

# 뇌파의 의사 결정 트리 분석과 가능성 기반 서포트 벡터 머신 분석을 통한 우울증 환자의 분류

## (EEG Classification for depression patients using decision tree and possibilistic support vector machines)

심우현<sup>1</sup>, 이기영<sup>1</sup>, 채정호<sup>2</sup>, 정재승<sup>1</sup>, 이도현<sup>1</sup>

<sup>1</sup>한국과학기술원 바이오시스템학과

<sup>2</sup>가톨릭 의과대학교 신경정신과

### 초 록

우울증은 가장 유병율이 높은 ‘기분 장애’(mood disorder)의 일종으로, 약 20%의 인구가 일생동안 우울증 증상을 한번쯤 경험한다. 이러한 우울증은 크게 ‘우울 장애’(major depressive disorder)와 ‘양극성 장애’(bipolar disorder)로 구분된다. 환자의 질병 분류에 따라 사용되는 약과 의학적 처방이 다르기 때문에, 우울증 환자의 빠르고 정확한 진단 및 분류는 매우 중요하다. 기존의 다면성 인성검사(MMPI)와 같은 통계적인 방법이 우울증 환자의 진단을 위해 사용돼 왔으나, 장시간의 집중력을 요구하기 때문에 집중력 저하의 특징을 보이는 우울증 환자들에게 적용하는데 어려움이 있다. 이 논문에서는 이러한 문제를 해결하고자, 빠른 측정이 가능하고 측정동안 집중력을 요하지 않는 EEG 데이터의 분석을 통해 우울증 환자의 분류를 시도하였다. EEG 채널 간 정보 흐름에서의 비선형성과 근사 엔트로피(approximate entropy)의 크기를 속성(attribute)으로 사용하여 데이터 마이닝 기법 중 의사 결정 트리(decision tree)와 가능성 기반 서포트 벡터머신(possibilistic support vector machines) 통해 분석을 수행하였다. 30명의 주요 우울장애환자와 24명의 양극성 장애 환자를 통해 위의 분석을 수행한 결과 의사 결정 트리의 경우 85.19%의 정확도를 가지며 분류해냈고, 가능성 기반 서포트 벡터머신의 경우 77.78%의 정확도를 보여줬다. 본 연구는 가능성 기반 서포트 벡터 머신 분석이 우울증 환자는 진단하고 분류하는데 유용하게 적용될 수 있는 가능성을 제시하고 있다.

**키워드:** 우울증, 비선형도 분석, 근사 엔트로피, 데이터 마이닝, 의사 결정 트리, 가능성 기반 서포트 벡터머신

### Abstract

Depression is the most common and widespread mood disorder. About 20% of the population might suffer a major, incapacitating episode of depression during their lifetime. This disorder can be classified into two types: major depressive disorders and bipolar disorder. Since pharmaceutical treatments are different according to types of depression disorders, correct and fast classification is quite critical for depression patients. Yet, classical statistical method, such as minnesota multiphasic personality inventory (MMPI), have some difficulties in applying to depression patients, because the patients suffer from concentration. We used electroencephalogram (EEG) analysis method for classification of depression. We extracted nonlinearity of information flows between channels and estimated approximate entropy (ApEn) for the EEG at each channel. Using these attributes, we applied two types of data mining classification methods: decision tree and possibilistic support vector machines (PSVM). We found that decision tree showed 85.19% accuracy and PSVM exhibited 77.78% accuracy for classification of depression, 30 patients with major depressive disorder and 24 patients having bipolar disorder.

**Keywords:** Depression, Nonlinearity Analysis, Approximate entropy, Decision tree, Possibilistic Support Vector Machines (PSVM)

### 서 론

인간의 생활이 점차 풍요해지고 여유로워짐에 따라, 인간

은 좀 더 행복한 삶을 추구하는데 시간을 보내게 되었다. 그런데 이러한 추세에도 불구하고 질병에 의해 어쩔 수 없이 행복을 영위하기 힘든 사람들이 있다. 그 중 대표적인 하나는 기분 장애(mood disorder)를 앓고 있는 환자이다. 기

분 장애(mood disorder)란 특정한 기분이 상하는 이유에 의해서건, 혹은 그러한 이유가 없이도 자신의 기분상태(mood condition)를 개인이 더 이상 제어 할 수 없는 현상이 오는 질병이다. 가장 흔한 예로 우울 증상을 들 수 있다. 이러한 기분 장애(mood disorder) 환자의 숫자는 점차 증가하고 있는 추세이며, 일생동안 우울증의 증상을 겪어 본 사람은 전체 인구의 약 20% 이상에 해당할 것이라고 예측하고 있다. 따라서 이러한 증상을 조기에 진단하고 질병을 판별하여 치료해야 하는 필요성이 늘어나고 있다 (Mark F.Bear et al., 2001).

이러한 기분 장애(mood disorder)는 크게 두 가지로 나뉘는데, 하나는 주요 우울 장애(major depressive disorder)이고, 또 다른 하나는 양극성 장애(bipolar disorder)이다. 주요 우울장애(major depressive disorder)의 경우 가장 일반적인 기분 장애(mood disorder)로써 1년에 인구분포의 약 5%에 해당하는 발병률을 보이고 있다. 이러한 주요 우울 장애(major depressive disorder)의 대표적인 증상으로는 식욕부진이나 식욕증진, 불면증이나 과대 수면증, 죄의식 혹은 자괴감, 집중력 저하, 죽음에 대한지속적인 생각 등이 있다. 양극성 장애(bipolar disorder) 환자의 경우 두 가지 증상(phase)을 번갈아가며 보이는데 조증(mania)과 울증(depression)이 그것이다. 조증(mania) 기간에는 과도한 자만심 혹은 자신감, 수면 욕구 감소, 광분성, 다양하고 과도한 정신활동, 목적을 향한 집착성 같은 활발한 정신적, 육체적 활동 증상을 보이고, 울증(depression)기간에는 주요 우울 장애(Major depressive disorder)와 비슷한 정신적, 육체적으로 침체된 증상을 보이게 된다. 조증(mania)의 증상 정도에 따라 양극성 장애(bipolar disorder)는 두 가지 타입으로 나뉘는데, 조증(mania)이 매우 심하고 전체인구의1%의 발병률을 보이는 Type I과, 조증(mania)이 약하고 0.6%의 발병률을 보이는 Type II가 그것이다 (Lewis L. Judd et al., 1998 Mark F.Bear et al., 2001; Biederman J, James RS., 2005).

이러한 기분장애(mood disorder)의 치료에 쓰이는 약은 질병의 종류에 따라 다르다. 예를 들어 주요 우울 장애(major depressive disorder)의 경우 항우울제(Antidepressant) 약을 사용한다. 항우울제의 대표적인 것으로는 삼환계 항우울제(tricyclic antidepressants) 계통의 아미트리프탈린(amitriptyline), 이미프라민(imipramine)이 있고, 선택적 세로토닌 재흡수 차단제(Selective Serotonin Reuptake Inhibitor) 계통의 프로작(prozac), 그리고 최근에는 이팩사(venlafaxine), 레메논(mirtazapine) 약품이 주로 쓰인다. 양극성 장애(bipolar disorder)의 경우 기분안정제(mood stabilizer)

와 항정신병제제(antipsychotics)가 치료에 병용 되는데, 기분안정제로는 리튬(Lithium), 발프로 익산(valproic acid), 카바마제핀(carbamazepine)이 대표적이고, 항정신병제제로는 헬로 페리딘(haloperidol), 클로르프로마진(chlorpromazine), 리스페달(risperidone)이 주로 쓰인다(William Bondareff et al., 2000 Sachs GS et al., 2000 Mark F.Bear et al., 2001)

따라서 우울 증세를 보이는 환자의 적절한 치료를 위해서는 정확한 질병진단과 분류가 필요하며, 이를 효과적으로 하는 방법이 요구된다. 현재 환자의 진단에 흔히 사용되는 방법은 다면적 인성 검사(MMPI) 계통의 통계적 분석을 하는 것이다. 즉, MMPI에 있는 500문항 이상의 문항(check list)에 대한 응답 분포를 파악하고, 이것을 기존 환자들의 응답 자료의 통계치와 비교하여 병을 진단하는 방법이다 (Wiener DN., 1948 Greene RL. 1991). 따라서 이러한 기존의 방법은 환자가 30분에서 45분가량 동안 500문항 이상이 되는 문제를 풀 수 있는 집중력(attention)이 필요하다. 하지만 우울증 환자의 경우 집중력(attention) 저하의 증상을 보인다(SV Faraone et al. 1997 Mark F.Bear et al., 2001). 따라서 정신적인 활동이 저하된 우울증 환자에게서 올바른 설문 결과를 예측하기 힘들다.

이러한 단점을 피하기 위해서 이 논문에서는 빠른 측정이 가능하고 측정동안 집중력을 요하지 않는 뇌파(EEG; electroencephalogram)의 분석을 통해 우울증 환자를 분류하였다. 우울증 증세 또한 뇌의 비정상적인 활동과 연관이 있음이 알려지면서 뇌의 활동 정도를 측정하는 방법이 우울증의 분류와 연관이 있을 것으로 생각되기 때문이다. 따라서 이 논문에서는, 우울 증상을 보이는 특정 환자에게서 뇌의 활동 정도를 뇌파(EEG)를 이용하여 측정하고, 그 데이터 분석을 이용하여 우울증 환자의 질병 분류에 사용할 수 있는 가능성을 보였다.

## 방법 및 알고리즘

### 1. 피험자

이 연구에서 우울증 분류에 사용된 환자는 주요 우울 장애(major depressive disorder) 환자 30명(나이: 평균=49.1666 표준편차 14.21045 남성: 10명 여성: 20명), 양극성 장애(bipolar disorder) 환자24명 (나이: 평균=37.75 표준편차 14.247 남성: 10명 여성: 14명)이다.

### 2. EEG 측정방법

EEG 측정은 총 16 위치(그림 1, F7, T3, Fp1, F3, C3, P3, O1, F8, T4, T5, T6, Fp2, F4, C4, P4, and O2)에 대하여 깨어있고 편하게 누워, 눈을 감은 상태에서 400Hz로 120초 동안 지속적으로 측정되었다(48000 data points at sampling frequency of 400 Hz).

교신저자 : 이도현 (Email: doheon@kaist.ac.kr)

본 논문은 과학기술부 국가지정연구실사업(2005-01450)의 지원으로 수행되었음. 연구 및 전산 시설은 정문술 바이오정보전자센터와 IBM SUR 프로그램의 도움을 받았음.

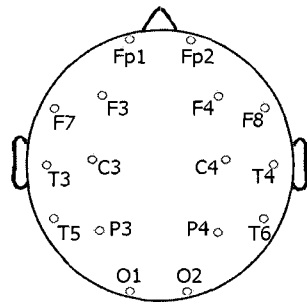


그림 1. 본 논문에서 사용된 16개의 채널 위치.

**3. 데이터 마이닝을 위한 속성 선택**

두 환자 그룹의 뇌파의 차이를 판별하기 위해 데이터 마이닝(datamining) 기법을 사용하였으며, 데이터 마이닝 기법을 위한 속성 값으로 뇌파(EEG) 채널사이 정보흐름에서의 비선형도(nonlinearity) 양과 근사 엔트로피(approximate entropy)의 크기를 속성으로 선택하였다. 즉, 16가지 채널의 가능한 두 채널간의 연결 120개( $_{16}C_2$ )의 정보 흐름(information flow)에서 Tanaka의 비선형도(nonlinearity) 측정기법을 이용하여 얻어진 비선형도(nonlinearity) 값, 그리고 16 채널 각각의 위치에서는 근사 엔트로피(approximate entropy)값을 측정하고, 이를 이용하여 총 136개(비선형도 120개 + 근사 엔트로피 16개)의 속성(attributes)을 생성하였다(Bruhn, Jorgen et al., 2000; Tanaka N et al., 2000).

뇌파의 비선형 분석결과, 질병을 앓고 있는 환자에게는 비선형적인 척도의 변화가 있음이 기존의 연구들에서 보고 되어 오고 있다(Micheloyannis S et al.,1998; J Jeong et al.,2001; Lehnertz et al., 2001; CJ Stam, 2005). 또한 비선형도(nonlinearity)는 뇌의 복잡도(complexity)와도 밀접한 연관이 있기 때문에 정보흐름(information flow)의 비선형도(nonlinearity)를 측정할 값은 의미를 가진다(Breakspear M. et al., 2003 Jeong-Ho Cahe et al., 2004). 하지만 가능한 두 채널간의 연결 120개( $_{16}C_2$ )에는 채널간의 거리가 멀다거나, 정보의 흐름이 적어 의미가 없는 연결이 다수 존재한다. 따라서 두 그룹 간의 정보흐름(information flow)의 비선형도(nonlinearity)의 값의 차이를 T-검증(T-test)하여 p값이(p-value) 0.3 초과인 채널 연결은, 두 환자그룹(주요 우울 장애와 양극성 장애)에서 통계적인 차이를 보이지 않는 것으로 간주하고 속성에서 삭제하였다. 그 결과 p값이 0.3 이하인 16개 연결만을 데이터마이닝(datamining)을 위한 속성(attribute)으로 사용하였다(Fp1-T3, F3-F4, F3-T3, F3-T5,

F3-F8, O1-C4, O1-P4, O1-O2, O1-T5, O1-F8, O1-T4, O1-T6, Fp2-T3, F4-T3, C4-F8, F7-T6).

근사 엔트로피(approximate entropy)의 경우 질병의 종류에 따라 다를 가능성이 있는 뇌 각 부분의 활성도를 측정하기 위해서 사용하였다(Steve Pincus, 1994 Bruhn, Jorgen et al., 2000). 근사 엔트로피(approximate entropy)의 경우 두 그룹 간에 차이를 T-검증을 통한 p값이(p-value) 앞의 비선형도(nonlinearity) 분석보다 작기 때문에 (평균: 0,1742 표준편차: 0.12816) 모두 데이터마이닝의 속성(attribute)으로 사용하였다. 따라서 데이터마이닝에 사용된 테이블에는 각 환자의 EEG별로 32개의 속성(attribute)을 가지고 있으며, 주요 우울 환자(major depressive disorder) 30명, 양극성 장애(bipolar disorder) 24명에 대하여 각각 30 \* 32, 24 \* 32 테이블을 생성하고 이를 이용하여 데이터 마이닝 테크닉을 통해 분석하였다.

**4. 데이터 마이닝 테크닉**

이 논문에서는 데이터마이닝(datamining)의 2가지 기술을 적용하여 환자를 분류하였다. 첫 번째 기법은 의사 결정트리(decision tree)로써, 환자를 구분하는데 쓰이는 속성(attribute)을 손쉽게 확인할 수 있고 해석 또한 용이하다. 두 번째 방법은 최근에 데이터의 클러스터링(clustering)과 분류(classification)에 많이 사용되고 있는 SVM(Support Vector Machine) 기법을 이용하였다. SVM기법은 의사 결정 트리(Decision tree)와는 달리 몇 가지 속성(attribute)에만 의존하지 않고, 입력 데이터의 차원을 변경하면서 속성(attribute)들의 분석하기 때문에 조금 더 신빙성 있는 결과를 예측할 수 있다(A Srivastava et al.,1999).

**분석 결과**

**1. 의사 결정 트리(Decision tree)**

이 논문에서는 의사 결정 트리(decision tree)의 그로잉 메소드(growing method) 알고리즘 중 QUEST를 이용하여 분석하였다(그림 2, SPSS사용. Significance level for splitting = 0.05, minimum parent node = 5, minimum child node = 5, maximum tree depth = 5 value interval = 10). 두 환자 그룹을 분류하는 첫 번째 속성(attribute) 값은 C3영역의 근사 엔트로피(approximate entropy)값의 크기로 0.34517896이하 그룹인 노드(node)1과 0.34517896초과 그룹인 노드(node)2로 나뉜다. 노드(node)1에서는 Fp1과 T3 영역 사이의 정보흐름 중 비선형도(nonlinearity)의 크기를 기준으로 다시 두 개의 그룹 노드3 (0.084677460이하), 노드4(0.084677460초과)로 나뉘어진다. 마찬가지로 노드(node)2의 경우에는 T5의 근사 엔트로피(approximate entropy)값의 차이로 노드5 (0.44590074이하) 와 노드6(0.44590074초과)로 구분 지어진다.

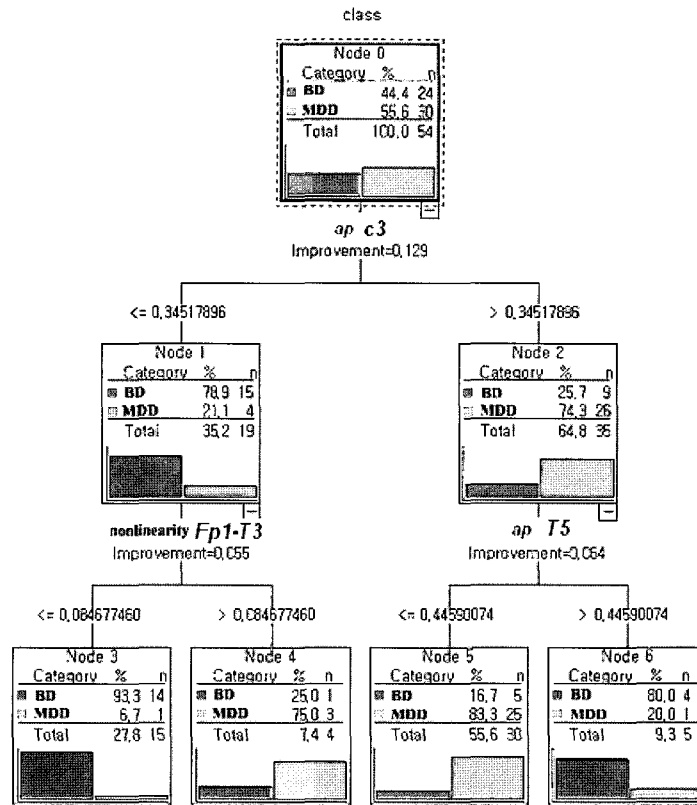


그림 2. 의사 결정 트리를 이용한 우울증환자의 분류. BD는 양극성 장애 환자, MDD는 주요 우울장애 환자를 뜻한다. ap 채널 이름은 채널이름에 해당하는 곳의 유사 엔트로피 값을 뜻하며, nonlinearity 채널이름-채널이름은 두 채널간의 nonlinearity의 값을 뜻한다.

노드(node)3에 속하는 그룹의 경우 93.3%가 양극성 장애에 해당되었으며, 노드(node)4에 속하는 그룹의 경우는 75%가 주요 우울증 환자로 판별되었다. 마찬가지로 노드(node)5에 속하는 그룹의 경우 83.3%가 주요 우울환자로 분류되었고, 노드(node)6의 그룹에서는 80%가 양극성 장애환자로 구분되었다. 환자의 분류의 전체적인 정확도는 약 85.185%로 측정되었다.

## 2. 가능성 기반 서포트 벡터 머신(possibilistic support vector machines)

이 논문에서 사용한 SVM 메소드는 최근에 발표된 PSVM(Possibilistic support vector machines)을 사용하였다 (KiYoung Lee, et al.,2005). 리브 원 아웃 메소드 (leave-one-out method)를 사용하여 뇌파 데이터를 분석하였다. PSVM 기법을 통해 두 개의 그룹을 분류한 결과 77.7778%의 정확도를 보였다.

## 결론

이 논문에서는 우울증 환자의 분류를 위한 뇌파(EEG)를 사용한 방법을 제시하였다. 뇌파(EEG)를 통한 분석은 환자가 단순히 뇌파(EEG) 측정 장치를 부착하고 1~2분 동안 누워있으면 측정이 완료되고, 이렇게 측정된 데이터를 분석기에 넣어 결과가 나오는데 5분 이하가 소요되는 테스트이다. 따라서 측정에만 30분에서 45분의 시간이 소요되는 MMPI(Minnesota multiphasic personality inventory) 계통의 통계분석 보다 빠른 측정이 가능하다. 결국 이러한 빠른 측정은 집중력에 문제가 있는 환자에게는 쓰이기 힘들었던 기존의 방법의 하나의 대안이 될 것이다.

하지만 뇌파(EEG)를 통한 측정은 임상적인 진단과 동시에 병행되어야 한다. 뇌파(EEG)를 통한 질병 진단이 아직까지 보편화가 되어있지 않고, 뇌파(EEG)의 어떠한 분석 기법이 가장 효율성이 있고 효율적인지 현 시점에서 밝혀지지

않았기 때문이다. 따라서 이 논문에서 제시된 뇌파(EEG) 분석 방법을 통해 환자의 질병 진단 빠르게 하되, 환자의 추후의 임상적인 증상에 기반을 두어 치료를 진행해가는 방향으로 사용되기를 제안한다. 마지막으로 뇌파(EEG)를 통한 질병의 분석 기법이 발전하기 위해 많은 분석 알고리즘의 개발 및 질병을 앓고 있는 환자의 뇌의 활동과 뇌파와의 관계에 대한 연구가 필요할 것이다.

### 참고문헌

- [1] A. Srivastava, EH Han, V. Kumar, and V. Singh. (1999) Parallel formulation of decisiontree classification algorithms. In *Data Mining and Knowledge Discovery: An International Journal* 3, pages 237-261
- [2] Biederman J, James RS. (2005) Advances in the neurobiology of pediatric bipolar disorder. *Biol Psychiatry* 1;58(7):515-6.
- [3] Breakspear M., Terry JR, Friston KJ, Harris AW, Williams LM, Brown K, Brennan J, Gordon E(2003) A disturbance of nonlinear interdependence in scalp EEG of subjects with first episode schizophrenia, *NeuroImage* 20:466 - 478
- [4] Bruhn J, Ropcke H, Rehberg B, Bouillon T, Hoeft A. (2000) Electroencephalogram Approximate Entropy Correctly Classifies the Occurrence of Burst Suppression Pattern as Increasing Anesthetic Drug Effect. *93(4):981-985*
- [5] CJ Stam (2005) Nonlinear dynamical analysis of EEG and MEG: review of an emerging field. *Clin Neurophysiol.*
- [6] Greene RL.(1991) *The MMPI-2/MMPI: An Interpretive Manual*. Boston, Mass: Allyn and Bacon.
- [7] Jeong, Jaeseung, Chae, Jeong-Ho, Kim, Soo Yong, Han, Seol-Heui (2001) Nonlinear dynamic analysis of the EEG in patients with Alzheimer's disease and vascular dementia, *Clin Neurophysiol*
- [8] Chae JH, Jeong J, Peterson BS, Kim DJ, Bahk WM, Jun TY, Kim SY, Kim KS. (2004) Dimensional complexity of the EEG in patients with posttraumatic stress disorder, *Psychiatry Research: Neuroimaging* 131:79-89
- [9] KiYoung Lee, Dae-Won Kim, Kwang H. Leea, Doheon Lee (2005) Possibilistic support vector machines, *Pattern Recognition.* 38:1325-1327
- [10] Lewis L. Judd, Hagop S. Akiskal, Jack D. Maser, Pamela J. Zeller, Jean Endicott, William Coryell, Martin P. Paulus, Jelena L. Kunovac, Andrew C. Leon, Timothy I. Mueller, John A. Rice, Martin B. Keller (1998) A Prospective 12-Year Study of Subsyndromal and Syndromal Depressive Symptoms in Unipolar Major Depressive Disorders, *Arch Gen Psychiatry* 55:694-700.
- [11] Lehnertz, Klaus, Andrzejak, Ralph G., Arnhold, Jochen, Kreuz, Thomas, Mormann, Florian, Rieke, Christoph, Widman, Guido, Elger, Christian E.(2001) Nonlinear EEG Analysis in Epilepsy:: Its Possible Use for Interictal Focus Localization, Seizure Anticipation, and Prevention., *Journal of Clinical Neurophysiology.* 18(3):209-222
- [12] Mark F.Bear Barry W. Connors, Michael A. Paradiso (2001) *Neuroscience Exploring The Brain* Lippincott Williams &Wilkins 2001 second edition.
- [13] Micheloyannis S, Flitzanis N, Papanikolaou E, Bourkas M, Terzakis D, Arvanitis S, Stam CJ. (1998) Usefulness of non-linear EEG analysis., *Acta Neurol Scand.*
- [14] Sachs GS, Printz DJ, Kahn DA, Carpenter D, Docherty JP.(2000) The Expert Consensus Guideline Series: Medication Treatment of Bipolar Disorder, *Postgrad Med.* 2000 Apr;Spec No:1-104.
- [15] Steve Pincus (1994) Approximate entropy (ApEn) as a complexity measure, *Chaos An Interdisciplinary Journal of Nonlinear Science*, 5(1): 110-117
- [16] Faraone, Stephen V., Biederman, Joseph, Mennin, Douglas, Wozniak, Janet, Spencer, Thomas(1997) Attention-Deficit Hyperactivity Disorder With Bipolar Disorder: A Familial Subtype? *Am Acad Child Adolesc Psychiatry*, 36(10):1378-1387.
- [17] Tanaka N, Okamoto H, Naito M (2000) Detecting and evaluating intrinsic nonlinearity present in the mutual dependence between two variables. *Physica D* 147:1-11
- [18] Wiener DN. (1948) Subtle and obvious keys for the MMPI. *J Consult Clin Psychol.* 12:164-170.
- [19] William Bondareff, Murray Alpert, Arnold J. Friedhoff, Ellen M. Richter, Cathryn M. Clary, Evan Batzar (2000) Comparison of Sertraline and Nortriptyline in the Treatment of Major Depressive Disorder in Late Life. *Am J Psychiatry* 157:729-736