논문 2024-4-25 http://dx.doi.org/10.29056/jsav.2024.12.25

필기를 이용한 알츠하이머병의 조기 진단을 위한 하이브리드 특징 선택 접근법

김영인*†

A Hybrid Feature Selection Approach for Early Detection of Alzheimer's Disease from Handwriting

Young-In Kim*†

요 약

전 세계적 고령화로 인한 알츠하이머병 환자의 급격한 증가는 공중보건의 심각한 우려를 야기하고 있으며, 이는 조기 진단의 필요성을 증가시키고 있다. 본 논문은 알츠하이머병을 위한 DARWIN(Diagnosis AlzheimeR WIth haNdwriting) 필기 데이터셋의 높은 특징 차원 문제를 해결하기 위해 효율적인 특징 선택을 위한 하이브리드 기법을 제안한다. 이 기법은 먼저 사전에 선정된 다양한 특징 선택 기법으로 다양한 특징 개수의 데이터셋을 추출하여 여섯 가지 분류 기법으로 실험한 후, 가장 우수한 성능을 보인 특징 선택 기법을 두 가지 선정하여, 이를 결합한 하이브리드 특징 선택하는 과정으로 구성된다. 두 가지 특징 선택 기법을 선정하기 위한 실험 결과, 상호정보량과 SHAP이 선정되었으며, 이를 결합한 하이브리드 특징 선택을 수행하고 선택된 특징으로 실험하였다. 그 결과, 이 기법으로 선택한 82개 특징만으로 정확도 93%의 성능을 구하였으며, 44개의 특징만으로도 92%의 정확도를 구하였다. 향후 연구에서는 선택된 특징들의 임상적 의미와 관련 작업을 심층 분석하고자 한다.

Abstract

The rapid increase in Alzheimer's disease patients due to global aging poses a significant public health concern, highlighting the urgent need for early diagnosis. This study proposes a hybrid feature selection technique to address the high-dimensional feature problem of the DARWIN (Diagnosis AlzheimeR WIth haNdwriting) handwriting dataset. The technique first extracts datasets with varying numbers of features using pre-selected feature selection techniques, experiments with six classification techniques, and then combines the two best-performing feature selection techniques into a hybrid approach. Experimental results identified Mutual Information and SHAP as the two optimal techniques, which were then combined to perform hybrid feature selection. Using the selected features, experiments demonstrated that the proposed technique achieved an accuracy of 93% with only 82 features and 92% with as few as 44 features. Future research will focus on the clinical relevance of the selected features and related tasks.

한글키워드 : 알츠하이머병, 조기 진단, 필기 데이터, 특징 선택, 하이브리드 기법

keywords : Alzheimer's Disease, Early Diagnosis, Handwriting Data, Feature Selection, Hybrid Technique

* 부산대학교 IT응용공학과
† 교신저자: 김영인(email: kimyi@pusan.ac.kr)

접수일자: 2024.12.03. 심사완료: 2024.12.14.
게재확정: 2024.12.20.

1. 서 론

최근 전 세계적인 고령화로 인한 알츠하이머병(AD: Alzheimer's disease) 환자의 급격한 증가로 공중보건의 심각한 문제가 발생하고 있으며, 2019년 기준 약 5,500만 명이 알츠하이머병을 앓고 있는 것으로 추정되어, 세계보건기구(WHO)는 이 숫자가 2050년까지 1억 3,900만 명에 이를 것으로 전망하며, 2019년 치매로 인한 연간 비용이 약 1.3조 달러에서 2030년에는 2.8조 달러를 넘어설 것으로 예측하고 있다[1]. 알츠하이머병을 포함한 치매는 환자와 돌봄 제공자, 그리고 의료 시스템에 막대한 부담을 안기고 있으므로 조기에 정확히 진단하여 적절한 치료 계획을 수립하고 치료하는 것은 매우 중요하다[2].

현재 알츠하이머병의 조기 진단과 모니터링을 위하여 주로 활용되는 신경영상 기법인 자기공명영상(MRI)과 양전자 방출 단층촬영(PET)은 뇌의 구조적 및 기능적 변화를 탐지하는 데 널리 사용되고 있으나, 고비용과 방사성 추적자의 사용으로 인한 문제로 일상적인 검사에는 부적합하다[3,4]. 또한 비침습적 기법으로 뇌파검사(EEG)와 자기뇌파검사(MEG)는 알츠하이머병의 초기 인지 장애와 관련된 뇌의 활동 변화를 실시간으로 포착할 수 있고 신경영상 기법에 비해 상대적으로 저렴하고 안전한 장점이 있으나, 두 기법 모두 뇌 활동의 정확한 위치를 분석하는데 한계가 있다[5].

최근에 웨어러블 기기와 다양한 센서를 활용한 생체 신호 분석 기법이 알츠하이머병 진단의 새로운 가능성을 보여, 필기, 보행, 언어, 수면 등의 패턴에 대한 생체신호를 수집하고 분석하는 연구가 활발히 진행되고 있다[3]. 센서 데이터는 일상생활에서 지속적으로 수집할 수 있어 알츠하이머병 환자의 조기 징후를 감지하는 데 유용할 수 있다[6,7]. 이러한 센서 데이터의 강점은 비침습적이고 상대적으로 적은 비용으로 지속적인 모니터링이 가능하다는 점이다. 더불어 센서 데이터는 기존 방법 보다 비교적 객관적이고 정량화된 정보를 제공할 수 있다.

최근 몇 년 동안에는 복잡한 인지 및 운동 활동으로 간단한 필기 과제를 수행하며 가속도계, 자이로스코프, 텐서 저항기와 같은 센서를 사용하여 필기 작업 중의 미세한 운동 및 압력 패턴의 데이터를 손쉽게 수집하는 연구가 이루어지고 있다. 필기는 보행이나 수면 패턴 분석에 비해 특수 장비나 장시간 모니터링이 불필요하며, 언어처럼 문화적, 교육적 배경에 따른 차이가 적어 객관적인 측정이 가능한 장점이 있으며, 미세한 운동 조절 능력과 시공간 지각 능력 등 다양한 인지 기능을 복합적으로 가지고 있어, 임상 환경이나 특수 장비없이 환자의 환경에서 자연스럽게 필기 작업을 수행하며 수집한 데이터와 기계학습의 알고리즘을 결합하면 알츠하이머병과 관련된 인지 변화를 반영하여 초기 징후를 효과적으로 감지할 수 있다[8]. 이와 관련하여 최근에는 다양한 기계학습 및 딥러닝 기법들이 제안되고 있으며, 특히 450개의 특징으로 구성된 공개 필기 데이터셋인 DARWIN[9,10]을 사용한 연구가 활발히 이루어지고 있으나, 많은 수의 특징으로 인해 연산 복잡도를 높이고, 진단 모델의 과적합 문제가 발생할 수 있어 분석에 어려움이 있다.

이와 같은 DARWIN 데이터셋을 활용한 연구로 [11]에서는 25개의 필기 작업 중에서 4개의 핵심 작업의 72개 특징을 사용하는 방법으로 SVM을 사용하여 80.43%의 정확도를 제시하였다. 한편, [12]에서는 기존 25개 필기 작업의 18개의 필기 특징에 Mean Azimuth와 Mean Slope을 추가로 제안하여 85.86%의 정확도를 구하였고, [13]에서는 이 데이터셋의 1차원 필기 특징을 2차원 이미지 특징으로 변환하고, 12개의 계층으로 구성된 새로운 CNN 모델을 제안하여 정확도 90.4%의 성능을 제시하였고, 다양한 기계학습과 딥러닝 모델들과의 성능 비교 실험을 수행하여 InceptionV3, ResNet152V2, MobileNetV2, Xception 등 8개의 기존 모델들과 비교 실험한

결과, 두 번째로 좋은 성능을 보인 Xception 모델보다 약 11.6%의 성능 향상을 그리고 가장 낮은 성능인 MobileNetV2의 정확도 59.6%와는 큰 성능 차이를 보임을 제시하였다. [14]는 딥러닝 트랜스포머 모델의 핵심 요소인 self-attention 메커니즘 기반으로 25개 전체 작업에 대해 평균 94.3%의 정확도를 달성하여 [13]의 CNN 기반 방법보다 약 4% 개선하였다. [15]는 앙상블 기계학습 모델을 제안하고 추출한 337개 특징들 중에서 각 분류기별로 최적의 특징을 선별하는 과정을 수행하고 최종적으로 스태킹 앙상블 모델로 97.14%의 높은 정확도를 달성하였지만, 최종 앙상블 모델의 성능 평가 과정에서는 교차 검증이 적용되지 않아 일반화 성능의 검증이 부족한 한계점이 있었다.

따라서 알츠하이머병의 조기 진단 정확도를 높이면서도 임상적으로 해석 가능하고 효율적인 특징 선택 기법에 대한 추가적인 연구가 필요하다. 특히 교차 검증을 통한 일반화 성능의 검증과 임상적 의미를 고려한 특징 선택, 그리고 실시간 처리가 가능한 효율적인 특징 선택 기법의 개발과 실험을 통한 연구가 필요하다.

본 논문에서는 다양한 특징 선택 기법과 여섯 가지 분류 알고리즘을 통한 실험을 수행하여 제안한 하이브리드 특징 선택 기법의 성능을 비교하는 실험을 한다. 2장은 사용한 데이터셋과 특징 선택 기법 및 분류 기법을 설명하며, 3장에서는 제안한 특징 선택 기법을 포함한 연구 방법을 설명한다. 4장에서는 실험 방법과 결과를 설명하며, 5장에는 결론을 기술한다.

2. 데이터셋과 특징 선택

2.1 데이터셋

DARWIN 데이터셋[9,10]은 174명의 피실험자로부터 수집된 필기 데이터로 구성된다. 알츠하이머병 환자군 89명(여성 46명, 남성 44명)과 건강한 대조군 85명(여성 51명, 남성 39명)으로 구성되며, 연령, 성별, 교육 수준, 직업 등을 기준으로 그룹 간 통계적 유사성을 확보하였다. 알츠하이머병 환자군의 평균 연령은 71.5세(±9.5), 평균 교육 연수는 10.8년(±5.1)이며, 대조군의 평균 연령은 68.9세(±12), 평균 교육 연수는 12.9년(±4.4)이다. 환자군의 질병 심각도는 MMSE (Mini-Mental State Examination), FAB (Frontal Assessment Battery), MoCA (Montreal Cognitive Assessment) 등의 임상 테스트를 통해 평가되었다.

데이터 수집은 WACOM Bamboo Folio 디지털 태블릿 상의 A4 용지에서 이루어졌으며, 각 피실험자는 25개의 필기 작업을 수행하였다. 필기 작업은 그래픽 작업, 복사 작업, 기억 작업, 받아쓰기 작업의 네 가지 유형으로 구분된다.

수집된 원시 데이터(펜의 x, y 좌표, 종이 접촉 여부, 공중 움직임, 시간 정보 등)는 18개의 정량적 특징으로 변환되었다. 주요 특징은 시간 관련 특징, 속도와 가속도 지표, 속도 변화율, 필압, GMRT(손의 안정성), 공간 확장성, Dispersion Index(필기 분산 정도) 등이다.

최종 데이터셋은 452개의 열로 구성되며, 첫 번째 열은 참가자 식별자, 마지막 열은 클래스 정보('P': 환자, 'H': 건강인)를 나타낸다. 나머지 열은 각 작업에서 추출된 특징 데이터를 포함한다. DARWIN 데이터셋은 필기를 통한 알츠하이머병의 초기 신호 포착 및 운동 제어와 인지 기능의 변화를 분석하는 데 유용한 데이터로 기계학습 모델 개발 및 평가를 위한 중요한 자료로 활용되고 있다.

2.2 특징 선택 기법

본 연구에서는 알츠하이머 치매 환자의 필기 데이터에서 효과적인 특징을 선택하기 위해 다음과 같은 다섯 가지의 특징 선택 기법을 사용하고자 한다.

2.2.1 F-value (ANOVA)

클래스 간 분산과 클래스 내 분산의 비율을 통해 F-value를 계산하여, 각 특징이 클래스간 분산을 얼마나 잘 설명하는지 평가한다[16]. 특징 선택 기법 중 단순하면서도 강력한 도구로, F-value는 모델에 독립적이며, 특징의 통계적 분포에 따라 계산되며, 계산 효율성도 높아 특징을 빠르게 평가할 수 있다. F-value는 식 (1)과 같이 정의된다.

$$F = (\text{클래스간분산})/(\text{클래스내분산}) \quad (1)$$
$$= (\sum_{k=1}^{K} n_k(\overline{x_k}-\overline{x})^2)/((\sum_{k=1}^{K}\sum_{i=1}^{n_k}(x_{i,k}-\overline{x_k})^2)$$

여기서 K는 클래스의 전체개수, n_k는 k번째 클래스의 데이터 개수, $\overline{x}_k$는 k번째 클래스의 평균, $\overline{x}$는 전체 데이터의 평균을 나타내며, $x_{i,k}$는 k번째 클래스에서 i번째 데이터를 나타낸다.

2.2.2 Random Forest Importance

Random Forest Importance는 다수의 의사결정 트리가 특징을 선택할 때 특징 간의 상호 작용을 포착하여 중복된 정보를 가진 특징들의 중요도를 낮게 할 수 있어, 각 특징이 모델 학습 과정에서 얼마나 중요한 역할을 했는지를 평가하는 임베디드 기법이다[17]. 특정 특징 j의 중요도는 식 (2)와 같이 정의된다.

$$Importance(j) = \sum_{t \in Trees} \triangle Gini_{j,t} \quad (2)$$

여기서 $(\triangle Gini_{j,t})$는 트리 (t)에서 Gini 불순도의 감소량을 나타낸다.

2.2.3 Mutual Information

Mutual Information은 모델 훈련이 필요없는 필터 기법으로 특징의 중요도를 계산한 후, 상위 중요도를 갖는 특징을 선택하는 독립적으로 특징을 평가하는 방식으로 클래스 구분에 중요한 특징을 선별하는 데 사용한다[18]. Mutual Information은 식 (3)과 같이 정의된다.

$$I(X,Y) = \sum_{x \in X} \sum_{y \in Y} p(x,y)\log(\frac{p(x,y)}{p(x)p(y)}) \quad (3)$$

여기서 p(x,y)는 X와 Y의 결합확률 분포를 나타내고, p(x), p(y)는 각각 X와 Y의 주변 확률 분포를 나타내, I 값이 클수로 변수간의 의존도가 높음을 의미한다.

2.2.4 SHAP (SHapley Additive exPlanations)

SHAP은 XAI(eXplainable AI) 기법의 하나로 게임 이론에서 유래된 shapley value를 기반으로 각 특징이 클래스 분류에 기여한 정도를 측정하는 기법[19]으로, 특징의 포함과 제외에 따른 모델의 출력의 변화량으로 각 특징이 미치는 기여도를 계산하는 특징 선택 기법이다. SHAP 값은 식 (4)와 같이 계산된다.

- shapley 값 계산:

$$\varnothing_j = \sum_{S \subseteq F\setminus\{j\}} \frac{|S|!(|F|-|S|-1)!}{|F|!} [f(S\bigcup\{j\})-f(S)] \quad (4)$$

- 특징 중요도 계산: $Importance(x_j) = |\varnothing_j|$
- 특징을 중요도 내림차순으로 정렬한 후,
  ① $|\varnothing_j|$의 상위 특징 k개를 선택:
  ② 누적 SHAP 값이 특정 임계값에 도달할 때까지 특징 선택:

$$\sum_{j=1}^{k}|\varnothing_j| \geq \alpha \cdot \sum_{j=1}^{d}|\varnothing_j|, \alpha \in (0,1) \quad (5)$$

여기서 F는 전체 특징 집합, S는 j-번째 특징을 제외한 임의의 부분 집합, f(S)는 S에 포함된 특징만 고려했을 때의 모델 예측 값, $|\varnothing_j|$는 각 특징 x_i에 대해 계산한 shapley value 즉 중요도이며, α는 누적 중요도의 임계값을 그리고 d는 전체 특징의 개수를 나타낸다.

2.2.5 Permutation Importance

Permutation Importance(순열 중요도)는 XAI의 도구 중 하나로, 모델 예측에 대한 특징의 기여도를 직관적으로 평가하고 설명하는 방법[20]으로, 특징을 한 개씩 무작위로 섞어 분류 성능의 변화를 평가하는 방식으로 특징 간에 상호작용이 많은 데이터에 적합하고 모델 독립적인 범용성이 높은 기법이다. 변수 x_j의 순열 중요도는 식 (6)과 같이 정의된다.

$$I_j = Score_{original} - Score_{shuffled_j} \quad (6)$$

여기서 I_j는 변수 x_j의 순열 중요도이며, $Score_{original}$은 변수 x_j를 유지한 상태에서의 모델 성능을 그리고 $Score_{shuffled_j}$는 변수 x_j의 값을 무작위로 섞어 입력했을 때의 모델 성능으로 이 두 가지 예측 성능을 비교하여 중요도를 계산한다.

F-value (ANOVA), Mutual Information과 Random Forest Importance는 기존 특징 선택 기법 중 검증된 성능과 초기 특징 선별 및 모델 학습 중 중요도 평가에 강점이 있으며, SHAP과 Permutation Importance은 XAI와 정보 이론 기반 접근법을 통해 해석 가능성과 비선형 관계를 강화할 수 있어 선정하였다. 이 기법들은 서로 다른 관점에서 특징의 중요도를 평가하여 상호 보완적인 정보를 제공하므로 이와 같은 특징 선택의 조합으로 DARWIN 데이터셋에서 알츠하이머 치매를 조기 진단하기 위하여 중요한 특징을 선택하는데 다각적 접근을 통해 각각의 특징 선택 기법의 한계를 보완하고, 보다 신뢰성 있는 특징 선택 기법을 개발하는데 기여하고자 하였다.

2.3 분류 모델

본 연구에서는 기존 연구[10-15]에서 우수한 성능을 보인 여섯 가지 분류 모델을 사용한다.

2.3.1. Logistic Regression

Logistic Regression[21]은 분류 작업에서 널리 사용되는 선형 모델로, 독립 변수와 종속 변수 간의 관계를 로지스틱 함수(Sigmoid Function)로 모델링한다. LR은 단순한 구조로 계산 비용이 낮아 대규모 데이터 처리에도 적합하고, 주로 이진 분류에 사용한다.

2.3.2. LightGBM

LightGBM(Light Gradient Boosting Machine)[22]은 Gradient Boosting Decision Tree(GBDT)의 경량화 버전으로, 대규모 데이터 처리에 최적화된 알고리즘으로 기존의 Level-Wise 방식보다 더 깊은 트리를 효율적으로 학습할 수 있다.

2.3.3. Random Forest

Random Forest(RF)[17]는 배깅(Bagging) 앙상블 방법론을 사용하는 알고리즘으로, 다수의 의사결정 트리를 학습시켜 최종 예측을 도출한다. 각 트리는 데이터 샘플과 특징 선택을 무작위로 하여 학습되며, 이를 통해 과적합의 가능성을 줄이고 잡음이나 이상치와 같은 데이터에 강한 모델을 제공한다.

2.3.4. Support Vector Machine (SVM)

SVM[23]은 데이터 분리를 위한 초평면을 찾는데 초점을 맞춘 선형 및 비선형 분류 모델이다. SVM은 커널 트릭(kernel trick)을 활용하여 비선형 데이터를 고차원 공간으로 변환함으로써 분류 가능성을 높인다.

2.3.5. Extra Trees

Extra Trees(Extremely Randomized Trees)[24]는 Decision Tree 기반의 앙상블 학습 알고리즘으로, Random Forest와 유사하지만 추

가적인 무작위성을 도입하여 다르게 작동한다. ET는 배깅 방식을 사용하여 각 트리 학습 시 데이터를 복원 추출 방식으로 샘플링하고, 노드 분할 기준을 무작위로 선택하여 모델의 과적합을 방지한다.

2.3.6. XGBoost

XGBoost[25]는 Gradient Boosting Decision Tree(GBDT)를 기반으로 한 앙상블 학습 알고리즘으로, 데이터 과학 대회에서 널리 사용되는 모델이다. XGBoost는 손실 함수 최적화에 2차 Taylor 전개를 적용하여 학습 속도와 모델 성능을 동시에 개선하며, 규제 항목 추가를 통해 과적합을 방지하며, 캐시 최적화와 데이터 샘플링 기술을 통해 메모리 사용량을 줄이고 계산 속도를 향상시킨다.

3. 연구 방법

본 연구에서는 필기 데이터셋을 사용하여 알츠하이머병의 조기진단을 위한 특징 선택 기법으로 DARWIN 데이터셋의 중요 특징을 선택하고 최적의 분류 성능을 도출하기 위해 다음과 같은 절차로 수행하고자 한다.

- 단계 1: 데이터 전처리 단계로 분석에 불필요한 식별자(ID) 열을 제거하고, P(환자)와 H(건강인)으로 구성된 클래스 레이블은 이진 형태로 인코딩하여 분류 분석에 적합한 형태로 변환한다. 모든 특징 값은 Standard Scaling을 통해 정규화하여 각 특징의 스케일 차이에 따른 영향을 제거한다.

- 단계 2: 특징 선택 단계로 2장에서 설명한 다섯 가지 특징 선택 기법을 사용하여, 구해진 특징 중요도 순서에 따라 특징 개수를 20, 50, 100, 200으로 개수의 변화에 따른 성능 변화를 평가할 수 있도록 데이터셋을 만든다. 이 단계에서는 2장의 설명과 같이 여섯 가지 분류 알고리즘을 사용하며, 최적의 하이퍼파라미터를 구하기 위하여 GridSearchCV를 사용하여 계층적 5겹 교차검증을 사용한다. 분류 알고리즘의 성능은 알츠하이머병의 조기 진단을 위하여 정확도, 민감도, 정밀도, F1-점수, AUC를 사용하여 종합적으로 평가하며, 특징 선택 기법과 분류 알고리즘 간의 최적 조합을 도출한다.

- 단계 3: 이 단계는 단계 2에서 구해진 최고 성능의 두 가지 기법을 결합한 하이브리드 특징 선택 기법이다. 이 기법은 우수한 성능을 보인 두 가지 특징 선택 기법의 장점을 결합하여 더욱 안정적이고 신뢰성 있는 특징을 선택하는 것을 목표로 한다. 이 단계의 특징 선택 과정은 다음과 같이 3 단계로 구성한다.

- 단계 3-1: 먼저 각 특징의 중요도를 계산한다. 특징 x에 대한 최종 중요도 I(x)는 첫 번째 특징 선택 기법(FST1)과 두 번째 특징 선택 기법(FST2)에서 도출된 값들을 [0,1] 범위로 정규화한 후, 이들의 산술평균으로 계산하며, 식은 (7)과 같다:

$$I(x) = \frac{FST1'(x) + FST2'(x)}{2} \qquad (7)$$

여기서 $FST1'(x)$와 $FST2'(x)$는 각각 [0,1] 범위로 정규화된 FST1과 FST2의 특징 중요도 값을 나타낸다.

- 단계 3-2: 점진적 특징 선택 단계로 최소 특징 수 k로 시작하여, 특징을 한 단계씩 추가하며 모델 성능 $S(n)$을 5-겹 교차 검증으로 평가하며, 이 과정은 다음 조건이 만족될 때까지 반복한다.

$$S(n+1) < \alpha \times S(n) \qquad (8)$$

여기서 $S(n)$은 현재 단계에서의 성능,

$S(n+1)$은 다음 단계에서의 성능이며 α는 성능 감소 허용치로 성능감소를 1%까지 허용한다면 0.99로 나타낸다. 이 조건을 통해 과적합을 방지하고, 모델 성능이 안정적으로 유지되는 특징 집합을 도출한다.

- 단계 3-3: 마지막 단계로 선택된 특징의 안정성을 검증하기 위하여 N회의 반복 검증을 통해 평가한다. 각 반복과정에서 각 특징 x의 안정성 점수 $Stability(x)$를 계산하며, 식은 (9)와 같다.

$$Stability(x) = \frac{Count(x)}{N} \quad (9)$$

여기서 $Count(x)$는 해당 특징이 선택된 횟수, N은 총 반복 횟수로 해당 특징이 선택된 횟수를 총 반복 횟수로 나눈 값으로 정의한다.

이러한 3단계 검증 과정을 통해 성능과 안정성을 동시에 고려한 특징 선택이 가능하도록 하여 특징 선택 과정의 신뢰도를 확보하고자 하였다. 이 기법이 기존 특징 선택 기법과 차별화된 점은 선정된 두 가지 특징 선택 기법의 중요도 값을 평균하여 각 특징의 중요도를 계산하는 것과 특징 선택의 신뢰도를 높이기 위하여 N회의 반복 검증을 통해 안정성 점수를 계산하여 최종 특징을 선택하고 명시적으로 성능 감소 허용치를 적용한다는 점이다.

4. 실험 및 결과

본 논문에서 윈도우 11의 WSL 환경에서 Ubuntu 24.04 LTS, Python 3.10.14, sklearn 1.5.1을 사용하였다.

본 연구에서는 다양한 특징 선택 기법과 특징 개수에 따라 정확도(accuracy), 민감도(Sensitivity) 등의 주요 성능 수치를 비교 분석하여 최적의 데이터셋과 설정을 도출하였다. 실험에 사용된 특징 선택 기법은 FC(F-value Classif), MI(Mutual Information), PI(Permutation Importance), RF(Random Forest), SHAP으로 표기하고 사용하였으며, 각 데이터셋은 20, 50, 100, 200의 특징 개수를 선택하여 특징 개수에 따른 변화를 관찰하였다.

구해진 24개 데이터셋의 분류 성능을 평가하기 위하여 여섯 가지 분류 모델인 LR(Logistic Regression), LB(LightGBM), RF(Random Forest), SVM(Support Vector Machine), ET(Extra Trees), 그리고 XB(XGBoost)로 표기하여 사용하고, 계층적 5겹 교차검증(stratified 5-fold cross-validation)과 최적 성능을 도출하기 위해 Grid SearchCV를 사용하여 정확도를 기준으로 하이퍼파라미터 최적화를 수행하였다. 표 1은 각 분류 알고리즘에 대한 GridSearchCV 설정을 정리한 것이다.

표 1. 분류 알고리즘별 GridSearchCV 하이퍼파라미터 탐색 범위
Table 1. hyperparameter search ranges for classification algorithms using GridSearchCV

분류모델	파라미터
LR	C: [0.1, 1, 10, 100] solver: ['lbfgs']
RF	n_estimators: [100, 200, 300] max_depth: [None, 5, 10, 20] min_samples_split: [2, 5, 10]
LB	learning_rate: [0.01, 0.1, 0.2] max_depth: [3, 6, 10] n_estimators: [100, 200]
XB	learning_rate: [0.01, 0.1, 0.2] max_depth: [3, 6, 10] n_estimators: [100, 200]
ET	n_estimators: [100, 200, 300] max_depth: [None, 5, 10, 20] min_samples_split: [2, 5, 10]
SVM	C: [0.1, 1, 10, 100] gamma: ['scale', 'auto'] kernel: ['rbf']

본 연구에서는 알츠하이머병의 조기 진단을 위하여 다섯 가지 특징 선택 기법과 네 가지 특징 개수(20, 50, 100, 200)에 따른 분류 모델의 성능을 평가하기 위해 알츠하이머병을 조기 진단하는 주요 지표로 정확도, 민감도, 정밀도, F1-점수, AUC의 다섯 가지 지표를 사용하여 데이터셋의 전체 성능을 평가하였으며, 그 결과는 그림 1과 표 2에 제시하였다.

그림 1과 표 2와 같이 MI의 200개 특징으로 구성된 데이터셋이 ET 모델에서 정확도 92.54%, 민감도 93.27%, F1-점수 92.74%, AUC 92.52%, 정밀도 92.47%의 성능을 기록하며 가장 우수한 성능을 보였다. 다음으로 SHAP의 200개 특징의 데이터셋은 ET 모델에서 정밀도가 94.56%로 우수하였다. SHAP의 100개 특징 데이터셋은 동일한 모델에서 AUC와 F1-점수에서 높은 성능을 보이며 평균 성능 92.63%로 우수한 성능을 보였다, 또한 ET를 사용한 분류 모델에서 높은 분류 성능을 구하였다. 따라서 MI와 SHAP 특징 선택 기법이 우수한 성능을 보여, 다음에 수행할 최적화된 조합을 사용하여 특징 선택을 하는 하이브리드 접근 방식으로 특징을 선택하였다.

MI의 특징 선택으로 구해진 200개 특징 데이터셋은 주로 필체 속도, 가속도, 압력 분포, 그리고 손 움직임의 흔들림(Mean Jerk)과 같은 동적 특징이 선택되었으며, 이러한 특징은 필체의 미세한 운동 조절 능력을 평가하여 신경학적 질환인 알츠하이머의 초기 진단에 효과적일 것으로 보인다. 반면 SHAP의 200개 특징의 데이터셋은 설명 가능성을 기반으로 중요 특징을 선택한 결과로 공간적 확장(Max X/Y Extension), 글씨의 분산도(Dispersion Index), 그리고 펜을 내리고 올리는 동작 횟수(Pendown Number)와 같은 공간적 및 시간적 특징을 주로 포함하여 알츠하이머 환자의 필체 분석에서 공간적 왜곡 및 시간적 불규칙성을 식별하는 데 유용한 것으로, MI 기반 특징은 주로 정량적 동적 특징에, 그리고 SHAP 기반 특징은 주로 정성적 설명력에 강점을 보이는 것으로 분석되었다. 이 분류 실험에서 가장 우수한 성능을 보인 분류 모델의 최적 하이퍼파라미터는 표 3과 같다.

표 2. 특징 선택 기법, 특징 개수, 모델에 따른 상위 10개 정확도 성능 비교
Table 2. accuracy performance by feature selection technique, feature count, and model

순위	특징 선택	특징 개수	모델	정확도	민감도	정밀도	F1-점수	AUC
1	MI	200	ET	92.538	93.268	92.468	92.744	92.516
2	SHAP	200	ET	92.521	91.046	94.561	92.482	92.582
3	SHAP	100	ET	92.521	91.046	94.444	92.563	92.582
4	MI	100	LB	91.966	93.268	91.856	92.38	91.928
5	PI	200	RF	91.966	93.333	91.542	92.302	91.961
6	PI	100	ET	91.966	91.046	93.601	92.106	91.993
7	RF	50	ET	91.966	89.935	94.496	91.908	92.026
8	MI	50	ET	91.95	91.046	93.667	91.998	91.993
9	MI	100	ET	91.95	91.046	93.203	92.063	91.993
10	SHAP	50	ET	91.95	89.935	94.444	91.881	92.026

표 3. MI(200)과 SHAP(200) 데이터셋에 대한 최적 모델 및 하이퍼파라미터
Table 3. best hyperparameters for MI(200) and SHAP(200) datasets

데이터셋	모델	최적 파라미터
MI (200)	ET	n_estimators: 100 max_depth: 10, min_samples_split: 2
SHAP (200)	ET	n_estimators: 100 max_depth: None, min_samples_split: 5

다음 단계로 앞 단계에서 우수한 성능을 보인 MI와 SHAP(ET)를 이용하여 3장의 하이브리드 특징 선택 기법을 반복 검증(N) 값은 10, 안전성 점수는 0.5 초과, 성능 감소 허용치는 0.99, 0.95, 0.9의 3가지로 특징 선택을 하였다. 그 결과로 각 성능 감소 허용치에 대하여 10개, 44개, 82개의 특징이 선택되었으며, 결과는 표 4와 같다.

표 4와 같이 상위 10개 특징은 주로 시간적 특징인 total_time과 air_time으로 구성되어 있으며, 상위 44개 특징에는 추가로 압력과 움직임 패턴(gmrt, speed) 관련 측정값이 추가되었다. 그리고 나머지 상위 82개의 38개 특징은 주로 위치 정보(extension)와 상세 움직임과 관련한 특징을 포함하였다.

다음으로 제안 방식으로 특징을 선택하여 구해진 세 가지 데이터셋과 기존 실험에서 우수한 성능을 보인 MI(200), SHAP(200), SHAP(100), MI (100), PI(200), PI(100), RF(50)을 가지고, 앞의 분류 실험과 동일한 방식으로 실험한 결과는 그림 2와 표 5와 같다.

실험 결과와 같이 제안한 특징 선택 기법(Proposed)은 82개의 특징을 사용한 ET 분류 모델에서 가장 높은 정확도인 93.66%를 구하였으며, 같은 데이터셋의 RF 모델에서도 동일한 정확도를 보였다. 특히 주목할 만한 점은 제안한 기법이 200개의 특징을 사용한 다른 기법들(MI: 92.54%, SHAP: 92.52%)보다 훨씬 적은 수의 특징으로도 더 우수한 성능을 보였다는 점이다.

세부적인 성능 지표를 살펴보면, Proposed의 82개 특징을 사용한 ET 모델은 93.27%의 민감도와 94.92%의 정밀도를 보여주었으며, F1-점수는 94.12%로 균형 잡힌 성능을 도출하였다. 특히 AUC 값이 93.79%로 나타나 전반적인 분류 성능이 우수하였다. Proposed의 82개 특징을 사용한 RF 모델의 경우에는 94.38%의 좀 더 높은 민감도를 보였으나, 정밀도는 93.97%로 약간 낮았다. MI(200)과 SHAP(200)은 92%로 82개 특징의 정확도보다 낮은 성능이며, 특히 44개 특징을 사용한 92.52%와 비교해도 유사하거나 낮은 수준이었다. 여기에 10개의 특징으로도 90.24%의 정확도를 보여, 특징 선택의 효율성이 있음을 보였다.

따라서 82개, 44개, 10개의 특징으로도 200개의 특징을 사용한 다른 기법 및 기존 관련 연구[11,12,13,14,15]에 비해서도 우수하거나 유사한 성능이 가능함을 제시하였다. 또한 모든 성능 지표에서 고른 분포를 보이며, 특히 F1-점수가 94% 이상을 유지한다는 점에서 분류의 안정성도 확인할 수 있었다.

5. 결 론

본 연구에서는 알츠하이머병의 조기 진단을 위한 DARWIN 필기 데이터셋에서 효과적인 특징 선택을 위한 하이브리드 특징 선택 기법을 제안하고 그 성능을 평가하였다. 제안 기법은 다양한 특징 선택 기법을 사용한 실험에서 우수한 성능을 보인 MI와 SHAP 특징 선택 기법을 결합하여 구성하였다.

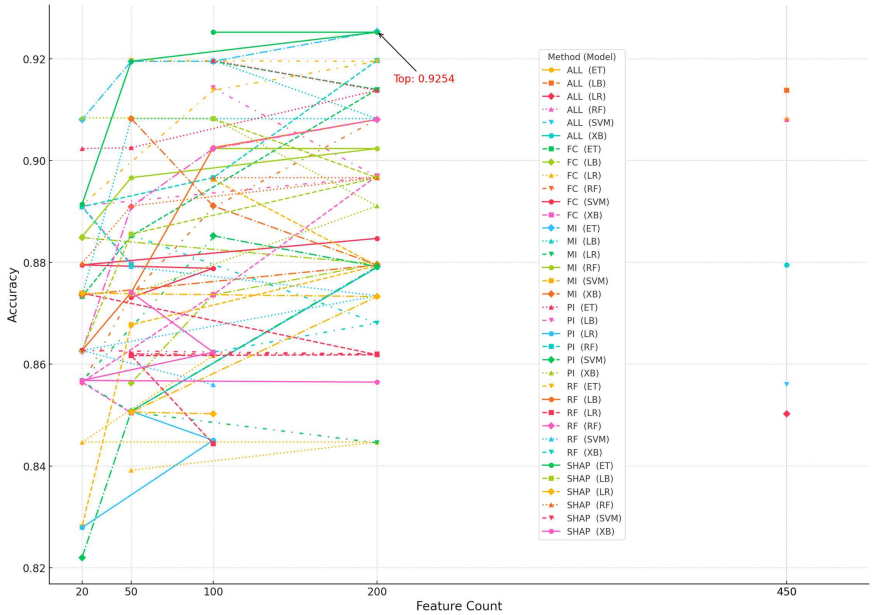

그림 1. 특징 선택 기법, 특징 개수, 모델에 따른 정확도 비교
Fig. 1. comparison of accuracy by feature selection technique, feature count and model

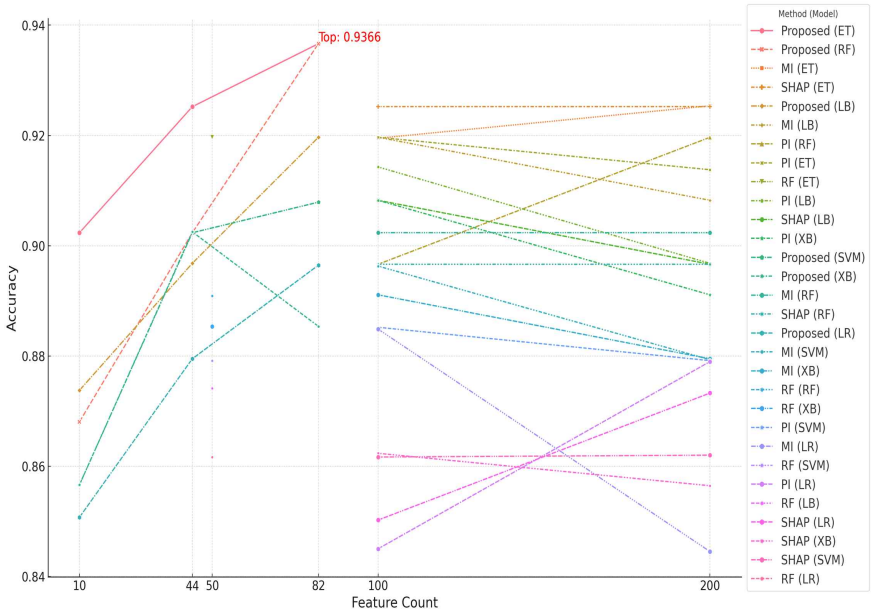

그림 2. 제안 특징 선택 기법과 기존 기법의 특징 개수, 모델에 따른 정확도 비교
Fig. 2. comparison of accuracy by proposed feature selection technique, feature count and model with existing techiniques

표 4. 성능 감소 허용치별 하이브리드 특징선택 기법으로 선택된 특징 목록
Table 4. selected features list by hybrid feature selection techniques according to performance degradation tolerance

성능감소 허용치 (α)	특징 개수	특징명
0.99	10	total_time13, total_time23, total_time15, air_time16, air_time15, pressure_var19, total_time22, num_of_pendown19, total_time9, air_time22
0.95	44	total_time15, air_time16, pressure_var19, num_of_pendown19, total_time23, air_time15, air_time22, total_time13, total_time9, air_time23, total_time22, paper_time9, paper_time23, disp_index23, total_time17, total_time16, pressure_mean21, total_time6, mean_gmrt17, mean_speed_on_paper10, gmrt_in_air23, air_time17, air_time19, gmrt_in_air17, pressure_mean9, total_time10, air_time21, disp_index22, paper_time11, gmrt_in_air7, mean_gmrt7, mean_jerk_in_air17, total_time2, paper_time15, total_time25, pressure_mean19, pressure_mean5, air_time5, paper_time22, mean_acc_in_air17, air_time20, paper_time6, max_x_extension19, air_time6
0.90	82	total_time15, air_time16, pressure_var19, num_of_pendown19, total_time23, air_time22, total_time13, total_time9, air_time15, total_time22, air_time23, paper_time9, disp_index23, total_time17, paper_time23, total_time6, total_time16, pressure_mean21, air_time21, mean_speed_on_paper10, mean_gmrt17, gmrt_in_air17, gmrt_in_air23, air_time19, air_time17, pressure_mean9, total_time10, total_time2, mean_jerk_in_air17, disp_index22, paper_time11, total_time25, paper_time15, pressure_mean5, pressure_mean19, paper_time10, gmrt_in_air7, mean_acc_in_air17, paper_time6, pressure_var5, paper_time22, mean_gmrt7, air_time20, mean_speed_in_air17, max_y_extension19, mean_gmrt14, air_time24, air_time5, air_time2, max_x_extension19, total_time18, total_time24, mean_speed_in_air23, air_time6, paper_time7, gmrt_on_paper10, air_time13, mean_speed_on_paper8, mean_gmrt19, max_y_extension21, mean_speed_in_air19, air_time11, paper_time17, mean_speed_on_paper19, mean_speed_on_paper7, paper_time19, mean_jerk_on_paper19, pressure_mean7, total_time19, pressure_mean8, total_time8, paper_time25, pressure_mean4, total_time7, disp_index6, paper_time12, mean_acc_on_paper21, gmrt_on_paper19, total_time11, total_time3, paper_time20, pressure_mean1

표 5. 제안 특징 선택 기법과 기존 기법의 상위 10개 정확도 비교 결과
Table 5. top 10 accuracy results comparing proposed feature selection techniques with existing techniques

순위	특징 선택	특징 개수	모델	정확도	민감도	정밀도	F1-점수	AUC
1	Proposed	82	ET	93.6639	93.268	94.9206	94.1176	93.7862
2	Proposed	82	RF	93.6639	94.3791	93.9683	92.9412	93.9737
3	MI	200	ET	92.5378	93.268	92.4683	91.7647	92.7437
4	SHAP	200	ET	92.521	91.0458	94.5614	94.1176	92.4816
5	Proposed	44	ET	92.521	91.0458	94.4444	94.1176	92.5628
6	SHAP	100	ET	92.521	91.0458	94.4444	94.1176	92.5628
7	Proposed	82	LB	91.9664	92.1569	92.3323	91.7647	92.2161
8	MI	100	LB	91.9664	93.268	91.8561	90.5882	92.3803
9	PI	200	RF	91.9664	93.3333	91.5418	90.5882	92.3024
10	PI	100	ET	91.9664	91.0458	93.6013	92.9412	92.1064

실험 결과, 82개의 특징만으로도 93.66%의 우수한 정확도를 구하였으며, 이는 200개의 특징을 사용한 기존 기법들 보다 우수한 성능을 보였다. 특히 44개의 특징만으로도 92.52%의 정확도를 보여, 특징 수를 약 78%를 감소시키면서도 다른 기법들과 동등한 수준의 성능을 유지할 수 있었다.

향후 연구에서는 이 기법을 다른 의료 데이터셋에 적용하여 일반화 가능성을 검증하고, 시계열 데이터셋으로 연구 범위를 확장하고자 한다. 더불어 과적합 가능성과 선택된 특징들의 임상적 의미를 심층적으로 분석하여, 알츠하이머병의 조기 진단을 위한 새로운 바이오마커 발굴에도 기여하고자 한다.

이 논문은 부산대학교
기본연구지원사업(2년)에 의하여 연구되었음.

참 고 문 헌

[1] World Alzheimer Report 2024: Global changes in attitudes to dementia, Alzheimer's Disease International, (2024.9)

[2] Ela Kaplan, Sengul Dogan, Turker Tuncer, Mehmet Baygin, Erman Altunisik, "Feed-forward LPQNet based Automatic Alzheimer's Disease Detection Model", Computers in Biology and Medicine, Vol. 137, (2021.10), DOI: 10.1016/j.compbiomed.2021.104828

[3] Alzheimer's disease facts and figures, Alzheimer's & dementia, the journal of the Alzheimer's Association, Vol. 18 No. 4, pp700 - 789, (2022), DOI:10.1002/alz.12638

[4] Dementia, (2024), World Health Organization, https://www.who.int/news-room/fact-sheets/detail/dementia

[5] Janice M. Ranson, Magda Bucholc, Donald Lyall, Danielle Newby, Laura Winchester, Neil P. Oxtoby, Michele Veldsman, Timothy Rittman, Sarah Marzi, Nathan Skene, Ahmad Al Khleifat, Isabelle F. Foote, Vasiliki Orgeta, Andrey Kormilitzin, Ilianna Lourida and David J. Llewellyn, "Harnessing the potential of machine learning and artificial intelligence for dementia research", Brain Inform, Vol. 10 No. 6, pp1-12, (2023), DOI: 10.1186/s40708-022-00183-3

[6] A. Godfrey, M. Brodie, K.S. van Schooten, M. Nouredanesh, S. Stuart, L. Robinson, "Inertial wearables as pragmatic tools in dementia", Maturitas, Vol. 127, pp12-17, (2019.9), DOI: 10.1016/j.maturitas.2019.05.010

[7] Xiaotong Sun, Xu Sun, Qingfeng Wang, Xiang Wang, Luying Feng, Yifan Yang, Ying Jing, Canjun Yang, Canjun Yang, Sheng Zhang, "Biosensors toward behavior detection in diagnosis of alzheimer's disease", Bioengineering and Biotechnology, Vol. 10, (2022.10), DOI: 10.3389/fbioe.2022.1031833

[8] Ikram Bazarbekov, Abdul Razaque, Madina Ipalakova, Joon Yoo, Zhanna Assipova and Ali Almisreb, "A review of artificial intelligence methods for Alzheimer's disease diagnosis: Insights from neuroimaging to sensor data analysis", Biomedical Signal Processing and Control, Vol. 92, (2024.6), DOI: 10.1016/j.bspc.2024.106023

[9] Nicole D. Cilia, Claudio De Stefano, Francesco Fontanella and Alessandra Scotto Di Freca, "An experimental protocol to support cognitive impairment diagnosis by using handwriting analysis", Procedia Computer Science, Vol. 141, pp.466-471, (2018.12), DOI: 10.1016/j.procs.2018.10.141

[10] Nicole D. Cilia, Giuseppe De Gregorio,

Claudio De Stefano, Francesco Fontanella, Angelo Marcelli and Alessandra Scotto Di Freca, “Diagnosing Alzheimer's disease from online handwriting: A novel dataset and performance benchmarking”, Engineering Applications of Artificial Intelligence, Vol. 111, (2022.3), DOI:10.1016/j.engappai.2022.104822

[11] Vincenzo Gattulli, Donato Impedovo, Giuseppe Pirlo and Gianfranco Semeraro, “Handwriting Task-Selection based on the Analysis of Patterns in Classification Results on Alzheimer Dataset”, IEEESDS'23: Data Science Techniques for Datasets on Mental and Neurodegenerative Disorders, pp18-29, (2023.6)

[12] Cansu Akyürek Anacur, Asuman Günay Yilmaz, Murat Aykut, “Handwriting Analysis for Alzheimer's Diagnosis: Feature Enrichment and Machine Learning Classification”, International Artificial Intelligence and Data Processing Symposium (IDAP), pp. 1-5, (2024.9), DOI:10.1109/IDAP64064.2024.10711121

[13] Pakize Erdogmus, Abdullah Talha Kabakus, “The promise of convolutional neural networks for the early diagnosis of the Alzheimer's disease”, Engineering Applications of Artificial Intelligence, Vol. 123, (2023.4), DOI: 10.1016/j.engappai.2023.106254

[14] L. Kang, X. Zhang, J. Guan, K. Huang and R. Wu, “Early Alzheimer's disease diagnosis via handwriting with self-attention mechanisms”, Journal of Alzheimer's Disease, Vol. 102, (2024.10), DOI:10.1177/13872877241283920

[15] Uddalak Mitra, Shafiq Ul Rehman, “ML-Powered Handwriting Analysis for Early Detection of Alzheimer's Disease”, pp. 69031-69050, IEEE Access, Vol.12, 15 (2024.5), DOI: 10.1109/ACCESS.2024.3401104

[16] Shaikh Shakeela, N Sai Shankar, P Mohan Reddy, T Kavya Tulasi, and M Mahesh Koneru, “Optimal ensemble learning based on distinctive feature selection by univariate ANOVA-F statistics for IDS”, International Journal of Electronics and Telecommunications, Vol. 67, No. 2, pp. 267-275, (2021.4), DOI: 10.24425/ijet.2021.135975

[17] Leo Breiman, “Random forests”, Machine Learning 45, pp. 5-32, (2001.10), DOI:10.1023/A:1010933404324

[18] Roberto Battiti, “Using mutual information for selecting features in supervised neural net learning”, IEEE Transactions on neural networks, Vol. 5, pp.537-550, (1994.7), DOI: 10.1109/72.298224

[19] Scott M. Lundberg, Su-In Lee, “A unified approach to interpreting model predictions”, Proceedings of the 31st International Conference on Neural Information Processing Systems, pp. 4768-4777, (2017.12), DOI: 10.48550

[20] André Altmann, Laura Toloşi, Oliver Sander and Thomas Lengauer, “Permutation importance: a corrected feature importance measure”, Bioinformatics, Vol. 26, (2010.5), pp. 1340 - 1347, DOI: 10.1093/bioinformatics/btq134

[21] Stanley Lemeshow, David W. Hosmer Jr., Rodney X. Sturdivant, Applied Logistic Regression 3rd., John Wiley & Sons, (2013), ISBN: 978-0-470-58247-3

[22] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye and Tie-Yan Liu, “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”, Advances in Neural Information Processing Systems 30, (2017)

[23] Vladimir Vapnik, The nature of statistical learning theory, Springer-Verlag, (1999)

[24] Pierre Geurts, Damien Ernst and Louis Wehenkel, “Extremely randomized trees”, Machine learning 63, pp. 3-42, (2006), DOI 10.1007/s10994-006-6226-1

[25] Chen, Tianqi, and Carlos Guestrin "Xgboost: A scalable tree boosting system", Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining, pp. 785-794, (2016),
DOI: 10.1145/2939672.2939785

저 자 소 개

김영인(Young-In Kim)

1996.2 명지대학교 컴퓨터공학과 박사
2007.8-2008.7 Univ. of Missouri 방문교수
2006.3-현재 : 부산대학교 교수
<주관심분야> 데이터베이스, 데이터마이닝