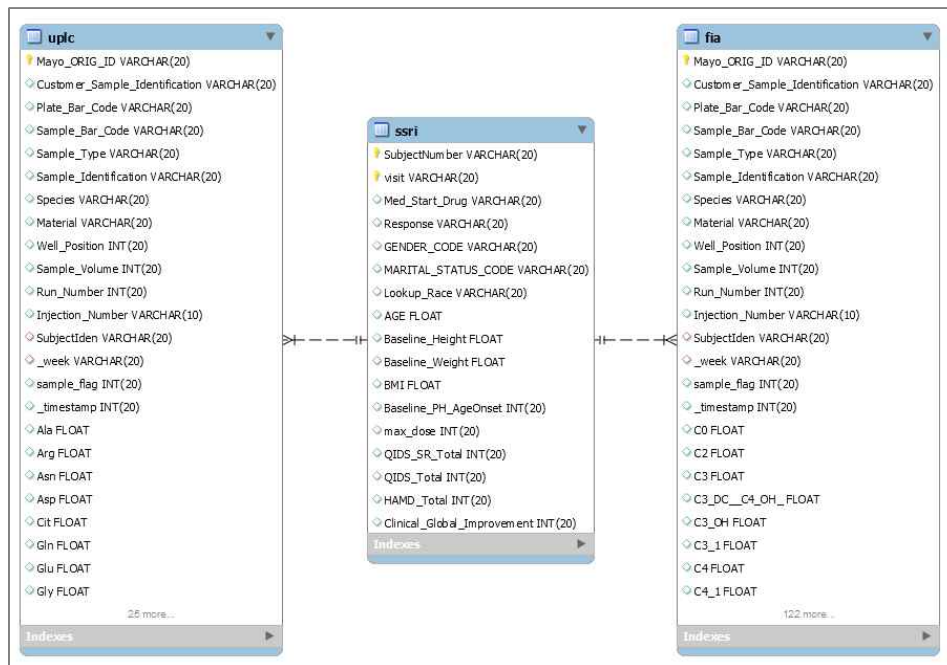


## 1. 데이터

- 우울증 환자의 설문 및 임상 분석 결과에 대한 데이터
- FIA 및 UPLC 데이터는 우울증 환자에 대한 임상 데이터 분석 결과 및 설문에 대해 나타내며, SSRI 데이터는 각 환자에 대해 기간별로 우울증 정도를 HAMD 수치로 나타내고 치료 여부를 Response를 통해서 나타냄. (Response : 호전, Non-Response : 비 호전)
- week 0, 4, 8로 기간을 나누어 HAMD를 측정하고 처음보다 50% 이상이 낮아졌을 경우 우울증이 치료된 것으로 판단함

### 1) E-R 다이어그램



### 2) FIA

- 데이터 수 : 578개
- Attribute 수 : 146개 (공통 16개)
- 결측치 : 없음

Attribute	추측 내용	데이터 예시
Mayo.ORIG.ID	기관 ID	1205206454, 1205208240
SubjectIden	환자 ID	SSRI275, SSRI276, SSRI261
...	...	...
week	치료 기간	week0, week4, week8
C0	임상 데이터 분석 결과	5.14699265148743, 5.10973303291835
SM.C26.0	"	-1.95670364004852, -2.38609643230444
...	...	...
SM.C26.1	"	-1.21725298078695, -1.42304343478627

### 3) UPLC

- 데이터 수 : 578개
- Attribute 수 : 49개 (  : 공통 16개)
- 결측치 : 없음

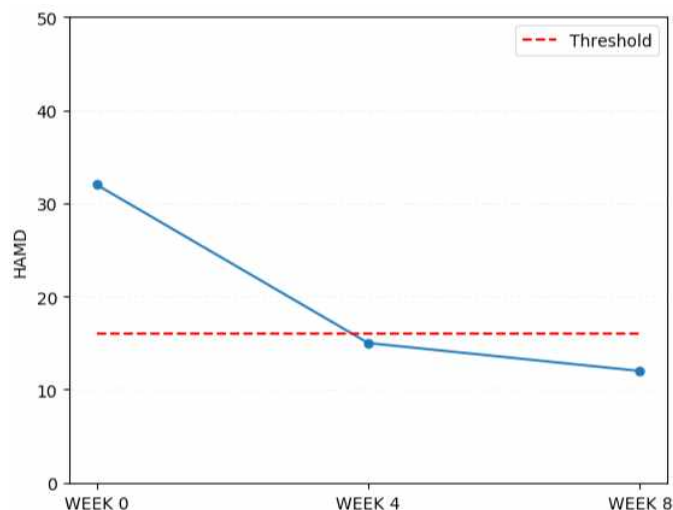
Attribute	추측 내용	데이터 예시
Mayo.ORIG.ID	기관 ID	1205206454, 1205208240
SubjectIden	환자 ID	SSRI275, SSRI276, SSRI261
...	...	...
week	치료 기간	week0, week4, week8
Ala	임상 데이터 분석 결과	8.2921162622033, 8.53846124494396
t4.OH.Pro	"	3.11986654726298, 3.28222382158747
...	...	...
Taurine	"	5.37736519775916, 6.75243185043716

### 4) SSRI

- 데이터 수 : 918개
- Attribute 수 : 17개
- 결측치 : Baseline\_Height - 150개, Baseline\_Weight - 177개, BMI - 177개,  
Clinical Global Improvement - 307개

Attribute	추측 내용	데이터 예시
SubjectNumber	환자 ID	SSRI007, SSRI051
visit	치료 기간	week0, week4, week8
Response	반응	Non-Response, Response
BMI	BMI 치수	33.07, 21.84
...	...	...
QIDS Total	QIDS 계	22, 4, 0
HAMD Total	우울증 수치	32, 11, 4

ex. SSRI357 환자에 대한 week 별 HAMD 치수



## 2. 테스트 데이터 구축

- FIA 및 UPLC 데이터에서 공통되는 16개의 attribute를 제외한 나머지 attribute와 우울증 여부를 나타내는 Response 값을 target으로 하는 데이터 구축
- FIA, UPLC, FIA + UPLC의 3가지 경우로 Attribute를 나눔
  - (a) FIA : (C0, C2, C3, ... , SM.C26.0, SM.C26.1, Response)
  - (b) UPLC : (Ala, Arg, Asn, ... , t4.OH.Pro, Taurine, Response)
  - (c) FIA + UPLC : (C0, C2, ... , SM.C26.1, Ala, Arg, ... , Taurine, Response)
- \* 사용된 Attribute 수는 각각 131, 34, 164개
- week 0의 경우, 우울증인 환자들을 나타내므로 Non-Response를 나타냄

### 1) 임상 데이터 분석 결과 및 설문 데이터를 통한 환자의 우울증 여부 예측

- 모든 데이터에 대해서 Response를 target으로 하여 우울증 여부를 예측하는 binary 문제를 구성함

	Attribute 수	데이터 수	Non-Response	Response
FIA	131	578	388	190
UPLC	34	578	388	190
FIA + UPLC	164	578	388	190

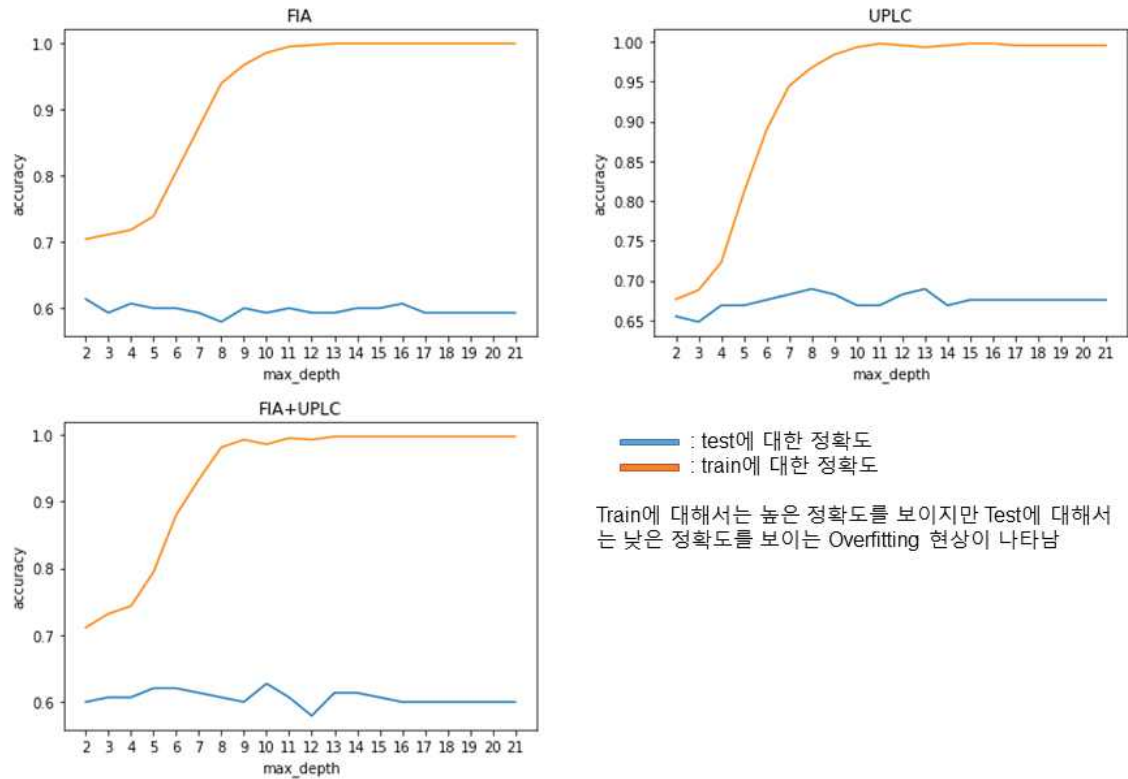
### 2) 치료 기간에 따른 우울증 치료 여부 예측

- week 0 ~ week 4까지의 우울증 치료 여부와 week 0 ~ week 8까지의 우울증 치료 여부를 비교하기 위한 데이터 구성

	데이터 수		
	Week 0	Week 4	Week 8
FIA	264	162	152
UPLC	264	162	152
FIA + UPLC	264	162	152
Non-Response	264	79	45
Response	0	83	107

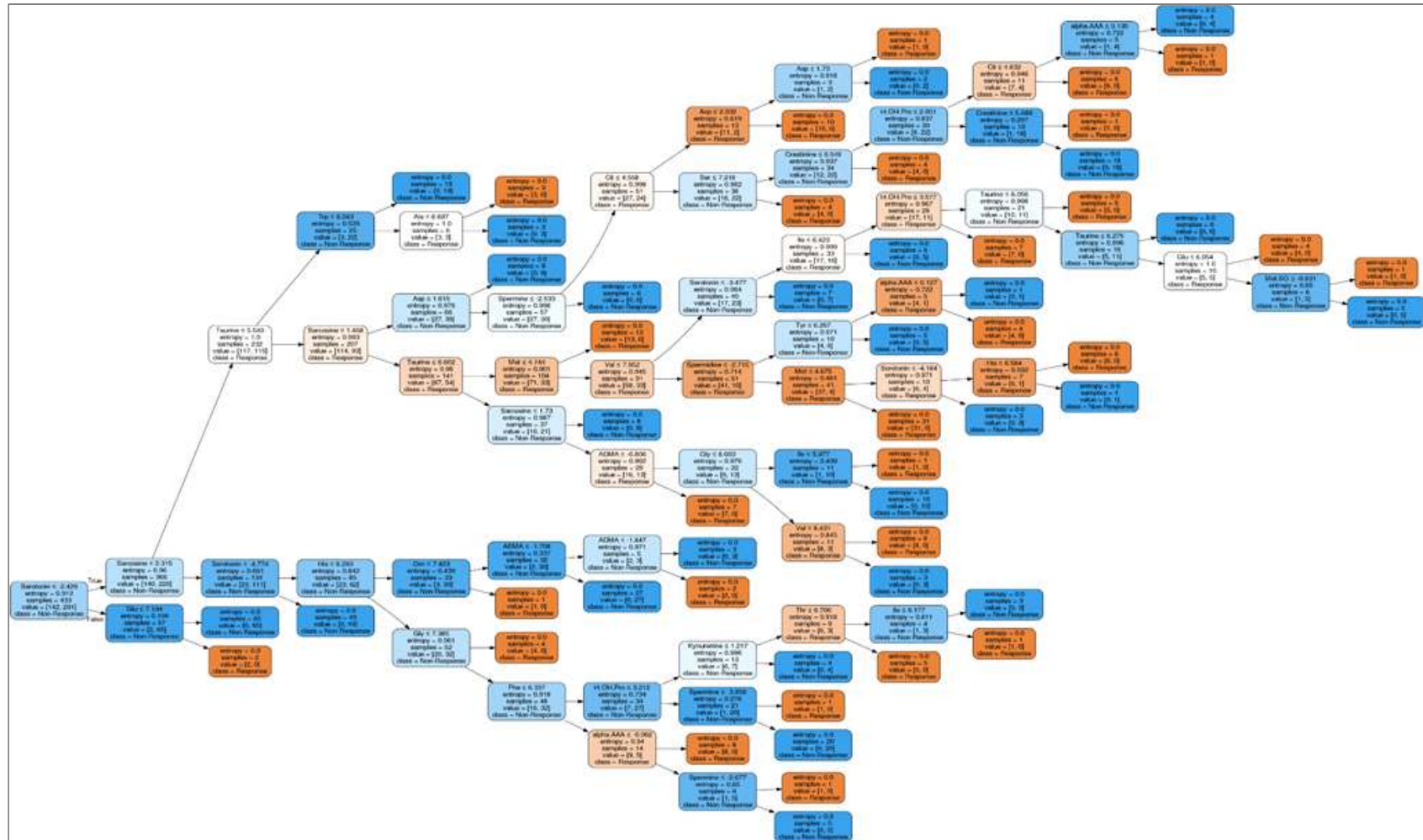
### 3. Decision Tree

#### 1) 성능 비교

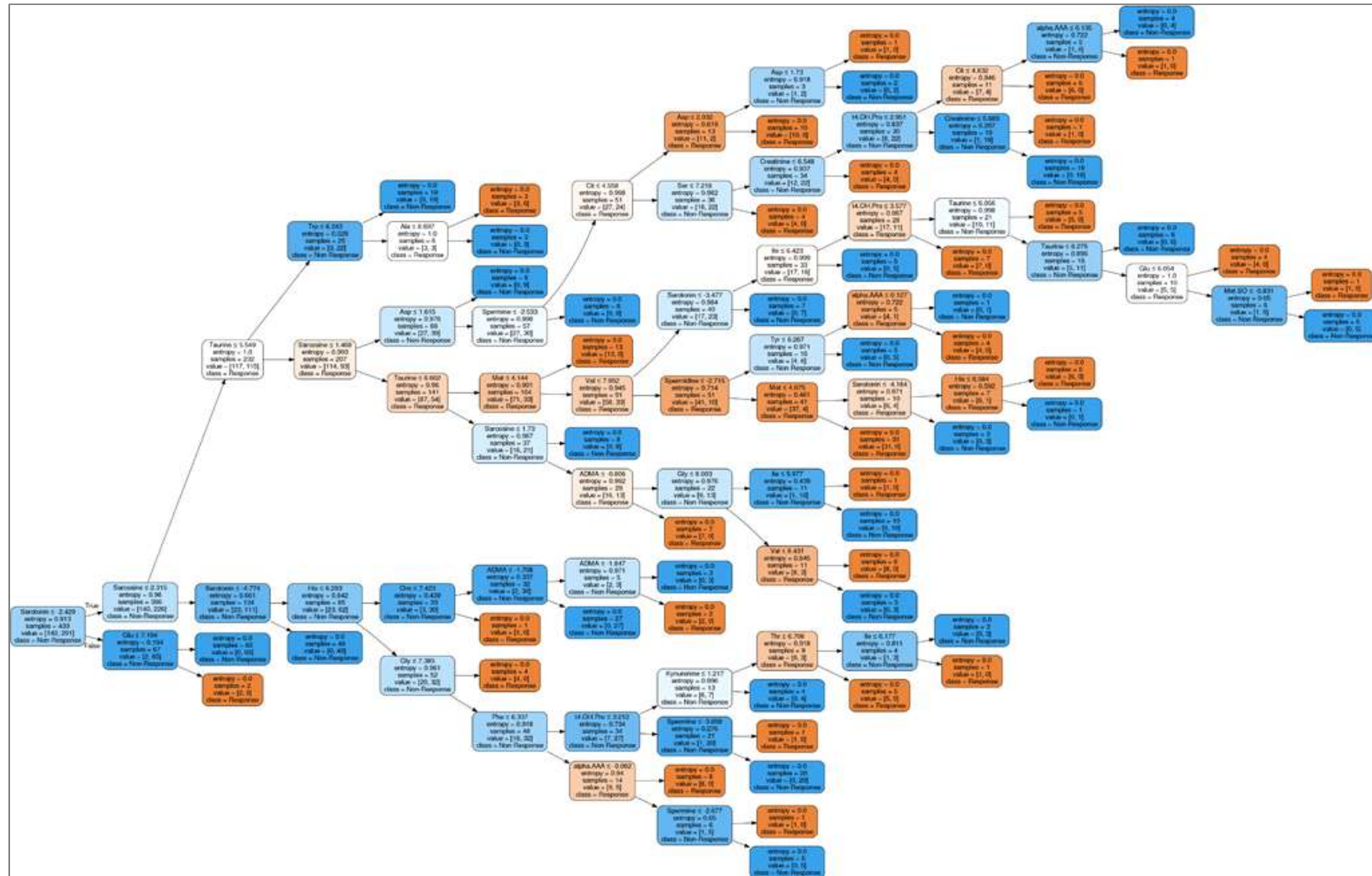


		Max_Depth												
		2	3	4	5	6	7	8	9	10	11	12	13	14
Train	fia	67.7%	68.8%	70.0%	76.2%	85.2%	91.2%	96.1%	98.8%	99.3%	99.5%	99.5%	99.5%	99.8%
	uplc	68.1%	71.4%	74.4%	79.2%	87.8%	93.8%	96.8%	98.4%	99.1%	98.8%	98.8%	98.8%	98.8%
	fia+uplc	67.0%	70.7%	73.4%	81.8%	89.4%	95.6%	97.0%	98.8%	98.8%	99.8%	99.8%	99.8%	99.8%
Test	fia	68.3%	69.0%	69.0%	69.0%	69.7%	70.3%	70.3%	71.7%	69.7%	70.3%	70.3%	70.7%	70.3%
	uplc	65.5%	65.5%	70.3%	71.7%	70.3%	70.3%	70.3%	69.7%	69.7%	69.7%	67.6%	69.0%	68.3%
	fia+uplc	65.5%	64.8%	64.8%	64.1%	64.1%	64.1%	64.1%	66.2%	66.2%	66.9%	66.2%	66.2%	65.5%

a) FIA



b) UPLC





c) FIA + UPLC

