



## Label-Free Blood Typing by Raman Spectroscopy and Artificial Intelligence

Jensen, Emil Alstrup; Serhatlioglu, Murat; Uyanik, Cihan; Hansen, Anne Todsén; Puthusserypady, Sadasivan; Dziegiel, Morten Hanefeld; Kristensen, Anders

*Published in:*  
Advanced Materials Technologies

*Link to article, DOI:*  
[10.1002/admt.202301462](https://doi.org/10.1002/admt.202301462)

*Publication date:*  
2024

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Jensen, E. A., Serhatlioglu, M., Uyanik, C., Hansen, A. T., Puthusserypady, S., Dziegiel, M. H., & Kristensen, A. (2024). Label-Free Blood Typing by Raman Spectroscopy and Artificial Intelligence. *Advanced Materials Technologies*, 9(2), Article 2301462. <https://doi.org/10.1002/admt.202301462>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

# Label-Free Blood Typing by Raman Spectroscopy and Artificial Intelligence

Emil Alstrup Jensen, Murat Serhatlioglu, Cihan Uyanik, Anne Todsén Hansen, Sadasivan Puthusserypady, Morten Hanefeld Dziegiel,\* and Anders Kristensen\*

Label-free blood typing by Raman spectroscopy (RS) is demonstrated by training an artificial intelligence (AI) model on 271 blood typed donor whole blood samples. A fused silica micro-capillary flow cell enables fast generation of a large dataset of Raman spectra of individual donors. A combination of resampling methods, machine learning and deep learning is used to classify the ABO blood group, 27 erythrocyte antigens, 4 platelet antigens, regular anti-B titers of blood group A donors, regular anti-A,-B titers of blood group O donors, and ABH-secretor status, from a single Raman spectrum. The average area under the curve value of the ABO classification is  $0.91 \pm 0.03$  and  $0.72 \pm 0.09$ , respectively, for the remaining traits. The classification performance of all parameters is discussed in the context of dataset balance and antigen concentration. Post-hoc scalability analysis of the models shows the potential of RS and AI for future applications in transfusion medicine and blood banking.

extremely subtle and extensive information about the composition of a sample.<sup>[1]</sup> In recent years, numerous applications of both ML<sup>[2]</sup> and DL<sup>[3]</sup> frameworks have been suggested, to translate the complexity of RS to a clinical setting. RS has been used for a wide range of applications in bioanalytics and blood analysis<sup>[4,5]</sup> including detection of glycated hemoglobin for diagnosis of diabetes,<sup>[6]</sup> in vivo measurements of oxygen saturation of hemoglobin,<sup>[7,8]</sup> and point of care detection of hemolysis.<sup>[9–12]</sup> RS has shown a potential to recognize distinct features of the different erythrocyte blood group antigens and thus distinguish between ABO blood group antigens, but has so far relied on comprehensive sample preparation steps applied to isolate specific substances

## 1. Introduction

The non-labeling nature of Raman spectroscopy (RS) makes it highly attractive for phenotyping in clinical medicine. RS provides a spectral fingerprint specific to the biological sample of interest. Supported by artificial intelligence (AI), (machine learning (ML) and deep learning (DL)), it is possible to extract

or components of the blood, thus limiting the throughput and the resulting dataset sizes. A laser tweezers Raman spectroscopy (LTRS) system has been proposed to probe single trapped erythrocytes,<sup>[13]</sup> and surface-enhanced Raman scattering (SERS) spectroscopy on purified globulins from blood plasma have been used to discriminate ABO blood types.<sup>[14,15]</sup> Raman analysis restricted to the individual substances has the advantage of reducing the simultaneous scattering and absorption factors that otherwise need to be considered, and produce features in a Raman spectrum, which are otherwise hidden by the strongly scattering hemoglobin in erythrocytes.<sup>[4,16,17]</sup> However, in many point-of-care applications and particularly in transfusion medicine, time is a decisive factor, and RS must be applied directly on whole blood to avoid time and labor consuming preparatory steps. Comprehensive work on RS on blood within flow-through tubes has shown the potential to acquire Raman spectra with minimal sample preparation in a fast and non-invasive way.<sup>[18,19]</sup> Finally, some previous studies of Raman blood analysis suffer from a low number of samples, subsequently limiting the access to large datasets for ML and DL analysis thus making the translation to clinical use difficult.<sup>[20]</sup>

ABO and RhD blood group determination for donors and patients is pivotal for safe transfusion medicine.<sup>[21]</sup> Determination of blood groups and subsequent matching is mandatory in transfusion medicine, and only RhD-negative erythrocytes are given to RhD-negative recipients. O RhD negative erythrocytes are universal in the sense that any recipient can receive them without the risk of reactivity with anti-A and anti-B antibodies and without

E. A. Jensen, A. T. Hansen, M. H. Dziegiel  
Department of Clinical Immunology  
Copenhagen University Hospital (Rigshospitalet)  
Blegdamsvej 9, 2100 København Ø, Denmark  
E-mail: Morten.Dziegiel@regionh.dk

E. A. Jensen, M. Serhatlioglu, C. Uyanik, S. Puthusserypady, A. Kristensen  
Department of Health Technology  
Technical University of Denmark  
Ørstedts Plads, Building 345C, 2800 Kongens Lyngby, Denmark  
E-mail: akri@dtu.dk

M. H. Dziegiel  
Department of Clinical Medicine  
University of Copenhagen  
Blegdamsvej 3, 2200 København N, Denmark

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/admt.202301462>

© 2023 The Authors. Advanced Materials Technologies published by Wiley-VCH GmbH. This is an open access article under the terms of the [Creative Commons Attribution](#) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

DOI: 10.1002/admt.202301462

**Table 1.** State-of-the-art blood typing methods..

Method	Techniques	Consumables	Sample Preparation	Scope	Time	Non-invasive
Manual agglutination [29,30]	Slide testing, tube testing	Antibody and antigen specific reagents	Whole blood, centrifugation	ABO, RhD	5–20 min	No
Automated Agglutination [29,30]	Microplate, microcolumn gel	Antibody and antigen specific reagents	Whole blood, centrifugation	ABO, Rh, Duffy, Kell Kidd, MNS	6–45 min	No
Paper based microfluidics (Agglutination) [29,31–33]	Dye-assisted, wax printing and dipping	Antibody treated paper, antigen specific reagents, indicator dyes (e.g., BCG)	Whole blood (forward typing), dilution	ABO, Rh, Duffy, Kell, P1, Lewis, Kidd, MNS	1–45 min	No
Genotyping [26,27]	PCR, sequencing	PCR consumables	PCR sample preparation, dilution	All	Hours	No
Surface plasmon resonance [29,34–36]	SPRi	Antibody specific reagents, SPR sensor chips	Washing, centrifugation	ABO, Rh, Duffy	10–15 min	No
Raman-based microfluidics	RS	None	None	All	1 min	Yes

the risk of inducing irregular antibodies to the RhD antigen. This makes the O RhD negative erythrocytes vital in emergency situations where the recipient blood group is unknown. However, only  $\approx 4\%$  of the population are O RhD negative. Therefore, blood banks strive to obtain fast determination of patients' blood groups and use donors with identical blood groups.<sup>[22]</sup> This motivates the development of fast and comprehensive analytical technology.

The immune system always produces antibodies against the ABO antigens that are not present on the surface of the RBCs of the person, hence the term regular anti-A and anti-B antibodies—and the interest in avoiding universal O type donors with excessive anti-A and anti-B antibody. Extended phenotyping of additional blood group systems is done as a supplement to the basic determination of blood groups for ABO and RhD. This is particularly important for patients who already have developed alloantibodies, potentially causing a hemolytic transfusion reaction (HTR).<sup>[23–25]</sup> A comparison to state-of-the-art blood typing methods is presented in **Table 1**. Most other methods are well established and tested with clinically relevant accuracies with both low cost and fast options. However, most rely on antibody and antigen specific reagents and are susceptible to the expression levels of certain antigens or limited by weak agglutination. Alternatively, PCR based genotyping has become state-of-the-art in extended blood type matching<sup>[26,27]</sup> but remains time-consuming and expensive on a larger scale due to the requirement of highly trained technicians and multiple reference laboratories.<sup>[28]</sup>

In this work, we address the analytical challenges of the consumable-heavy serological methods as well as the time-consuming and expensive genotyping methods, using a flow cell protocol to acquire RS from 271 whole blood samples with no pre-analytical preparative steps. Based on a single Raman spectrum for each sample measured in 1 min, we aim to simultaneously map out 35 blood group traits and determine titers of regular blood group antibodies.

Notably, a trait is not a direct measurement of a low concentration antigen, but rather a characteristic of the donor that could encompass several distinct components in the blood. The identity and molar concentration of the entire set of informative analytes are not clear, but it is assumed that the entire collection of informative molecular moieties is far more abundant than the molar concentration of the single product representing the molec-

ular end-point of a trait. A similar approach of using a multi-parameter association has recently been reported for proteomics and blood groups ABO and Rh.<sup>[37]</sup> A strict association between molar concentrations of the blood group antigens and the performance of RS is not to be expected. Consequently, the complex correlation between reference blood traits and RS is analyzed by multivariate statistics, ML and DL, and the challenge of detecting a rare trait is addressed by resampling methods such as random undersampling, adaptive synthetic (ADASYN)<sup>[38]</sup> oversampling and synthetic minority (SMOTE)<sup>[39]</sup> oversampling, as well as using ensemble learning.<sup>[40,41]</sup> The observed scaling of accuracy with size of training set supports the principal feasibility to achieve clinically relevant accuracy on determining blood traits by RS and AI trained on 30–60 000 donor samples—with a perspective to radically improve transfusion medicine via fast and easy access to all clinically relevant traits for donors and patients.

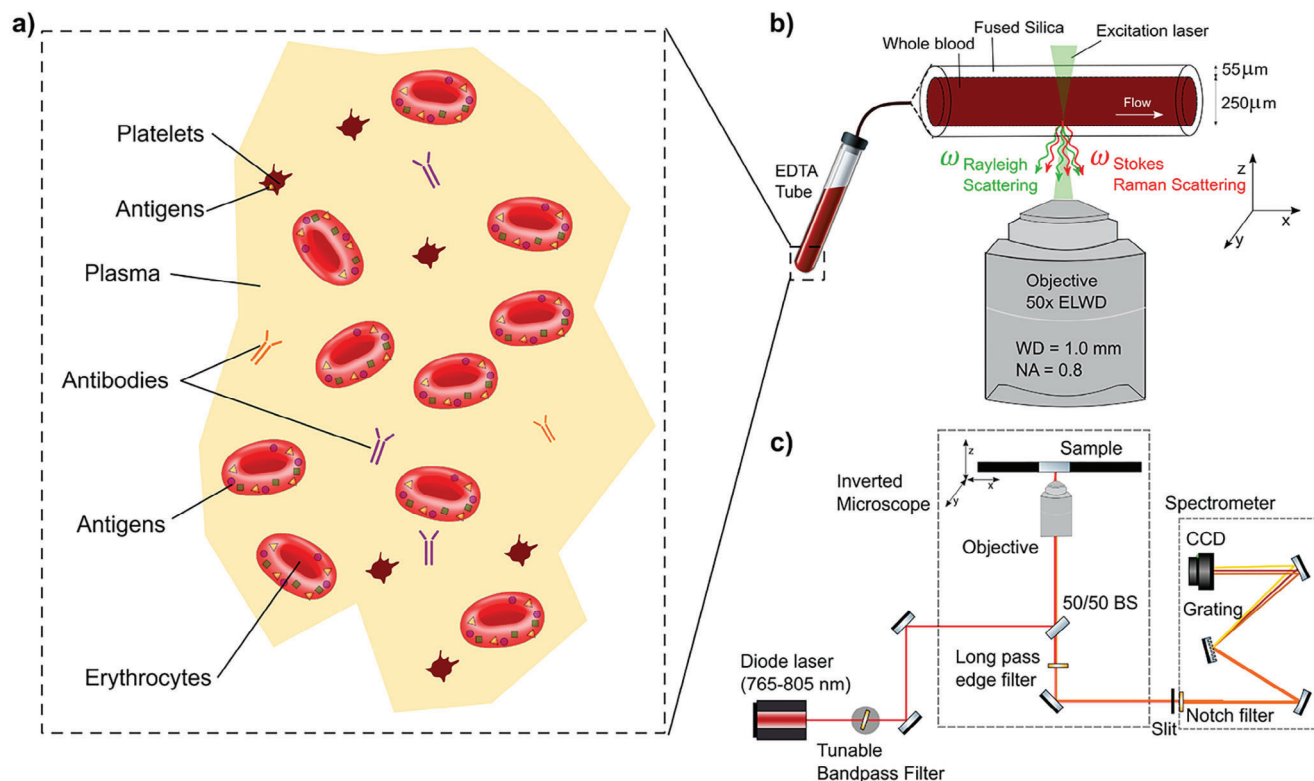
## 2. Experimental Section

### 2.1. Ethical Approval

Blood was collected from voluntary blood donors after obtaining informed written consent for use of blood as normal material. Samples were anonymized and accompanied by reference data (ABO, phenotype predicted from genotype and hematology reference data). According to legislation, this use of anonymized donor samples as normal material does not require approval by an ethical committee.

### 2.2. Microfluidics

Whole blood samples in EDTA tubes were collected from donors at the blood bank in Copenhagen University Hospital and stored at 4 °C for no longer than 60 h before Raman measurements were done. The main source of spectral change over time can be attributed to oxygenation of hemoglobin, which was visible in regions related to oxygenation and deoxygenation bands (1200–1230  $\text{cm}^{-1}$  and 1500–1660  $\text{cm}^{-1}$ ).<sup>[42]</sup> A short window from the time blood is collected to a Raman spectrum was acquired, limits



**Figure 1.** a) Antigens on the surface of erythrocytes, platelets, in plasma, and antibodies in plasma all determine various blood groups and other traits for a specific donor. b) A sample of whole blood is collected in EDTA with no pre-analytical sample preparation and drawn directly through a fused silica microcapillary by a syringe pump during the acquisition of the Raman spectrum. c) The RS setup is built on an inverted microscope: The 785 nm excitation laser source is focused to a spot inside the micro-capillary and Raman scattered light is collected by the microscope objective, filtered by notch and long pass edge filters and delivered to a spectrometer.

the contamination of oxygenation, although some variance between donors can be expected (see Figure S2, Supporting Information). The blood samples were drawn directly from the EDTA tubes into a fused silica microcapillary with an inner diameter of 250  $\mu\text{m}$  and outer diameter of 360  $\mu\text{m}$  (TSP250350, Polymicro Technologies LLC) by a syringe pump (Harvard Apparatus 11 Plus) to measure Raman spectra while flowing the sample as shown in Figure 1a–c. This flow cell configuration was chosen to avoid elaborate sample preparation steps and multiple disposable sample holder parts, to make the setup suitable for high throughput applications<sup>[43]</sup> in comparison with previously reported Raman-based methods, where sample preparation has been the rate-limiting step. In addition, the flow of the sample accommodated the fact that stationary blood samples were vulnerable to photodegradation and laser-induced denaturation of erythrocytes.<sup>[44]</sup> Photodegradation is in large part due to the strong absorption of hemoglobin at blue and green wavelengths, however, even at near-infrared excitation and low laser powers, photodamage of hemoglobin can occur. Heme aggregation due to thermally or laser induced denaturation creates spectral inhomogeneity in the measurements, which ultimately will disturb the blood type analysis.<sup>[45]</sup> A constant flow rate of 10  $\mu\text{L min}^{-1}$  was found to be optimal in terms of limiting the needed sample volume per measurement while avoiding photo-bleaching and -degradation in the sample (see Figure S1, Supporting Information). A maximum of 70  $\mu\text{L}$  of whole blood was used per Raman

spectrum. In-between sample exchange the capillary was cleaned by a bleach-based cleaning solution (Cat.no. BD 340345, BD Biosciences) followed by a buffer solution consisting of 1000 $\times$  0.5 mM EDTA mixed with phosphate buffered saline in a 1:1000 ratio to avoid any EDTA gradients and clot formation. The capillary was cleaned for  $\approx$ 5 min in-between sample exchange. Sample collection and measurements were carried out over several weeks using consistent sample loading and cleaning protocols.

### 2.3. Raman Spectroscopy

The flow cell was mounted on a motorized translational stage of an inverted microscope (Nikon-Ti) with a 50 $\times$  air objective (NA 0.8, WD 1 mm). A diagram of the optical setup is shown in Figure 1c. A tunable 765–805 nm diode laser (Toptica DLC DL-pro) was used for Raman excitation at 785 nm and sidebands were filtered using a tunable band-pass edge filter installed with a fixed angle with respect to the incident light. The laser source was then coupled into the inverted microscope in a backscattering configuration and focused by the objective to a small spot inside the fused silica capillary. A laser power of 25 mW at the excitation spot was used. A long-pass edge filter (785RS-25, Semrock) and a notch filter were inserted to suppress Rayleigh scattering such that only Raman scattered light entered the spectrometer. The microscope field of view was imaged by a Shamrock 303i imaging

spectrometer, with an integrated Newton 920 deep-cooled back-illuminated CCD camera (Andor Technology). A 300  $\mu\text{m}$  wide slit was positioned at the entrance of the spectrometer to suppress light away from the focus of the laser spot. A grating of 1200 lines  $\text{mm}^{-1}$  was applied to image the spectral distribution of focused Raman scattered light onto horizontal positions of the CCD camera, while vertical positions corresponded to scattered light at various positions at the entrance slit of the spectrometer. Positions that were spatially offset from the focus of the laser excitation spot collect Raman scattered light from sample volume that was not in focus with the focal plane of the microscope.<sup>[9]</sup> The capillary was aligned perpendicular to the optical axis of the microscope objective with the center of the capillary in the  $xy$ -plane. Raman signal from the fused silica capillary walls was suppressed by focusing the excitation laser along the  $z$  direction until fused silica Raman peaks disappear and signal from the blood samples was maximized. The position of the laser spot with respect to the sample flow was calibrated by conducting a  $z$ -scan, tracking fused silica bands (495, 606  $\text{cm}^{-1}$ ) and hemoglobin bands, such that a sample volume 10  $\mu\text{m}$  from the inner capillary wall can be probed consistently (see Figure S3, Supporting Information). The short penetration depth of the laser in the whole blood sample forces the measurement position to be relatively close to the capillary wall, where a cell-free region formed in close proximity to the capillary walls,<sup>[46,47]</sup> creating a narrow region without fused silica background. A  $z$ -scan calibration ensured that no unwanted variability in the spectra from glass signal, varying cell concentration and flow velocity was introduced. The flow of the sample enabled probing of a large sample volume and long integration time without damaging the blood cells. An integration time of 60 s was used in each measurement such that a relatively large sample volume can be analyzed, and sufficient SNR achieved ( $6.8 \pm 1.7$ ). For each measurement the spectral range of 393–1869  $\text{cm}^{-1}$  was used with the spectral dispersion varying from 0.63 to 0.49  $\text{cm}^{-1}$  per pixel. Wavenumber calibration was performed using the 520  $\text{cm}^{-1}$  vibration of a single crystalline silicon wafer sample.

## 2.4. Dataset

The Raman dataset was collected from whole blood samples in EDTA tubes from 69 type A, 69 type B, 70 type O, and 63 type AB donors. Three separate Raman measurements were conducted on each donor sample resulting in a total of 813 hyperspectral Raman images with 2663  $\times$  256 pixel values. The row with maximum intensity in the spectral region of interest was extracted for further pre-processing. The reference dataset consisted of serologically determined ABO and RhD blood groups for all the donors in the dataset and antibody parameters determined by serology for blood group A and O. Smaller subsets of the total donor cohort had an additional set of 51 blood group antigens predicted by genotyping. Only 38 of the traits were deemed suitable for classification purposes in terms of providing sufficient occurrences of both positive and negative donors for principal component analysis (PCA) and support vector machine (SVM) analysis. 27 of the traits were erythrocyte antigens (molecular end-point of trait), 4 were platelet antigens (molecular end-point of trait), and 2 were ABO antibody (Ab) titers, anti-B (aB), and anti-A,-B (aAB). The anatomical location of the trait, blood group system, frequency

of the antigen, and molar antigen concentration are outlined in Table 2.

All blood group antigens and haematological parameters were provided by the donor database at the Copenhagen University Hospital blood bank. The blood group antigen testing was carried out by PCR methods<sup>[27]</sup> and haematology parameters were determined by a commercial haematology analyzer (Sysmex).

## 2.5. Spectral Pre-Processing

To maximize the performance of the classification algorithms, a series of pre-processing steps were carried out on the Raman spectra. Fluorescence and measurement noise create a background drift in the spectra that needed to be accounted for, such that any spectral differences due to thermal fluctuations and other environmental factors are avoided. Background baselines for each spectrum were removed using an asymmetrically reweighted penalized least squares smoothing (arPLS) algorithm that effectively corrects noisy baselines while maintaining an accurate peak height estimation.<sup>[48]</sup> The arPLS algorithm can be tuned by a regularization parameter,  $\lambda$ , which was treated as a hyperparameter when training the classifiers. An outline of the importance of  $\lambda$  can be found in Figure S4, Supporting Information.

## 2.6. SVM Baselines and Training Details

SVM were used as the baseline classification method for all blood traits due to their efficient implementation and high performance on small datasets.<sup>[49]</sup> The four ABO blood groups were divided into 6 pairwise sub-classification problems, while the rest of the parameter determinations were reduced to a binary classification of either positive or negative. The input dataset consisted of 813 vectors, each with 2663 intensity variables. Analyzing data in a high-dimensional space was both computationally costly and often required large amounts of data to obtain reliable results. PCA was used to reduce the high dimensional (2663 intensity variables) input data to 15 principal components (PCs) before SVM classification. The number of PCs was treated as a hyperparameter when training the SVM classifiers. The SVM classifier used a radial basis function (RBF) kernel to project the input space into a higher dimensional space, such that the non-linearly separable classes can be distinguished. A kernel coefficient,  $\gamma$ , and a SVM regularization parameter,  $C$ , were then optimized along with the rest of the hyperparameters using fivefold cross-validation. The area under the receiver operating characteristics curves (AUC-ROC), balanced accuracy (BA), F1 score, precision, sensitivity/recall, and specificity were used to evaluate the performance of the models (see Section 2.9). BA was used as the loss function when optimizing the models due to the imbalanced nature of most of the parameters (see Section 2.7). The open source Python module Scikit-learn<sup>[50]</sup> was used for all SVM analysis.

## 2.7. Dataset Imbalance and Ensemble Learning

The donors were selected to have a balanced ABO blood group dataset and consequently the majority of the other parameters

**Table 2.** Overview of the different antigens, the cell types or substances in which they are found, the designation of the corresponding blood group system, number of antigens per cell, and concentration in  $\mu\text{m}.$ .

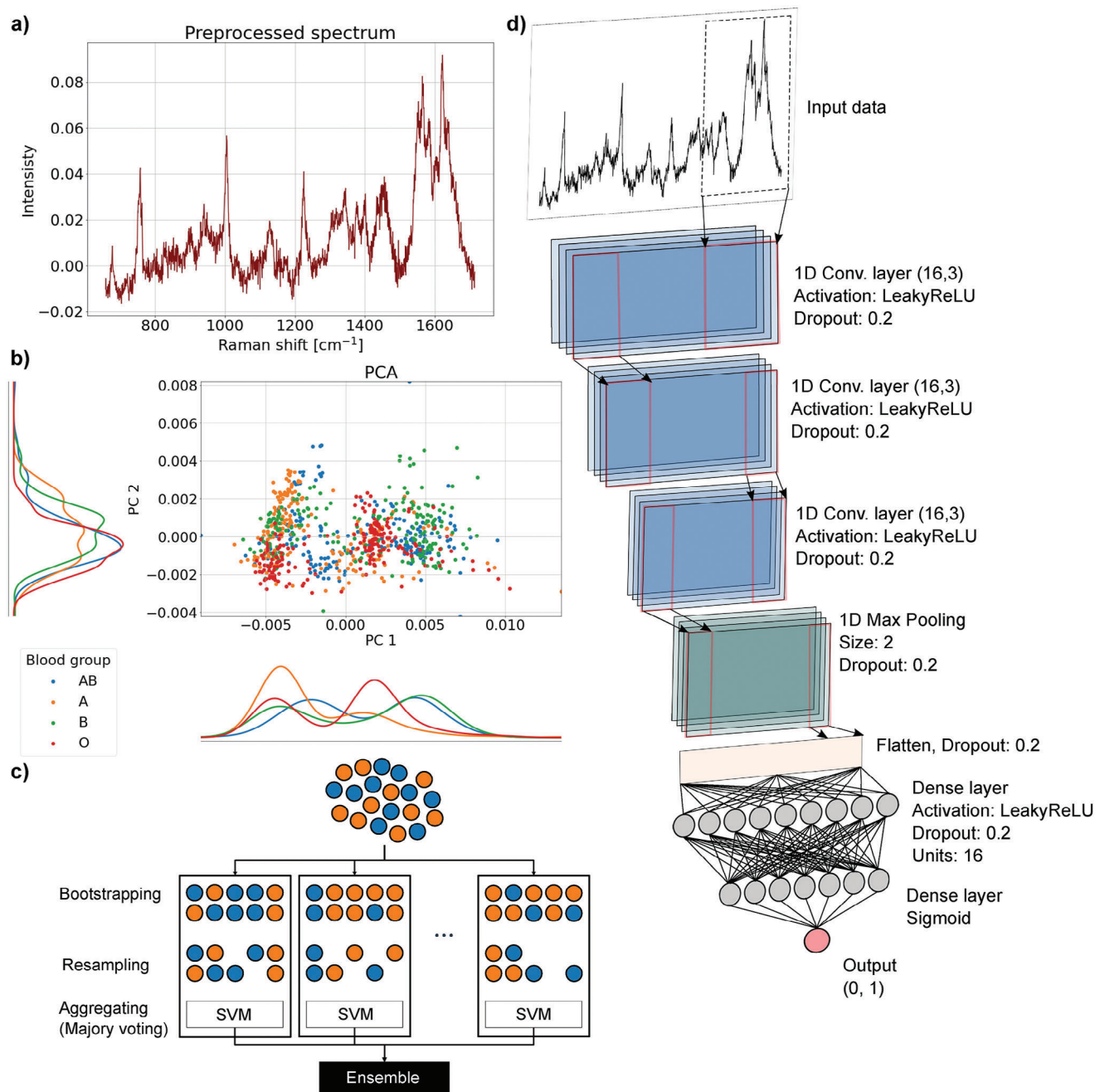
Antigen	Anatomical location	System	Number of antigens per cell	Concentration, $\mu\text{m}$
A1	Erythrocyte	ABO	$800 \cdot 10^3 - 1200 \cdot 10^3$	6.6–10.0
A2	Erythrocyte	ABO	$100 \cdot 10^3 - 400 \cdot 10^3$	0.83–3.3
A1B (A antigen)	Erythrocyte	ABO	$500 \cdot 10^3 - 900 \cdot 10^3$	4.2–7.5
A2B (A antigen)	Erythrocyte	ABO	$< 100 \cdot 10^3$	<0.83
B	Erythrocyte	ABO	$750 \cdot 10^3$	6.3
A1B (B antigen)	Erythrocyte	ABO	$430 \cdot 10^3$	3.6
ABH	Secretions, plasma	Secretor	-	-
C, c	Erythrocyte	Rh	$50 \cdot 10^3$	0.42
Cw	Erythrocyte	Rh	$14 \cdot 10^3 - 32 \cdot 10^3$	0.075–0.28
Dd	Erythrocyte	Rh	$9 \cdot 10^3 - 23 \cdot 10^3$	0.075–0.192
DD	Erythrocyte	Rh	$14 \cdot 10^3 - 33 \cdot 10^3$	0.12–0.28
E	Erythrocyte	Rh	$27.5 \cdot 10^3$	0.23
e	Erythrocyte	Rh	$25 \cdot 10^3$	0.21
Co <sup>a</sup> , Co <sup>b</sup>	Erythrocyte	Colton	$120 \cdot 10^3 - 160 \cdot 10^3$	1.00–1.33
Kp <sup>a</sup> , Kp <sup>b</sup>	Erythrocyte	Kell	$3.5 \cdot 10^3 - 18 \cdot 10^3$	0.029–0.150
K, k	Erythrocyte	Kell	$3.5 \cdot 10^3 - 18 \cdot 10^3$	0.029–0.150
Kn <sup>a</sup> , Kn <sup>b</sup>	Erythrocyte	Knops	20–1500	0.00016–0.012
M, N	Erythrocyte	Glycophorin A	$1000 \cdot 10^3$	8.33
S, s	Erythrocyte	Glycophorin B	$170 \cdot 10^3 - 250 \cdot 10^3$	1.4–2.1
Fy <sup>a</sup> , Fy <sup>b</sup>	Erythrocyte	Duffy	$6 \cdot 10^3$	0.050
Jk <sup>a</sup> , Jk <sup>b</sup>	Erythrocyte	Kidd	$14 \cdot 10^3$	0.12
Do <sup>a</sup> , Do <sup>b</sup>	Erythrocyte	Dombrock	-	-
Di <sup>a</sup> , Di <sup>b</sup>	Erythrocyte	Diego	$1000 \cdot 10^3$	8.3
Le <sup>a</sup> , Le <sup>b</sup>	Erythrocyte, secretions, plasma	Lewis	-	-
Lu <sup>a</sup> , Lu <sup>b</sup>	Erythrocyte	Lutheran	$1.64 \cdot 10^3 - 4.07 \cdot 10^3$	0.0137–0.0339
Yt <sup>a</sup> , Yt <sup>b</sup>	Erythrocyte	Cartwright	$7 \cdot 10^3 - 10 \cdot 10^3$	0.0583–0.0833
HPA-1a,b	Platelet	HPA-1	$35 \cdot 10^3 - 52 \cdot 10^3$	0.0146–0.0217
HPA-5a,b	Platelet	HPA-5	$3 \cdot 10^3 - 5 \cdot 10^3$	0.00125–0.0021
HPA-15a,b	Platelet	HPA-15	$1 \cdot 10^3$	0.000417

have a highly imbalanced distribution of positive and negative instances as they approximately reflect the balance in the population. Most ML models trained on an imbalanced dataset were susceptible to producing models which were biased toward the majority class. SVMs, especially, have been shown to produce separating hyperplanes ignorant toward the minority class, generating more false negative predictions.<sup>[51]</sup> Different re-sampling methods were applied to balance the datasets before classification. Random under-sampling, random over-sampling, and synthetic over-sampling methods, SMOTE<sup>[39]</sup> and ADASYN<sup>[38]</sup> were applied on the training data in each cross-validation evaluation to solve the problem of class imbalance. Additionally, bagging ensembles of SVM models were used instead of a single estimator, to further improve the classification performance.<sup>[40]</sup> Each individual SVM was trained independently on subsets of the training dataset, and an aggregate prediction was determined by majority voting (see **Figure 2c**). The subsets were randomly chosen by a bootstrap technique such that a spectrum can be used repeatedly in the training of multiple models, and each subset was balanced by resampling before training.<sup>[41]</sup> The size of the ensemble (number of  $n$  estimators) was treated as a hyper-parameter when opti-

mizing the models. The open source Python module imbalanced-learn<sup>[52]</sup> was used for all imbalance and ensemble learning analysis.

## 2.8. CNN Architecture

When potentially increasing the dataset to thousands of donors and combining all traits in a unified classification task, using a single or ensembles of non-linear SVMs was practically infeasible.<sup>[53]</sup> CNNs were established as one of the most utilized DL methods for pattern recognition and image classification,<sup>[54]</sup> and various CNN architectures were proposed as an efficient and accurate classifier of spectroscopic data.<sup>[55]</sup> The simple 1D CNN model used in this study consisted of three convolutional layers and a pooling layer for feature extraction, followed by two fully connected dense layers for classification. A block diagram of the architecture can be seen in **Figure 2d**. To avoid overfitting, a dropout of 0.2 was used in between each block, meaning 20% of the nodes were left out randomly from one layer to the next. A single 1D max pooling layer was applied after the last convolutional



**Figure 2.** a) Each Raman spectrum is baseline corrected and normalized. b) The dimension of the input data is reduced to a few principal components (PCs) before the models are trained. c) The input data is then used to train an ensemble of SVM models where subsets are created with replacement from the total dataset and balanced by resampling. The final prediction is computed by majority voting. d) Alternatively the preprocessed Raman spectra are used directly as input data in a 1D convolutional neural network (CNN) model, with 3 convolutional layers, a max pooling layer and two fully connected dense layers (see Section 2.8 for details). The number of filters ( $F$ ) and the kernel size ( $K$ ) in each convolutional layer are denoted ( $F, K$ ).

layer, to reduce the dimension of the feature map before classification. The CNN architecture hyperparameters, including the number of filters and kernel size in each convolutional layer as well as the number of hidden units in fully connected dense layers, were optimized using the Hyperband tuning algorithm.<sup>[56]</sup> The open source Python modules Keras<sup>[57]</sup> and TensorFlow<sup>[58]</sup> were used for all DL analysis.

## 2.9. Performance Metrics

The metrics used to evaluate the classification performance of the traits are defined as follows

$$\text{Sensitivity} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (1)$$

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}}, \quad (2)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (3)$$

$$\text{BA} = \frac{1}{2} (\text{Sensitivity} + \text{Specificity}), \text{ and} \quad (4)$$

$$F1 = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (5)$$

Here TP, TN, FP, and FN are true positive, true negative, false positive, and false negative predictions, respectively. The AUC is the area under the curve defined by the true positive rate (TPR) and false positive rate (FPR) for arbitrary values of  $\theta$  (posterior probability between 0 and 1)

$$\text{TPR}(\theta) = \frac{\text{TP}(\theta)}{\text{TP}(\theta) + \text{FN}(\theta)}, \quad (6)$$

$$\text{FPR}(\theta) = \frac{\text{FP}(\theta)}{\text{FP}(\theta) + \text{TN}(\theta)}. \quad (7)$$

## 2.10. Overfitting and Data Leakage

In order to prevent overfitting and data leakage, pipeline frameworks in both Scikit-learn<sup>[50]</sup> and Imbalanced-learn<sup>[52]</sup> were utilized to make sure that feature selection were completely separated from the validation data, that is, the PCA was recomputed using only the training data for each run in the cross-validation scheme and then the validation data was projected into that subspace. Similarly, the resampling methods were implemented within each fold and not across all cross-validation folds. Using BA as the metric for model selection ensured that the models were not biased toward the majority class. Both training scores and cross-validation scores were computed in the scalability analysis to show that the performance can be generalized. Too large a discrepancy between training and cross-validation curves indicated overfitting on the training data. In that case the SVM regularization parameter,  $C$ , was decreased to limit model complexity by preventing the weights of the models from becoming too large.

## 3. Results and Discussion

### 3.1. ABO Blood Group Classification

As mentioned, most bands in the whole blood Raman spectra can be attributed to vibrations of the hemoglobin molecule. The peaks are identified and assigned to different vibrational modes in **Figure 3a**. The PC loadings in **Figure 3b** suggest that certain Raman shifts contribute significantly more than others to the most descriptive PCs. In **Figure 3c** the intensity distribution at 24 of the most interesting bands are plotted. Distributions of spectra belonging to each of the ABO blood groups are plotted separately to account for any differences between the groups. Distinct differences can be seen at the Pyrrole (Pyr) breathing modes ( $\nu_{15}$ ,  $\nu_6$ ,  $\nu_{46}$ ) as well as at the asymmetric ( $\nu_{44}$ ,  $\nu_{30}$ ) and symmetric ( $\nu_{41}$ ,  $\nu_4$ ,  $\nu_{12}$ ) Pyr half-ring modes, and the Pyr quarter-ring mode ( $\nu_{20}$ ). Similarly, differences can be seen at the Phe skeletal C–C mode ( $898 \text{ cm}^{-1}$ ), the deformation modes of amino acid side chains

( $\delta(\text{CH}_2/\text{CH}_3)$ ) and amide I. Other porphyrin stretching modes such as  $\nu(\text{C}_m\text{H})$  ( $\nu_{13}$ ,  $\nu_{42}$ ,  $\nu_{21}$ ) and  $\nu(\text{C}_c\text{C}_m)_{\text{asym}}$  ( $\nu_{10}$ ) seem to be ABO blood group dependent as well. Differences at vibrational modes specific to Porphyrin between blood groups are consistent with the existing study on single trapped erythrocytes.<sup>[13]</sup> Positive and negative distributions of the intensity values at the same 24 bands for the rest of the traits can be found in **Figure S5**, Supporting Information.

The ABO blood group classification is carried out without any resampling or ensemble methods as the dataset is already balanced by construction. The performance results of the SVM (RBF) model are plotted in **Figure 4** as ROC curves and confusion matrices using fivefold cross-validation repeated ten times, and all performance metrics are summarized in **Table 3**. The AUC values are  $0.92 \pm 0.03$ ,  $0.88 \pm 0.04$ ,  $0.93 \pm 0.02$ ,  $0.87 \pm 0.04$ ,  $0.90 \pm 0.03$ , and  $0.95 \pm 0.02$  for AB-A, AB-B, AB-O, A-B, A-O, and B-O discrimination, respectively, matching previously reported values.<sup>[13–15]</sup> Type AB, A and B seem to be more difficult to discriminate which could be explained by the fact that antigen A and antigen B differ only by having the sugar *N*-acetylgalactosamine and galactose attached at the terminus, respectively.

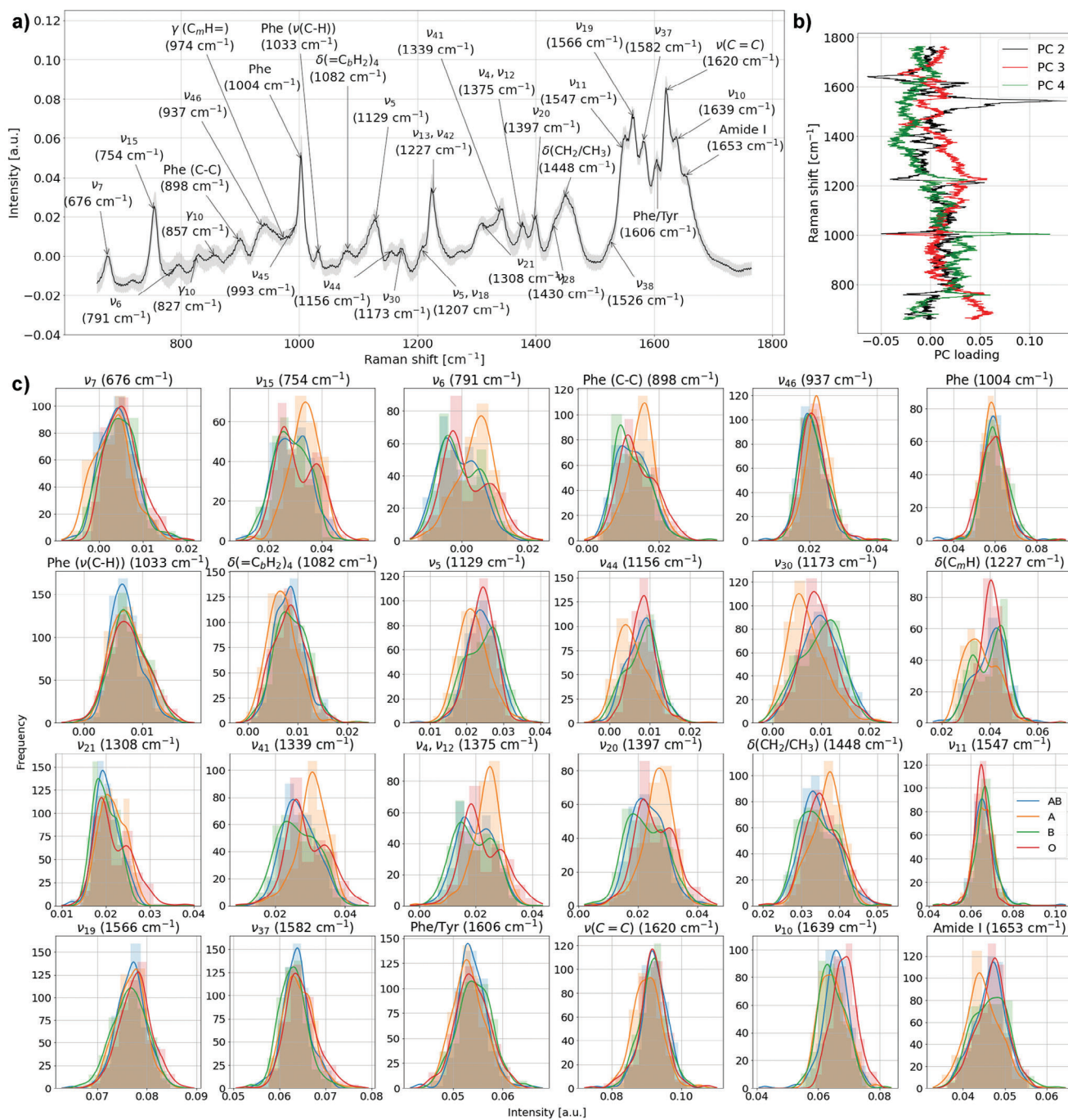
The determinations of ABO Ab titers are given as regular anti-B titers of blood group A donors  $\geq 10$  (positive) or  $< 10$  (negative), designated the aB Ab trait, and regular anti-A,-B titers of blood group O  $\geq 50$  (positive) or  $< 50$  (negative) designated the aAB Ab trait. As outlined in **Figure 5** and **Table 3** the aB trait is classified with an AUC of  $0.80 \pm 0.06$  on relatively limited reference dataset of size 243, showing great potential of extending RS and ML to antibody testing. The antibody titer laboratory tests are equivalent to a specific level of antibodies in the donor plasma, thus a quantitative determination in contrast to the rest of the antigen traits. Using RS to routinely measure quantitative levels of specific antibodies in donor plasma would be a highly attractive clinical tool for blood banks. Strong homology of anti-B in type A donors has previously been reported,<sup>[61]</sup> which supports the claim that there is a measurable difference in the Raman signal between donors below and above 10 anti-B (aB).

The aAB reference dataset is highly imbalanced (24% positives) resulting in an AUC of  $0.72 \pm 0.07$  but a poor F1 score of  $0.52 \pm 0.06$ , suggesting the applied resampling and ensemble methods have not been able to completely prevent underfitting. The Se (dominant) and se (recessive) traits are the ABH-secretor status of the donor, referring to a donor's ability to make ABO antigens in their plasma and secretions. Classification of Se shows promising results using random undersampling with an AUC of  $0.88 \pm 0.07$  despite significant class imbalance (see **Figure 5**).

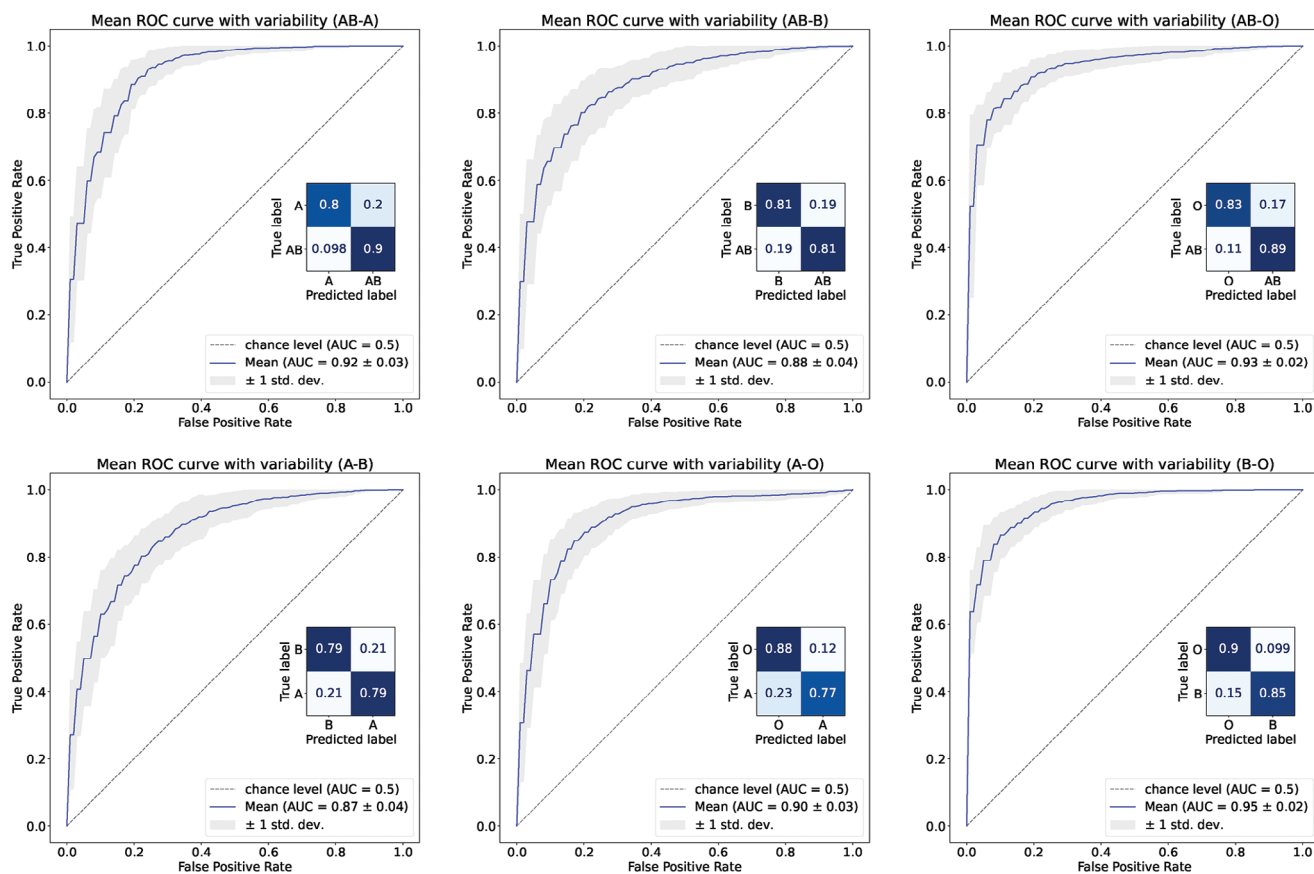
### 3.2. Erythrocyte and Platelet Antigen Traits

In order to maintain as large a dataset as possible, each erythrocyte and platelet antigen trait is classified as either negative or positive, rather than a homozygote/heterozygote discrimination, for example,  $\text{Jk}^a\text{Jk}^a$  and  $\text{Jk}^b\text{Jk}^b$ . Intuitively, homozygote discrimination would be easier due to a more distinct difference. However, the number of spectra available per class would be reduced significantly, and the nature of the RS ensures that both forms present in the heterozygote are clearly visible in the spectral data.





**Figure 3.** a) Peak assignments of blood components in a normalized average spectrum of all measurements. Assignments where  $\nu$  is followed by a number subscript are based on a labeling scheme developed for metalloporphyrins,<sup>[4,59,60]</sup> and constitute various vibrations in the porphyrin, the main structure of hemoglobin.  $\nu_{15}$ ,  $\nu_6$ ,  $\nu_{44}$ ,  $\nu_{30}$ ,  $\nu_{41}$ ,  $\nu_4$ ,  $\nu_{12}$ ,  $\nu_{20}$ , and  $\nu_{38}$  are all in-plane stretching modes belonging to Pyr, the outer ring structure of Porphyrin.  $\nu_7$  is in-plane deformation mode belonging to Pyr.  $\nu_{45}$ ,  $\nu_5$ ,  $\nu_{18}$ ,  $\nu_{13}$ ,  $\nu_{42}$ ,  $\nu_{21}$ ,  $\nu_{28}$ ,  $\nu_{11}$ ,  $\nu_{19}$ ,  $\nu_{37}$ , and  $\nu_{10}$  are other modes belonging to the Heme b structure in hemoglobin. More specifically in-plane stretching and deformation vibrations of C–H and C–C at various  $\alpha$ ,  $\beta$  and meso (m) positions in porphyrin. Protein assignments include phenylalanine (Phe), an amino acid which is present in both hemoglobin and other membrane proteins, deformation modes  $\delta$ (CH<sub>2</sub>/CH<sub>3</sub>) from amino acid side chains and amide I. Tyr refers to the amino acid Tyrosine. b) PC loadings for a few important PCs, describing the importance of each intensity variable. c) Distributions of the intensity values at 24 of the most significant peaks according to (a) and (b) for all Raman spectra.



**Figure 4.** ROC curves and confusion matrices for the ABO classification results. A ROC curve is produced for each of the five validation splits and the average is calculated. Similarly, the confusion matrices are averaged over five iterations of CV splits.

The distribution of positives and negatives for each trait is given in Table 3 as the fraction of positives in both the population<sup>[62]</sup> and the actual dataset for training. Four classification performance examples of erythrocyte antigens, S, Do<sup>b</sup>, Fy<sup>b</sup>, and Jk<sup>a</sup>, are presented in Figure 5. 50.1%, 87.1%, 78.9%, and 79.1% of donors in each dataset are genotyped positive, respectively. The highly imbalanced traits Do<sup>b</sup>, Fy<sup>b</sup>, and Jk<sup>a</sup> are all randomly undersampled to balance the datasets before classification, resulting in promising AUC values of  $0.75 \pm 0.1$ ,  $0.72 \pm 0.05$ , and  $0.74 \pm 0.05$ , and F1 scores of  $0.86 \pm 0.03$ ,  $0.75 \pm 0.06$ , and  $0.75 \pm 0.05$ , despite the significant reduction of training dataset size. The S antigen dataset is almost perfectly balanced from the start and is classified with an AUC of  $0.73 \pm 0.04$  without any resampling of the training dataset.

### 3.3. Dataset Balance and Antigen Concentration

Traits with very few positive observations (<10%) Kp<sup>a</sup>, Lu<sup>a</sup>, Cw, K, and Yt<sup>b</sup> are resampled during training using ADASYN and have promising AUC values  $0.93 \pm 0.08$ ,  $0.86 \pm 0.07$ ,  $0.83 \pm 0.07$ ,  $0.72 \pm 0.15$ , and  $0.67 \pm 0.19$ . However, all traits suffer from low precision and consequently low F1 scores (see Table 3) due to very few available positives in the test splits, and the general performance seems to vary significantly in each iteration of the repeated

cross-validation (see Figure 6). This can be attributed to a lack of positive values when testing the performance in each cross-validation split, and illustrates the challenge of constructing large validation datasets for the rare antigens. Antigen e, which is positive for 98.1% of the donors, performs extremely well on all metrics, and oversampling seems to sufficiently describe the minority (negative) class.

As expected, the ABO blood groups are easiest to classify as the datasets are both balanced and the molar antigen concentration among the highest (see Table 2). Other high concentration antigens, RhD, M, and N do, however, not perform as well with AUC values of  $0.7 \pm 0.05$ ,  $0.61 \pm 0.07$ , and  $0.58 \pm 0.06$ , suggesting that a balanced distribution of positives and negatives is more critical than the molar antigen concentration during classification. HPA-15 is the antigen with the lowest density of 1000 molecules per platelet which with a platelet density of  $250 \times 10^9$  per L corresponds to 417 pM. The two homozygous forms have comparable frequencies of 0.26 and 0.24 and the amino acid difference between the two antigens is a tyrosine to serine substitution. Despite the antigen location on the relative sparse platelets, a set of metrics above the average is observed (HPA-15a AUC =  $0.76 \pm 0.05$  and HPA-15b AUC =  $0.72 \pm 0.08$ ). Similarly, the more abundant platelet antigens HPA-1b (1.46–2.17 nM) and HPA-5b (1.25–2.1 nM) are classified with above average AUC values of  $0.68 \pm 0.05$  and  $0.78 \pm 0.08$ .

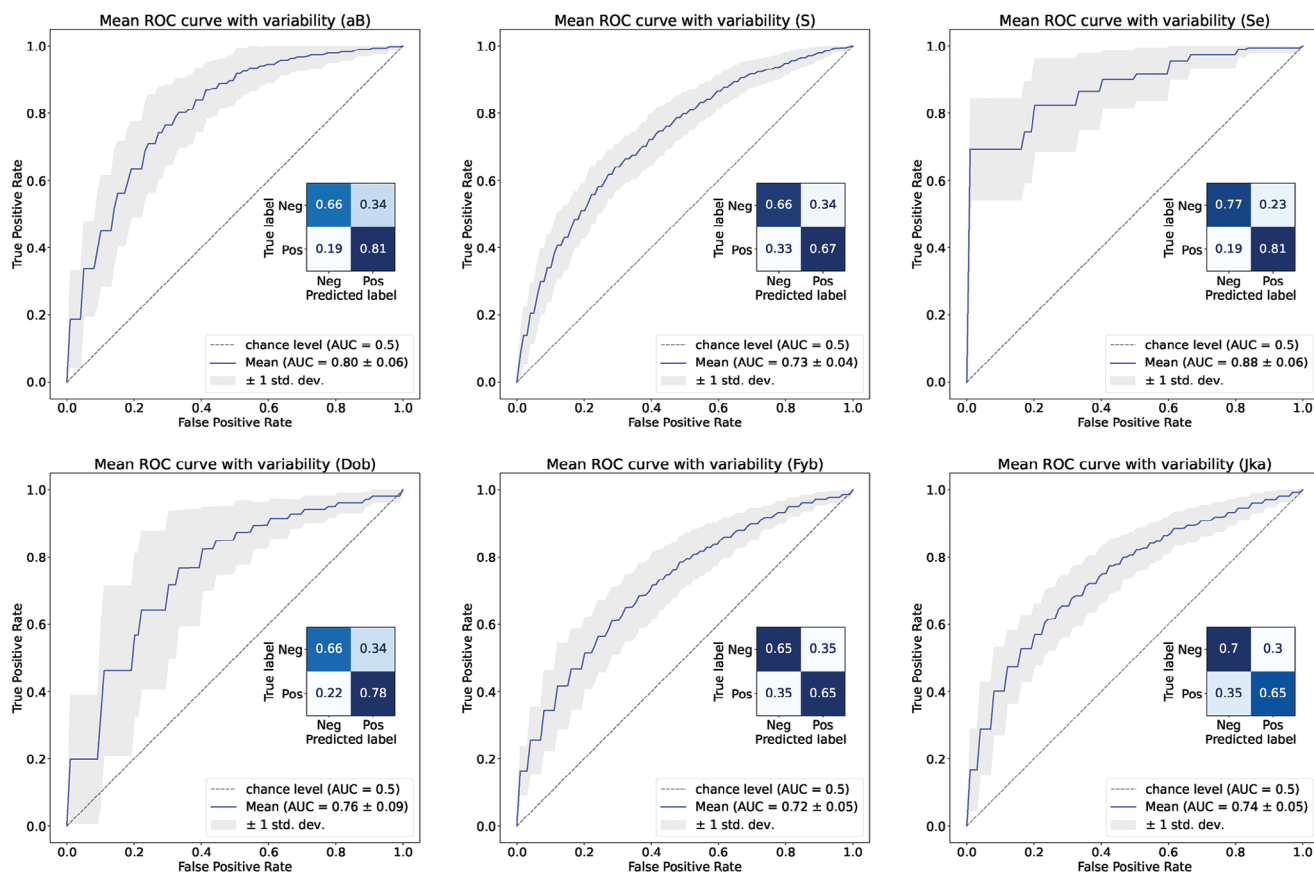
**Table 3.** Bagging ensembles with either a single or 50 estimators and a SVM based estimator. Each dataset is either not resampled, under- or oversampled. Hyperparameters are optimized by fivefold cross-validation using BA as the loss function. The cross-validation scores are presented in the table..

Antigen	Balance (Pop.)	Spectra	Balance (Data)	BA	AUC	F1	Recall	Precision	Specificity
AB-A	-	396	0.477	0.85±0.04	0.92±0.03	0.86±0.03	0.9±0.05	0.83±0.04	0.8±0.02
AB-B	-	396	0.477	0.81±0.04	0.88±0.04	0.82±0.04	0.81±0.05	0.83±0.06	0.81±0.03
AB-O	-	399	0.474	0.86±0.04	0.93±0.02	0.87±0.03	0.89±0.05	0.86±0.04	0.83±0.02
A-B	-	414	0.500	0.79±0.05	0.87±0.04	0.79±0.05	0.79±0.07	0.8±0.05	0.79±0.02
A-O	-	417	0.496	0.82±0.04	0.9±0.03	0.82±0.04	0.77±0.06	0.87±0.05	0.88±0.02
B-O	-	417	0.496	0.88±0.03	0.95±0.02	0.88±0.03	0.85±0.05	0.9±0.05	0.9±0.01
C	0.701	627	0.657	0.57±0.05	0.62±0.05	0.71±0.04	0.71±0.08	0.71±0.03	0.43±0.01
Cw	0.022	549	0.056	0.71±0.1	0.83±0.07	0.27±0.09	0.57±0.21	0.18±0.06	0.85±0.01
c	0.795	630	0.814	0.64±0.05	0.68±0.06	0.76±0.03	0.67±0.05	0.88±0.03	0.6±0.06
Co <sup>b</sup>	0.051	375	0.112	0.75±0.07	0.8±0.09	0.41±0.07	0.71±0.14	0.29±0.06	0.78±0.0
Do <sup>a</sup>	0.878	372	0.669	0.6±0.06	0.63±0.06	0.71±0.06	0.69±0.08	0.74±0.04	0.51±0.03
Do <sup>b</sup>	0.576	372	0.871	0.73±0.07	0.75±0.1	0.86±0.03	0.78±0.05	0.94±0.02	0.67±0.09
E	0.284	633	0.299	0.57±0.04	0.59±0.05	0.45±0.04	0.61±0.08	0.36±0.04	0.53±0.01
e	0.976	627	0.981	0.92±0.08	0.94±0.08	0.94±0.01	0.88±0.03	1.0±0.0	0.96±0.13
Fy <sup>a</sup>	0.666	621	0.667	0.58±0.04	0.6±0.05	0.7±0.04	0.68±0.06	0.72±0.03	0.47±0.01
Fy <sup>b</sup>	0.822	603	0.789	0.65±0.06	0.72±0.05	0.75±0.06	0.66±0.09	0.88±0.03	0.64±0.02
Jk <sup>a</sup>	0.764	618	0.791	0.68±0.04	0.74±0.05	0.75±0.05	0.65±0.07	0.89±0.02	0.7±0.02
Jk <sup>b</sup>	0.736	606	0.705	0.63±0.05	0.68±0.05	0.73±0.04	0.67±0.06	0.8±0.03	0.6±0.03
K	0.080	633	0.068	0.68±0.08	0.72±0.15	0.23±0.05	0.69±0.19	0.14±0.03	0.67±0.03
Kr <sup>b</sup>	0.049	177	0.102	0.57±0.11	0.72±0.12	0.21±0.19	0.24±0.22	0.21±0.22	0.89±0.01
Kp <sup>a</sup>	0.020	450	0.027	0.88±0.12	0.93±0.08	0.33±0.12	0.85±0.26	0.22±0.09	0.91±0.01
Le <sup>a</sup>	0.22	135	0.200	0.64±0.1	0.78±0.11	0.4±0.2	0.38±0.22	0.51±0.26	0.91±0.01
Le <sup>b</sup>	0.72	138	0.696	0.67±0.09	0.78±0.1	0.83±0.06	0.87±0.09	0.8±0.06	0.48±0.08
Lu <sup>a</sup>	0.063	453	0.060	0.76±0.11	0.86±0.07	0.3±0.09	0.71±0.24	0.2±0.06	0.82±0.02
M	0.78	579	0.807	0.58±0.06	0.61±0.07	0.75±0.06	0.68±0.08	0.84±0.03	0.48±0.04
N	0.72	441	0.728	0.57±0.06	0.58±0.06	0.6±0.07	0.49±0.08	0.79±0.05	0.64±0.03
P1	0.79	138	0.674	0.57±0.08	0.54±0.12	0.43±0.13	0.31±0.13	0.83±0.16	0.84±0.03
RhD	0.84	777	0.828	0.6±0.05	0.7±0.05	0.83±0.03	0.8±0.04	0.87±0.02	0.41±0.05
S	0.535	591	0.501	0.66±0.04	0.73±0.04	0.67±0.04	0.67±0.08	0.67±0.05	0.66±0.01
s	0.824	594	0.924	0.59±0.08	0.65±0.09	0.79±0.06	0.69±0.08	0.94±0.02	0.49±0.09
Se	0.849	177	0.847	0.78±0.1	0.88±0.07	0.87±0.04	0.81±0.07	0.95±0.04	0.75±0.14
se	0.627	177	0.661	0.7±0.08	0.73±0.08	0.7±0.08	0.59±0.09	0.86±0.08	0.81±0.06
Yt <sup>b</sup>	0.067	372	0.089	0.67±0.1	0.67±0.19	0.27±0.08	0.67±0.2	0.17±0.06	0.67±0.01
HPA-1b	0.231	387	0.326	0.62±0.06	0.68±0.05	0.52±0.07	0.64±0.12	0.44±0.06	0.61±0.0
HPA-5b	0.154	387	0.155	0.69±0.08	0.78±0.08	0.46±0.12	0.52±0.15	0.43±0.11	0.87±0.0
HPA-15a	0.759	378	0.762	0.69±0.06	0.76±0.05	0.78±0.04	0.71±0.06	0.87±0.04	0.66±0.07
HPA-15b	0.741	378	0.802	0.65±0.07	0.72±0.08	0.83±0.03	0.79±0.05	0.87±0.03	0.5±0.09
aAB	-	350	0.240	0.69±0.06	0.72±0.07	0.52±0.07	0.61±0.12	0.46±0.07	0.76±0.0
aB	-	243	0.556	0.73±0.06	0.8±0.06	0.77±0.06	0.81±0.09	0.75±0.05	0.65±0.03

For future construction of datasets containing the entire cohort of genotyped donors at Copenhagen University hospital ( $\approx 35\,000$  and growing by 10 000 each year), only 2% would be expected to be positive for the antigen Kp<sup>a</sup>, meaning less than 1000 positive spectra can be collected. Even fewer donors would be available for a specific phenotype such as homozygote Kp<sup>a</sup>Kp<sup>a</sup> (0.04%), so to create a sufficiently large validation dataset, oversampling and data augmentation of training data are necessary.

### 3.4. Scalability and Learning Curves

The sample size of the training dataset remains one of the most significant parameters of model selection and performance evaluation in the field of ML and DL.<sup>[63]</sup> The selected model must be able to capture the complexity of the problem, and its own parameters must be able to be estimated through the available observation points. The total number of model parameters are affected by various factors. Since the model is expected to



**Figure 5.** ROC curves and confusion matrices for classification results of selected traits.

represent the underlying data point distribution, the model could be extremely simple or complex, which decreases or increases the total number of model parameters, respectively. Generally, a large number of observation points improves the estimation quality of the model parameters and the generalization ability of the model. Low complexity ML algorithms perform well on small datasets but tend to plateau when the training data reaches a certain size, while the performance of DL algorithms is superior with larger amounts of data.<sup>[64]</sup>

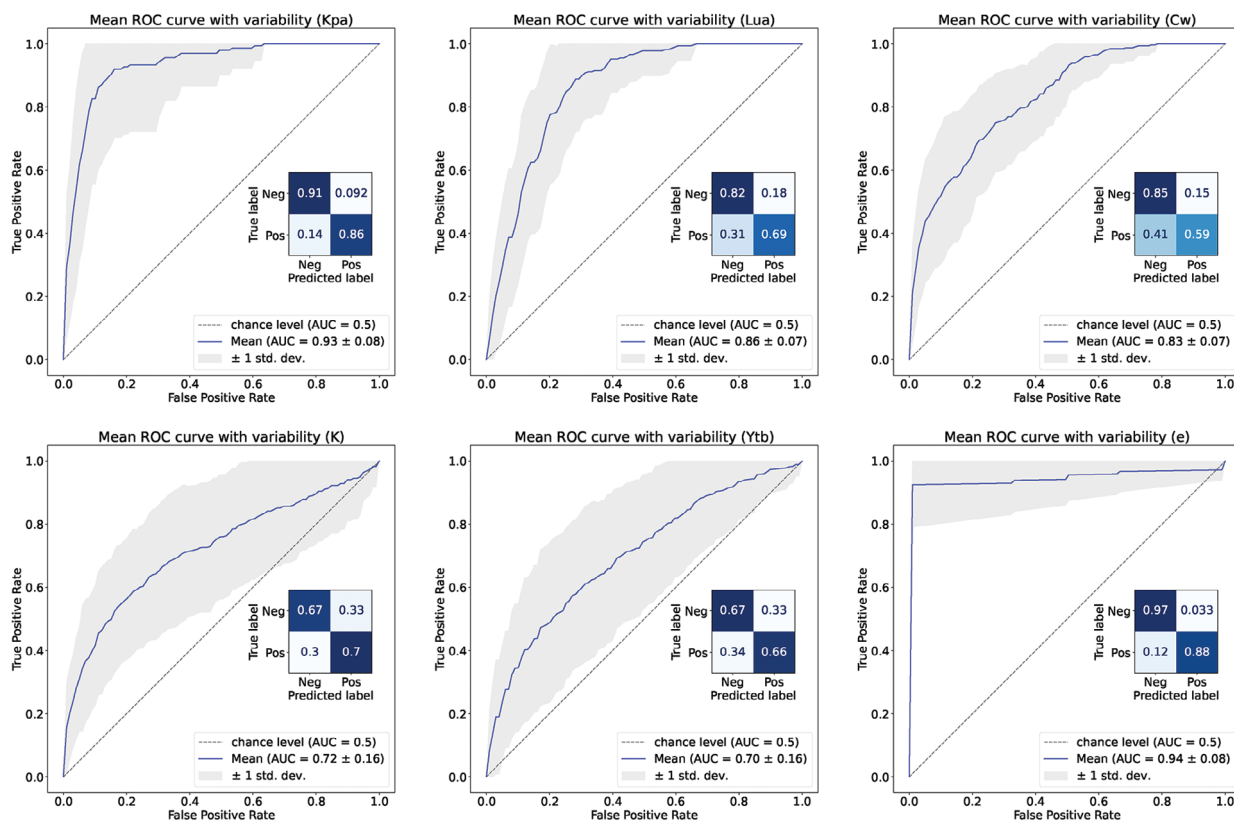
As reported in a study on the scaling of ML and DL models trained on the UKBiobank brain images dataset,<sup>[65]</sup> one of the largest biomedical datasets in the world, higher model complexity does not necessarily improve the performance of a classifier significantly, even when increasing the sample-size. It depends entirely on the presence of learnable non-linearity in a dataset, whether going from a linear SVM, to a shallow non-linear kernel SVM (RBF) or a deep nonlinear neural network (CNN) will improve the accuracy. If such non-linearity is accessible in a dataset, the performance of a SVM model with a non-linear kernel is expected to surpass a linear SVM model. This seems to be the case for our dataset as seen in Figure 8b where both the SVM (RBF) and CNN model clearly outperforms a linear SVM model.

To give an estimate of the potential performance of the models when increasing the dataset significantly (more than 30 000), a post-hoc sample-size analysis of the collected datasets based on cross-validation is conducted.<sup>[66]</sup> Fivefold cross-validated training

and test scores are generated using different training test sizes. Plotting the performance as function of the number of observations in the training subset then provides an approximation of the learning process of the model and a saturating inverse power law is fitted to each of the learning curves<sup>[67–69]</sup>

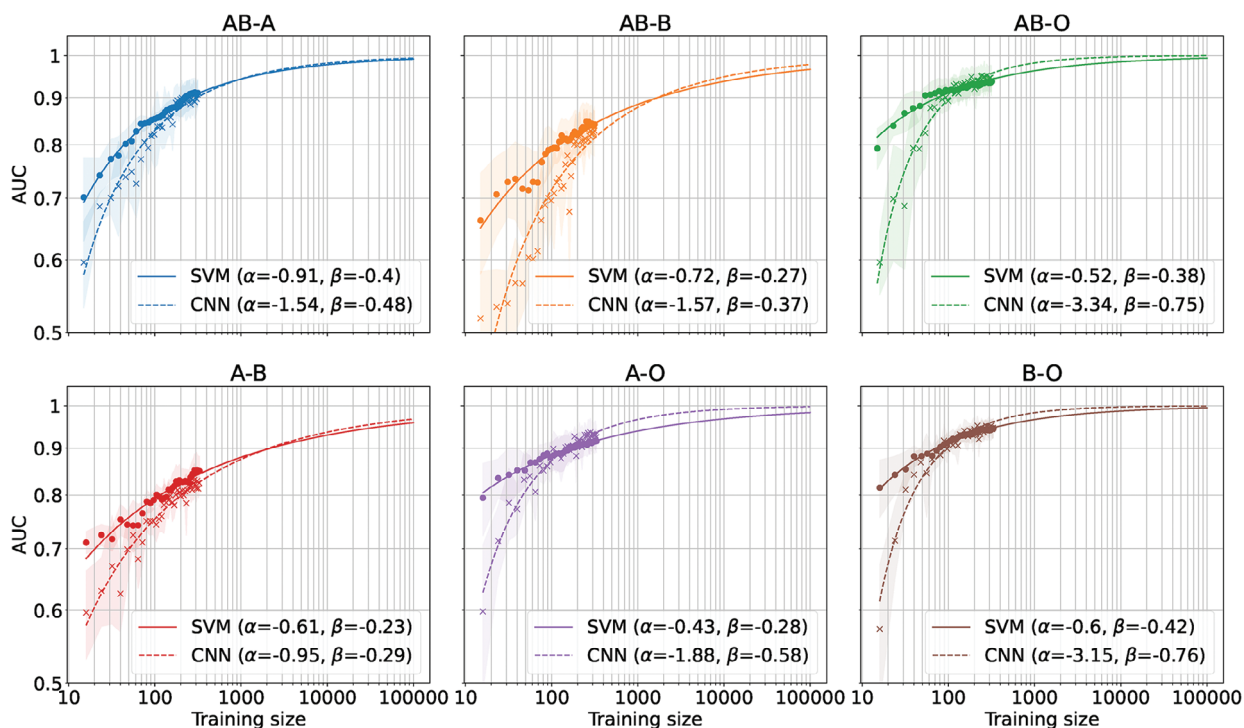
$$y = f(x; \alpha, \beta) = 1 + \alpha x^\beta \quad (8)$$

The learning rate,  $\alpha$ , and decay rate,  $\beta$  are estimated by non-linear least squares. In order to have reasonably balanced data, a scaling comparison of the SVM (RBF) and CNN model is carried out on each of the ABO discrimination datasets, and the results are presented in Figure 7. The learning curves suggest that at a certain training size the CNN model starts to outperform the SVM (RBF) model and approach clinically relevant AUC values faster. To reach a combined (three steps) 1 error in 500 000 determinations of donors, a single determination accuracy of 0.99 would be required.<sup>[70]</sup> The learning potential does not seem to have stagnated at the current dataset size, suggesting a larger dataset could improve the performance. The potential level of improvement should be investigated further by collection of a large dataset. Both hemoglobin and albumin are high abundance components in whole blood and dominate the Raman spectra, which could mask signal from other analytes providing valuable information to determine a trait. To fully exploit the sensitivity and specificity of RS, blood components could be

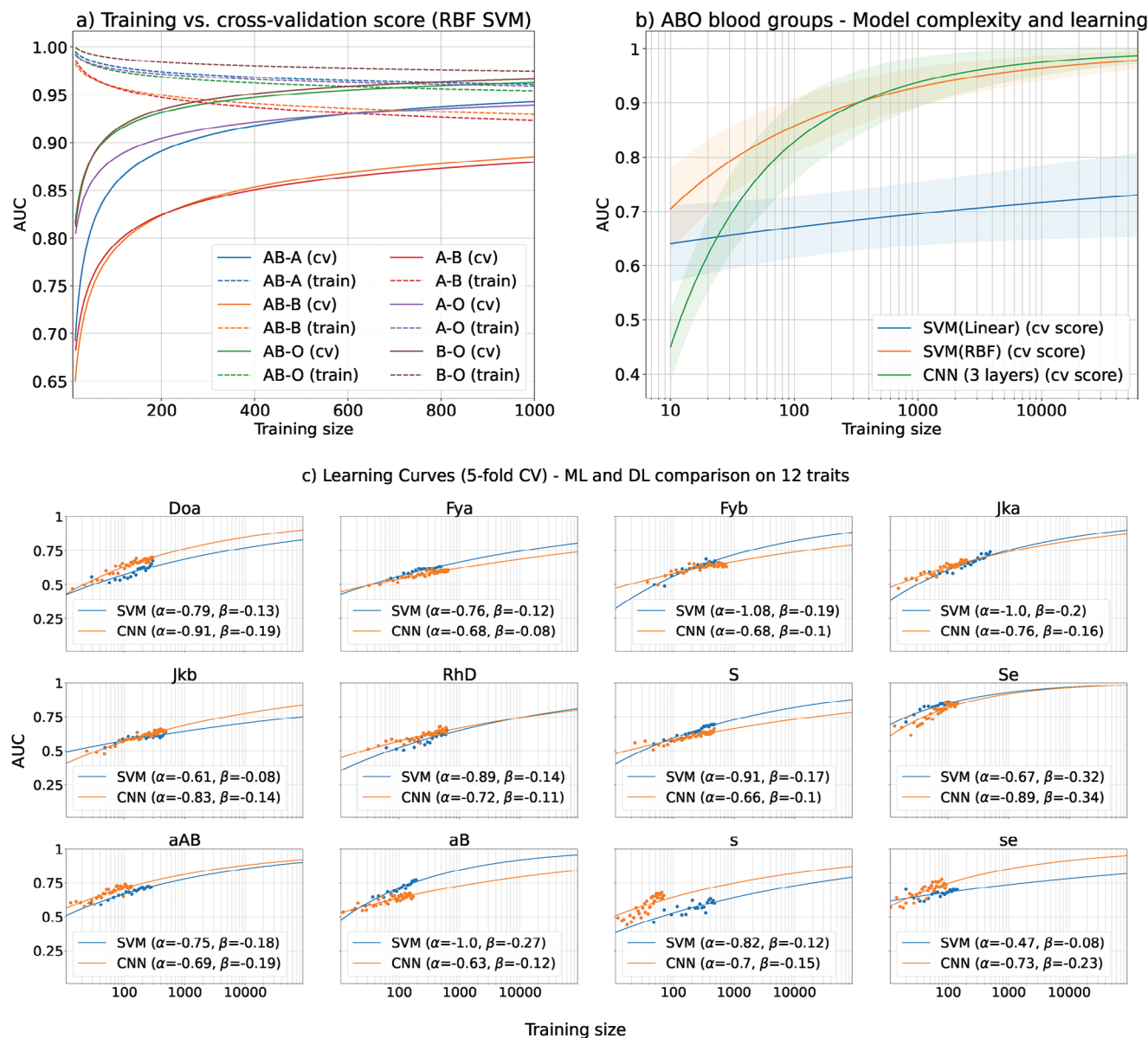


**Figure 6.** ROC curves and confusion matrices for classification results of traits with very few positives (Kp<sup>a</sup>, Lu<sup>a</sup>, Cw, K, and Yt<sup>b</sup>) or very few negatives (e).

### Learning Curves (5-fold CV) - ML and DL



**Figure 7.** Learning curves computed by varying the training size during fivefold cross-validation. Scaling comparison between a SVM model with an RBF kernel and a CNN model.



**Figure 8.** a) ABO learning curve fits for both the training and cross-validation score. b) Comparison between ABO learning curve fits using a linear SVM model, non-linear SVM model and a CNN model. c) Average learning curve fits for 12 selected additional traits. Learning curves of the remaining traits can be found in Figure S6, Supporting Information and include poor fits due to imbalance, small datasets, or outlier learning rates at low training sizes.

spatially separated in-line such that multiple analytes can be probed independently without introducing sample preparation steps or compromising the label-free nature of the method.<sup>[71,72]</sup> Additionally, a more efficient collection of Raman scattered light will improve SNR without increasing the exposure time, and on-line suppression of fluorescence background using modulation of the excitation wavelength<sup>[73,74]</sup> or shifted-excitation Raman difference spectroscopy (SERDS)<sup>[75]</sup> could increase the amount of accessible information in the Raman spectra significantly. Using multiple excitation wavelengths could further improve the detection level of low concentration analytes, by probing several resonances and pre-resonances with electronic states in the blood components.<sup>[76]</sup>

Learning curves are computed for 12 additional traits and presented in Figure 8c. All 12 traits show better performance when increasing the training size, however, superior scaling of the CNN model is not clear from this analysis. For some traits (Do<sup>a</sup>, Jk<sup>b</sup>, RhD, aAB, s, and se) the CNN model out-performs the non-linear kernel SVM model, both at current dataset size and at the projected large training sizes. For other traits, (Fy<sup>a</sup>, Fy<sup>b</sup>, Jk<sup>a</sup>, S, Se, and aB), the simple CNN model used in this study does not seem to improve the classification performance.

Raman spectra are high dimensional data so a model trained on a small dataset risk leading to over-fitting and it can be difficult to accurately estimate the learning ability of the model. Dimensionality reduction methods mitigates this problem,<sup>[77]</sup> but

for small training and subsequently test datasets (<100 samples per class), bias and high uncertainty in the testing strategy can falsely indicate a model's learning rate and overestimate the performance at low training sizes.<sup>[78,79]</sup> To examine the generalizability of the models the training and cross-validation scores are plotted in Figure 8. The training and cross-validation scores approach each other as the training size increases, implying that the SVM (RBF) models do not overfit on the ABO datasets, however, a larger number of donors is needed to examine this further.

## 4. Conclusion

We demonstrated that using RS on whole blood in a fused silica micro-capillary enables determination of multiple blood types in one single, fast, label-free measurement with no pre-analytical preparation. A peak analysis of the Raman spectra showed that hemoglobin dominates the Raman scattered signal of whole blood at 785 nm excitation, as expected. ABO blood groups were pairwise discriminated with an average AUC of  $0.91 \pm 0.03$  using SVM-based classification matching existing studies on Raman based ABO blood typing.<sup>[13–15]</sup> However, the direct use of whole blood significantly increases the throughput of our method and lack of reproducibility due to any heterogeneity of SERS substrates is avoided. Post-hoc scalability analysis indicated that increasing the training size to the entire donor cohort at Copenhagen University Hospital (30–60 000) would improve the performance, potentially enabling clinically relevant values, and scaling of the performance of a simple three-layer CNN model supports the potential of increasing the dataset size in the future.

We demonstrated a correlation between the presence of various erythrocyte and platelet antigens and the Raman spectrum of whole blood. Spectra containing rare frequency erythrocyte antigens such as Kp<sup>a</sup>, Lu<sup>a</sup>, Cw, K, and Yt<sup>b</sup> were classified with an average AUC of  $0.81 \pm 0.09$  by oversampling the minority class during training. The class imbalance presented a challenge in the development of the ML and DL models and a larger validation dataset should be used to improve the model stability and minimize the test variance. Highly imbalanced traits such as Do<sup>b</sup>, Fy<sup>b</sup>, and Jk<sup>a</sup> were randomly undersampled resulting in AUCs of  $0.76 \pm 0.09$ ,  $0.72 \pm 0.05$ , and  $0.74 \pm 0.05$ . The presence of low molar concentration platelet antigens HPA-1b, HPA-5b, HPA-15a, and HPA-15b were correlated to the Raman spectra and classified with an average AUC of  $0.73 \pm 0.4$ . Furthermore, Anti-B titers of blood group A donors (aB) above and below 10 were classified with an AUC of  $0.80 \pm 0.06$  and the dominant ABH-secreter status (Se) were classified with an average AUC of  $0.88 \pm 0.7$ . We emphasize that this is a proof-of-concept study and significant efforts need to be made to achieve clinically relevant performance. The specificity and sensitivity of the method are limited by high abundance analytes, which dominate the Raman spectra. In addition to increasing the amount of training data, efforts need to be made on emphasizing low concentration analytes, using methods such as in-line separation of blood components, background fluorescence suppression, or multi-excitation RS to truly enable clinical use.

Expansion of our approach to the entire available donor cohort requires particular handling of rare frequency antigen donors, using resampling, data augmentation and repeated measurements in the training process, such that the number of validation

samples are sufficient. Our study shows the feasibility of developing RS and AI as an accurate and fast clinical tool to determine a range of donor traits that are otherwise time-consuming and expensive to procure.

## Supporting Information

Supporting Information is available from the Wiley Online Library or from the author.

## Acknowledgements

This work was supported by the NovoNordisk Foundation (Grant reference numbers 18OC0034948 and NNF0070888). The authors acknowledge Danielle Poulsen (Copenhagen University Hospital) for reference data extraction.

## Conflict of Interest

The authors declare no conflict of interest.

## Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

## Keywords

blood typing, machine/deep learning, micro-capillary fluidics, precision transfusion medicines, Raman spectroscopy

Received: October 14, 2023  
Revised: November 12, 2023  
Published online: December 6, 2023

- [1] F. Lussier, V. Thibault, B. Charron, G. Q. Wallace, J. F. Masson, *TrAC Trends Anal. Chem.* **2020**, *124*, 115796.
- [2] N. Blake, R. Gaifulina, L. D. Griffin, I. M. Bell, G. M. H. Thomas, *Diagnosics* **2022**, *12*, 1491.
- [3] R. Luo, J. Popp, T. Bocklitz, *Analytica* **2022**, *3*, 287.
- [4] C. G. Atkins, K. Buckley, M. W. Blades, R. F. Turner, *Appl. Spectrosc.* **2017**, *71*, 767.
- [5] D. Cialla-May, C. Krafft, P. Rösch, T. Deckert-Gaudig, T. Frosch, I. J. Jahn, S. Pahlow, C. Stiebing, T. Meyer-Zedler, T. Bocklitz, I. Schie, V. Deckert, J. Popp, *Anal. Chem.* **2022**, *94*, 86.
- [6] J. Lin, J. Lin, Z. Huang, P. Lu, J. Wang, X. Wang, R. Chen, *J. Innov. Opt. Health Sci.* **2014**, *7*, 1350051.
- [7] I. P. Torres Filho, J. Terner, R. N. Pittman, L. G. Somera III, K. R. Ward, T. Filho, L.-n. G. Somera III, K. R. Ward Hemoglobin, *Am. J. Physiol. Heart Circ. Physiol.* **2005**, *289*, 488.
- [8] I. P. Torres Filho, N. M. Nguyen, R. Jivani, J. Terner, P. Romfh, D. Vakhshoori, K. R. Ward, *J. Surg. Res.* **2016**, *201*, 425.
- [9] M. Matthiae, A. Kristensen, *Opt. Express* **2019**, *27*, 3782.
- [10] R. Gautam, J. Y. Oh, M. B. Marques, R. A. Dluhy, R. P. Patel, *Lab Med.* **2018**, *49*, 298.
- [11] M. Z. Vardaki, C. G. Atkins, H. G. Schulze, D. V. Devine, K. Serrano, M. W. Blades, R. F. Turner, *Analyst* **2018**, *143*, 6006.

- [12] K. Buckley, C. G. Atkins, D. Chen, H. G. Schulze, D. V. Devine, M. W. Blades, R. F. Turner, *Analyst* **2016**, *141*, 1678.
- [13] D. Lin, Z. Zheng, Q. Wang, H. Huang, Z. Huang, Y. Yu, S. Qiu, C. Wen, M. Cheng, S. Feng, *Opt. Express* **2016**, *24*, 24750.
- [14] J. Wang, J. Lin, Z. Huang, L. Sun, Y. Shao, P. Lu, W. Shi, J. Lin, R. Chen, *Opt. Health Care Biomed. Opt. V* **2012**, *8553*, 855323.
- [15] J. Wang, Y. Y. Zeng, J. Q. Lin, L. Lin, X. C. Wang, G. N. Chen, Z. F. Huang, B. H. Li, H. S. Zeng, R. Chen, *Laser Phys.* **2014**, *24*, 6.
- [16] Y. Ozaki, A. Mizuno, H. Sato, K. Kawauchi, S. Muraishi, *Appl. Spectrosc.* **1992**, *46*, 533.
- [17] B. Schrader, B. Dippel, S. Fendel, S. Keller, T. Löchte, M. Riedl, R. Schulte, E. Tatsch, *J. Mol. Struct.* **1997**, *408–409*, 23.
- [18] P. J. Caspers, G. W. Lucassen, G. J. Puppels, *Biophys. J.* **2003**, *85*, 1.
- [19] D. Qi, A. J. Berger, *Applied Optics* **2007**, *46*, 1726.
- [20] M. Paraskevaidi, B. J. Matthew, B. J. Holly, B. J. Hugh, C. P. Thulya, C. Loren, C. Stjohn, G. Peter, G. Callum, K. G. Sergej, K. Kamila, K. Maria, L. M. Kássio, M. H. L. Pierre, P. Evangelos, P. Savithri, A. A. John, S. Alexandra, S. Marfran, S. S. Josep, T. Gunjan, W. Michael, W. Bayden, *Appl. Spectrosc. Rev.* **2021**, *56*, 804.
- [21] K. Landsteiner, *Wien. Klin. Wochenschr.* **2001**, *113*, 20.
- [22] G. Garratty, S. A. Glynn, R. McEntire, *Transfusion* **2004**, *44*, 703.
- [23] S. Zalpuri, J. J. Zwaginga, S. le Cessie, J. Elshuis, H. Schonewille, J. G. van der Bom, *Vox Sanguinis* **2012**, *102*, 144.
- [24] S. Zalpuri, D. Evers, J. J. Zwaginga, H. Schonewille, K. M. De Vooght, S. Le Cessie, J. G. Van Der Bom, *Transfusion* **2014**, *54*, 1981.
- [25] L. Helander, C. R. Alquist, In *Essentials of Blood Product Management in Anesthesia Practice* (Eds: C. S. Scher, A. D. Kaye, H. Liu, S. Perelman, S. Leavitt), Springer, Cham **2021**, pp. 11–20.
- [26] J. Portegys, G. Rink, P. Bloos, E. A. Scharberg, H. Klüter, P. Bugert, *Transfus. Med. Hemother.* **2018**, *45*, 331.
- [27] G. R. Krog, K. Rieneck, F. B. Clausen, R. Steffensen, M. H. Dziegiel, *Transfusion* **2019**, *59*, 3264.
- [28] M. G. Quirino, C. M. Colli, L. C. Macedo, A. M. Sell, J. E. L. Visentainer, *Hematol., Transfus. Cell Therap.* **2019**, *41*, 44.
- [29] H. Y. Li, K. Guo, *Blood Group Test.* **2022**, *9*, 827619.
- [30] A. Mujahid, F. L. Dickert, *Sensors* **2016**, *16*, 51.
- [31] S. Ebrahimi Fana, M. Paknejad, M. Aminian, *Biomed. Microdev.* **2021**, *23*, 3.
- [32] H. Zhang, X. Qiu, Y. Zou, Y. Ye, C. Qi, L. Zou, X. Yang, K. Yang, Y. Zhu, Y. Yang, Y. Zhou, Y. Luo, *Sci. Transl. Med.* **2017**, *9*, 381.
- [33] H. Zhang, R. Liu, Q. Li, X. Hu, L. Wu, Y. Zhou, G. Qing, R. Yuan, J. Huang, W. Gu, Y. Ye, C. Qi, M. Han, X. Chen, X. Zhu, Y. Deng, L. Zhang, H. Chen, H. Zhang, W. Gao, Y. Liu, Y. Luo, *ACS Nano* **2021**, *15*, 4.
- [34] C. Pipatpanukul, S. Takeya, A. Baba, R. Amarit, A. Sombonkaew, B. Sutapun, P. Kitpoka, M. Kunakorn, T. Sriksirin, *Biosens. Bioelectron.* **2018**, *102*, 267.
- [35] W. L. Then, M. I. Aguilar, G. Garnier, *Biosens. Bioelectron.* **2015**, *73*, 79.
- [36] W. L. Then, H. McLiesh, M. I. Aguilar, G. Garnier, *Biomed. Microdev.* **2016**, *18*, 6.
- [37] F. Di Cesare, L. Tenori, C. Luchinat, E. Saccenti, *J. Proteome Res.* **2022**, *21*, 11.
- [38] H. He, Y. Bai, E. A. Garcia, S. Li, in *Proc. of the Int. Joint Conf. on Neural Networks*, IEEE, Piscataway, NJ **2008**, pp. 1322–1328.
- [39] N. V. Chawla, K. W. Bowyer, L. O. Hall, W. P. Kegelmeyer, *J. Artif. Intell.* **2002**, *16*, 321.
- [40] H. C. Kim, S. Pang, H. M. Je, D. Kim, S. Y. Bang, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, Vol. 2388, Springer, Berlin, Heidelberg **2002**.
- [41] S. Wang, X. Yao, in *2009 IEEE Symp. on Computational Intelligence and Data Mining, CIDM 2009*, **2009**, pp. 324–331.
- [42] C. G. Atkins, H. G. Schulze, D. Chen, D. V. Devine, M. W. Blades, R. F. Turner, *Analyst* **2017**, *142*, 12.
- [43] M. Herbig, A. Isozaki, D. Di Carlo, J. Guck, N. Nitta, R. Damoiseaux, S. Kamikawaji, E. Suyama, H. Shintaku, A. R. Wu, I. Nikaido, K. Goda, *Nat. Methods* **2022**, *19*, 633.
- [44] B. Hansson, C. H. Allen, S. Qutob, B. Behr, B. Nyiri, V. Chauhan, S. Murugkar, *Biomed. Opt. Express* **2019**, *10*, 2275.
- [45] P. Lemler, W. R. Premasiri, A. DelMonaco, L. D. Ziegler, *Anal. Bioanal. Chem.* **2014**, *406*, 193.
- [46] D. A. Fedosov, B. Caswell, A. S. Popel, G. E. Karniadakis, *Microcirculation* **2010**, *17*, 615.
- [47] H.-Y. Chang, A. Yazdani, X. Li, K. A. Douglas, C. S. Mantzoros, G. E. Karniadakis, *Biophys. J.* **2018**, *115*, 1371.
- [48] S.-J. Baek, A. Park, Y.-J. Ahn, J. Choo, *Analyst* **2015**, *140*, 250.
- [49] C. Cortes, V. Vapnik, *Mach. Learn.* **1995**, *20*, 3.
- [50] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, *J. Mach. Learn. Res.* **2011**, *12*, 2825.
- [51] R. Batuwita, V. Palade, in *Imbalanced Learning*, Wiley, Hoboken, NJ **2013**, pp. 83–99.
- [52] G. Lemaître, F. Nogueira, C. K. Aridas, *J. Mach. Learn. Res.* **2017**, *18*, 1.
- [53] J. Liu, M. Osadchy, L. Ashton, M. Foster, C. J. Solomon, S. J. Gibson, *Analyst* **2017**, *142*, 4067.
- [54] L. Alzubaidi, J. Zhang, A. J. Humaidi, A. Al-Dujaili, Y. Duan, O. Al-Shamma, J. Santamaría, M. A. Fadel, M. Al-Amidie, L. Farhan, *J. Big Data* **2021**, *8*, 53.
- [55] J. Acquarelli, T. van Laarhoven, J. Gerretzen, T. N. Tran, L. M. Buydens, E. Marchiori, *Anal. Chim. Acta* **2017**, *954*, 22.
- [56] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, A. Talwalkar, *J. Mach. Learn. Res.* **2016**, *18*, 1.
- [57] F. Chollet, *Keras*, <https://github.com/fchollet/keras> **2015**, Accessed January 2023.
- [58] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, X. Zheng, In *Proceedings of the 12th USENIX Symposium on Operating Systems Design and Implementation, OSDI 2016*, ACM, New York **2016**, pp. 265–283.
- [59] M. Abe, T. Kitagawa, Y. Kyogoku, *J. Chem. Phys.* **1978**, *69*, 4526.
- [60] B. R. Wood, P. Caspers, G. J. Puppels, S. Pandiancherri, D. McNaughton, *Anal. Bioanal. Chem.* **2007**, *387*, 1691.
- [61] M. Kaneko, Y. Kato, H. Horiuchi, M. Osawa, *Immunol. Lett.* **2003**, *86*, 45.
- [62] M. E. Reid, C. Lomas-Francis, M. L. Olsson, *The Blood Group Antigen*, Academic Press, San Diego, CA **2012**.
- [63] S. Raudys, A. Jain, *IEEE Trans. Pattern Anal. Mach. Intell.* **1991**, *13*, 252.
- [64] W. Ng, B. Minasny, W. d. S. Mendes, J. A. M. Demattê, *SOIL* **2020**, *6*, 565.
- [65] M.-A. Schulz, B. T. T. Yeo, J. T. Vogelstein, J. Mourao-Miranada, J. N. Kather, K. Kording, B. Richards, D. Bzdok, *Nat. Commun.* **2020**, *11*, 4238.
- [66] I. Balki, A. Amirabadi, J. Levman, A. L. Martel, Z. Emersic, B. Meden, A. Garcia-Pedrero, S. C. Ramirez, D. Kong, A. R. Moody, P. N. Tyrrell, *Can. Assoc. Radiol. J.* **2019**, *70*, 344.
- [67] R. L. Figueroa, Q. Zeng-Treitler, S. Kandula, L. H. Ngo, *BMC Med. Inf. Decis. Making* **2012**, *12*, 8.
- [68] J. Cho, K. Lee, E. Shin, G. Choy, S. Do, *arXiv:1511.06348* **2015**.
- [69] A. Rokem, Y. Wu, A. Lee, *bioRxiv:196659* **2017**.
- [70] W. A. Flegel, *Blood Transfus.* **2013**, *11*, 6.
- [71] S. Tripathi, Y. V. B. Kumar, A. Agrawal, A. Prabhakar, S. S. Joshi, *Sci. Rep.* **2016**, *6*, 26749.



- [72] M. Matthiae, X. Zhu, R. Marie, A. Kristensen, *Analyst* **2019**, *144*, 602.
- [73] A. C. De Luca, M. Mazilu, A. Riches, C. S. Herrington, K. Dholakia, *Anal. Chem.* **2010**, *82*, 738.
- [74] A. De Luca, K. Dholakia, M. Mazilu, *Sensors* **2015**, *15*, 13680.
- [75] M. T. Gebrekidan, C. Knipfer, F. Stelzle, J. Popp, S. Will, A. Braeuer, *J. Raman Spectrosc.* **2016**, *47*, 198.
- [76] A. P. Lister, C. J. Highmore, N. Hanrahan, J. Read, A. P. S. Munro, S. Tan, R. N. Allan, S. N. Faust, J. S. Webb, S. Mahajan, *Anal. Chem.* **2022**, *94*, 669.
- [77] W. Jia, M. Sun, J. Lian, S. Hou, *Complex Intell. Syst.* **2022**, *8*, 2663.
- [78] C. Beleites, U. Neugebauer, T. Bocklitz, C. Krafft, J. Popp, *Anal. Chim. Acta* **2013**, *760*, 5.
- [79] A. Vabalas, E. Gowen, E. Poliakoff, A. J. Casson, *PLOS One* **2019**, *14*, e0224365.