIBJA_TYPE2	-0.1182240	0.1521450	-0.777	0.43720

GAIBJA_TYPE3	-0.1778537	0.1525411	-1.166	0.24374
GAIBJA_TYPE4	-0.1490888	0.1526009	-0.977	0.32866
GAIBJA_TYPE5	-0.1220548	0.1520309	-0.803	0.42214
GAIBJA_TYPE6	-0.1637090	0.1520231	-1.077	0.28163
GAIBJA_TYPE7	-0.1465397	0.1529953	-0.958	0.33824
GAIBJA_TYPE8	-0.1561429	0.1519166	-1.028	0.30412
GAIBJA_TYPE9	-0.1090312	0.1525242	-0.715	0.47476
GAIBJA_TYPE10	-0.1049850	0.1526160	-0.688	0.49157
GAIBJA_TYPE11	-0.1115845	0.1525695	-0.731	0.46461
GAIBJA_TYPE12	-0.1267056	0.1515247	-0.836	0.40311
GAIBJA_TYPE13	-0.1482625	0.1521157	-0.975	0.32981
GAIBJA_TYPE14	-0.1066787	0.1520687	-0.702	0.48304
GAIBJA_TYPE15	-0.1073321	0.1518110	-0.707	0.47962
GAIBJA_TYPE16	-0.1535327	0.1513237	-1.015	0.31038
GAIBJA_TYPE17	-0.1537364	0.1512208	-1.017	0.30941
GAIBJA_TYPE18	-0.1826924	0.1508287	-1.211	0.22590
GAIBJA_TYPE19	-0.1273864	0.1509048	-0.844	0.39866
GAIBJA_TYPE20	-0.1641510	0.1504899	-1.091	0.27546
CMPR_DSB_GRADE1	0.3486189	0.3315744	1.051	0.29316
CMPR_DSB_GRADE2	0.1307750	0.3291501	0.397	0.69117
CMPR_DSB_GRADE3	0.1941716	0.3307119	0.587	0.55716
CMPR_DSB_GRADE4	0.0691715	0.3245187	0.213	0.83122
CMPR_DSB_GRADE5	-0.0097080	0.3283861	-0.030	0.97642
CMPR_DSB_GRADE6	-0.0305502	0.3260551	-0.094	0.92536
MAIN_DSB_TYPE1	-0.7246948	0.3287494	-2.204	0.02758 *
MAIN_DSB_TYPE2	-0.7507092	0.3207277	-2.341	0.01932 *
MAIN_DSB_TYPE3	-0.6210021	0.3406293	-1.823	0.06839 .
MAIN_DSB_TYPE4	-0.6919098	0.3502611	-1.975	0.04832 *
MAIN_DSB_TYPE5	-0.7102458	0.3387068	-2.097	0.03609 *
MAIN_DSB_TYPE6	-0.5588425	0.3541252	-1.578	0.11466
MAIN_DSB_TYPE8	-0.0533168	0.3694396	-0.144	0.88526
MAIN_DSB_TYPE9	-0.7022864	0.3524377	-1.993	0.04639 *
MAIN_DSB_TYPE10	-0.4388342	0.4741646	-0.925	0.35479
MAIN_DSB_TYPE11	-0.2857190	0.4100245	-0.697	0.48596
MAIN_DSB_TYPE12	-0.1075384	0.3899759	-0.276	0.78276



MAIN_DSB_TYPE14	0.3234314	0.3590291	0.901	0.36774	
MAIN_DSB_TYPE15	NA	NA	NA	NA	
CMPR_DSB_GRADE_L1	-0.6412150	0.0762870	-8.405	< 2e-16	***
CMPR_DSB_GRADE_L2	-0.4138457	0.0658978	-6.280	3.90e-10	***
CMPR_DSB_GRADE_L3	-0.0269842	0.0998080	-0.270	0.78690	
CMPR_DSB_GRADE_L4	0.0828406	0.0971159	0.853	0.39373	
CMPR_DSB_GRADE_L5	0.2344642	0.0901870	2.600	0.00938	**
CMPR_DSB_GRADE_L6	NA	NA	NA	NA	
MAIN_DSB_TYPE_L1	0.9986585	0.4592555	2.175	0.02975	*
MAIN_DSB_TYPE_L2	1.0045994	0.4513732	2.226	0.02612	*
MAIN_DSB_TYPE_L3	0.9300549	0.4663984	1.994	0.04623	*
MAIN_DSB_TYPE_L4	0.9092364	0.4728870	1.923	0.05461	.
MAIN_DSB_TYPE_L5	0.8573336	0.4654410	1.842	0.06558	.
MAIN_DSB_TYPE_L6	0.9120381	0.4750709	1.920	0.05498	.
MAIN_DSB_TYPE_L8	NA	NA	NA	NA	
MAIN_DSB_TYPE_L9	1.0241321	0.4723752	2.168	0.03024	*
MAIN_DSB_TYPE_L10	0.4956341	0.5936255	0.835	0.40383	
MAIN_DSB_TYPE_L11	0.6606035	0.5676756	1.164	0.24464	
MAIN_DSB_TYPE_L12	0.8271075	0.5537816	1.494	0.13540	
MAIN_DSB_TYPE_L14	NA	NA	NA	NA	
MAIN_DSB_TYPE_L15	NA	NA	NA	NA	
L_tube_feeding1	0.2212807	0.0206276	10.727	< 2e-16	***
pneumonia1	0.1273601	0.0133689	9.527	< 2e-16	***
Infarction1	-0.0122228	0.0196568	-0.622	0.53412	
SAH1	-0.0106228	0.0243292	-0.437	0.66241	
SDH1	-0.0472920	0.0244103	-1.937	0.05280	.
ICH1	-0.0315587	0.0169921	-1.857	0.06338	.
IVH1	0.1542423	0.0282812	5.454	5.35e-08	***
Focal_small1	-0.0253457	0.0142873	-1.774	0.07617	.
multifocal_large1	0.1244061	0.0164098	7.581	4.61e-14	***
Acute_recent1	0.0245411	0.0154867	1.585	0.11316	
Chronic_old1	-0.0011319	0.0171331	-0.066	0.94733	
Temporal1	-0.0043119	0.0251829	-0.171	0.86406	
Cerebellum1	0.0033030	0.0195581	0.169	0.86590	
BG1	-0.0002034	0.0153103	-0.013	0.98940	

```

brain_stem1      0.0633013  0.0236369  2.678  0.00745 **
  Thalamus1      -0.0175057  0.0213025  -0.822  0.41128
MCA1             0.0051089  0.0153418  0.333  0.73915
Aneurysm1       -0.0405892  0.0175911  -2.307  0.02111 *
  Coil1          -0.1553086  0.0322934  -4.809  1.59e-06 ***
  craniotomy1    -0.0337415  0.0259615  -1.300  0.19382
smoke_yn2        0.0246290  0.0141229  1.744  0.08128 .
smoke_yn3        0.0139132  0.0180868  0.769  0.44181
drink_yn2        -0.0306466  0.0188640  -1.625  0.10436
drink_yn3        -0.0089123  0.0156998  -0.568  0.57030
ALBUMIN          -0.1067363  0.0146361  -7.293  3.92e-13 ***
BUN              0.0013029  0.0008668  1.503  0.13295
CR               0.0122646  0.0082284  1.491  0.13620
GLUCOSE_AC       0.0004966  0.0001519  3.269  0.00109 **
  HB             -0.0205240  0.0038181  -5.375  8.26e-08 ***
  HBA1C          -0.0034725  0.0055561  -0.625  0.53202
jaewon_GW        0.0010886  0.0005718  1.904  0.05703 .
total_cowork     0.0013161  0.0014254  0.923  0.35589
pum_cowork       0.0229780  0.0095228  2.413  0.01589 *

```

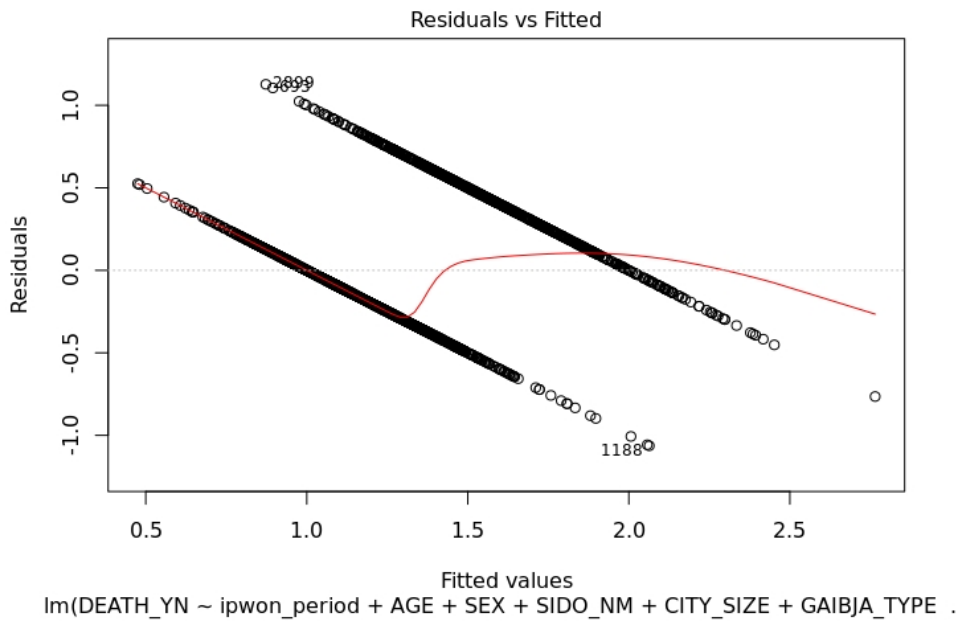
---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

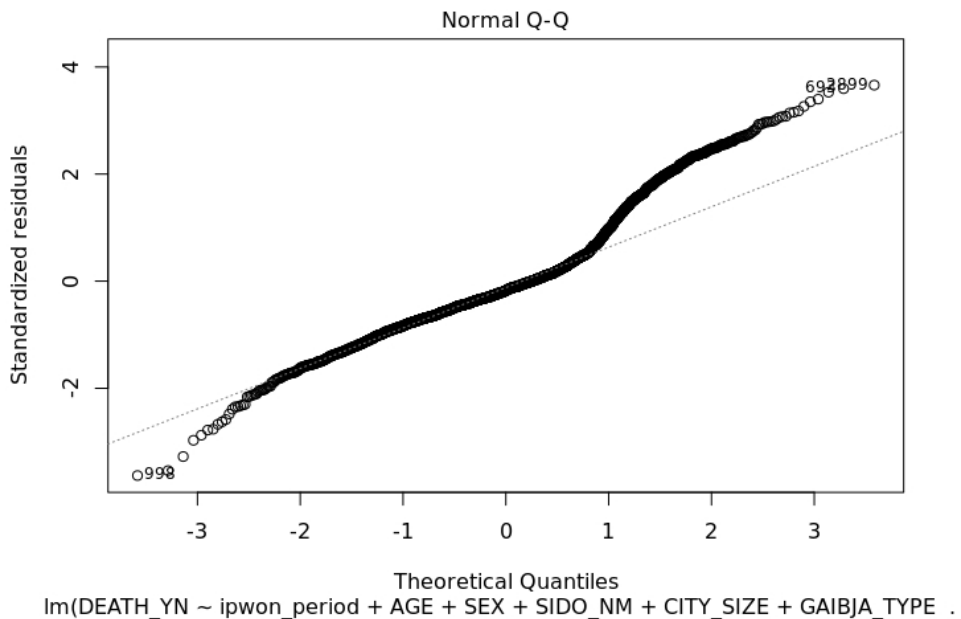
Residual standard error: 0.3114 on 2838 degrees of freedom  
Multiple R-squared: 0.4821, Adjusted R-squared: 0.4618  
F-statistic: 23.8 on 111 and 2838 DF, p-value: < 2.2e-16

```
> plot(result)
```

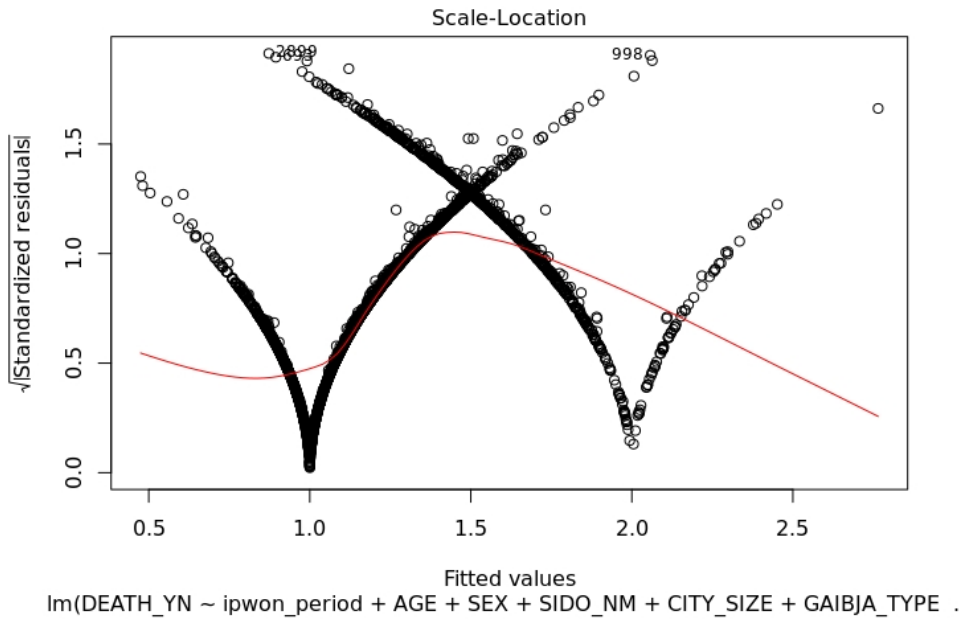
Hit <Return> to see next plot:



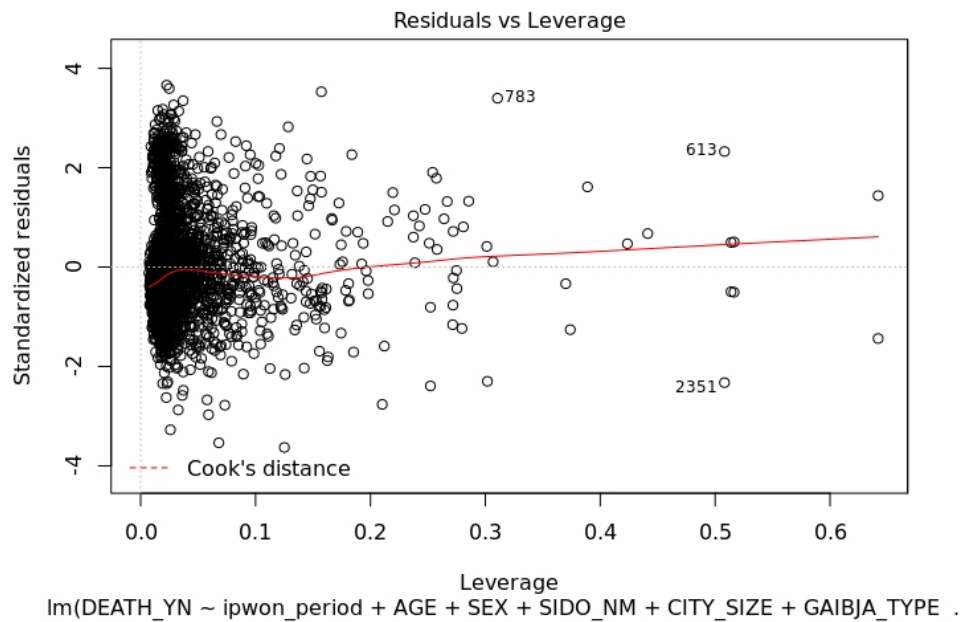
Hit <Return> to see next plot:



Hit <Return> to see next plot:



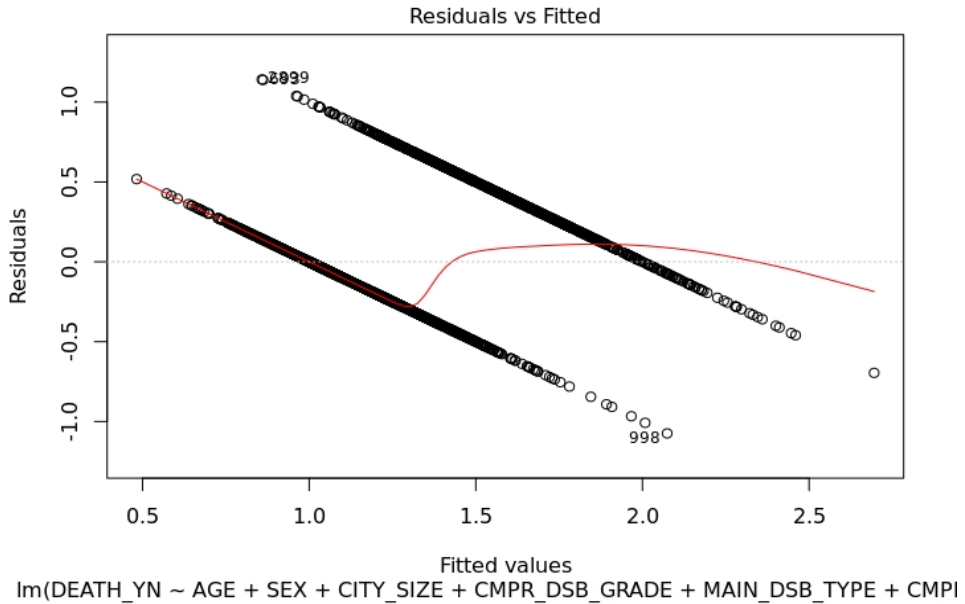
Hit <Return> to see next plot:



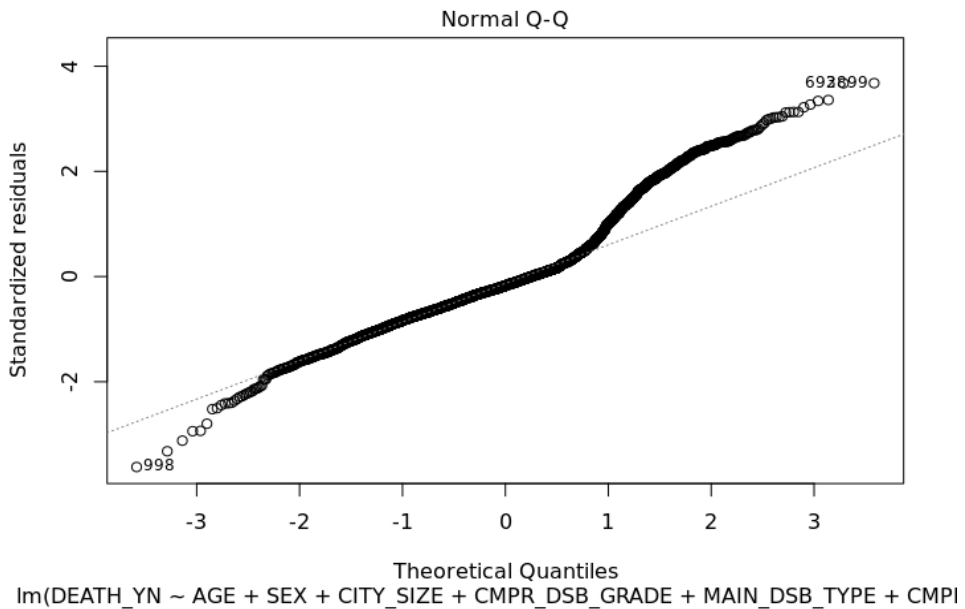
```
result_2=step(result, direction="backward")
```

```
> plot(result_2)
```

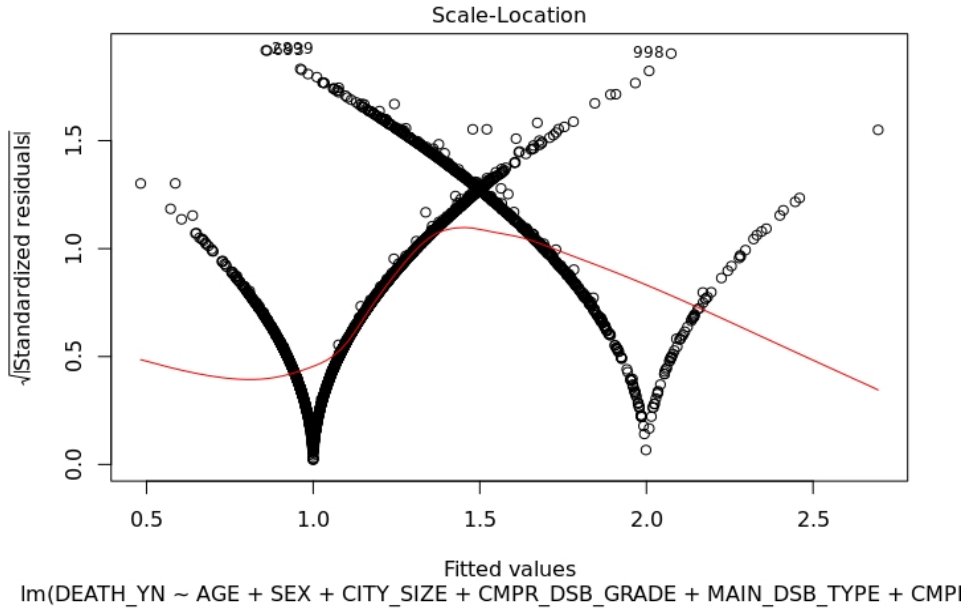
Hit <Return> to see next plot:



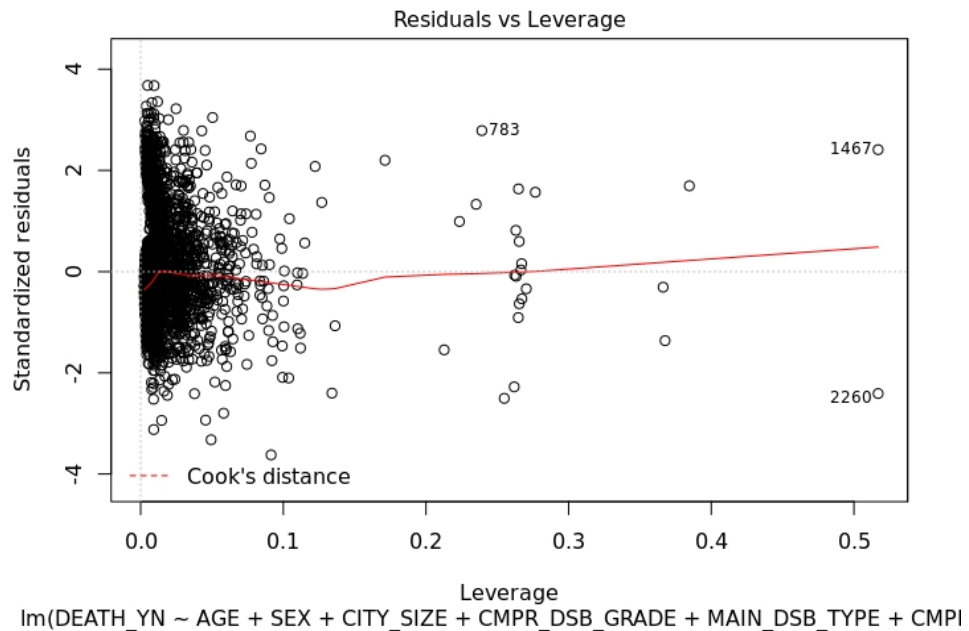
Hit <Return> to see next plot:



Hit <Return> to see next plot:



Hit <Return> to see next plot:



Summary(result\_2)

Call:

```
lm(formula = DEATH_YN ~ AGE + SEX + CITY_SIZE + CMPR_DSB_GRADE
+
    MAIN_DSB_TYPE + CMPR_DSB_GRADE_L + L_tube_feeding +
pneumonia +
    SDH + ICH + IVH + Focal_small + multifocal_large +
Acute_recent +
    brain_stem + Aneurysm + Coil + ALBUMIN + BUN + CR +
GLUCOSE_AC +
    HB + jaewon_GW + pum_cowork, data = Data_final)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.0743	-0.1928	-0.0541	0.1123	1.1418

Coefficients: (2 not defined because of singularities)

Estimate Std. Error t value Pr(>|t|)

(Intercept)	1.3680880	0.0782912	17.474	< 2e-16	***
AGE	0.0061796	0.0005195	11.896	< 2e-16	***
SEX2	-0.0612800	0.0139626	-4.389	1.18e-05	***
CITY_SIZE2	0.0520778	0.0210338	2.476	0.013346	*
CITY_SIZE3	0.0634164	0.0449182	1.412	0.158110	
CITY_SIZE4	0.1005160	0.0502460	2.000	0.045542	*
CITY_SIZE5	0.0131021	0.0177450	0.738	0.460360	
CITY_SIZE6	0.0236277	0.0164407	1.437	0.150786	
CITY_SIZE7	0.1901250	0.0338433	5.618	2.12e-08	***
CITY_SIZE8	0.0082392	0.0898882	0.092	0.926974	
CMPR_DSB_GRADE1	0.8533718	0.2424986	3.519	0.000440	***
CMPR_DSB_GRADE2	0.6446583	0.2379081	2.710	0.006774	**
CMPR_DSB_GRADE3	0.6973487	0.2413028	2.890	0.003882	**
CMPR_DSB_GRADE4	0.5675386	0.2340676	2.425	0.015383	*
CMPR_DSB_GRADE5	0.5070147	0.2334651	2.172	0.029960	*
CMPR_DSB_GRADE6	0.4893608	0.2322049	2.107	0.035164	*
MAIN_DSB_TYPE1	-0.2325934	0.2247912	-1.035	0.300891	

MAIN_DSB_TYPE2	-0.2617657	0.2254521	-1.161	0.245709
MAIN_DSB_TYPE3	-0.1967296	0.2282152	-0.862	0.388739
MAIN_DSB_TYPE4	-0.2847167	0.2268039	-1.255	0.209455
MAIN_DSB_TYPE5	-0.3022788	0.2328855	-1.298	0.194400
MAIN_DSB_TYPE6	-0.1352572	0.2419078	-0.559	0.576118
MAIN_DSB_TYPE8	-0.5492589	0.2908209	-1.889	0.059039.
MAIN_DSB_TYPE9	-0.2093353	0.2339812	-0.895	0.371039
MAIN_DSB_TYPE10	-0.3424211	0.2687771	-1.274	0.202767
MAIN_DSB_TYPE11	-0.0223481	0.2744162	-0.081	0.935099
MAIN_DSB_TYPE12	0.3098591	0.2742072	1.130	0.258562
MAIN_DSB_TYPE14	-0.1718252	0.2750153	-0.625	0.532162
MAIN_DSB_TYPE15	NA	NA	NA	NA
CMPR_DSB_GRADE_L1	-0.6416134	0.0741586	-8.652	< 2e-16 ***
CMPR_DSB_GRADE_L2	-0.4252258	0.0650711	-6.535	7.50e-11 ***
CMPR_DSB_GRADE_L3	-0.0494609	0.0971815	-0.509	0.610824
CMPR_DSB_GRADE_L4	0.0881195	0.0951354	0.926	0.354391
CMPR_DSB_GRADE_L5	0.2163423	0.0888736	2.434	0.014982 *
CMPR_DSB_GRADE_L6	NA	NA	NA	NA
L_tube_feeding1	0.2203344	0.0198854	11.080	< 2e-16 ***
pneumonia1	0.1236117	0.0130919	9.442	< 2e-16 ***
SDH1	-0.0511272	0.0225680	-2.265	0.023557 *
ICH1	-0.0262236	0.0157092	-1.669	0.095162 .
IVH1	0.1521735	0.0271270	5.610	2.22e-08 ***
Focal_small1	-0.0305576	0.0130281	-2.346	0.019068 *
multifocal_large1	0.1269455	0.0160205	7.924	3.25e-15 ***
Acute_recent1	0.0198919	0.0131377	1.514	0.130106
brain_stem1	0.0599641	0.0228015	2.630	0.008588 **
Aneurysm1	-0.0519983	0.0156830	-3.316	0.000926 ***
Coil1	-0.1461262	0.0288353	-5.068	4.28e-07 ***
ALBUMIN	-0.1111647	0.0143398	-7.752	1.24e-14 ***
BUN	0.0012650	0.0008482	1.491	0.135978
CR	0.0155330	0.0075404	2.060	0.039492 *
GLUCOSE_AC	0.0004571	0.0001219	3.749	0.000181 ***
HB	-0.0197426	0.0037176	-5.311	1.18e-07 ***
jaewon_GW	0.0008715	0.0004291	2.031	0.042364 *



```

pum_cowork      0.0266370  0.0081010  3.288 0.001021 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.3111 on 2899 degrees of freedom
Multiple R-squared:  0.4719, Adjusted R-squared:  0.4628
F-statistic:  51.8 on 50 and 2899 DF,  p-value: < 2.2e-16

anova(result, result_2)

Analysis of Variance Table

Model 1: DEATH_YN ~ ipwon_period + AGE + SEX + SIDO_NM +
CITY_SIZE + GAIBJA_TYPE +
  CMPR_DSB_GRADE + MAIN_DSB_TYPE + CMPR_DSB_GRADE_L +
MAIN_DSB_TYPE_L +
  L_tube_feeding + pneumonia + Infarction + SAH + SDH + ICH +
  IVH + Focal_small + multifocal_large + Acute_recent + Chronic_old
+
  Temporal + Cerebellum + BG + brain_stem + Thalamus + MCA +
  Aneurysm + Coil + craniotomy + smoke_yn + drink_yn + ALBUMIN +
  BUN + CR + GLUCOSE_AC + HB + HBA1C + jaewon_GW + total_cowork +
  pum_cowork
Model 2: DEATH_YN ~ AGE + SEX + CITY_SIZE + CMPR_DSB_GRADE +
MAIN_DSB_TYPE +
  CMPR_DSB_GRADE_L + L_tube_feeding + pneumonia + SDH + ICH +
  IVH + Focal_small + multifocal_large + Acute_recent + brain_stem
+
  Aneurysm + Coil + ALBUMIN + BUN + CR + GLUCOSE_AC + HB +
  jaewon_GW + pum_cowork
Res.Df  RSS  Df Sum of Sq    F Pr(>F)
1   2838 275.17
2   2899 280.58 -61   -5.4091 0.9146 0.6636

```

연구보고서 NHIMC-2021-PR-021

## 일산병원과 건강보험공단 자료를 결합한 인공지능 기반 뇌졸중 및 뇌손상 환자의 발병 후 예후 예측 모델

발행일	2023년 5월 10일
발행인	현재룡, 김성우
편집인	이해종, 이천균
발행처	국민건강보험공단 건강보험연구원, 일산병원 연구소
주소	경기도 고양시 일산동구 일산로 100
전화	031) 900-6986, 6987
팩스	0303-3448-7105~7
인쇄처	지성프린팅 (02-2278-2490)



*h·well*  
국민건강보험



(우)10444 경기도 고양시 일산동구 일산로 100(백석1동 1232번지)  
대표전화 1577-0013 / 팩스 031-900-0049  
[www.nhimc.or.kr](http://www.nhimc.or.kr)

## 일산병원과 건강보험공단 자료를 결합한 인공지능 기반 뇌졸중 및 뇌손상 환자의 발병 후 예후 예측 모델

N a t i o n a l   H e a l t h   I n s u r a n c e   S e r v i c e   I l s a n   H o s p i t a l