

Discrimination of handwriting marks written by pencil and mechanical pencil using SEM-EDS and machine learning

Choong-Sik Lee[★], Phil-Sang Ahn, Seung-Chan Roh, and Hwan-Oh Kim

Department of Image, Document and Fingerprint, Defense Forensic Service, Ministry of National Defense,
Itawon-ro 22, Seoul 04383, Korea

(Received March 10, 2025; Revised April 16, 2025; Accepted April 17, 2025)

SEM-EDS와 머신러닝을 이용한 연필과 기계식펜으로 기재된 필적의 판별

이충식[★] · 안필상 · 노승찬 · 김환오

영상문서지문과, 국방부조사본부 과학수사연구소
(2025. 3. 10. 접수, 2025. 4. 16. 수정, 2025. 4. 17. 승인)

Abstract: In forensic document examination, it is often necessary to determine whether a handwriting mark has been forged or altered by identifying whether it was written with a pencil or a mechanical pencil. Pencil and mechanical pencil leads are composed of graphite or activated carbon mixed with clay or resin, depending on the type of pencil. Particularly, the difference in the clay can be a good criteria for distinguishing between pencil leads and mechanical pencil leads. In this study, the surface of handwriting marks written with 14 types of pencils and 15 types of mechanical pencil leads was analysed using SEM-EDS. Clay is composed of minerals such as talc, pyrophyllite, kaolinite, smectite and illite, etc., with the primary elements being silicon (Si), aluminium (Al), magnesium (Mg), and oxygen (O). In pencil handwriting mark, the presence of clay resulted in an increasing trend of $Mg < Al < Si$ elemental content, while in mechanical pencil handwriting mark, the weight % of Mg, Al, and Si were significantly lower, with Al being particularly minimal or absent. Additionally, carbon (C) from the pulp and calcium (Ca) from calcium carbonate filler were considered as background components due to the influence of the paper. Linear discriminant analysis (LDA) and support vector machine (SVM) were applied to the ratios of Al/Mg, Al/Si, Si/Ca, and Ca/C to classify the handwriting mark. This study provides a crucial method for distinguishing between pencil and mechanical pencil in handwriting mark, contributing to the verification of handwriting addition, forgery or alteration in the forensic document written by pencil-type instrument.

요약: 법과학적 문서감정에서 연필 종류의 필적이 위조 또는 변조되었는지 확인하기 위해 연필(pencil) 또는 기계식펜(mechanical pencil) 중 어떤 필기구로 기재되었는지 결정할 경우가 있다. 연필심과 기계식

[★] Corresponding author

Phone : +82-(0)2-748-1920 Fax : +82-(0)2-748-1888

E-mail : cshj35@gmail.com

This is an open access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

펜심은 흑연 또는 활성탄을 점토(clay) 또는 수지(resin)와 혼합하여 제조된다. 특히 점토의 차이는 연필심과 기계식펜심을 판별하는 좋은 기준이 될 수 있다. 본 연구는 14종 연필과 15종 기계식펜심으로 기재된 필적 표면을 SEM-EDS로 분석하였다. 점토는 탈크(talc), 파이로필라이트(pyrophyllite), 카올리나이트(kaolinite), 스멕타이트(smectite), 그리고 일라이트(illite) 등의 광물들로 구성되며, 주요 원소성분은 실리콘(Si), 알루미늄(Al), 마그네슘(Mg), 산소(O) 등이다. 연필 필적에서 점토의 존재는 $Mg < Al < Si$ 순서로 원소 함량을 증가시키는 경향을 보였으나, 기계식펜 필적에서 Mg, Al, Si의 함량이 적고 특히 Al의 함량이 매우 적거나 없었다. 또한 종이의 영향을 고려하여 종이 필프의 탄소(C)와 충전제(filler)로 사용되는 탄산칼슘의 칼슘(Ca)을 바탕성분으로 포함하였다. 선형판별분석(LDA)과 서포트벡터머신(SVM)을 적용하여 Al/Mg, Al/Si, Si/Ca 및 Ca/C의 상대비율을 기반으로 필적을 분류하는 모델을 확립하였다. 본 연구결과는 연필 또는 기계식펜으로 기재된 필적을 판별하는 중요한 방법을 제공하며, 연필류 필기구로 작성된 법과학 문서 감정에서 필적의 가필, 위조 또는 변조 여부를 검증하는데 기여할 것이다.

Key words: handwriting #1, pencil #2, mechanical pencil #3, LDA #4, SVM #5

1. 서 론

본 연구는 연필 종류의 필기구로 기재된 필적에서 연필과 기계식펜 중 어떤 필기구로 기재되었는지 확인이 가능한 법과학적 분석방법을 검토하였다. 법과학적 문서감정에서 연필 또는 기계식펜으로 작성된 필적을 필획의 형태로 구별하기는 거의 불가능하다. 연필심과 기계식펜심은 구성물질과 제조과정이 매우 상이하다.¹ 현대적인 형태의 연필심은 1795년 Nicolas Jacques Conte가 발명하였다. 흑연(graphite)과 점토(clay)를 혼합하고 열처리하여 굳혀서 강도를 유지하였다.^{2,3} 연필심의 주요 성분은 흑연과 점토이고 흑연 대 점토 비율이 연필의 경도(hardness)를 결정한다. 점토 비율이 높을수록 단단해지고 흑연 비율이 높을수록 물러진다. 점토는 물과 섞으면 가소성을 보이고 천연의 지표상에서 산출되는 미세한 광물을 의미한다. 점토 광물은 결정구조와 화학조성에 따라 talc($Mg_3Si_4O_{10}(OH)_2$), pyrophyllite($Al_2Si_4O_{10}(OH)_2$), kaolinite($Al_2Si_2O_7(OH)_2$), smectite($Al_2Si_2O_7(OH)_4$), illite($K_2Al_4(Al_2Si_6)O_{20}(OH)_4$) 등으로 구성되어 있고, 실리콘(Si), 알루미늄(Al)과 마그네슘(Mg)이 주요 원소성분이다.⁴ 연필(pencil)은 점토와 흑연, 물 등을 섞어서 시간을 들여 잘 갠 후에 건조시켜 단단하게 구워서 연필심을 제조한다. 이것을 나무틀에 넣어 연필을 완성한다. 기계식펜(mechanical pencil)은 19세기 후반 금속 나사 구조를 이용해 연필심을 고정하는 방식으로 개발되었고, 샤프펜이라 불리우는 현대식 기계식펜은 1915년 “에버샤프 펜슬(Eversharp Pencil)” 이름으로 개발되었다.⁵ 기계식펜심은 흑연과 고분자 수지를 섞어서 굵고 왁스(wax)를 스며들게

해서 매끄럽게 만든다. 이 과정을 통해 기계식펜심은 가늘면서 쉽게 부러지지 않고 필기감이 부드러운 심으로 만들어진다. 또한 첨가제인 왁스는 펜의 이동을 부드럽게 하고 종이 위에 연필흔적(pencil mark)을 유지하게 한다. 연필 필적에서 연필심 종류를 구별하기 위한 연구들이 매우 적게 보고되었다. Cantu 등은 연필심 자체에서 왁스 성분을 박층크로마토그래피(thin layer chromatography, TLC)로 구별할 수 있었다. 그러나 종이에 쓰여진 필적에서 연필심과 왁스를 충분히 추출할 수 없었고, 종이 성분이 TLC 식별을 심각하게 방해하였다.⁶ Zoro 등은 충전제(filler)가 거의 또는 전혀 없는 종이에 기재한 연필 필적을 굽어내고 왁스 성분을 추출하여 질량분석법으로 분석하였다. 17종 연필심이 4개 그룹으로 분류되었으나 충전제가 포함된 일반적인 종이에서는 방해성분에 의해 분석에 문제가 있었다.⁷ Chen 등은 경도가 다른 4종류(6B, 4B, 2H, 4H)의 연필들로 종이에 적고 레이저유도플라즈마 분석법(laser induced breakdown spectroscopy, LIBS)으로 분석하였다. 탄소(C), 마그네슘(Mg), 알루미늄(Al), 칼슘(Ca), 실리콘(Si), 철(Fe) 등의 원소성분을 분석하고 BP-신경망과 PCA-KNN 모델을 상호 보완적으로 적용한 연구가 보고되었다.⁸ 또한 Denman 등에 의하면 연필심 자체를 ICP-MS와 ToF-SIMS를 이용하여 5종 연필 제품의 차이를 PCA 통계분석으로 연구하였다.⁹ 그러나 종이 위에 연필과 기계식펜으로 기재된 필적을 구별하는 연구들은 보고되지 않았다. 따라서, 이 연구는 종이 위에 기재된 필적 표면을 주사전자현미경-에너지 분산분광법(scanning electron microscope-energy dispersive spectroscopy, SEM-EDS)으로 분석하고 머신러닝

(machine learning)으로 선형판별분석(linear discriminant analysis, LDA)^{10,11}과 서포트벡터머신(support vector machine, SVM)을 적용하여, 점토 성분의 유무에 따른 원소성분 차이를 특성으로 연필과 기계식펜으로 기재된 필적을 판별하는 법과학적 분류법을 확립하였다.

2. 재료 및 방법

2.1. 시료 및 기구

시료는 한국에서 판매되는 14종 연필과 15종 기계식펜심을 구매하여 사용하였다(Table 1). 연필과 기계식펜심 제조사는 한국의 문화(Munhwa), 동아(DONG-A)와 제노(Xeno), 스위스의 에델바이스(EDELWEISS), 미국의 프리즈마칼라(PRISMACOLOR), 일본의 미쓰비시(MITSUBISHI), 펜텔(pentel)과 톰보우(Tombow), 독일의 스테들러(STEADTLER) 등이다. 필적을 기재한 종이는 “University of Yensei” 노트(PENCIA, Korea)를 사용하였다. SEM 시료는 1 mm 지름의 Harris MICRO-PUNCH®(Ted Pella Inc. USA)로 필적으로부터 5개씩 채취하고 SEM 전용 stub에 부착하였다. Stub에 채취된 분석시료들은 SEM-EDS(Evo Ma15, Zeiss, Germany/XFlash 6130, Bruker, Germany)로 분석하였다. 분석조건은 가속전압 25 kV, 작동거리 15 mm, SEM 확대비율 130배 내외, BSD 검출기 Medium으로 설정하였다. 점토 광물들의 화학식과 종이의 구성성분을 고려하여 주요 원소성분의 정규화된 무게 % (normalized weight %)를 분석하였다.

2.2. 실험 방법

종이 위에 연필 또는 기계식펜으로 필적을 기재한 후, 마이크로펀치로 채취된 분석시료를 SEM stub에 부착한 후 SEM-EDS로 분석하였다. 이때 연필과 기계식펜의 필적이 남은 부분은 필압과 흑연의 영향으로 대전현상이 감소하였고, 그 결과로 필적이 남겨진 위치를 SEM에서 선택할 수 있었다. 먼저, 필적이 기재된 종이재질을 바탕성분으로 분석하고 연필심과 기계식펜심 자체를 분석하였다. 14종 연필과 15종 기계식펜심으로 종이 위에 필적을 기재하고 필적 부분을 마이크로펀치(지름 1mm)로 각각 5개씩 반복하여 분석시료를 채취하였다. 채취된 분석시료를 SEM stub에 부착하고 SEM-EDS로 분석하였다. 점도광물 등의 화학조성과 종이의 구성성분을 고려하여 주요 원소성분의 무게 %를 확인하였다. 연필과 기계식펜의 필적을 판별하기 위한 머신러닝(machine learning)으로 Matlab 소프트웨어(Mathworks, USA)의 선형판별분석(LDA)과 서포트벡터머신(SVM)으로 분류모델을 생성하고 최적화하였다. LDA는 모든 클래스의 샘플데이터 공분산 행렬이 동일하다는 가정에 근거하고 직교변환을 사용하는 지도학습 데이터분석 방법이다. LDA는 측정 데이터들이 특정 클래스에 할당되는 확률을 측정한다.^{10,13} 이 연구에서 LDA는 4개의 주요 변수들인 Si/Ca, Al/Si, Al/Mg와 Ca/C 이 더욱 차이 나도록 선형 결합을 수행하여 정준상관변수(canonical variable)를 얻었다. LDA는 연필 필적 또는 기계식펜 필적 클래스 간 분산의 거리를 최대화하고 클래스 내 데이터의

Table 1. Commercial pencils and mechanical pencil leads used for handwriting mark

Abb.	Pencil Manufacturer	Brand and Product Name	Abb.	Mechanical Pencil Lead Manufacturer	Brand and Product Name
P1	Munhwa	DEOJON hi-mic (HB)	S1	DONG-A	XQ CERAMICS 301 (H)
P2	Munhwa	DEOJON OMR (2B)	S2	DONG-A	XQ CERAMICS 301 (B)
P3	Munhwa	DEOJON hi-mic (4B)	S3	DONG-A	XQ GOLD 1000 (HB)
P4	Munhwa	DEOJON hi-mic (2B)	S4	DONG-A	XQ GOLD 502 (HB)
P5	DONG-A	Fable (4B)	S5	Pentel	Ain STEIN (HB)
P6	DONG-A	Fable (2B)	S6	Pentel	Ain STEIN (B)
P7	EDELWEISS	CARANDACHE (2B)	S7	STAEDTLER	Mars micro carbon (2H)
P8	EDELWEISS	CARANDACHE (HB)	S8	STAEDTLER	Mars micro carbon (H)
P9	Tombow	Hi-precision DRAFTING (2B)	S9	STAEDTLER	Mars micro carbon (HB)
P10	PRISMACOLOR	VERITHIN 747 Bk	S10	STAEDTLER	Mars micro carbon (B)
P11	MITSHUBISHI	Hi-uni (4B)	S11	MITSHUBISHI	Uni (HB)
P12	STAEDTLER	Yellow pencil 134 (B)	S12	DONG-A	Newen XQ SUPREME (HB)
P13	STAEDTLER	Noris ergosoft Nr.153 (2B)	S13	Tombow	MONO graph Tombow (B)
P14	None	None	S14	Tombow	MONO graph Tombow (HB)
			S15	Xeno	polymer Lea (HB)

분산을 최소화하기 위해 4개 변수들을 최적 결합으로 변환되었다.¹³ LDA 판별기준(criterion)을 위해 후진 단계적선택법(backward stepwise)을 적용하였다. LDA 모델에서 최적의 고유값(eigenvalue)과 분산(variance)을 얻었다. 또 다른 방법으로 SVM은 클래스 간 최대 여백(margin)을 갖는 초평면을 탐색하여 분류를 수행하는 모델이다.¹⁴ SVM에서 Si/Ca, Al/Si, Al/Mg를 변수로 선형 및 가우시안 커널(gaussian 또는 radial basic function) SVM을 이용하여 하이퍼파라미터(hyperparameter)인 비용(cost, C)과 감마(gamma)를 최적화하였다. C는 마진과 오차 허용간의 균형을 조정하여 데이터의 오분류 비용을 제어하고 gamma는 비선형 분류인 가우시안 커널함수 등의 스케일을 제어한다.^{15,16}

3. 결과 및 토의

3.1. SEM-EDS 분석

먼저 종이와 연필심의 원소성분을 분석하였다. 종이에서 C, O, Mg, Al, Si, Ca 등 성분이 주로 검출되었다. C와 O의 근원은 주로 펄프의 주성분인 셀룰로오스(cellulose)와 헤미셀룰로오스(hemi-cellulose)이다.^{17,18} 칼슘(Ca)은 종이의 충전제인 탄산칼슘(CaCO_3)의 조성 성분이고 제지공정 역사에서 산성 초지가 중성 및 알칼리성 초지로 전환되면서 사용량이 증가되어 왔다.¹⁹ 실험에 사용된 종이의 원소성분 특징은 Mg와 Si의 함량이 작으며 상호 상대비가 유사하였다. 종이를 SEM 이미지로 관찰할 때 대전현상이 나타나지만, 연필 필적과 기계식펜 필적에서는 필압과 흑연의 영향으로 대전현상이 감소하였다. 따라서 대전현상이 없는 부분을 필적 부분으로 확인할 수 있었고 BSD 검출기의 SEM 이미지에서 필적 부분과 종이 부분을 구분하기 어려운 문제를 해결하였다.

연필심의 원소성분은 대체로 C, O, Na, Mg, Al, Si, S, K, Ca, Ti, Fe 등이 검출되었다. 점토광물은 결정구조와 결정화학적 차이에 따라 talc - $\text{Mg}_3\text{Si}_4\text{O}_{10}(\text{OH})_2$, pyrophyllite - $\text{Al}_2\text{Si}_4\text{O}_{10}(\text{OH})_2$, kaolinite - $\text{Al}_2\text{Si}_2\text{O}_7(\text{OH})_2$, smectite - $\text{Al}_2\text{Si}_2\text{O}_7(\text{OH})_2$, illite - $\text{K}_2\text{Al}_4(\text{Al}_2\text{Si}_6)\text{O}_{20}(\text{OH})_4$ 등으로 구성되고 화학식에 따라 Si, Al, Mg, O가 주요 원소성분으로 확인되었다. 종이 위 연필 필적은 O, Mg, Al, Si, Ca의 상대적 함량이 증가하고 종이의 연필심이 혼합된 원소성분 특성을 나타내었다. 특히 C, Mg, Al, Si 성분 함량이 높았고 이는 흑연과 점토성분의 주요 구성성분이다. 또한 종이 충전제인 탄산칼슘의 영향으로 Ca의 상대적 함량이 크게 증가하였다. 연필

심이 종이 표면을 덮기 때문에, 연필 필적에서 종이의 주요 성분인 O의 세기가 종이에 비해 작아지고, 연필심에 비해 커지는 경향을 나타내었다. 연필 필적에서 가장 큰 특징은 함량세기가 $\text{Mg} < \text{Al} < \text{Si}$ 순서로 커지는 경향을 나타내었다. 따라서 머신러닝 분석을 위해 점토 성분의 영향을 반영하는 변수로 Al/Si, Al/Mg를 선정하였다. 추가적으로 연필 종류 심(lead)이 종이 표면으로 전이될 때, 연필흔적 전체에 불균질하게 전이되는 특성을 가지는 것도 이해해야 한다.

기계식펜심은 주요 원소성분으로 C와 Si, 기계식펜심 종류에 따라 미량성분으로 P, Na, S, Ca, Cl, Fe 등 다양한 원소성분들이 검출된다. 종이 위 기계식펜

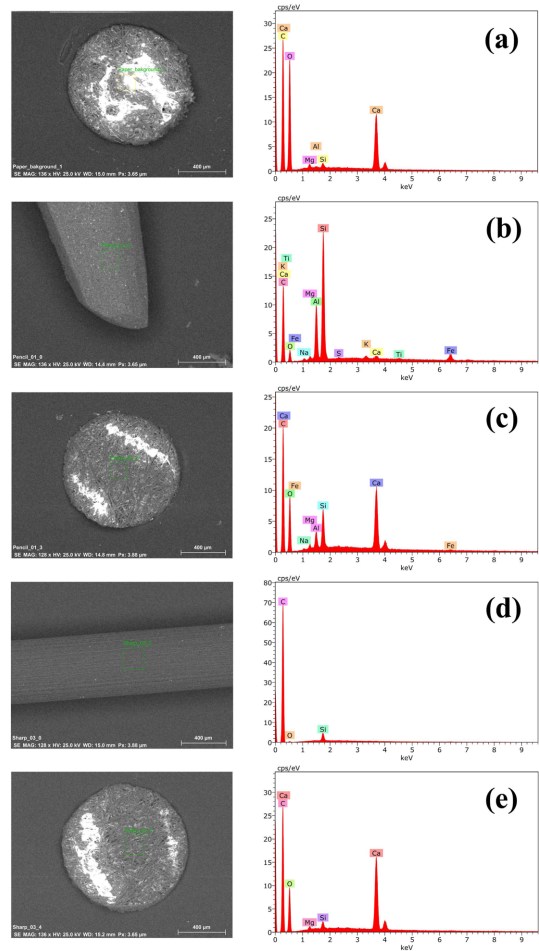
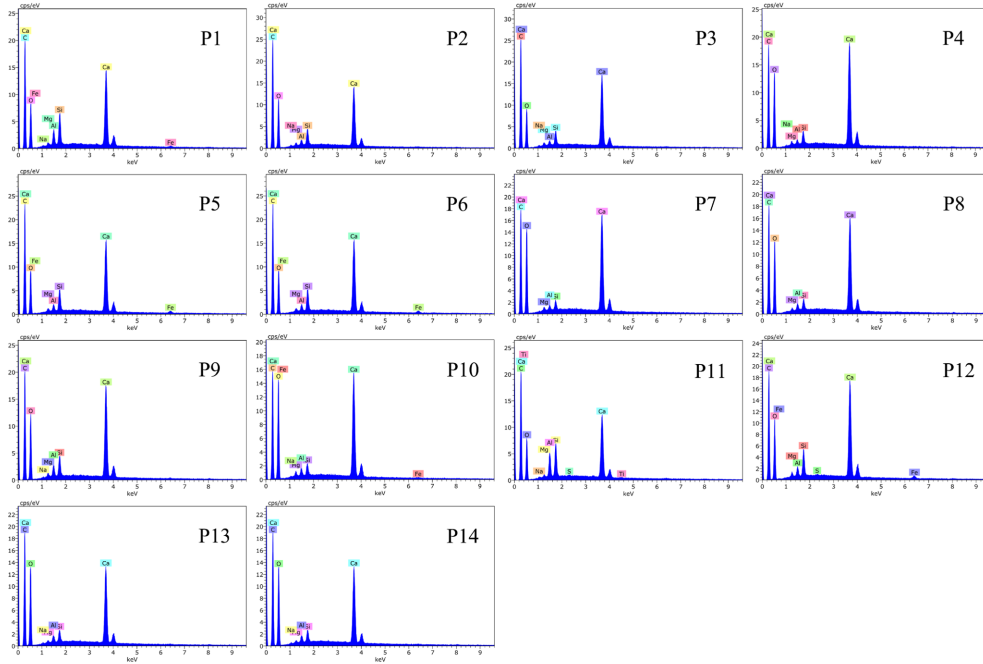


Fig. 1. SEM image and EDS spectra: (a) paper, (b) pencil lead (Munhwa DEOJON hi-mic HB), (c) pencil mark on paper, (d) mechanical pencil lead (DONG-A XQ GOLD 1000 HB), (e) mechanical pencil mark on paper.

필적은 O, Si, Ca의 상대적 함량이 증가하고 종이와 기계식펜심이 혼합된 원소성분 특성을 나타내었다.

특히 Si의 함량이 작게 증가하였다. 미량 성분 P, S, Cl 등은 종이의 바탕신호에 묻혀 기계식펜 필적에서

(a) Pencil handwriting



(b) Mechanical pencil handwriting

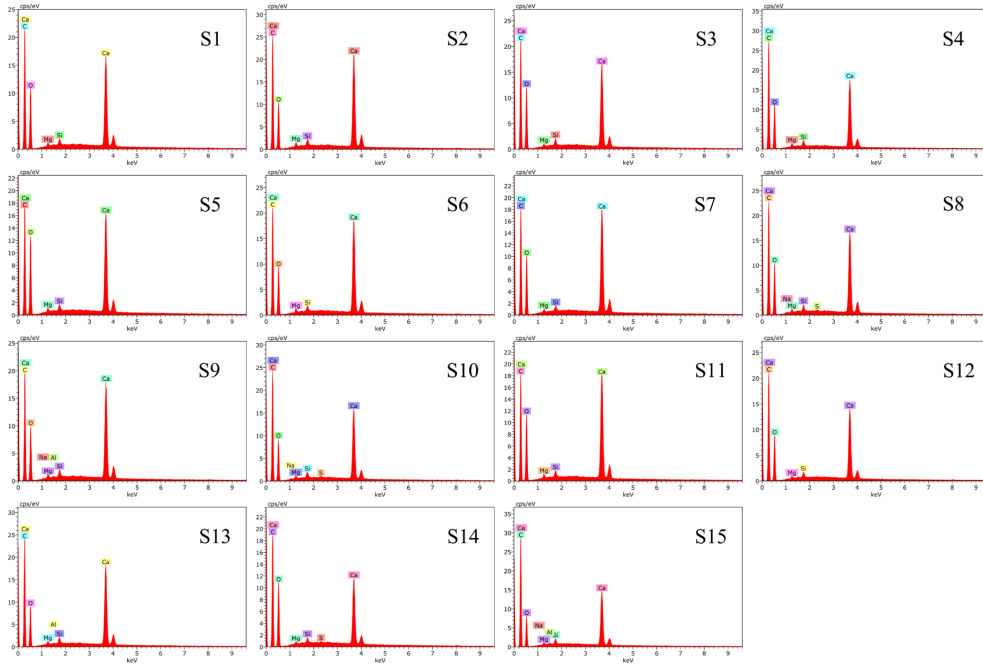


Fig. 2. SEM-EDS spectra: (a) the handwriting mark of pencil P1~P14 (blue), (b) the handwriting mark of mechanical pencil S1~S15 (red).

검출되지 않았다. 또한 기계식펜 필적은 종이의 영향으로 O와 Ca이 증가하였다. 기계식펜 필적에서 가장 큰 특징은 Al이 거의 없기 때문에 함량세기가 $Al < Mg < Si$ 순서로 커지는 경향을 보이고 연필심 필적과 차이를 나타내었다. 특히 Al 함량이 Mg와 Si에 비해 상대적으로 작거나 없기 때문에 위에서 선정한 변수 Al/Si, Al/Mg는 기계식펜 필적에서도 큰 의미를 가진다. 종이, 연필심, 연필 필적, 기계식펜심과 기계식펜 필적에 대한 SEM 이미지와 EDS 스펙트럼에 대한 예시를 Fig. 1에 나타내었다.

모든 14종 연필 필적과 15종 기계식펜 필적에 대한 EDS 스펙트라를 Fig. 2에 나타내었다. 연필 필적과 기계식펜 필적의 주요 차이점은 다음과 같은 원소성분 상대비로 구별할 수 있었다. 먼저 필적에서 점토성분의 존재 여부를 확인할 수 있는 Mg, Al, Si의 상대적 비율의 차이, 즉 Al/Si와 Al/Mg를 변수로 결정하였다. 또한 연필 필적에서 점토성분에 의한 Si 성분이 상당히 증가하지만, 기계식펜에서는 실리콘 오일에 의한 Si 성분이 작은 증가만을 나타낸다. 따라서 종이의 바탕 성분인 Ca과 점토 또는 실리콘 오일의 Si 성분에 의한 영향을 반영하여 Si/Ca를 추가적인 변수로 고려하였다. 14종 연필과 15종 샤프심의 Al/Si, Al/Mg, Si/Ca의 3차원 plot을 Fig. 3에 나타내었다. 직관적으로 기계식펜 필적에서 Si/Ca, Al/Mg과 Al/Si은 작은 값을 가진다. 이는 기계식펜 필적에는 Al이 거의 없고, 종이의 바탕 성분인 Ca에 비해 Si가 상대적으로 작다는 것을 의미한다. 반대로 연필 필적에서 Al 성분이 분명히 존재하고, Mg에 비해 Al이 큰 값을 가진다. 또한 연필 필적

에서 Si/Ca는 작은 값에서 큰 값까지 넓게 분포한다. 마지막으로 종이의 C와 Ca, 흑연의 C에 대한 영향을 고려하기 위해 Ca/C를 추가적인 변수로 선정하였다. 따라서 Al/Si, Al/Mg, Si/Ca 또는 Ca/C를 변수로 하여 LDA와 SVM의 분류 모델을 구성하였다.

3.2. 선형판별분석(LDA)

LDA는 법과학에서 잉크, 종이, 연소잔해물, 토양, 유리 등 거의 모든 성분분석 분야에 응용되고 있다.^{11,20-23} 이 연구에서 LDA는 연필과 기계식펜 클래스간 분산을 최대화하고 각 클래스내 분산을 최소화하였다. LDA는 주요 원소성분의 상대비를 4개 변수(Al/Si, Al/Mg, Si/Ca, Ca/C)로 설정하였다. 위에서 설명한 것처럼 Al/Si, Al/Mg는 점토 성분의 영향을 반영한다. 점토가 미포함되었으나 간혹 실리콘 오일의 Si 성분이 포함된 기계식펜 필적에서 Si/Ca는 연필 필적보다 작게 나타나는 현상을 반영하였다. 또한 Ca/C는 종이의 영향을 추가로 반영하기 위한 변수이다. 14종 연필 필적, 15종 기계식펜 필적을 각각 5번씩 분석하여 총 145 데이터를 획득하였다. 전체 데이터를 훈련데이터와 시험데이터로 분할(holdout=0.2)하여 모델을 생성하고 5겹 교차 검증(5-fold cross-validation)을 수행하였다. 정확도, F1-score와 교차검증 오차율(error rate)로 모델 성능을 평가하였다. LDA 결과, 3개의 정준변수를 얻었고 이 정준변수들의 3D plot을 Fig. 4에 나타내었다. 또한 판별함수들의 고유값(eigenvalue)과 개별 % 분산(individual % variance)을 Table 2에 나타내었다. 여기서

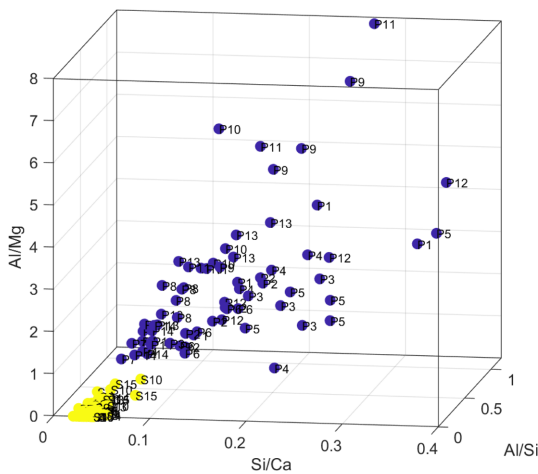


Fig. 3. 3D plot of pencil handwriting mark (blue) and mechanical pencil mark (yellow).

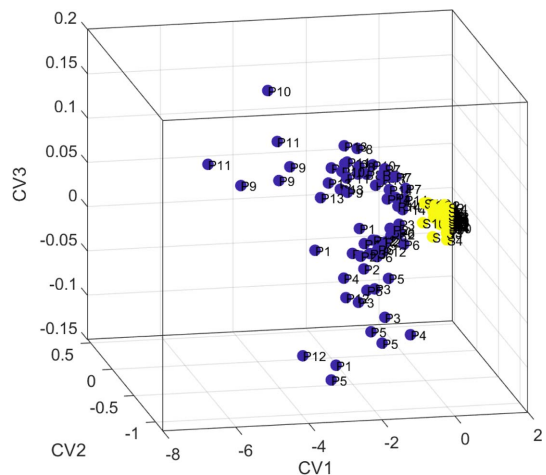


Fig. 4. 3D plot of canonical variables in LDA. Pencil handwriting mark (blue) and mechanical pencil mark (yellow).

Table 2. Eigenvalues and variance of discriminant functions

Discriminant function	Eigenvalue	Individual % variance
DF1	5.24E+3	100
DF2	9.61E-14	0.0

고유값은 클래스에 대한 판별 점수(discriminant score)에서 제곱합간(between sums of squares, BSS)과 제곱합내(within sums of squares, WSS)의 비율이다. 상대적으로 큰 고유값은 좋은 판별함수를 나타낸다. 판별함수 1(식 (1))은 고유값이 5,241로 총 분산의 거의 100%를 설명하고 있다. 따라서 1개의 판별함수만으로도 연필 필적과 기계식펜 필적을 판별할 수 있는 LDA 모델이 생성되었다. 판별함수 1과 2는 다음과 같다.

$$DF1 = -0.048 \text{ Si/Ca} - 0.280 \text{ Al/Si} - 0.959 \text{ Al/Mg} + 0.003 \text{ Ca/C} \quad (1)$$

$$DF2 = -0.004 \text{ Si/Ca} + 0.960 \text{ Al/Si} - 0.280 \text{ Al/Mg} + 0.003 \text{ Ca/C} \quad (2)$$

고유값이 5,241로 가장 큰 판별함수인 DF1에서 계수들을 고려하면, Al/Mg와 Al/Si가 (-) 방향으로 가장 큰 영향을 주고 있다. 이는 위에서 설명한 것처럼 기계식펜 필적에서 Al이 거의 없고, 연필 필적은 Al 성분이 분명히 존재하고 Al이 Mg에 비해 매우 크다는 것을 명확하게 설명하고 있다. 이 LDA 모델은 매우 높은 정확도(accuracy = 1)와 분류성능(F1 score = 1) 그리고 좋은 교차검증 오차율(error rate = 0%)로 14종 연필 필적과 15종 기계식펜 필적을 판별할 수 있었다.

3.3. 서포트벡터머신(SVM)

SVM은 클래스 간 최대 여백을 갖는 초평면을 탐색하여 클래스 분류를 수행하는 모델이다. SVM도 LDA 처럼 화학적 성분분석에 의한 클래스를 분류할^{24,25} 뿐만 아니라 디지털포렌식, 필자인식, 얼굴인식, 사이버포렌식, 사망원인 같은 다양한 법과학분야에서 응용되고 있다.²⁶⁻²⁹ SVM에서 변수 Ca/C의 영향이 매우 작고 초평면을 3차원에 시각화하기 위해 Al/Si, Al/Mg, Si/Ca 만을 변수로 설정하였다. SVM에서도 LDA와 같이 14종 연필 필적, 15종 기계식펜 필적을 각각 5번씩 분석한 145 데이터를 훈련데이터와 시험데이터로 분할(holdout = 0.2)하였다. 선형과 가우시안(또는 RBF) 커널을 적용하여 특징 공간으로 투영 후 초평면을 결정하고 모델을 생성하고 5겹 교차검증(5-fold cross-validation)을 수행하였다. 모델 성능은 정확도, F1-score와 교차검

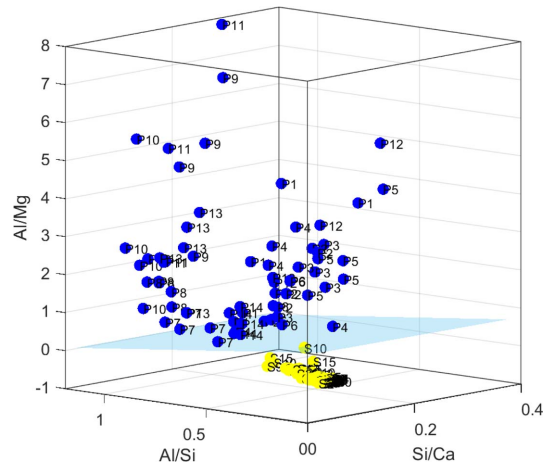


Fig. 4. 3D plot of canonical variables in LDA. Pencil handwriting mark (blue) and mechanical pencil mark (yellow).

Table 3. Diagnostics of C and gamma in SVM

SVM method	C	gamma	Cross-validation error rate (%)
Linear	30.8	N.A.	0
Gaussian (Radial Basic Function)	106.2	1.9	0.7

증 오차율(error rate)로 평가하였다. SVM의 하이퍼 파라미터인 비용(cost, C)과 gamma를 최적화하였다. C와 gamma는 각각 학습 데이터의 오분류 비용과 비선형 분류의 가우시안 커널을 제어한다. C는 정규화 파라미터로 마진과 오차 간의 오류를 제어하고, gamma는 커널에서 데이터 사이의 거리 척도를 조정한다. SVM의 선형 커널에서 C는 30.8로 최적화되었고 선형 SVM에서 연필과 기계식펜을 분류하는 초평면을 Fig. 5에 시각화하였다. 또한 가우시안 커널 SVM에서 C는 106.2, gamma는 1.9로 최적화된 모델을 생성하였다(Table 3). 특히, 큰 C와 gamma는 모델이 데이터에 민감하고 복잡한 결정경계를 생성하여 과적합 가능성이 존재하지만, 이 가우시안 SVM 모델은 매우 높은 정확도(accuracy = 1)와 분류성능(F1-score = 1) 그리고 좋은 교차검증 오차율(error rate = 0.7%)로 14종 연필 필적과 15종 기계식펜 필적을 판별할 수 있었다.

4. 결 론

본 연구는 점토 성분의 존재 여부에 따른 원소성분의

차이를 확인하여 연필 필적과 기계식펜 필적을 판별하는 법과학적 시험방법을 제안하였다. SEM-EDS와 머신러닝을 이용한 판별에서 가장 중요한 요인은 점토광물의 존재 유무였다. 연필 필적은 점토성분의 영향으로 $Mg < Al < Si$ 원소성분의 무게 %가 상대적으로 증가하는 경향이 있고 Al/Si , Al/Mg 의 특징적인 상대비가 확인되었다. 반대로 기계식펜 필적에서 점토성분의 부재로 Mg , Al , Si 의 함량이 적으며, 특히 Al 의 함량이 매우 적거나 없는 것이 특징이었다. 따라서 Al/Si , Al/Mg , Si/Ca , Ca/C 원소성분 상대비를 LDA에 적용하여 판별함수와 분류 모델을 구하였다. 그 결과, 연필 필적과 기계식펜 필적 2개 클래스에 대하여 1개의 판별함수만으로도 전체 분산의 거의 100%를 설명할 수 있었다. 또한 Si/Ca , Al/Si , Al/Mg 원소성분 상대비를 SVM에 적용하여 연필과 기계식펜 필적을 판별할 수 있도록 C 와 γ 를 최적화한 SVM 모델을 생성하였다. 그 결과 SEM-EDS와 머신러닝(LDA와 SVM)을 적용하여 높은 정확도와 분류성능 그리고 좋은 교차 검증 오차율을 가지는 연필과 기계식펜 필적의 분류 모델을 확립하였다. 본 연구는 문서감정에서 연필 또는 기계식펜으로 기재된 필적을 판별하여 필적의 가필, 위조, 또는 변조 여부를 확인하는데 새롭고 중요한 역할을 할 것으로 기대한다.

감사의 글

이 연구는 국방부조사본부 자체 연구사업으로 수행하였다. 제지공정에 대하여 조언을 주신 서울지방국세청의 정창우님께 감사드린다. 이 논문에 대한 유익한 조언과 세심하게 심사해주신 심사위원님께서도 감사드린다.

References

1. W. Yoshiyuki and W. Sadami, 'Amazing Technology of Stationery', KADOKAWA, Tokyo, 2014, Korean translation by UU press, Seoul, 16-35 (2017).
2. F. L. Encke, *J. Chem. Educ.*, **47**(8), 575-576 (1970). <https://doi.org/10.1021/ed047p575>
3. E. H. Voice, *Trans. Newcomen Soc.*, **27**(1), 131-141 (1949). <https://doi.org/10.1179/tns.1949.012>
4. K. Pye, 'Geological and Soil Evidence: Forensic Applications', CRC Press, New York, 2007.
5. H. Petroski, 'THE PENCIL: A History of Design and Circumstance', Chapter 17. Getting the Point, ALFRED A. KNOPE, New York, 1989.
6. S. Cain, A. A. Cantu, R. Brunelle, and A. Lyter, *J. Forensic Sci.*, **23**(4), 643-661 (1978). <https://doi.org/10.1520/JFS10720J>
7. J. A. Zoro and R. N. Totty, *J. Forensic Sci.*, **25**(3), 675-478 (1980), 675-478 (1980). <https://doi.org/10.1520/JFS11273J>
8. Y. J. Chen, U. Z. Liu, B. Y. Han, and W. J. Yu, *Optik*, **259**, 169008 (2022). <https://doi.org/10.1016/j.ijleo.2022.169008>
9. J. A. Denman, I. M. Kempson, W. M. Skinner, and K. P. Kirkbride, *Forensic Sci. Int.*, **175**, 123-129 (2008). <https://doi.org/10.1016/j.forsciint.2007.05.017>
10. A. Kher, M. Mulholland, E. Green, and N. Reedy, *Vib. Spectrosc.*, **40**(2), 270-277 (2006). <https://doi.org/10.1016/j.vibspec.2005.11.002>
11. M. Oravec, A. Beganović, L. Gál, M. Čeppan, and C. W. Huck, *Forensic Sci. Int.*, **299**, 128-134 (2019). <https://doi.org/10.1016/j.forsciint.2019.03.041>
12. L. Qu and Y. Pei, *Processes*, **12**(7), 1382 (2024). <https://doi.org/10.3390/pr12071382>
13. W. P. Gardner, 'Statistical Analysis Methods for chemists', Royal Society of Chemistry, Cambridge, 1997.
14. H. Li, 'Machine Learning Methods', Springer, Wiesbaden, 2024.
15. Y. J. Lee, S. Y. Won, S. B. Park, and H. J. Kim, *Herit. Sci.*, **12**, 373 (2024). <https://doi.org/10.1186/s40494-024-01460-6>
16. S. W. Hwang, T. K. Lee, H. B. Kim, H. W. Chung, J. G. Choi, and H. M. Yeo, *Holzforchung*, **76**(1), 1-13 (2022). <https://doi.org/10.1515/hf-2021-0051>
17. E. Małachowska, M. Dubowik, P. Boruszewski, J. Łojewska, and P. Przybysz, *Sci. Rep.*, **10**, 19998 (2020). <https://doi.org/10.1038/s41598-020-77101-2>
18. I. Haq, P. Mazumder, and A. S. Kalamdhad, *Bioresour. Technol.*, **312**, 123636 (2020). <https://doi.org/10.1016/j.biortech.2020.123636>
19. R. W. Hagemeyer, 'Pigments for Paper', TAPPI press, Georgia, 139-155 (1997).
20. C. Y. Liong, L. C. Lee, K. Osman, and A. A. Jemain, *AIP Conf. Proc.*, **1750**, 060017 (2016). <https://doi.org/10.1063/1.4954622>
21. A. Allen, M. R. Williams, and M. E. Sigman, *Forensic Chem.*, **16**, 100188 (2019). <https://doi.org/10.1016/>

- j.forc.2019.100188
22. M. Baron, J. Gonzalez-Rodriguez, R. Croxton, R. Gonzalez, and R. Jimenez, *J. Appl. Spectrosc.*, **65**(10), 1151-1161 (2011). <https://doi.org/10.1366/10-06197>
 23. S. Goswami and E. J. Wegman, *J. Stat. Sci. Appl.*, **4**(03-04), 65-84 (2016). <https://doi.org/10.17265/2328-224X/2015.0304.001>
 24. G. Cioccia, R. Wenceslau, M. Ribeiro, G. S. Senesi, J. Cabral, G. Nicolodelli, C. Cena, and B. Marangoni, *Microchem. J.*, **207**, 112142 (2024). <https://doi.org/10.1016/j.microc.2024.112142>
 25. P. V. Renterghem, P. E. Sottas, M. Saugy, and P. V. Eenoo, *Anal. Chim. Acta*, **768**, 41-48 (2013). <https://doi.org/10.1016/j.aca.2013.01.003>
 26. U. Islam, H. S. Alwageed, M. M. U. Farooq, I. Khan, F. A. Awwad, I. Ali, and M. R. Abonazel, *Sensors*, **23**(12), 5626 (2023). <https://doi.org/10.3390/s23125626>
 27. M. Tapiador, J. Gómez, and J. A. Sigüenza, IEA/AIE' 2004: Proceeding of the 17th international conference on Innovations in applied artificial intelligence, Canada, Ottawa, 625-632 (2004). https://doi.org/10.1007/978-3-540-24677-0_64
 28. M. H. Rahman, M. A. Bashar, F. H. M. Rafi, T. Rahman, and F. Mitul, 2013 International Conference on Informatics, Electronics and Vision (ICIEV), Dhaka, Bangladesh, 1-6 (2013). <https://doi.org/10.1109/ICIEV.2013.6572576>
 29. G. Mujtaba, L. Shuib, R. G. Raj, R. Rajandram, and K. Shaikh, *J. Forensic Leg. Med.*, **57**, 41-50 (2018). <https://doi.org/10.1016/j.jflm.2017.07.001>