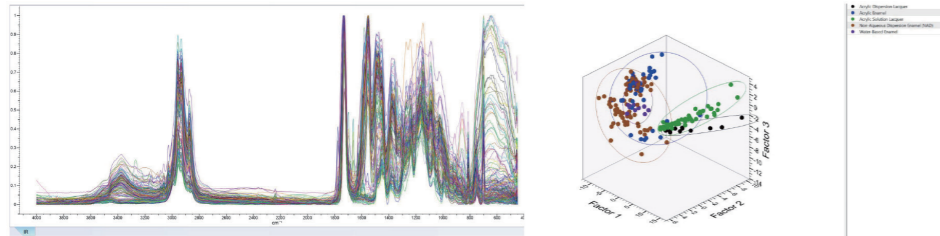


## 데이터셋 1: 자동차 도장 샘플의 IR 분광 분석

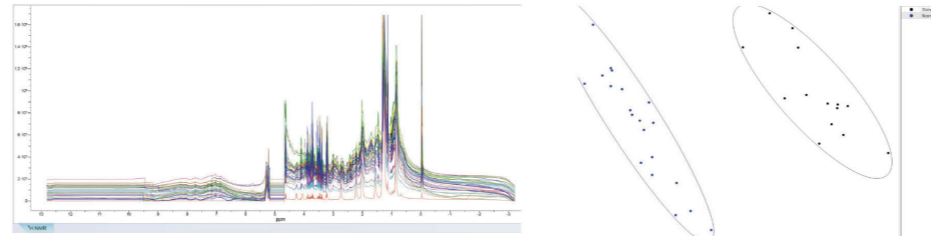
- **데이터셋:** Wiley “IR - Automobile Paint Chips” 데이터베이스의 초기 200개 IR 스펙트럼
- **PCA 파라미터:** 전처리 “mean-center” ; Y 변환 “2차 미분 (포인트 수 11)”
- **결과:** PCA 분석을 통해 도장 샘플은 에나멜과 라커 페인트 유형에 해당하는 두 개의 뚜렷한 클러스터로 성공적으로 분리되었다 (그림 1). 3D 스코어 플롯은 이러한 분리를 명확히 보여주며, 페인트 유형 간의 중첩을 최소화되었다. 이러한 클러스터링을 통해, 구축된 PCA 공간에 투영함으로써 미지의 도장 샘플을 자동으로 분류할 수 있다.



**그림 1 :** 자동차 도장 칩 200개 IR 스펙트럼에 대한 PCA 분석  
 왼쪽: 200개의 IR 스펙트럼  
 오른쪽: 에나멜과 라커 페인트가 명확히 분리된 3D PCA 스코어 플롯

## 데이터셋 2: 혈액 샘플의 <sup>1</sup>H NMR 분석

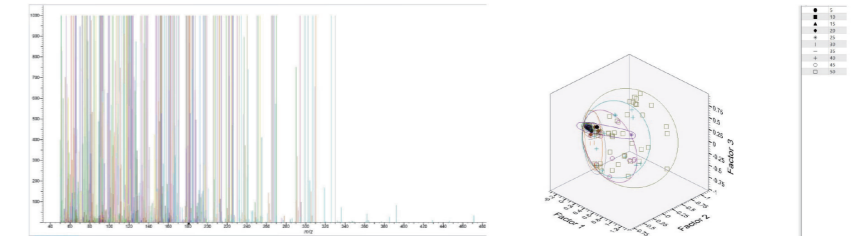
- **데이터셋:** 당뇨 환자 및 정상 혈액 샘플에서 획득한 34개의 <sup>1</sup>H NMR 스펙트럼
- **PCA 파라미터:** 전처리 “mean-center”; Y 변환 “divide-by Sample-2Norm”; <sup>1</sup>H NMR 스펙트럼의 해상도를 낮추기 위한 “Intelligent binning” 기법 적용
- **결과:** PCA 분석을 통해 2D 스코어 플롯에서 당뇨 환자와 정상 혈액 샘플 간의 뚜렷한 분리가 확인되었다 (그림 2). 또한 로딩 플롯을 통해 당뇨 상태와 관련된 바이오마커를 식별할 수 있다.



**그림 2 :** 당뇨 vs 정상 혈액 샘플의 PCA 분석  
 왼쪽: <sup>1</sup>H NMR 스펙트럼 데이터  
 오른쪽: 샘플 분류를 보여주는 PCA 2D 스코어 플롯

## 데이터셋 3: LC-MS tandem 질량 분석

- **데이터셋:** “Wiley Registry of Tandem Mass Spectral Data: MSforID” 데이터베이스의 유사 구조 검색 결과 200건을 사용
- **PCA 파라미터:** 전처리 “mean-center” ; Y 변환 “divide-by Sample Max”
- **결과:** 3D PCA 스코어 플롯에서 충돌 에너지와 스펙트럼 클러스터링 간의 명확한 관계가 확인되었다 (그림 3). 높은 충돌 에너지의 스펙트럼은 더 넓게 분산되었으며, 낮은 충돌 에너지의 스펙트럼은 밀집된 클러스터(검은 점)를 형성하였다. 이로 인해 더욱 다양하고 차별화된 질량 스펙트럼이 생성되었다. 이러한 패턴은 높은 충돌 에너지에서 분자 단편화가 증가함을 반영한다.



**그림 3 :** tandem 질량 분석 데이터의 PCA 분석  
 왼쪽: 200개 MS 스펙트럼  
 오른쪽: 에너지 의존적 클러스터링을 보여주는 PCA 3D 스코어 플롯

## 초록 (Abstract)

본 연구에서는 주성분 분석(PCA)을 활용하여 다학제적 분석 기법으로부터 도출된 스펙트럼 데이터셋을 종합적으로 고찰하였다. 분석 결과, 단일 스펙트럼 분석으로는 식별이 어려운 잠재적 패턴과 변수 간의 상관관계를 성공적으로 규명하였다. 이는 PCA가 고차원 데이터로부터 핵심 인사이트를 추출함으로써 과학적 발견을 가속화하는 유용한 도구임을 시사한다. 또한, 본 연구는 PCA가 스펙트럼 클러스터와 고유 특성 간의 상관관계를 효과적으로 식별함을 입증하였으며, 결과적으로 분석 워크플로우의 최적화 및 효율성 제고에 기여하는 실질적 가치를 제시한다.

## 서론 (Introduction)

주성분 분석(PCA)은 고차원 데이터셋의 변동성을 최대한 보존하면서 차원을 축소하여 복잡한 데이터를 단순화하는 강력한 통계적 방법론이다. 특히 분광학 데이터에 PCA를 적용할 경우, 기존의 개별 스펙트럼 분석으로는 식별하기 어려운 잠재적 패턴과 트렌드를 효과적으로 도출할 수 있다. 본 포스터에서는 IR, NMR, MS를 포함한 다양한 데이터베이스 컬렉션에 PCA를 적용한 사례를 제시한다. 또한, 본 기법은 Raman, UV-Vis 및 크로마토그래피 데이터 등 광범위한 분석 영역으로 확장 적용이 가능함을 시사한다.

## 방법 및 결과(Method & Results)

본 연구의 모든 데이터 분석은 KnowItAll 2026 소프트웨어의 TrendFinder 애플리케이션을 사용하여 수행되었다. 서로 다른 분석 기법 전반에 걸친 PCA(주성분 분석)의 범용적 활용 가능성을 검증하기 위해, 각기 다른 특성을 가진 세 가지 독립적인 데이터셋을 선정하여 분석을 진행하였다.

## 결론 (Conclusion)

본 연구를 통해 주성분 분석(PCA)이 대규모 분광 데이터셋의 처리 및 해석에 있어 매우 강력한 통계적 도구임을 입증하였다. PCA는 기존의 개별 스펙트럼 분석법을 보완하여 데이터 내 잠재된 트렌드를 효과적으로 시각화하며, 이는 수학, 컴퓨터 공학, 그리고 화학적 통찰이 결합된 다학제적 분석의 정수를 보여준다. 본 연구는 PCA가 다양한 분광 기법에서 상호 보완적인 분석 도구로서 높은 효율성을 가짐을 확인하였으며, 특히 다음의 세 가지 측면에서 유의미한 성과를 거두었다.

- **분류 성능 :** 사전 정보가 배제된 상태에서도 이질적인 샘플 유형을 명확히 식별 및 구분함.
- **방법론적 범용성 :** 다중 분광 분석 기법 전반에 걸쳐 일관된 분석 유효성을 확보함.
- **물리적 인사이트 :** 단순한 통계 수치를 넘어 실험 파라미터와 스펙트럼 특성 간의 실질적인 상관관계를 규명함.

본 결과는 패턴 인식 및 시료 분류를 통해 분석 워크플로우를 향상시키는 데 있어, PCA의 광범위한 적용 가능성을 시사한다.