Highlights

Pen Ink Library: An interactive database of writing instruments based on Vis-NIR reflection spectra and optical properties of inks

- New non-destructive methodology for identification of inks from handwriting strokes.
- Interactive database containing measured optical parameters and MSP-Vis-NIR spectra of 718 blue and black writing instruments.
- Semi-automatic tool for searching writing instruments in the database.
- Evaluation of computer-based methods for classification and comparative analysis of ink samples.

Pen Ink Library: An interactive database of writing instruments based on Vis-NIR reflection spectra and optical properties of inks

ARTICLE INFO

Keywords:
handwritten document examination
writing ink
interactive database
machine learning
microspectrophotometry
optical analysis

ABSTRACT

Many tasks in forensic examination of handwritten documents require identification of writing instruments that have ink of similar properties as the ink found on a questioned document. In this paper, we propose a new methodology for non-destructive identification of inks based on optical properties and reflectance spectra of the ink, measured from handwriting strokes. Building on this methodology, we develop an interactive database that we call the "Pen Ink Library", which lists 718 various writing instruments and enables systematic comparison and semi-automatic search of writing instruments, using the measured characteristics of their ink. To highlight the significance and applicability of the database, we additionally exploit the large amounts of collected measurements to design computer-based data analysis methods for classification and comparative analysis of ink samples. For validation of the semi-automatic search functionality of the Pen Ink Library we performed a series of blind tests using twenty randomly selected writing instruments. Here, an instrument with the same brand and model was found in nine cases, and an instrument with a different brand and model, but with identical spectrum and optical parameters, was found in five cases. Cross-validation of the computer-based data analysis methods on the measurements from the database yielded above 90 % accuracy of the classification method and 5.3 % to 12.7 % error rate of the comparative analysis method.

1. Introduction

In forensic examination of handwritten documents, it is often necessary to determine whether an original document has been altered by addition or overwriting of text portions. In other instances, the objective is to decide whether the handwritten text on multiple documents was made using the same writing instruments. Another common task is to assess whether a signature on a printed document was executed after printing, or if it was already present on a blank paper as a "bianco" signature, and the text was only printed later. Resolving these issues frequently requires the preparation of reference samples, using materials (in particular inks) similar to those found in the questioned documents. To accomplish this, a collection of writing instruments is put together, from which an instrument with ink of similar characteristics to those present on the questioned documents can be selected.

Various analytical methods can be employed to distinguish writing inks [1]. High performance thin layer chromatography (HPTLC) [2, 3] is considered to posses the highest discriminatory power, particularly for the analysis of dye. High-performance liquid chromatography (HPLC) and gas chromatography-mass spectrometry (GC-MS) provide information on the binders, additives, pigments and dyes [4], and are also useful for the analysis of solvents with the goal of determining the age of the analyzed document [5, 6]. Additionally, numerous instrumental techniques, summarized comprehensively by Gorziza [7], can be employed for the analysis of inks. One drawback of many of these methods is the necessity to manually extract a sample of the ink from paper, which causes damage to the questioned document. In numerous cases, the party requesting the forensic examination disagrees with the use of destructive analyses. In that situation non-destructive methods, such as Raman spectroscopy, and, for certain types of writing inks, attenuated total reflectance Fourier transform infrared spectroscopy (ATR-FTIR), are available. These methods can be employed to analyze samples in terms of the utilized dyes or pigments without significant disruption of the integrity of the document.

In many laboratories specializing on document examination, instrumental chemical analysis methods, as the ones discussed above, may be unavailable or are solely used for material differentiation rather than identification. Consequently, optical methods, that involve microscopic and spectral evaluation of the samples, can serve as a valuable alternative. This includes microspectrophotometric measurements in the visible and near-infrared region (MSP-Vis-NIR), or in the ultraviolet and visible region (MSP-UV-Vis) [8]. Using MSP-Vis-NIR in the reflectance mode, spectra can be measured without any degradation of the questioned document. Although the obtained spectrum cannot be

ORCID(s):

used to precisely identify the individual components of the dye or pigment mixture, it has a high discriminatory power, particularly when combined with mathematical or chemometric techniques [9, 10, 11, 12, 13].

In this paper, we present a new methodology for non-destructive identification of inks, based on standardized procedures for evaluation of optical properties and measurement of MSP-Vis-NIR spectra of handwriting strokes. Building on this methodology, we develop an interactive database that we call the "Pen Ink Library".[1] This database accommodates a large collection of blue and black-ink writing instruments, where each record contains detailed information about the manufacturer and model of the corresponding writing instrument, an MSP-Vis-NIR spectrum, and several additional optical parameters that we describe in the subsequent section. The user of the interactive database can upload the MSP-Vis-NIR spectrum of an ink sample and perform a semi-automatic search for a group of writing instruments with similar characteristics in the database. Finally, to further emphasize the practicality of the database, we demonstrate how the collected spectra can be used to design: machine learning methods for automatic classification of unknown inks into one of the internal categories within the database; and statistical methods for comparative analysis of questioned and reference writing instruments.

2. Material and methods

In this section we first describe techniques that were used for collection and processing of the ink samples within the database. Then we introduce several methods for characterization of optical properties of inks. Finally, we discuss how the MSP-Vis-NIR spectrum of each ink sample was measured.

2.1. Samples

Since 2012, the author's home institution has been building a collection of ink samples from various writing instruments that are available on the domestic market. This collections consists of 718 ink samples of writing instruments with blue (442 samples) and black (276 samples) ink. These include ballpoint pens, gel pens, rollerball pens, liners, fiber-tip markers, and fountain pens. Each sample is labeled with a registration number, which is linked to basic information in the database, such as the brand, model, descriptive details, color, country of origin, ink type, erasability, point type, point width, and year of sample acquisition. Ink samples of selected writing instruments were collected repeatedly over the years to verify the consistency of their composition across different batches.

Samples of the ink were prepared by manually drawing strokes in the form of loops and lines on a white office paper of $80\,\mathrm{g\,m^{-2}}$ weight. Two variants of each sample were collected, differing in the stroke pressure: light pressure and heavy pressure. The samples are stored between sheets of blank paper in dark folders under normal office conditions. Fresh writing is added and measured after two or three years to assess any differences between older and fresher records.

2.2. Optical methods

For optical evaluation of the samples we used standard equipment available in most forensic laboratories. This includes a stereoscopic microscope with magnification of up to 50× (Zeiss STEMI 2000), a digital microscope with coaxial illumination, polarization, and a magnification of 350× or higher (HIROX with MXG-2500REZ lens), and a video-spectral comparator for image analysis across different spectral regions, including the analysis of IR luminescence and IR ink absorption (Projectina DocuCenter Expert).

The following eight parameters were assessed using the optical methods for each ink sample:

- Apparent viscosity
- Width of light strokes
- Width of heavy strokes
- Color hue
- IR absorption
- IR luminescence
- Changes in luminescence over time

[1] For access to the database, please contact the first author.

- Reflection

A detailed description of the optical parameters is provided below in separate subsections.

2.2.1. Apparent viscosity

With an optical microscope, two types of inks can be distinguished. Inks with *high viscosity*, resembling a paste, tend to adhere to the surface of the paper and along its fibers. There may be areas between the fibers that the ink does not cover and it is not absorbed deep into the paper. Inks with *low viscosity*, such as liquid or gel inks, diffuse into the paper and fill the gaps between the fibers. We do not distinguish between gel inks and liquid inks (rollerball pens, fiber-tip liners, or fountain pens) since they are difficult to distinguish under the microscope [14].

2.2.2. Colour hue

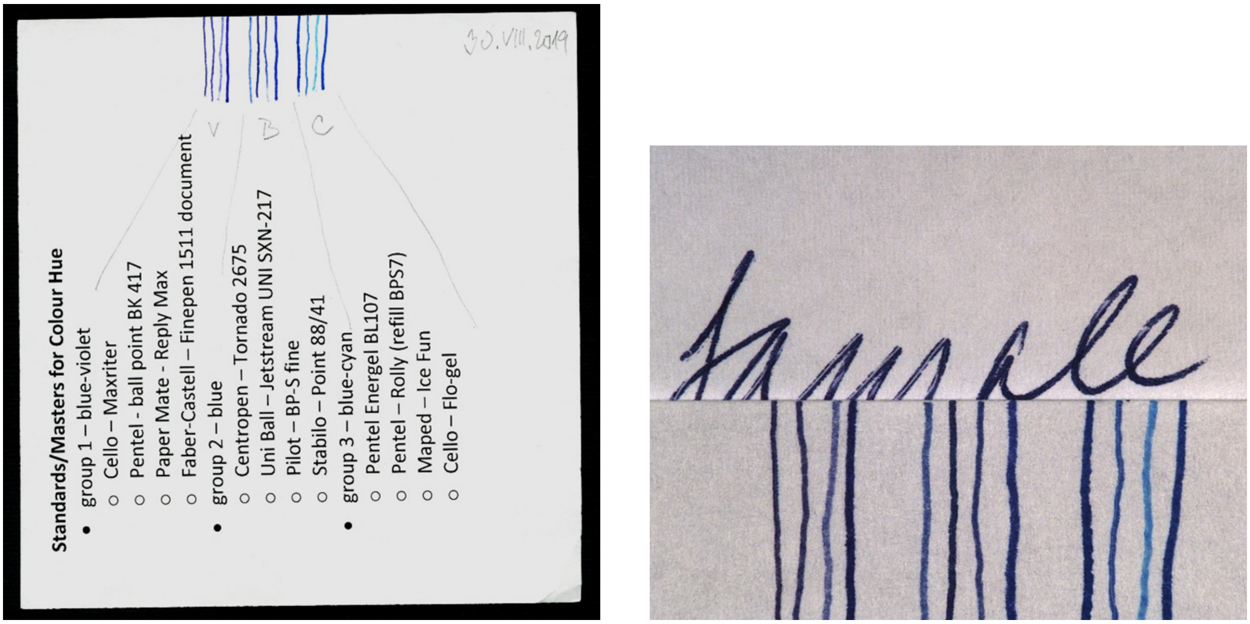
For evaluating the color hue of blue inks, three categories were established: *blue-violet*, *blue*, and *blue-green* (cyan). All three categories are defined by a comparative standard which consists of strokes created with selected writing instruments. For each category of blue inks, four representative writing instruments from the collection were chosen. When classifying an unknown ink sample these four representatives need to be viewed as a whole (see Figure 1). The comparative standard was prepared on an appropriate substrate compatible with the analyzed sample, making sure that the color tone of the substrate has a similar effect on the perceived ink color as the paper carrying the ink sample. The comparison of hues was conducted using an optical stereoscopic microscope to avoid color distortion caused by computer displays. The prepared comparative standard was then stored in a dark environment for future reference. In cases where it is difficult to confidently assign a hue to one of the categories, it is recommended to search both considered categories in the database.

The following writing instruments have been selected for each category:

- Group 1 – blue violet
  - Cello – Maxriter
  - Pentel - ball point BK 417
  - Paper Mate - Reply Max
  - Faber-Castell – Finepen 1511 document
- Group 2 – blue
  - Centropen – Tornado 2675
  - Uni Ball – Jetstream UNI SXN-217
  - Pilot - BPE-GP-CFL-BG 1107, or BP-S fine
  - Stabilo – Point 88/41
- Group 3 – blue cyan
  - Maped - Gel Freewriter, or Pentel – Energel BL107
  - Pentel – Rolly (or refill BPS7)
  - Maped – Ice Fun
  - Cello – Top Gel, or Flo-gel

The hue of black ink strokes was analyzed at 350x magnification under polarized light. Four categories of black inks were established:

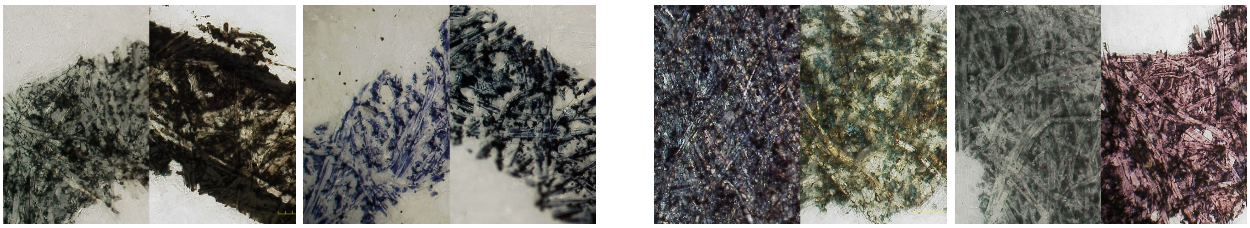
- Black – black and gray hues (e.g., Stabilo Performer+, Centropen Paint Marker 9211). Examples are shown in the Figure 2a.
- Dark blue – strokes with dark blue to dark purple tones (e.g., Herlitz Ballpoint Pen, Cello Pronto). Examples are shown in the Figure 2b.



(a) (b)

Figure 1: Real example of color hue examination: a) the colour hue comparative standard with 3 groups and 12 blue ink instruments; and b) a detailed view showing a handwriting sample together with the blue strokes on the comparative standard. (Reproduction of colour hues may be inaccurate).

- Spotty – black and gray strokes with (rare) spots of blue color. Some inks may have spots of red hue, but if they are present along with blue ones, they fall into the "spotty" category (e.g., Herlitz My Pen, Centropen 2637 M OHP Permanent). Examples are shown in the Figure 2c.
- Red – strokes with red hue or black and gray strokes with red spots (e.g., Faber-Castell Broadpen 1554, Pilot Frixon Ball). Examples are shown in the Figure 2d.



(a) (b) (c) (d)

Figure 2: Detail of black inks for categories: a) black, b) deep blue, c) spotty, and d) red.

2.2.3. *IR absorption*

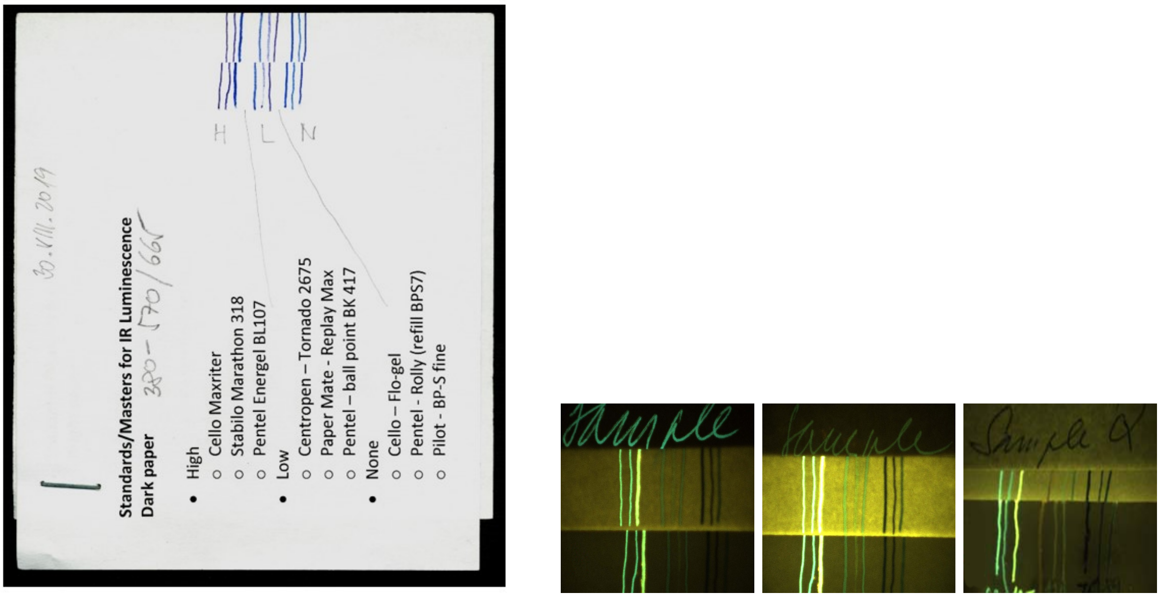
Based on absorption of radiation in the infrared (IR) range with wavelengths of 850 nm to 1000 nm (also termed the near-IR range), we discriminate ink strokes which absorb IR, absorb it partially and don't absorb it. An ink stroke which appears dark in the near-IR range of the spectrum classifies in the category *Yes*. If the writing stroke absorbs the radiation partially, resulting in a weakened but still visible stroke (in the near-IR range) it falls within *Middle* category. If the writing stroke does not absorb the radiation at all and disappears completely, or only its outline remains visible due to pressure indention, it falls within the *No* absorption category.

2.2.4. *IR luminescence*

To evaluate luminescence, a comparative standard with three categories of standards is used to ensure balanced exposure. We discriminate inks with: *high*, *low*, and *none* luminescence. The comparative standard is created using selected writing instruments on two types of paper—office paper, which appears dark in the corresponding luminescence range of the spectrum, and office paper that emits luminescence in that range. The luminescence of ink is overlapped by luminescence from the luminescent papers, which could lead to confusion when analyzing an unknown ink sample. The same comparative standard can be used for both blue and black writing strokes (see Figure 3).

Writing strokes are evaluated in the excitation range of 380 nm to 570 nm and with 665 nm emission filter. The following writing instruments were chosen as representatives for each category:

- High luminescence – strokes appear to be glowing (in the luminescence range)
  - Cello Maxriter
  - Stabilo Marathon 318
  - Stabilo Gel Refill (or Pentel Energel BL 107)
- Low luminescence – strokes appear bright but without a glowing effect
  - Paper Mate 300 Gel (or Centropen Tornado 2675)
  - Paper Mate Replay Max
  - Pentel BK 417 Ballpoint (refill BKS7E)
- None luminescence – strokes appear dark
  - Maped Freewriter Gel (or Cello Flo Gel)
  - Pentel Rolly
  - Pilot Ballpoint Pen BP-S



(a) (b) (c) (d)

Figure 3: Example of the luminescence examination: a) the luminescence comparative standard, a stroke sample with b) high, c) low, and d) no luminescence.

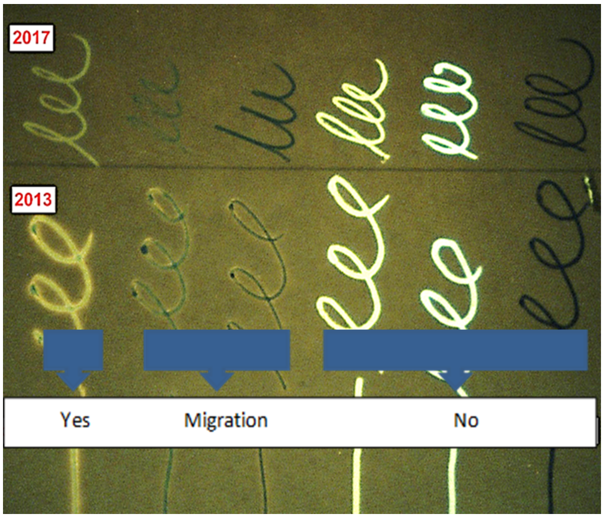


Figure 4: Change of the luminescence over time. The strokes above the black line were freshly written at the time of evaluation. The three categories are indicated on the white strip.

2.2.5. *Changes in luminescence over time*

The luminescent effect may not be stable for all inks. Any change in luminescence should be evaluated several years after creating the ink strokes on paper and storing them under normal office conditions (humidity 30 % to 60 %, temperature 18 °C to 30 °C, and storage in dark compartments in filing cabinets). With respect to changes in luminescence, we discriminate the three categories, that are listed below and shown in Figure 4.

- No change – the luminescence intensity (and quality) remains the same
- Significant change – the luminescence intensity increases over time
- Luminescence migration – over time the luminescence increases at the stroke outline

Recently measured samples, where it is still too soon to measure any changes in luminescence, are labeled as "To be analyzed" in the database.

2.2.6. *Reflection*

Reflection refers to the metallic sheen of the ink observed on a steady writing stroke (where the amount of ink deposit is neither reduced nor increased) under diffused light, compared to the image seen under polarized light at sufficient magnification (250x). Three categories of reflection were established:

- High reflection – strong reflections with colour change to bronze, red, purple - metallic character
- Low reflection – There is a colour change but only locally on the surface of the fibres. There is not as much change in tone between fibers, the gloss is lower
- None reflection – no colour change, rather only a basic gloss of the paper fibres can be observed

2.3. ViS-NIR Spectroscopy

For the measurement of the reflectance spectra of writing strokes on paper using Vis-NIR spectroscopy, OceanOptics Flame S fiber optic spectrophotometer was used in the range of 400 nm to 900 nm with a Zeiss lens at 5x magnification, 0.170 mm diaphragm size, 8 measurements averaging capture and spectrum smoothing. Before a measurement session, the device was calibrated using a white standard (spectralon) to determine the maximal and (with the light turned off) minimal intensity. Measurement of each spectrum then consisted of two steps. First, the spectrum of the paper (background spectrum) was measured, and then the spectrum of the ink was measured (if possible, at the intersection of the writing strokes, where there is double the amount of ink). To minimize the impact of the paper type, we then subtracted the paper spectrum from the ink spectrum as we explain later in Section 2.3.2.

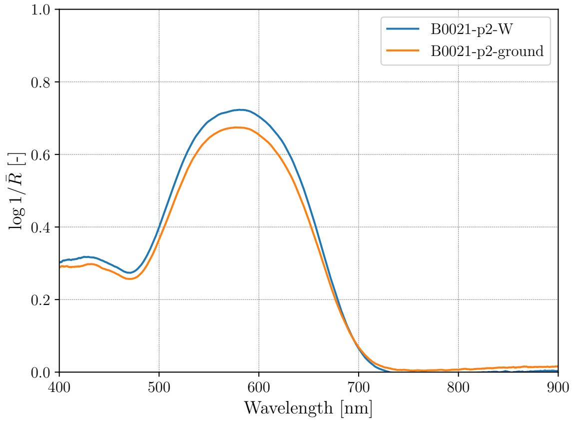


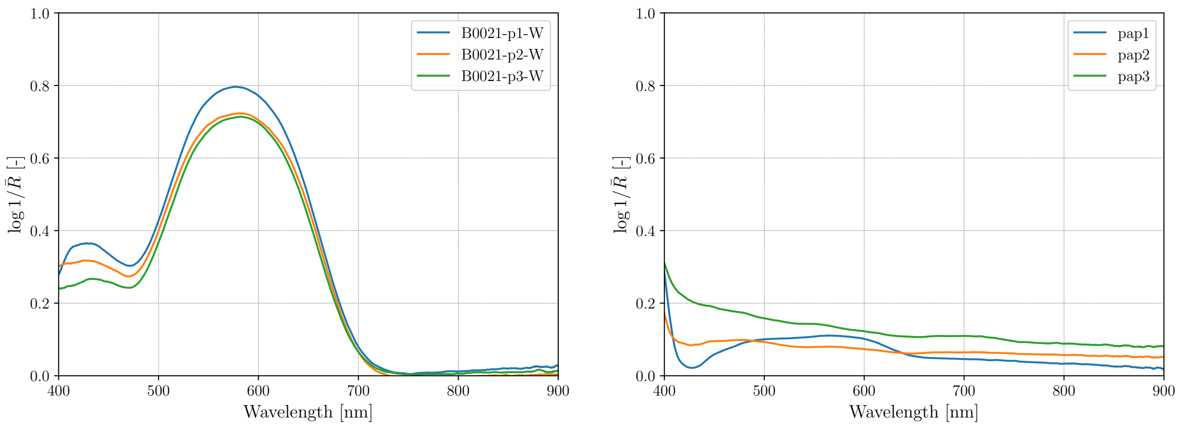
Figure 5: Influence of the white calibration. Ink spectrum measured with white standard (spectralon) with processing (subtraction of paper spectrum) – red line, and spectrum measured with background setting as 100 % reflection – blue line.

2.3.1. Primary experiments

In the development of the method, the influence of various factors on the measured data was first verified.

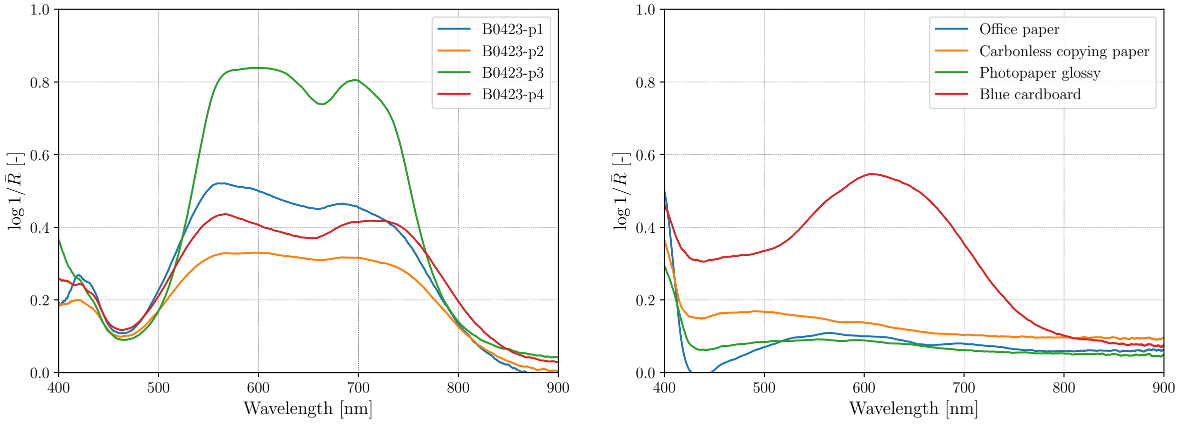
Measurement calibration. To ensure higher measurement stability, the spectra in the database were measured with the spectrophotometer calibrated to a white standard (spectralon) and with the paper spectrum subtracted, as we discussed above. An alternative approach is to simply set the reflectance of a blank area of the paper of the respective sample as the 100 % white. As shown in Figure 5, the measured spectra resulting from the two calibration approaches do not differ significantly. We thus suggest, that for the analysis of a questioned sample, it is sufficient to use the second (simpler) approach.

Influence of paper type. The use of different types of office paper does not have significant impact on the observed spectra, although slight deviations may occur in the region around 415 nm, where the fluorescence of certain papers intervenes (see Figure 6). Photographic paper with a glossy surface provides spectra with sharper peaks. Tinted paper affects certain parts of the spectrum. NCR (no carbon required) paper may also affect the spectra of certain inks (see Figure 7).



(a) (b)

Figure 6: Influence of different office paper: a) Spectra of Pilot frixion pen on paper 1, 2 and 3, b) spectra of paper: pap 1 – JET multifunctional extra white, pap 2 – recycled, pap 3 – PROFI copy paper.



(a) (b)

Figure 7: Influence of other types of paper: a) spectra of Uni-ball Air Micro (no. B0423) blue pen on four different papers b) spectra of different papers.

Influence of ink concentration. The amount of ink deposit on the paper only affects the overall intensity of the reflectance spectra. The optimal position for measuring the spectra is at the intersection of the writing strokes where there is double the amount of ink (see Figure 8). Writing instruments, such as fiber-tip pens, transfer more ink. In such cases, the reflectance intensity is too high, and individual peaks are not discernible, so it is advisable to measure the peripheral part of the writing stroke, where the ink concentration may be lower.

Influence of record age. Ink from writing strokes that have not undergone visible degradation does not exhibit significant changes in the reflectance spectra.

2.3.2. Data processing

Before further analysis, the measured spectra were processed as follows [9, 11, 13]. The measured reflectance values $R \in (0, 1]$ are normalized as $\bar{R} = R/R_0$ to supress the reflectance of the paper $R_0 \in (0, 1]$. Subsequntly, for visualization and futher processing the normalized reflectance values are converted as $D = \log 1/\bar{R}$. In the range of 400 nm to 900 nm, the spectra are sampled on a uniform grid of $N = 251$ points. Each converted spectrum is thus represented by the sequence $D[n]$, for $n = 1, 2, \dots, N$.

3. Pen Ink Library Application

The Pen Ink Library was developed as a web application based on a client-server concept. The client provides a graphical user interface (frontend) which communicates via the internet with a server that contains most of the functionality and evaluation logic (backend). This includes the functionality for analyzing and comparing spectra of inks, and for handling any database queries. The Pen Ink Library implements a user-friendly interface for navigation through the samples in the database and perform semi-automatic search of writing instruments based on MSP-Vis-NIR spectra and optical properties of the ink. In this section, we briefly describe the user interface of the application and explain the inner workings of the integrated semi-automatic search tool.

3.1. User Interface

The user interface is part of the frontend which runs in the user's browser. It includes all the functionality that is necessary for graphical display of tables, graphs, matrices, and images. After successful login, the user can perform the following activities:

- Perform analysis of an uploaded spectra.
- Compare different samples based on their spectra.

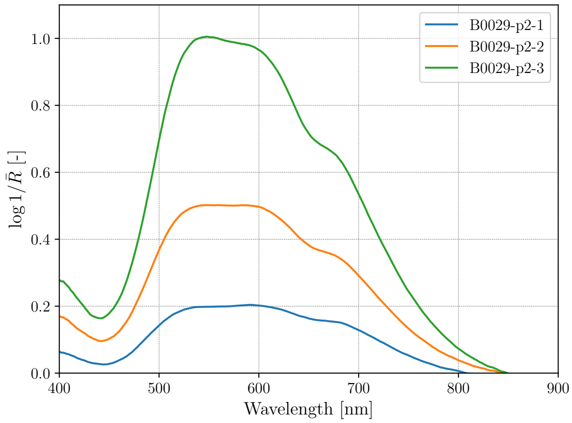


Figure 8: Spectra of Schneider xpress measured from a single pen stroke, double pen stroke and 4 times intersecting stroke.

- Filter and display writing instruments according to their spectra and optical parameters.
- Browse and (if she has administrative rights) edit the database of writing instruments.
- Manage her user account or (if she has administrative rights) the accounts of other users.

In the case of using a different brand or model of microspectrophotometer with a different file format for MSP-Vis-NIR spectral data, the user can download and use our template to convert their data to a format compatible with our application. Screenshots of user interface can be found in Section S1 of our supplementary material.

3.2. Semi-automatic Search Tool

As discussed in the introduction, the iteractive database implements a semi-automatic tool for searching writing instruments with ink of similar characteristics to that of a questioned sample. The search is based in part on a distance metric between two spectra and additional information (optical parameters) provided by the user. We remark, that, apart from the distance metric considered below (Euclidean distance), also other distance/similarity metrics have been tested—all yielding very similar results.

In the first step, the user uploads a MSP-Vis-NIR spectrum, which is internally converted into the $\log 1/\bar{R}$ spectrum $D[n]$, with $n = 1, 2, \ldots, N$, as discussed in Section 2.3.2. A normalization step is then performed, which ensures that the spectrum is non-negative and sums up to one, i.e.,

$$\bar{D}[n] = \frac{D[n] - D_{\min}}{D_{\text{sum}}}, \quad n = 1, 2, \ldots, N, \tag{1}$$

where

$$D_{\min} \triangleq \min_{n=1,2,\ldots,N} D[n], \tag{2}$$

$$D_{\text{sum}} \triangleq \sum_{n=1}^{N} D[n] - N\, D_{\min}. \tag{3}$$

This step minimizes variability between the spectra caused by inconsistent ink deposit on the paper and calibration errors. Subsequently, the Euclidean distance

$$d_k \triangleq \sqrt{\sum_{n=1}^{N} (\bar{D}_k[n] - \bar{D}[n])^2}, \quad k = 1, 2, \ldots, K, \tag{4}$$

is calculated between the questioned spectrum and all K (normalized) spectra $\bar{D}_k[n]$, $k = 1, 2, \ldots, K$, within the database. The writing instruments from the database are then sorted by increasing d_k and displayed to the user.

In the second step, the user is allowed to enter optical parameters of the searched writing instrument to narrow down the search. After entering each new property the list of sorted writing instruments is filtered to remove instrument that do not agree with the entered property value. All optical parameters that the user can specify were discussed in Section 2.2.

4. Computer-based Data Analysis Methods

The abundance of data provided by our database, can be exploited to design computer-based data analysis methods that can answer several questions encountered in forensic practice. Here we present two methods that utilize the MSP-Vis-NIR spectra from the database. The first method, based on a convolutional neural network, can be used to automatically assign a spectrum of unknown writing instrument (not contained in the database) to one of the internal classes that we introduce subsequently. The second method, based on a multivariate statistical model, can be used to judge whether two unknown spectra come from the same writing instrument.

4.1. Classification Using CNNs

Convolutional neural networks (CNNs) play a significant role in machine learning, particularly for processing and analyzing spatially correlated data, such as image data or, in our context, MSP-Vis-NIR spectra. Through convolutions, these networks efficiently capture spatial and temporal relationships in data. A typical CNN comprises several layers, including convolutional layers and fully connected layers—the latter interpret the extracted features from the convolutional layers for tasks like classification.

For ink identification, we employed the ResNet [15] (Residual Network) architecture, which addresses the vanishing gradient problem. ResNet utilizes residual blocks that enable direct signal transmission across multiple layers without alteration, a feature implemented through so called skip connections. Our ResNet in modification known as ResNet-18 has been adapted for use with 1D signals. In the Figure 9 is a schematic representation of the used network. The main motivation of using the CNN as the first step is its ability to learn even tiny differences that are difficult to recognize by traditional methods.

4.2. Comparative Analysis Using TLL Model

In this section we present a multivariate comparative method based on a two-level linear (TLL) model [16, Sec 7.6.2] and show how it can be applied to compare pen inks based on their MSP-Vis-NIR spectra. A similar approach has been used, e.g., in [13] and [17].

Let us consider the following hypothetical situation. A questioned document, written by an unknown writing instrument, is provided together with a comparative document written by one specific instrument. Based on the MSP-Vis-NIR spectra, extracted from both documents, our goal is to quantify the evidence in support of the following propositions [13]:

- The prosecutor's proposition H_p states that the ink from the questioned documents and the ink from the comparative document are from the same writing instrument.
- The defense proposition H_d states that the ink from the questioned document is from some writing instrument, from the relevant population of writing instruments, other than the instrument which left the ink on the comparative document.

Let $\boldsymbol{y}_n$, for $n = 1, 2, \ldots, N$, denote feature vectors extracted from a MSP-Vis-NIR spectrum of one specific writing instrument. Under the TLL model, $\boldsymbol{y}_n$ is given by

$$\boldsymbol{y}_n = \boldsymbol{\theta} + \boldsymbol{\epsilon}_n, \quad n = 1, 2, \ldots, N, \tag{5}$$

where $\boldsymbol{\epsilon}_n \overset{iid}{\sim} \mathcal{N}(\mathbf{0}, \boldsymbol{U})$ and $\boldsymbol{\theta} \sim \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{C})$. Here the hyper-parameters $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{\mu}$, denote the within-source covariance matrix, between-source covariance matrix, and between-source mean, respectively. The probability density function (pdf) of the feature matrix $\boldsymbol{Y} \triangleq \left(\boldsymbol{y}_1\ \boldsymbol{y}_2\ \cdots\ \boldsymbol{y}_N\right)$, given $\boldsymbol{\theta}$, is thus

$$p(\boldsymbol{Y}|\boldsymbol{\theta}) = \prod_{n=1}^{N} \mathcal{N}(\boldsymbol{y}_n; \boldsymbol{\theta}, \boldsymbol{U}). \tag{6}$$

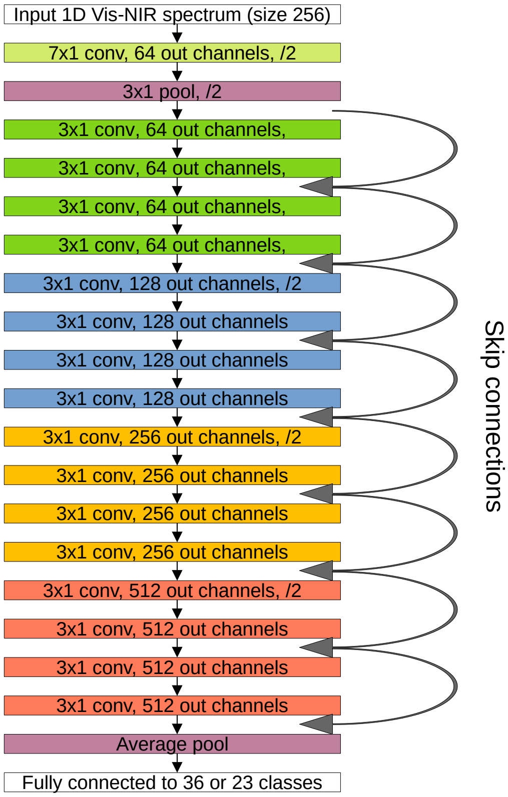


Figure 9: ResNet-18-based architecture used for blue ink classification. Contrary to classical ResNet-18 we used 1D convolutions in each layers. Last fully connected layer shown in the figure corresponds to classification of 36 classes of blue ink spectral numbers or 23 classes based on pen IDs. The number of classes depends on evaluated scenario described in 5. Gray bent arrows represent skip connections.

In the two level model, each pen is characterized by the parameter $\boldsymbol{\theta}$ which determines the mean of the feature vectors extracted from the MSP-Vis-NIR spectra of the ink.

To quantify the evidence in support of either proposition we use the likelihood ratio

$$\Lambda(\boldsymbol{Y}_{\mathrm{q}}, \boldsymbol{Y}_{\mathrm{c}}) = \frac{p(\boldsymbol{Y}_{\mathrm{q}}, \boldsymbol{Y}_{\mathrm{c}} | H_{\mathrm{p}})}{p(\boldsymbol{Y}_{\mathrm{q}}, \boldsymbol{Y}_{\mathrm{c}} | H_{\mathrm{d}})}, \tag{7}$$

based on the feature matrices from the questioned document $\boldsymbol{Y}_{\mathrm{q}} = \left(\boldsymbol{y}_{\mathrm{q},1}\ \boldsymbol{y}_{\mathrm{q},2}\ \cdots\ \boldsymbol{y}_{\mathrm{q},N_{\mathrm{q}}}\right)$ and the comparative document $\boldsymbol{Y}_{\mathrm{c}} = \left(\boldsymbol{y}_{\mathrm{c},1}\ \boldsymbol{y}_{\mathrm{c},2}\ \cdots\ \boldsymbol{y}_{\mathrm{c},N_{\mathrm{c}}}\right)$. Here the likelihood function under hypothesis H_{p} is based on the assumption that the questioned and comparative feature vectors have the same (but unknown) mean, which is then integrated out as a nuisance parameter, i.e.,

$$p(\boldsymbol{Y}_{\mathrm{q}}, \boldsymbol{Y}_{\mathrm{c}} | H_{\mathrm{p}}) = \int p(\boldsymbol{Y}_{\mathrm{q}} | \boldsymbol{\theta}) p(\boldsymbol{Y}_{\mathrm{c}} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{8}$$

The likelihood function under hypothesis H_{d}, on the other hand, is based on the assumption that the questioned and comparative feature vectors have different (unknown) means, that are also integrated out, i.e.,

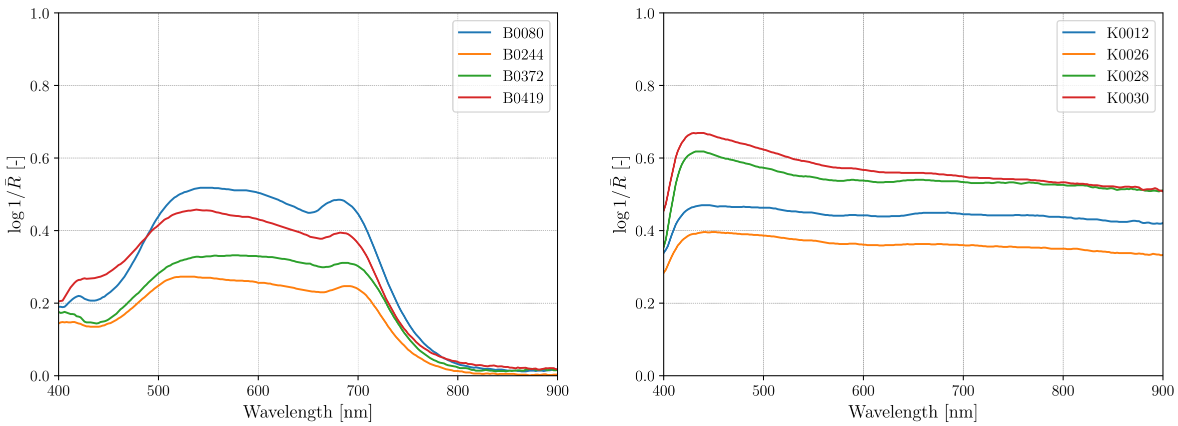
$$p(\boldsymbol{Y}_{\mathrm{q}}, \boldsymbol{Y}_{\mathrm{c}} | H_{\mathrm{d}}) = \int p(\boldsymbol{Y}_{\mathrm{q}} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \int p(\boldsymbol{Y}_{\mathrm{c}} | \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}. \tag{9}$$

Both (8) and (9) have a closed form, which can be found in [18]. A value of the likelihood ratio $\Lambda(\boldsymbol{Y}_{\mathrm{q}}, \boldsymbol{Y}_{\mathrm{c}})$ larger than one signifies that the evidence supports proposition H_{p}, whereas value smaller than one signifies that the evidence supports proposition H_{d}.

In practice, hyper-parameters $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{\mu}$ are typically unknown, and they must be estimated from a training set of feature vectors. If the dataset is balanced, i.e., categories (writing instruments) have equal number of samples, simple closed-form estimators are available [18]. Unfortunately, they do not apply to an unbalanced dataset as the one we consider. For that reason, to estimate the hyper-parameters, we developed an iterative expectation-maximization (EM) algorithm that we describe in Appendix A. In contrast to the closed form estimators of [18], the EM algorithm can, furthermore, avoid certain numerical issues which occur, e.g., when the samples within a category happen to be linearly-dependent.

5. Results and Discussion

For the purpose of validation of the semi-automatic search tool and performance evaluation of the classification and comparative analysis methods we consider each record in the database to be associated with a "spectrum number" and a "pen ID". The spectrum number, with value in the range SB01–SB36 for blue ink instruments and SK01–SK23 for black ink instruments, identifies records with similar ink spectra.The spectral numbers were assigned manually by a forensic expert, with the goal of grouping instruments which have the same (or indistinguishably similar) ink spectrum.The pen ID corresponds to the unique identifier that is assigned to each writing instrument as it is entered into the database.



(a) (b)

Figure 10: The most frequent waveform of (a) blue ink spectrum is spectrum number SB11 and (b) black ink spectrum is spectrum number SK05.

5.1. Semi-automatic Search Tool

In order to validate the functionality of the tool for semi-automatic search of writing instruments implemented in the application, a series of blind tests was conducted with 10 blue and 10 black ink instruments. Thirteen samples were created using instruments from our collection that are already contained in the database, and seven samples were created using random instruments that are not contained in the database (an instrument of the same brand and model might still be contained). Each sample was labeled with a unique id, QB1–QB10 for blue ink instruments, and QK1–QK10 for black ink instruments. An expert, unaware of the brand and model of the instruments used for creation of the samples, was then asked to measure the MSP-Vis-NIR spectra and evaluate optical parameters of the samples, enter these measurements into the application, and perform a search for matching instruments in the database. The results of the blind tests are presented in Table 1. Here the column "First rank result" shows the first ranking instrument in the search results. If an instrument of the same brand and model has ranked at worst third in the search results, we state that the correct instrument was found. If an instrument with the same spectrum number and optical parameters as the searched instrument has obtained a rank or 1–10, we state that the correct group of instruments was found. Alternatively, if the correct instrument ranked worse than third, and simultaneously the correct group of instruments ranked worse than tenth, we state that the search result was incorrect. The three possible outcomes are indicated in the

Table 1
Results of the blind tests conducted for validation of the semi-automatic search tool. We remark that QK10 has a unique spectrum that does not compare to any of the spectra in the database. Evaluation of the search result for this instrument was thus impossible similar to the "Unknown" writing instruments.

Searched writing instrument				Search result		
ID	Database ID	Brand	Type	First rank result	Evaluation	Rank
QB1	B0306	Schneider	Slider Touch	Perro Sissy BP1405	Incorrect	-
QB2	-	Pilot	V ball grip GRP	Pilot V ball grip GRP	Correct	1
QB3	B0426	Carioca	OOPS erasable	M&G iEras POP	Group	2
QB4	-	Parker	Ink cartridge	Waterman standard	Incorrect	-
QB5	B0359	Dolphin	Lineplus 420 BP	Lineplus refill for Dolphin 420 BP	Correct	1–2
QB6	B0310	Pilot	BPS-GP	Pilot RFT-4-F for ballpoint pen	Group	10
QB7	-	Solidly	Techjob Office	Unknown ballpoint	Incorrect	-
QB8	B0013	Uniball	Jetstream SXN217	Uniball Jetstream SXN217	Correct	1
QB9	B0189	Maped	Dark	Maped Dark	Correct	1
QB10	B0193	BIC	Cristal gel	Herlitz Gel click	Group	6
QK1	K0256	Stedtler	Ball 423	Stedtler Ball 423	Correct	1
QK2	K0214	Schneider	Maxx 278	Schneider Maxx 278	Correct	1
QK3	K0138	Cello	Top Gel	Cello Top Gel	Correct	1
QK4	K0245	Kores	K-marker thin	Centropen 2846 permanent	Group	2
QK5	K0242	FranklinCovey	Ballpoint pen refill	Pentel ballpoint BK 437	Group	3
QK6	K0002	Pilot	Frixion ball	Pilot Frixion ball Clicker	Correct	1–3
QK7	-	Unknown	-	Unknown ballpoint	-	-
QK8	-	Pentel	EnerGel	Pentel EnerGel	Correct	1
QK9	-	Unknown	-	Kores K1 - F	-	-
QK10	-	Centropen	2651	-	-	-

column “Evaluation” using the labels “Correct”, “Group”, and “Incorrect”. The column “Rank” additionally indicates the rank that was assigned to the searched instrument in the search results. As we can see, in twelve cases the exact instrument, or an instrument within the same group, was found and received a rank of 1–3 in the search results. In two cases instruments within the same group were found with rank 6 and 10. In six cases, the correct instrument was found on a lower rank or was not found (twice a corresponding instrument was not in the database, twice the manufacturer of the searched instrument was not known). For unbranded writing instruments recorded as “Unknown”, the accuracy of the result cannot be verified. With increasing size of the database, samples with similar spectral profiles yet different brand and model occur more frequently. In that case, it is often possible to narrow down the search results based on differences in optical parameters. Figures comparing the spectra of the instruments in the search results can be found in Section S2 of our supplementary material.

When judging whether two ink samples come from the same writing instrument, it is necessary to consider their uniqueness with regard to the population distribution of inks. To illustrate this, let us examine the uniqueness of the samples within the database. In the wide range of different writing instruments, there are inks whose spectra are very rare, with a frequency of occurrence below 1 %. The discriminating power of such inks is very high. In contrast, around 14 % of blue-ink instrument in the database have very similar spectra and their discriminating power is thereby accordingly lowered (see Figure 10a). Similar phenomenon can be observed in the black-ink instruments, where around 19 % of black inks in the database have a very similar spectra (see Figure 10b). We conjecture, that this occurs because many manufacturers, especially the ones who do not brand their products, use inks of one common composition. In these cases it is impossible to distinguish individual manufacturers or brands of writing instruments based on the spectra and optical parameters of their ink. On the other hand, we also observed differences in the spectra of the ink between instrument model of the same brand in the database. We also observed that instruments of the same brand and model, which we acquired for our collection in different years, are characterized by more or less the same spectra. This indicates, that after starting the production of a certain model, the manufactures are unlikely to change the composition of the ink.

Table 2
Precision and accuracy of the 1D ResNet-18-based classifier averaged over 10 cross-validation cycles.

Classes	Num. of classes	Precision	Accuracy
spectrum number	36	**0.956**	**0.937**
pen ID	23	**0.932**	**0.903**

5.2. Classification Using CNNs

In this section we describe the settings and results of classification of ink samples, using the CNNs of Section 4.1. For training and testing the network we use two labeled datasets created from the records in the database. The first dataset uses the pen ID as a label and consists of blue ink writing instruments that were measured at least five times. This dataset contains 165 samples of 23 different writing instruments. The second dataset uses the spectrum number as a label and consists of blue ink writing instruments for which at least five records with the same spectrum number exist in the database. We divided each dataset into 10 non-overlapping partitions, and in order to assess the classification accuracy, we employed cross-validation where nine partitions were used for training and one partition for testing during 10 separate training and testing cycles.

For training of the network, we used the cross-entropy loss and the Adam optimizer with an initial learning rate of 0.01, which was gradually reduced to 0.0001 over approximately 200 epochs. The epoch count was determined in preliminary experiments where we observed the changes in the loss function value during training. The evaluation results are presented in form of precision, and accuracy metrics in Table 2. The tabulated values represent the average of the considered metrics over the 10 cross-validation cycles.

We can see the slightly better performance when the second dataset (using spectrum number as a label) is used. This is because it contains a larger amount of samples in each class. However, the difference of only few percents is not significant and overall performances of CNN are above 90 %.

5.3. Comparative Analysis Using TLL Model

The accuracy of the statistical comparative method of Section 4.2 was evaluated on a subset of the records in the database with the use of the pen IDs. The subset consists of blue ink writing instruments that were measured at least five times, i.e., at least five records with the same pen ID exist in the database. The resulting subset of the database consists of 165 records of 23 different instruments. The requirement of having at least five spectra was found to be necessary for the feature extraction method based on linear discriminant analysis (described below) to work properly. The comparative analysis method itself does not have this requirement.

We experimented with two methods for feature extraction: CIE XYZ, proposed in [13], and linear discriminant analysis: In both cases the spectra are normalized using (1)–(3) before feature extraction.

- CIE XYZ – The feature vector is populated by the CIE XYZ tristimulus values calculated from the normalized spectrum $\bar{A}[n]$. For the calculation we used the CIE 1931 2° standard observer and the standard illuminant D65 (see [19, Ch. 3] for details of the calculation). Depending on whether we use all three values (X, Y, and Z) or a pair of the values (e.g., X and Y) the length of the feature vector is either three or two.

- Linear discriminant analysis – The feature vector is populated by a projection of the normalized spectrum $\bar{A}[n]$ onto a lower-dimensional space. Linear discriminant analysis (LDA) is used to find a projection with maximal between-class variance relative to within-class variance on a training dataset of normalized spectra [20, Sec. 4.3.3]. The length of the feature vector can be chosen arbitrarily.

In Table 3 we report the result of cross-validation on the described database subset. The presented metrics, i.e., false positive rate (FPR), false negative rate (FNR), and error rate (ER), are evaluated over all possible splits, where one split corresponds to a validation partition, which contains only samples of two selected classes, and training partition, which contains all the remaining samples. The training partition is used to find the LDA projection and to estimate the hyper-parameters $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{\mu}$. The validation partition, on the other hand, is used to calculate the error metrics. Using the CIE XYZ feature vectors we obtained the best performance in terms of ER, when all tristimulus values are used. LDA features yield lower ER and FPR than CIE XYZ features. The FNR is, nevertheless, higher. The optimal feature vector length for LDA is 6 for the considered dataset.

Table 3
Performance of multivariate analysis based on different types of feature vectors: CIE XYZ using the XYZ, XY, XZ, and YZ values, and LDA using 4, 5, 6, 7, and 8 dimensional feature vector. The tabulated values represent the false positive rate (FPR), false negative rate (FNR), and error rate (ER) in percent.

CIE XYZ

val.	FPR	FNR	ER
XYZ	6.2	8.1	7.2
XY	16.5	8.6	12.7
XZ	12.1	7.7	10.0
YZ	13.6	6.9	10.4

LDA

len.	FPR	FNR	ER
4	2.8	10.2	6.3
5	1.8	10.6	6.0
6	1.5	9.3	5.3
7	1.3	10.5	5.7
8	1.3	10.9	6.0

6. Conclusion

We proposed a new methodology for non-destructive identification of inks based on optical properties and MSP-Vis-NIR spectra, measured from handwriting strokes. Building on this methodology, we designed the Pen Ink Library interactive database which showcases a collection of 718 writing instruments and enables efficient semi-automatic search of specific instruments, or group of instruments, based on the characteristics of their ink. Characteristics that we used for this purpose include the MSP-Vis-NIR spectra as well as several optical parameters, such as apparent viscosity, IR luminescence, etc. Through a series of blind tests we showed that a combination of MSP-Vis-NIR spectra with a few optical parameters is typically sufficient to identify the exact instrument or at least a group of instruments with similar ink. Additionally, we demonstrated that the database can be used to build computer-based data analysis methods for automatic classification of unknown spectra, or to assess whether two spectra come from the same writing instrument. Cross-validation on the presented dataset have shown promising performance of the used methods.

We stress that the Pen Ink Library is designed in a way that allows uploading MSP-Vis-NIR spectra in a standardized format, which is independent of the specific measuring device. Apart from the described use cases, the database can also be helpful when one is concerned about determining the uniqueness of a specific ink in the local population of writing instruments.

A. Estimation of $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{\mu}$

For estimation of the hyper-parameters $\boldsymbol{U}$, $\boldsymbol{C}$, and $\boldsymbol{\mu}$ from an unbalanced dataset of feature vectors, we developed the following EM algorithm. Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K$ denote a training dataset of feature matrices, where each feature matrix $\boldsymbol{Y}_k = (\boldsymbol{y}_{k,1} \cdots \boldsymbol{y}_{k,N_k})$ represents the k-th pen, whose parameter vector is denoted as $\boldsymbol{\theta}_k$. Our dataset is unbalanced, thus the number of feature vectors N_k varies with $k = 1, 2, \ldots, K$.

The EM algorithm is an iterative algorithm that attempts to find hyper-parameters which maximize the marginal likelihood [21]

$$p(\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_K; \boldsymbol{U}, \boldsymbol{C}, \boldsymbol{\mu}) = \prod_{k=1}^{K} \int p(\boldsymbol{Y}_k | \boldsymbol{\theta}_k; \boldsymbol{U}) p(\boldsymbol{\theta}_k; \boldsymbol{\mu}, \boldsymbol{C}) \mathrm{d}\boldsymbol{\theta}_k. \tag{10}$$

Here the dependence of the pdfs on the hyper-parameters is indicated after the semi-colon. One iteration of our EM algorithm consists of the following two steps.

E-step: For all $k = 1, 2, \ldots, K$, compute the mean and covariance matrix of the posterior pdf $p(\boldsymbol{\theta}_k | \boldsymbol{Y}_k; \hat{\boldsymbol{U}}, \hat{\boldsymbol{C}}, \hat{\boldsymbol{\mu}}) = \mathcal{N}(\boldsymbol{\theta}_k; \hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\Sigma}}_k)$, given the latest estimates $\hat{\boldsymbol{U}}$, $\hat{\boldsymbol{C}}$, and $\hat{\boldsymbol{\mu}}$ of the hyper-parameters. That is [22, Sec. 9.9]

$$\hat{\boldsymbol{\theta}}_k = (N_k \hat{\boldsymbol{U}}^{-1} + \hat{\boldsymbol{C}}^{-1})^{-1} (N_k \hat{\boldsymbol{U}}^{-1} \bar{\boldsymbol{y}}_k + \hat{\boldsymbol{C}}^{-1} \hat{\boldsymbol{\mu}}), \tag{11}$$

$$\hat{\boldsymbol{\Sigma}}_k = (N_k \hat{\boldsymbol{U}}^{-1} + \hat{\boldsymbol{C}}^{-1})^{-1}, \tag{12}$$

where $\bar{\boldsymbol{y}}_k \triangleq \frac{1}{N_k} \sum_{n=1}^{N_k} \boldsymbol{y}_{k,n}$.

M-step: Update the estimates of the hyper-parameters via

$$\hat{\boldsymbol{\mu}}^\star = \frac{1}{K}\sum_{k=1}^{K}\hat{\boldsymbol{\theta}}_k, \tag{13}$$

$$\hat{\boldsymbol{C}}^\star = \frac{1}{K}\sum_{k=1}^{K}\left(\hat{\boldsymbol{\Sigma}}_k + (\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\mu}}^\star)(\hat{\boldsymbol{\theta}}_k - \hat{\boldsymbol{\mu}}^\star)^{\mathrm{T}}\right), \tag{14}$$

$$\hat{\boldsymbol{U}}^\star = \frac{1}{N}\sum_{k=1}^{K}\left(N_k\hat{\boldsymbol{\Sigma}}_k + \sum_{n=1}^{N_k}(\boldsymbol{y}_{k,n} - \hat{\boldsymbol{\theta}}_k)(\boldsymbol{y}_{k,n} - \hat{\boldsymbol{\theta}}_k)^{\mathrm{T}}\right), \tag{15}$$

where $N = \sum_{k=1}^{K} N_k$. Convergence of the algorithm can be detected by watching the change of the hyper-parameter estimates between consecutive iterations and stopping as soon as the change is small. After convergence, we take $\hat{\boldsymbol{U}}^\star$, $\hat{\boldsymbol{C}}^\star$, and $\hat{\boldsymbol{\mu}}^\star$ as the final estimates of the hyper-parameters. In our experiments, the algorithm typically converged in less than 20 iterations.

References

[1] V. Causin, R. Casamassima, C. Marega, P. Maida, S. Schiavone, A. Marigo, and A. Villari. The discrimination potential of ultraviolet-visible spectrophotometry, thin layer chromatography, and Fourier transform infrared spectroscopy for the forensic analysis of black and blue ballpoint inks. *J. Forensic Sci.*, 53(6):1468–1473, Nov. 2008. doi: 10.1111/j.1556-4029.2008.00867.x.

[2] C. Neumann, R. Ramotowski, and T. Genessay. Forensic examination of ink by high-performance thin layer chromatography—the united states secret service digital ink library. *J. Chromatogr. A*, 1218(19):2793–2811, May 2011. doi: 10.1016/j.chroma.2010.12.070.

[3] C. Roux, M. Novotny, I. Evans, and C. Lennard. A study to investigate the evidential value of blue and black ballpoint pen inks in Australia. *Forensic Sci. Int.*, 101(3):167–176, May 1999. doi: 10.1016/S0379-0738(99)00021-3.

[4] G. Germinario, S. Garrappa, V. D'Ambrosio, I. D. van der Werf, and L. Sabbatini. Chemical composition of felt-tip pen inks. *Anal. Bioanal. Chem.*, 410:1079–1094, Nov. 2017. doi: 10.1007/s00216-017-0687-x.

[5] J. H. Bügler, H. Buchner, and A. Dallmayer. Age determination of ballpoint pen ink by thermal desorption and gas chromatography–mass spectrometry. *J. Forensic Sci.*, 53(4), July 2008. doi: 10.1111/j.1556-4029.2008.00745.x.

[6] T. A. Leal, C. Ferreira, A. Quintas, and A. Bernardo. Dating inks on paper through chromatographic analysis of volatile compounds: A mini-review. *Ann. Med.*, 51(1), May 2019. doi: 10.1080/07853890.2018.1562750.

[7] R. P. Gorziza, C. M. B. de Carvalho, T. Korndörfer, R. S. Ortiz, M. González, L. B. Leal, T. Trejos, and R. P. Limberger. Blue and black ballpoint pen inks: A systematic review for ink characterization and dating analysis. *Braz. J. Forensic Sci. Med. Law Bioeth.*, 8(3):113–138, May 2019. doi: 10.17063/bjfs8(3)y2019113.

[8] J. Wilson, G. Laporte, and A. Cantu. Differentiation of black gel inks using optical and chemical techniques. *J. Forensic Sci.*, 49(2):364–70, Feb. 2004. doi: 10.1520/JFS2003262.

[9] L. Gál, M. Oravec, P. Gemeiner, and M. Čeppan. Principal component analysis for the forensic discrimination of black inkjet inks based on the Vis–NIR fibre optics reflection spectra. *Forensic Sci. Int.*, 257:285–292, Dec. 2015. doi: 10.1016/j.forsciint.2015.09.011.

[10] V. A. G. da Silva, M. Talhavini, J. J. Zacca, B. R. Trindade, and J. W. B. Braga. Discrimination of black pen inks on writing documents using visible reflectance spectroscopy and PLS-DA. *J. Braz. Chem. Soc.*, 25(9):1552–1564, Sept. 2014. doi: 10.5935/0103-5053.20140140.

[11] R. Kumar and V. Sharma. A novel combined approach of diffuse reflectance UV–Vis-NIR spectroscopy and multivariate analysis for non-destructive examination of blue ballpoint pen inks in forensic application. *Spectrochim. Acta Part A Mol. Biomol. Spectrosc.*, 175:67–75, Mar. 2017. doi: 10.1016/j.saa.2016.12.008.

[12] A. C. de Almeida Assis, F. I. Romano Inácio, J. S. S. de Melo, and C. Farinha. Writing instruments inks: Microspectrophotometry forensic analysis and characterization. *Eur. Police Sci. Res. Bull.*, (16):187–207, 2017.

[13] A. Martyna, D. Lucy, G. Zadora, B. M. Trzcinska, D. Ramos, and A. Parczewski. The evidential value of microspectrophotometry measurements made for pen inks. *Anal. Methods*, 5(23):6788–6795, Sept. 2013. doi: 10.1039/C3AY41622D.

[14] L. A. Mohammed and L. Cunningham. Chapter 7 - Effects of writing instruments and constraints on signatures. In L. A. Mohammed, editor, *Forensic Examination of Signatures*, pages 97–117. Academic Press, London, UK, 2019. doi: 10.1016/B978-0-12-813029-2.00007-4.

[15] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proc. CVPR*, pages 770–778, Las Vegas, NV, USA, June 2016. doi: 10.1109/CVPR.2016.90.

[16] C. Aitken, F. Taroni, and S. Bozza. *Statistics and the Evaluation of Evidence for Forensic Scientists*. Wiley, Hoboken, NJ, USA, third edition, 2021. doi: 10.1002/0470011238.

[17] C. E. H. Berger. Inference of identity of source using univariate and bivariate methods. *Sci. Justice*, 49(4), Dec. 2009. doi: 10.1016/j.scijus.2009.03.003.

[18] C. Aitken and D. Lucy. Evaluation of trace evidence in the form of multivariate data. *Appl. Statistics*, 53(1), June 2004. doi: 10.1046/j.0035-9254.2003.05271.x.

[19] J. Schanda. *Colorimetry: Understanding the CIE System*. Wiley, Hoboken, NJ, USA, 2007. doi: 10.1002/9780470175637.

[20] T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer, New York, NY, USA, second edition, 2009. doi: 10.1007/978-0-387-84858-7.

[21] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc. Ser. B*, 39(1):1–37, Sept. 1977. doi: 10.1111/j.2517-6161.1977.tb01600.x.

[22] M. H. DeGroot. *Optimal Statistical Decisions*. Wiley, Hoboken, NJ, USA, 2004. doi: 10.1002/0471729000.