

랜덤 포리스트를 이용한 비제어 급성 출혈성 쇼크의 흰쥐에서의 생존 예측

최준열^{1,2} · 김성권^{1,3} · 구정모^{1,3} · 김덕원^{1,2,3}

¹연세대학교 의과대학 의학공학교실, ²연세대학교 의과대학 의과학과, ³연세대학교 생체공학협동과정

A Survival Prediction Model of Rats in Uncontrolled Acute Hemorrhagic Shock Using the Random Forest Classifier

J. Y. Choi^{1,2}, S. K. Kim^{1,3}, J. M. Koo^{1,3} and D. W. Kim^{1,2,3}

¹Department of Medical Engineering, Yonsei University College of Medicine

²Brain Korea 21 Project for Medical Science, Yonsei University College of Medicine

³Graduate Program in Biomedical Engineering, Yonsei University

(Received August 20, 2012. Accepted September 10, 2012)

Abstract: Hemorrhagic shock is a primary cause of deaths resulting from injury in the world. Although many studies have tried to diagnose accurately hemorrhagic shock in the early stage, such attempts were not successful due to compensatory mechanisms of humans. The objective of this study was to construct a survival prediction model of rats in acute hemorrhagic shock using a random forest (RF) model. Heart rate (HR), mean arterial pressure (MAP), respiration rate (RR), lactate concentration (LC), and peripheral perfusion (PP) measured in rats were used as input variables for the RF model and its performance was compared with that of a logistic regression (LR) model. Before constructing the models, we performed 5-fold cross validation for RF variable selection, and forward stepwise variable selection for the LR model to examine which variables were important for the models. For the LR model, sensitivity, specificity, accuracy, and area under the receiver operating characteristic curve (ROC-AUC) were 0.83, 0.95, 0.88, and 0.96, respectively. For the RF models, sensitivity, specificity, accuracy, and AUC were 0.97, 0.95, 0.96, and 0.99, respectively. In conclusion, the RF model was superior to the LR model for survival prediction in the rat model.

Key words: hemorrhagic shock, random forest, logistic regression, variable selection, survival prediction

1. 서 론

2004년 외상으로 인한 사망자는 전 세계적으로 약 500만 명으로 보고되었으며, 2020년에는 800만 명까지 증가할 것이라 예측하였다[1,2]. 이러한 외상으로 인한 사망자 가운데 1/3의 직접적인 원인은 출혈성 쇼크이다[2]. 한국의 경우, 최근 8년 동안 응급실에서 다발성 장기손상으로 사망한 환자들 중 74%는 저혈량성 쇼크가 원인이다[3]. 쇼크란 조직에

필요한 산소 요구량과 산소 공급 간의 불균형에 의해 유발되는 임상 증후군을 말하며 이러한 조직의 산소 불균형 상태는 모든 형태의 쇼크에서의 근본적인 원인이라 할 수 있다[4].

현재 임상에서는 출혈성 쇼크를 진단하기 위해 환자의 심박수(heart rate, HR), 혈압(blood pressure), 호흡수(respiration rate, RR), 소변량(urine output), 의식상태(mental status) 등을 확인한다. 그러나 이는 보상기전으로 인하여 실제 환자의 상태를 조기에 정확하게 진단하는데 어려움이 있다[5,6]. 한편, 최근에는 조직세포 허혈(tissue ischemia)을 나타내는 젖산 농도(lactate concentration, LC)와 미세순환을 보여주는 관류(peripheral perfusion, PP) 측정을 통해 출혈성 쇼크를 진단하는 연구가 발표되었다[5,7]. 그러나 젖산 농도와 관류를 이용한 출혈성 쇼크의 조기 진단에 있

Corresponding Author : 김덕원

서울특별시 서대문구 연세로 50 연세대학교 의과대학 의학공학교실
TEL: +82-2-2228-1916 / FAX: +82-2-364-1572

E-mail: kdw@yuhs.ac

연세대학교 의과대학 2011년도 교내연구비에 의하여 이루어 졌음
(6-2011-0087).

어 기존의 생리학적 지표에 비해 우수하지 않다는 보고도 있다[8]. 따라서 출혈성 쇼크의 중증도를 진단하는데 가장 영향력 있는 지표를 보다 객관적으로 연구할 필요성이 있으며 하나의 생체 신호에 의존하기보다 다양한 생체징후에 대한 복합적인 분석이 필요하다.

최근 기계학습을 이용하여 수학적으로 출혈성 쇼크의 중증도를 진단하기 위해 다양한 시도를 하였다. Kim 등은 제어출혈(volume controlled hemorrhage)을 일으킨 흰쥐의 심박수, 혈압, 호흡수, 직장체온(rectal temperature)을 측정하여 로지스틱 회귀분석(logistic regression) 모델을 이용한 생존예측 모델을 제시하였고, 높은 정확도를 보여주었다[9]. Jang 등은 제어 출혈을 일으킨 흰쥐에서 5 fold cross validation을 이용한 인공신경망(artificial neural network)과 지지벡터 기계(support vector machine)의 생존예측 모델들을 제시하여 모델의 신뢰도를 높였다[10]. 그러나 현재 기계학습을 이용한 대부분의 선행연구들은 제어출혈을 일으킨 동물 모델에서 행해졌으며 임상과의 근접성을 향상시키기 위해서는 비제어 출혈(uncontrolled hemorrhage)을 일으킨 모델에 관한 연구가 필요하다. 또한 변수 선정 과정(variable selection)을 거치지 않아 최적화 된 예측모델이라 하기 어려우며, 분류기 사용에 있어, 복잡한 추가설정이 필요한 단점이 있다. 한편 Breiman에 의해 최근에 개발된 랜덤 포리스트(random forest, RF) 분류기는 기존의 다른 분류기에 비해 과잉접합(overfitting)을 하지 않는 장점이 있으며 손쉬운 사용법과 높은 분류 정확도를 보여준다. 또한 이진 분류뿐만 아니라 다중 분류도 가능하여 세분화된 질병의 감별 진단의 도구로서 제시되고 있다[11-13].

본 연구에서는 흰쥐를 대상으로 비제어 출혈을 유도한 후 심박수, 평균 동맥압(mean arterial pressure, MAP), 호흡수, 젖산 농도 및 관류를 측정된 값을 바탕으로, 랜덤 포리스트 분류기를 이용한 생존 예측 모델을 제시하였다. 랜덤 포리스트 분류기의 객관적인 성능 평가를 위해 임상에서 분류기 평가의 표준 방법인 로지스틱 회귀분석 모델을 함께 구축하여 평가하였다. 생존 예측 모델 구축에 앞서 변수 선정 과정을 거쳐 가장 영향력 있는 변수를 선정한 후 최적화 된 모델을 제시하였다.

II. 본 론

1. 실험동물준비

본 실험에서는 체중 300-360 g의 수컷 흰쥐(Sprague-Dawley rats) 36 마리를 사용하였다. 개체 36마리를 임의의 세 군으로 나누고, 15분 동안 출혈량을 체중 100 g당 2.0, 2.5, 3.0 mL로 하여, 각 군당 12마리씩 배정하였다. 이는 각각 전체 혈액량 중 32.8%, 41.0%, 49.2%의 출혈에 해당되

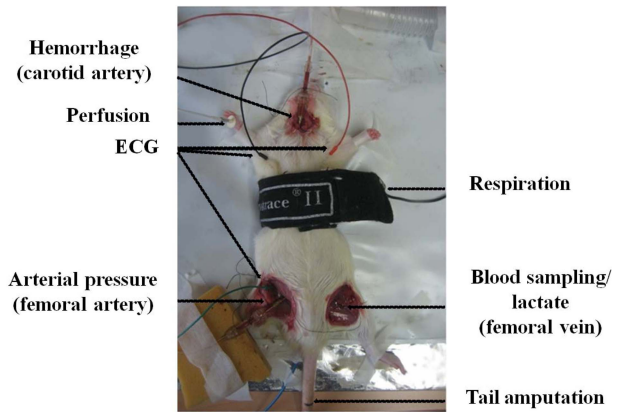


그림 1. 실험 장면의 사진

Fig. 1. Photo of experimental setup

는 양이다[14]. 배정된 흰쥐는 Isoflurane 전용 호흡 마취기(RC2, VetEquip, Pleasanton, CA)로 마취 후 앙와위(supine position)로 고정하였다. 그림 1과 같이 제어 출혈, 동맥혈압 측정 및 채혈을 위해 오른쪽 목의 측면과 양측 서혜부를 절개 하였다. 제어 출혈을 위해 우측 경동맥에 24 gauge 카테터(Becton Dickinson, Franklin Lakes, NJ)를 삽입하였으며, 출혈 유도 펌프(Pump 11 plus, Harvard Apparatus, South Natick, MA)에 연결하였다. 동맥 혈압 측정을 위해 우측 대퇴 동맥에 24 gauge 카테터를 삽입한 후, 혈압 센서(1620 Pressure®, MSI Sensors, Australia)에 연결하였다. 좌측 대퇴 정맥은 젖산 농도 측정을 위한 채혈을 위해 노출하였다.

2. 생체신호 측정

심전도, 평균 동맥압, 호흡은 아날로그-디지털 신호 변환기(PowerLab 8/30, AD Instruments, Colorado Springs, CO)로 입력 받아 컴퓨터 소프트웨어(LabChart 6 Pro, AD Instruments)로 분석하였다. 모든 신호는 1 kHz의 샘플링 주파수로 측정하였다. 심전도는 침습적인 바늘 전극으로 측정하여 생체신호증폭기(Animal bio amp®, ML136, AD Instruments)로 증폭하였다. 평균 동맥압 측정을 위한 혈압센서는 자체 제작한 증폭기와 연결하였다. 실험 전 혈압센서와 증폭기를 통해 출력되는 신호를 수은 혈압계를 이용하여 보정하였다. 호흡수는 흉곽에 호흡벨트(MLT1132, AD Instrument)를 이용하여 흉곽의 단면적 변화를 이용하여 측정하였다.

관류는 laser Doppler perfusion monitor (LDPM) 장비인 PeriFlux system 5000 (Perimed, Sweden)과 측정 프로브(Probe 407, Perimed)를 사용하여 흰쥐의 오른쪽 앞발바닥에서 32 Hz의 샘플링 주파수로 측정하였다. LDPM 기기는 780 nm 파장의 레이저 빛을 이용하여 피하 0.5~1.0 mm의 미세혈류에 반사되는 혈액의 적혈구의 개수와 속도

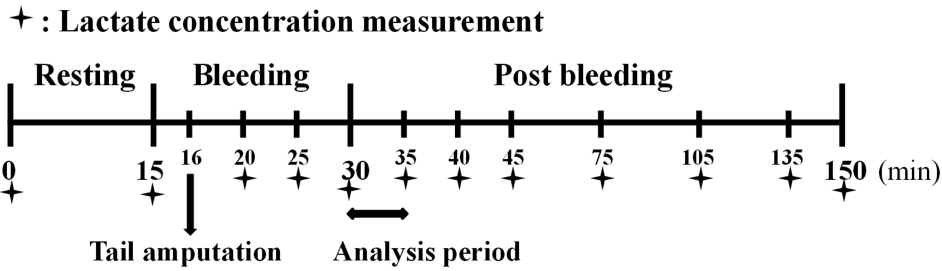


그림 2. 실험 과정 및 분석 구간
Fig. 2. Experimental protocol and data analysis period

의 곱에 비례하는 관류로 표현된다[15]. 측정된 신호는 컴퓨터 소프트웨어(Perisoft for window, Perimed)로 저장 및 분석하였다.

채혈은 좌측 대퇴정맥에서 31 gauge의 인슐린 주사기 (BD Ultra-Fine II, Becton Dickinson)를 삽입하여 반복 하였다. 채혈된 혈액은 휴대용 젖산 농도 측정기(Lactate Pro LT-1710, ARKRAY, Japan)로 젖산 농도를 측정하였다. 측정기의 측정 범위는 0.8~23.3 mmol/L이다. 1회 측정 당 필요 혈액은 5 μ L이며, 분석 시간은 60초가 소요되었다. 그림 2와 같이 젖산 농도 측정은 실험 시작 시점인 0분, 실험 시작 후 15분에서 45분 구간에서는 5분 간격으로, 그 후로는 30분 간격으로 측정하였고, 실험 종료 시점인 150분 혹은 사망시점에서 측정하였다. 연속적인 젖산 농도 값을 계산하기 위하여 선형 보간법을 사용하였으며 컴퓨터 소프트웨어 (LabVIEW 2009, National Instruments, Austin, TX)로 구현하였다[16].

3. 실험 과정

동물실험을 위한 준비 후 그림 2와 같이 15분간의 안정기를 가졌다. 안정기 후 우측 경동맥에서 출혈 펌프를 통해 각 군 당 15분 동안 체중 100 g당 2.0, 2.5, 3.0 mL의 제어 출혈을 유도하였다. 제어 출혈이 시작한 1분 후 인 16분에서 꼬리 길이의 75%를 절단하여 비제어 출혈을 유도하였다. 꼬리의 절단면은 응고방지를 위해 헤파린 나트륨과 생리 식염수가 혼합된 실린지에 담갔으며, 무게 변화를 통해 꼬리에서 출혈량을 계산하였다. 제어 출혈 종료 시점인 30분에 꼬리 절단면을 봉대를 이용하여 지혈을 하였다. 치명적인 출혈성 쇼크 상태인 환자 중 절반 이상은 외상 후 120분 이내에 발생한다고 하여[17], 실험은 출혈 종료 후 120분까지 진행하였으며, 사망은 평균 동맥압이 0 mmHg로 측정되는 시점으로 규정하였다[18].

4. 데이터 분석

그림 2와 같이 데이터는 출혈 종료 후 5분 동안 분석하였다. 이는 응급실에서 출혈 환자의 지혈 과정 및 지혈 후

표 1. 생존 여부와 훈련 및 평가 데이터 세트 수
Table 1. Number of survival and death data sets with training and testing

Group	Training set	Testing set	Total set
Survival set	54	21	75
Death set	71	29	100
Total set	125	50	175

환자 관찰 등 실제 응급 상황을 모사하기 위함이었다[10]. 생존예측 모델을 위한 입력 변수로는 심박수, 평균 동맥압, 호흡수, 젖산 농도 및 관류이며 모든 데이터는 1분씩 구간을 나누어 평균값을 구하였다. 각 분당 심박수, 평균 동맥압, 호흡수, 젖산 농도 및 관류를 하나의 세트로 볼 때 지혈 후 5분간의 데이터를 분석하였으므로 각 쥐 당 5개의 데이터 세트를 분석하였다. 따라서 총 데이터 세트의 개수는 표 1과 같이 한 마리(분석구간 중 사망)를 제외한 총 35마리의 데이터인 175(1 set/min \times 5 min \times 35 rats)개 이다. 이 중 75개의 데이터 세트는 생존군이며 100개의 데이터 세트는 사망군에 해당된다. 랜덤 포리스트 분류기를 위한 데이터 분석은 MATLAB Version 2011 (Mathworks Inc, Natick, MA)을 이용하였다.

(1) 로지스틱 회귀 분석

로지스틱 회귀 분석은 이분화 된 데이터에 대하여 결과를 예측하는 함수식을 생성한다. 식 (1)은 회귀 함수식이며 P(x) 값은 0과 1사이의 값을 가지며 임의의 경계를 기준으로 이분화 된 데이터의 결과를 예측한다.

$$P(x) = \frac{1}{1 + e^{-(\alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \dots + \beta_n x_n)}} \quad (1)$$

식 (1)에서 $\beta_1 \sim \beta_n$ 는 입력 변수($x_1 \sim x_n$)의 회귀계수이며, $e^{\beta_1} \sim e^{\beta_n}$ 값은 변수의 위험도를 나타낸다. 일반적으로 로지스틱 회귀 분석은 표준 방법으로서 질병의 진단으로 사용되는 분류기의 성능 평가에 사용된다[12]. 본 연구에서는 출혈성 쇼크의 생존 예측 모델을 새롭게 제안하는 랜덤 포리스트의

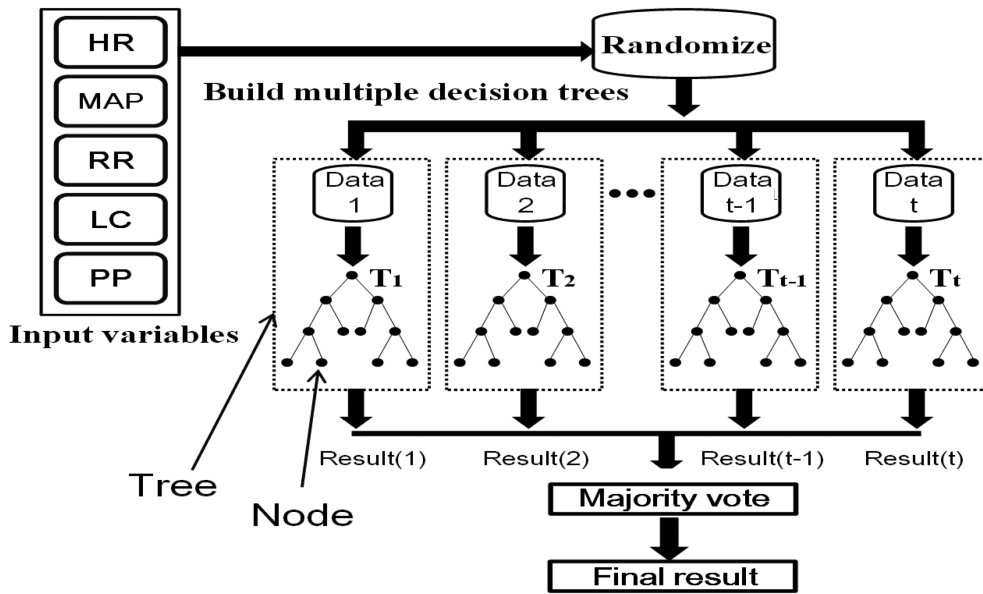


그림 3. 랜덤 포리스트 예측 모델의 입력 변수와 알고리즘

Fig. 3. Algorithm of the random forest model and their variables. HR: heart rate, MAP: mean arterial pressure, RR: respiration rate, LC: lactate concentration, PP: peripheral perfusion

성능을 로지스틱 회귀 분석모델과 비교하여 평가하였다.

(2) 랜덤 포리스트

랜덤 포리스트는 앙상블 기반의 분류 알고리즘으로, 그림 3과 같이 다수의 독립적인 의사 결정 나무들로 구성된다. $Data_1 \sim Data_{t-1}$ 는 의사 결정 나무($T_1 \sim T_t$)를 생성하기 위한 데이터 집합으로 입력 변수로부터 무작위로 배정받는 bagging idea를 이용한다. 배정받은 데이터는 랜덤으로 노드(node) 안에 들어가 조건을 형성하는데 그 조건의 여부에 따라 가치를 확장하여 다음 노드를 형성한다. 이러한 과정이 반복되어 한 개의 의사 결정 나무가 완성되며 사용자가 지정한 나무의 수만큼 독립적인 나무를 형성한다. 랜덤 포리스트의 최종 분류 결과는 다수결 원칙에 의거하여 각 독립적인 나무들의 결과를 종합하여 결정된다[12].

랜덤 포리스트의 단점은 생성된 의사 결정 나무의 내부를 관찰 할 수 없어 예측 결과 이외에 정보를 얻을 수 없다는 것이다[19]. 그러나 모델 생성 시 나무의 수와 노드 안에 들어가는 조건의 수만 결정하므로 비교적 사용하기에 간편하다는 장점이 있다[12]. 본 연구에서는 현재 가장 많이 사용되고 있는 나무의 수(10, 100, 200, 300, 400, 500)를 변경하며 랜덤 포리스트 모델을 생성하였다[12,20]. 노드 조건의 수는 입력변수의 개수에 따라 자동적으로 1개로 고정되었다[11].

5. 입력 변수 선정 및 모델 생성

생존 예측 모델 생성 및 성능 평가를 위하여 표 1과 같이 랜덤하게 전체 데이터 세트의 71%($n = 125$)를 훈련 데

이터 세트(training set)로, 29%($n = 50$)를 평가 데이터 세트(testing set)로 각각 나누었다. 그림 4와 같이 모델의 최적화를 위한 입력 변수 선정 과정은 훈련 데이터 세트 안에서 이루어지며, 선정된 변수를 갖는 최종 모델은 평가 데이터 세트에서 검증된다.

심박수, 평균 동맥압, 호흡수, 젖산 농도, 관류 모두 랜덤 포리스트의 변수 선정 과정의 후보로 사용되었다. 랜덤 포리스트의 변수 선정은 변수 간의 생존 사망에 대한 우선순위를 결정한 후에 가장 중요하지 않은 변수부터 하나씩 제거해가며 모델의 정확도를 비교하는 후진 제거법(backward elimination method)을 이용하였다. 최종 선택된 변수는 가장 적은 개수일 때 가장 높은 정확도를 보여주는 변수들이다. 변수 선정을 위한 우선순위 결정은 Breiman의 방법을 이용하였으며[11] 모델들의 평균 정확도는 5 fold cross validation을 이용하여 신뢰도를 높였다.

로지스틱 회귀 분석의 변수 선정 과정은 전진 선택법(forward stepwise method)을 이용하였다. 전진 선택법은 변수를 하나씩 회귀 방정식에 삽입해가며 생존 사망에 대한 통계적 유의성을 검증한 후 유의한 차이가 있으면 방정식에 그 변수를 포함하는 방법이다. 이 과정은 더 이상 통계적으로 유의한 변수가 없을 때까지 진행 된다. 로지스틱 회귀 분석의 변수 선정은 SPSS 18.0 (SPSS Inc, Chicago, IL)을 이용하였다.

본 연구에서는 평가 데이터 세트를 이용하여 선택된 변수들을 입력변수로 하는 생존 예측 모델들의 민감도(sensitivity), 특이도(specificity), 정확도(accuracy) 및 컴퓨터 수행 시간을

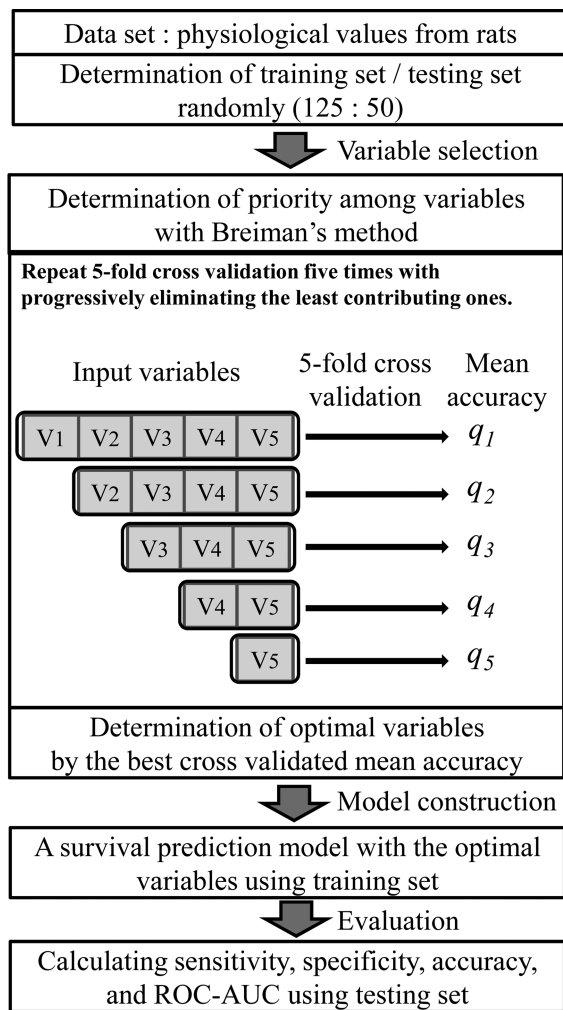


그림 4. 랜덤 포리스트 후진제거 변수 선정 과정 및 생존 예측 모델 생성의 흐름도

Fig. 4. Flowchart of backward elimination variable selection and survival prediction model for the random forest

계산하였다. 또한 랜덤 포리스트와 로지스틱 회귀 분석 모델들의 ROC-AUC (area under the receiver operating characteristic curve)를 계산하여 모델의 성능을 비교하였다. ROC-AUC는 MATLAB 소프트웨어를 이용하여 계산하였다.

III. 결 과

Brieman의 방법을 이용한 랜덤 포리스트의 우선순위는 차례로 관류, 호흡수, 평균 동맥압, 심박수, 젖산 농도이다. 표 2는 변수 선정 과정을 통해 선정된 변수들이다. 랜덤 포리스트와 로지스틱 회귀 분석 모두 호흡수와 관류를 최종 변수로 선정하였다. 표 3은 나무 수에 따른 랜덤 포리스트와 로지스틱 회귀 분석 모델의 AUC, 민감도, 특이도, 정확도 및 수행 시간이다. 랜덤 포리스트 모델들의 AUC, 민감

표 2. 랜덤 포리스트와 로지스틱 회귀 분석에서 선택된 변수
Table 2. Selected variables for the random forest and logistic regression models

Variables	RF						LR
	The number of trees						Forward
	10	100	200	300	400	500	
HR							
MAP							
RR	*	*	*	*	*	*	*
LC							
PP	*	*	*	*	*	*	*

RF: random forest, LR: logistic regression, HR: heart rate, MAP: mean arterial pressure, RR: respiration rate, LC: lactate concentration, PP: peripheral perfusion

표 3. 랜덤 포리스트와 로지스틱 회귀 분석에서의 생존 예측 모델의 성능
Table 3. Performance of the random forest and logistic regression survival prediction models

	RF						LR
	The number of trees						Forward
	10	100	200	300	400	500	
ROC-AUC	0.99	0.99	0.99	0.99	0.99	0.99	0.96
Sensitivity	0.97	0.97	0.97	0.97	0.97	0.97	0.83
Specificity	0.95	0.95	0.95	0.95	0.95	0.95	0.95
Accuracy	0.96	0.96	0.96	0.96	0.96	0.96	0.88
Time (sec)	0.06	0.07	0.08	0.08	0.09	0.10	1.38

RF: random forest, LR: logistic regression
ROC-AUC: area under the receiver operating characteristic curve

도, 특이도, 정확도는 0.99, 0.97, 0.95, 0.96이었으며 나무 수와 관계없이 동일한 결과를 보여주었다. 수행 시간은 나무의 수가 늘어남에 따라 증가하였다. 로지스틱 회귀 분석 모델의 AUC, 민감도, 특이도, 정확도는 0.96, 0.83, 0.95, 0.88이었으며, 수행 시간은 1.38초였다. 모든 랜덤 포리스트 모델들이 로지스틱 회귀 분석 모델에 비해 좋은 성능과 빠른 수행 시간을 보였다. 그림 5는 나무 수 100개인 랜덤 포리스트와 로지스틱 회귀 분석 모델의 ROC 커브이다.

IV. 결 론

본 연구에서는 급성 출혈성 쇼크의 흰쥐에서의 생존 예측 모델을 기존의 임상에서 가장 많이 사용하고 있는 로지스틱 회귀 분석과 새롭게 개발된 랜덤 포리스트를 통해 제시하였고 변수 선정 과정을 통해 모델을 최적화하였다. 그 결과 랜덤 포리스트 모델이 로지스틱 회귀 분석 모델보다 우수한 성능 및 빠른 수행 시간을 보여주었다. 현재 랜덤 포리스트를 이용한 출혈성 쇼크의 생존예측 모델은 연구된 바 없다. 따

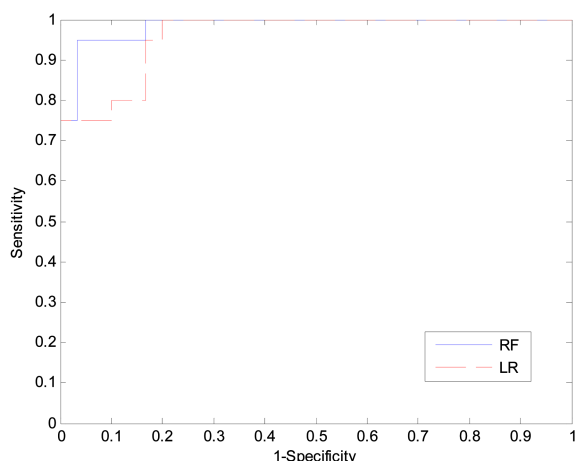


그림 5. 랜덤 포리스트(나무 수 = 100)와 로지스틱 회귀분석 생존 예측 모델의 ROC 커브

Fig. 5. ROC curves of random forest with 100 trees and logistic regression for survival prediction model

라서 본 연구는 출혈성 쇼크의 생존예측에 있어 랜덤 포리스트의 가능성을 보여주었다. Archer 등은 랜덤 포리스트 모델 생성 시 예측 결과의 안정성을 위해 나무의 수를 충분히 크게 설정 하여야 한다고 보고한 반면[21,22], Chen 등은 더 많은 나무의 수가 더 높은 성능을 보장하지 않는다고 보고하였다[20]. 본 연구에서는 나무의 수가 모델의 성능에 영향을 끼치지 않는 것을 확인하였다.

랜덤 포리스트 모델들의 컴퓨터 수행 시간은 로지스틱 회귀 분석모델 보다 약 10배 이상 빠른 것을 확인하였다. Albert 등의 연구에서는 10,000개의 데이터 세트에서 랜덤 포리스트 모델을 생성하는데 1분밖에 소요 되지 않았다[22]. 이러한 빠른 데이터 처리 속도는 추 후 생물정보학(bioinformatics)과 같은 방대한 자료를 처리하고 평가하는 데 매우 중요한 역할을 할 것이라 사료된다.

변수 선정 과정은 모델을 효율적으로 최적화 하는데 매우 중요하다. 본 연구에서는 심박수, 평균 동맥압, 호흡수, 젖산 농도 및 관류를 모델의 초기 변수로 사용하였으나 5 fold cross validation 변수 선정 과정 후 2개의 변수인 호흡수와 관류로 랜덤 포리스트 모델을 최적화하였다. 따라서 변수 선정 과정을 수많은 유전 정보 중 특정 질병에 영향력 있는 유전 정보를 찾는 데 이용한다면 가장 효율적으로 질병을 감별 및 진단할 수 있을 것이다.

본 연구에서 관류는 출혈성 쇼크의 생존 예측에 영향력 있는 변수로 선정되었다. 그러나 관류 측정을 위한 LDF (laser Doppler flowmetry)는 고가이며 동잡음에 취약한 단점이 있다[23]. 따라서 추후 관류 측정 시 고정시킬 수 있는 프로브에 대한 연구가 필요하다. 본 연구는 출혈에 따른 세분화된 중증도 예측이 아닌 생존 예측 모델을 제시하였기 때문에 임상적용에 제한점이 있다. 또한 일반적인 동물 실험

모델은 마취 상태에서 진행이 되며, 이는 심혈관 기능, 대사 기능 등에 영향을 줄 수 있기 때문에, 본 연구 결과를 임상에 바로 적용하기에는 무리가 있다[24]. 따라서 추 후 실제 임상 데이터를 이용한 출혈성 쇼크 환자의 생존 예측 모델 뿐만 아니라 더 나아가 다중 랜덤 포리스트 분류기를 이용한 다양한 출혈에 따른 세분화된 중증도를 예측하는 모델을 구축한다면 다수의 외상환자 발생 시 신속히 쇼크의 경중을 판단하여 위험군 환자들을 우선적으로 치료할 수 있을 것이라 사료된다.

참고문헌

- [1] World Health Organization, *World health statistics 2010*, Geneva, Switzerland: World Health Organization Press, 2010, pp. 62-70.
- [2] C.J. Murray and A.D. Lopez, "Alternative projections of mortality and disability by cause 1990-2020: global burden of disease study," *Lancet*, vol. 349, no. 9064, pp. 1498-1504, 1997.
- [3] C.D. Deakin and I.R. Hicks, "AB or ABC: pre-hospital fluid management in major trauma," *J Accid Emerg Med*, vol. 11, no. 3, pp. 154-157, 1994.
- [4] V.J. Markovchick and P.T. Pons, *Emergency medicine secrets, 4th edition*, Philadelphia, USA: Mosby, 2006, pp. 28-32.
- [5] M. Wilson, D.P. Davis, and R. Coimbra, "Diagnosis and monitoring of hemorrhagic shock during the initial resuscitation of multiple trauma patients: a review," *J Emerg Med*, vol. 24, no. 4, pp. 413-422, 2003.
- [6] H.R. Guly, O. Bouamra, M. Spiers, P. Dark, and T. Coats et al., "Vital signs and estimated blood loss in patients with major trauma: testing the validity of the ATLS classification of hypovolaemic shock," *Resuscitation*, vol. 82, no. 5, pp. 861-867, 2010.
- [7] R.P. Dutton, "Current concepts in hemorrhagic shock," *Anesthesiology clinics*, vol. 25, no. 1, pp. 23-34, 2007.
- [8] G.J. Pestel, K. Fukui, O. Kimberger, H. Hager, and A. Kurz et al., "Hemodynamic parameters change earlier than tissue oxygen tension in hemorrhage," *Journal of Surgical Research*, vol. 160, no. 2, pp. 288-293, 2010.
- [9] D.W. Kim, J.L. Choi, and Y.S. Park, "Survival prediction in rats with fixed-volume hemorrhage using a logistic regression equation," *Shock*, vol. 33, Suppl. 1, pp. 14, 2010.
- [10] K.H. Jang, J.L. Choi, T.K. Yoo, M.K. Kwon, and D.W. Kim, "Survival prediction of rats with hemorrhagic shock using support vector machine," *J of KOSOMBE*, vol. 33, no. 1, 2012.
- [11] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5-32, 2001.
- [12] C.H. Hsieh, R.H. Lu, N.H. Lee, W.T. Chiu, and M.H. Hsu et al., "Novel solutions for an old disease: diagnosis of acute appendicitis with random forest, support vector machines, and artificial neural networks," *Surgery*, vol. 149, no. 1, pp. 87-93, 2010.
- [13] R. Diaz-Uriarte and S. Alvarez de Andres, "Gene selection and classification of microarray data using random forest," *BMC Bioinformatics*, vol. 7, no. 3, pp. 1-13, 2006.
- [14] J.L. Choi, J.Y. Choi, W.H. Lee, M.K. Kwon, I.C. Park, and

- D.W. Kim, "Development of a new index for mortality prediction due to hemorrhagic shock using lactate concentration and perfusion in S-D rats," *J of The Korean Society of Emergency Medicine*, vol. 23, no. 1, 2012.
- [15] M.J. Leahy, F.F. de Mul, G.E. Nilsson, and R. Maniewski, "Principles and practice of the laser-Doppler perfusion technique," *Technol Health Care*, vol. 7, no. 2-3, pp. 145-162, 1999.
- [16] D.A. Roberts, J.B. Holcomb, B.E. Parker, J.L. Sondeen, and A.E. Pusateri et al., "The use of polynomial neural networks for mortality prediction in uncontrolled venous and arterial hemorrhage," *J Trauma*, vol. 52, no. 1, pp. 130-135, 2002.
- [17] L.G. Not, R.B. Marchase, N. Fulop, C.A. Brocks, and J.C. Chantham, "Glucosamine administration improves survival rate after severe hemorrhagic shock combined with trauma in rats," *Shock*, vol. 28, no. 3, pp. 345-352.
- [18] A. Takasu, T. Sakamoto, and Y. Okada, "Effect of induction rate for mild hypothermia on survival time during uncontrolled hemorrhagic shock in rats," *J Trauma*, vol. 61, no. 6, pp. 1330-1335, 2006.
- [19] A.M. Prasad, L.R. Iverson, and A. Liaw, "Newer classification and regression tree techniques: bagging and random forests for ecological prediction," *Ecosystems*, vol. 9, no. 2, pp. 181-199, 2006.
- [20] X.W. Chen and M. Liu, "Prediction of protein-protein interactions using random decision forest framework," *Bioinformatics*, vol. 588, no. 3, pp. 424-432, 2008.
- [21] K.J. Archer and R.V. Kimes, "Empirical characterization of random forest variable importance measures," *Computational Statistics & Data Analysis*, vol. 52, no. 4, pp. 2249-2260, 2008.
- [22] J. Albert, E. Aliu, H. Anderhub, P. Antoranz, and A. Armada et al., "Implementation of the random forest method for the imaging atmospheric cherenkov telescope MAGIC," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 588, no. 3, pp. 424-432, 2008.
- [23] M.L. Kaiser, A.P. Kong, E. Steward, M. Whealon, and M. Patel et al., "Laser Doppler imaging for early detection of hemorrhage," *J Trauma*, vol. 71, no. 2, pp. 401-406, 2011.
- [24] W.J. Meijerink, P.E. Molina, and N.N. Abumrad, "Mammalian opiate alkaloid synthesis: lessons derived from plant biochemistry," *Shock*, vol. 12, no.3, pp. 165-173, 1999.