

태아심박동자료의 발육제한증 진단을 위한 신경망 모형

차경준¹⁾, 황선호²⁾

요 약

본 연구에서는 자궁 내 태아의 발육지연이 주산기 사망률 및 이환율을 증가시키는 고위험 임신의 한 예로써, 태아 발육제한증과 관련한 비선형적인 자료를 통계적인 방법으로 접근하는데 초점을 두었다. 이에 정상태아와 발육제한증 태아를 판별하기 위한 분석을 실시함에 있어 신경망 이론 중 하나인 다층 퍼셉트론 모형으로 예측하고자 하였다.

주요용어 : Multilayer perceptron, FHR, IUGR

1. 서론

고위험임신의 한 예로서 태아발육제한증(Intrauterine Growth Restriction; IUGR)과 관련한 자료를 분석하였다. 자궁 내 태아의 발육지연은 주산기사망률 및 이환율을 증가시키는 중요한 요인중의 하나로, 이의 원인에는 여러 가지가 있으나 대부분이 태반의 기능이 어떠한 원인-예를 들면 임신성 고혈압 등으로 저하되어 태아로 가는 혈류 및 영양공급의 감소로 인한 이차적인 결과로 생각되고 있다. 이러한 태아의 안녕 상태를 예견하는 중요한 지표로 태아의 심박동 변이도의 분석은 최근 많이 이용되는 것으로 특히 태아 가사상태나 저산소증, 태아산혈증으로 인한 중추신경계의 기능 저하가 있는 경우에는 심박동 변이도가 감소됨은 잘 알려져 있다. 따라서, 태아의 심박동 자료를 토대로 하여 정상태아와 발육지연태아 진단하기 위한 분석을 실시함에 있어 신경망 이론에서 지도학습 중 하나인 다층 퍼셉트론과 연결강도를 변경하는 알고리즘으로 오류역전파 알고리즘에 적용하고자 한다.

2. 사례 분석

2.1 태아 심박동 자료

본 연구에서 두 그룹을 판별하기 위해 다음과 같은 변수를 이용하여 분석하였다.

1. 평균 태아심박동(Mean Based Fetal Heart Rate)
2. FHR 변이도(amplitude; AMP)
3. 분만주수(Delivery weeks)
4. Ponderal Index($g/cm^3 \times 100$); 태아의 체형을 표준화하기 위한 개개인의 키와 몸무게의 특

1) 한양대학교 수학과 교수

2) 한양대학교 대학원 수학과

별한 비로 정의한다.

5. Approximate entropy; 정상적인 심박동은 많은 다양한 조절시스템에 의해서 이루어지는 매우 케이오틱한(chaotic)신호로 심박동의 복잡성(complexity)과 변동성(fluctuation)은 생체의 건강상태를 반영하는 중요한 인자가 된다.

태아심박동과 제 변수 자료는 한양대학교 산부인과에서 개발된 FHR분석시스템인 HYFM-I & II(1998)를 이용하여 수집 처리되었고, 자료의 통계적 분석은 SAS(Statistical Analysis System, version 8.1, USA)를 이용하였고, 자료의 판별분석을 위한 신경망 프로그램은 C-언어를 사용하여 개발하였다.

2.2 신경망 이론

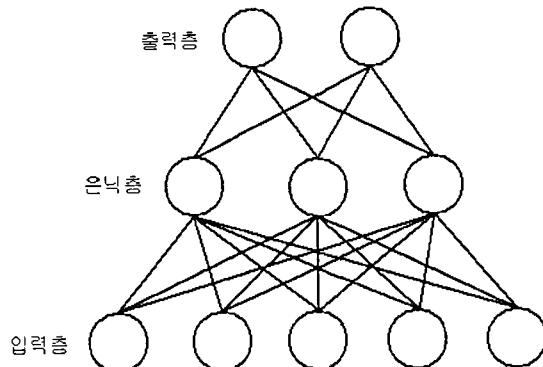
신경망이란 인간의 뇌를 흉내내 구성한 네트워크로 그 학습능력과 추론능력이 매우 뛰어난 것으로 알려져 있다. 신경망은 단순한 처리인자들인 뉴론(neuron)의 전이함수(transfer function)와 신경망의 구조를 나타내는 계층(layer)의 수, 처리인자들간의 연결상태(connection) 및 연결강도(weights)로 구성되어 있다. 이들을 적절하게 조정하는 학습규칙에 따라 여러 가지의 형태를 갖게 된다.

신경망은 입력변수(input)와 출력변수(output)간의 아무런 수학적 공식이 존재하지 않거나, 문제를 해결하는 공식이나 알고리즘, 정확한 이론적 모형이 없는 경우와 설명보다는 예측이 더 중요한 경우 예측 문제에 있어 유용하기 때문에 최근 몇 년 동안 신경망은 통계적인 방법과 결합되기 시작하여 자료로부터 복잡한 비선형관계를 발견해 냈으므로써 통계적 모형으로 매우 유용하다는 평가를 받고 있다.

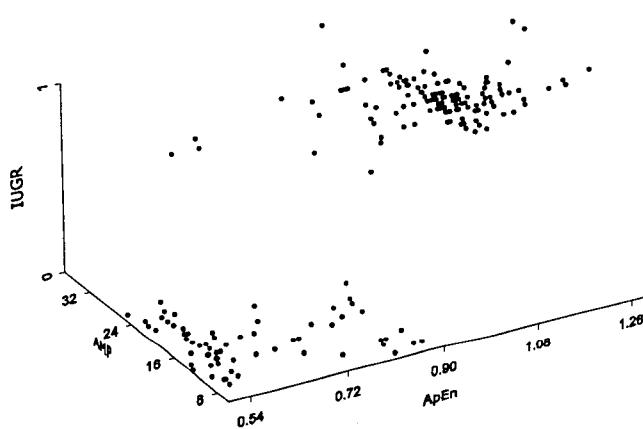
신경망의 가장 대표적인 구조는 <그림 1>에서 보여주는 다층 퍼셉트론(Multilayer Perceptron; MLP)모형이다. 이는 입력층(input layer)과 은닉층(hidden layer), 출력층(output layer) 방향으로 연결되며, 각 층 내의 연결과 출력층에서 입력층으로는 직접적으로 연결되지 않는 전방향(feedforward) 구조이다.

MLP 모형은 입력층과 출력층 사이의 중간층과 각 마디(unit)의 입력 출력 특성을 비선형화 함으로써 네트워크의 능력을 최대화한다.

신경망에 쓰이는 학습은 크게 지도학습과 자율학습으로 나눌 수 있다. 지도학습은 입력된 값에 대해 출력값이 정해져 있기 때문에 학습에서 피드백(feedback)을 줄 수 있는 반면에 자율학습은 입력된 값에 대해 출력값이 정해져 있지 않기 때문에 층 내에서 비슷한 처리요소간의 상호작용에 의하여 군집화(clustering)하면서 유사한 변수끼리 묶는 작업을 하게 된다.



<그림 1> 다층 퍼셉트론 아키텍쳐

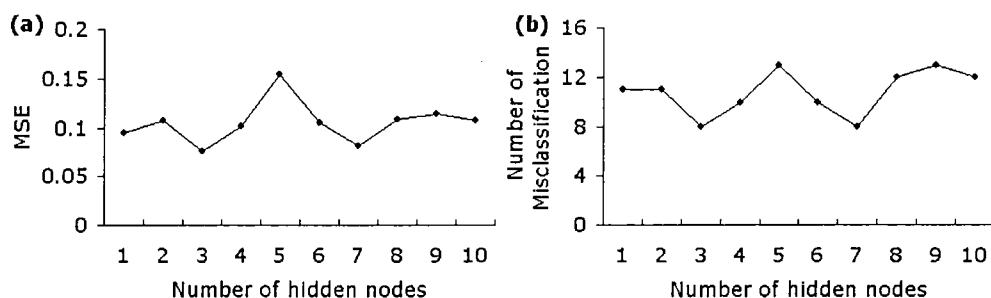


<그림 2> 정상태아군($N = 133$)과 발육제한증태아군($N = 67$)의 FHR 변이도와 근사엔트로피에 대한 산점도

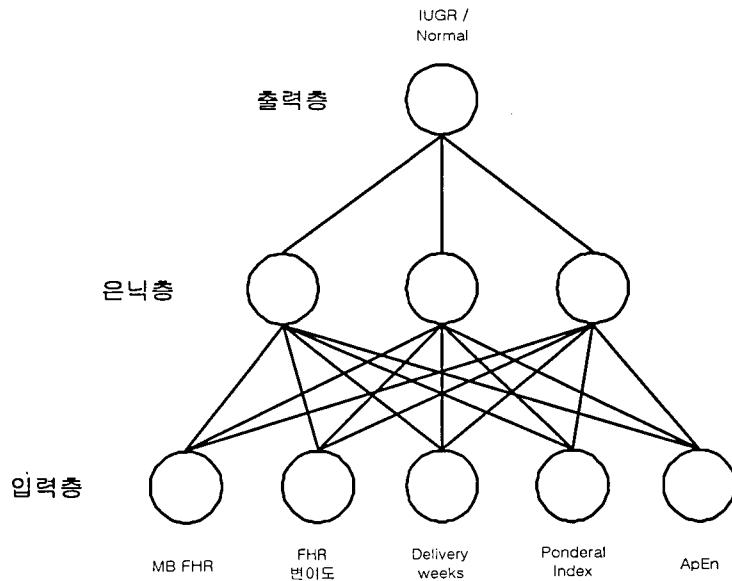
2.3 MLP 모형과 분석결과

본 연구에서 사용되어지는 변수로는 평균 태아심박동수, FHR 변이도, 분만주수, 근사 엔트로피, Ponderal Index 등의 5개를 독립변수로 선정하였고 발육제한증태아군과 정상태아군의 두 그룹을 판별하고자 하였다.

FHR 자료는 강한 비선형성을 가지고 있어 현재까지도 통계적 분석을 위한 특별한 모형이 제안되지 못하였다. 이에, 정상태아와 발육제한증태아를 판별하기 위한 분석으로 신경망 이론 중 다층 퍼셉트론을 C 언어로 개발하여 분석하였다. 다층 퍼셉트론 프로그램에 적용하기 위해 위에서 제시한 5개의 독립변수와 정상태아군과 발육제한증태아군 두 그룹을 명목형인 종속변수로 사용하였고, 분석에 있어서는 300 예의 데이터 중에서 200 예를 훈련데이터셋(training dataset)



<그림 3> (a) 은닉노드의 개수에 따른 검증데이터의 관측값과 예측값의 MSE
(b) 은닉노드의 개수에 따른 검증데이터의 오분류 개수



<그림 4> 태아 심박동 자료의 최적 신경망 모형

으로, 100 예를 검증데이터셋(test dataset)으로 하였고 기본적으로 은닉계층의 개수는 1개로 고정시켰다.

<그림 2>는 두 그룹에 대한 독립변수 중에서 두 변수 FHR 변이도와 균사엔트로피에 대한 두 그룹의 산점도이다. 그림에서 FHR 변이도가 높을수록 그리고 균사엔트로피가 높을수록 정상태아군으로 나타났고 반대로 낮은 FHR 변이도와 낮은 균사엔트로피에서 발육제한증 태아군으로 나타났다. 신경망모형을 최적화하기 위하여 1개의 은닉계층에 대하여 1개~10개의 노드를 갖는 신경망 모형을 훈련시킨 후 검증데이터로 실제값과 예측값간의 *MSE*를 구하였고, 또한 이에 대한 오분류의 개수를 비교하였다. <그림 3>은 각각의 노드에 대한 실제값과 예측값간의 *MSE*와 오분류를 나타내었다. 그 결과, *MSE*와 오분류에서 노드의 개수에 따라 거의 일치된 형태를 가졌으며, 특히 은닉노드가 3개일 경우에 가장 낮은 *MSE*와 오분류를 가짐을 알 수 있었다. 따라서 5개의 입력 노드와 3개의 은닉노드, 1개의 출력 노드를 갖는 다층 퍼셉트론을 가장 적합한 모형으로 선택하였다. <그림 4>는 최적의 신경망 모형 아키텍쳐이다.

이 신경망 모형과 전통적인 통계적 모형을 비교하기 위하여 짐단의 수가 두 개인 경우 사용되는 로지스틱 판별 모형을 태아 심박동 자료에 대하여 실시하였다.

<표 1>은 검증데이터 100예를 은닉노드가 3개로 훈련되어진 다층 퍼셉트론 아키텍쳐와 로지스틱 판별모형에 의해 실행되어진 분류 결과이다. 관측치는 정상태아군, 발육제한태아군 각각 67, 33예 이고, 이 중에서 신경망모형의 경우에는 정상태아 67예 중 61예를 정상태아로, 발육제한태아 33예 중 31예를 발육제한태아로 예측하였다. 즉, 총 100예 중 총 8예가 오분류 되었고, 여기서 실제 정상태아에서 6예를, 실제 발육제한태아에서 2예가 오분류 되었음을 알 수 있다. 이에 비해 로지스틱 판별 모형에서는 정상태아 67예 중 63예를 정상태아로, 발육제한태아 33예 중 28예를 발육제한태아로 예측하였다. 총 100예 중 총 9예가 오분류 되었고, 여기서 실제 정상태아에서 4예를, 실제 발육제한태아에서 5예가 오분류 되었음을 알 수 있다. 3개의 은닉노드를 갖는 신경망모형의 오분류율은 8%(8/100)이고 로지스틱 판별 모형의 오분류율은 9%(9/100)

<표 1> 3개의 은닉노드를 갖는 MLP 모형과 로지스틱 판별 모형의 100 예의 검증데이터에 대한 분류행렬

예측값 \ 관측값	정상태아	발육제한증태아	Total
정상태아	61(63)	2(5)	63(68)
발육제한증태아	6(4)	31(28)	37(32)
Total	67	33	100

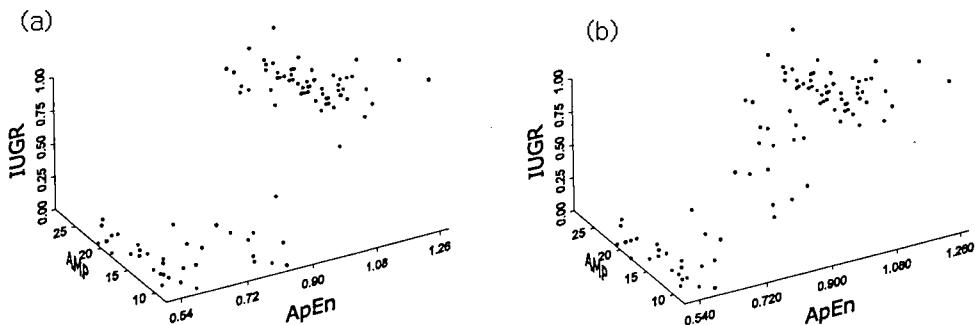
()의 값은 로지스틱 판별분석의 예측결과

이다. 그리고 신경망 모형에서는 정상태아군에서, 로지스틱 판별모형에서는 발육제한증에서 각각 오분류율이 높았으며 또한 이 두 모형 중 신경망모형의 오분류율이 로지스틱 판별 모형의 오분류율보다 더 낮음을 알 수 있다.

<그림 5>는 검증데이터 100예를 3개의 은닉노드로 훈련되어진 신경망모형과 로지스틱 판별모형에 의해 실행되어진 분류 결과에 대하여 독립변수 중 FHR 변이도와 근사엔트로피에 대한 두 그룹의 산점도이다. 예측값은 정상태아군에 속할 확률값으로 출력되는데, 정상태아일 경우는 1에 가까운 값으로 발육제한태아일 경우는 0에 가까운 값으로 출력된다. <그림 5>의 (a)는 신경망 모형의 예측값의 분포이고, (b)는 로지스틱 모형의 예측값의 분포이다. 각각의 분포를 보면 신경망의 모형의 경우는 예측확률값이 정상태아와 발육제한태아의 구분이 확실하게 나타난다. 즉 정상태아의 예측확률은 대부분이 1에 가까운 값이고, 발육제한태아의 예측확률은 대부분이 0에 가까운 값이다. 그러나 로지스틱 판별 모형의 경우는 조금 애매한 확률값을 가짐을 볼 수 있다.

3. 결론 및 향후과제

신경망 이론 중 다층퍼셉트론을 이용하여 고위험임신의 한 예로서 태아발육제한증(Intrauterine



<그림 4> 정상태아군과 발육제한증태아군에 대한 각 모형의 예측값의 산점도

(a) ; 정상태아군과 발육제한증태아군에 대한 신경망 모형($MSE = 0.07584$)

(b) ; 정상태아군과 발육제한증태아군에 대한 로지스틱 판별 모형($MSE = 0.08023$)

Growth Restriction; IUGR)군과 정상태아군을 판별하기 위한 분석을 태아심박동에 관한 자료를 각 독립변수(평균 태아심박동수, FHR 변이도, 분만주수, 근사 엔트로피, Ponderal Index)로 사용하였고, 여러 신경망 모형을 제시하였다. 은닉계층은 1개로 고정시킨 후 은닉노드의 개수에 따라 여러 모형을 구현하여 최종적으로 5개의 입력노드, 3개의 은닉노드 그리고 1개의 출력노드를 갖는 아키텍쳐를 선택하였다($MSE = 0.07584$, 오분류율 = 0.08). 또한 최적 모형으로 선택된 신경망 모형과 로지스틱 판별 모형을 비교해 보면 로지스틱 판별 모형은 MSE 가 0.08023 이고 오분류율은 0.09가 되었다. 특히 신경망 모형에서는 정상태아군에서, 로지스틱 판별모형에서는 발육제한증태아군에서 각각 오분류율이 높은 결과를 얻었는데, 실제 진단에 있어서는 발육제한증태아군에서의 오분류가 더 낮아야 더 효과적인 진단을 할 수 있을 것이라 사료된다. 또한 예측확률에 대한 분포는 조금 차이를 가짐을 알 수 있었다. 즉 신경망 모형은 예측확률에 있어서 확실한 구분을 보여준 반면, 로지스틱 판별 모형은 종종 애매한 예측확률값을 보여주었다.

사례 분석의 예로서 산과학에서 고위험임신 중 태아발육제한증에 적용하였던 바, 이러한 매우 비선형적인 자료는 전통적인 통계적 방법인 로지스틱 판별 모형보다 신경망 모형이 조금 더 우수함을 알 수 있었고, 정상적인 태아군과 발육제한 태아군을 성공적으로 구별할 수 있는 상당히 정확한 분류율을 제공하였다. 이로써, 한양대학교 산부인과에서 개발된 FHR분석시스템인 HYFM-I & II(1998)를 이용하여 자료를 수집 처리하여 신경망 모형에 적용함으로써 발육제한증의 일련의 진단방법으로 제시할 수 있다.

참고문헌

- [1] 김대수(1992), “신경망 이론과 응용(I)”, 하이테크정보.
- [2] 황정혜, 박문일, 염명걸(1997), “근사엔트로피(Approximate Entropy) : 발육지연 태아에서의 새로운 심박동변이도의 분석지표”, 대한산부인과학회지, 제40권, 1호, pp 92-99.
- [3] D. Michie, D. J. Spiegelhalter, C.C. Taylor(1994), "Machine Learning, Neural and Statistical Classification", Ellis Horwood, New York.
- [4] K. Maeda, M. Utsu, A. Makio, M. Serizawa, Y. Noguchi, T. Hamada, K. Mariko, F. Matsumoto.(1998), "Neural network computer analysis of fetal heart rate", Journal of Maternal Fetal Investigation, Vol. 8, pp 163-171.
- [5] Simon Haykin(1999), "Neural networks 2nd ed", Prentice Hall, New Jersey.
- [6] Warren S. Sarle.(1994), "Neural networks and statistical models", Proceedings of the 19th Annual SAS Users Group International Conference.