



Unsupervised Clustering of Late Preterm Infants in Terms of Developmental Outcome

Ho Kim, MD, MS^{1,2}, Jun Hwan Song, MD, PhD^{1,2}, Seung Soo Kim, MD, PhD¹

¹Department of Pediatrics, Soonchunhyang University Cheonan Hospital, Cheonan; ²Regional Newborn Intensive Care Center, Soonchunhyang University Cheonan Hospital, Cheonan, Korea

Objective: This study aimed to derive the subtype of late preterm infants (gestational age, 34 to 36 weeks) according to their developmental outcomes.

Methods: We retrospectively investigated the medical records of premature infants who had undergone developmental testing and were discharged from a single regional newborn intensive care center. We used 5 domains (motor, language, cognition, social-emotional, adaptive behavior) of the Korean version of the Bayley scale of infant and toddler development III (K-Bayley III) to group subjects. K-means clustering (KM), hierarchical clustering, and density-based spatial clustering of applications with noise were used. We used the average silhouette index (ASI) and Calinski-Harabasz (C-H) score as evaluation metrics.

Results: KM showed the best performance (ASI, 0.25; C-H score, 58.83) and revealed 3 clusters. Cluster 1 (need observation) showed low normal scores in K-Bayley III Scales, and cluster 2 (excellent development) showed high normal scores. In contrast, cluster 3 (global delay) showed delayed or borderline scores other than the social-emotional scale. Maternal age ($P<0.01$), number of fetuses ($P=0.03$), prenatal steroid use ($P=0.01$), pH ($P<0.01$), and base excess ($P=0.03$) showed a statistical significance among the 3 clusters.

Conclusion: The authors found 3 phenotypes with distinct developmental outcomes among late preterm infants and discovered variables necessary for their prediction. If the target group, requiring developmental testing, can be screened early by using these predictors, it may be beneficial in improving the developmental prognosis of late preterm infants.

Key Words: Premature infants, Child development, Cluster analysis, Unsupervised machine learning

Received: 11 April 2022

Revised: 11 May 2022

Accepted: 18 May 2022

Correspondence to

Seung Soo Kim, MD, PhD
Department of Pediatrics,
Soonchunhyang University Cheonan
Hospital, 31 Suncheonhyang 6-gil,
Dongnam-gu, Cheonan 31151, Korea

Tel: +82-41-570-2160

Fax: +82-41-572-4996

E-mail: equalkss@schmc.ac.kr

Copyright© 2022 The Korean Society of
Perinatology

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided that the original work is properly cited.

서론

재태 연령 34주에서 36주 사이의 후기 미숙아(late preterm infant)는 최근 빠르게 늘고 있으며, 국내 미숙아 중 차지하는 비율이 78% 이상으로 보고되어 있다.¹ 그동안 후기 미숙아는 신경 발달 예후가 좋을 것으로 생각되어 왔으며, 상대적으로 적절한 추적관찰이 이루어지지 않는 경우가 많았다.² 하지만 37명의 후기 미숙아를 대상으로 한 국내 연구에서는 24.3%가 학동기 지능지수(intelligence quotient)가 70에서 85로 경계성(borderline) 범위에 해당되었으며, 65%가 주의력 부족 행동 과다 장애(attention deficit hyperactivity disorder) 선별검사서 이상 소견을 보인 것으로 보고하기도 하였다.² 이렇듯 최근 후기 미숙아에서도 신경 발달에 대한 추적관찰 필요성이 대두되고 있지만, 아직 국내외에 후기 미숙아의 신경 발달 예후에 대한 연구는 많지 않은 실정이다.³ 기존 연구들은 주로 재태 연령 31주 미만의 초기 미숙아나 출생체중 1,500 g 미만의 극소 저출생 체중아(very low birth weight infant)를 대상으로 삼았으며, 여러 발달 예후 지표 중 일부를 결과 변수로 선택해서 예측 모형을 만드는 형태가 많았다.⁴ 하지만 이러한 형태의 예측 모형은 복잡하고 다양한 발달 예후 지표들을 아우르기는 어렵다는 한계를 가진다. 최근 이에 대한 대안으로 군집화

(clustering)나 잠재계층분석(latent class analysis)을 이용해 균일한 집단으로 여겨졌던 대상군을 여러가지 하부 그룹으로 나누는 연구들이 증가하고 있으며, 미숙아의 발달 예후와 관련해서도 아직은 많지 않지만 관련 연구들이 나타나기 시작했다.^{5,6}

본 저자들은 비지도 기계학습(unsupervised machine learning) 방법인 군집화를 이용해서 발달 예후에 따른 후기 미숙아의 표현형을 분류하여, 밀접한 추적관찰이 필요한 대상의 조기 선별에 필요한 정보를 얻고자 본 연구를 수행하였다.

방법

1. 대상

본원 신생아집중치료실에서 입원 치료 후 퇴원한 출생 시 재태 연령 34주 이상에서 37주 미만인 영유아 중, 2016년 9월 1일부터 2022년 2월 28일까지 교정연령 6개월에서 16개월 사이에 한국형 베이리 영유아 발달검사 3판(Korean version of Bayley scale of infant and toddler development, K-Bayley-III) 검사를 받은 156명을 대상으로 했다.⁷ K-Bayley-III의 5가지 척도 중 하나라도 결측값이 있는 경우는 대상에서 제외했다. 본 연구는 순천향대학교 천안 병원 임상시험심사위원회의 승인을 받았다(SCHCA2022-04-031).

2. 인구통계학적 특성 및 혈액 표지자, K-Bayley-III 검사 결과

대상군의 의무기록을 바탕으로 산모의 연령, 영유아의 성별, 태아 수, 조기양막파수, 임신고혈압, 임신당뇨, 산전 스테로이드 투여, 산전 황산마그네슘(MgSO₄) 투여, 재태 연령, 출생체중, 아프가 점수(1분, 5분), 출생 체온, 신생아중환자실(neonatal intensive care unit, NICU) 입원기간과 입원 첫 날 전체혈구계산(complete blood count) 검사에서 백혈구(white blood cell) 수, 적혈구크기 분포폭(red blood cell distribution width), 평균혈소판용적(mean platelet volume)과 원대 출생아의 경우 출생 1시간 이내, 원대 출생아의 경우 입원 1시간 이내의 혈액가스분석(blood gas analysis, BGA)의 pH와 염기과잉(base excess, BE), 그리고 호흡곤란증후군(respiratory distress syndrome)과 뇌초음파로 관찰된 뇌출혈 여부를 후향적으로 조사했다. 또한 재태 연령에 따른 출생체중의 표준 점수(Z score)를 구했다.⁸ 그리고 교정연령 6개월에서 16개월의 K-Bayley-III 운동척도(motor scale), 언어척도(language scale), 인지척도(cognition scale), 사회정서척도(social-emotional scale) 및 적응행동척도(adaptive behavior scale) 결과도 의무기록에서 후향적으로 조사했다.

3. 기술적 통계 분석 및 예측 모형의 개발과 평가

본 분석 전에 재태 연령에 따라 G34 (34주 0일에서 34주 6일), G35 (35주 0일에서 35주 6일), G36 (36주 0일에서 36주 6일)의 세 그룹으로 분류하고, 태아기 및 출생 시, NICU 입원 과거력, 입원 시 검사실 검사 결과, 입원기간과 K-Bayley-III 검사 시 교정연령과 그 결과 등의 변수들의 대표값과 통계적 차이를 분석했다. 일표본 Kolmogorov-Smirnov 검정을 사용하여 정규분포 여부를 분석하고, 정규분포를 따르는 경우는 대표값을 평균±표준편차로 표시하고 정규분포를 따르지 않는 경우는 대표값을 중위수(사분범위)로 표시하였다. 정규분포를 따르는 연속형 변수들의 분석에는 독립 일원 배치분산 분석(one-way ANOVA) 후 Bonferroni 사후 검정(post hoc analysis)을 시행했으며, 정규분포를 따르지 않는 연속형 변수들의 분석에는 Kruskal-Wallis 검정 후 Wilcoxon-Mann-Whitney 검정의 P값을 Bonferroni 방법으로 교정하여 사후 검정을 했다. 범주형 자료들은 명(%)으로 표시하였으며, 카이제곱 검정을 사용해 분석했다.

군집화는 (1) 분석 자료의 전처리(data preprocessing)와 (2) 군집화 분석(cluster analysis), (3) 군집 간 변수들의 통계적 차이 및 효과 크기(effect size) 분석의 세 단계로 이루어졌다.

(1) 전처리 과정에서는, 군집에 사용할 K-Bayley-III의 5가지 척도 중 하나라도 결측치(missing value)가 있는 대상(sample)을 분석에서 제거했다. 다음으로 각 변수의 값을 표준화(standardization)시켜, 평균이 0이고 분산이 1이 되도록 변환했다.

(2) 군집화 분석 단계에서는 먼저 열지도(heat map)를 통해 각 변수 간의 상관관계를 알아보았다(Fig. 1). 이후 k평균 군집화(K-means clustering, KM)와 계층적 군집화(hierarchical clustering, HC), 밀도 기반 군집화(density-based spatial clustering of applications with noise, DBSCAN)의 세 가지 방법을 사용해 군집화를 수행했으며, 평균 실루엣 지수(average silhouette index)와 Calinski-Harabasz 점수를 평가지표로 사용해서 적절한 군집의 개수와 군집화 방법을 선택했다.⁹

실루엣 점수는 한 표본과 표본이 속한 군집과 가장 가까운 군집 내의 표본들 간의 거리 평균($b(\bar{x}_i)$)에서 표본과 표본이 속한 군집 내 다른 표본들 간의 거리 평균($a(\bar{x}_i)$)을 빼고, 이 값을 두 값 중 큰 것으로 나누어 구했다.⁹

$$s(\bar{x}_i) = \frac{b(\bar{x}_i) - a(\bar{x}_i)}{\max(a(\bar{x}_i), b(\bar{x}_i))}$$

Calinski-Harabasz 점수는 전체 표본 수(M)에서 군집의 수(k)를 빼고, 이것을 군집의 수(k)에서 1을 뺀 수로 나눈 후, 다시 여기에 전체 자료의 중심(centroid)에 대한 각 군집의 중심들의 분산(between cluster dispersion, BCD)을 군집내 중심에 대한 각 표본들의 분산의 합(within cluster dispersion, WCD)으로 나눈 값을 곱하여 구

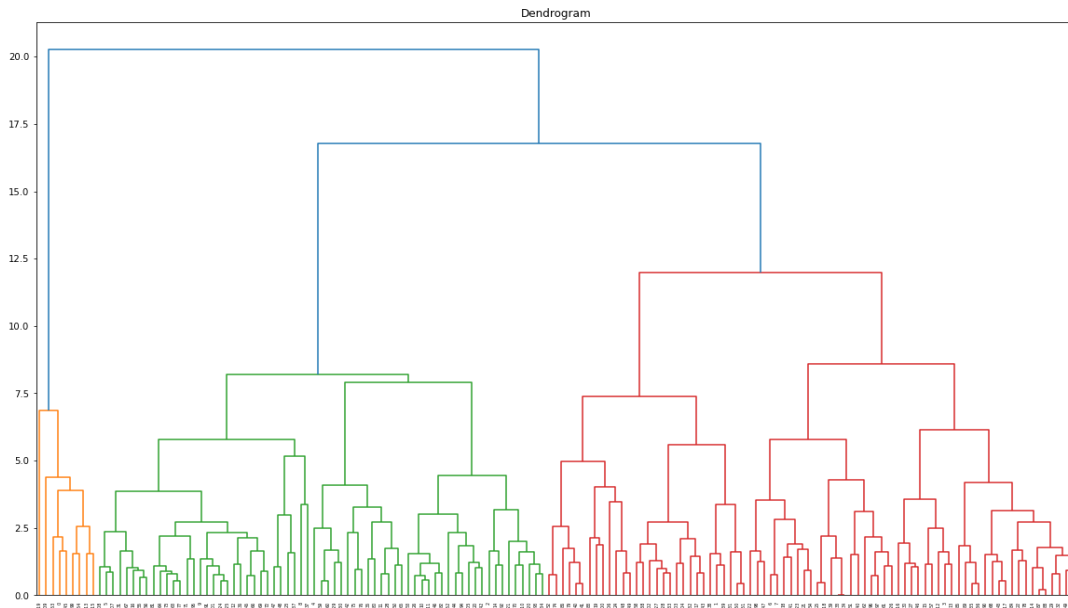


Fig. 1. Heat map for 5 scales of Korean Bayley scales of infant development III (K-Bayley-III). Except for the social-emotional scale, the remaining scales of K-Bayley III showed a moderate correlation with each other.

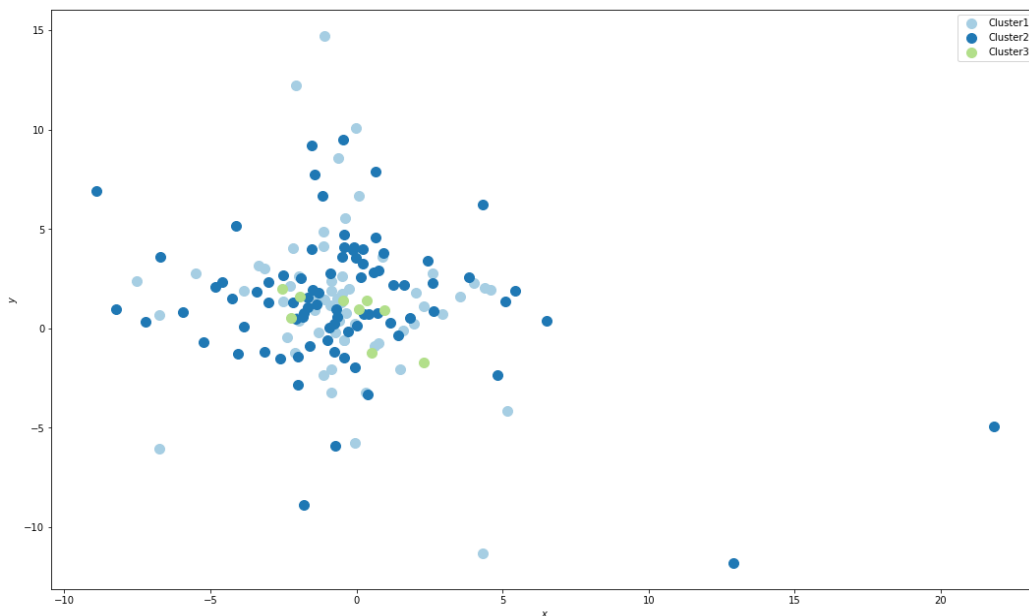


Fig. 2. Dendrogram of hierarchical clustering. The hierarchical clustering algorithm showed the best performance on 3 clusters.

했다.⁹

$$CH_k(X, Y_{pred}) = \frac{M - k}{k - 1} \cdot \frac{BCD(k)}{WCD(k)}$$

실루엣 점수(범위 -1에서 1)는 값이 클수록 이상적인 분류가 이루어진 상황으로 해석했으며, Calinski-Harabasz 점수 역시 값이 클수록 분류가 잘 수행된 것으로 해석했다.⁹ KM과 HC에서는 격자 검색

색(grid search) 방법으로 여러 군집 수를 대입한 후 가장 우수한 평가지표를 보여준 값을 최종 군집 수로 결정했으며, DBSCAN에서는 이웃의 수를 입력 변수의 2배인 10으로 설정한 상태에서 K-최근접 이웃(K-nearest neighbors) 알고리즘을 이용해 구한 거리에 팔꿈치 법(elbow method)을 적용해 군집화 반경(epsilon) 값을 결정했다.⁹

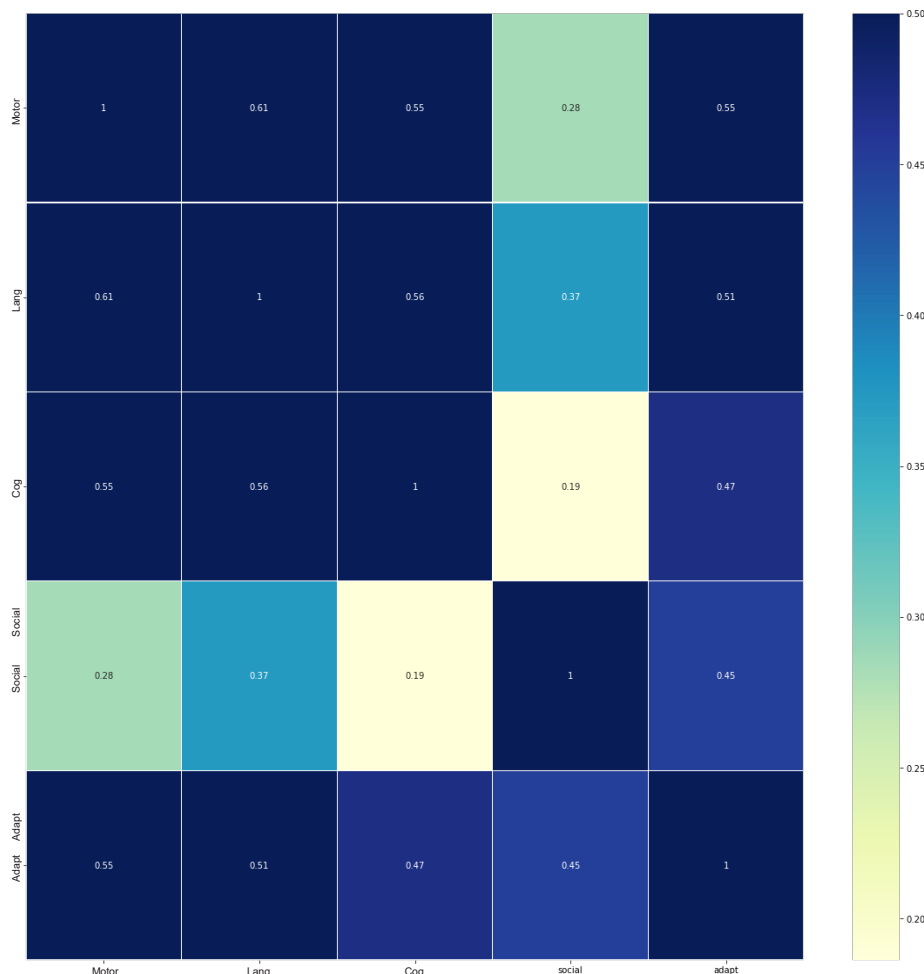


Fig. 3. t-distributed stochastic neighbor embedding (t-SNE) plot with cluster by k-means clustering. Three clusters classified using the k-means clustering algorithm are displayed on the t-SNE plot drawn using 5 scales of Korean Bayley scales of infant development III.

최종 결정된 초매개변수(hyperparameter)는 다음과 같았다.

- A. K-means clustering: number of groups=3
- B. Hierarchical clustering: number of groups=3 (Fig. 2)
- C. DBSCAN: epsilon=2, minimum points=10, metrics=minkowski, p=4

(3) 최종 선택된 군집 별로 각 변수 간의 차이를 알아보고, 통계적으로 유의한 차이를 보이는 변수의($P < 0.05$) 효과 크기를 구했다. 정규분포를 따르는 연속형 변수에는 Cohen's d (범위 0에서 1)를 사용했으며, 정규분포를 따르지 않는 경우에는 Cliff's delta (범위 -1에서 1)를 사용했다. 범주형 자료들은 Cramer's V를 사용해서 군집과 변수 간의 상관관계의 크기를 알아보았다. 마지막으로 직관적인 이해를 위해 군집화에 사용된 변수들의 차원을 2차원으로 축소한 t-SNE (t-distributed stochastic neighbor embedding) 그림에 분류된 군

집을 표시하여 제시하였다(Fig. 3).

4. 통계 프로그램 및 패키지

기술통계 분석에는 IBM SPSS ver. 25.0 (IBM Corp., Armonk, NY, USA)을 이용했으며, 분석 자료의 전 처리 및 군집화 분석, 효과 크기 계산, t-SNE 그림에는 python 3.6과 패키지 sklearn 0.23.2, scipy 1.5.4, matplotlib 3.2.1, pandas 0.25.3, numpy 1.16.4, cliffs-delta 1.0.0을 이용했다.

결과

1. 기술적 통계

최종적으로 156명의 대상군 중 K-Bayley III의 언어척도 값이 누

Table 1. Patient Characteristics Stratified by Gestational Age Group

Characteristic	Missing (%)	G34	G35	G36	P-value
Total		75	54	26	
Age of mother (yr)		32 (29, 36)	31 (30, 35)	34 (30, 38)	0.46
Sex (boy)		40 (53.3)	30 (55.6)	13 (50.0)	0.90
No. of fetus		1 (1, 2)	1 (1, 2)	1 (1, 2)	0.82
Premature rupture of membrane	4.5	14 (18.9)	9 (17.6)	2 (8.7)	0.51
Pregnancy induced hypertension	4.5	9 (12.2)	11 (21.6)	2 (8.7)	0.23
Gestational diabetes mellitus	4.5	16 (21.6)	5 (9.8)	3 (13.0)	0.19
Prenatal steroid	5.8	60 (81.1)	32 (64.0)	13 (59.1)	0.04
Prenatal MgSO ₄	5.8	40 (54.1)	21 (42.0)	4 (18.2)	0.01
Birth weight (g)		2,160 (1,790, 2,380) ^{a)}	2,290 (2,050, 2,550)	2,370 (2,080, 2,680) ^{a)}	0.01
Birth weight (z score)		-0.4323±1.0838	-0.5479±0.9153	-0.8327±0.8295	0.21
Vaginal delivery		8 (10.7)	9 (16.7)	3 (11.5)	0.59
Apgar score (1 min)	1.9	6 (5, 7) ^{b)}	7 (6, 8)	8 (6, 9) ^{b)}	0.01
Apgar score (5 min)	1.9	8 (7, 9)	9 (8, 10)	9 (7, 10)	0.16
Body temperature (1st)		36.9 (36.7, 37.0)	36.8 (36.5, 37.0)	36.8 (36.6, 37.0)	0.31
pH (1st)		7.3±0.1	7.3±0.1	7.3±0.1	0.50
Base excess (1st)		-4.5 (-6.0, -2.9)	-4.8 (-6.9, -3.3)	-3.7 (-6.4, -1.7)	0.28
White blood cell (1st)		11,106±2,989	11,622±3,439	11,249±3,834	0.57
Red cell distribution width (1st)		16.0 (15.3, 17.3)	16.2 (15.4, 17.5)	15.9 (15.5, 17.1)	0.76
Mean platelet volume (1st)	0.6	9.6±0.7	9.7±0.7	9.6±0.6	0.76
Respiratory distress syndrome		12 (16.0)	5 (9.3)	2 (7.7)	0.38
Intracranial hemorrhage	6.5	10 (13.9)	5 (10.2)	5 (20.8)	0.47
Length of admission (day)		17 (11, 26)	11 (8, 15)	10 (7, 14)	<0.01
Age at K-Bayley-III (mo)		9 (9, 10)	9 (8, 10)	9 (8, 10)	0.71
K-Bayley-III					
Cognition scale		90.0 (82.5, 95.0)	90.0 (85.0, 95.0)	90.0 (85.0, 95.0)	0.30
Language scale		97.0 (91.0, 100.0)	97.0 (91.0, 97.0)	97.0 (91.0, 100.0)	0.44
Motor scale		97.0 (91.0, 107.0)	97.0 (91.0, 103.0)	94.0 (86.5, 103.0)	0.57
Social-emotional scale		100.0 (95.0, 110.0)	95.0 (85.0, 110.0)	95.0 (85.0, 105.0)	0.02
Adaptive behavior scale		95.0±9.3	94.9±9.6	91.8±8.9	0.33

Non-parametric variables are presented as median (interquartile range) and statistical testing performed using Kruskal-Wallis test. Parametric variables are presented as mean±standard deviation and statistical testing performed using one-way ANOVA. Categorical variables are presented as number (%) and statistical testing performed using the chi-square test. The percentage of missing values is zero unless otherwise shown.

G34, gestational age 34 weeks 0 days to 36 weeks 6 days; G35, gestational age 35 weeks 0 days to 35 weeks 6 days; G36, gestational age 36 weeks 0 days to 36 weeks 6 days; K-Bayley-III, Korean version of Bayley scale of infant and toddler development III.

^{a)}<0.05 in Bonferroni *post hoc* test. ^{b)}<0.05 in Wilcoxon-Mann-Whitney test followed by manual adjustment for P-value by Bonferroni method.

략된 1명을 제외한 155명의 대상군이 분석에 포함되었다(Table 1). G34군은 75명(48.4%), G35군은 54명(34.8%)이었으며, G36군은 26명(16.8%)이었다. G34군과 G35군, G36군 사이에서 산전 스테로이드 사용($P=0.04$), 산전 황산화마그네슘 사용($P=0.01$), 출생체중($P=0.01$), 1분 아프가 점수($P=0.01$) 및 K-Bayley-III의 사회정서척도($P=0.02$)에서 통계학적인 차이를 보였다. 사후 검정에서는 G34군과 G36군 사이에서 출생체중과 1분 아프가 점수가 통계적으로 차이

Table 2. Evaluation Metrics of each Approach

Approach	No. of clusters	ASI	C-H score	No. of excluded outliers
K-means clustering	3	0.25	58.83	NA
Hierarchical clustering	3	0.25	58.79	NA
DBSCAN	1	0.59	18.47	2

ASI, average Silhouette index; C-H score, Calinski-Harabasz score; DBSCAN, density-based spatial clustering of applications with noise; NA, not applicable.

($P < 0.05$)를 보였다.

2. 군집화

KM은 군집 수 3개에서 평균 실루엣 지수 0.25, Calinski-Harabasz 점수 58.83로 가장 좋은 성능을 보여주었으며, HC도 군

집 수 3개에서 평균 실루엣 지수 0.25, Calinski-Harabasz 점수 58.79로 좋은 성능을 보여주었다(Table 2, Fig. 3). DBSCAN은 평균 실루엣 지수 0.59, Calinski-Harabasz 점수 18.47을 보여주었으며, 1개의 군집과 2개의 noise points를 분류했다. 종합적으로 가장 우수한 성능을 보인 KM 방법으로 최종 군집을 결정했다.

Table 3. Patient Characteristics Stratified by Cluster by K-means Clustering

	Missing (%)	Cluster 1	Cluster 2	Cluster 3	P-value
Summary		Need observation	Excellent development	Global delay	
Total		63	83	9	
Age of mother (yr)		35 (30, 37) ^{a)}	31 (29, 35) ^{b)}	29 (25, 35) ^{b)}	<0.01
Gestational age (wk)		35.1 (34.4, 35.7)	34.9 (34.3, 35.4)	35.3 (34.5, 36.1)	0.42
Sex (boy)		35 (55.6)	43 (51.8)	5 (55.6)	0.90
No. of fetus		1 (1, 2) ^{a)}	1 (1, 2) ^{b)}	1 (1, 1) ^{b)}	0.03
Premature rupture of membrane	4.5	10 (16.9)	14 (17.5)	1 (11.1)	0.89
Pregnancy induced hypertension	4.5	10 (16.9)	10 (12.5)	2 (22.2)	0.63
Gestational diabetes mellitus	4.5	13 (22.0)	9 (11.3)	2 (22.2)	0.21
Prenatal steroid	5.8	43 (74.1)	60 (75.0)	2 (25.0)	0.01
Prenatal MgSO ₄	5.8	27 (46.6)	37 (46.3)	1 (12.5)	0.17
Birth weight (g)		2,227.0±434.5	2,224.6±456.9	2,435.6±393.7	0.39
Birth weight (z score)		-0.5989±0.9438	-0.5390±1.0226	-0.1320±1.0563	0.42
Mode of delivery (vaginal delivery)		6 (9.5)	12 (14.5)	2 (22.2)	0.47
Apgar score (1 min)	1.9	7 (6, 8)	6 (5, 8)	5 (4, 8)	0.23
Apgar score (5 min)	1.9	9 (8, 9)	9 (8, 10)	8 (6, 9)	0.15
Body temperature (1st)		36.8 (36.6, 37.0)	36.8 (36.7, 37.0)	36.8 (36.5, 37.2)	0.26
pH (1st)		7.3±0.1 ^{c)}	7.3±0.1	7.2±0.1 ^{c)}	< 0.01
Base excess (1st)		-4.0 (-5.5, -2.2) ^{b)}	-4.9 (-6.4, -2.9)	-6.2 (-8.3, -4.8) ^{b)}	0.03
White blood cell (1st)		11,198±3,208	11,406±3,418	11,004±2,759	0.90
Red cell distribution width (1st)		16.2 (15.3, 17.5)	15.9 (15.3, 17.3)	16.6 (15.7, 18.0)	0.61
Mean platelet volume (1st)	0.6	9.6 (9.2, 10.2)	9.6 (9.3, 10.0)	9.8 (9.3, 10.5)	0.76
Respiratory distress syndrome		5 (7.9)	13 (15.7)	1 (11.1)	0.37
Intracranial hemorrhage	6.5	7 (12.3)	13 (16.5)	0 (0)	0.36
Length of Admission (day)		15 (10, 21)	12 (9, 18)	11 (10, 24)	0.40
Age at K-Bayley-III (mo)		9 (8, 10)	9 (8, 10)	11 (9, 14)	0.11
K-Bayley-III					
Cognition scale		85.0 (80.0, 90.0) ^{a)}	95.0 (90.0, 95.0) ^{a)}	65.0 (57.5, 74.5) ^{a)}	<0.001
Language scale		91.0 (89.0, 97.0) ^{a)}	100.0 (97.0, 103.0) ^{a)}	77.0 (75.5, 79.0) ^{a)}	<0.001
Motor scale		94.0 (88.0, 97.0) ^{a)}	100.0 (97.0, 110.0) ^{a)}	64.0 (50.5, 73.0) ^{a)}	<0.001
Social-emotional scale		85.0 (80.0, 95.0) ^{b)}	105.0 (100.0, 115.0) ^{a)}	90.0 (82.5, 100.0) ^{b)}	<0.001
Adaptive behavior scale		88.0 (83.0, 94.0) ^{a)}	100.0 (96.0, 103.0) ^{a)}	75.0 (69.5, 86.0) ^{a)}	<0.001

Non-parametric variables are presented as median (interquartile range) and statistical testing performed using Kruskal-Wallis test. Parametric variables are presented as mean±standard deviation and statistical testing performed using one-way ANOVA. Categorical variables are presented as number (%) and statistical testing performed using the chi-square test. The percentage of missing values is zero unless otherwise shown.

K-Bayley-III, Korean version of Bayley scale of infant and toddler development III.

^{a)}<0.05 in in Wilcoxon-Mann-Whitney test followed by manual adjustment for P-value by Bonferroni method with every other group. ^{b)}<0.05 in in Wilcoxon-Mann-Whitney test followed by manual adjustment for P-value by Bonferroni method with one other group. ^{c)}<0.05 in Bonferroni *post hoc* test.

3. 군집간 변수 값의 차이와 효과 크기

군집 1 (관찰 필요)은 K-Bayley-III점수에서 정상에서 정상 하한(lower normal)을 보였다. 군집 2 (우수 발달)는 K-Bayley-III에서 정상에서 정상 상한(upper normal)을 보였다. 군집 3 (전반적 지연)은 사회정서영역을 제외한 나머지 4개 영역에서 지연 및 경계성을 보였다. 각 발달 영역은 세 그룹 간에 통계적인 차이($P<0.001$)를 보였으며, 사후 검정에서도 관찰 필요군과 전반적 지연군 사이의 사회정서영역 외에는 통계적으로 유의한 차이($P<0.05$)를 보였다. 나머지 변수 비교에서는 산모의 연령($P<0.01$), 태아의 수($P=0.03$), 산전 스테로이드 사용($P=0.01$), 출생 후 1시간 이내 BGA의 pH ($P<0.01$) 및 BE ($P=0.03$)에서 통계적인 차이를 보였다. 사후 검정에서는 관찰 필요군과 우수 발달군, 관찰 필요군과 전반적 지연군 사이의 산모 연령과 태아 수에서 통계적인 차이($P<0.05$)를 보였으며, 관찰 필요군과 전반적 지연 군사이에서는 pH와 BE에서 통계적인 차이($P<0.05$)를 보였다(Table 3).

각 군집을 분류하는 데 작용하는 각 변수의 효과 크기의 분석에서는 BE가 높은 효과 크기(군집 1 대 군집 2: -0.92, 군집 1 대 군집 3: -0.91, 군집 2 대 군집 3: -0.98)를 보여주었으며, pH는 군집 2 대 군집 3에서 높은 효과 크기(0.81)를 보여주었다(Table 4).

고찰

후기 신생아를 대상으로 한 본 연구에서 출생 주수에 따라 분류한 그룹(G34, G35, G36) 사이에서는 K-Bayley-III의 척도 중 사회정서

영역을 제외한 나머지 네 가지 영역에서는 통계적으로 유의한 차이를 보이지 않았다. 하지만 군집화 방법(KM)으로 분류한 세 그룹(관찰 필요, 우수 발달, 전반적 지연) 간에는 5가지 척도 모두에서 통계적으로 유의한 차이가 존재했으며, 군집화에 사용하지 않은 산모의 연령, 태아의 수, 산전 스테로이드 사용, 출생 후 1시간 이내 BGA의 pH 및 BE의 변수에서도 세 그룹 간의 통계적인 차이를 보여주었다. 반면 재태 연령이나 출생체중, 재태 연령에 따른 출생체중의 표준 점수, 아프가 점수(1분, 5분) 등은 세 그룹 간에 통계적으로 유의한 차이를 보이지 않았다.

최근 정밀의료(precision medicine)와 이에 따른 맞춤형료(personal medicine)가 의료계의 화두가 되면서, 그동안 단일한 집단으로 생각했던 환자군을 좀더 세분화된 하부그룹 혹은 표현형으로 분류하는 연구들이 많이 이루어지고 있다.¹⁰ 미숙아에서도 재태 연령 및 출생체중 등의 전통적인 분류 기준 외에 다양한 변수를 사용해서 세분화된 표현형을 찾아내려는 노력이 늘고 있다.^{5,6,10-18} 기존의 연구들은 주로 고위험 임신 및 출산을 예방하는 목적으로 주로 산모의 산전 병력과 태아의 주산기 병력을 변수로 사용한 분류가 많았다. 최근에는 미숙아의 발달에 초점을 맞춘 분류 연구도 발표되기 시작했으나, 아직은 그 수가 많지 않다.^{5,6} 특히 본 연구와 같이 재태 연령 34주에서 36주 사이 후기 미숙아의 발달 예후를 기준으로 표현형을 분류한 연구는 국내외에서 발표된 적이 없었다.⁴ 기존 미숙아 발달 예후 연구들은 대상군의 측면에서는 상대적으로 불량한 예후를 보이는 재태 연령 31주 이하의 초기 미숙아를 대상으로 한 연구가 대부분이었으며, 일부 재태 연령 32주에서 33주 사이의 중기 미숙아를 대상으로 한 연구들이 있었다.⁴ 예후 변수(outcome)의 측면에서는 상술한 소수의 표현형 연구들 외에는 주로 여러 결과 변수들 중 하나 이상에서 문제를 가진 경우를 예측하는 모형을 만들거나, 여러 결과 변수 별로 개별적인 예측 모형을 만드는 방식을 사용한 연구가 대부분이었다.⁴⁻⁶ 이러한 형태의 연구들은 다양한 예후 변수들을 전체적으로 아우르는 측면에서는 한계를 가질 수 밖에 없다.

새로운 표현형을 찾으려는 연구에서 가장 많이 사용되고 있는 것은 군집화로 대표되는 비지도 기계학습과 잠재계층분석과 같은 구조방정식 모형이다. 두 가지 방법 모두 연구자가 미리 지정하지 않은 하부그룹으로 연구대상군을 분류한다.^{9,19} 이중 군집화는 유사도(similarity)에 근거하여 표본들을 군집으로 구분하여 묶는 비지도 기계학습 방법으로 대규모의 데이터에도 적용할 수 있어 최근 사용이 늘고 있다. 우리에게 익숙한 주성분 분석(principal component analysis)이나 요인분석(factor analysis)이 변수들을 묶어서 자료를 축소하는 방법이라면, 군집화는 사례 혹은 표본들을 묶어서 자료를 축소하는 방법이다.⁹ 이러한 자료의 축소를 통해 우리의 자료의 복잡성을 줄이고, 표본들 간 관계의 패턴을 더욱 명확하게 파악할 수 있게 된다.²⁰ 군집화는 크게 분할 기반(partition based), 연

Table 4. Estimated Effect Size According to Cluster by K-means Clustering

	Reference: cluster 1		Reference: cluster 2
	Cluster 2	Cluster 3	Cluster 3
Age of mother ^{a)}	0.26	0.45	0.21
Number of fetus ^{a)}	0.17	0.37	0.19
Prenatal steroid ^{b)}		0.25	
pH (1st) ^{c)}	0.19	0.13	0.81
BE (1st) ^{a)}	-0.92	-0.91	-0.98
K-Bayley-III			
Cognition ^{a)}	-0.49	0.94	1.00
Language ^{a)}	-0.54	0.97	1.00
Motor ^{a)}	-0.54	0.94	0.99
Social-emotional ^{a)}	-0.75	-0.18	0.65
Adaptive behavior ^{a)}	-0.78	0.69	0.99

BE, base excess; K-Bayley-III, Korean version of Bayley scale of infant and toddler development III.

^{a)}Cliff's delta (range -1 to 1). ^{b)}Cramer's V. ^{c)}Cohen's d (range 0 to 1).

결 기반(linkage based), 밀도 기반(density based) 등으로 분류할 수 있다.⁹ 분할 기반 군집화는 평균값(KM) 혹은 중앙값(k-medoids clustering)을 중심으로 거리가 가까운 표본들을 같은 군집으로 분류한다.⁹ 이 중 KM 알고리즘은 오랜 시간 진화해온 기법으로, 간단한 원리를 이용해 군집화를 수행하고 간단한 수정으로 매우 유연하게 결점을 극복할 수 있다는 장점을 가진다.²⁰ 이 덕분에 KM 알고리즘이 개발된 이후에도 군집화 기법은 꾸준히 발달해 왔지만, KM 알고리즘은 여전히 가장 많이 사용되는 군집화 기법이다.²⁰ 본 연구에서도 KM 알고리즘을 사용한 군집화가 평균 실루엣 지수 및 Calinski-Harabasz 점수를 종합적으로 고려했을 때 가장 우수한 성능을 보여주었으며, 이 군집들은 우리가 가진 인식에도 잘 부합되었다. 연결 기반 군집화의 하나인 HC는 시작 단계에서는 각각의 표본들을 하나의 독립된 군집으로 간주하고, 이 중 비유사성이 가장 낮은 두 군집을 하나로 통합한다. 이후 모든 표본이 하나의 군집으로 통합될 때까지 위의 과정을 반복한다.⁹ 밀도 기반 군집화인 DBSCAN은 군집 안에 포함된 표본의 개수 즉 밀도를 바탕으로 군집을 정의한다. 우선 임의의 출발점에서 지정된 거리 ϵ (에타)내의 이웃을 추출한다. 여기에 지정된 최소 숫자(minimum points, n_{min}) 이상의 표본이 포함되어 있다면 군집으로 편입하고, 거리 바깥의 표본은 이상치(noise)로 지정한다. 새로 편입된 표본에 대해서도 동일 과정을 되풀이하여 군집의 영역을 확장해 간다. 이를 모든 표본을 방문할 때까지 반복한다. 군집의 수를 미리 지정할 필요는 없지만, ϵ 와 n_{min} 은 연구자가 결정해야 한다.⁹ DBSCAN은 알고리즘의 특성 상 t-SNE 그림에서 뚜렷이 구분되는 기하학적인 모양을 보여주는 자료에서 좋은 성능을 보이는 것으로 알려져 있으며, 본 연구에서는 유용한 군집을 분류해내지 못했다. 본 연구에서는 KM과 HC 모두 세 개의 군집으로 분류했을 때 가장 좋은 성능을 보여주었으며, 이는 본 연구에서 최종적으로 채택한 세 개의 표현형이 다른 알고리즘에서도 일관되게 나타난다는 방증이 라 하겠다.

기존 신생아 발달 예측 연구에서 예측 변수로 자주 사용되었던 pH 및 BE, 산모의 연령이 본 연구에서도 각 표현형 간에 차이를 보이는 것으로 나타났으며, 특히 이 중 BE와 pH의 효과 크기가 큰 것으로 나타났다.⁴ 반면 재태 연령, 출생체중, 아프가 점수 등은 각 표현형 간에 통계적으로 유의한 차이를 보이지 않았으며, 이는 후기 미숙아의 재태 연령이나 출생체중, 아프가 점수 등이 비교적 균일하고, 발달 예측에 영향을 줄 수 있는 절단 값(cut-off)보다 이미 높기 때문에 벌어지는 현상으로 생각된다.⁴ 하지만 대상군이 비교적 작은 본 연구의 특성 상 각 표현형 간의 비교를 위해 사용한 비모수적인 추정 방법의 한계에서 기인한 현상일 가능성도 염두에 두어야 하겠으며, 이는 추후 대규모 코호트 연구에서 확인이 필요할 것으로 생각된다.

2018년 Johnson 등⁵은 1,139명의 미숙아(재태 연령 32-36주)를 대상으로 교정연령 24개월의 발달 및 행동 문제 등의 변수에 잠

재계층모형을 적용하여 시행한 연구에서 3개의 잠재계층(healthy outcome, term-born phenotype, preterm phenotype)을 보고 하였다. 이는 각각 본 연구에서 찾아낸 3개의 표현형(우수 발달, 관찰 필요, 전반적 지연)과 대응하는 측면이 많다. 결과적으로 방법론과 대상군이 다른 연구에서도 본 연구와 비슷한 표현형이 관찰된 것으로 이해할 수 있으며, 본 연구의 결과를 강력하게 뒷받침하는 증거로 볼 수 있다.

본 연구는 단일 기관의 의료 기록을 바탕으로 한 후향적 단면 연구로의 한계를 가진다. 또한 상대적으로 대상군의 수가 충분하지 않아 충분한 수의 변수를 군집화에 사용하지 못했다는 한계도 지닌다. 그리고 후향적 단면연구의 한계 상 전향적으로 수집한 평가데이터를 사용한 임상검증(clinical validation) 또한 이루어지지 못했다. 또한 발달검사를 시행한 연령이 교정연령 6개월에서 16개월 사이로 비교적 균일하지 않다는 한계 또한 지닌다. 하지만 후기 신생아 중 발달 예측에서 뚜렷한 차이를 보이는 세 가지 표현형을 찾아냈으며, 이의 예측에 필요한 변수들을 발굴해냈다는 것에 본 연구의 의의를 둘 수 있다. 이 예측 변수들을 활용하여 발달 검사가 필요한 대상군을 조기 선별할 수 있다면, 후기 미숙아의 발달 예측 향상에 도움이 될 것으로 생각된다. 나아가 본 연구를 바탕으로 추후 대규모 코호트 자료를 이용한 연구와 충분한 임상검증이 이루어진다면, 국가적인 차원의 맞춤형 발달 검사 계획 수립에도 도움이 될 것이다.

ORCID

Ho Kim, <https://orcid.org/0000-0003-1444-5628>

Jun Hwan Song, <https://orcid.org/0000-0002-9330-9140>

Seung Soo Kim, <https://orcid.org/0000-0003-4276-9777>

Conflict of Interest

No potential conflict of interest relevant to this article was reported.

Acknowledgments

Following was results of a study on the "HPC Support" Project, supported by the 'Ministry of Science and ICT' and NIPA.

Authors' Contributions

Conceptualization: HK; Data curation: HK, JHS; Formal analysis: HK, SSK; Investigation: all author; Methodology: HK, SSK; Project administration: HK, SSK; Resources: HK, JHS; Visualization: HK, SSK; Writing-original draft: all authors; Writing-review & editing: all authors.

References

- 1) Park SH, Kim JS, Lim DO. Secular trend of gestational age specific preterm birth rate in Korean singleton and multiple birth: 1997-98, 2013-14. *J Health Info Stat* 2017;42:10-5.
- 2) Jin JH, Yoon SW, Song J, Kim SW, Chung HJ. Long-term cognitive, executive, and behavioral outcomes of moderate and late preterm at school age. *Clin Exp Pediatr* 2020;63:219-25.
- 3) Kim ES. What do we know about the long-term cognitive and behavioral outcomes of school-aged children who were born moderate to late preterm? *Clin Exp Pediatr* 2020;63:215-6.
- 4) Crilly CJ, Haneuse S, Litt JS. Predicting the outcomes of preterm neonates beyond the neonatal intensive care unit: what are we missing? *Pediatr Res* 2021;89:426-45.
- 5) Johnson S, Waheed G, Manktelow BN, Field DJ, Marlow N, Draper ES, et al. Differentiating the preterm phenotype: distinct profiles of cognitive and behavioral development following late and moderately preterm birth. *J Pediatr* 2018;193:85-92.
- 6) Villar J, Restrepo-Mendez MC, McGready R, Barros FC, Victora CG, Munim S, et al. Association between preterm-birth phenotypes and differential morbidity, growth, and neurodevelopment at age 2 years: results from the INTERBIO-21st newborn study. *JAMA Pediatr* 2021;175:483-93.
- 7) Weiss L, Oakland T, Aylward G. Bayley-III clinical use and interpretation. Elsevier Inc.; 2010.
- 8) Fenton TR, Sauve RS. Using the LMS method to calculate z-scores for the Fenton preterm infant growth chart. *Eur J Clin Nutr* 2007;61:1380-5.
- 9) Bonaccorso G. Hands-on unsupervised learning with Python: implement machine learning and deep learning models using Scikit-Learn, TensorFlow, and more: Birmingham: Packt Publishing; 2019.
- 10) Kim KP. Applying precision medicine in clinical practice. *Korean J Med* 2020;95:382-6.
- 11) Barros FC, Papageorgiou AT, Victora CG, Noble JA, Pang R, Iams J, et al. The distribution of clinical phenotypes of preterm birth syndrome: implications for prevention. *JAMA Pediatr* 2015;169:220-9.
- 12) Hendryx M, Chojenta C, Byles JE. Latent class analysis of low birth weight and preterm delivery among Australian women. *J Pediatr* 2020; 218:42-8.
- 13) MacBean V, Lunt A, Drysdale SB, Yarzi MN, Rafferty GF, Greenough A. Predicting healthcare outcomes in prematurely born infants using cluster analysis. *Pediatr Pulmonol* 2018;53:1067-72.
- 14) Markasz L, Olsson KW, Holmstrom G, Sindelar R. Cluster analysis of early postnatal biochemical markers may predict development of retinopathy of prematurity. *Transl Vis Sci Technol* 2020;9:14.
- 15) Mukhopadhyay S, Underwood MA. Phenotyping preterm infants at birth to predict infection risk. *Pediatr Res* 2021;90:508-9.
- 16) Souza RT, Cecatti JG, Passini R Jr, Pacagnella RC, Oliveira PF, Silva CM; Brazilian Multicentre Study on Preterm Birth study group. Cluster analysis identifying clinical phenotypes of preterm birth and related maternal and neonatal outcomes from the Brazilian multicentre study on preterm birth. *Int J Gynaecol Obstet* 2019;146:110-7.
- 17) Strobel NA, Richardson A, Shepherd CC, McAuley KE, Marriott R, Edmond KM, et al. Modelling factors for Aboriginal and Torres Strait Islander child neurodevelopment outcomes: a latent class analysis. *Paediatr Perinat Epidemiol* 2020;34:48-59.
- 18) Villar J, Papageorgiou AT, Knight HE, Gravett MG, Iams J, Waller SA, et al. The preterm birth syndrome: a prototype phenotypic classification. *Am J Obstet Gynecol* 2012;206:119-23.
- 19) Collins LM, Lanza ST. Latent class and latent transition analysis: with applications in the social, behavioral, and health sciences. New Jersey: Wiley; 2013.
- 20) Lantz B. Machine learning with R: expert techniques for predictive modeling to solve all your data analysis problems. Yun SJ, translator. 2nd ed. Seoul: Acorn Publishing Co.; 2017.