인공신경망을 이용한 조산 단태아의 분만 임신주수 예측

이지윤¹ · 조수정¹ · 정은진¹ · 이광식² · 김승우³ · 김호연¹
 조금준¹ · 홍순철¹ · 오민정¹ · 김해중¹ · 안기훈¹

고려대학교 의과대학 산부인과교실¹, 고려대학교 안암병원 AI센터², 리스크솔루션 인공지능 연구팀³

Prediction of Gestational Age at Birth using an Artificial Neural Networks in Singleton Preterm Birth

Jee-Yun Lee¹ · Soo-Jung Jo¹ · Eun-Jin Jung¹ · Kwang-Sig Lee² · Seung-Woo Kim³ · Ho Yeon Kim¹
 Geum Joon Cho¹ · Soon-Cheol Hong¹ · Min-Jeong Oh¹ · Hai-Joong Kim¹ · Ki Hoon Ahn¹

*Department of Obstetrics & Gynecology, Korea University College of Medicine¹
 AI Center, Anam Hospital, Korea University College of Medicine²
 Artificial Intelligence Research Team, RiskSolutions³*

Purpose: The objective of the present study was to predict the gestational age at preterm birth using artificial neural networks for singleton pregnancy.

Methods: Artificial neural networks (ANNs) were used as a tool for the prediction of gestational age at birth. ANNs trained using obstetrical data of 125 cases, including 56 preterm and 69 non-preterm deliveries. Using a 36-variable obstetrical input set, gestational weeks at delivery were predicted by 89 cases of training sets, 18 cases of validating sets, and 18 cases of testing sets (total: 125 cases). After training, we validated the model by another 12 cases containing data of preterm deliveries.

Results: To define the accuracy of the developed model, we confirmed the correlation coefficient (R) and mean square error of the model. For validating sets, the correlation coefficient was 0.839, but R of testing sets was 0.892, and R of total 125 cases was 0.959. The neural networks were well trained, and the model predictions were relatively good. Furthermore, the model was validated with another dataset of 12 cases, and the correlation coefficient was 0.709. The error days were 11.58±13.73.

Conclusion: In the present study, we trained the ANNs and developed the predictive model for gestational age at delivery. Although the prediction for gestational age at birth in singleton preterm birth was feasible, further studies with larger data, including detailed risk variables of preterm birth and other obstetrical outcomes, are needed.

Key Words: preterm delivery, artificial neural networks

서 론

교신저자: 안기훈, 서울특별시 성북구 인촌로 73

고려대학교 의과대학 산부인과교실

전화번호: 02-920-6777, 팩스번호: 02-922-5357

E-mail: akh1220@hanmail.net

접수일자: 2018년 8월 1일, 심사일자: 2018년 9월 3일

게재확정일자: 2018년 9월 4일

Copyright © 2018 by The Korean Society of Maternal and Child Health

조산은 임신주수 20~37주 미만에 분만이 진행된 경우를 일컫는 것으로 이로 인한 저체중, 성장 지연 및 신경계 손상 등은 신생아의 주산기 이환율과 관련하여 전 세계적인 문제로 대두되고 있다. 세계보건기구에 따르면 조산의 발생률은 전 세계적으로 5%에서 15%에 이르며 매년 약 1,500만 명의 조산아가 태어나는 것으로 보고 있다. 이들 중 100만 명이 넘는 아

동이 매년 조산과 관련된 합병증으로 사망하며 생존한 조산아들 중에서도 많은 수가 일생에 걸쳐 장애나 시력, 청력과 관련된 후유증을 갖게 된다(Beck et al., 2010).

우리나라의 경우 역시 분만 연령이 증가함에 따라 조산율도 상승하고 있는 추세로 2016년 기준해 조산율은 7.2%로 2006년 4.9%에 비해 2.3% 증가했다(Statistics Korea, 2017). 이처럼 고령임신, 난임으로 인한 인공 수정 등 고위험 임신의 증가와 함께 조산율이 증가함에 따라 현재까지 예측에 관한 많은 연구가 수행되어 왔다. 그러나 조산에 기여하는 많은 요인들로 인해 조산의 예측은 여전히 정확도가 떨어지는 것이 사실이다(Vinken et al., 2009; Lee et al., 2011; Morken et al., 2014; Zhu et al., 2017).

특히 조산이 되는 임신주수가 작아질수록 신생아의 이환율과 사망률이 급격히 증가함은 이미 널리 알려져 있기 때문에 조산이 되는 산모에 있어 분만을 하는 임신주수를 미리 예측할 수 있다면 신생아의 예후를 향상시키면서 사회경제적 비용을 낮출 수 있는 중요한 일이 된다.

인공신경망(Artificial Neural Networks, ANNs) 모형은 현재 의학의 여러 분야에서도 사용되고 있는 것으로 인간의 뇌를 통해 문제를 처리하는 방법과 유사한 방법으로 문제를 해결하기 위해 채택하고 있는 통계학적 학습 알고리즘이다(Lisboa PJG, 2002; Schmidhuber J, 2015). 이러한 인공신경망을 의학적으로 이용한 예로는 ST분절 급상승 심근경색 환자들의 단기 사망률 예측 혹은 제2형 당뇨병의 진단을 향상을 연구한 논문 등이 있다(Lim et al., 2012; Kho & Kim, 2018).

Catley 등은 이러한 인공신경망을 이용해 무증상의 산모들에서 조산의 위험도를 예측하고자 하였다. 8개의 산과적 변수 즉, 분만시 나이, 현재 분만시 태아수, 이전 조산 횟수, 이전 만삭 분만 횟수, 전체 분만한 횟수, 신생아의 성별, 모유수유 여부, 임신 20주 이후 흡연력을 입력하여 조기 분만의 고위험 모델을 학습시켜 민감도가 54.8% 임을 확인하였으나 고위험의 정의가 무엇이었는지 명확하지 않고 임신 33주를 “매우 이른 조산”으로 정의하는 등 현재 산과학적인 정의와는 다른 기준을 이용하였다. 또한 기존 연구에서 밝혀진 조산의 위험인자로 알려져 있는 변수가 적절히 사용되지 않았고 조산 여부만 확인하고자 하였다는 한계점이 있다(Catley et al., 2006).

이에 본 연구는 이전 연구에서 알려져 있는 조산의 위험인자를 확인 후 단태아에 대해 인공신경망을 학습하여 조산의 분만 예정 주수를 예측가능한지 확인하고자 하였다.

연구대상 및 방법

1. 연구설계

본 연구는 2017년 1월부터 2018년 5월 사이에 K병원에서 분만이 완료된 산모를 대상으로 한 후향적 연구로 분만 전 시행한 혈액검사와 초음파 검사, 질 도말 검사, 문진을 통해 분만에 영향을 미칠 수 있는 인자들을 분석하여 인공 신경망을 학습한 뒤 분만이 이루어진 당시의 임신 주수를 예측 가능한지 인공 신경망을 통해 연구조사하였다(Lockwood CJ & Kucyzynski E., 2001; Torchin et al., 2016). 인공 신경망 학습에는 만삭과 조산 분만을 모두 포함하여 산모의 모집단을 반영하도록 학습 범주를 선정하였으며 이를 토대로 조산 분만의 임신 주수 예측의 가능성 여부를 확인하고자 하였다.

2. 연구대상

본 연구는 분만을 위해 입원한 단태아 임부 중 입력값을 모두 포함하고 있는 대상자 137명을 추출하였다. 이들 중 125명은 인공신경망의 학습에 사용되었고 나머지 12명은 별도의 검증을 위한 자료 분석에 사용되었다. 학습에 이용된 125명은 만삭분만과 조산분만을 모두 포함하고 있으며 그 기준 주수는 37주 0일로 정의하였다. 본 연구는 의학연구심의위원회(K 2018-1662-001)의 승인을 받고 진행하였다.

대상자의 입력값은 분만시기에 영향을 미칠 수 있는 인자들을 고려하여 산모의 나이, 분만력, 유산력, 조산의 기왕력, 이전 자궁 소파술, 임신 전 체질량 지수, 전자간증, 임신성당뇨병, 갑상선기능저하증, 자궁경부원추절제술의 기왕력, 이번 임신 중 자궁경부 봉축술 유무, 만삭전조기양막파열 혹은 조기양막파열, 보조생식술 여부, 질출혈, 조기분만진통 여부, 산전 기형아 선별 검사, 분만 입원시 혈압, 백혈구 수치, 적혈구 침강속도, 갑상선자극호르몬, 방광염 여부, 질 도말 검사, 임신 중 자궁경부 길이, 분만 전 가장 최근의 추정 태아 체중을 바탕으로 학습하였다(Andersen et al., 1990; Berkowitz & Papiernik, 1993; Cassell et al., 1993; Goldenberg et al., 2000; Iams et al., 2002; Dilek et al., 2007; Morken et al., 2011; Cnattingius et al., 2013; Liong et al., 2013). 임신 중 자궁경부 길이는 주로 2삼분기에 측정된 값으로 입력하였으며 만삭 분만의 경우 태아 정밀초음파시에 측정된 값으로 입력하였다. 조기분만 진통은 입원 후 30분 동안 적어도 4회 이상의

규칙적인 자궁 수축으로 정의하였으며 이러한 산모들에 대하여 입원 후 자궁 수축 억제제 투여하였고 34주 미만에서 만삭 전 조기 양막파열이 확인되었을 때는 자궁수축억제제 사용 및 항생제 치료를 시작하였다.

3. 연구방법

1) 인공신경망 모델링

인공신경망은 머신러닝(Machine learning) 기법 중 보편적으로 사용되는 방법이다. 이 방법은 생물학적 신경망을 인공적으로 구조화하여 자료 간의 상관관계를 찾아내는 특징을 가지고 있다(Schmidhuber, 2014). Fig. 1은 인공신경망의 구조이며 앞서 언급하였듯이 입력층(Input layer), 은닉층(Hidden layer), 그리고 출력층(Output layer)으로 구성된다. 이런 각 층은 신경(Neuron)으로 연결되어 있는데 생물학에서의 신경이 인공신경망에서는 가중치(Weight)와 편의(Bias)이며, 학

습된다는 것은 입력과 출력의 상관성을 가장 잘 모사하는 가중치와 편의를 찾는 것이다. 첫 번째 네트워크 n_1 은 다음과 같이 표현된다.

$$n_1 = W_1 \cdot p + b_1$$

여기서 $p(R \times 1)$ 는 R 개의 입력변수, $W_1(S^1 \times R)$ 과 $b_1(S^1 \times 1)$ 은 S^1 개의 신경을 가진 은닉층의 가중치와 편이다. 은닉층의 출력 $a_1 = f_1(n_1)$ 은 첫 번째 네트워크 n_1 을 변수로 하는 f_1 함수로 표현된다. 두 번째 네트워크 n_2 는 앞 식과 유사하게 표현된다.

$$n_2 = W_2 \cdot a_1 + b_2$$

여기서 $W_2(S^2 \times S^1)$ 과 $b_2(S^2 \times 1)$ 은 S^2 개의 신경을 가진 출력층의 가중치와 편이다. 출력층의 출력 $a_2 = f_2(n_2)$ 이다. f_1 과 f_2 는 전이함수(Transfer function)이며, 실제 모델에서 다양한 선형 함수와 비선형 함수를 사용한다.

Table 1은 본 연구에서 사용한 인공신경망 모델의 특성을 보여준다. 학습 알고리즘은 역전파로 널리 알려진 Levenberg-

Fig. 1. Architecture of artificial neural networks.

Table 1. Characteristics of the Artificial Neural Network Model

Item	Characteristics
Training algorithm	Levenberg-Marquardt
Type of network	Multi-layer feedforward network
Transfer function	Tan-sigmoid and linear in the hidden and output layer
Input p (36)	Maternal age, BMI, previous preterm delivery, parity, previous abortion, ART, PPRM (or PROM), vaginal spotting, D&C, preeclampsia, GDM, hypothyroidism, previous conization, cervical cerclage, prenatal screening test, preterm contraction, SBP, DBP, WBC, ESR, TSH, UA bacteria, UA WBC, vaginal culture(11), cervical length, last EFW
Output a_2	Gestational age at birth
Number of hidden layer	1
Number of neuron in the hidden layer	Variable (from 16 to 25)
Performance	Correlation coefficient, mean square error
Criterion of performance	Maximum correlation coefficient
Data division	Random (train 70%, validation 15%, test 15%)

Abbreviations: BMI, body mass index; ART, assisted reproductive technology; PPRM, preterm premature rupture of membrane; PROM, premature rupture of membrane; D&C, dilation and curettage; GDM, gestational diabetes; SBP, systolic blood pressure; DBP, diastolic blood pressure; WBC, white blood cell counts; ESR, erythrocyte sedimentation rate; TSH, thyroid stimulating hormone; UA, urinalysis; EFW, estimated fetal weight.

Marquardt 알고리즘이고 네트워크의 구조는 다중 피드포워드(multi-layer feedforward network)이다. 출력값은 분만 당시의 임신주수이며 은닉층의 신경 개수는 16에서 25사이 가변적으로 바뀐다. 모델 성능은 상관계수와 평균제곱근오차로 확인되며 자료 분할은 무작위적으로 학습 70% (89건), 검증 15% (18건), 시험 15% (18건)으로 나뉜다. 전체 자료의 70%만 학습에 사용되며, 검증 15%는 학습을 계속 진행할 것 인지를 통제하는 자료이고 시험 15%는 학습된 네트워크의 성능을 시험하는데 사용된다.

Fig. 2는 인공신경망 모델링 과정을 보여주는 절차이다. 초기 네트워크는 16개 신경과 목표상관계수 0.950으로 설정하였다. 그리고 전체 자료를 임의로 학습 70%, 검증 15%, 시험 15%로 나눴다. Table 1에서 정의한 네트워크 구조에서 역전파 알고리즘으로 입력층의 변수를 삽입하여 학습을 시작한다. 이 때 최대 에폭은 200으로 제한한다. 여기서 에폭은 학습의 반복횟수이다. 대부분 최대 에폭에 이르기 전에 학습된 네트워크의 성능이 개선되지 않기 때문에 멈춘다. Fig. 3은 에폭에 따른 검증단계가 어떻게 확인되는지 보여준다. 이렇게 학습이 완료된 네트워크의 성능을 계산하여 목표값보다 작으면 재학습한다. 재학습 시 변경되는 것은 자료 분할과 신경의 개수이다(Kim et al., 2015).

2) 인공신경망 성능검증

인공신경망 모델의 성능은 상관계수(correlation coeffi-

Fig. 2. Flow chart of the artificial neural network modeling.

Table 2. Characteristics of Pregnant Women (Continue)

	Preterm	Fullterm	Total	p-values*
Total	56	69	125	
Age (years) ± SD	32.4±4.4	32.3±3.6	32.3±3.9	0.941
20~24, n (%)	3 (5.4)	0	3 (2.4)	
25~29	8 (14.3)	16 (23.2)	24 (19.2)	
30~34	25 (44.6)	35 (50.7)	61 (48.8)	
35~39	18 (32.1)	16 (23.2)	34 (27.2)	
40~44	2 (3.6)	2 (2.9)	4 (3.2)	
Parity, n (%)				0.911
0	33 (58.9)	42 (60.9)	75 (60.0)	
1	21 (37.5)	24 (34.8)	45 (36.0)	
2	2 (3.6)	3 (4.3)	5 (4.0)	
Previous preterm delivery, n (%)	9 (16.1)	2 (2.9)	11 (8.8)	0.0097
Previous abortion, n (%)				0.175
0	39 (69.6)	44 (63.8)	83 (66.4)	
1~2	15 (26.8)	25 (36.2)	40 (32.0)	
≥3	2 (3.6)	0	2 (1.6)	
D&C, n (%)	8 (14.3)	11 (15.9)	19 (15.2)	0.798

Table 2. Characteristics of Pregnant Women (Continue)

	Preterm	Fullterm	Total	p-values*
Prepregnancy BMI (Kg/m ²)	21.4±3.1	21.6±3.8	21.5±3.5	0.824
<18.5, n (%)	6 (10.7)	12 (17.4)	18 (14.4)	
18.5~22.9	37 (66.1)	41 (59.4)	78 (62.4)	
23.0~24.9	8 (14.3)	6 (8.7)	14 (11.2)	
25.0~29.9	3 (5.4)	8 (11.6)	11 (8.8)	
30.0~34.9	2 (3.6)	2 (2.9)	4 (3.2)	
ART, n (%)	1 (1.8)	9 (13.0)	10 (8.0)	0.021
Previous conization, n (%)	1 (1.8)	2 (2.9)	3 (2.4)	0.686
Prenatal screening test, n (%)				0.147
Positive	3 (5.4)	9 (13.0)	12 (9.6)	
Negative	53 (94.6)	60 (87.0)	113 (90.4)	
Cervical length (cm)	2.4±1.3	3.5±0.9	2.8±1.3	<0.001
<2.5, n (%)	28 (50.0)	8 (11.5)	43 (34.4)	
≥2.5	28 (50.0)	61 (88.4)	82 (65.6)	
Last EFW (g)	1,989.2±538.5	3,228.1±419.3	2,673.1±779.6	<0.001
<1000, n (%)	4 (7.1)	0	4 (3.2)	
1,000~1,499	10 (17.9)	0	10 (8.0)	
1,500~2,499	32 (57.1)	2 (2.9)	34 (27.2)	
≥2,500	10 (17.9)	67 (97.1)	77 (61.6)	
Results of vaginal culture, n (%)				
<i>Ureaplasma</i>	21 (37.5)	13 (18.8)	34 (27.2)	0.0197
<i>Mycoplasma</i>	3 (5.4)	2 (2.9)	5 (4.0)	0.485
<i>G.vaginalis</i>	5 (8.9)	0	5 (5.0)	0.011
<i>Chlamydia</i>	2 (3.6)	1 (1.4)	3 (1.6)	0.441
<i>Candida spp.</i>	4 (7.1)	14 (20.3)	18 (14.4)	0.037
<i>E.coli</i>	4 (7.1)	8 (11.6)	12 (9.6)	0.401
<i>GBS</i>	2 (3.6)	5 (7.2)	7 (5.6)	0.374
<i>E.faecalis</i>	9 (16.1)	9 (13.0)	18 (14.4)	0.632
<i>K.pneumoniae</i>	3 (5.4)	3 (4.3)	6 (4.8)	0.793
Gestation at birth (weeks)	32.9±3.0	38.5±1.2	36.0±3.5	<0.001
<24+0, n (%)	1 (1.8)	—	1 (0.8)	
24+0~27+6	4 (3.2)	—	4 (3.2)	
28+0~31+6	8 (14.3)	—	8 (6.4)	
32+0~36+6	43 (76.8)	—	43 (34.4)	
≥37+0	—	69 (100)	69 (55.2)	
Neonate weight (g)	2,074.1±591.3	3,222.2±445.7	2,707.8±769.8	<0.001
<1,000, n (%)	5 (8.9)	—	5 (4)	
1,000~1,499	5 (8.9)	—	5 (4)	
1500~2499	32 (57.1)	4 (5.8)	36 (28.8)	
≥2,500	14 (25)	65 (94.2)	79 (63.2)	

*T-test or Chi-square test for p-values.

Abbreviations: SD, standard deviation; D&C, dilation and curettage; BMI, body mass index; ART, assisted reproductive technology, EFW, estimated fetal weight; *G.vaginalis*, *Gardnerella vaginalis*; *E.coli*, *Escherichia coli*; GBS, Group B *Streptococcus*; *E.faecalis*, *Enterococcus faecalis*; *K. pneumoniae*, *Klebsiella pneumoniae*.

cient)와 평균제곱오차(mean square error)를 통해 확인한
다. 상관계수는 다음 식으로 산정한다.

$$R = \frac{COV(X_1, X_2)}{\sigma_{X_1} \sigma_{X_2}} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^N (x_i - \bar{x})^2 \sum_{i=1}^N (y_i - \bar{y})^2}}$$

여기서 COV 와 σ^2 는 공분산과 표준편차이고 x_i 는 목표값(target)이고 y 는 인공신경망모델로 예측되는 출력값(output)으로 본 연구에서 분만시 임신 주수로 산정하였다.

평균제곱오차는 다음 식으로 계산되며 Fig. 4는 에폭에 따른 학습, 검증, 그리고 시험의 평균제곱오차를 보여준다. 여러 에폭에서 가장 좋은 결과를 보여주는 것이 2번째 에폭임을 알 수 있다(Kim et al., 2015).

$$mse = \frac{\sum_{i=1}^N (y_i - x_i)^2}{N}$$

결 과

대상자의 입력값은 모집단을 대표하는지 여부를 판단하기 위해 수치로 표시되는 나이, 임신전 체질량 지수, 혈압, 혈액 검사 결과, 자궁경부 길이, 분만 전 추정 태아 체중, 분만 주수가 모집단의 분포를 따르는지 검증하여 Fig. 5와 같이 적합도를 확인하였다.

그 외의 입력값은 유무의 여부를 판단하여 0 혹은 1, 필요시 2로 분류하여 입력하였다. 대상자의 인구학적 특성을 살펴

Fig. 3. Validation checks according to the number of epoch (an example).

Fig. 4. Mean square error according to the number of epoch (an example)

보기 위해 조산과 만삭으로 나누어 분석하였으며 Table 2에서 조산은 56건, 만삭은 69건으로 모두 125건의 분만이 학습에 이용되었다. 평균 나이는 32.33 ± 3.94 로 두 군에서 큰 차이를 보이지 않았으며 임신 전 체질량 지수 역시 21.51 ± 3.49 로 유사함을 보였다. 또한 자궁경부 길이와 추정 태아 체중 및 질 도말 검사 결과를 확인할 수 있는데 자궁경부 길이는 2.5 cm를 기준으로 조산일 경우 평균 2.43 ± 1.31 으로 짧은 것으로 나타났으며 만삭의 경우 3.45 ± 0.92 로 2.5 cm 보다 길었고 통계학적으로 유의하게 차이가 있는 것으로 나타났다($p < 0.001$). 또한 비율을 보았을 때도 2.5 cm 미만의 조산의 경우 50.0% (28명)로 나타났으며 만삭의 경우 11.5% (8명)의 비율로 나타났다. 질 도말 검사 결과 조산의 산모에서는 *Ureaplasma*, *Mycomplasma*, *Gardnerella vaginalis*, *Chlamydia*와 같은 균종이 비교적 높은 비율을 보였으며 그 중 *Ureaplasma*와 *Gardnerella vaginalis*는 조산산모에서 유의하게 높은 것으로 나타났다. *Candida*는 만삭 산모에서 높은 것으로 보였으나 임상적 의미는 없을 것으로 판단하였다.

Table 3에서는 산모의 기저질환 및 자궁경부 봉축술여부,

Table 3. Obstetrical Complications in Pregnant Women

	Preterm	Full term	Total
Total	56	69	125
Preeclampsia, n (%)			
Yes	6 (10.7)	3 (4.3)	9 (7.2)
No	50 (89.3)	66 (95.7)	116 (92.8)
GDM, n (%)			
Yes	3 (5.4)	8 (11.6)	11 (8.8)
No	53 (94.6)	61 (88.4)	114 (91.2)
Hypothyroidism, n (%)			
Yes	9 (16.1)	9 (13.0)	18 (14.4)
No	47 (83.9)	60 (87.0)	107 (85.6)
Cervical cerclage, n (%)			
Yes	8 (14.3)	5 (7.2)	13 (10.4)
No	48 (85.7)	64 (92.8)	112 (89.6)
PPROM or PROM, n (%)			
Yes	19 (33.9)	7 (10.1)	26 (20.8)
No	37 (66.1)	62 (89.9)	99 (79.2)
Vaginal spotting, n (%)			
Yes	9 (16.1)	1 (1.4)	10 (8.0)
No	47 (83.9)	68 (98.6)	115 (92.0)
Preterm contraction, n (%)			
Yes	28 (50.0)	7 (10.1)	35 (28.0)
No	28 (50.0)	62 (89.9)	90 (72.0)

Abbreviations: GDM, gestational diabetes; PPROM, preterm premature rupture of membrane; PROM, premature rupture of membrane.

질출혈, 만삭전 조기양막파수, 조기분만진통 등에 대해 분석하였으며 조산의 경우 조기분만진통의 비율이 50.0% (28명)로 만삭 산모의 10.1% (7명)보다 높았다.

Fig. 6은 인공지능경망 모델의 성능을 보여준다. 전체 125개의 자료 중에서 학습 자료(training set)는 70%인 89개이고 검증 자료(validation set)와 시험 자료(test set)는 각각 15%인 18개이다. 그림에서 보듯이 학습에 사용된 자료는 예측 결과와 거의 일치하고 있는데 이는 자기 자신을 학습에 사용했기 때문이다(Fig. 6A). 학습 중단을 결정하는 검증 자료의 상관계수는 0.839로 상당히 양호하게 분석되었다(Fig. 6B). 학습에 전혀 개입하지 않는 테스트 자료의 상관계수는 0.892로 나타났다(Fig. 6C). 모든 자료에 대한 상관계수는 0.959로

초기 목표치인 0.95와 유사한 결과이다.

Fig. 7과 Fig. 8은 인공지능경망 모델링에 사용하지 않은 12건의 조산 사례에 대해 모델 검증을 수행한 것이다. 모델 검증은 상관계수를 사용하여 확인하였으며, 개발된 인공지능경망 모델의 상관계수는 0.709로 비교적 우수한 예측능력을 보여주고 있다. Fig. 8은 관찰주수(target)와 인공지능경망 모델의 예측주수(predicted output)를 비교한 그림으로 전체적으로 예측값이 측정값보다 다소 크게 산정되지만 경향성은 유사하며 예측을 비교적 잘 수행하고 있다. Table 4에서는 12사례의 입력값과 출력값을 정리한 표로 입력값을 통해 분만 주수를 추정한 결과를 수치로 나타내었으며 평균 11.58 ± 13.73 의 오차율을 나타내었다.

Fig. 5. Distribution estimation of the data from the prenatal charts. Goodness of fit test was performed on the data using Kolmogorov-Smirnov test, χ^2 (Chi-square) test, and Anderson-darling test. Each goodness of fit test was confirmed to be appropriate, if the level of significance (α) was >0.05 . Abbreviations: BMI, body mass index; SBP, systolic blood pressure; WBC, white blood cell counts; ESR, erythrocyte sedimentation rate; EFW, estimated fetal weight.

Fig. 6. Correlation coefficients of training sets (A, 89 cases), validating sets (B, 18 cases), testing sets (C, 18 cases) and total (D, 125 cases).

Fig. 7. Correlation coefficient of 12 validation test cases.

Fig. 8. Comparison between the target and the predicted output with 12 cases

고찰

조산을 유발하는 인자는 매우 다양한 것으로 알려져 있다. 우리나라 출산율은 급속히 감소하고 있는 추세이나 조산의 비율은 증가하고 있는 실정으로 조산율을 낮추기 위한 노력은 계속 되어야 할 것이다. 그러나 이러한 조산을 미리 예측할 수 있다면 그에 대한 적절한 대처가 가능할 것으로 판단할 수 있

Table 4. Comparison between the Target and the Predicted Output with 12 Cases

	Input						GA at birth (weeks)	Output (weeks)	Error (days)
	Age (years)	BMI (kg/m ²)	CL (cm)	EFW (g)	Cx	VC			
Case 1	27	16.4	3.8	1,940	—	—	34+1	35+2	+8
Case 2	33	16.9	2.3	2,172	PP	—	34+2	36+6	+16
Case 3	28	25.8	1.9	2,501	PP	—	34+4	36+1	+13
Case 4	30	16.9	1.9	2,472	—	<i>Ureaplasma</i> <i>E.coli</i>	34+5	35+4	+6
Case 5	38	24.2	1.3	2,531	+ ¹⁾	—	34+5	32+0	-19
Case 6	35	28.1	4.2	3,042	+ ²⁾	—	36+1	41+0	+34
Case 7	33	16.2	2.5	2,423	+ ³⁾	<i>Ureaplasma</i> <i>Candida</i>	36+5	39+2	+18
Case 8	30	19.7	0.48	1,279	+ ⁴⁾	—	28+1	31+0	+20
Case 9	25	19.5	0.2	2,883	—	—	35+2	37+6	+18
Case 10	30	22.0	0.8	2,251	PP	<i>Ureaplasma</i>	34+5	33+5	-7
Case 11	37	23.0	1.76	1,652	+ ⁵⁾	—	31+3	34+2	+20
Case 12	33	21.3	1.55	1,690	PP	<i>Ureaplasma</i>	32+0	33+5	+12

¹⁾ previous preterm delivery. ²⁾ previous preterm delivery, hypertension, GDM, hypothyroidism. ³⁾ GDM. ⁴⁾ hypothyroidism. ⁵⁾ previous preterm delivery, GDM.

Abbreviations: BMI, body mass index; CL, cervical length; EFW, estimated fetal weight; Cx, obstetrical complications; VC, vaginal culture; GA, gestational age; PP, preterm premature rupture of membrane; *E.coli*, *Escherichia coli*.

다. 산모에 대한 산전 관리 및 태어날 신생아의 건강과 신생아 중환자실 치료 여부에 대한 예측을 함으로써 비용 및 신생아 이환률 또한 감소시킬 수 있을 것이다(Behrman & Butler, 2007).

본 연구에서 분만 시기에 영향을 미칠 수 있는 인자로 산모의 나이, 분만력, 유산력, 조산의 기왕력, 이전 자궁 소파술, 임신 전 체질량 지수, 전자간증, 임신성당뇨병, 갑상선기능저하증, 자궁경부원추절제술의 기왕력, 이번 임신 중 자궁경부 봉축술 유무, 만삭전 조기양막파열, 보조생식술 여부, 질출혈, 조기분만진통 여부, 산전 기형아 선별 검사, 분만입원시 혈압, 혈액 검사를 통한 염증수치, 방광염 여부, 질 도말 검사, 임신 중 자궁경부 길이, 분만 전 가장 최근의 추정 태아 체중을 입력값으로 채택하였다. 그러나 이는 조산을 예측할 수 있는 원인 중 일부만을 포함하고 있으며 최근 연구에서 조기진통 및 분만의 예측 인자로 알려진 자궁경부에서의 태아섬유결합소(fetal fibronectin) 검사 결과는 대상자의 수가 적어 포함하지 못 하였다(Bruijn et al., 2016). 조기분만의 경우 조기 분만진통과 만삭 전 조기 양막파열이 본 연구에서도 높은 비중을 차지하고 있는데 그 원인으로 자궁 내 감염 및 염증이 연관성이 있다고 알려져 있으며 그에 대한 병리 기전에 대한 연구 역시 많이 이루어지고 있다(Goldenberg et al., 2000). 그에 대한 지표로 혈액 검사를 통한 백혈구 수치와 적혈구 침강속도, 방광염 여부, 질 도말검사를 통한 균 동정을 시행해 입력값으로 채택

하였으나 그에 따른 적절한 시기의 항생제 치료 여부, 조기 분만진통시 자궁수축억제제 사용 여부 등은 입력값에 포함하지 않았다. 또한 C-반응성단백질(C reactive protein, CRP) 검사의 경우 만삭 분만 산모의 경우 실시 되지 않은 경우가 많아 본 연구의 입력값으로는 제외되었다. 그러나 C-반응성단백질 검사의 경우 적혈구 침강속도에 비해 염증반응에 대한 민감도가 높고 반응속도가 빨라 추후 입력값으로 포함할 경우 염증수치를 반영하여 학습의 효율을 올리는데 효과적일 것이다.

Zhu와 그의 동료들이 최근 조산 예측을 위한 새로운 모델링을 시도한 연구에 따르면 조산과 연관된 8가지 인자를 분석하여 양성 소견이 증가될 때마다 조산의 가능성이 높아진다고 보고하였다. 8가지 인자인 이전 조산력, 임신 전 체질량지수, 초음파를 통한 자궁경부의 탄성도(cervix tissue elasticity), 자궁경부의 밀도(densitometry of cervix), 자궁경부 개대, 자궁경부길이, 태아섬유결합소, 자궁경부 및 질 내 인터루킨-1β(interleukin-1β)이며 이들 인자의 양성 결과에 따라 “≥2×양성인자” 인 경우 상대적으로 높은 82.6%의 민감도와 88.1%의 특이도를 확인하였다고 보고 하였다. 또한 상기 연구에서는 만삭전 조기양막파수와 자궁경부 개대가 3 cm 이상, 생식기관의 감염, 질출혈, 조기분만진통, 고혈압, 당뇨 등과 같은 산과적 합병증을 제외시킨 상태에서 예측한 연구로써 의미가 있다 할 수 있다(Zhu et al., 2017).

본 연구에서는 산모의 나이, 임신 전 체질량 지수, 백혈구수

치, 적혈구 침강속도, 갑상선자극 호르몬, 분만 입원시 혈압, 자궁경부 길이, 태아 체중 예측값을 제외하고 모두 0과 1 혹은 2로 입력되었는데 인공신경망 학습시 이러한 수치의 입력은 학습의 정확도를 감소시킨다. 따라서 향후 연구의 정확도를 높이기 위해서는 체내 염증 정도를 나타내는 적혈구 침강계수를 포함한 C-반응성 단백질, 인터루킨-1, 인터루킨-6, 인터루킨-8, 종양괴사인자(tumor necrosis factor, TNF) 및 태아 섬유결합소의 정량과 같은 수치화된 정보를 입력값으로 추가한다면 정확도가 높아질 것으로 예상된다(Goldenberg et al., 2000). 또한 항생제 투여와 자궁수축억제제의 사용은 분만일시를 연장시키는 효과가 있으므로 투여 약물의 종류 및 사용 일수를 입력값으로 추가하는 것 역시 출력값의 정확도를 향상시킬 수 있을 것이다. 산모의 분만 방법 역시 학습에 영향을 미칠 수 있는데 특히 제왕절개의 경우 정규 수술과 응급 수술의 경우가 있으므로 추후 구분하여 분석할 필요가 있다. 학습의 정확도를 높이기 위해서는 제왕절개와 질식 분만을 분리하여 질식 분만 중에서도 자발적 분만 진행의 경우에 대해 조산과 만삭 산모를 학습한다면 향상된 결과값을 확인할 수 있을 것으로 사료된다.

인공신경망 학습에 사용하지 않은 별도의 조산 분만 12 건의 사례에 대한 상관계수는 0.709로 비교적 신뢰할 만한 수치를 나타내었으며 인공신경망 학습에 이용된 125건 중 테스트로 사용된 18건의 분만에 대한 상관계수 0.892보다는 낮게 나타났다. 이는 본 연구에서 학습한 분만 건수의 개수가 상대적으로 적고, 산모의 자발적인 진통에 의한 분만 뿐 아니라 진통 전 시행한 제왕절개 분만 역시 포함되어 있기 때문이다. 특히 Table 4에서 6번째 사례의 경우 예측값이 43주로 계산되었는데 예측 분만 주수는 임상적으로 유도분만이 시행되는 41주를 최대 임신주수로 가정하여 예측값으로 채택하였다. 또한 예측분만주수가 높게 나온 것은 2삼분기에 측정된 자궁경부길이 길고 임신성당뇨의 영향으로 태아추정체중 역시 다른 산모에 비해 크게 측정되었기 때문으로 보인다. 8번 사례에서는 입원 당시 자궁경부개대가 60% 가량 진행되어 분만이 급속히 이루어졌으며 이를 반영하기 위한 지표로 추후 입원당시 Bishop score도 추가하여 학습하게 된다면 예측 분만 주수의 정확성을 향상시킬 수 있을 것이다. 11번 대상자의 경우 태아 모니터링 상에서 반복되는 심박수 저하로 응급제왕절개술로 분만이 이루어졌으며 이는 학습시 인자로 포함되지 않아 예측 분만 일수와 실제 분만일 사이에 상대적으로 큰 오차가 생긴 것으로 판단된다. 따라서 상기와 같은 이유 등으로 인해 상관

계수 및 오차 일수의 평균과 표준편차에 상당한 영향을 미친 것으로 보인다.

그리고 학습 세트의 상관계수와 검증, 시험 세트에서 보인 상관계수 사이에 0.10 이상의 큰 차이를 보였는데 이는 과적합(overfitting)이 일어났다고 하며 이를 보완하기 위해서는 자료를 확대하고 훈련을 강화해서 모형의 과적합을 해소하고 모형의 확장성(generalization)을 개선하는 것이 앞으로의 연구 과제이다.

본 연구는 125건의 임부를 대상으로 한 소규모 연구로, 인공신경망 학습 시 자료가 방대할수록 정확도가 향상된다는 점에서 본 연구의 한계가 있으며, 분만에 영향을 미치는 다양한 인자들을 모두 포함하지 못했다는 제한점이 있다. 또한 분만 시기를 결정함에 있어 임상의의 임상적 판단이 영향을 미칠 수 있고 연구대상자 선정에 있어서 대학병원급 산모들을 대상으로 하여 선택 편이 발생 여부도 배제할 수 없다.

그러나 본 연구는 분만 전 시행한 혈액검사와 초음파 검사, 질 도말 검사, 요검사 및 산과력, 과거력, 산과적 합병증 등의 입력값을 통해 인공신경망을 학습하여 분만시기를 예측하고자 시도하였다는 점에서 그 의의를 찾을 수 있다. 이를 바탕으로 향후 보완된 측정값을 포함한 대규모 데이터베이스를 토대로 인공신경망을 학습시켜 분석한다면 조산의 시기를 예측할 뿐 아니라 조산에 미치는 비중 인자를 분석하여 그에 따른 고위험 산모의 개별적 집중치료 역시 가능할 것으로 보여 연구적 가치가 있을 것으로 사료된다.

결 론

본 연구는 125건의 단태아를 분만한 임부를 대상으로 만삭 전 조기양막파수, 조기분만진통, 자궁경부 상태와 염증 정도, 감염 여부, 임신관련 합병증 등을 입력값으로 인공지능을 학습하여 조기 분만 여부를 판별할 뿐 아니라 분만 주수를 예측하고자 하였다. 125건 중 순수 학습에 이용된 89건으로 학습의 상관계수는 1.00으로 나타났고 모델링 내 18건을 테스트로 시행한 상관계수는 0.892로 나타났다.

인공신경망 학습 모델링과 무관한 조산 분만 건수 12사례에 대한 분만주수 예측의 상관계수는 0.709였으며, 실제 분만 주수와 오차일 수는 11.58 ± 13.73 였다. 이는 실제 산모들에게 투여된 자궁수축억제제 및 항생제 치료 및 제왕절개 일 등이 고려되지 않아 크게 34일까지의 분만일 수 차이를 보였다. 이러한 오차는 본 연구에서 사용된 0과 1의 입력값을 정

량적인 수치로 보완하고 치료일 수, 치료방법, 분만 방법, 제왕 절개의 적응증, 태아상태에 대한 정보 등과 같은 입력값을 추가하여 대규모 데이터를 바탕으로 인공신경망을 학습한다면 향후 조기 분만의 임신 주수를 예측할 뿐 아니라 분만에 영향을 미치는 가장 비중 있는 변수를 확인하여 그에 대한 적절한 치료 및 대처도 가능할 것으로 보인다.

참고문헌

- Andersen HF, Nugent CE, Wanty SD, Hayashi RH. Prediction of risk for preterm delivery by ultrasonographic measurement of cervical length. *Am J Obstet Gynecol* 1990;163(3):859-67.
- Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, Requejo JH, et al. The worldwide incidence of preterm birth: a systematic review of maternal mortality and morbidity. *Bulletin of the World Health Organization* 2010;88:31-8.
- Behrman RE, Butler AS. *Preterm Birth: Causes, Consequences, and Prevention*. The National Academies press. Washington DC 2007;255-8.
- Berkowitz GS, Papiernik E. Epidemiology of preterm birth. *Epidemiologic Reviews* 1993;15:414-43.
- Bruijn MM, Kamphuis EI, Hoesli IM, Tejada BM, Loccufer AR, K hnert M. et al. The predictive value of quantitative fibronectin testing in combination with cervical length measurement in symptomatic women. *Am J Obstet Gynecol* 2016;215:791-8.
- Cassell GH, Waites KB, Watson HL, Crouse DT, Harasawa R. *Ureaplasma urealyticum* intrauterine infection: role in prematurity and disease in newborns. *Clin Microbiol Rev* 1993;6(1):69-87.
- Catley C, Frize M, Walker RC, Petriu DC. Predicting high-risk preterm birth using artificial neural networks. *IEEE Trans Inf Technol Biomed* 2006;10(3):540-9.
- Cnattingius S, Villamor E, Johansson S. Maternal Obesity and Risk of Preterm Delivery. *JAMA* 2013;309(22):2362-70.
- Dilek TUK, Yazici G, Gurbuz A, Tasdelen B, Gulhan S, Dilek B. Progressive Cervical Length Changes versus Single Cervical Length Measurement by Transvaginal ultrasound for Prediction of Preterm Delivery. *Gynecol Obstet Invest* 2007;64:175-9.
- Goldenberg RL, Hauth JC, Andrews WW. Intrauterine infection and preterm delivery. *N Engl J Med* 2000;342:1500-7.
- Iams JD, Newman RB, Thom EA, Goldenberg RL, Mueller-Heubach E, Moawad A. Frequency of Uterine Contractions and the Risk of Spontaneous Preterm Delivery. *N Engl J Med* 2002;346:250-5.
- Kho BW, Kim MK. The Implemetation of Artificial Neural Network Model for improving the Diagnosis Accuracy of Type 2 Diabetes. *Proceed Symposium Korean Inst commun Inform Sci* 2018;1:849-50.
- Kim SW, Melby JA, Nadal-Caraballo NC, Ratcliff J. A time-dependent surrogate model for storm surge prediction based on an artificial neural network using high-fidelity synthetic hurricane modeling. *Nat Hazards* 2015;76:565-85.
- Lee KA, Chang MH, Park MH, Park H, Ha EH, Kim YJ, et al. A model for prediction of spontaneous preterm birth in asymptomatic women. *J Womens Health* 2011;20:1825-31.
- Lim KH, Ryu KS, Park SH, Shon HS, Rhy KH. Short-term Mortality Prediction of Recurrence Patients with ST-segment Elevation Myocardial Infarction. *J Korea Soc Comp Inform* 2012;17(10):145-53.
- Liong S, Quinzio MD, Fleming G, Permezel M, Rice GE, Georgiou HM. Prediction of spontaneous preterm labour in at-risk pregnant women. *Reproduction* 2013;146:335-45.
- Lisboa PJG. A review of evidence of health benefit from artificial neural networks in medical intervention. *Neural Netw* 2002;15:11-39.
- Lockwood CJ, Kuczyński E. Risk stratification and pathological mechanism in preterm delivery. *Paediatr Perinat Epidemiol* 2001;15(2):78-89.
- Morken NH, Gunnes N, Magnus P, Jacobsson B. Risk of spontaneous preterm delivery in a low-risk population: the impact of maternal febrile episodes, urinary tract infection, pneumonia and ear-nose-throat infections. *Eur J Obstet Gynecol Reprod Biol*. 2011;159:310-4.
- Morken NH, K llen K, Jacobsson B. Predicting risk of spontaneous preterm delivery in women with a singleton pregnancy. *Paediatr Perinat Epidemiol* 2014;28(1):11-22.
- Schmidhuber J. Deep Learning in Neural Networks: An overview. *Neural Netw* 2015;61:85-117.
- Statistics Korea, Final results of Birth statistics in 2016.
- Torchin H, Ancel PY. Epidemiology and risk factors of preterm birth. *J Gynecol Obstet Biol Reprod* 2016;45(10):1213-30.
- Vinken M, Rabotti C, Mischi M, Oei SG. Accuracy of Frequency-Related Parameters of the Electrohysterogram for Predicting Preterm Delivery: A Review of the literature. *Obstet Gynecol Surv* 2009;64(8):529-41.
- Zhu Y, Peng G, Tian G, Qu X, Xiao S. New model for predicting preterm delivery during the second trimester of pregnancy. *Scientific Reports* 2017;7:11294.