

## Development of Discriminant Model of PIH Pregnant using Decision Tree\*

Young-Sun Park<sup>1</sup>, Hang-Suk Choi<sup>2</sup>, Young-Koun Lee<sup>3</sup>,  
Kyung-Joon Cha<sup>4</sup>, Sung-Hoon Lee<sup>5</sup>, Moon-Il Park<sup>6</sup>

### Abstract

The various methods have been studied to develop discriminant model for Pregnancy Induced Hypertension(PIH) as high risk pregnant. In this study, we adapt the approximate entropy which is the non-linear chaotic measuring method. Then, we develop the system to discriminant PIH pregnant using QUEST with S-PLUS.

*Keywords* : Decision Tree, QUEST algorithm, PIH, Approximate Entropy.

### 1. 서론

현대사회의 발전과 더불어 최근에 다양한 분야에서 의사결정에 관한 문제를 해결하기 위해 여러 가지 기법들이 모색되고 있다. 그 중에서 1960년대 이후 꾸준히 발전하여 최근까지 새로운 알고리즘이 개발되고 있는 것이 의사결정나무모형이다.

의사결정나무모형은 분류 또는 예측이 가능하여 분석과정의 설명이 필요한 경우 더욱 유용하게 사용되는 기법이다. 더욱이, 결과의 용이함과 누구나 쉽게 이해할 수 있고 설명되어질 수 있어 고객의 의사결정, 신용 점수화(credit scoring), 시장분석(market analysis) 등과 더불어 다양한 분야에서 이용되고 있다(Frawley, 1991).

---

\*This work was supported by Hanyang University Research Fund, 2003.

<sup>1</sup>Research Professor, Research Institute of Natural Sciences, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea. E-mail : pppppys@ihanyang.ac.kr

<sup>2</sup>Ph.D student, Dept. of Mathematics, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

<sup>3</sup>Msc student, Dept. of Mathematics, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

<sup>4</sup>Professor, Dept. of Mathematics, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

<sup>5</sup>Dept. of Obstetrics and Gynecology, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

<sup>6</sup>M.D., Dept. of Obstetrics and Gynecology, Hanyang Univ., 17 Haengdang-dong, Seongdong-Gu, Seoul, 133-791, Korea.

본 연구에서는 산과학에서 고위험 임신의 한 예로서 산모의 임신성고혈압증(Pregnancy Induced Hypertension; PIH)과 관련된 태아의 판별모형 개발을 위하여 의사결정나무 모형 중 하나인 QUEST algorithm을 활용하였다.

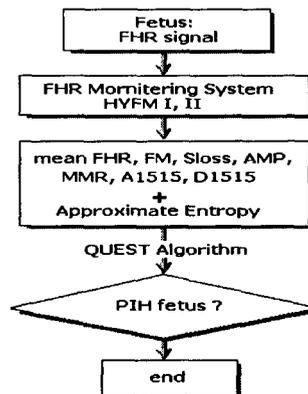
실제로 임신성고혈압과 관련된 질환인 만성고혈압(Chronic Hypertensive Vascular Disease; CHVD) 및 경증 및 중증 전자간증(Mild and severe preeclampsia) 등은 주요한 모성사망의 원인중 하나로서 Precht(1967)에 의하면 이러한 위험요인이 두 개 이상 동시에 일어나는 경향이 있으며 서로 함께 작용할 때 그 결과를 악화시키는 상승작용을 하고 있다고 하였다.

현재까지 알려진 태아 성장에 미치는 요소로는 자궁 및 태반을 통한 혈류, 태아의 인슐린 (insulin), 그리고 포도당 등이 알려져 있으며(Vorherr, H., 1982), 자궁 및 태반의 순환장애에 기인한 태아의 저산소증이 발육제한태아의 주원인으로 알려져 있다.

그러나 지금까지 태아의 상태를 진단하는 방법으로 산모의 자궁고를 측정하거나, 초음파단층촬영에 의하여 태아의 체중을 측정하는 방법(Neilson, J.P., 1984; Geirsson, R.T., 1984) 등이 이용되고 있는데, 이와 같은 방법들은 판별할 수 있는 기준이 불명확하며, 더욱이 태아 및 산모에게 좋지 않은 영향을 끼칠 수 있는 침습적인 방법으로서 최근까지 이에 대한 많은 연구들이 진행되고 있다. 이에 반하여, 태아의 심박동(Fetal Heart Rate; FHR) 변이도(variability) 분석은 태아 가사상태나 저산소증, 태아산혈 증으로 인한 중추신경계의 기능저하가 있는 경우로 활용되고 있으며, 이는 비 침습적인 방법으로서 태아의 생체 신호만을 이용하는 기법으로 잘 알려져 있다.

이에 본 연구에서는 <그림 1>에서와 같이 HYFM-I, II에서 수집된 자료인 평균태아심박동수 (mean FHR), 태동수(Fetal Movement; FM), 기록손실도(Signal loss; Sloss), FHR 변이도(variability)인

박동간격(Mean Minute Range; MMR)과 진동폭(Amplitude; AMP), 임신주수(NST week), FHR 증가/감소 횟수(A1515/ D1515) 등의 기본 자료와 더불어 비선형적인 카오스 정량화 방법인 근사엔트로피(Approximate Entropy)를 추가변수로 사용하여, 최종적으로 의사결정나무 중 QUEST algorithm으로 판별모형을 개발하였다.



<그림 1> 임신고혈압증(PIH)과 관련한 병적태아의 진단시스템 흐름도.

## 2. 태아 심박동 분석시스템과 연구방법

본 절에서는 임신성고혈압증 태아의 판별모형 개발 연구를 위해, 태아 심박동 분석시스템(FHR monitoring system)과 비선형적 방법인 카오스(chaos) 정량화, 그리고 데이터마이닝 기법 중 의사결정나무(decision tree)에 대한 소개를 한다.

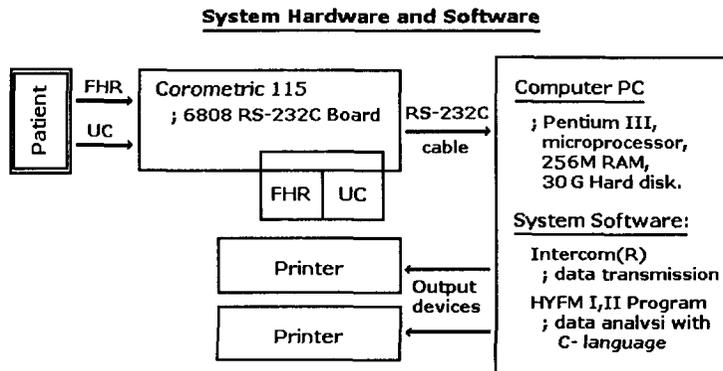
### 2.1. 태아 심박동 분석시스템

산과영역에서 분만 전에 태아의 상태를 정확히 평가한다는 것은 산모의 산전검사에서 거의 필수적인 과정이며, 현대 산과학에서 가장 중요한 과제이다. 그동안 이러한 목적으로 ‘전자태아심박동-자궁수축 감시법’이 연구되어 왔으며, 외국에서는 이미 태아심박동(FHR)의 제 변수값의 추출에서 새로운 알고리즘을 적용한 소프트웨어의 개발과 함께, 컴퓨터를 이용한 객관적인 판정을 시도한 많은 연구들이 있었다(Dawes 등, 1985; Searle 등, 1988; 박문일 등, 2001).

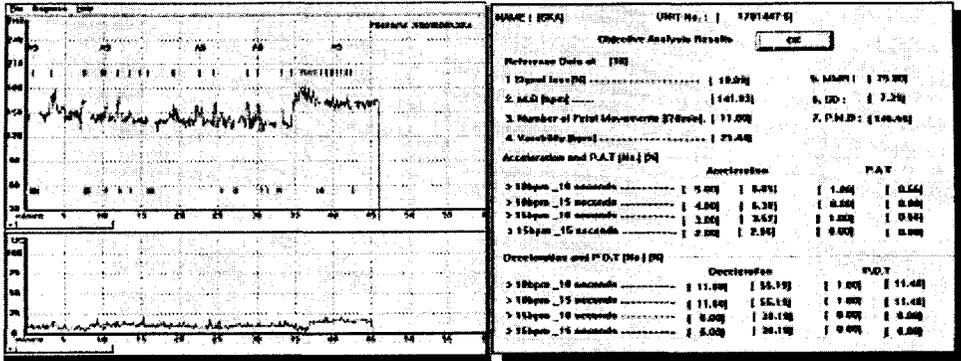
현재 분만 전 태아상태평가에 있어서 ‘전자태아심박-자궁수축 감시법’ 중에서도 대표적인 검사는 비수축성검사(Non-Stress Test; NST)인데, 이러한 검사는 현재 고위험도 임신에서는 물론, 최근에는 정상임신에서도 광범위하게 이용되고 있다.

이러한 분석은 컴퓨터시스템이 필수적인데, 현재 영국에서 Redman 등 (1993)에 의해 개발된 ‘옥스퍼드시스템’이 상용화되어 있다. 국내에서도 FHR 분석시스템을 개발하여 사용하고 있는 것이 바로 <그림 2>의 구조를 갖는 Hanyang Fetal Monitoring System(HYFM, 1998)이다(박문일 등, 2001). HYFM에서 사용되는 전자태아 심박 자궁수축감시 장치는 Corometric 115 model(USA)이다.

태아심박동의 각 변수와 각 자료 처리과정으로 우선, 기록손실도(loss of signal)의 계산(박문일 등, 2001)이 이루어지며, 기록손실로 처리된 자료를 제외한 후 나머지 자료를 이용하여 기본태아심박동수가 계산 된다. 다음으로, 태아심박동증가시의 각 조합은 태아심박동의 증가가 지속된 시간과 증가폭의 4가지 조합으로서 설정한 후 비교한다. 마지막으로 변이도 계산은 Root mean square, coefficient



<그림 2> 태아심박동(FHR) 분석시스템의 하드웨어와 소프트웨어의 구조도



<그림 3> FHR 곡선의 예(left)와 분석시스템 HYFM II의 결과화면(right)

of variation, MMR, AMP 등 네 가지의 변수로 측정된다(박문일 등, 2001). 이렇게 측정된 자료는 <그림 3>에서 보는바와 같이 1999년 1월 2일(Filename: 99010201.ska)에 한양대학교 산부인과에 내원한 산모로서 임신 28주에 NST검사를 받은 HYFM의 몇 가지 결과와 함께 표출된다.

## 2.2. 근사엔트로피(Approximate Entropy)

카오스 이론은 Poincare(1854~1912)에 의하여 제기되었으며 자연의 불규칙성(irregularity), 비예측성(non-prediction) 그리고 복잡성(complexity) 등을 다루며 자연의 복잡성 속에 숨어있는 보편적 질서구조와 생성원리를 이해하는 분야이다(Freeman 등, 1990).

일반적으로 카오스 정량화 방법에는 불확정성 및 정보이론에 입각한 확률과정적 엔트로피가 Shanon(Pincus, 1992)에 의해 제기되면서 불규칙성과 복잡성 등을 정량화한 근사엔트로피(Approximate Entropy; ApEn)가 있고(Davidson, 1992), 시간지연(time delay) 방법을 이용하여 매립 공간(embedding space)에서 시스템의 자취를 재구성하는 상관차원(correlation dimension)(Packard 등, 1980; Takens, 1980)이 있다. 이밖에 Hurst(1951)에 의해서 제안된 허스트지수와 최근에 Peng(1994)에 의해 개발된 fluctuation 지수 등이 있다.

본 연구에서는 임신성고혈압증 태아의 판별을 위해 비선형적인 지표로 카오스 정량화 방법 중 에서 먼저, 근사엔트로피(ApEn)를 활용하였으며, 근사엔트로피는 다음과 같은 연산방법을 이용한다.

태아심박수의  $N$ 개의 자료점  $u(1), u(2), \dots, u(n)$  이 있다고 하자. 2개의 입력매개변수인  $m$ 과  $r$ 이  $ApEn(m, r, N)$ 을 계산하기 위해 고정한다(여기서  $m$ 은 매립차원,  $r$ 은 상태공간을  $1/r$ 등분으로 나타내는 변수,  $N$ 은 전체 자료수).

ApEn을 정의하기 위해 먼저  $\{u(i)\}$ 로부터  $x(i)=[u(i), \dots, u(i+m-1)]$ 로 정의되는  $x(i)$  벡터를 만든다. 이 벡터는  $i$ 번째 점에서 시작되는  $m$ 개의 연속적인  $u$ 값을 나타낸다.  $x(i)$ 와  $x(j)$  사이 거리를 나타내는  $d[x(i), x(j)]$ 는 각 스칼라 성분에 존재하는 최대거리로 정의된다. 이때

$x(1), x(2), \dots, x(N-m+1)$ 을 사용하여 각  $i \leq N-m+1$ 에서

$$C_i^m(r) = \frac{\{d[x(i), x(j)] \leq r \text{ 이 되는 } x(j) \text{ 의 개수}\}}{(N-m+1)}$$

가 되며,  $C_i^m(r)$ 는 관용오차(tolerance)  $r$ 내에  $m$ 차원의 기존 벡터의 어떤 형태가 반복되는 빈도(frequency) 또는 규칙성(regularity)의 확률이 된다. 또

$$\Phi^m(r) = \left\{ (N-m+1)^{-1} \sum_{i=1}^{N-m+1} \ln C_i^{m+1} \right\}$$

로 정의하며, 이때  $\ln$ 은 자연대수이다. 즉,  $\Phi^m(r)$ 은  $\ln C_i^m(r)$ 의 평균으로 정의될 때, 근사 엔트로피 ApEn은 다음과 같이 정의된다.

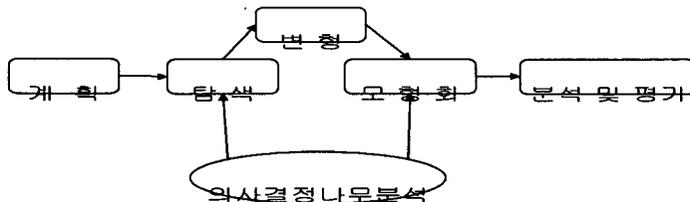
$$ApEn(m, r, N) = \Phi^m(r) - \Phi^{m+1}(r), \quad ApEn(m, r) = \lim_{N \rightarrow \infty} [\Phi^m(r) - \Phi^{m+1}(r)].$$

즉, 근사엔트로피는  $m$ 개의 관찰점 동안의 어떤 패턴이 다음  $m+1$ 개의 관찰점 동안에도 계속 유사한 확률을 갖고 있느냐 하는 logarithm화한 수이다. 따라서 규칙성의 가능성이 클수록 ApEn은 크고 불규칙성을 갖는 것일수록 ApEn은 작다.

### 2.3 데이터마이닝 기법과 의사결정나무

다양한 형태의 데이터로부터 의사결정에 유용한 정보 및 지식을 발견하려는 일련의 데이터분석 및 모형 선정과정을 데이터마이닝이라고 한다.

데이터마이닝에서 의사결정 과정은 계획, 탐색, 모형화, 분석 및 평가로 구성되며 <그림 4>와 같은 단계로 구성된다. 의사결정나무(decision tree)는 특히 탐색과 모형화에 유용한 통계적 기법이다. 의사결정나무는 의사결정규칙(decision rule)을 나무구조로 도표화하여 분류(classification)와 예측(prediction)을 수행하는 분석방법이다.



<그림 4> 데이터마이닝 과정에서 의사결정나무의 분석구조

의사결정나무는 Sonquist와 Morgan(1963)이 그 효시라고 할 수 있는데 AID (Automatic Interaction Detection)라는 컴퓨터 프로그램과 함께 의사결정나무의 역사가 시작되었다. 1980년대 이후 지금까지

지 CHAID(Kass, 1980), CART(Breiman, Friedman, Olshen and Stone, 1984), C4.5(Quinlan, 1993)와 QUEST(Loh and Shih, 1997)의 알고리즘이 널리 사용되고 있다.

의사결정나무는 누락된 관측 값에 대한 처리가 다른 모델보다 우수하고 변수들 간의 교호작용(interaction)의 설명과 처리가 용이하다는 장점이 있다. 반면에 의사결정나무에서는 선형(linear) 또는 주효과(main effect) 모형에서와 같은 결과를 얻을 수 없다는 한계점이 있다.

한편, 본 연구에서 사용되는 QUEST(Quick, Unbiased, Efficient, Statistical Tree)는 빠르게(Quick) 변수선택 편의를 줄이기 위해(Unbiased) 변수선택의 단계와 선택된 변수에 기초한 분리기준을 찾는 단계를 나누는 특징을 가지고 있다. 더불어 이지분리(binary split)를 수행하는 알고리즘으로, 변수선택 편의(bias)나 계산시간을 줄이는 방법으로 관측치의 수가 많거나 복잡한 데이터에 효과적이다.

### 3. 실증분석 및 결과

1999년 1월부터 2001년 12월까지 한양대학교 부속병원 산부인과를 내왕한 산모 중, 분만전 비수축 검사(Nonstress Test; NST)를 20분 시행 받은 산모 2,187명 중에서 임신 30-35주에 해당되는 쌍태 임신과 발육제한증(IUGR), 전치태반(placenta previa), 태반조기박리(abruptio placenta), 태아 곤란증(fetal distress), 둔위(breech), 기형(fetal anomaly), 당뇨병(GDM) 그리고 태아빈맥(getal tachycardia) 등 비정상 요소를 제외한 정상태아 200명과 목표변수 임신성고혈압(PIH)이 있는 태아 100명을 최종 연구대상으로 하였다. 설명변수로는 NST 측정 주수(NST weeks), Sloss, 평균 태아 심박동(Mean FHR), FHR 변이도(AMP, MMR), FHR 증감 횟수(A1515, D1515) 그리고 근사 엔트로피(ApEn)를 사용하였다.

태아심박동과 제 변수 자료는 한양대학교 산부인과에서 개발된 FHR분석 시스템인 HYFM-II를 이용하여 수집 처리되었고, 의사결정 알고리즘인 QUEST는 S-PLUS 2000을 사용하여 작성하였으며, 데이터의 통계분석은 SAS version 8.2를 이용하였고, 그룹간 유의성 검정은 Student's *t*-test를 적용하였으며, 나무모형 형성을 위한 training data는 임신성고혈압 66명, 정상 134명을 사용하였으며, 모형 검증을 위한 test data는 임신성고혈압 34명, 정상 66명을 대상으로 하였다.

임신성고혈압태아군 ( $N=100$ )과 정상태아군 ( $N=200$ ) 두 그룹 간 설명변수 차이에 대한 결과는 <표 1>과 같다. 즉, 근사엔트로피(approximate entropy)의 경우 정상태아와 임신성고혈압태아에 대하여 각각  $1.01(\pm 0.12)$ ,  $0.67(\pm 0.15)$ 로서 임신성고혈압태아 보다 정상태아가 통계적으로 유의하게 높게 나타났다 ( $p < 0.0001$ ), 또한, FHR 변이도인 AMP에서는 각각 16.98, 16.01이었으며, MMR은 49.54, 47.17으로 정상태아와 임신성고혈압태아 간의 유의한 차이가 없는 것으로 나타났다.

이와 같은 분석을 거쳐 확정된 유의한 변수와 의학적으로 활용 가능한 변수로 데이터마이닝 기법인 의사결정나무로 정상태아와 임신성고혈압태아에 대한 분류와 예측을 시행하였다. <그림 6>은 S-QUEST를 이용하여 training data 200명에 의해 나온 결과를 나무모형으로 나타낸 것이다. <그림

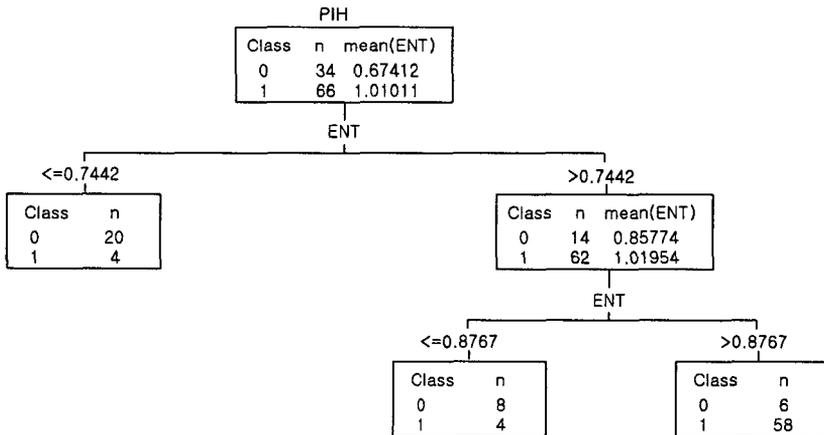
6>에서 근사 엔트로피는 임신성고혈압이 있는 태아를 설명하는 변수이며, 이는 <표 1>에서 살펴본 통계적 유의성 검정 결과와 일치한다.

<표 1> 정상 및 임신성고혈압증 태아의 심박동에 관한 통계적 지표

	PIH(N=100)	Normal(N=200)	p-value
NST weeks(wks)	32.17±1.34	32.04±1.30	0.3039
Sloss(%)	4.02±6.07	4.75±6.21	0.3609
Mean FHR(bpm)	140.62±6.42	141.51±5.42	0.4017
AMP(bpm)	16.01±4.07	16.98±4.12	0.0811
MMR(msec)	47.17±12.70	49.54±13.44	0.1042
A1515(no./20min)	1.71±2.19	1.68±2.31	0.9496
D1515(no./20min)	0.32±0.95	0.33±1.01	0.9345
Approximate Entropy	0.67±0.15	1.01±0.12	<.0001

임신성고혈압을 진단하기 위한 유의한 설명변수로 근사 엔트로피로 나타났다. 즉, 근사 엔트로피가 0.7442이하이면 임신성고혈압 태아로 판단되며, 근사 엔트로피가 0.8767이상이면 태아는 정상인 태아로 판단되고, 근사 엔트로피가 0.7442이상이고 0.8767이하이면 임신성고혈압 태아로 판단되었다.

한편, 근사 엔트로피가 0.7442이상이고 0.8767이하이면 임신성고혈압태아로 판단되었으나, 정상태아 4명과 임신성고혈압 태아 8명으로 표본의 수가 작기 때문에 쉽게 판단을 내릴 수 없으며, 따라서 앞에서 제시한 설명변수로는 진단이 어렵기 때문에 다른 방법이 필요로 한다.



<그림 5> S-QUEST로 분석한 결과를 구성한 나무 모형

<표 2> S-QUEST 실행한 결과에 대한 오분류표(Test Set)

관측	0 (PIH)	1 (정상)	분류정확%
0 (PIH)	28	6	82%
1 (정상)	8	58	88%

<표 2>는 S-QUEST를 이용하여 training data에 의해 도출된 모형을 test data에 적용한 오분류표이다. 임신성고혈압태아군 (N=34)에 대하여 82%의 예측력을 보였으며, 정상태아군 (N=66)에 대

하여도 88%의 높은 예측 결과를 보였다.

#### 4. 결론 및 고찰

본 연구는 생물정보학의 한 분야로서 산과학의 고위험 임신의 한 예로 정상태아와 임신성고혈압태아 두 그룹간 분류를 위해, 카오스 정량화인 근사엔트로피(ApEn)를 이용하여, 의사결정나무 분류모형으로 정상태아와 임신성고혈압태아를 분류하였다.

근사엔트로피란 심혈관계의 복잡성(complexity), 불규칙성(irregularity), 무작위성(randomness)을 정량화하여, 미래의 상태를 예측하는 수학적 알고리즘이다. Pincus 등(1992)에 의하면 ApEn는 전체 심박동수가 어떤 패턴을 가질 확률을 나타내는 것이며 이 확률이 낮다는 뜻은 두 개의 심박동수를 알았을 때 그 다음 세 번째 심박동수의 방향(증가 또는 감소)과 동시에 그 증감의 크기를 예측하기 어렵다는 것이다. 이것은 또한 심박동수를 발생시킨 과정이 random process일 가능성이 크며(방향의 불확실성), 동시에 nonlinear process(증가, 감소정도의 불확실성)일 가능성이 크다는 것을 의미한다.

두 그룹 간 통계적인 비교분석 결과, ApEn의 값이 정상태아에서  $1.01(\pm 0.12)$ 이고, 임신성고혈압태아에서  $0.67(\pm 0.15)$ 로 임신성고혈압태아가 낮게 나타났으며, 임신 30주에서 35주간 모든 구간에서도 같은 결과를 나타내었다 ( $p < 0.0001$ ).

본 연구에서 의사결정나무를 적용 분석한 결과, 정상 태아는 88%, 임신성고혈압태아는 82%로 정확하게 분류하였다. 이와 같은 결과는 생물통계학 분야에 의사결정나무를 적용하여 소기의 목적을 달성하였던 바, 이와 관련한 활발한 연구의 필요성을 갖는다. 또한, 본 연구 결과를 이용하여 출산 전에 생체신호만을 이용하여 임신성고혈압증과 관련된 태아를 진단하는 과정을 나무구조로 도표화하여 표현함으로써 보다 쉽게 진단이 가능하게 되고, 출산 전에 태아의 상태를 쉽게 판단할 수 있다. 출산 전에 태아의 건강상태를 빠르게 판단하므로 건강한 태아를 임신한 산모는 편안하게 출산 할 것이며, 비정상적인 태아를 임신한 산모는 출산 전에 태아의 치료가 시도되어 보다 근원적인 치료가 가능할 것이다.

본 연구에서 임신성고혈압증과 관련한 태아를 설명하는 변수로 근사엔트로피가 가장 중요한 변수로 분석되었는데, 앞에서 살펴본 바와 같이 근사엔트로피 0.7442이상 0.8767이하인 구간에선 근사엔트로피 만으로 정확한 진단이 어렵기 때문에, 좀 더 유의한 통계적 알고리즘 개발이 요구된다고 하겠다.

#### 참고문헌

- [1] Breiman, L. et al.(1984), *Classification and Regression Trees*, Wadsworth, Belmont.
- [2] Davidson, SR., Rankin, JHG., Martin, JCB., et al.(1992). Fetal heart rate variability and behavior state: analysis by power spectrum. *Am J. Obset. Gynecol.* Vol. 167, pp. 717-722.



- [3] Dawes, GS. Redman, CWG., Smith, JH. (1985), Improvements in the registration and analysis of fetal heart records at the bed side. *Br J Obstet Gynecol*, 92, pp. 317.
- [4] Frawley, W.J. et al.(1991), *Knowledge discovery in databases*, AAAI Press/The MIT Press.
- [5] Freeman, WJ., Skarda, CA. (1990). *Mind/Brain Science. Neuroscience on philosophy of Mind. Chap.7 in Lepore E, van Gulick R(eds) John Searle and his, Critics.* Oxford, Blackwell. pp. 377-428.
- [6] Geirsson, R.T. et al.(1984), Diagnosis of intrauterine growth retardation using ultrasound, *Clin Obstet Gynecol*, Vol. 11, pp. 457-479.
- [7] Harvey, D. et al.(1982), Abilities of children who were small for gestational age babies, *Pediatrics*, Vol. 69, pp. 296-300.
- [8] Hurst, HE.(1951) Long term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers*, Vol. 166, pp. 770-808.
- [9] Kass, G.V.(1980). An exploratory technique for investigating large quantities of categorical data, *Applied Statistics*, Vol. 29, pp. 119-127.
- [10] Loh, W.Y. and Shih, Y.S.(1997), Split selection methods for classification tree, *Statistica Sinica*, Vol. 7, pp. 815-840
- [11] M.I. Park, J.H. Hwang, K.J. Cha, Y.S. Park, S.K. Koh. (2001), Computerized analysis of fetal heart rate parameters by gestational age. *International Journal of Gynecology & Obstetrics*. 74, pp. 157-164.
- [12] Neilson, J.P. et al.(1984), Screening for small for dates fetuses: A controlled trial. *Br Med J*, Vol. 289, pp.1179-82.
- [13] Packard, NH., Crutchfield, JP., Farmer, JD., Shaw, RS.(1980). Geometry from a Time Series, *Phys. Rev. Let.* Vol. 45, pp. 712-716.
- [14] Peng, CK., Buldyrev, SV., Havlin, S., et al(1994). On the mosaic organization of DNA sequences. *Phys Rev E*, Vol. 49, pp. 1685-1689.
- [15] Pincus, SM.(1992), Approximating Markov chains, *Proc. Natl. Acad. Sci. USA*. Vol. 89, pp. 4432-36.
- [16] Precht HFR.(1967), Neurological sequelae of prenatal and perinatal complications. *BMJ*, 14, pp. 763-70.
- [17] Quinlan, J.R.(1993), *C4.5: Programs for machine learning*. Morgan Kaufmann, Los Altos.
- [18] Redman, CWG. (1993), Communicating the significance of the fetal heart rate record to the user. *Br J Obstet Gynaecol*, 100, pp. 24.
- [19] Searle, JR., DeVoe, LD., Phillips, MC. et al. (1988), Computerized analysis of resting fetal heart rate tracings. *Obstet Gynecol*, 71, pp. 407.
- [20] Sonquist, J. A. and Morgan, J. N. (1963). Problems in the analysis of survey data and a proposal, *Journal of American Statistical Association*, Vol. 58, 415-434.
- [21] Takens, F.(1980). *Detecting Strange Attractors in Turbulence, volume 898 of Lecture Notes in Mathematics (Dynamical Systems and Turbulence, Warwick)*, pp. 366-381.
- [22] Vorherr, H.(1982), Factors influencing fetal growth. *Am J Obstet Gynecol*, Vol. 142, pp. 577-88.