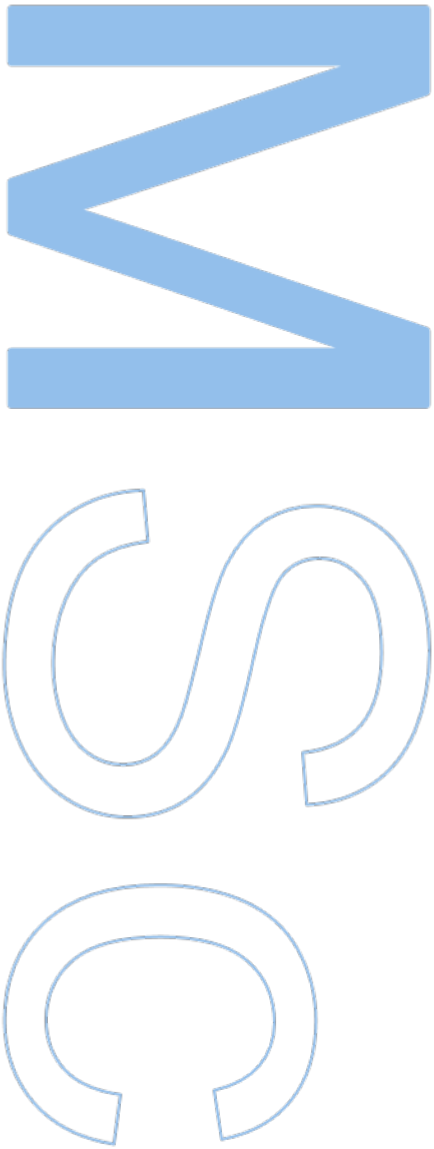
Performance validation of an AI spectroscopy PoC hematology device through hyperspectral microscopy



Ricardo Jorge Fernandes Tavares
Master's Degree in Medical Physics
Department of Physics and Astronomy
2022

Supervisor
Rui da Costa Martins, Senior Researcher,
Applied Photonics Research Center – INESC TEC


U.PORTO
FACULDADE DE CIÊNCIAS
UNIVERSIDADE DO PORTO

Sworn Statement

I, Ricardo Jorge Fernandes Tavares, enrolled in the Master Degree in Medical Physics at the Faculty of Sciences of the University of Porto hereby declare, in accordance with the provisions of paragraph a) of Article 14 of the Code of Ethical Conduct of the University of Porto, that the content of this dissertation reflects perspectives, research work and my own interpretations at the time of its submission.

By submitting this dissertation, I also declare that it contains the results of my own research work and contributions that have not been previously submitted to this or any other institution.

I further declare that all references to other authors fully comply with the rules of attribution and are referenced in the text by citation and identified in the bibliographic references section. This dissertation does not include any content whose reproduction is protected by copyright laws.

I am aware that the practice of plagiarism and self-plagiarism constitute a form of academic offense.

Ricardo Jorge Fernandes Tavares

November 2022

Acknowledgements

As with everything in life, no moment is truly enjoyed or appreciated without the ones that made it possible. Therefore, I would like to take some time to express my sincere gratitude to all of those who subtly hide behind each sentence, each piece of data and each hour spent working.

This thesis would not be possible without the work and guide of my supervisor, professor Rui Martins, from INESC TEC, to whom I thank the opportunity of working on this compelling and ground breaking project. My appreciation as well to my team mate, Luís Monteiro, for his work and for always being available to help and share a laugh.

I would like to thank FCUP and its distinguished teachers, for the knowledge and passion they convey, without which, I wouldn't have reached this far in my academic journey. I'm specially thankful to professor Pedro Teles, Course Director of the Master's degree in Medical Physics, for his patience and for doing everything in his reach to help his students succeed.

My appreciation to the Centro Hospitalar Veterinário do Porto for working with the team and providing the blood samples essential to this research.

To all my friends and people I met along the way, thank you for making the hard moments more bearable and for enriching my life with memories, experiences, laughs and wisdom. As happiness is only real when shared.

Lastly, but most importantly, I express my warmest and deepest gratitude to my parents, for always showing unconditional love and for supporting me, morally and financially, in every step of my life. To you, I'm eternally in debt, and may your efforts echo through my successes.

Resumo

As tecnologias usadas em analisadores hematológicos padrão e de última geração não atingem a acessibilidade necessária para atender às necessidades globais de saúde pública e veterinária. O problema está no foco atual da investigação, que dá prioridade a dispositivos para uso laboratorial, com componentes volumosos e dispendiosos, o que impossibilita a sua aplicação em ambiente doméstico ou no campo. Isto limita bastante o alcance benéfico e potencial do hemograma completo.

O objetivo desta tese é mostrar como a microscopia hiperespetral pode ser usada para melhorar um dispositivo *point-of-care* (PoC) de espetroscopia sem reagentes, que usa inteligência-artificial para obter parâmetros de hemograma completo. Esfregaços de sangue foram analisados ao microscópio e os seus espetros obtidos. Com esses espetros, foi feita uma análise de componentes principais, onde três grupos distintos puderam ser facilmente identificados: cães saudáveis, cães com infeção e cães com anemia e infeção – mostrando assim uma alta capacidade de separação de amostras intrínseca ao método. Previsões quantitativas para eritrócitos, hemoglobina e hematócrito foram feitas, comparando com um analisador hematológico de veterinária padrão (Mindray BC-5000 Vet) tanto para o dispositivo PoC como para o método microscópico. Os coeficientes de determinação (R^2) e os erros absolutos médios percentuais (MAPE) foram calculados. Ambos os métodos mostraram forte concordância entre si e em relação ao analisador padrão, com o PoC (R^2~0,9; MAPE~6%) apresentando melhor desempenho que o método microscópico (R^2~0,85; MAPE~10%). A visão detalhada da amostra no método microscópico permite uma melhor separação das amostras, enquanto que o maior volume de sangue analisado no dispositivo PoC produz resultados quantitativos melhores. Isto mostra que existe simultaneamente equivalência e complementaridade entre os dois métodos, propondo que sejam utilizados em conjunto para o desenvolvimento da tecnologia.

Identificaram-se falhas no método do microscópio: baixa representatividade da amostra e dependência dos resultados com a qualidade do esfregaço. Aumentar a área analisada e automatizar a formação do esfregaço foram sugeridos como soluções para esses problemas. Para trabalho futuro, foi proposto estudar a relação entre os nível microscópico e o macroscópico, a fim de criar um mapa que permita o uso da separabilidade superior do método do microscópio pelo dispositivo PoC.

Palavras-chave: [Hematologia, Quimiometria, Hemograma Completo, Espetroscopia, Inteligência-Artificial, Saúde, Veterinária, Microscopia Hiperspetral, *Point-of-Care*]

Abstract

Standard and state-of-art hematology analyzer technology fail to achieve the accessibility requirements to fulfill the global needs of health care and veterinary. The problem is rooted in the current focus of research development, that prioritizes benchtop devices for laboratory use, often with bulky parts and expensive components, creating a gap for in-house and in-field applications, which massively restricts the benefits and potential of the complete blood count test.

The aim of this thesis is to explore the capabilities of hyperspectral microscopy and how they can be used to improve a reagentless spectroscopy point-of-care (PoC) device that uses artificial intelligence to obtain complete blood count parameters. Blood smears were analysed under the microscope and their spectral behaviour acquired. From these spectra, a principal component analysis was done, where three distinct groups could be easily identified: healthy dogs, dogs with infection and dogs with anemia and infection – thus showing a high sample differentiation capability intrinsic to the method. Quantitative predictions for red blood cells, hemoglobin and hematocrit were done, comparing back to a standard veterinary hematology device (Mindray BC-5000 Vet) for both the PoC device and the microscope method. The coefficients of determination (R^2) and mean absolute percentage errors (MAPE) were calculated. Both methods showed strong agreement between each other and in respect to the standard analyzer, with the PoC (R^2~0.9; MAPE~6%) having a better performance then the microscope method (R^2~0.85; MAPE~10%). The detailed sample view on the microscope method allows for better sample separation, while the bigger blood volume analysed in the PoC device yields better quantitative results. This shows that there's simultaneously an equivalence and complementarity to these methods, advocating for their cooperative use to the development of the technology.

Flaws in the microscope method were identified: low sample representativeness and results dependence on blood smear quality. Increase in analysed area and automation of smear formation were suggested as solutions to these problems, respectively. For future work, it was proposed studying the relationship between the microscopic and macroscopic levels, in order to create a map that allows the use of the heightened separability of the microscope method by the PoC device.

Keywords: [Hematology, Chemometrics, Complete Blood Count, Spectroscopy, Artificial-Intelligence, Health Care, Veterinary, Hyperspectral Microscopy, Point-of-Care]

Table of Contents

List of Tables

List of Figures

List of Abbreviations

AI	Artificial-Intelligence
ASVCP	American Society for Veterinary Clinical Pathology
BAS	Basophils
BLL	Beer-Lambert Law
CBC	Complete Blood Count
CovM	Covariance Mode
CV	Coefficient of variation
EOS	Eosinophils
Fcup	Faculty of Sciences of the University of Porto
HbCO	Carboxyhemoglobin
HbO_2	Oxyhemoglobin
HbR	Deoxyhemoglobin
HCT	Hematocrit
Hgb	Hemoglobin
HO	Hold-out
INESC TEC	Institute for Systems and Computer Engineering, Technology and Science
IoT	Internet of Things
LED	Light Emitting Diode
LOC	Lab-on-a-chip
LYM	Lymphocytes
MAPE	Mean Absolute Percentage Error
MCH	Mean Corpuscular Hemoglobin
MCV	Mean Corpuscular Volume
MetHb	Methemoglobin
MON	Monocytes
NEU	Neutrophils
PCA	Principal Components Analysis
PCT	Procalcitonin
PLT	Platelet
PMT	Photomultiplier
PoC	Point-of-care
PoCT	Point-of-care technology
RBC	Red Blood Cell
RDW	Red Cell Distribution Width

SBC	Single-Board Computer
SLAI	Self-learning artificial-intelligence
SWNIR	Short-wave near-infrared radiatioin
UP	University of Porto
UV	Ultraviolet radiation
Vis	Visible radiation
WBC	White Blood Cell

1. Introduction

1.1. Background of the study

With increasing exposure to high levels of stress, pollutants, carcinogens and overall poor health habits, healthcare is a growing concern among individuals in developed countries [1]. The lack of resources, exacerbated by overpopulation, climatic changes and pollution, make healthcare of paramount importance for developing countries too, as inaccessibility to medical equipment and disease predisposition are a reality [2]. Thankfully, due to the development of technology, devices have been growing more accurate and attainable, allowing for tests that asses various health conditions in no time and with actionable results. Having accessibility as one of the main driving forces of technology development, smaller and cheaper devices became the focus of researchers, in order to provide answers to the in-house and in-field scenarios. Some reach further, developing wearable devices [3]. Portability and price, as well as connectivity by medical Internet-of-Things (IoT) networks, allow for close follow-ups, constant monitoring of people with chronic diseases and widespread health check-ups in regions with insufficient health resources [4]. Many tests that were once lab exclusive, are now available, almost in the literal sense, in the palm of our hands, as Point-of-Care (PoC) technology starts to emerge.

Despite of all the progress made in different technologies to bring healthcare closer to the general public, small steps have been taken towards accessibility in the hematology analyzer technology used in complete blood count (CBC) tests. Most advances focus on bench top designs with little to no portability and expensive [5] components, making them unfeasible for in-house and in-field applications [6][7][8][9] [10]. CBC tests can help diagnose anemia, infections, blood cancers, immune system diseases, side effects to drugs and a big number of other pathologies, making it the most requested laboratory test in hospitals and a mandatory tool in healthcare [11].

In the veterinary area, a similar necessity for Point-of-Care technology (PoCT) can be noticed. With globalization and urbanization rising, the trade of animals and animal products is increasing, as is their exposure to zoonotic diseases, due to greater contact with other animals and humans. Rapid and wide-ranging spread of emerging diseases could have catastrophic effects on people, animals and the economy. Traditional laboratory-based testing is usually laborious and expensive, requiring

deeply trained personnel and convoluted work logistics. By introducing PoCTs in veterinary, costs could be greatly reduced and time-efficiency increased.

Veterinary shares most of the same benefits of hematology analyzer technology as human healthcare does, with the CBC test being of undeniable importance in the area. It too faces the same ongoing inconveniences, with current and state-of-art technology not fulfilling the desired accessibility characteristics. However, the situation is aggravated, as most emerging technologies focus their applicability on human blood. This is contradictory, since most research starts on animal trials, on account of animal blood samples, namely dogs and cats, being easier to come, with less health and ethical complications involved. In this sense, both veterinary and human developments on the applications of this technology could be suffering a setback.

Therefore, upon this brief delineation, it's imperative that the CBC test becomes available ubiquitously, regardless of country development and socioeconomic status of the patients. New technology for hematology analyzers needs to be researched in order to achieve the desired costs and portability, without trading accuracy and relevance. The same efforts should be put into the veterinary field, as this brings benefits across multiple planes. Only this way, the CBC's benefit can achieve it's full reach.

1.2. Base work

As an answer to the technology gaps pointed out in the previous subchapter, Martins et al. introduced a "Point-of-care Vis-SWNIR spectroscopy" device, designed to acquire spectra of bodily fluids, in particular, whole-blood [12][13][14][15]. Self-learning artificial intelligence (SLAI) is then used to search for systematic and stable covariance between blood composition and spectral features in order to unscramble the information and identify and quantify blood constituents.

The device uses reusable plug-in capsules with mirrored ends in order to maximize the optic path in blood. It only takes a small amount of it (< 10μL) to be able to acquire the spectrum of the whole-blood sample, which allows for fingerprick sampling (or earprick in animals). All controls and displays can be accessed through the Internet of Things (IoT) interface, which can be opened on the computer or smartphone. So far, algorithm implementation is dependent of an external computer, but future improvements intend to embed the analysis process in the internal computer

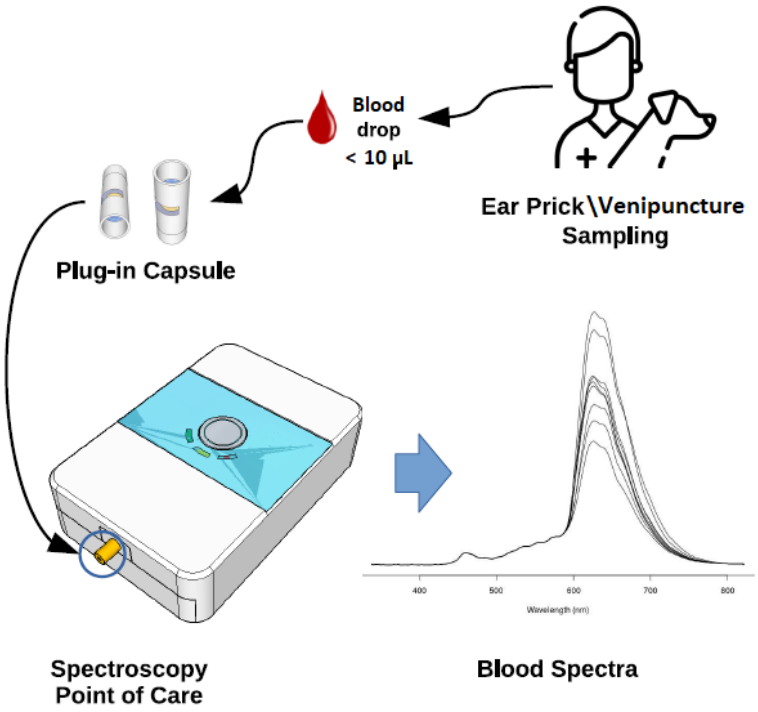


Figure 1: Schematic representation of the PoC approach. A minimal amount of blood can be used, allowing for ear prick or venipuncture sampling. Utilizing the plug-in reusable capsules, whole-blood spectra can be acquired in no time.

of the device. A power LED is used as a source of light and a 7-core optical fiber is responsible of transmittance of light and collection of the spectrum, which is then acquired by a miniaturized spectrometer. A small single-board computer is used to administrate all the tasks and a battery supplies energy for all components. Due to these features, the device is highly portable and the acquisition process exceptionally simple. The components are relatively inexpensive as well, with the reusable capsules being the biggest investment.

To fit the highly interferant nature of blood spectral data, a SLAI method was developed, overcoming the gaps of state-of-art chemometrics algorithms. This method, called Covariance Mode (CovM) search, searches for groups of samples in the feature space that contain the same interference information characteristics, that is, samples whose correlation between spectral features and composition have the same covariance eigenvector; a covariance mode. These CovMs have stable covariance and represent the correlated variability between certain spectral feature and certain combination of compositional variables. With this information, precise quantitative measurements can be done on unknown samples, by finding which covariance mode (or combination of modes) best fits the sample in the feature space.

Table 1: Performance parameters from previous work.

	RBC (10^{12}/L) Dog / Cat	Hgb (g/dL) Dog / Cat	HCT (%) Dog / Cat	WBC (10^{9}/L) Dog	PLT (10^{9}/L) Dog
R	0.94 / 0.98	0.95 / 0.99	0.93 / 0.98	0.88	0.84
SE	0.54 / 0.56	12.86 / 5.72	4.43 / 2.36	6.92	61.2
LV	1 / 1	1 / 1-2	1-2 /1-2	1	1-3
MAPE	6.4 / 5.7	7.1 / 4.1	4.4 / 1.7	25.4	24.7

R-Pearson correlation coefficient of Measured vs Predicted plot; SE-Standard error of regression; LV-Number of latent variables that correlate the compositional variable to spectral variability; MAPE-Mean absolute percentage error. Data acquired from [13][14][16]

So far, only experiments with dog and cat blood have been made, and over a partial number of CBC parameters, namely: hemoglobin (Hgb), red blood cells (RBC) count, hematocrit (HTC), white blood cells (WBC) count and platelet (PLT) count. Throughout all these parameters, SLAI has shown strong correlation and accuracy relative to measurements done by standard hematology analyzers, outperforming, at the same time, all of the most commonly used state-of-art chemometrics methods. In table 1, some of the performance parameters obtained in previous works are summarized.

1.3. Relevance

One limitation of most chemometrics algorithms, with no exception to the one here presented, is that right predictions within the sample population don't necessarily mean causality. For this technology to become successful in analysing a broad range of unknown samples, it's important to rule out sampling-biased correlation among compositional variables and find true causality with spectra characteristics. Hence, in this thesis, I search deeper the causality between constituent concentration and spectral features of dog whole-blood, by exploring this same relationship on a macroscopic level. The SLAI algorithm was applied to spectra of blood smears obtained by microscope. By getting a closer look of the constituents, it's possible to better amplify and isolate their contribution to the spectra and understand how their interference translates to the macroscopic level. This is crucial to validate previous results, as it is to improve the predictions, by teaching the algorithm to search preferable principal components and covariance modes. By working with animal blood, the adaptation capabilities and versatility of the device are demonstrated, with direct pay-off for the veterinary field.

On a global standpoint, a truly portable and inexpensive PoC device, that is able of obtaining accurate and actionable information, is of utmost importance to reach CBC's true potential and fulfil public health's and veterinary necessities. The use of spectroscopy allows for miniaturized and cheap components with low energy consumption (batteries can be used to feed the system) and simplified acquisition process (robust inner works and user-friendly interaction). Computing resources can then be focused on implementing AI within the device itself and the development of an IoT platform to further improve portability and user-friendliness. It's reagent-less nature and small amount of blood needed enable reusability of the plug-in capsules, cutting the costs additionally, and simplify the sample preparation process (no extensive training would be needed to use the device). This technology could be applied in in-house scenarios, enabling periodic checks on chronic diseased or bedridden patients. It could be easily transported to areas with shortage of medical resources in order to assess the health state of the population. On these premisses, primary care, nursing and epidemic control would all benefit greatly from this technology. Some of the concepts could even be extended to wearable technology, with smartwatches or smartphones serving as host devices to attachable gadgets that sample and acquire the spectral data, which is then analysed by the former, further increasing the closeness of healthcare. Regarding veterinary, this technology could help prevent the mass spread of diseases among livestock by keeping close track on animal health in a time-efficient and non-expensive manner, which in turn is a great advantage for human health and the economy. Our closest friends would also benefit from these technologies, since the expenses of pet care would be cut down and the process facilitated to lessen suffering and inconvenience.

As for my self-benefit, the work here developed, due to it's multidisciplinary nature, taught me valuable things that have a carry over to other research projects and applications. Spectroscopy is a widely used technique across multiple areas within and outside the field of medical physics. The same is true with artificial-intelligence, being more and more present in recent technologies. Therefore, by working with them, I got a better understanding of their combined potential and possible applications, which translates in an easier "get used to" in future works. Working with blood open my horizons to the area of hematology; it's extensive diagnostic abilities, current technologies and their deficiencies. I am now familiarized with laboratory techniques to prepare and analyse blood samples, as well as recognizing how different diseases manifest themselves through blood hemogram parameters. Working together with Prof. Rui Martins, as my supervisor, and Luís Monteiro, as my thesis colleague, showed me

the importance of communication, trust and individual and collective work within a team. I'm therefore wiser and better prepared for the future.

1.4. Automated hematology

Prior to 1950, blood counts were done manually using a hemocytometer and a microscope. Not only this methods was very time-consuming and laborious, it was susceptible to great inaccuracies due to sample preparation and human perception. In 1953, Wallace H. Coulter patented the Coulter principle, which gave rise to the first commercially available counting instrument by 1956. With the first whole-blood bench top automated analyzer being introduced in 1968, a new era of hematology was born: the automated hematology [17]. Since then, big improvements were made to refine this technology: the introduction of flow cytometry allowed for 5-part differentials; embedded automated colorimetric tests made possible to measure hemoglobin concentratioin; the use of fluorescent analysis increased the accuracy of the differentials; improvements in computational powered automated the process further by flagging certain conditions upon cell analysis. Nowadays, an automated hematology analyzer is a quintessential device for any medical analysis laboratory, as the information it provides is of great value for healthcare. Consequently, researches keep pushing technology forward by either trying to improve on standard technology or by creating new one to replace it. A short review on current and state-of-art technology is presented next, in order to get up to date with the automated hematology situation and give context to the reasons that motivated the chain of work that led to this thesis.

1.4.1. Complete Blood Count (CBC)

The complete blood count is the most commonly performed test in hospitals. It allows to count and differentiate the various types of cells contained in blood, as well as measure chemical and physical parameters related to them. With this knowledge, health professionals can then assess the general health condition and infer about the nutritional state of patients. Most importantly, this test is essential for diagnosing anemia, infections, allergies, immunodeficiencies, blood cancers (eg. Leukemia), acute hemorrhagic states and monitoring side effects of certain drugs, which makes it essential in the healthcare picture [11].

For the cell counts, in blood, we have red blood cells (RBC), platelets (PLT) and white blood cells (WBC), with the last ones splitting into 5 different types: neutrophils

(NEU), eosinophils (EOS), basophils (BAS), monocytes (MON) and lymphocytes (LYM). As for the chemical parameters, a CBC measures the hemoglobin concentration (Hgb), the mean corpuscular hemoglobin (MCH), that is, the mean hemoglobin per RBC, and procalcitonin concentration (PCT). There are also physical parameters that can be measured, related to the size and distribution of cells in the blood, like the mean corpuscular volume (MCV), red cell distribution width (RDW) and hematocrit (HCT).

1.4.2 Current Technology

Nowadays, a standard hematology analyzer condenses different types of techniques in a single automated machine to measure the desired parameters. Commonly, these techniques are: flow cytometry, for counting and differentiating WBCs, the Coulter principle, to count and determine the volume of RBCs and PLTs, and a colorimetric test, to measure Hgb concentration [18][19][20][21].

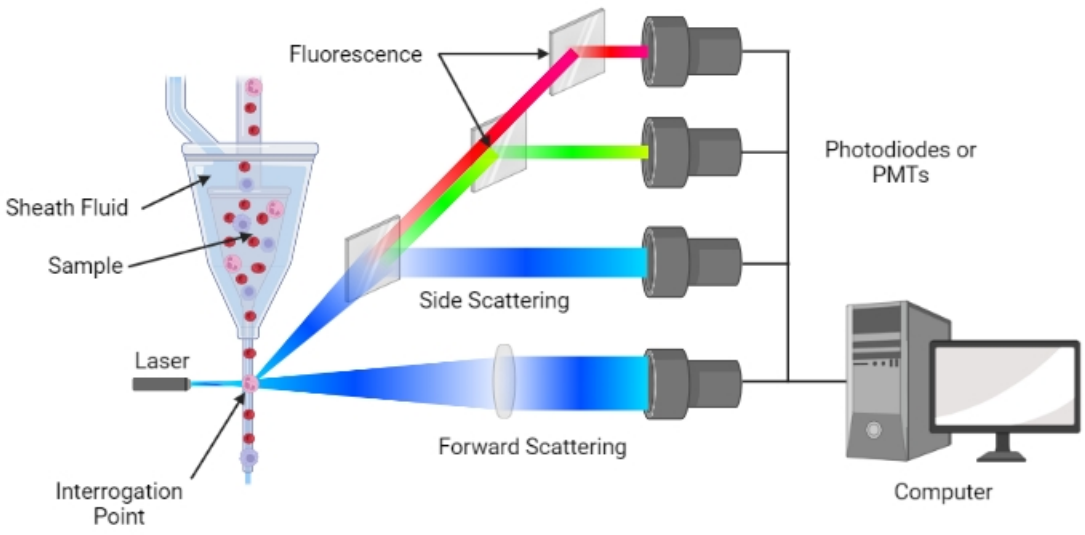


Figure 2: Working principles of flow cytometry. Cell morphology and complexity dictate scattering and fluorescence patterns, allowing for the distinction of different cells.

- *Flow Cytometry*: It consists of 3 main components: fluidics, optics and electronics. The fluidics system consists of a sheath fluid that is pressurized and running in a laminar flow around the sample fluid. When they join together, the sheath fluid delivers and focuses the sample in a way that a single file of cells is created, enabling them to be analyzed one by one. The optical system consists of lasers, that interact with the cells producing scattering and fluorescence, and optical collectors, like photomultiplier tubes or photodiodes, to capture the scattered and fluorescent radiation. When the sample reaches the intersection with the laser, also known as interrogation point, forward scattering, side scattering and fluorescence is produced. The amount of forward scattering is proportional to the size of the cell, with bigger cells producing more of it. On the other hand, side scattering is proportional to shape and internal cell complexity, with more complex cells producing more side scattering due to more interactions with its organelles. The presence of fluorophores, either endogenous or exogenous, can be used to collect further information on cell structure, helping with differentiation. The electronics are then responsible for collecting and processing all the data, identifying different groups of scattering and fluorescence patterns and labeling them as a specific cell based on that.

- *Coulter Principle*: In a Coulter counter, a tube with a small hole is immersed in an electrolyte solution with suspended non-conducting particles. One electrode is placed inside the tube and other on the outside, creating a current path through the orifice when a voltage is applied. A pump is used to suck the solution from inside the tube, making the particles pass through the hole. When a particle crosses the hole, it displaces its volume in electrolyte solution, increasing the impedance of the electrical path by an amount proportional to that volume. That change creates a pulse in current or voltage, depending on the set-up, which can be used to count the number of particles passing through and their volume. Controlling the flow of solution that is being pumped, the concentration of particles from the sample can be determined. This principle found great success in the hematology field and it’s still the standard method for counting RBCs and PLTs.

- *Colorimetric test*: In this test, hemoglobin is diluted in a reagent solution to create a hemoglobin compound that is color stable and strictly obeys Beer-Lambert’s law. The absorbance of the solution is then compared to a known standard solution from which the hemoglobin concentration can be determined.

These methods form the standard basis of modern hematology analysis.

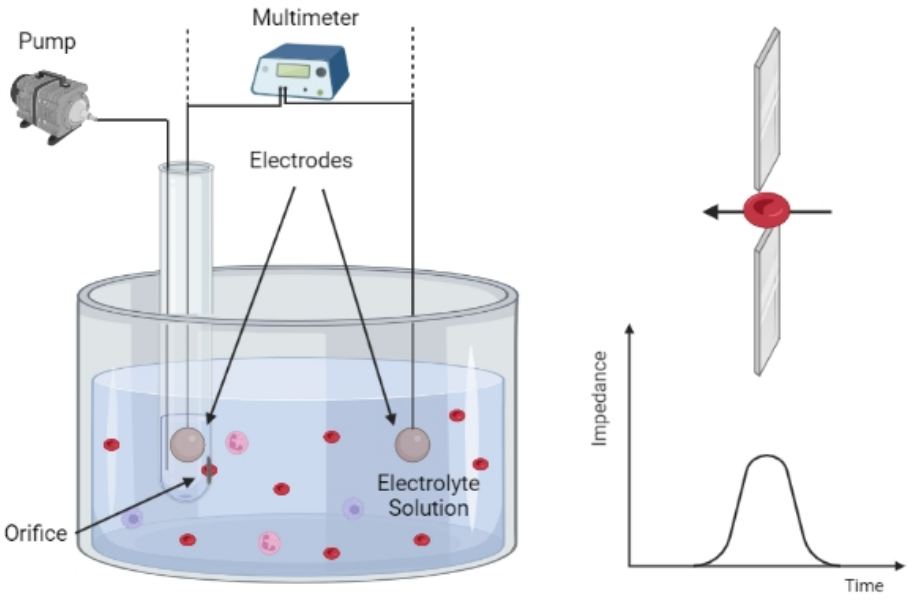


Figure 3: Working principles of a Coulter counter. The displacement of electrolyte volume by the cell when passing through the aperture results in a sudden increase in current path impedance. Each impedance peak and respective shape is used to count and differentiate cells based on size.

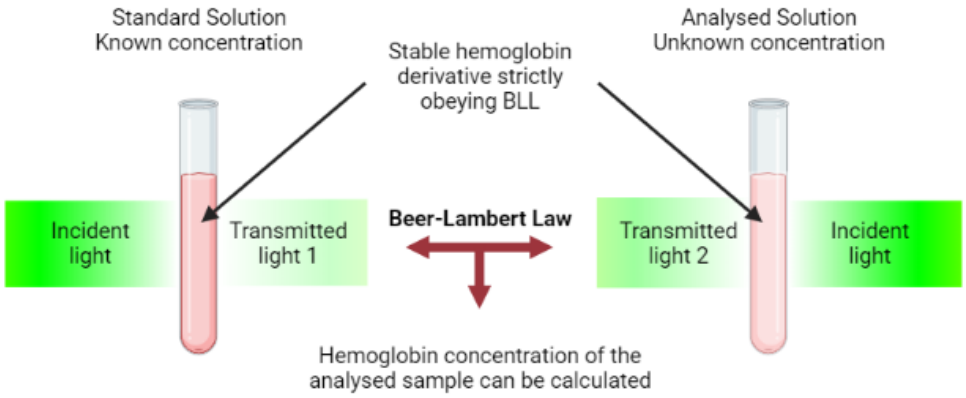


Figure 4: Working principles of colorimetric tests. A color stable hemoglobin derivative is obtained by chemical reaction. Through the BLL law, its absorbance can be compared with a standard solution to obtain its concentration.

With the automation and optimization of these techniques, it's possible to merge them in a single self-operating hematology analyzer, allowing for decision-making information in a matter of minutes. Studies report satisfying intra-assay coefficients of variation (CV) of 1-6% across most parameters for the most common brands of hematology analyzers [22]. For inter-assay CVs, parameters like RBC, Hgb, HCT, MCV

also show good agreement, with around 4% for all of them. Yet, they're not without flaws. WBC count and differentials and PLT count were reported to disagree between brands up to 11% and 19%, respectively [23][24]! With these parameters being of paramount importance on evaluating immune system response and stroke risk, manual tests for evaluating them have to be used to confirm these results, which increases significantly the time needed for results. Abnormal or immature cells are also not identifiable by most of these machines (some recent models have flagging capabilities to identify them) [6].

However, the real downside of this technology is that, the limiting factors in speed and convenience for this test is not the analysis itself, but rather the logistics surrounding it, mainly the collection of blood samples through venipuncture and the need of a dedicated lab to analyze them. As an example, in a pediatric or veterinary setting, fear and involuntary movements can make the collection of blood through venipuncture difficult for the healthcare professional and uncomfortable for the patient. For bedridden patients, the sampling would have to be done at home and then processed back in the lab, increasing the wait time for the results, or by moving the patient to the hospital/clinic for the test, something that is very inconvenient and strenuous for them. Also, in developing countries, some areas don't have quick access to hospitals or labs, or often face themselves with facilities under stress, due to the overflowing of patients, making it impossible to have results on actionable time.

With these inconveniences in hand, how is the current research and state-of-art technology trying to deal with them?

1.4.3. State-of-art

In order to push technology further, researchers have to choose between building on existing methods and their known advantages, or coming up with new ones, with different assets of their own. Whichever the case is, the aim is always to improve aspects like accuracy, time consumption, accessibility, user-friendliness, convenience or workflow. From my review on the forefront of hematology analyzer technology, there are four main branches being intensively researched as of today: Microscopic image analysis by computer vision, microfluidics or lab-on-a-chip (LOC) technology, improved standard methods and limited parameter devices.

- *Computerized microscopic image analysis* [7][9][10]: A fully automated microscope is used to acquire highly magnified pictures of stained blood cells. Multiple wavelength channels in brightfield and fluorescent microscopy allow for multispectral images of the

cells, extracting both chemical and morphological information. AI image analysis algorithms are then applied to identify and quantify these cells based on their spectral print and morphology. Recent evaluations of this technology report high accuracy parameters, great flagging capabilities, no need of additional calibrations, user-friendliness, due to automation, and high potential for further improvement, by implementation of better AI algorithms. However, the high costs of the components and their limited miniaturization capability make this technology impractical for patient-owned and in-field scenarios. Also, accuracy depends highly on sample preparation, as this technology can only successfully work with images of highly diluted stained blood cells in a monolayer. This makes the use of lab-on-a-chip technology almost an indispensable ally if user-friendliness and accuracy is to be achieved.

- *Lab-on-a-chip / Microfluidics*: A lab-on-a-chip is device that integrates into a single chip/cartridge multiple laboratory functions, using only small amounts of sample volume (down to the picoliter). They're made up of microchannels and small chambers that are responsible for the transportation, mixing, storing and preparation of samples and reagents without need of further manual intervention. They can also include electrodes, valves, pumps, electronics, electric and/or magnetic fields, etc., depending on the desired function. The increasing trend of LOCs in technology is justified by their seemingly endless functions and many advantages; high applicability in PoC or low resource scenarios; use of very small sample and reagent volumes; user-friendly due to automation; multiple analysis can take place in a single chip at the same time; low fabrication costs; high process control; no need to house or washout reagents in analysis equipments. The downside to this technology is that, with microfluidics, there can be unexpected effects due to surface roughness and capillary effects on the sub-milimetric scale, jeopardizing sample preparation and consequently the results. As for availability, the majority of this technology is not yet ready for commercialization. Also, LOCs are rarely used on their own; usually they are only responsible for sample management; all acquisition and analysis is made by external devices, often bulky, which reduces portability. Examples of state-of-art methods that use microfluidics are:

- *Viscoelastic flow cytometry* [7][25][26]: It takes advantage of the overpowering shear and wall lift forces that arise in viscoelastic fluids flowing through microchannels to focus suspended particles in a single file or monolayer.

- *LOC Coulter counter* [27][28][29]: LOC technology is applied to collapse all the Coulter counter setup into a single chip. A focusing method (microfluidics, optical or acoustic, usually) is used to focus the particles into a single file. Electrodes within the microchannels are then used to pick up on the fluctuating impedance.

- *Improved standard technology* [6]: Brands and researchers reach for higher accuracy, better flagging capabilities and extended number of measurable parameters within the existing technology. With these improvements, new applications and diagnostics are possible using only hematology analyzers. Some proposed applications are: early identification of marrow regeneration, prediction of total platelet recovery after chemotherapy and diagnosis and monitoring of microangiopathies. This brings major benefits to the clinical setting, but doesn't improve the accessibility gap, since these devices are expensive and bulky.

- *Limited parameter devices* [30][31][32][33]: As the name suggests, limited parameter devices are gadgets that measure a very small number of blood parameters, usually one to three. Even though there are many of these devices being explored, usually for hemoglobin measurement, the principles and methods used by each one are almost exclusively developed by the corresponding research teams, giving rise to a vast array of existing technologies. A strong point of these devices is that, most designs focus on portability and user-friendliness, as they are intended for PoC applications. Also, the restricted number of parameters measured heighten the potential for better accuracy, due to there being more degrees of freedom to achieve parameter specific optimization. However, this is a disadvantage as well, since for a complete blood count, multiple devices would be needed.

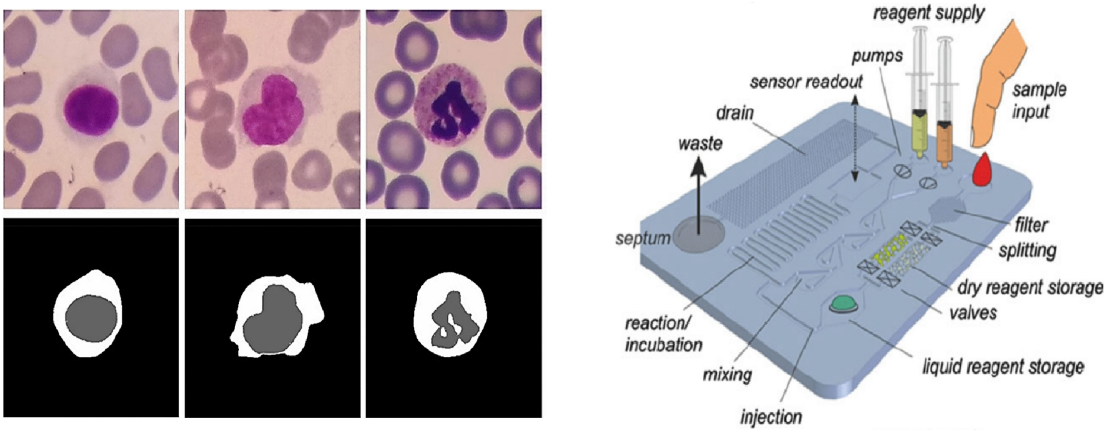


Figure 5: State-of-art technology. (Left) Cell complexity can be used for segmentation by computer vision algorithms in order to differentiate and identify cells. In this case, different types of leucocytes are presented, showing different morphology and inner structures. (Right) A microfluidics lab-on-a-chip cartridge with passive components. Channels conduct the sample through each chamber for mixing, filtering and analysis.

Table A.1 in Attatchments compiles some of the existing state-of-art devices, the technology they implement and some performance parameters.

From this review on current and state-of-art technology, it's clear that there is in fact a gap in PoC devices to act on in-house and in-field scenarios. Most designs, old or new, focus mainly of bench top applications with expensive components, or cartridge based applications that are either dependent of bulky external devices, or have a limited amount of measurable parameters.

To fix this issue, new concepts and methods must be explored in the field of hematology analyzer technologies, specifically aiming to make them cheap and portable, without trading for accuracy and relevance. Rui Martins proposed using spectroscopy together with SLAI to achieve that goal, developing his own PoC spectroscopy device and SLAI algorithm for whole-blood analysis. In this work, I extend on those concepts and solidify previous results by applying the SLAI algorithm to spectra obtained by microscopy.

1.5. Theoretical background

To fully understand the potential of the technology in hand and the relevance of this work, it is first needed to comprehend the concepts behind it. What is the true aim of point-of-care technology? How can spectroscopy and artificial intelligence fulfil that aim? How can spectroscopy acquire the rich data contained in blood? How does AI transform that data into valuable information? The answers to these questions are the core of the method here developed and therefore it's fundamental to shine some light on them. I now go over some of the key theoretical concepts prevalent throughout all the development of this thesis.

1.5.1. Point-of-Care Technology

Point-of-care technologies are technologies capable of giving fast and actionable information at the time and place of care. Their use spans a multitude of fields, like chemistry, agriculture and geology, but it's in healthcare that they find greater demand. They exclude the need of a dedicated lab and some, even the direct intervention of a healthcare professional. This allows for closer and more frequent screenings, giving better understanding of patient health to act upon [34]. The implied

portability of PoCTs facilitate the screening in low medical resourced areas too, which is a crucial step for improving developing countries' public health.

In primary care, the need to improve quality of care, health outcomes and financial feasibility is the main drive of PoCTs innovation. For home care, the main drivers of innovation are the ability for self-management, facilitated home nursing, closer screenings for chronic diseases and allowing remote consultations with higher diagnosing abilities. For emergency medical services, PoCTs would increase the number of possible tests on site and during transportation due to portability, cutting shorter assessment times and therefore increasing the chances of a better outcome. PoCTs could also help control pandemic situations to a greater extent, as regular self-screens would get individuals up to date on their health status avoiding further spread of the disease.

In livestock veterinary, animal-side tests to help control diseases and assess overall animal health are greatly desired, in particular, for remote or impoverished geographical regions, where lab-based tests would not be easily available, implying steep expenses and delayed results. For pet care, PoCTs promises to reduce testing costs and animal handling and suffering, which is highly sought after from both professionals and owners.

Even though greater focus has been put in the development of PoCTs in recent years, this concept is nothing new, as several devices capable of measuring vital signs, or other health parameters, have already been marketed and are present in the daily lives of many patients, in particular, for the measurement of blood glucose, blood pressure and heart rate. There is an increasing trend to try to combine these technologies with smartphones or smartwatches, in order to make the monitoring of these parameters more accessible and comfortable. The demand and offer of wearable technologies are exponentially rising, which reflects the approval by the population [35].

There are several working principles used by POCTs, depending on the parameters to be measured. However, there are common points between all of them: the components that make up the system must be miniaturizable and relatively inexpensive; the device should be easy to use and the testing methods must be robust, ensuring relevant and reliable results. For these reasons, two very commonly used technologies are spectroscopy and artificial intelligence.

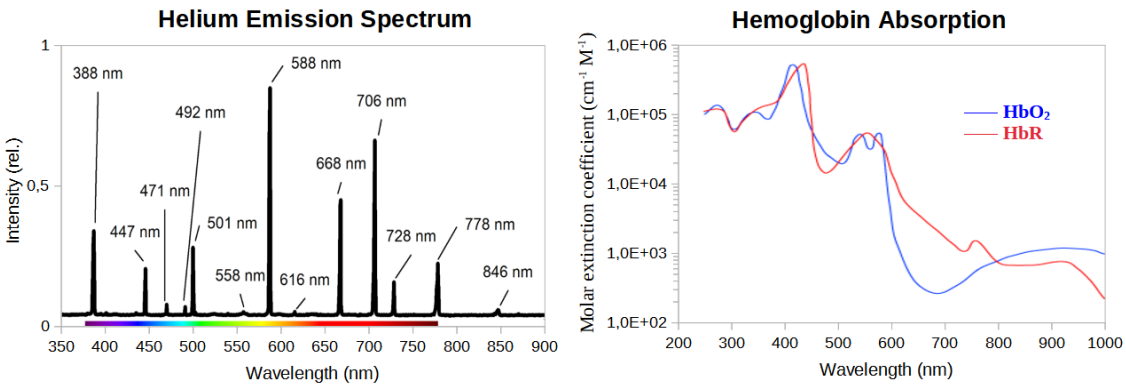


Figure 6: Comparison between an atomic spectrum and a molecular spectrum. (Left) Helium emission spectrum. The individual peaks can be easily singled out, giving a print of the element. (Right) Hemoglobin absorption spectra in its oxygenated (HbO_2) and reduced (HbR) forms. The spectrum is a continuum across the wavelengths and spanning multiple orders of magnitude. Instead of peaks there are bands.

1.5.2. UV-Vis Spectroscopy

Spectroscopy is a technique used to study the absorption and emission of electromagnetic waves in matter to further characterize or identify analytes on their spectral print. This is possible due to the property of atoms and molecules of having energy states that can be excited by radiation or other forms of energy, producing absorption and emission spectra that are unique to each chemical species. In the ultraviolet and visible range, the energy states excited in single atoms correspond to electron transitions between orbitals. This creates spectra consisting of discrete spikes that can be, for the most part, easily singled out [36]. However, for molecules, electron orbital transitions can be accompanied by the excitation of molecular rotational and/or vibrational energy states, that are very close together energetically. This blends the discrete excitation spikes into broad and continuous absorption or emission bands, creating continuous spectra that still allow for characterization and identification [37], but are prone to changes depending on many other variables. Figure 6 illustrates the differences between atomic and molecular spectra.

The appeal of UV-Vis spectroscopy is that it is possible to find a linear relationship between absorbance and analyte concentration for most solutions, allowing for quantitative measurements: this is called the Beer-Lambert Law (BLL) [38]. As light crosses the sample, constituents interact with it by absorption or scattering, enriching the spectrum with information about the molecular structure of constituents and the composition of the sample. Due to the plentiful information contained in the spectra and the correlation between spectral features and constituent concentration, spectroscopy is one of the pillars in analytical chemistry, and it's clear how this could

be useful in blood cell counts as well, since blood constituents have distinct spectral prints themselves, which in theory, should allow for their quantification.

In spectroscopy, 4 main components can be identified: the light source, to irradiate the sample, the light transfer system, on which light is carried from source to detector, the detector, to collect and convert the light into spectra, and a computer, to interpret and display the data. When it comes to the implementation of this technique into PoCT, a few small and relatively cheap components are all that is needed. For the light source, LEDs or laser diodes can be used due to their small proportions. For the transfer system, it's very common to use optical fiber, which enables bending the optical path into small areas. For the detector, micro spectrometers can be used; some of them have dimensions comparable to a coin. As for the computer, most single-board computers (SBCs) nowadays have more than enough computational power to administrate all the tasks required, including connecting to a IoT platform in order to interpret, display and save the acquired information.

1.5.3. Chemometrics

As mentioned in the previous chapter, one advantage of spectroscopy is that it is possible to find a linear relationship between absorbance and analyte concentration due to the BLL. With this relationship, quantitative measurements can be done, which is highly desirable in analytical chemistry and could be very useful for hematology. For a single analyte, the concentration can easily be determined by direct application of this law. For solutions with just a few known components, simple multiple component quantitative analysis methods can be applied. These usually consist in the preparation of standard solutions with varying amounts of the analytes and interferents. These standards are then to be used to calibrate the machines, that use systems of linear equations based on the BLL, to determine the concentration of analytes and interferents [39]. This is usually enough for laboratory prepared samples, where separation and purification techniques can be used to reduce the number of constituents to a small number, allowing for all of them to be accounted in the models, as well as in the preparation of standards. However, with complex solutions containing a large amount of constituents, like most biological samples, the production of standards is nearly impossible, as separating the components and accounting for them all is extremely laborious or even unachievable. This problem led to the development of chemometrics.

Chemometrics is the discipline that applies mathematical, statistical and computational methods to the analysis of chemical data. The main objective of chemometrics is to extract as much information as possible from the chemical data obtained, usually by finding a correlation between two or more physico-chemical properties of the sample; which for the work here presented are spectral features and analyte concentration. What separates chemometric methods from the previously mentioned methods is that, chemometrics doesn't depend on a causal physical model to understand effects and correlate identities. Instead, it revokes the need of causality and finds correlation by pattern recognition, prediction and factorization [40][41]. Only then is a model created to fit the results and give them causal meaning (otherwise they would be of no value).

The advantage to this approach is that all information contained in the data is included in the analysis: the entire range of wavelengths of the spectra can be used; it's not needed to filter or separately account for interferents or impurities; even noisy data can be dealt with and included. Chemometrics takes in raw mixed data and gives back also raw mixed data, but in a format that can be better demystified. This way, it can provide information that traditional methods simply cannot, as physical models will oversimplify the complexity of the sample and of the processes involved. This fits perfectly with biological samples, since information about a particular constituent is present in different scales and distributed along the range of wavelengths. Also, most interferents and inpurities are either unaccountable or simply unknown. In the end, all the rich information contained in biological samples is preserved and used, allowing for relevant and actionable results.

However, as a consequence of this seemingly miraculous simplicity, chemometrics faces itself with some challenges that are imperative to overcome if valuable results are to be attained. To begin with, a good population representativity is mandatory to exclude false correlations between variables. For that, sampling should be large and differentiated enough so that all independent variables are presented as such to the algorithm. Any falsely correlated variables would immediately devalue the results, as there would be no guarantee that the method is actually able to identify and quantify at least one of those variables. Then, there's the fact that the math, programming and calibrations involved in these methods are considerably more complex than traditional methods. Even though there are multiple standard chemometrics algorithms being extensively used in the field, their performance depends noticeably with application. Ultimately, an application specific algorithm is to be developed for ideal performance. This is usually the biggest challenge, specially for

biological samples, as interferences are scattered throughout the entire spectra and in multiple scales, requiring to be unscrambled. Lastly, the results obtained need to have a causal interpretation or model for them to be validated and sound. No confidence can be placed in the results if there's no logical proof that the algorithm is in fact finding, through correlation, a causal link between two identities. This is particularly essential in the healthcare scenario, as blindly trusting patient health on pattern recognizing or statistical algorithms is simply not acceptable. This can be aggravated if the two previous requisites are overlooked.

Chemometrics is often paired with AI, as many algorithms benefit amply from its techniques, for example, heuristics, machine learning, machine vision and artificial neural networks. When it comes to implementing artificial intelligence, all that is needed is a computer with enough computational power to be able to run the algorithms and store the necessary data. With the ever increasing computational power of SBCs, most are perfectly capable of AI, taking very little space and being relatively cheap, which makes the integration of AI and chemometrics into PoC technology a favourable match.

1.5.4. Blood spectroscopy

Blood is a complex biological fluid permeated by valuable information about the individual's health status. Due to its large number of constituents and intricate chemical interactions, it's difficult to create mathematical models that relate different physico-chemical identities to each other, and this is not less true when relating spectral features to constituents concentration. Analyte information is mixed together with that of interferents through multi-scaled interferences and matrix effects distortions that span the entire considered wavelength range. The major challenge of implementing spectroscopy in blood cell counts is then, to unscramble this information in a way that quantitative predictions can be made.

The major spectral features of blood come from the hemoglobin contained in red blood cells. Hemoglobin, in particular its oxygenated derivative, oxyhemoglobin (HbO_2), is the overwhelming absorber in blood due to it's high concentration and absorption peaks in the UV-Vis region. The remaining constituents will overlap their spectra to that of hemoglobin, resulting in the final observed spectra. When two or more constituents have overlapping absorption bands, they cannot be individually measured and are said to be interfering. For oxyhemoglobin, two interfering regions can be identified: in the 400-450 nm region, its biggest interferent is bilirubin; in the

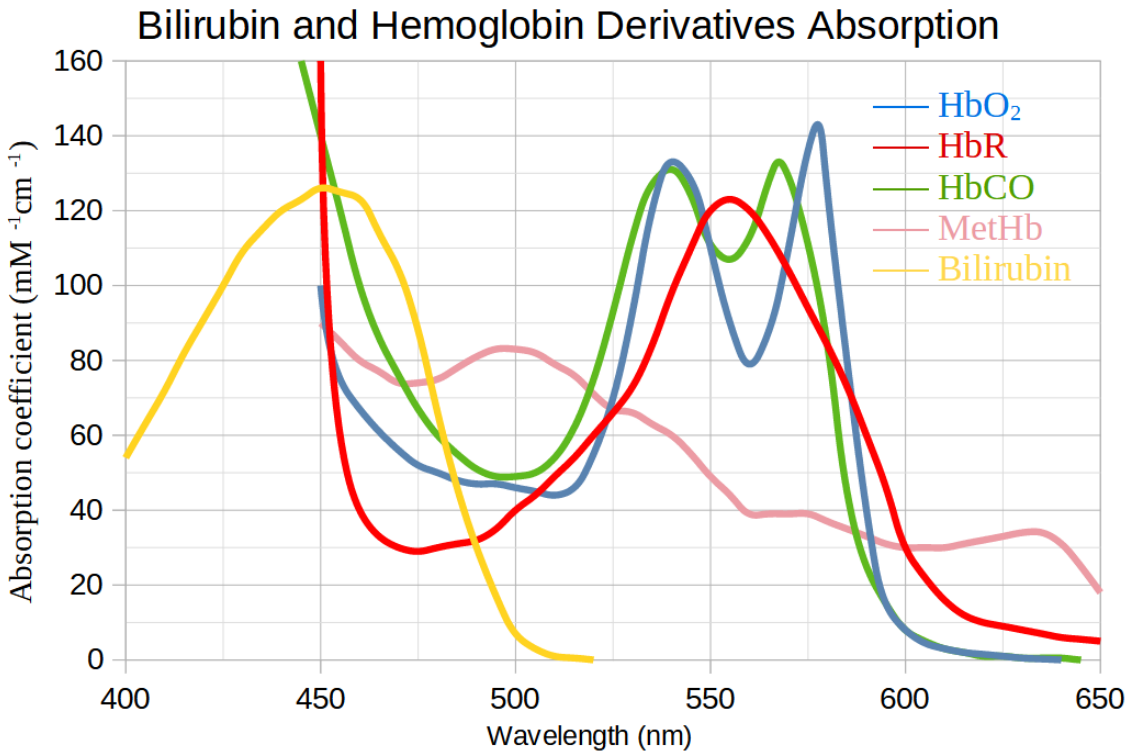


Figure 7: Hemoglobin interferents spectra. Hemoglobin derivatives have similar absorption peaks which makes them highly interferent, even at lower concentrations. The wide absorption peak of bilirubin also creates big interferences throughout most of the lower portion of the visible range, making it extra hard to unscramble hemoglobin's spectral information.

525-580 nm region, it interferes with its derivatives. Nevertheless, this latter region is the one that shows biggest variance in spectral features due to the overpowering absorption of hemoglobin, and will likely be the one picked by the algorithms to correlate spectral variance to hemoglobin concentration, RBC counts and HCT.

2. Materials and methods

2.1. Sampling and conditioning

Dog blood samples were collected at Centro Hospitalar Veterinário do Porto. These samples resulted from daily clinical practice and were taken through venipuncture of the jugular vein by qualified personnel. To guarantee blood integrity, the samples were stored in K3-EDTA tubes, collected daily and analyzed on the same day; only samples with less than 8 hours from blood extraction were analyzed. Transportation was made in a thermal bag with a cold accumulator to preserve the blood until it reached the lab for analysis. Before the spectra acquisition procedure, the samples were carefully homogenized by rocking the tube gently to assure the representativeness of the blood drop extracted.

A total of 145 dog blood samples were used in this study. This total was then separated into two groups: the training set and the prediction set. The training set consisted of 85 samples, while the prediction set consisted of 60 samples. This separation was done based on sample collection order, which doesn't jeopardize the analysis validity due to the randomness of blood composition granted from daily clinical practice.

2.2. CBC parameters and benchmark

CBC parameters were determined by the veterinary hematology analyzer Mindray BC-5000 Vet. This device uses flow cytometry combined with tri-angle laser scatter and chemical dye for WBC differential count, electric impedance method for RBC and PLT counts and a cyanide-free colorimetric method for hemoglobin determination.

The CBC parameters of the training set were used by the algorithm to relate spectral variability to composition variability. As for the parameters of the prediction set, these were used to compare with the predictions made by the methods.

2.3. Spectroscopy

- *PoC*: The device used consists of a Vis-SWNIR PoC spectrometer prototype controlled through an IoT interface. It uses a reusable plug-in capsule with a mirrored

cap, to trap a blood drop (~5 µL) in a thin layer. White light is then shone on it to acquire the whole blood spectrum. It contains a power LED of color temperature 4500 K to produce the white light, a 7-core fiber, with the 6 outer cores used for transmittance and the inner core for collection, a spectrometer and a small computer to manage all the commands fed through the IoT interface. To acquire the spectra, a 5 µL blood drop was set on the glass inside the capsule using an automatic pipette. The mirrored cap was carefully screwed on so that the blood drop would get trapped between the mirror and the glass and form a thin layer. The capsule was then coupled to the PoC and the spectrum acquired. Through the IoT interface, integration time was adjusted and new acquisitions were made until the spectrum spanned the entire detectable range on the intensity axis without saturation. The spectrum was then saved, the capsule removed from the PoC and the mirrored cap unscrewed. The capsule's parts were washed with deionized water, dripping to a solution of water and bleach, and then dried off with paper, being ready to be used again on other samples. The process was repeated until spectra from all samples were acquired.

- *Microscope*: Zeiss Axio Vert A1 with the Axiocam 208 color camera. A neutral filter was used on the LED and no filter was used after the objective. The numerical aperture was set to maximum. Using an automatic pipette, a 5 µL blood drop was set in a glass slide, proceeding to prepare the blood smear with the slide dragging method, trying to maintain similar slide angles and drag speeds to assure consistency among all samples and restrict other variability sources. The blood smear was then positioned under the microscope and a representative homogeneous monolayer region was searched. Once found, the spectra were acquired through the ocular by an optical fiber connected to the spectrometer. A magnification of 200x with integration times of 130ms were used, acquiring 3 consecutive spectra. LED intensity was set to optimize spectrum signal/noise ratio inside the linearity zone of the spectrometer. With the camera and using the Zeiss Zen 3.5 microscopy program, the same region was photographed, adjusting the RGB histograms with the Min/Max option and the LED intensity to avoid saturation. These steps were repeated until all the blood samples were analyzed. After that, the slides were immersed in bleach and washed with water afterwards. Before being used the next day, each slide was again cleaned with ethanol to remove any remaining residue.

2.4. Covariance Mode search

To correlate spectral features with sample composition, a SLAI algorithm was developed by Rui Martins: the Covariance Mode (CovM) search method. The method works by performing the following steps:

i. Feature space optimization: An initial rough form of information unscrambling is done by optimizing the feature space through selection of adequate features and transforms. This allows to separate and identify groups with similar characteristics at the spectral and compositional level based on their coordinates in the feature space.

ii. Covariance Mode search: From the feature space, the method searches a pre-determined number of neighbors of a given sample, creating groups and storing those with high correlation and low number of latent variables. Covariance within these sub-groups is then maximized by adding and removing samples until a stable covariance group is found. This sub-group is considered a covariance mode when one latent variable is sufficient for providing a small standard error. This is repeated for all samples in feature space until all CoVMs are found. At this point, it's known how spectral features and composition change accordingly for each group, since the covariance eigenvectors is determined. Then, to predict the composition of an unknown sample, interpolation of the covariance eigenvector of the most similar group is used.

The advantage of searching for CovMs is that, by finding the gradient of information between variable and spectral features, precise quantitative measurements can be made. Also, due to the decomposition of the information to the feature space and the equivalence of latent structure of the spectra and composition, there is a facilitated interpretation of interference.

2.5. Qualitative test

The first step in this analysis is to check the ability of the microscopic method to separate samples based on their composition. To do so, a basic principal components analysis (PCA) was done, retaining only the two first main components. This separates the samples based on spectral features. Consulting the CBC parameters, it's possible to understand if this spectral separation is related to composition differences or not; clustering of similarly composed samples indicate separation ability. Raw spectral data and images of the analysed blood smears also provide valuable information to interpret the results and so are also to be used.

2.6. Quantitative test

The extent of the usefulness of hyperspectral microscopy to the development of the PoC device lies in it's ability (or lack of) to predict CBC parameters that are simultaneously comparable to standard technology and the PoC device.

With the PCA of the microscopic data working as an initial separation of the samples, it's then necessary to correlate microscopic spectral features to composition. This is done by applying the CovM search algorithm to the training set and finding the groups of stable covariance and single latent variable. After that, the algorithm is ready to make predictions of the CBC parameters. Only hemoglobin related parameters were acquired, that is, RBC, Hgb and HCT; minor parameters like WBC and PLT didn't get enough representativeness in the microscopic method to be able to get predictions of them. The same process of prediction was done with the data acquired with the PoC device.

The predictions done by both methods were then benchmarked against the Mindray BC-5000 Vet analyzer by doing a linear regression in the scatter plot and calculating the coefficient of determination (R^2) and the mean absolute percentage error (MAPE).

3. Results and discussion

3.1. Qualitative test

From the spectra obtained through microscopy, a principal component analysis was done to try to identify clusters of spectra that share similar spectral information. Figure 8 presents the results from the PCA. In it, three distinct clusters can be identified. By analysing the CBC results it's possible to trace back what each cluster represents. Table 2 shows the CBC parameters from 3 samples picked, one per cluster. The red cluster, containing most of the samples, represents dogs with CBC parameters inside the reference range. The blue cluster represents dogs with low values of RBC, Hgb and HCT (Anemia) and high values of WBC (Infection). The green cluster represents dogs with normal levels of RBC, Hgb and HCT, but high values of WBC. As for the outliers, it wasn't possible to drop them in a specific category. They were either samples with different blood composition who didn't get a number of other samples to reach enough representativeness, or were samples whose analysis were compromised by blood integrity or smear quality.

From this analysis, it's possible to conclude already that the microscope method has a very strong intrinsic ability to separate samples based on their composition. But it's important to understand where this heightened ability comes from.

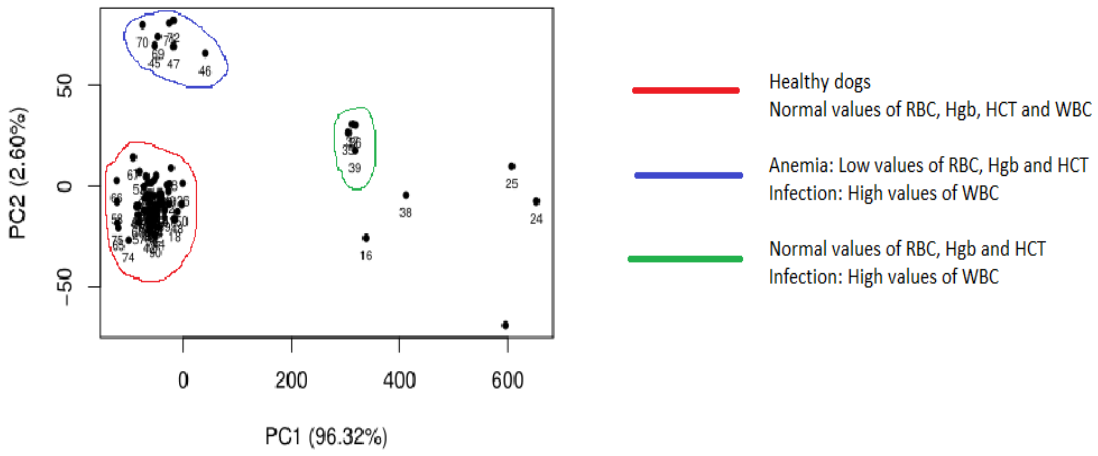


Figure 8: Results from the principal component analysis. Its possible to identify 3 distinct clusters. Red cluster corresponds to healthy dogs. Blue cluster to dogs with anemia and infections. Green cluster dogs with infections only.

Table 2: Hemogram values of Red, Green and Blue clusters samples.

Sample Number	RBC (10^{12}/L)	HGB (g/dL)	HCT (%)	WBC (10^{9}/L)	PLT (10^{9}/L)
20 R	7.78	197	0.499	8.69	287
35 G	5.74	132	0.342	94.5	590
47 B	3.10	62	0.164	26.50	28

To further explore microscopy's ability to qualitatively separate samples based on spectral information and interpret the results obtained, one sample from each cluster was picked and their spectra and microscopic blood smear images were compared (figure 9). Healthy dog blood samples are characterized by densely packed red blood cells and a reduced probability of finding a white blood cell in a randomly picked area. This is due to the extremely low proportion of WBCs to RBCs in non infectious situations (1:1000). That is represented by the red spectra and image in figure 9. For infectious cases, the increased count of WBCs makes it more likely to find one in a randomly selected representative area on the blood smear, as it's possible to see in both blue and green framed images. If infection is the only condition present, then no other major alterations are expected to be noticed in the blood smear; RBCs still behave in the same manner, showing high density and homogeneity in their distribution (green framed image). As for the spectral behaviour, little to no change, when compared to healthy dogs, is seen in the region of hemoglobin's peak absorbance (around 525 to 580 nm). However, after 580 nm, as hemoglobin's absorbance drops significantly, a slight drop in transmittance can be seen in the spectra of dogs with infection, due to the presence of WBCs. This is consistent with previous results, that show that, even though WBC have higher absorbance in the UV region, their spectral variability is better noticed in the higher wavelength region due to less interference with hemoglobin [16]. As for anemia, it manifests itself by a low density of RBCs, often forming clumps that create an inhomogeneous environment, as it can clearly be seen in the blue framed image. The spectral changes of anemia are pretty evident as well. With hemoglobin as the major absorber in blood, its reduction results in an increase in transmitted light in the spectra as a whole, which widens and pushes the spectra up. This effect is specially noticed around hemoglobin's absorption peaks (around 525 to 580 nm). These changes can be seen in the blue spectrum.

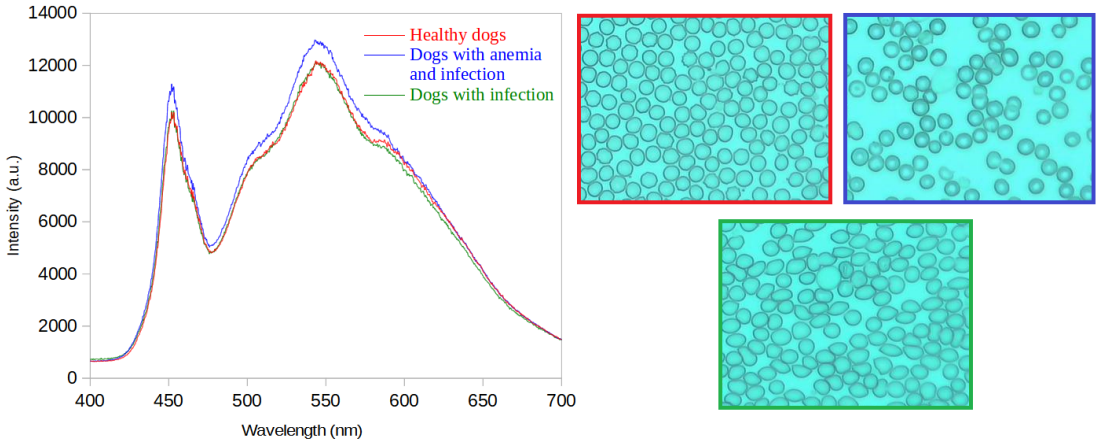


Figure 9: Spectra and microscopic image comparison for one sample in each cluster. Red – Healthy dogs; Blue – Dogs with anemia and infection; Green – Dogs with infection. Blood from dogs with anemia have higher transmittance in the range 525 to 580 nm. Dogs with infections have subtle spectral variability for wavelenghts over 580 nm.

With this analysis, it's possible to understand that, due to the microscope method's detailed view of the sample, components manifest themselves in a more noticeable way, creating spectral variability that is perfectly noticed just by looking and even more when using computers and algorithms to unscramble this information. Therefore, it's concluded that the method, not only is able to separate samples based on their composition, it does so it with an intrinsically high capability.

3.2. Quantitative test

Quantitative predictions can further strengthen the concept of whole-blood spectra having all the necessary information for composition determination. To compare each method's capacity in quantifying CBC parameters, the SLAI CoVM search algorithm was used to predict the values of RBC, Hgb and HCT on the prediction set and regress them to the measurements done with the Mindray BC-5000 Vet. This is presented in figure 10. Performance parameters from each regression line are presented in table 3. The PoC method shows a strong correlation for all variable, with R^2~0.89 for RBC, R^2~0.91 for Hgb and R^2~0.93 for HCT. MAPE values were all below the maximum total error of 10% for these parameters, as stipulated by the American Society for Veterinary Clinical Pathology [42]. As with previous works, the PoC method shows good predicability of CBC parameters. As for the microscopic method, good correlation was obtained with R^2~0.86 for RBC, R^2~0.84 for Hgb and R^2~0.88 for HCT. The MAPE values rounded the 10%.

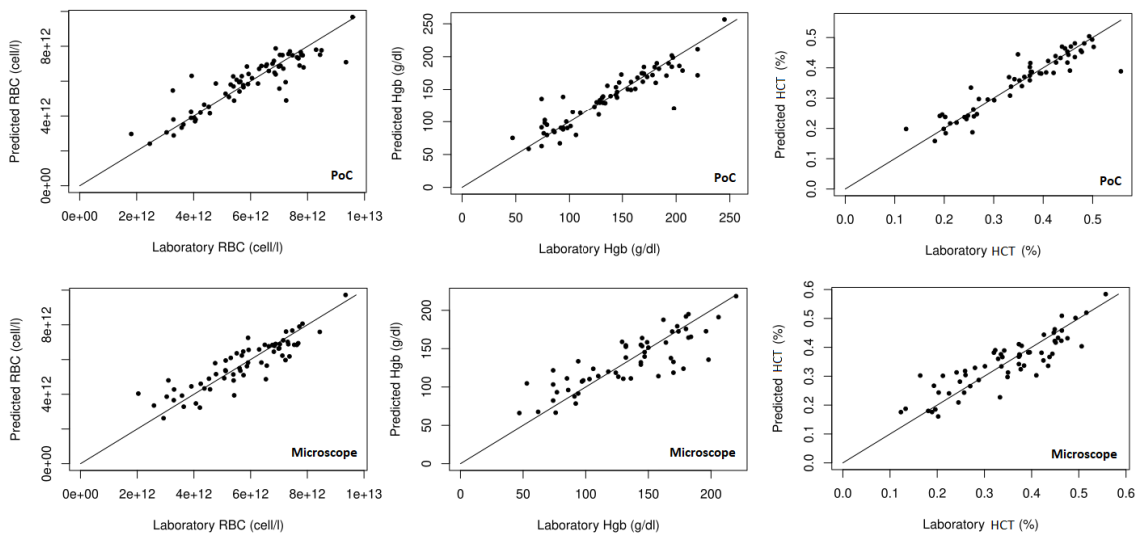


Figure 10: Linear regression results by PoC method and Microscope method. These quantitative predictions where regressed against the values obtained by the Mindray BC-5000 Vet. Good correlation is shown from both methods

Table 3: Performance parameters for PoC method and Microscope method.

Equipment	Parameter	RBC	Hgb	HCT
PoC	R^2	0.8945	0.9141	0.9300
	MAPE	9.62	4.76	4.24
Microscope	R^2	0.8561	0.8350	0.8766
	MAPE	9.10	10.37	9.70

These results show that, it's possible to make quantitative predictions of major blood constituents, with the microscopic method, that are comparable to both standard technology and the PoC device in development. This also implies that the information present at the microscopic level and the macroscopic level is equivalent and therefore possible to correlate with each other.

3.3. Complementarity

Although the information shared between the microscopic and macroscopic levels is equivalent, there are noticeable differences in the performance these levels allow. To begin with, the PCA results obtained with the microscope method get a much higher sample separation, compared to a typical PCA score obtained with the PoC method [13][14][16]. Then, it's also seen that the quantitative predictions obtained with

the PoC device have higher accuracy than the microscope method. One last major difference can be seen by analysing the raw spectra obtained by the two methods.

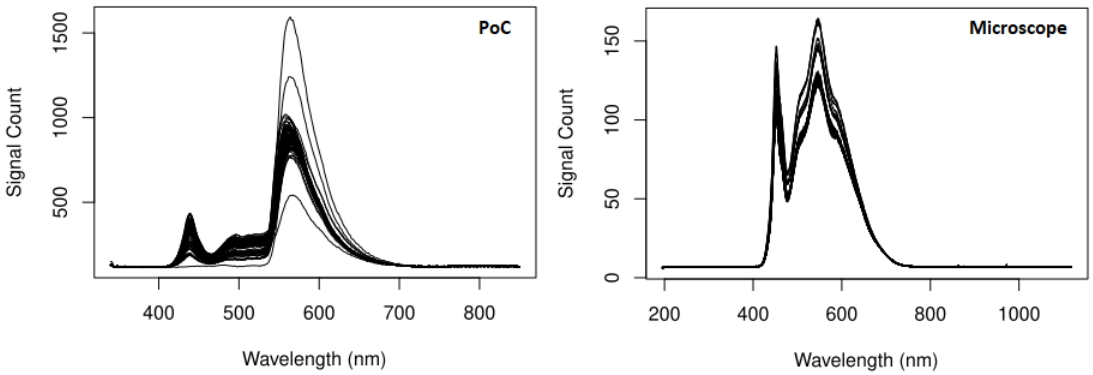


Figure 11: Comparison between PoC spectra and Microscope spectra. The same constituents manifest themselves in a different manner for each method, due to their different views of the sample.

In figure 11, the spectra obtained with the PoC device and with the microscope are presented side by side. Their overall shape is evidently different, as this depends of the light source spectrum characteristics of each device. However, what's important to realize is that, even though the constituents per sample are the same, the way the sample as a whole shapes the spectra on each method is completely different. In the microscopic spectra, it's possible to notice separation among three different sample groups, just like in the PCA. The same is not true for the PoC device, as most spectra are very close together.

The microscopic method looks deep into the blood structure, reducing the light interactions to a single monolayer of blood. This detailed view of the sample, makes each constituent stand out to a greater degree, allowing to notice their presence in the spectrum. For this reason, samples can be separated almost immediately by their spectra and even more after a principal component analysis, or other type of features separation and optimization. However, since this close up look of the sample covers a very small area of the smear, there's very low representativeness of the constituents. This makes quantitative predictions inferior or sometimes even impossible, as it was the case with WBCs.

On the other side, there's the PoC device, which looks to a whole volume of blood (even though small). Light interactions are high, making interferences and matrix effects dominate the spectral behaviour, scrambling the information heavily. This makes both spectra and PCA scores very close together, with only subtle differences,

and therefore hard to separate with simple resources. At the same time, the generalized view of the sample allows for a very good representation of blood composition. The spectra contains all the information needed to quantify constituent concentration. After that, is a matter of implementing the right algorithms to extract that information and get accurate predictions. Optimizing the algorithms, however, is not a simple task.

This analysis leads to an important conclusion: the information obtained with the microscope method and with the Poc device is, simultaneously, equivalent and complementary, advocating for their cooperative use on the development of the technology.

4. Recommendations

Knowledge is of no use if one cannot be critical about it and put it to use. It's important to look back to the thoughts, methods and results developed in this work and reflect upon them. In this chapter, suggestions of improvement and future work are given, to push forward the capabilities of the technology here explored.

4.1. Improvements

The methodology of sample analysis used in the microscopic method is very simple. Blood smears are made for each sample, a monolayer region is found and the spectra of the constituents in a small area is obtained. But, this methodology has two flaws that need to be solved in order to get past the proof of concept point and start using hyperspectral microscopy in the development of the device.

- *Representativeness*: Due to the small area of analysis, there's low representativeness of the blood components, diminishing the quality of the quantitative results or even making it impossible to get any. A solution to this problem would be increasing the area of analysis by creating either a mosaic image or a continuous analysis sweep of the monolayer region of the smear. This would allow the inclusion of a greater count of constituents without decreasing the detailed view that the method gets. Major absorbing components would get better quantitative results and minor components like WBCs and PLTs would be allowed a certain degree of predictability.

- *Blood smear quality*: Blood smear quality can alter the results obtained, as it's directly linked to cell ratios, distribution and density, light interactions and other factors important to make the blood smear represent the whole sample and keep consistency between samples. One way of dealing with this problem is by creating a blood smear protocol and quality standard that can guarantee consistency from sample to sample. This has the drawback of needing specialized personnel and resulting, certainly, in a slower rate of analysis. Another way is by using microfluidics cartridges, like it's already being used by other state-of-art devices, to automatically and consistently create a monolayer of blood that can be then analysed. The drawback is, being expensive in case a high number of samples are analysed.

4.2. Future work

To follow up on this work, it would be interesting and certainly beneficial to study the relationship between the microscopic and macroscopic levels, developing a model that allows predicting the results of one through the other. By creating this mapping, the PoC method could use the high detail of the microscopic method, in an indirect way, to facilitate the unscrambling of information. This will subsequently translate into a greater ease for the SLAI to find the covariance modes and, consequently, to make the best quantitative predictions, as there is a much better isolation of the different spectral characteristics. At the same time, one will get to know better the microscopic spectral behavior of the constituents, which would then make it easier to give a causal interpretation to the results obtained at the macroscopic level.

Conclusion

This research aimed to give a proof of concept on the analytical capabilities of hyperspectral microscopy and how it can be used in the development of a reagent-less point-of-care hematology device with the potential to fulfil the gaps of accessibility seen in both standard and state-of-art tecnologies.

From qualitative analysis of spectra from dog blood smears, obtained through microscopy, it was seen that, on this level, spectral data has high detail information on blood composition, allowing for a very strong differentiation between samples, unlike previously seen by the PoC device. Quantitative results showed that both PoC and microscopy methods are able to accurately predict major absorbing CBC parameters, exhibiting an equivalence in information between the microscopic and macroscopic levels. The better quantitative performance of the PoC method is credited to the more extensive view it has on the sample allowing for greater representativeness. It was seen therefore that the methods share equivalent but complementary information, advocating for their combined use to develop the technology.

To tackle the methodology problems with the microscopy method, a quality standard for blood smears should be implemented. This could perhaps be achieved by creating a smear formation protocol or by using microfluidics cartridges to create the monolayer in an automated and consistent manner. As for the representativeness of the analysed area, this could be solved by creating a mosaic picture or a continuous analysis sweep in order to include a greater number of counts.

For future work, the direct correlation between macroscopic and microscopic spectra could be explored in order to develop a model that allows mapping the results at one level to the other. This map could then be used to join the high differentiability of the microscopic method to the high representativeness of the PoC device, resulting in much better predictions of CBC parameters.

A truly portable and cheap device able to obtain accurate and actionable results from a small amount of blood is paramount to reach the demands of healthcare and veterinary on a global scale. That would translate in more frequent health check-ups for chronically ill patients, easier assessment of health in low resource areas and a faster action on combating the spread of diseases on humans and animals. Hyperspectral microcopy in a shortcut to that objective and the beginning is this.

References

[1] McMichael AJ. The urban environment and health in a world of increasing globalization: issues for developing countries. Bull World Health Organ. 2000;78(9):1117-1126.

[2] Vasan A, Friend J. Medical Devices for Low- and Middle-Income Countries: A Review and Directions for Development. J Med Device. 2020 Mar 1;14(1):010803. doi: 10.1115/1.4045910. Epub 2020 Feb 5. PMID: 32328210; PMCID: PMC7164506.

[3] Guk K, Han G, Lim J, Jeong K, Kang T, Lim E-K, Jung J. Evolution of Wearable Devices with Real-Time Disease Monitoring for Personalized Healthcare. Nanomaterials. 2019; 9(6):813. https://doi.org/10.3390/nano9060813

[4] Farahani, B., Firouzi, F., Chakrabarty, K. (2020). Healthcare IoT. In: Firouzi, F., Chakrabarty, K., Nassif, S. (eds) Intelligent Internet of Things. Springer, Cham. https://doi.org/10.1007/978-3-030-30367-9_11

[5] Briggs, C., Kimber, S. and Green, L. (2012), Where are we at with point- of- care testing in haematology?. Br J Haematol, 158: 679-690. https://doi.org/10.1111/j.1365-2141.2012.09207.x

[6] Buttarello M, Plebani M. Automated blood cell counts: state of the art. Am J Clin Pathol. 2008;130(1):104-116. doi:10.1309/EK3C7CTDKNVPXVTN

[7] Ben-Yosef Y, Marom B, Hirshberg G, D'Souza C, Larsson A, Bransky A. The HemoScreen, a novel haematology analyser for the point of care. J Clin Pathol. 2016;69(8):720-725. doi:10.1136/jclinpath-2015-203484

[8] Zelmer KC, Bauer N, Moritz A. Evaluation of the scil vCell 5, a novel laser- and impedance-based point-of-care hematology analyzer, for use in dogs and cats. J Vet Diagn Invest. 2022;34(3):504-517. doi:10.1177/10406387221083621

[9] Bachar N, Benbassat D, Brailovsky D, et al. An artificial intelligence-assisted diagnostic platform for rapid near-patient hematology. Am J Hematol. 2021;96(10):1264-1274. doi:10.1002/ajh.26295

[10] Gasparin, A.T., Araujo, C.I.F., Schmitt, P. et al. Hilab system, a new point-of-care hematology analyzer supported by the Internet of Things and Artificial Intelligence. Sci Rep 12, 10409 (2022). https://doi.org/10.1038/s41598-022-13913-8

[11] Celkan TT. What does a hemogram say to us?. Turk Pediatri Ars. 2020;55(2):103-116. Published 2020 Jun 19. doi:10.14744/TurkPediatriArs.2019.76301

[12] Barroso TG, Ribeiro L, Gregório H, Santos F, Martins RC. Feasibility of Total White Blood Cells Counts by Visible-Near Infrared Spectroscopy. Chemistry Proceedings. 2021; 5(1):77. https://doi.org/10.3390/CSAC2021-10434

[13] T.G. Barroso, L. Ribeiro, H. Gregório, F. Santos, R.C. Martins, Point-of-care Vis-SWNIR spectroscopy towards reagent-less hemogram analysis, Sensors and Actuators B: Chemical, Volume 343, 2021, 130138, ISSN 0925-4005, https://doi.org/10.1016/j.snb.2021.130138.

[14] Barroso TG, Ribeiro L, Gregório H, Santos F, Martins RC. Visible–Near-Infrared Platelets Count: Towards Thrombocytosis Point-of-Care Diagnosis. Chemistry Proceedings. 2021; 5(1):78. https://doi.org/10.3390/CSAC2021-10435

[15] Barroso, Teresa & Ribeiro, Lénio & Gregório, Hugo & Neves Dos Santos, Filipe & Martins, Rui. (2021). Quantification of Red Blood Cells and White Blood Cells using Scattering Information Extraction from Vis-SWNIR Spectra: Towards Spectroscopy Point-of-Care Cell Counts.

[16] Barroso TG, Ribeiro L, Gregório H, Silva FM, Santos F and Martins RC. Total White Blood Cells Counts by Visible-NearInfrared Spectroscopy Reagent-less Point-of-Care. Submitted to Chemosensors. 2022.

[17] Wright D, Lakos G. THE EVOLUTION OF WHITE BLOOD CELL DIFFERENTIAL TECHNOLOGIES. MAPSS optical cell detection technology. https://hematologyacademy.com/abbotttechnology/mapss-technology/. Published 2022. Accessed October 25, 2022.

[18] DeNicola DB. Advances in hematology analyzers. Top Companion Anim Med. 2011;26(2):52-61. doi:10.1053/j.tcam.2011.02.001

[19] Aysun Adan, Günel Alizada, Yağmur Kiraz, Yusuf Baran & Ayten Nalbant (2017) Flow cytometry: basic principles and applications, Critical Reviews in Biotechnology, 37:2, 163-176, DOI: 10.3109/07388551.2015.1128876

[20] 1.Don M. The Coulter Principle: Foundation of an Industry. JALA: Journal of the Association for Laboratory Automation. 2003;8(6):72-81. doi:10.1016/s1535-5535(03)00023-6

[21] Karakochuk, C.D., Hess, S.Y., Moorthy, D., Namaste, S., Parker, M.E., Rappaport, A.I., Wegmüller, R., Dary, O. and (2019), Measurement and interpretation of

hemoglobin concentration in clinical and field settings: a narrative review. Ann. N.Y. Acad. Sci., 1450: 126-146. https://doi.org/10.1111/nyas.14003

[22] Becker, M., Moritz, A. and Giger, U. (2008), Comparative clinical study of canine and feline total blood cell count results with seven in-clinic and two commercial laboratory hematology analyzers. Veterinary Clinical Pathology, 37: 373-384. https://doi.org/10.1111/j.1939-165X.2008.00085.x

[23] Hu X, Li Y, Xu L, et al. Study on the accuracy of automated hematology analyzers in Shanghai. Southeast Asian J Trop Med Public Health. 2002;33 Suppl 2:1-5.

[24] Kawai Y, Takeuchi K, Shimizu N, et al. Rinsho Byori. 1999;47(4):343-352.

[25] Kumar, T., Ramachandraiah, H., Iyengar, S.N. et al. High throughput viscoelastic particle focusing and separation in spiral microchannels. Sci Rep 11, 8467 (2021). https://doi.org/10.1038/s41598-021-88047-4

[26] Zhou J, Papautsky I. Viscoelastic microfluidics: progress and challenges. Microsyst Nanoeng. 2020 Dec 14;6:113. doi: 10.1038/s41378-020-00218-x. PMID: 34567720; PMCID: PMC8433399.

[27] Dekker, S., Isgor, P.K., Feijten, T. et al. From chip-in-a-lab to lab-on-a-chip: a portable Coulter counter using a modular platform. Microsyst Nanoeng 4, 34 (2018). https://doi.org/10.1038/s41378-018-0034-1

[28] Romen Rodriguez-Trujillo, Oscar Castillo-Fernandez, Miquel Garrido, Martin Arundell, Antoni Valencia, Gabriel Gomila, High-speed particle detection in a micro-Coulter counter with two-dimensional adjustable aperture, Biosensors and Bioelectronics, Volume 24, Issue 2, 2008, Pages 290-296, ISSN 0956-5663, https://doi.org/10.1016/j.bios.2008.04.005.

[29] Grenvall C, Antfolk C, Bisgaard CZ, Laurell T. Two-dimensional acoustic particle focusing enables sheathless chip Coulter counter with planar electrode configuration. Lab Chip. 2014 Dec 21;14(24):4629-37. doi: 10.1039/c4lc00982g. Epub 2014 Oct 10. PMID: 25300357.

[30] Luo, J., Chen, C. and Li, Q. (2020), White blood cell counting at point-of-care testing: A review. ELECTROPHORESIS, 41: 1450-1468. https://doi.org/10.1002/elps.202000029

[31] Sanchis-Gomar F, Cortell-Ballester J, Pareja-Galeano H, Banfi G, Lippi G. Hemoglobin Point-of-Care Testing: The HemoCue System. Journal of Laboratory Automation. 2013;18(3):198-205. doi:10.1177/2211068212457560

[32] Nagy, M., Fazaeli, S., van Oerle, R. et al. Evaluation of the analytical performance of the PC100 platelet counter. Thrombosis J 19, 29 (2021). https://doi.org/10.1186/s12959-021-00283-w

[33] Mohendra Roy, Geonsoo Jin, Dongmin Seo, Myung-Hyun Nam, Sungkyu Seo, A simple and low-cost device performing blood cell counting based on lens-free shadow imaging technique, Sensors and Actuators B: Chemical, Volume 201, 2014, Pages 321-328, ISSN 0925-4005

[34] Christopher P Price, Larry J Kricka, Improving Healthcare Accessibility through Point-of-Care Technologies, Clinical Chemistry, Volume 53, Issue 9, 1 September 2007, Pages 1665–1675, https://doi.org/10.1373/clinchem.2006.084707

[35] Julia E. Mück, Barış Ünal, Haider Butt, Ali K. Yetisen, Market and Patent Analyses of Wearables in Medicine, Trends in Biotechnology, Volume 37, Issue 6, 2019, Pages 563-566, ISSN 0167-7799, https://doi.org/10.1016/j.tibtech.2019.02.001.

[36] Svanberg, S. (2004). Atomic Structure. In: Atomic and Molecular Spectroscopy. Advanced Texts in Physics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-18520-5_2

[37] Svanberg, S. (2004). Molecular Structure. In: Atomic and Molecular Spectroscopy. Advanced Texts in Physics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-18520-5_3

[38] Brian C. Smith, 2 - Single Analyte Analysis, Editor(s): Brian C. Smith, Quantitative Spectroscopy: Theory and Practice, Academic Press, 2002, Pages 43-84, ISBN 9780126503586, https://doi.org/10.1016/B978-012650358-6/50003-8.

[39] Brian C. Smith, 3 - Multiple Components I: Least Squares Methods, Editor(s): Brian C. Smith, Quantitative Spectroscopy: Theory and Practice, Academic Press, 2002, Pages 85-124, ISBN 9780126503586, https://doi.org/10.1016/B978-012650358-6/50004-X.

[40] KÁROLY HÉBERGER, Chapter 7 - Chemoinformatics—multivariate mathematical–statistical methods for data evaluation, Editor(s): Károly Vékey, András Telekes, Akos Vertes, Medical Applications of Mass Spectrometry, Elsevier, 2008, Pages 141-169, ISBN 9780444519801, https://doi.org/10.1016/B978-044451980-1.50009-4.

[41] Brian C. Smith, 4 - Multiple Components II: Chemometric Methods and Factor Analysis, Editor(s): Brian C. Smith, Quantitative Spectroscopy: Theory and Practice,

Academic Press, 2002, Pages 125-179, ISBN 9780126503586, https://doi.org/10.1016/B978-012650358-6/50005-1.

[42] Nabity MB, Harr KE, Camus MS, Flatland B, Vap LM. ASVCP guidelines: Allowable total error hematology. Vet Clin Pathol. 2018;47(1):9-21. doi:10.1111/vcp.12583

[43] HB-801 (2022) HemoCue America. Available at: https://www.hemocue.us/hb-801/ (Accessed: November 15, 2022).

[44] WBC (2022) HemoCue America. Available at: https://www.hemocue.us/wbc/ (Accessed: November 15, 2022).

Attachments

Attachment 1

State-of-Art Devices

Some of the most promising hematology PoC devices are presented next, together with their used technology and some performance parameters.

A few notes to consider while consulting this table:

- The parameters summarized in this table have the purpose of comparison with the device presented in this thesis, both to previous work or the one here developed. For this reason, some values are rounded and only a restricted amount of parameters are presented, corresponding to those tested so far by the team: RBC, Hgb, HCT, WBC, PLT. For an extended list of performance parameters, it is advised the consultation of the documents referenced for each device.

- Each research team or brand uses its own performance parameters and methods. Therefore, the information here contained might no be directly comparable. For a better understanding of the performance parameters meaning and the methodology used to obtain them, the referenced documents should be consulted.

Table 4: Performance parameters of state-of-art devices.

HemoScreen [7]					
Computer vision; AI; Viscoelastic flow cytometry; Disposable microfluidics cartridges					
	RBC	Hgb	HCT	WBC	PLT
Correlation (r)	0.98	0.98	0.98	0.97	0.98
Slope (m)	0.953	0.977	1.009	0.813	1.085
CV (%)	2.83	3.06	3.19	7.36	6.73

Sight OLO [9]					
Computer vision; AI; Disposable LOC cartridges; Passive process of blood monolayer formation					
	RBC	Hgb	HCT	WBC	PLT
Correlation (r)	0.99	0.99	0.98	0.997	0.98
Slope (m)	1.019	1.031	1.030	1.011	1.006
CV (%)	2.2	1.9	2.2	4.3	5.4

Hilab [10]					
Computer vision; AI; Disposable microfluidis cartridges					
	RBC	Hgb	HCT	WBC	PLT
Accuracy (%)	99.3	99.1	98.7	98.0	99.8
Specificity (%)	93.0	99.7	96.3	93.5	99.9
Sensibility (%)	99.7	100.0	98.9	98.6	99.8
CV (%)	4.15	1.3	-	10.97	7.24

HemoCue Hb 801 [43]					
Single parameter device; Spectroscopy; Absorption of HBG in whole-blood					
	RBC	Hgb	HCT	WBC	PLT
Correlation (r^2)	-	0.92	-	-	-
Slope	-	0.96	-	-	-
CV (%)	-	1.0	-	-	-

HemoCue WBC [44]					
Single parameter device; Computer vision; AI; Single use cartridges					
	RBC	Hgb	HCT	WBC	PLT
Correlation (r^2)	-	-	-	0.991	-
Slope (b_1)	-	-	-	1.006	-
CV (%)	-	-	-	5.4	-

PC100 [32]					
Single Parameter Device; “patented optical imaging technology”					
	RBC	Hgb	HCT	WBC	PLT
Correlation (r)	-	-	-	-	0.98
Slope	-	-	-	-	1.002
Percentage Difference in 95% CI	-	-	-	-	-4.82 to 1.31

CMOS [33]					
Computer vision; AI; RBC, WBC and differential					
	RBC	Hgb	HCT	WBC	PLT
Correlation (R^2)	0.99	-	-	0.92	-
Slope	0.943	-	-	0.739	-