



(19) 대한민국특허청(KR)  
(12) 등록특허공보(B1)

(45) 공고일자 2021년04월05일  
(11) 등록번호 10-2235934  
(24) 등록일자 2021년03월30일

(51) 국제특허분류(Int. Cl.)  
G01N 21/17 (2006.01) G01N 21/84 (2006.01)  
G06N 3/08 (2006.01)  
(52) CPC특허분류  
G01N 21/17 (2013.01)  
G01N 21/84 (2013.01)  
(21) 출원번호 10-2019-0150069  
(22) 출원일자 2019년11월21일  
심사청구일자 2019년11월21일  
(65) 공개번호 10-2020-0069222  
(43) 공개일자 2020년06월16일  
(30) 우선권주장  
1020180156028 2018년12월06일 대한민국(KR)

(73) 특허권자  
세종대학교 산학협력단  
서울특별시 광진구 능동로 209 (군자동, 세종대학교)  
(72) 발명자  
손기선  
서울특별시 강남구 언주로146길 18, 4동 302호  
이진웅  
부산광역시 해운대구 운봉길 77-7  
(74) 대리인  
특허법인아이엠

(56) 선행기술조사문헌  
Felipe Oviedo 등, Fast classification of small X-ray diffraction datasets using physics-based data augmentation and deep neural networks, NIPS 2018, pp1-11, 2018.11.20.\*  
JP2017522563 A  
JP2004502934 A  
JP2000516342 A  
\*는 심사관에 의하여 인용된 문헌

전체 청구항 수 : 총 5 항

심사관 : 정치영

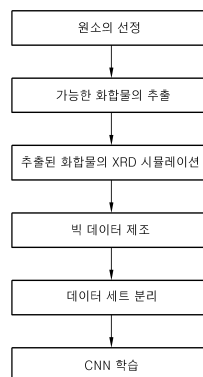
(54) 발명의 명칭 물질의 판별 및 분석 방법

(57) 요약

본 발명은 무기물 및 유기물 등 다양한 물질에 포함된 원소, 상(phase), 조성 등의 판별 및 분석이 용이한 방법에 관한 것이다.

본 발명에 따른 방법은, (a) 2 이상의 원소를 선택하는 단계; (b) 상기 2 이상의 원소로 생성 가능한 것으로 분 (뒷면에 계속)

대표도 - 도1



석된 복수의 화합물의 데이터를 수집하는 단계; (c) 상기 수집된 복수의 화합물 각각에 관한 이미지 또는 스펙트럼 형태의 분석 데이터를 준비하는 단계; (d) 상기 복수의 화합물 중에서, 2종 이상의 화합물을 선택하여 소정의 혼합 비율로 혼합하고, 상기 각각 이미지 또는 스펙트럼 데이터를 상기 소정의 혼합 비율에 맞추어 혼합 가공한 것을 포함하여 훈련 데이터를 만드는 단계; (e) 상기 훈련 데이터를 사용하여 기계 학습을 수행하는 단계; 및 (f) 실제 물질로부터 얻은 이미지 또는 스펙트럼 형태의 분석 데이터를 상기 기계 학습을 통해 획득한 모델을 사용하여 판별 및/또는 분석하는 단계를 포함한다.

(52) CPC특허분류

*G06N 3/08* (2013.01)

*G01N 2021/1765* (2013.01)

*G01N 2021/8405* (2013.01)

이 발명을 지원한 국가연구개발사업

과제고유번호	1711065337
부처명	과학기술정보통신부
과제관리(전문)기관명	한국연구재단
연구사업명	미래소재디스커버리지원
연구과제명	휴리스틱스 전산기반 기능성 무기소재 창제 연구
기 여 율	1/1
과제수행기관명	세종대학교 산학협력단
연구기간	2018.02.04 ~ 2019.02.03

---

**명세서**

**청구범위**

**청구항 1**

- (a) 2 이상의 원소를 선택하는 단계;
  - (b) 상기 2 이상의 원소로 생성 가능한 것으로 분석된 복수의 화합물의 데이터를 수집하는 단계;
  - (c) 상기 수집된 복수의 화합물 각각에 관한 이미지 또는 스펙트럼 형태의 분석 데이터를 준비하는 단계;
  - (d) 상기 복수의 화합물 중에서, 2종 이상의 화합물을 선택하여 소정의 혼합비율로 혼합하고, 상기 각각 이미지 또는 스펙트럼 데이터를 상기 소정의 혼합 비율에 맞추어 혼합 가공한 것으로 훈련 데이터를 만드는 단계;
  - (e) 상기 훈련 데이터를 사용하여 기계 학습을 수행하는 단계; 및
  - (f) 실제 물질로부터 얻은 이미지 또는 스펙트럼 형태의 분석 데이터를 상기 기계 학습을 통해 수득한 모델을 사용하여 판별 및/또는 분석하는 단계를 포함하고,
- 상기 (f) 단계를 통해, 2종 이상의 화합물이 혼합되어 있는 혼합물로부터 각 화합물의 상과 상 분율을 판별하는, 물질의 판별 및 분석방법.

**청구항 2**

제1항에 있어서,  
 상기 복수의 화합물의 분석 데이터는, 기존에 분석되어 있는 화학분석, 물질분석 데이터를 이용하는, 물질의 판별 및 분석방법.

**청구항 3**

제1항에 있어서,  
 상기 복수의 화합물에 관하여 각각의 이미지 또는 스펙트럼 형태의 분석 데이터를 준비하는 단계는, 실제 분석된 결과를 사용하거나, 각각의 화합물에 관한 물질 정보를 이용하여 프로그램을 사용하여 해당 분석 이미지 또는 스펙트럼을 시뮬레이션하여 획득하는, 물질의 판별 및 분석방법.

**청구항 4**

제1항에 있어서,  
 상기 이미지 또는 스펙트럼 형태의 분석 데이터는, XRD 데이터, XPS 데이터, IR 데이터 또는 투과전자현미경 회절 패턴인, 물질의 판별 및 분석방법.

**청구항 5**

제1항에 있어서,  
 상기 기계 학습은 CNN(Convolutional Neural Network)을 통해 수행되는, 물질의 판별 및 분석방법.

**발명의 설명**

**기술 분야**

[0001] 본 발명은 무기물 및 유기물 등 다양한 물질에 포함된 원소, 상(phase), 조성 등의 판별 및 분석을 효율적으로 수행할 수 있도록 하는 방법에 관한 것이다.

**배경 기술**

[0003] 분말 분석 기술에 기초한 물질 발견 과정에서 가장 빈번하게 직면하는 상황 중 하나는 알려지지 않은 다상 화합

물의 식별 및 정량화와 관련이 있다.

- [0004] 일반적으로 물질에 포함된 원소, 상(phase), 성분 등의 분석에는, XRD, XPS, EDS 등의 다양한 분석장비가 사용되고 있다. 그런데, 상기 분석장비를 통해 수득한 분석결과는 해당 기술분야에 전문적인 지식과 많은 경험을 가진 사람이 상당한 시간을 들여야만 정확한 분석결과를 얻는 경우가 많다.
- [0005] 물질 분석 시간을 단축하기 위하여, 개발되어 있는 다양한 분석 소프트웨어가 있다. 예를 들어, 이론 결정학적 이론을 기반으로 하는 알고리즘을 통한 분석이 있으나 충분한 정확도를 제공하지 못한다. 또한, 기계 학습 방법을 이용하는 것도, 훈련 데이터셋이 소규모이고, 이론적인 지식에 기반한 과도한 피쳐 엔지니어링을 사용하기 때문에, 기존의 규칙 기반의 분석 소프트웨어와 큰 차이가 없는 결과를 제공한다.
- [0006] 또한, 기존의 분석 소프트웨어는, 해당 물질을 포함하는 제조 공정 라인에서 실시간으로 제조되고 있는 물질에 관한 정보를 정확하게 도출할 수 있는 것은 거의 없다고 할 수 있다.

**선행기술문헌**

**특허문헌**

- [0008] (특허문헌 0001) 대한민국 공개특허공보 제10-2018-0073819호

**발명의 내용**

**해결하려는 과제**

- [0009] 본 발명의 과제는 무기물 및 유기물 등 다양한 물질에 포함된 원소, 상(phase) 또는 함량 등의 판별과 분석이 단시간 내에 높은 정확도로 이루어질 수 있도록 하는, 물질의 판별 및 분석 방법을 제공하는데 있다.

**과제의 해결 수단**

- [0011] 상기 과제를 해결하기 위해 본 발명은, (a) 2 이상의 원소를 선택하는 단계; (b) 상기 2 이상의 원소로 생성 가능한 것으로 분석된 복수의 화합물의 데이터를 수집하는 단계; (c) 상기 수집된 복수의 화합물 각각에 관한 이미지 또는 스펙트럼 형태의 분석 데이터를 준비하는 단계; (d) 상기 복수의 화합물 중에서, 2종 이상의 화합물을 선택하여 소정의 혼합 비율로 혼합하고, 상기 각각 이미지 또는 스펙트럼 데이터를 상기 소정의 혼합 비율에 맞추어 혼합 가공한 것을 포함하여 훈련 데이터를 만드는 단계; (e) 상기 훈련 데이터를 사용하여 기계 학습을 수행하는 단계; 및 (f) 실제 물질로부터 얻은 이미지 또는 스펙트럼 형태의 분석 데이터를 상기 기계 학습을 통해 수득한 모델을 사용하여 판별 및/또는 분석하는 단계를 포함하는, 물질의 판별 및 분석 방법을 제공한다.
- [0012] 상기 방법에 있어서, 상기 2 이상의 원소를 선택하는 단계에 있어서, 바람직한 원소의 수는 2 ~ 10이고, 보다 바람직한 원소의 수는 3 ~ 8이다.
- [0013] 상기 방법에 있어서, 상기 복수의 화합물의 데이터는 현재까지 수행되어 있는 화학분석, 물질분석 데이터를 이용할 수 있다. 예를 들어, 무기물의 경우, ICSD에 등록된 화합물을 사용할 수 있다.
- [0014] 상기 방법에 있어서, 상기 복수의 화합물에 관하여 각각의 이미지 또는 스펙트럼 형태의 분석 데이터를 준비하는 단계는, 실제 분석된 결과를 사용하거나, 각각의 화합물에 관한 물질 정보를 이용하여 프로그램을 사용하여 해당 분석 이미지 또는 스펙트럼을 계산한 것을 사용할 수도 있다.
- [0015] 상기 방법에 있어서, 상기 소정의 혼합 비율은 예를 들어, A 화합물, B 화합물, C 화합물이 3종이 존재할 경우, 상기 3종의 화합물의 함량이 질량%로 0.5% 단위로 변경되도록 하고, 그에 맞추어 이미지 또는 스펙트럼 데이터를 가공하는 형태일 수 있으나, 분석할 물질의 특성 및 구비되는 데이터의 사정에 맞추어 조절될 수 있으며, 반드시 이에 제한되는 것은 아니다.
- [0016] 상기 방법에 있어서, 상기 이미지 또는 스펙트럼 형태의 분석 데이터는 기계 학습이 가능한 형태라면 특별히 제한되지 않으며, 예를 들어, XRD 데이터, XPS 데이터, IR 데이터, 투과전자현미경 회절 패턴 등이 사용될 수 있다.
- [0017] 상기 방법에 있어서, 상기 기계 학습은 다양한 형태의 알고리즘이 적용될 수 있으며, 예를 들어

CNN(Convolutional Neural Network)을 통해 수행될 수 있다.

**발명의 효과**

- [0019] 본 발명에 따르면, 임의의 원소로 이루어진 성분 시스템을 설정하고, 이 시스템으로 가능하며 알려진 모든 화합물의 분석 데이터를 이용하여, 상기 모든 화합물의 혼합 비율을 미리 상정하여 이미지 또는 스펙트럼 데이터를 가공하여 훈련 데이터셋으로 기계 학습을 한 후, 기계 학습의 결과물인 예측모델을 통해서, 실제 물질의 판별 및/또는 분석이 이루어지므로, 단시간 내에 정확한 분석 결과를 도출할 수 있다. 예를 들어, 철강 공장 라인에서, XRD 데이터 결과를 본 발명에서 수득한 피팅 결과를 사용할 경우 실시간으로 상(phase) 분석이 가능하여 불량 발생 여부를 확인할 수 있게 된다.
- [0020] 또한, 본 발명에 따른 물질의 판별 및/또는 분석 방법은, 종래 개발된 이론 기반의 소프트웨어 방식 등에 비해, 다양한 성분이 혼합되어 있는 실제 물질의 상 분석 및 상 분율의 분석을 높은 정확도로 수행할 수 있게 한다.

**도면의 간단한 설명**

- [0022] 도 1은 본 발명의 실시예에 따른 물질의 판별 방법에 관한 흐름도이다.
- 도 2는 본 발명의 실시예에서 사용된 화합물을 나타낸 것이다.
- 도 3은 본 발명의 일 실시형태에 따라, ICSD 결정구조 데이터로부터 XRD 구성 파라미터를 이용하여 특정 화합물의 XRD 데이터를 계산하여 도출하는 단계를 설명하는 도면이다.
- 도 4는 분석하고자 하는 성분을 포함하는 무기 화합물인  $Al_2O_3$ ,  $Li_2O$ ,  $SrO$  및  $SrAl_2O_4$ 의 실제로 분석한 XRD 패턴과, 상기 방법을 통해 시뮬레이션된 XRD 패턴을 비교한 것이다.
- 도 5는 각 화합물의 혼합 데이터를 도출하는 과정을 나타낸 것이다.
- 도 6은 도 5에서 도출된 혼합 비율에 맞추어 XRD 데이터를 가공하는 과정을 나타낸 것이다.
- 도 7은 혼합 XRD 데이터를 도출하는 구체적인 조건을 나타낸 것이다.
- 도 8은 본 발명의 실시예에서 딥러닝에 사용한 제1 CNN 아키텍처를 나타낸 것이다(a는 CNN2, b는 CNN3).
- 도 9는 본 발명의 실시예에서 딥러닝에 사용한 제2 CNN 아키텍처를 나타낸 것이다.
- 도 10은 'Dataset\_800k\_org'를 사용한 경우 학습 코스트 및 정확도를 나타낸 것이다.
- 도 11은 'Dataset\_800k\_rand' 및 'Dataset\_180k\_rand'를 사용한 경우 학습 코스트 및 정확도를 나타낸 것이다.

**발명을 실시하기 위한 구체적인 내용**

- [0023] 이하 본 발명의 실시예에 대하여 첨부된 도면을 참고로 그 구성 및 작용을 설명하기로 한다. 하기에 본 발명을 설명함에 있어, 관련된 공지 기능 또는 구성에 대한 구체적인 설명이 본 발명의 요지를 불필요하게 흐릴 수 있다고 판단되는 경우에는 그 상세한 설명을 생략할 것이다. 또한, 어떤 부분이 어떤 구성요소를 '포함'한다고 할 때, 이는 특별히 반대되는 기재가 없는 한 다른 구성요소를 제외하는 것이 아니라 다른 구성요소를 더 포함할 수 있는 것을 의미한다.
- [0024] 도 1은 본 발명의 실시예에 따른 물질의 판별 방법에 관한 흐름도이다. 도 1에 도시된 것과 같이, 본 발명의 실시예에 따른 물질의 판별 방법은, 분석하고자 하는 물질을 구성하는 원소를 선정하는 단계와, 선정된 원소로 가능한 화합물에 관한 정보를 추출하는 단계와, 추출된 각 화합물에 대한 XRD 데이터를 시뮬레이션하는 단계와, 시뮬레이션된 XRD 데이터를 소정 비율로 혼합하여 기계학습용 빅데이터를 제조하는 단계와, CNN 아키텍처를 통해 기계학습을 수행하는 단계를 포함하여 이루어진다.
- [0025] 원소 선정 및 화합물 데이터의 수집
- [0026] 본 발명의 실시예에서는, 분석하고자 하는 물질에 포함되는 원소로, 무기 발광 물질의 전형적인 성분인 스트론튬(Sr), 알루미늄(Al), 리튬(Li) 및 산소(O)를 선정하였다.
- [0027] 다음으로, 상기 4가지 원소를 포함하는 화합물 정보를 무기결정구조 데이터(ICSD, Inorganic Crystal Structure Data)를 통해 확인한 결과 총 174개가 등재되어 있으며, 이중 5개는 잘못된 것이므로 제외하였다. 여기에 2018

년 ICSD 버전에는 포함되지 않았으나, 최근에 발견된 4원계 화합물로 LED용 발광 재료로 유망한  $\text{Sr}_2\text{LiAlO}_4$ 를 추가하였다.

- [0028] 이러한 과정을 통해, 스트론튬(Sr), 알루미늄(Al), 리튬(Li) 및 산소(O)를 포함하는 총 170개의 화합물 정보를 추출하였으며, 추출된 무기 화합물을 도 2에 나타내었다.
- [0029] 본 발명의 실시예에서는 분석하고자 하는 물질을 무기물로 선정하여 ICSD 데이터를 이용하였으나, 다른 형태의 물질의 경우, 다른 형태의 데이터를 사용할 수 있음은 물론이다.
- [0031] 화합물 분석을 위한 XRD 데이터 시뮬레이션
- [0032] 이상과 같은 화합물의 분석에는 다양한 분석방법이 사용될 수 있으므로, 화합물의 분석에 적용하고자 하는 분석 방법을 선택하고 관련 데이터의 확보가 필요하다. 물질의 분석에는 XRD 데이터, XPS 데이터, 또는 IR 데이터 등 다양한 데이터가 사용될 수 있는데, 본 발명의 실시예에서는, 무기물의 상(phase)을 판별하는 분석에 일반적으로 사용되고 있는 XRD를 물질 판별용 분석 데이터로 선택하였다.
- [0033] 다음으로, 170개 화합물에 관한 XRD 데이터가 있을 경우, 해당 XRD 데이터를 사용할 수도 있으나, 본 발명의 실시예에서는 충분한 XRD 데이터가 없더라도, 기계학습에 충분한 효과를 구현할 수 있도록, 시뮬레이션 방법을 통해 XRD 데이터를 준비하였다.
- [0034] 도 3은 ICSD 결정구조 데이터로부터 특정 화합물의 XRD 데이터를 시뮬레이션하는 과정을 나타낸 것이다.
- [0035] 본 발명의 실시예에서는 백그라운드 시뮬레이션을 위한 총 6개의 무작위 파라미터를 선택하고, 피크 모양 시뮬레이션을 위해서는 총 5개의 무작위 파라미터를 사용하며, 랜덤 노이즈 데이터를 부여하는 방법을 통해, 실제와 매우 유사한 시뮬레이션된 XRD 패턴 데이터를 도출하였다. 이때, ICSD 결정 정보로부터 XRD 패턴을 도출할 때 격자상수를 무작위로 변형시켜 보다 많은 수의 XRD 패턴을 도출할 수 있도록 하였다.
- [0036] 시뮬레이션된 XRD 패턴은, ICSD에서 얻은 구조 팩터(structural factor) 및 열적 팩터(thermal factor)와, 다중도(multiplicity), 로렌츠 폴라리제이션 팩터(Lorentz and polarization factor), 흡수도(absorption) 및 우선 방향(preferred orientation)을 고정 변수로 사용하고, 동시에 피크 프로파일(Caglioti 및 믹싱 매개 변수), 배경 및 화이트 노이즈와 같은 조정 가능한 매개 변수를 사용하여 얻었다.
- [0037] 구체적으로, ICSD에 제시된 시메트리(symmetry) 데이터로부터 다중도를 얻고, 입사 빔에 흑연 모노 크로메이터가 장착된 브래그-브란타노(Bragg-Brantano) 지오메트리를 사용하여 일반 실험용 XRD에 폴라리제이션 보정이 적용되었으며, 우선 배향은 존재하지 않는 것으로 간주하였다. 또한, 실험실 규모의 X-선 회절계에 흔히 나타나는 파라미터 값을 참조하여 조정 가능한 변수(parameter)를 무작위로 할당하였다.
- [0038] 도 4는 분석하고자 하는 성분을 포함하는 무기 화합물인  $\text{Al}_2\text{O}_3$ ,  $\text{Li}_2\text{O}$ ,  $\text{SrO}$  및  $\text{SrAl}_2\text{O}_4$ 의 실제로 분석한 XRD 패턴과, 상기 방법을 통해 시뮬레이션된 XRD 패턴을 비교한 것이다. 도 4에서 위에 표시된 패턴(녹색)은 실제 실험을 통해 분석된 결과를 나타낸 것이고, 아래에 표시된 패턴(갈색)은 상기 방법으로 시뮬레이션된 XRD 패턴을 나타낸 것이다. 도 4에 나타낸 것과 같이, 시뮬레이션된 XRD 패턴과 실제 XRD 패턴 사이에 차이를 구분하기 어려운 정도로 매우 유사한 형태를 나타내었다. 이를 통해, 본 발명의 실시예에 따라 시뮬레이션된 XRD 패턴은 기계 학습용 훈련 데이터에 적용될 수 있다고 보여지며, 이는 시뮬레이션 XRD 패턴을 이용한 기계 학습을 통해 수립한 모델을 사용하여 실제 물질의 정확한 판별로 이어지는 결과를 통해서도 확인된다.
- [0039] 이상과 같은 과정을 통해, 총 170개 화합물에 대한 다양한 시뮬레이션 XRD 패턴을 도출하였다.
- [0041] 훈련 데이터셋 작성
- [0042] 도 5는 각 화합물의 혼합 데이터를 도출하는 과정을 나타낸 것이고, 도 6은 도 5에서 도출된 혼합 비율에 맞추어 XRD 데이터를 가공하는 과정을 나타낸 것이고, 도 7은 혼합 데이터를 도출하는 구체적인 조건을 나타낸 것이다.
- [0043] 도 5에 도시된 것과 같이, A 화합물, B 화합물, C 화합물을 이용한 1중(unary), 2중(binary), 3중(ternary) 혼합물에 대한 조성 선택과 혼합 비율의 할당을 통해 다양한 성분과 혼합 비율을 가진 혼합물이 도출되도록 하였다.
- [0044] 2중 및 3중 혼합은, 도 2에 나타낸 그룹을 기반으로, 170개가 아닌 38개의 클래스로 구성된 조합을 기반으로 생성하였다. 그런 다음 클래스를 선택할 때마다 각 구성 요소에 170개의 변형물 중 하나를 무작위로 할당하였다.



예를 들어,  $Al_2O_3$ 가 관심 혼합물의 성분으로 선택되었을 때, 74개의 변형 중 하나를  $Al_2O_3$ 에 무작위로 할당하였다. 이를 통해  $Al_2O_3$ 는 74개의 변형이 적어도 한 번은 나타나도록 하였다. 또한,  ${}_{38}C_3 \times 21 + {}_{38}C_2 \times 9 + {}_{38}C_1$  (183,521)의 상이한 혼합물에 대해, 3중 혼합물은 21개의 고정된 조성을 가지고, 2중 혼합물은 9개의 고정된 조성을 가진다. 즉, 혼합물당 평균 4배 반복을 초래할 수 있으나, 많은 변형이 있는 일부 혼합물에 대해서만 반복적인 선택을 채택했기 때문에, 전체 데이터세트에서 동일한 혼합물(즉, 동일한 성분 및 동일한 분율)을 거의 찾을 수 없었다. 이러한 방식으로 데이터 준비를 하지 않으면, 38개의 클래스 각각이 동일한 선택 기회를 가지기 때문에, 예를 들어 변형상이 없는  $Li_2O_2$ 에 비해 74개의 변형상들이 있는  $Al_2O_3$ 의 각 상이 선택될 가능성이 훨씬 적어지는 문제가 발생할 수 있다.

[0045] 이상과 같은 과정을 통해 혼합 비율이 도출되면, 도 6에 도시된 방법을 통해, 혼합물의 XRD 패턴을 도출한다. 도 6에 도시된 바와 같이, A 화합물로 시물레이션된 XRD 패턴과 B 화합물로 시물레이션된 XRD 패턴과 C 화합물로 시물레이션된 XRD 패턴을, 혼합물의 혼합 비율에 맞추어 XRD 패턴을 더하는 방법을 통해, 혼합물의 XRD 패턴을 도출하였다.

[0046] 혼합물의 시물레이션된 XRD 패턴은, 2중 또는 3중 혼합물 조합을 구성하기 위해 170개 엔트리에서 2 ~ 3개의 화합물을 선택할 때마다, 피크 프로파일이 무작위로 선택되게 하여 총 800,942개의 XRD 패턴 사이에 동일한 것이 존재하지 않도록 하였다. 이 과정에 사용된 구체적인 조건은 도 7에 나타내었다. 이때, 동일한 클래스(class)에 속하는 엔트리로 구성된 2중 또는 3중 혼합물 조합은 모두 제외하였다. 또한, 임의로 선택된 배경 및 백색 잡음이 각 혼합물에 적용되었다.

[0047] 또한, 도출된 XRD 패턴은, 무작위로 훈련용 데이터세트(600,942 패턴), 유효성 검사용 데이터세트(100,000 패턴) 및 테스트 데이터세트(100,000 패턴)로 각각 분할되었다.

[0048] 또한, 데이터 유사성이 없도록 랜덤 조성 선택 및 반복 선택이 없는 조건으로, 상기한 방법에 비해 축소된 데이터세트를 별도로 준비하였고, 이 축소된 데이터세트의 혼합물 수는  ${}_{38}C_3 \times 21 + {}_{38}C_2 \times 9 + {}_{38}C_1$  (183,521)로 줄었다.

[0049] 이상과 같은 과정을 통해 3종류의 데이터세트가 준비되었다. 도 5의 위에 나타난 것과 같이, A 화합물, B 화합물, C 화합물의 3중 21개, 2중 9개를 등간격으로 나누고, 도 2에 나타난 물질 중에 ICSD 개수가 많이 포함되어 있는 혼합물은 도 7에 나타난 조건에 따라 반복하여 생성한 800,942개의 XRD 패턴 데이터를 'Dataset\_800k\_org'라고 한다.

[0050] 도 5의 아래에 나타난 것과 같이, A 화합물, B 화합물, C 화합물의 3중 21개, 2중 9개를 랜덤하게 나누고, 도 2에 나타난 물질 중에 ICSD 개수가 많이 포함되어 있는 혼합물은 도 7과 같은 조건에 따라 반복하여 생성한 800,942개의 XRD 패턴 데이터를 'Dataset\_800k\_Rand'라고 한다.

[0051] A 화합물, B 화합물, C 화합물의 3중 21개, 2중 9개를 랜덤하게 나누고, 반복 횟수 없이 생성된 183,521개의 XRD 패턴 데이터를 'Dataset\_180k\_Rand'라고 한다.

[0053] CNN을 적용한 딥러닝

[0054] 본 발명의 실시예에서는 데이터세트 크기가 크고, 다상 XRD 패턴을 다루므로, 이의 분석에 적합한 CNN을 적용하여 기계학습을 수행하였다.

[0055] 도 8 및 도 9는 본 발명의 실시예에서 딥러닝에 사용한 2가지 CNN 아키텍처를 나타낸 것이다.

[0056] 도 8 및 도 9에 나타난 것과 같이, CNN 아키텍처는, 여러 컨볼루션 레이어(convolution layer)와 풀링 레이어(pooling layer)와 3개의 연속적으로 연결된 전결합층(fully connected layer)으로 구성된다.

[0057] 구체적으로, 도 8에 도시된 것과 같이, 2개의 컨볼루션 레이어가 있는 아키텍처를 'CNN\_2'라고 하고, 도 9에 도시된 것과 같이, 3개의 컨볼루션 레이어가 있는 아키텍처를 'CNN\_3'라고 한다. 이 두가지 아키텍처에 있어서, 필터 수, 커널 크기, 풀링 크기, 스트라이드의 구성은 도 8 및 도 9에 도시된 것과 같다.

[0058] 한편, 도 8에 나타난 것과 같이, 본 발명의 실시예에서는, 스트라이드가 컨볼루션 레이어 중 하나의 풀링 크기보다 넓은 것을 채택하였는데, 이러한 구성은 기존의 풀링보다 현저하게 빠른 수렴을 제공하였다. 한편, CNN2 및 CNN3의 풀링 크기와 동일한 스트라이드로도 테스트한 결과 수렴은 상대적으로 느리지만 정확도는 실질적으로 동일한 수준으로 나타났다.

[0059] 또한, 컨볼루션 레이어를 위한 활성화 함수와 전결합층을 위한 선형 함수로 정류선형유닛(ReLu)이 채택되었고,

드롭 아웃은 전결합층에 대해서만 구현되었다. 전결합층에 대한 최종 활성화 기능은 시그모이드 기능이며, 그에 따른 코스트(또는 손실) 기능은 교차 엔트로피 기능이었다. 입력은 4501×1 벡터를, 출력은 38×1 벡터로 하였다. 또한, Adam 옵티 마이저를 사용하였다. 상 식별을 위한 모든 에포크(epoch)에 대해 실행 속도는 0.001로 고정되었다.

[0060] 이상과 같은 CNN 아키텍처로 훈련 데이터셋을 사용하여 예측모델을 도출하였다.

[0061] 한편, 본 발명의 실시예에서 적용한 CNN 아키텍처는 다양하게 변형될 수 있고 본 실시예에서 제시한 아키텍처에 제한되지 않는다. 나아가 반드시 CNN만을 이용할 필요도 없으며, 예를 들어 순환 신경망(Recurrent Neural Network, RNN)도 사용될 수 있으며, 특히 시뮬레이션으로 도출된 XRD 패턴의 실측 XRD 패턴과의 차이를 보완하기 위해서는 생성적 적대 신경망(Generative Adversarial Network, GAN) 기술도 적용될 수 있다.

[0063] CNN 예측 모델을 이용한 상 판별 결과

[0064] 이상에서 설명한 훈련 데이터셋을 사용하여 CNN2 및 CNN3 아키텍처를 통한 학습 결과를 통해 도출된 예측 모델을 사용하여, 상 판별(phase identification)이 가능한지를 테스트하였다.

[0065] (1) Dataset\_800k\_org

[0066] 도 10은 'Dataset\_800k\_org'를 사용한 경우 학습 코스트 및 정확도를 나타낸 것이다. 도 9에서 확인되는 것과 같이, CNN2 아키텍처나 CNN3 아키텍처 모두 유효성 검사 결과 정확도가 거의 100%에 도달하였고, CNN2 및 CNN3에 대해 유효성 검증 코스트가 각각 0.007 및 0.0018로 감소하였다.

[0067] 'Dataset\_800k\_org'를 사용하여 훈련한 CNN2 및 CNN3의 테스트를 위하여, 훈련 데이터셋과 겹치지 않도록 100,000개의 시뮬레이션 XRD 패턴을 준비하였고, 또한 3종의 화합물 Li<sub>2</sub>O-SrO-Al<sub>2</sub>O<sub>3</sub>과 SrAl<sub>2</sub>O<sub>4</sub>-SrO-Al<sub>2</sub>O<sub>3</sub>을 다양한 비율로 건식 혼합하여 각각 50개의 혼합물 샘플을 제조하여, 실제 XRD 패턴을 획득하여 상 판별 테스트에 사용하였다.

[0068] 아래 표 1은 상기 2가지 테스트 데이터셋을 사용하여 테스트한 결과를 나타낸 것이다.

표 1

[0069]

CNN 아키텍처		Dataset_80K_org	
		CNN2	CNN3
상 판별 (2 Epochs)	시뮬레이션 XRD 테스트 데이터셋		
	100,000 패턴	99.60%	100%
	실제 XRD 테스트 데이터셋		
	Li <sub>2</sub> O_SrO_Al <sub>2</sub> O <sub>3</sub> (50 패턴)	100%	100%
	SrAl <sub>2</sub> O <sub>4</sub> _SrO_Al <sub>2</sub> O <sub>3</sub> (50 패턴)	97.33%	98.67%

[0071] 상기 표 1에 나타난 것과 같이, 시뮬레이션된 테스트 데이터셋은 CNN2 및 CNN3 아키텍처에 모두 테스트 정확도가 거의 완벽한 수준으로 나타났다. 또한, Li<sub>2</sub>O-SrO-Al<sub>2</sub>O<sub>3</sub> 데이터셋의 테스트 결과도 CNN2 및 CNN3 모두에 대해 정확도 100%로 완벽하게 일치하였다.

[0072] 한편, SrAl<sub>2</sub>O<sub>4</sub>-SrO-Al<sub>2</sub>O<sub>3</sub> 데이터셋의 경우 테스트 정확도가 다소 저하되는 결과를 나타내는데, 이는 혼합원료로 사용된 SrAl<sub>2</sub>O<sub>4</sub>(SAO) 분말에 일정량의 불순물 상이 포함된 것에 기인하는 것으로 확인되었으며, 이는 약간의 잘못 예측된 결과가 CNN 모델의 오류가 아니라는 점을 의미한다.

[0073] 이러한 테스트 결과로부터, 본 발명의 실시예에 따른 방법이, 다양한 물질이 혼합된 다성분계 물질에 포함된 상(phase)의 판별에 있어서, 매우 높은 정확도를 제공할 수 있음을 알 수 있다.

[0075] (2) Dataset\_800k\_rand / Dataset\_180k\_rand에 의한 상 식별

[0076] 데이터셋 'Dataset\_800k\_rand'와 'Dataset\_180k\_rand'는 상기 데이터셋과 달리, 2종 및 3종 혼합물을 설정



함에 있어서, 임의적인 방법을 사용한 것으로, 훈련 데이터셋을 준비할 때 혼합물 데이터를 생성할 때 블랜딩 방법의 차이에 따른 영향을 비교하기 위한 것이다.

[0077] 도 11에 나타난 것과 같이, Dataset\_800k\_org를 사용하여 CNN을 수행한 경우와 실질적으로 동일한 코스트 및 정확도가 달성되었다.

[0078] 아래 표 2는 상기한 테스트 데이터셋을 이용하여 테스트한 결과를 나타낸 것이다.

표 2

[0079]

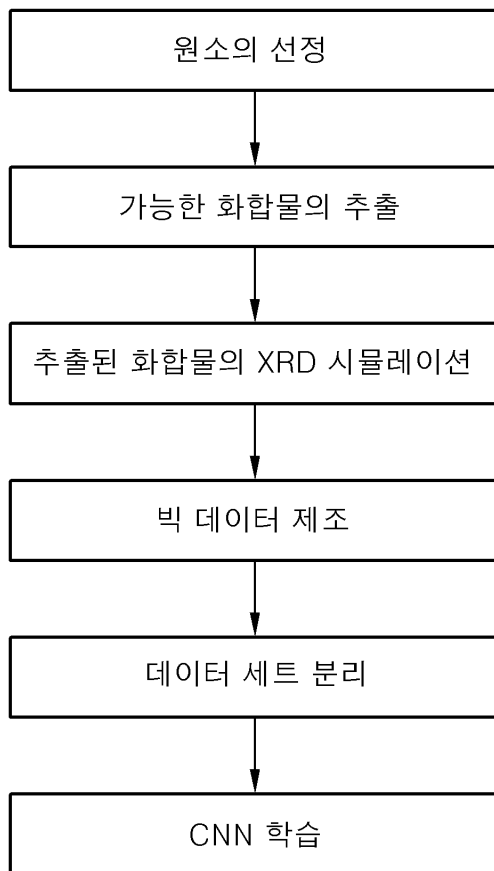
CNN 아키텍처		Dataset_80K_rand	Dataset_18K_rand
		CNN3	CNN3
상 관별 (2 Epochs)	시뮬레이션 XRD 테스트 데이터셋		
	100,000 패턴	100%	(23,000 패턴) 99.76%
	실제 XRD 테스트 데이터셋		
	Li <sub>2</sub> O-SrO-Al <sub>2</sub> O <sub>3</sub> (50 패턴)	100%	98.67%
	SrAl <sub>2</sub> O <sub>4</sub> -SrO-Al <sub>2</sub> O <sub>3</sub> (50 패턴)	98%	97.33%

[0081] 표 2와 같이, 시뮬레이션된 테스트 데이터셋과 실제 Li<sub>2</sub>O-SrO-Al<sub>2</sub>O<sub>3</sub> 테스트 데이터셋은 'Dataset\_800k\_rand'로 훈련할 때 거의 100% 정확도를 나타내었다. 한편, 'Dataset\_180k\_rand'로 훈련하면 시뮬레이션된 테스트 데이터셋과 실제 테스트 데이터셋에서 모두 테스트 정확도가 다소 저하되는 경향을 나타내는데 이는 훈련 데이터셋의 크기가 상대적으로 작은 것에 기인하는 것으로 보인다.

[0082] 이상과 같은 결론부터, 훈련 데이터셋을 준비할 때 혼합물의 블랜딩 방법의 차이에 따른 영향은 거의 없으며, 데이터셋의 크기에 따라 정확도에 약간의 차이가 있음을 알 수 있다.

도면

도면1



도면2

type	개수	group 숫자
Corundum-Al2O3 (Al-O)	74	0
Spinel-defect (Al-O)	9	1
AlFeO3 (Al-O)	3	2
Ga2O3 (Al-O)	2	3
gamma (Al-O)	2	4
Ag3bO3 (Al-O)	1	5
Cr2Mg4O4 (Al-O)	1	6
Fluorite-CaF2 (Al-O)	1	7
NaCl (Al-O)	1	8
Rh2S3 (Al-O)	1	9
무소속 (Al-O)(23660)	1	10
무소속 (Al-O)(40200)	1	11
Fluorite-CaF2, Li2O (Li-O)	14	12
Li2O2 (Li-O)	1	13
무소속 (Li-O)(24143)	1	14
무소속 (Li-O)(108886)	1	15
NaCl (Sr-O)	7	16
CaC2 (Sr-O)	6	17
LiAlO2-gamma (Li-Al-O)	7	18
Li5AlO4 (Li-Al-O)	2	19
spinel-LiFe5O8 (Li-Al-O)	2	20
Spinel-defect (Li-Al-O)	1	21
Delafossite-NaCrS2 (Li-Al-O)	1	22
LiFeO2-alpha (Li-Al-O)	1	23
Li5GaO4-alpha (Li-Al-O)	1	24
SrAl2O4 (Sr-Al-O)	5	25
M-type-ferrite#CaAl12O19 (Sr-Al-O)	4	26
Sr4Al14O25 (Sr-Al-O)	3	27
BaMg5O4 (Sr-Al-O)	3	28
BaAl2O4 (Sr-Al-O)	2	29
Ca9Al6O18 (Sr-Al-O)	2	30
CaAl4O7 (Sr-Al-O)	2	31
무소속 (Sr-Al-O) (57177,200671)	2	32
CaAl3O7 (Sr-Al-O)	1	33
Sr5Al11O17 (Sr-Al-O)	1	34
Sr10Ga6O19 (Sr-Al-O)	1	35
무소속 (Sr-Al-O)	1	36
Sr2LiAlO4	1	37

도면3

### ICSD 결정구조 Data


```
#(C) 2018 by FIZ Karlsruhe - Leibniz Institute :
rights reserved,
data_173206-ICSD
_database_code_ICSD 173206
_audit_creation_date 2008-08-01
_chemical_name_systematic 'Lithium oxide'
_chemical_formula_structural 'Li2 O'
_chemical_formula_sum 'Li2 O1'
_chemical_name_structure_type Fluorite-CaF2
_exptl_crystal_density_diffrn 2.02
_publ_section_title
'Walton, A.'
'Edwards, P.P.'
_cell_length_a 4.6128(4)
_cell_length_b 4.6128(4)
_cell_length_c 4.6128(4)
_cell_angle_alpha 90.
_cell_angle_beta 90.
_cell_angle_gamma 90.
_cell_volume 98.15
_cell_formula_units_Z 4
_symmetry_space_group_name_H-M 'F m -3 m'
```

### XRD 구성 Parameter (Random choice)

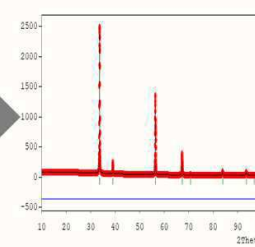
1. Background Parameter (6개)

ε <sub>0</sub>	ε <sub>1</sub>	ε <sub>2</sub>	ε <sub>3</sub>	ε <sub>4</sub>	ε <sub>5</sub>	
Coefficients	5.12	-48.857	34.387	21.309	-48.358	79.889
2. Peak Parameter (5개)

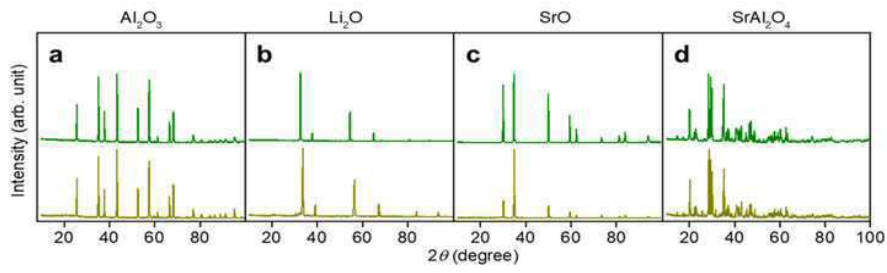
PWHM Parameters			
U	V	W	
Coefficients	0.412229	-0.390814	0.111677

Shape Parameters		
Eta_0	X	
Coefficients	0.483130	0.000000
3. Random noise


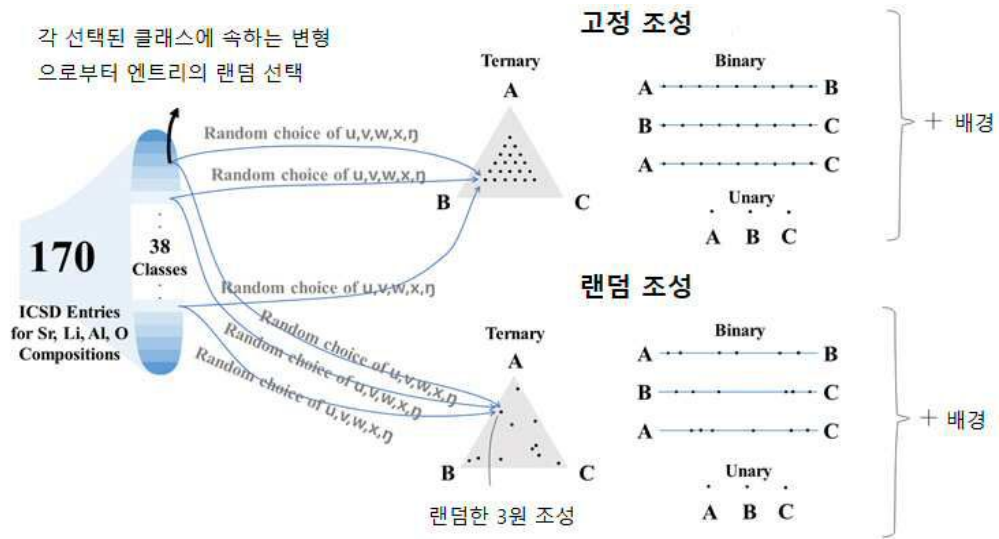
### 계산된 XRD Pattern



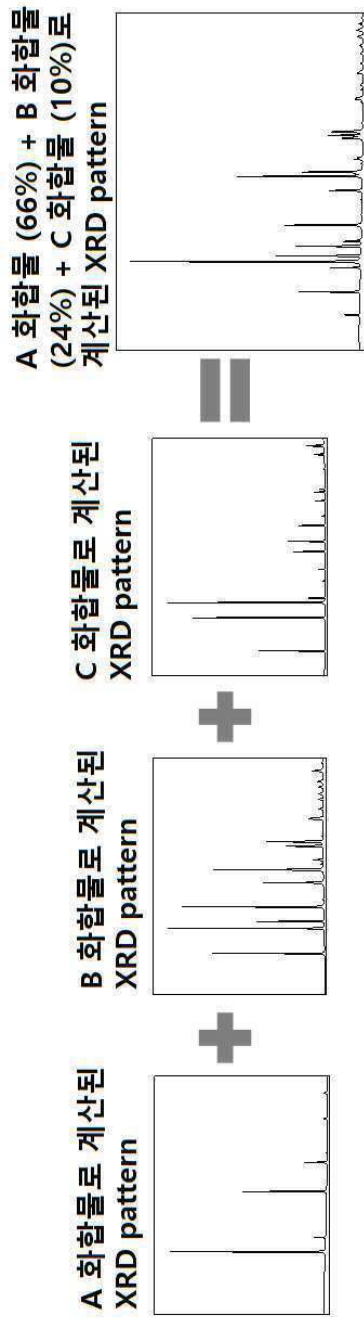
도면4



도면5



도면6

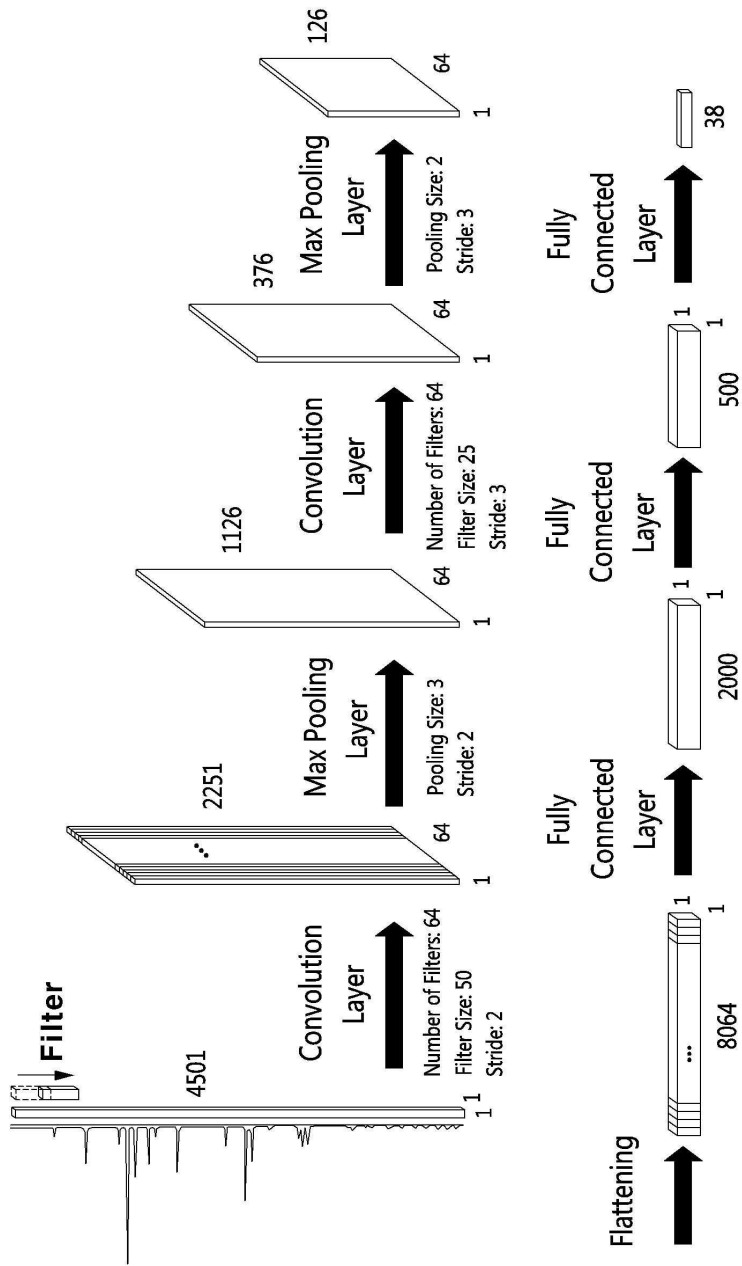


도면7

범위		데이터갯수(범위)	ternary갯수	ternary반복횟수	합계
86이상	97이하	49	21	32	32928
76이상	85이하	617	21	16	207312
10이상	75이하	2205	21	8	370440
3이상	9이하	5565	21	1	116865
					727545
범위		데이터갯수(범위)	binary갯수	binary반복횟수	합계
binary는 범위 없음		703	9	11	69597
범위		데이터갯수(범위)	single 갯수	single 반복횟수	합계
single는 범위 없음		38	1	100	3800
					총 합계
					800942

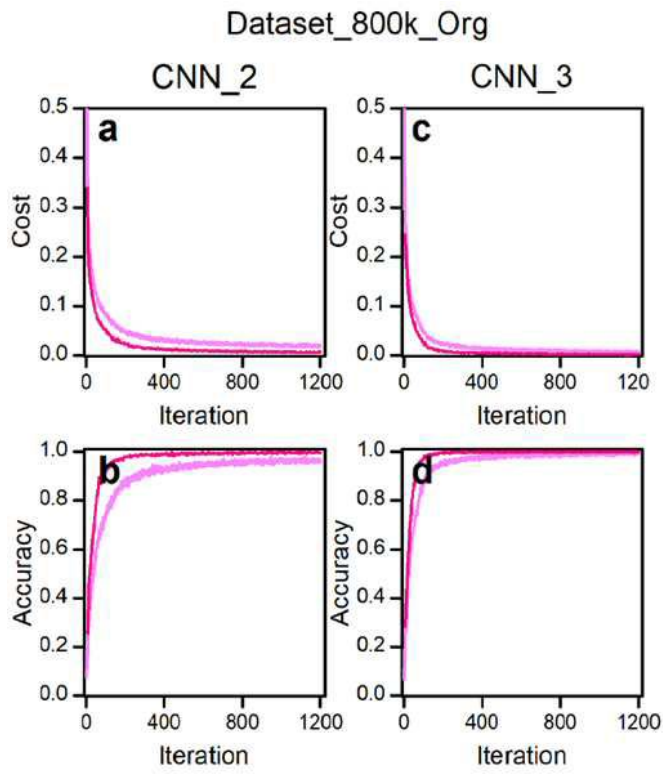


도면8





도면10



도면11

