

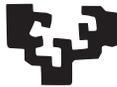
Photonic technology for diagnosis of perinatal asphyxia

Ion Olaetxea Azkarate-Askatsua

Directed by Prof. Andreas Seifert and Prof. Joseba Zubia

A thesis presented for the degree of
Doctor of Philosophy

eman ta zabal zazu



Universidad
del País Vasco

Euskal Herriko
Unibertsitatea

Department of Communications Engineering
University of the Basque Country
April 2022

Resumen

La asfixia perinatal, una de las complicaciones más comunes en partos de alto riesgo, se caracteriza por un intercambio de gases anómalo que da lugar a la falta de oxígeno en tejidos y la correspondiente insuficiencia orgánica. En efecto, un evento prolongado de asfixia puede causar un daño cerebral conocido como encefalopatía hipóxica-isquémica (HIE) que es una de las principales causas de secuelas neurológicas a largo plazo, parálisis cerebral y muerte neonatal.

Según la Organización Mundial de la Salud (OMS), en el año 2017 se produjeron 2.5 millones de muertes neonatales, de las cuales el 24% fueron debidos a complicaciones durante el periodo intraparto como es la asfixia perinatal. Mientras, entre los que sobreviven, se estima que la HIE afecta a 1.5 por cada 1000 a nivel global, aunque este número aumenta de forma considerable en países en vías de desarrollo. En España, diferentes estudios sitúan el ratio de incidencia en torno a 1 por cada 1000, lo que significa que cada año 400 recién nacidos vivos desarrollan HIE.

La monitorización fetal juega un papel decisivo en la identificación y prevención de la asfixia perinatal y, por lo tanto, en las muertes neonatales relacionadas y los consecuentes daños neurológicos, como el HIE. Sin embargo, la efectividad de los métodos de diagnóstico actuales como la cardiotocografía (CTG), o el muestreo de sangre fetal en calota (MSF), han sido puestos en duda en la atención obstétrica ya que no posibilitan un diagnóstico exitoso, unívoco y completamente eficaz, en un proceso

médico tan sensible socio-sanitariamente como es el parto.

Distintos estudios han asociado la CTG, que evalúa el bienestar del feto mediante el registro de la frecuencia cardíaca fetal y las contracciones uterinas, con un aumento de cesáreas y partos instrumentales. A pesar de la recomendación de la Organización Mundial de la Salud (OMS) de que el número de cesáreas no debe superar el 15%, la media mundial se sitúa en torno al 22%, llegando incluso al 40% en algunos países. La OMS también alerta de que una tasa de cesáreas superior al 10% no está asociada con una reducción de la tasa de mortalidad de madres y neonatos, sino todo lo contrario.

El MSF pretende reducir el número de cesáreas innecesarias mediante un procedimiento complementario que respalde el diagnóstico de la asfixia perinatal. El diagnóstico se realiza en base a la medición de pH o lactato en sangre, obtenida por medio de una punción en la calota fetal. Un déficit de aporte de oxígeno induce el metabolismo anaerobio de la glucosa que deriva en acidosis metabólica. A pesar de ser el parámetro de referencia, el pH, por sí solo, no discrimina entre acidosis respiratoria y metabólica. Aunque la determinación del lactato en calota fetal durante el parto ha sido estudiada desde los años 70 y se hayan publicado resultados comparables al pH, su uso no se ha generalizado. En cualquier caso, además de ser un procedimiento incómodo para la madre, el MSF consiste en una medida invasiva, discontinua y con alto índice de fallo, en el que el resultado, que tarda varios minutos en obtenerse, no ofrece una imagen a tiempo real de lo que está sucediendo.

En definitiva, la falta de evidencia y el elevado número de casos neonatales adversos compromete la funcionalidad de los métodos actuales y pone en entredicho su uso clínico para el diagnóstico de la asfixia perinatal. Tanto los profesionales clínicos como los usuarios de los servicios sanitarios demandan un control más riguroso y, al mismo tiempo, menos invasivo en el parto. Sobre todo en partos de riesgo, cuyo porcentaje se ha visto incrementado de manera exponencial en los últimos años debido al aumento de la edad materna o la existencia de cesáreas previas.

Ante esta situación, se observa una clara oportunidad para el desarrollo de un nuevo método de diagnóstico para la monitorización de la asfixia perinatal. En esta tesis se propone una alternativa, basada en espectroscopia Raman, que consiste en la combinación de una sonda específica que se introduciría en la vagina y se colocaría sobre la calota fetal, y algoritmos de aprendizaje automático. La espectroscopia de

Raman aporta información de gran valor mediante la interacción de la luz láser con los distintos tejidos. Los datos espectroscópicos constituyen un espacio multidimensional con información de todos los constituyentes de la muestra bajo prueba, como pueden ser la sangre, la piel, el líquido intersticial, las capas lipídicas, el hueso y/o todos aquellos componentes del cráneo del bebé que pueden ser alcanzados por el haz del láser. En otras palabras, a diferencia del estado del arte donde un único parámetro como el pH o el lactato sirve para la toma de decisión, esta tecnología tiene en cuenta la totalidad de la anatomía accesible por la luz incidente. Mediante modelos predictivos se identifican aquellos patrones característicos de un evento de asfixia. De esta manera, se obtiene un diagnóstico más preciso y fiable mediante una monitorización no-invasiva, continua y a tiempo real de la salud del bebé.

Con el fin de cumplir el objetivo establecido, la tesis se ha desarrollado en base a los siguientes capítulos.

Mientras el **Capítulo 1** introduce el tema y la motivación de la tesis doctoral, el **Capítulo 2** describe la teoría fundamental requerida para la correcta interpretación de los capítulos posteriores.

Capítulo 3. El objetivo de este capítulo es demostrar el gran valor diagnóstico que se puede lograr al considerar múltiples parámetros que describan el cuadro clínico generalizado. Para ello, se ha llevado a cabo un estudio retrospectivo de los datos experimentales obtenidos por Héctor Lafuente durante su tesis doctoral. Con un objetivo distinto al de su tesis, en el siguiente capítulo se investigan parámetros dinámicos obtenidos mediante análisis de sangre a cerdos recién nacidos sobre los que se simula un episodio de asfixia perinatal. En primer lugar, el análisis univariado de los datos nos permite relacionar aquellos parámetros biológicos más relevantes con el grado de asfixia. Posteriormente, se ha analizado de manera cuantitativa el poder predictivo del pH y el lactato como parámetros actuales de referencia ("gold-standard") para la identificación de la hipoxia-isquemia. Además, se propone el uso de algoritmos de aprendizaje automático como nueva estrategia para examinar la correlación y la importancia de los diferentes parámetros en su conjunto. El valor diagnóstico de los modelos predictivos desarrollados ha sido analizado.

Capítulo 4. Este capítulo evalúa la aplicabilidad de la espectroscopia Raman junto el aprendizaje automático para el análisis cuantitativo y cualitativo del pH y del lactato

en fluidos corporales complejos como la sangre o el plasma. Este estudio representa un salto significativo hacia el diagnóstico de la asfixia perinatal, para el cual $\text{pH} < 7.20$ y $[\text{lactato}] > 4.8 \text{ mM}$ son los umbrales establecidos para iniciar una intervención. Para excluir efectos secundarios de un medio complejo, la tecnología ha sido validada primero *in vitro* en muestras acuosas. Se establece un protocolo que considera todo el proceso, desde la preparación de muestras, las mediciones de espectroscopia Raman, hasta la evaluación de los datos. Posteriormente, se ha investigado la variación del pH y el lactato *ex vivo* en muestras de sangre y plasma provenientes de cerdos domésticos.

Capítulo 5. El desarrollo de un prototipo de una sonda Raman específica para una monitorización fetal adecuada supone un reto y juega un papel principal en la futura comercialización de la tecnología. Dada la existencia de un mercado de sondas ópticas amplio y maduro, la experiencia y el conocimiento al respecto es más que evidente. Por ello, como parte de la validación de la tecnología y previo paso a la realización de un estudio preclínico, se ha realizado un evaluación de diferentes sondas Raman comerciales que aporten información relevante y ofrezcan una visión acerca de las características que puedan ser de utilidad en el futuro desarrollo de la sonda. En particular, este capítulo se centra en la comparación de la sonda "EmVision Raman Lensed probe" (EmVision LLC) y de la sonda "RamanProbeTM" (RPS785, Inphotonics).

Capítulo 6. El nuevo método de diagnóstico, que combina la espectroscopia Raman con el uso de algoritmos de aprendizaje automático, es validado *in vivo*. Este capítulo se centra en la fase preclínica en la que se encuentra el proyecto, en el que la sonda RamanProbeTM, previamente descrita en el Capítulo 5, es evaluada para la identificación de un episodio de hipoxia-isquemia. La tecnología es analizada como una alternativa real para una monitorización no-invasiva, continua y a tiempo real que de soporte en la toma urgente de decisiones en el cuidado obstétrico.

Capítulo 7. Esta tesis doctoral está integrada en el marco de un proyecto de investigación con gran capacidad innovadora, ya que surge de una necesidad clínica real e incluye un enfoque impulsado por el mercado. Desde el grado de preparación más bajo, la tecnología ha sido elevada hasta un TRL6 en el marco de esta tesis doctoral. Además de su origen, este capítulo describe los aspectos innovadores, destaca la importancia socio-económica, y ofrece una visión del análisis de mercado y el plan estratégico.

Abstract

Worldwide, the number of newborn deaths counts for 2.5 million, of which 24% are caused by intrapartum complications such as perinatal asphyxia, one of the most common medical disorders in high-risk births. Characterized by insufficient oxygen supply to fetal tissue, a prolonged episode of perinatal asphyxia is responsible for neurological damage known as hypoxic-ischemic encephalopathy. Fetal monitoring plays a major role in the diagnosis of perinatal asphyxia, and consequent neonatal adverse outcomes. However, performance of current standard methods, such as Cardiotocography (CTG) or Fetal Scalp Blood Sampling (FSBS) have been questioned in obstetric care. While CTG has been related with an increase of cesarean sections and instrumental deliveries, pH, or alternatively lactate, obtained from FSBS have demonstrated limited evidence for reduction of cesarean deliveries and adverse neonatal outcome. The method, which operates in an invasive, non continuous and intermittent fashion, does not provide a real picture of the clinical status of the fetus, leading to a high misclassification rate.

The primary goal of this thesis is the development of a non-invasive clinical tool for continuous and real-time monitoring of asphyxia during delivery. The new technology combines Raman spectroscopy, which is a highly specific vibrational spectroscopy method, with machine learning algorithms. Equipped with an application-specific probe, the technology takes into account the systemic picture of physiological variations or anomalies, compared to state-of-the-art, where a single parameter, as pH or lactate, serves as base for decision-making. This results in a much more sensitive and stable

identification of pathological states and prediction of specific parameters that will support in an innovative way immediate medical decision-making.

In order to fulfill the objective, the thesis has been divided in the following chapters.

While **Chapter 1** introduces the theme and the motivation of the present doctoral thesis, **Chapter 2** describes the fundamental theory that is required to correctly interpret subsequent chapters.

Chapter 3. In this chapter, the objective is to demonstrate that higher diagnostic value could be achieved by considering multiple parameters that allow for an overall clinical picture. Thus, a retrospective study is carried out with experimental data obtained by Héctor Lafuente Echeverría in his doctoral thesis. With a different purpose, in the following chapter dynamic parameters from blood gas of newborn piglets undergoing simulated perinatal asphyxia are investigated. In a first step, the most relevant biochemical parameters are related with the severity of the asphyxia state by univariate analysis methods. In a next step, the predictive power of pH and lactate as current gold standards for the identification of hypoxic-ischemic events is quantitatively analyzed. As a new approach, multiparametric machine learning methods are introduced to examine the correlation and importance of different parameters as a combined dataset. With such techniques, new classification models are developed and their diagnostic value is examined.

Chapter 4. In the present chapter, the applicability of Raman spectroscopy together with machine learning is considered for quantitative and qualitative analysis of both pH and lactate in complex body fluids, such as blood and plasma. This approach represents a significant step toward diagnosis of perinatal asphyxia, where $[\text{lactate}] > 4.8 \text{ mM}$ and $\text{pH} < 7.20$ are defined as standards for intervention. To exclude secondary effects of complex media, the technology is firstly validated *in vitro* with a set of aqueous samples. A protocol has been developed for the determination of physiological pH and lactate, considering the entire process chain from sample preparation over Raman measurements to data evaluation. Following the protocol, the variation of pH and lactate *ex vivo* in blood and plasma samples from domestic pigs is then investigated.

Chapter 5. Prototyping and development of hardware components of a specific Raman probe for accurate fetal monitoring represents a challenge and plays a major role in the future commercialization of the technology. Given the existence of a mature

and wide market of optical probes, valuable practical and technical expertise is already available. Therefore, as part of the validation of the technology and prior to conducting a preclinical study, an evaluation of different commercial Raman probes has been performed to give insight into essential characteristics our customized Raman probe will require. Particularly, this chapter describes the comparison of the "EmVision Raman Lensed probe" (EmVision LLC) and the "RamanProbe™" (RPS785, Inphotonics).

Chapter 6. In this chapter, the new diagnostic tool, which combines Raman spectroscopy with machine learning, is validated *in vivo*. In this pre-clinical phase, the RamanProbe™ described in Chapter 5 is tested for identification of hypoxic-ischemic events. The technology is evaluated as a genuine alternative for supporting in an innovative way immediate medical decision-making by non-invasive, continuous and real-time monitoring of perinatal asphyxia.

Chapter 7. This doctoral thesis is embedded in a research project with high innovative power, as it originates from real clinical requirements and includes from the beginning market-driven needs and real-life applications into its research. Starting at lowest technology readiness level (TRL), the technology was elevated to TRL6 within the framework of this doctoral thesis. Apart from its origin, this chapter describes the innovative aspects, highlights the socio-economic importance, and gives insight into market analysis and strategic planning.

Abbreviations

AB	adaptive boosting
ADP	adenosine diphosphate
aEEG	amplitude integrated electroencephalogram
ALS	asymmetric least squares
ATP	adenosine triphosphate
AUROC	area under the receiver operating characteristic curve
BC	basal condition
CBF	cerebral blood flow
CO ₂	carbon dioxide
CP	phosphocreatine
CTG	cardiotocography
ECG	electrocardiogram
EMSC	extended multiplicative signal correction
FADH ₂	reduced flavin adenine dinucleotide
FHR	fetal heart rate
FSBS	fetal scalp blood sampling

GB	gradient boosting
GTP	guanosine triphosphate
H_2CO_3	carbonic acid
H_2O	water
HCO_3^-	bicarbonate
HI	hypoxia-ischemia
HI _{p1}	immediate post-hypoxia-ischemia
HI _{p2}	interval 15-30 min after hypoxia-ischemia event
HI _{p3}	60 min after hypoxia-ischemia event
HI _p	post-hypoxia-ischemia
HIE	hypoxic-ischemic encephalopathy
LOOCV	leave-one-out cross-validation
LOSOCV	leave-one-subject-out cross-validation
NA	numerical aperture
NAD ⁺	oxidized nicotinamide adenine dinucleotide
NADH	reduced nicotinamide adenine dinucleotide
NH ₄ ⁺	ammonium
NPV	negative predictive value
O ₂	oxygen
PBS	phosphate-buffered saline
PCA	principal component analysis
pCO ₂	partial pressure of carbon dioxide
PLS	partial least squares
PLS – AB	partial least squares – adaptive boosting
PLS – DA	partial least squares – discriminant analysis
PLS – GB	partial least squares – gradient boosting
PLS – RF	partial least squares – random forest
PLS – SVM	partial least squares – support vector machine

PLS – XGB	partial least squares – extreme gradient boosting
pO ₂	partial pressure of oxygen
PPV	positive predictive value
R ²	coefficient of determination
RF	random forest
RMSEP	root mean square error of prediction
SNR	signal-to-noise ratio
SVM	support vector machine
TRL	technology readiness level
WHO	world health organization
XGB	extreme gradient boosting

Contents

Resumen	iv
Abstract	vii
Abbreviations	xi
1 Introduction	1
1.1 Motivation	1
1.2 State-of-the-art	2
1.3 Novel approach	4
2 Theoretical background	7
2.1 Metabolic pathways	8
2.1.1 Phosphocreatine or CP-ATP system	8
2.1.2 Anaerobic metabolism	8
2.1.3 Aerobic metabolism	9
2.2 Acid-base balance	10
2.3 Hypoxic-ischemic encephalopathy	12
2.4 Raman spectroscopy	13
2.5 Data preprocessing	14
2.5.1 Artifacts in Raman signal	15
2.5.2 General preprocessing methods	16
2.6 Machine learning	16
2.6.1 Model construction and optimization	17

2.6.2	Model evaluation	18
2.6.3	Regression algorithm	20
2.6.4	Classification algorithm: Single classifier	20
2.6.5	Classification algorithm: Ensemble learning	20
3	Physiological changes induced by hypoxia-ischemia	23
3.1	Material & Methods	24
3.1.1	Sample preparation and data acquisition	24
3.1.2	Statistical analysis and classification	25
3.2	Results	27
3.2.1	Transient induction of hypoxia-ischemia	27
3.2.2	Physiological parameters from blood gas analysis	27
3.2.3	Development of multiparametric predictive models	36
3.3	Discussion	40
3.4	Conclusion	42
4	pH and lactate sensing in body fluids	43
4.1	Material & Methods	44
4.1.1	Customized Raman spectroscopy system	44
4.1.2	Sample preparation and acquisition of Raman spectra	44
4.1.3	Preprocessing of Raman spectra	46
4.1.4	Multivariate analysis	47
4.2	Results	48
4.2.1	pH and lactate monitoring in aqueous solutions	48
4.2.2	pH and lactate monitoring in blood	53
4.2.3	Lactate monitoring in blood plasma	59
4.2.4	Evaluation of predictive models	62
4.3	Conclusion	62
5	Optimization of a Raman probe system	63
5.1	Comparison of commercial probes	64
5.1.1	Optical and mechanical characteristics	64
5.1.2	Measurement performance of the Raman probes	66
5.2	<i>In vivo</i> validation of the probe system	69
5.2.1	Optimization of the focal position	69
5.2.2	Tissue differentiation	70
5.2.3	Pre- & post-mortem differentiation	71

6	<i>In vivo</i> non-invasive monitoring of hypoxia-ischemia	75
6.1	Material & Methods	76
6.1.1	Experimental procedure	76
6.1.2	Data acquisition: blood gas analysis and Raman spectra	77
6.1.3	Preprocessing of Raman spectra	79
6.1.4	Data analysis	80
6.2	Results	82
6.2.1	Induction of hypoxia-ischemia	82
6.2.2	pH and lactate as gold standard parameters	82
6.2.3	Predictive models based on Raman spectra	88
6.3	Discussion	97
6.4	Conclusion	102
7	Technology transfer	103
7.1	Origin of the project	104
7.2	Proposed diagnostic procedure	104
7.3	Innovation & impact	105
7.3.1	Major innovation	105
7.3.2	Social impact	106
7.4	Market opportunity	107
7.5	Roadmap	108
	Final summary	112
	Dissemination of results	114
	Acknowledgment	116
	Bibliography	117

CHAPTER 1

Introduction

1.1 Motivation

Perinatal asphyxia, which is one of the most common complications in high-risk births, is a serious medical disorder, characterized by abnormal gas exchange rates that result in oxygen deprivation in fetal tissue during labor and, ultimately, in organ failure [1]. Prolonged perinatal asphyxia is responsible for brain damage, known as hypoxic-ischemic encephalopathy (HIE), which may cause long-term neurological sequelae, cerebral palsy, and neonatal death [1, 2].

According to the World Health Organization (WHO), 24% of the estimated 2.5 million neonatal deaths in 2017 were due to intrapartum complications, such as birth asphyxia [3]. Although the HIE incidence varies from study to study, it is considered that it affects on average 1.5 per 1000 live births worldwide [4]. This number increases considerably in preterm infants or in developing countries with poor obstetric care [2, 4]. In Spain, the incidence rate stands at approximately 1 per 1000 live births, meaning that every year around 400 newborns develop HIE [5].

1.2 State-of-the-art

Intrapartum fetal monitoring plays a major role in the identification and prevention of perinatal asphyxia, and hence, related deaths or neurological impairments, such as HIE. Assessment of fetal heart rate (FHR) is widely used as indicator of fetal oxygen deprivation and adverse fetal state through continuous electronic monitoring by cardiotocography (CTG), as described in Figure 1.1 [6, 7].

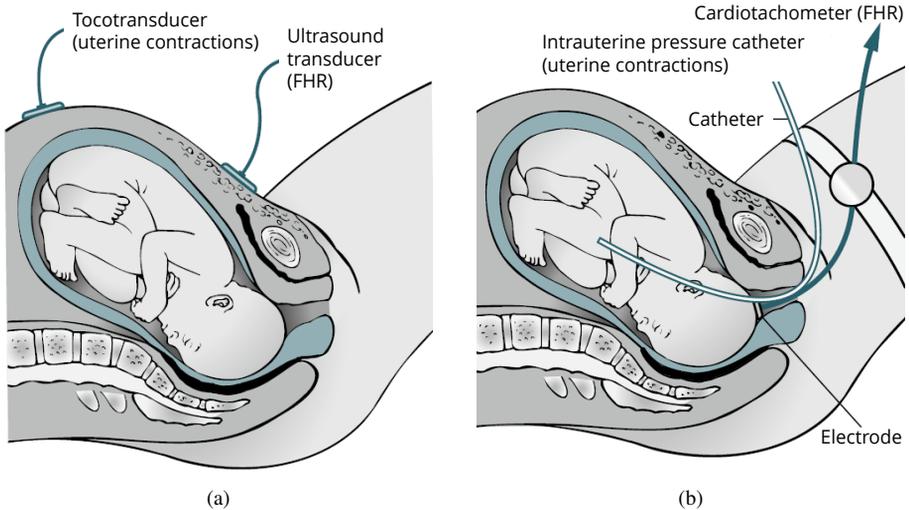


Figure 1.1: **(a)** External non-invasive cardiotocography. A Doppler ultrasound transducer is placed on the maternal abdomen to monitor the fetal heart rate. A tocotransducer or tocodynamometer is placed above the umbilicus to record the frequency and the duration of uterine contractions. **(b)** Internal invasive cardiotocography. Monitoring of fetal heart rate is obtained by attaching a spiral electrode to the fetal scalp. The leg plate on the thigh of the mother avoids electrical interference. An intrauterine pressure transducer is introduced in the uterus of the mother to record the frequency, the duration and the intensity of uterine contractions. Figures are adapted from [8].

FHR can be externally recorded, by placing a Doppler ultrasound transducer on the maternal abdomen, or internally, by attaching an electrode to the fetal scalp. Despite providing a more reliable and accurate surveillance, the rupture of the amniotic membranes is required for internal recording of FHR. Simultaneous information about the frequency and the duration of uterine contractions is obtained by a second pressure transducer placed below the uterine fundus. Likewise, this second transducer can also be substituted by an intrauterine pressure catheter [7, 9]. Reassuring or non-reassuring

CTG patterns prompt the clinical procedure to follow. A non-reassuring CTG can represent an adaptive response of the fetus to an asphyxia episode. However, considering the variability between individual fetal responses and the inter-observer disagreement on a CTG interpretation, non-reassuring patterns do not always reflect a fetal distress [10]. The effectiveness of FHR to identify perinatal asphyxia is actually questioned due to an increasing number of cesarean sections and operative deliveries, which are often unnecessary [6, 11]. Despite WHO's recommendation to not exceed a rate of 15% cesarean sections, the worldwide average is around 22%, sometimes reaching 40% [12], with no concomitant reduction of maternal and neonatal mortality [13, 14]. A gradual trend to advanced maternal age increases high-risk births and the probability of clinical complications and adverse neonatal outcomes [15, 16].

Prolonged asphyxia is responsible for anaerobic metabolism and consequently of metabolic acidosis. Analysis of intradermal pH by means of fetal scalp blood sampling (FSBS) is commonly used for fetal surveillance as an additional test to avoid unnecessary interventions. Blood gas analysis is the predominant tool for identifying acid-base imbalances that reflect the current state of the newborn. In practice, blood pH measurements entail high failure rates and often lead to misclassification of fetal clinical condition, which has drawn clinical interest to the use of lactate as an alternative indicator [11, 17, 18]; lactate is overproduced during anaerobic metabolism. Lactate testing presents adequate properties to be successfully undertaken as it allows for faster sampling of smaller blood volumes, which are 2-5 μL for lactate compared to 30-60 μL for pH. However, there is limited or no evidence that pH or lactate FSBS reduces the number of cesarean deliveries or the adverse neonatal outcomes [7, 11, 18].

Moreover, abnormal values of blood pH and lactate may have multiple underlying mechanisms [19]; thus, a diagnostic test based on fluctuations of a single parameter poses an evident risk. Higher diagnostic value could be achieved by considering multiple parameters that allow for an overall clinical picture. Continuous collection and availability of a multitude of clinical information represents a unique opportunity for improved personalized assistance. However, handling, integrating, and understanding this complex information is still a major challenge in healthcare. Machine learning algorithms, based on multivariate analysis of clinical data, have demonstrated high potential for statistical analysis and identification of abnormal patterns [20, 21]. Computer-aided systems that assist obstetricians in fetal surveillance by correctly visualizing

1 and interpreting the CTG [22, 23, 24] or postnatal electroencephalographic signal [25] have already been reported with different degrees of success and performance.

Current clinical diagnostics, as blood gas analysis, voltammetric-based or liquid chromatography detection methods, require arterial blood sampling and operate in an invasive and non-continuous manner. Moreover, these techniques require significant time for analysis, are quite expensive, expose healthcare professionals to patients' blood, and result in iatrogenic blood loss [26]. Considering these limitations, an effort has been put on developing continuous perinatal asphyxia, or alternatively pH and lactate, monitoring systems to allow for robust decision-making and timely clinical interference in obstetric care medicine. The capabilities and strengths of biophotonic methods distinctly showcase the importance that spectroscopy gains in medical diagnostics [27, 28]. Although some attractive methods based on microwave sensors [29] or near-infrared spectroscopy-based sensors [30] for non-invasive monitoring of pH and lactate have been proposed, so far none of them has shown enough predictive accuracy or reliability in clinical use.

1.3 Novel approach

Raman spectroscopy, as a promising and emerging photonic technique for diagnostics, has become of general interest in medical research. The inelastic scattering of the incident photons due to specific vibrational modes of individual molecules provides a chemical fingerprint of the sample being studied, hence, delivering highly specific and diagnostically valuable information. The great ability to non-invasively provide high molecular selectivity makes Raman spectroscopy an ideal method for chemical quantification in medical diagnostics [31]. Moreover, Raman instrumentation is sufficiently mature for studying the human organism and physiological parameters. In fact, several studies have already demonstrated that Raman spectroscopy could be used for rapid biomedical analysis *ex vivo* as well as for continuous monitoring *in vivo* [32, 33, 34].

The interpretation of Raman spectra is far from being straightforward. Due to the generally weak Raman signal and high fluorescence of biomolecules by nature, as well as the overlapping of spectral features from different constituents present in a sample, peak identification and accurate prediction constitute a serious challenge. Nevertheless, the combination of Raman spectroscopy with a proper sequence of data

preprocessing methods and machine learning algorithms results in a very powerful technology for unveiling hidden features. Machine learning algorithms have the capacity to extract valuable knowledge for building up predictive models to be applied to new data sets. The potential of such algorithms for quantification and classification of complex Raman spectroscopy signals has been demonstrated extensively [35]. Although such a combination has already been used to analyze pH and lactate for different purpose, as for example food control [36] or sensing of aqueous solutions [37, 38], clinical requirements are still not fulfilled, particularly not for obstetric approach.

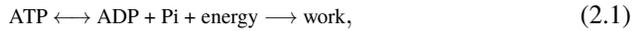
The primary goal of this endeavor is the development of a new diagnostic tool for monitoring perinatal asphyxia. Our technology represents a photonic non-invasive approach that works in continuous mode, delivering real-time diagnostics. We combine Raman spectroscopy, featured by specifically tailored probes, with machine learning as powerful diagnostic tool. The method is not limited to examine a single parameter as pH or lactate, it considers the entirety of physiological parameters that are accessible by Raman spectroscopy, and hence, delivers a systemic picture measured on molecular level. Moreover, we look in real-time at the development of physiology, thus, at the pathological severity. Continuous monitoring of a variety of biochemical parameters allows for the identification of physiological anomalies caused by perinatal asphyxia, and supports in an innovative way immediate medical decision-making.

Theoretical background

For a correct interpretation of the thesis, some fundamental theory needs to be explained. Hence, this chapter summarizes the most important concepts of the different fields addressed throughout the document. First, physiology-related notions, such as metabolic pathways, acid-base equilibrium and hypoxic-ischemic encephalopathy, are analyzed to better understand the cause and the consequences of perinatal asphyxia. The basis of Raman spectroscopy as a photonic biosensing technique is then described, Finally, the importance of data preprocessing and the concept of machine learning for development of predictive models are addressed.

2.1 Metabolic pathways

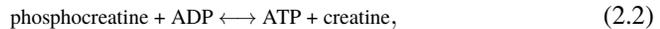
The human body requires a constant supply of energy to guarantee a proper functioning of the organism. The hydrolysis of adenosine triphosphate (ATP) to adenosine diphosphate (ADP) is the so-called “energy currency” of the body. Breakdown of a phosphate group from ATP,



provides the fuel for physiological activity. However, the availability of ATP is limited to few seconds. Thus, other sources of energy and inorganic phosphate (P_i) are required to guarantee quick and continuous regeneration of ATP in working cells. Nutrients, such as carbohydrates, fats and proteins are broken down into carbon dioxide (CO_2) and water (H_2O) molecules while energy is ultimately transferred to the activated energy carrier molecule ATP. The selection of the type of fuel, and therefore the metabolic pathway, varies depending on the physiological circumstances. Three main metabolic pathways are distinguished [39].

2.1.1 Phosphocreatine or CP-ATP system

Phosphocreatine (CP) is the first source of energy when the cell runs out of ATP. Stored in muscles, only one step is needed to separate the phosphate group from the creatine

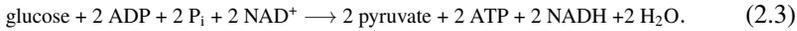


enabling a rapid regeneration of ATP. Due to the reversibility of the equation from CP into creatine, the CP-ATP system is fundamental in tissues with high fluctuations of energy demand such as the brain [40, 41]. Despite its high potential to transfer a high-energy phosphate group, CP is shortly supplied and is therefore able to sustain the regeneration of ATP only for a few seconds [42].

2.1.2 Anaerobic metabolism

The oxidation of each glucose molecule into two smaller molecules of pyruvate is called glycolysis. Electrons are transferred from glucose to the coenzyme nicotinamide adenine dinucleotide, on its oxidized form NAD^+ , generating two NADH (H for hydrogen) high-

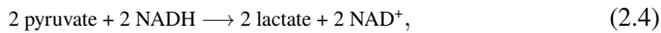
energy electron carriers without the involvement of oxygen (O_2)



This sequence of reactions, known as substrate-level phosphorylation, occurs in the cytosol, an intracellular fluid where organelles are suspended, and produces two molecules of ATP [39].

Fermentation: Excretion of lactate

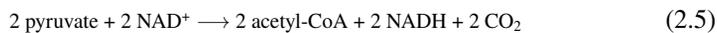
In anaerobic conditions, pyruvate and NADH remain in the cytosol. NADH gives up its electrons to pyruvate to produce lactate



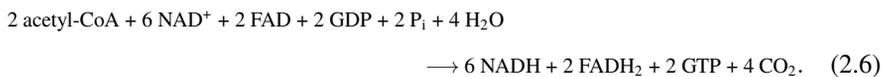
while regeneration of NAD^+ maintains the glycolytic activity as the principal source of ATP at low levels of oxygen. This energy pathway is called fermentation [39]. Even though it has long been considered a waste or even a harmful end-product, recent studies confirmed that lactate plays a major role in distinct metabolic pathways [43, 44].

2.1.3 Aerobic metabolism

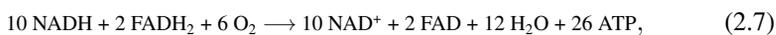
Under aerobic conditions, ATP synthesis is not at substrate level but in an organelle called mitochondria. Pyruvate enters the mitochondria



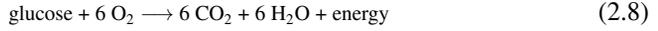
and is converted into CO_2 , NADH and acetyl-CoA, where most of the energy is preserved. Oxidation to CO_2 and further energy carrier molecules, as guanosine triphosphate (GTP), reduced flavin adenine dinucleotide (FADH_2) or NADH, continues with a series of reactions known as Krebs cycle or citric acid cycle, summarized as



Energy retained in carrier molecules are finally used to produce ATP by transferring high-energy electrons to oxygen within the electron transport chain. The generation of ATP with involvement of oxygen



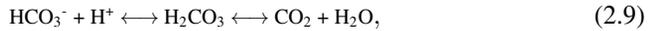
which occurs in the mitochondrial membrane, is termed oxidative phosphorylation. Unlike anaerobic respiration, where only a throughput of two ATP molecules is achieved, complete oxidation of a glucose molecule under aerobic conditions is much more efficient and produces about 30 ATP molecules in total.



Moreover, other fuels as fats and proteins can also be converted and oxidized by intermediate steps of aerobic metabolic pathway [39].

2.2 Acid-base balance

When the limit of the mechanisms responsible for maintaining pH homeostasis is exceeded, physiological acid-base equilibrium is disrupted. Minor variations in the normal extracellular pH range – 7.35 - 7.45 – may have irreversible consequences. For this reason, accurate blood gas interpretation is a challenge to identify metabolic disorders [45]. Both intracellular and extracellular buffer systems can provide a defense response to pH variations. Even though phosphate and proteins are also used for intracellular buffering, the bicarbonate (HCO_3^-)\carbonic acid (H_2CO_3) system



is the primary solution of the interstitial fluid surrounding the cells. In addition to buffering both acidic and alkaline medium, fluctuations in hydrogen concentration are independently regulated by kidneys and lungs [46, 47].

Respiratory system Sensitive chemoreceptors to pH or CO_2 , controlling the alveolar ventilation, are activated to increase or decrease the tidal volume and the respiratory rate. Consequently, pCO_2 is adjusted and pH restored to normal values. The concentration of carbonic acid determines the direction of the process [46, 47].

Renal system Regulation of bicarbonate by the kidneys also maintains the acid-base equilibrium and allows for pH imbalance correction. A decrease of pH activates the reabsorption of filtered bicarbonate and the production of new bicarbonate by acid excretion into urine in the form of titratable acids (urinary buffers) and

ammonium NH_4^+ . Conversely, an increase of pH leads to reduction of the reabsorbed bicarbonate, which is excreted through the urine. Unlike rapid respiratory system, renal compensation is a slow process [46, 47].

When mechanisms responsible for acid-base regulation are overwhelmed, body pH homeostasis is impaired. If prolonged, acid-base disorders might lead to severe consequences. In this study, only disorders related to a decrease in blood pH are addressed. Two types of acidosis are distinguished.

Respiratory acidosis When CO_2 is accumulated in blood, pH is reduced as the equation 2.9 leans towards the left, increasing the concentration of hydrogen ions. Acidosis is related to any disorder interfering with alveolar ventilation [17, 45].

Metabolic acidosis Characterized by a deficiency of bicarbonate concentration, acidosis might be motivated by different reasons; an overproduction or a decreased excretion of acid load, and renal or gastrointestinal bicarbonate loss [17, 45].

Acid-base equilibrium is restored either by correction, modifying the system that is directly affected by a disturbance, or by compensation, triggering an alternative system as a response to the disturbance (Table 2.1).

Table 2.1: Characteristics of respiratory and metabolic acidosis disorders.

Type of acidosis	Consequence	Correction	Compensation
Respiratory	CO_2 accumulation	Reduce CO_2 by increasing tidal volume and respiratory rate.	Increase HCO_3^- reabsorption and production. Increase acid excretion.
Metabolic	HCO_3^- reduction	Increase HCO_3^- reabsorption and production. Increase acid excretion.	Reduce CO_2 by increasing tidal volume and respiratory rate.

2.3 Hypoxic-ischemic encephalopathy

2 Birth is often related to a more or less pronounced asphyxia event, accompanied by a number of physiological changes. Even though a low-oxygen environment is common in intrauterine conditions, with fetal partial pressure of oxygen (pO_2) close to 25-30 mmHg, the fetus has additional adaptive mechanisms when a severe oxygen deficiency takes place. Reduction of the heart rate, decrease of fetal motility, such as respiratory movements, and redistribution of blood flow, mainly towards the heart and the brain, are considered self-protective behaviors against temporary deprivation of nutrients and oxygen. A reduced metabolic demand is pursued to counterbalance a decreased energy supply. However, if the impairment is prolonged in time, perinatal asphyxia might result in hypoxic-ischemic encephalopathy (HIE) and consequent neurological disorder or death [48, 49].

HIE refers to a brain damage characterized by a deficient cerebral blood flow during the perinatal period. This impairment is mainly caused by an interruption of the placental blood circulation (ischemia) and subsequent reduced cerebral blood flow, or an inadequate oxygen supply at tissue level (hypoxia). Mostly, a combination of both is observed [50, 51].

The pathophysiology of HIE is characterized by three distinct phases after the hypoxic-ischemic insult. A primary energy failure begins at the acute phase, characterized by an ATP depletion caused by oxygen deficit. Efficiency of the anaerobic metabolic pathway in terms of ATP is considerably reduced, as explained in Section 2.1. Moreover, failure of the aerobic metabolism leads to overproduction of lactate and accumulation of hydrogen ions, mainly consumed in the oxidative phosphorylation process. The combination of an increased lactate production and a reduced blood pH caused by tissue hypoxia is called Type A lactic acidosis [43], the most common cause of metabolic acidosis. Apart from monitoring acidosis, reduction of blood pH has also been considered a contributing adaptive response to metabolic arrest, due to its inhibitory effect in enzymatic activity. When correct brain perfusion is restored, a latent period followed by a secondary energy failure, associated with augmented seizure activity, occurs. Interventions to avoid further damage can be carried out during the latency phase, lasting about 6 hours. The degree of recovery right after the hypoxic-ischemic insult plays a major role and is reflected both in the latency period and in the brain

damage provoked by the secondary phase of injury. A tertiary phase involves long-term consequences during the months or even years after the HI event [51, 52].

2.4 Raman spectroscopy

Raman spectroscopy, which belongs to the family of vibrational spectroscopic techniques, spectrally resolves inelastically scattered photons. Vibrational techniques are widely used to provide fundamental information about molecular structure in a rapid, straightforward, reproducible and non-invasive manner. Moreover, the negligible sample preparation makes it ideal for study of biological samples. Particularly, Raman spectroscopy relies on the interaction of radiation, in the form of monochromatic light, with specific vibrational modes of molecules [53, 54].

When a sample is irradiated with a laser, both Rayleigh and Raman scattering processes are involved as depicted in Figure 2.1, the latter being a much rare phenomenon with an occurrence around 1 in 10^8 . The molecule that absorbs an incident photon is excited to a virtual energy state. If the energy of the re-emitted photon remains unchanged after the collision with the molecule, the process is known as elastic scattering or Rayleigh scattering. However, loss or gain of energy provoked by the interaction between the molecule and the photon, results in a frequency shift in the scattered light, the so-called Raman effect. Depending on their initial vibrational state, molecule excitation will give raise to Stokes (loss of energy) or anti-Stokes Raman scattering (gain in energy) [53, 54].

The energy shift is characteristic of vibrational modes associated with specific functional groups, thus providing valuable information about the sample to be analyzed. Therefore, the Raman spectrum, depicted as the scattered intensity of the Raman shifts from the incident wavelength, represents a distinctive biological fingerprint [31]. The intensity of the Raman signal is described by,

$$I_R \propto \nu^4 I_0 N \left(\frac{\partial \alpha}{\partial Q} \right)^2, \quad (2.10)$$

where ν is the frequency of the incident beam, I_0 is the intensity of the incident beam, N represents the number of Raman-active molecules, and $\left(\frac{\partial \alpha}{\partial Q} \right)$ indicates the change in the polarizability of the molecules as a function of a vibrational mode [53].

The above equation unveils the relevance of selecting a particular laser frequency

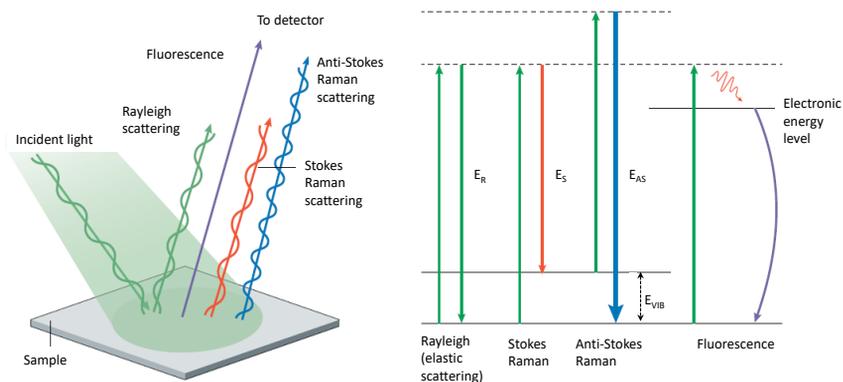


Figure 2.1: Excitation of the sample molecules with an incident laser results in different types of interactions. Momentary absorption of a photon excites the molecule from a ground state into a virtual state. Transition from this virtual state generates a new scattered photon. Incident photons might be reflected with the same energy E_R (Rayleigh or elastic scattering) or can alternatively lose E_S or gain E_{AS} some energy (Raman or inelastic scattering). Raman scattering coming from a molecule that is initially in a ground vibrational state is called Stokes Raman, while scattering coming from molecules in an excited state is known as anti-Stokes Raman. When the energy of the incident photon matches an electronic energy level, the molecule is excited to a higher level. As some energy is dissipated as heat or vibrations, the wavelength of the re-emitted photon is higher. This process is called fluorescence. Figure is adapted from [55].

for the sample excitation, resulting in a significant improvement of the signal at smaller wavelengths. However, high-energy laser (small wavelength) sources can produce strong fluorescent background and sample photodamage yielding disturbance of the Raman signal. To avoid such effects, lasers emitting at 785 nm, 830 nm and 1064 nm are extensively used for biological studies, as they additionally provide a higher penetration depth into the sample [31, 56].

2.5 Data preprocessing

The complexity of chemical molecules in addition to the diversity of biological samples makes the analysis and correct interpretation of measurements challenging. Therefore, a number of different data preprocessing steps of raw Raman spectra are necessary to reliably extract underlying chemical and structural information from the samples and mitigate noise and unwanted background.

2.5.1 Artifacts in Raman signal

Cosmic ray features

Detectors of Raman spectroscopy systems are generally sensitive to cosmic rays that sporadically produce spike artifacts in the raw spectra. Although these spikes are easy to distinguish by eye, it is not that simple to eliminate them without modifying adjacent parts of the spectra [57]. A simple technique to remove cosmic ray features consists in taking the median of several measurements from the same sample, due to the low probability of another cosmic ray appearing at exactly the same wavenumber in consecutive spectra.

High frequency noise

Raman spectra are often affected by high frequency artifacts, that need to be filtered out. Shot, read or dark current noise generated during collection and conversion of photons by the charge-coupled device (CCD) are common sources of noise [58]. Savitzky-Golay filtering is an appropriate low-pass filtering technique based on window polynomial fitting. The smoothing of spectra is performed by successively fitting a polynomial to a subset of data points by least-squares [57, 59].

Fluorescent background

Intrinsic fluorescence of molecules, often more intensive than the Raman peaks themselves, can strongly impede a correct analysis of the sample. A corrected Raman signal can be obtained by subtracting a smoothed background signal from the spectra by means of the asymmetric least squares (ALS) method [60]. Unlike ordinary least square, ALS gives priority to positive residual errors to guarantee the integrity of Raman peaks [57].

Multiplicative artifacts

Physical events, as light scattering, can also cause multiplicative effects or scattering artifacts, such as fluctuating background. Even though such artifacts disruptively affect the interpretation of Raman spectra, extended multiplicative signal correction (EMSC) allows us to get rid of those varying interfering features. Corrected spectra are obtained

by scaling all measurements to the mean spectrum of the whole data set. EMSC extends linear regression to multiple regression with additional non-linear terms to adequately model the chemical variability of the spectra [61].

Variance uncertainty

Since most of multivariate analysis methods are based on the variance of specific features of data, it is important to center the data relative to a reference point, like the mean. Translating the origin of the coordinate system to the center of gravity of the data avoids uncertainties of the variance of the mean [62].

2.5.2 General preprocessing methods

Principal component analysis (PCA) is a mathematical method used for data compression while retaining most of the information. The dimensionality of the data is reduced by projection into a new set of orthogonal independent variables, called principal components, which are linear combinations of the original variables. Principal components are arranged in descending order according to the direction of greatest variance, providing insight into the amount of information they contain [57, 62].

PCA can also be used for exploratory analysis, but due to the lack of any prior knowledge about the data, its discrimination and quantification power are limited [57].

Standardization addresses the difference in the order of magnitude between variables. Scale disparity directly affects the performance of a classifier. Units are corrected by standardization so that each variable is equally considered, resulting in an improved discrimination in certain circumstances. For standardization, each variable is mean-centered and divided by its standard deviation [63].

2.6 Machine learning

Machine learning consists in computational processes designed to emulate human capacity to learn and process information from the surrounding in order to perform a particular task. This is achieved by exploring the dependencies of observed features

and fitting the structure of the data into models which are then understood and applied by end-users for new unknown inputs. In recent years, machine learning has been successfully applied in many different fields, like finances, entertainment, preference identification, autonomous vehicles, psychology or biomedical applications [64, 65].

Based on the nature of the learning procedure, machine learning is classified into supervised, unsupervised, semi-supervised or reinforcement learning. In the case of supervised learning, input data is paired with a label. Relationship between the label and the independent variables allows for identification of the distinctive characteristics despite the similarities that could exist between classes (classification) or conditions (regression). Unlike supervised learning, unsupervised algorithms do not have an associated label but rather identify clusters based on patterns that occur with some regularity. Semi-supervised learning is the combination of preceding learning techniques. It addresses applications where only a minor size of the data is labeled. Many real-world processes where labeling might be an intensive or even limited task fit in semi-supervised learning. Information from labeled data is used in the learning procedure of the unlabeled part. Reinforcement learning is based on the interaction of a learning agent with the surrounding environment. Sequence of actions are rewarded based on trial and error scenario. Considering the consequence of previous actions, reinforcement learning interprets the environment to maximize the reward and achieve an optimal action [64, 65].

Machine learning algorithms have brought the ability to carry out tedious and time-consuming tasks more efficiently. Moreover, the capacity to handle huge amount of complex data allows for identification of subtle patterns that are negligible to human appreciation. The potential of such algorithms for quantification of complex Raman spectroscopy signals has been demonstrated extensively [35].

2.6.1 Model construction and optimization

Machine learning algorithms are able to learn from input data through repeated experience. This learning and adaptation process is known as training. Available data is commonly divided into a training and testing set. Even though there is not an established rule, test samples often represent around $1/3 - 1/4$ of all data, depending on the data size [63, 64]. The predictive model is built and calibrated with the training set, while the

testing set is used to evaluate how well the model generalizes and to quantify its performance. Generalization is a concept that refers to the ability of the model to perform well not only with the training set but also with new or unseen data [64, 65]. Cross-validation enables the optimization of model parameters. This optimization technique divides the training set into k folds, where each of the folds is used for validation only once and one at a time while the remaining $k-1$ folds are used for training. Parameters are tuned based on best cross-validated performance [64, 66].

2.6.2 Model evaluation

A predictive model is evaluated based on its yield for predicting unknown data. The predictive power is calculated according to statistical metrics [65, 67]. The main goal of a regression model is to find the best fitting curve that maps the input data with the continuous output. The root mean square error of prediction (RMSEP) calculates the deviation of predicted values from reality [65], defined as,

$$\text{RMSEP} = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n-1}}, \quad (2.11)$$

where \hat{y}_i is the predicted output, y_i is the real value and n is the number of samples. This quantity is used as an indicator of the performance of the predictive model.

Classification problems seek the decision boundary that maximizes the distinction among classes or discrete values. Hence, measuring RMSEP brings no added value. Alternatively, a confusion matrix, see Table 2.2, records the assignment of data to a particular category: true negative (TN), true positive (TP), false negative (FN) or false positive (FP). The evaluation of model performance is represented by scalar metrics calculated based on these records [67, 68].

Table 2.2: Outcome of correct and incorrect predictions sorted by a confusion matrix or contingency table.

		Predicted	
		Negative (0)	Positive (1)
Actual	Negative (0)	TN	FP
	Positive (1)	FN	TP

Some prominent figures of merit of such confusion matrix are listed subsequently.

Sensitivity or recall refers to the proportion with a positive condition that are predicted

as positive by the model. It is defined as,

$$\text{Sensitivity} = \frac{TP}{TP + FN}. \quad (2.12)$$

Specificity refers to the proportion without a positive condition that are predicted as negative by the model. It is defined as,

$$\text{Specificity} = \frac{TN}{TN + FP}. \quad (2.13)$$

Positive predictive value (PPV) or precision refers to the proportion predicted as positive by the model, that really has a positive condition. It is defined as,

$$\text{PPV} = \frac{TP}{TP + FP}. \quad (2.14)$$

Negative predictive value (NPV) refers to the proportion predicted as negative by the model, that does not have a positive condition. It is defined as,

$$\text{NPV} = \frac{TN}{TN + FN}. \quad (2.15)$$

Accuracy refers to the ratio between correct predictions and the total number of cases to be analyzed. It is defined as,

$$\text{Accuracy} = \frac{TP + TN}{P + N}, \quad (2.16)$$

where P and N stand for all real positive and negative cases, respectively. The sum of both values represents the total number of predictions.

F-measure also refers to prediction accuracy. Defined as,

$$F_1 = 2 \cdot \frac{\text{precision} \cdot \text{recall}}{\text{precision} + \text{recall}}, \quad (2.17)$$

F-measure is the harmonic average of precision and recall. Unlike accuracy, F-measure gives more importance to FN and FP being useful for imbalanced datasets.

Area under receiver operating characteristic (AUROC) curve measures the separability between classes as a probability that a randomly chosen positive example is ranked higher than a randomly chosen negative example. The ROC curve is represented by plotting sensitivity against 1 - specificity for different discrimination thresholds.

2.6.3 Regression algorithm

Partial least squares (PLS) is a powerful multivariate regression technique used for supervised analysis and for the development of predictive regression models for chemical quantification. The main goal of PLS is the prediction of a set of responses Y , as a matrix of outcomes, from a set of predictors X , the matrix of observations. Similar as in PCA, PLS extracts a new set of orthogonal components, called latent vectors, in the direction of maximum variance that simultaneously decomposes the data. Unlike PCA, PLS additionally maximizes the covariance between between X and Y at the same time [69, 70].

2.6.4 Classification algorithm: Single classifier

Partial least squares – Discriminant analysis (PLS – DA) is an extension of PLS. Despite being initially introduced for regression to handle continuous variables, PLS also evolved for classification purposes. In PLS – DA, output categorical variables are recoded into discrete numbers, typically 0 and 1. Due to the inherent function of PLS, resulting predictions rather take values close to these two integers. Class membership is determined according to an arbitrary cut-off point [71].

Support vector machine (SVM) separates positive and negative samples by finding a hyperplane that maximizes the margin or distance between the hyperplane and data points. The use of a Kernel function enables distinction of linearly non-separable data by transforming data to a higher dimensional space [72].

2.6.5 Classification algorithm: Ensemble learning

The objective of ensemble learning is to combine weak or base learners with low predictive performance to build much stronger and accurate models. The combination of weak learners helps to minimize the variance of a single model prediction.

Bagging (bootstrap aggregating) consists in generating multiple subsets, i.e. randomly sampling with replacement of the original training dataset. A weak learner is simultaneously and independently applied to each subset and trained in parallel. Final classification is obtained by voting, that is aggregating all individual predictions.

Random forest (RF) fits multiple uncorrelated deep trees to bootstrapped subsets. These trees are then combined to provide a more consistent prediction. Apart from sampling over observations when building individual trees, the random selection of features guarantees a lower variance, i.e. overfitting, and hence, a more robust model [73].

Boosting algorithms represent another alternative of ensemble learning. Although the principle is the same as bagging, in this case, weak learners are sequentially trained with multiple subsets in an iterative manner. Special attention is given to observations that have been misclassified by previous learners.

Adaptive boosting (AB) redefines the weight of the samples at each iteration. Thus, distribution of samples is changed by putting a larger weight on wrong samples. Final classification is obtained by voting of weak learners' predictions, which are also weighted based on their individual performance [74].

Gradient boosting (GB) gives importance to misclassified data by fitting a new learner to the residuals of the predecessor at each iteration. Final classification is done through a gradient descent optimization to minimize the loss function [74, 75].

Extreme gradient boosting (XGB) is an optimized implementation of Gradient boosting. Performance and efficiency of the algorithm are improved by employing parallel computing, regularization terms and more accurate approximations [76].

Physiological changes induced by hypoxia-ischemia

In this chapter, the objective is to demonstrate that higher diagnostic value could be achieved by considering multiple parameters that allow for an overall clinical picture. Experimental data is the result of the doctoral thesis of Héctor Lafuente Echeverría [77] and serves as base for a retrospective study, which was not subject of his thesis. In the following chapter, dynamic parameters from blood gas of newborn piglets undergoing simulated perinatal asphyxia are investigated. In a first step, the most relevant biochemical parameters are related with the severity of the asphyxia state by univariate analysis methods. In a next step, the predictive power of pH and lactate as current gold standards for the identification of hypoxic-ischemic events is quantitatively analyzed. As a new approach, multiparametric machine learning methods are introduced to examine the correlation and importance of different parameters as a combined dataset. With such techniques, new classification models are developed and their diagnostic value is examined.*

*Published as: Hector Lafuente, Ion Olaetxea, Ana Valero, Francisco Jose Alvarez, Ander Izeta, Ibon Jaunarena, and Andreas Seifert. Identification of hypoxia-ischemia by chemometrics considering systemic changes of the physiology. *IEEE Journal of Biomedical and Health Informatics*, 2022, DOI:10.1109/JBHI.2022.3142190

3.1 Material & Methods

3.1.1 Sample preparation and data acquisition

22 piglets, 1 to 3 days old, were obtained from a local vendor (Arri-Turri farm, Alava, Spain) with pathogen-free status. Piglets were anesthetized with a perfusion of phen-tanil, propofol, and midazolam in dextrose 5% (0.004, 3, and 0.5 mg/kg/h, respectively) and paralyzed by a perfusion of vecuronium (3 mg/kg/h) administered through an ear vein. Animals were then intubated, and positive pressure ventilated (Bourns BP200, CA). Continuous three-lead electrocardiogram (ECG) was recorded. The right common carotid flow (Qcar) was measured by an ultrasonic flow probe (Transonic Systems Inc., NY), as a representation of cerebral blood flow (CBF). The femoral artery was cannulated to monitor blood pressure (Ominare CMS24; HP, Göbblingen, Germany) and to obtain blood samples. Blood oxygen saturation was monitored by transcutaneous pulse oximetry. Furthermore, brain activity was monitored using a two channel bed amplitude integrated electroencephalogram (aEEG) monitor (BRM2, BrainZ Instruments, Auckland, New Zealand). Rectal temperature was maintained between 37.5 and 38.5°C with heating lamps. Hypoxia-ischemia (HI) was induced by clamping both carotid arteries with vascular occluders and lowering the fraction of inspired oxygen (FiO₂) to 8 - 10% over 20 minutes. HI was considered for 20 minutes of the flat traces of aEEG (< 4 μV) (grade V). After HI, carotid blood flows were restored, and the inspired fraction of oxygen was returned to 21%. At the end of the experiment, the piglets were euthanized with an intravenous injection of potassium chloride.

Table 3.1 shows the 18 physiological parameters that were measured (ABL5, Radiometer, Copenhagen, Denmark) from arterial blood gas samples of newborn piglets. The database analyzed consists of 98 measurements divided in two subsets, normoxia or basal condition (BC group, $n = 49$) and post-hypoxia-ischemia (HI_p group, $n = 49$). In the BC group, blood gas analysis was randomly performed during 30 minutes of stabilization after intubation. The HI_p group was formed by three subsets: immediate post-hypoxia-ischemia (HI_{p1}, $n = 22$), at the interval 15 - 30 minutes after hypoxia-ischemia event (HI_{p2}, $n = 17$) and 60 minutes after hypoxia-ischemia (HI_{p3}, $n = 10$). This partitioning was included to study the ability of both, gold standard methods and proposed models, to distinguish an HI event even during the recovery process. Considering

the lower probability to detect an HI event as the recovery time passes, the number of samples per interval were included accordingly.

Table 3.1: Physiological parameters of newborn piglets measured by blood gas analysis.

T (°C) (37 - 39.6)	[Na ⁺] (mM) (118 - 153)	COHb (%) (0 - 4.1)
pH (7.07 - 7.49)	[K ⁺] (mM) (118 - 153)	MetHb (%) (0 - 4.1)
[Lactate] (mM) (0.7 - 12.9)	[Ca ²⁺] (mM) (0.66 - 1.45)	HHb (%) (-3.9 - 84.1)
pCO ₂ (mmHg) (31 - 57)	[Glucose] (mM) (45 - 159)	SO ₂ (%) (5.5 - 99.1)
[HCO ₃ ⁻] (mmHg) (11.5 - 30.4)	[tHb] (g/dL) (4.5 - 11.1)	[Base excess] (mM) (-16.8 - 5.1)
pO ₂ (mmHg) (9 - 140)	O ₂ Hb (%) (15.1 - 100)	Hematocrit (%) (14 - 33)

3.1.2 Statistical analysis and classification

Univariate analysis

The most basic form of data analysis studies single variables with the objective to describe specific attributes of the data. Box-and-whisker plots were used to depict data distribution. Non-parametric Mann-Whitney U-test was applied to determine statistical significance among groups, while Spearman's correlation is used to measure the monotonic relationship between physiological parameters and the clinical state. Finally, predictive power of lactate and pH for a binary diagnosis of HI was examined by means of statistical metrics derived from a confusion matrix, including sensitivity, specificity, accuracy, positive and negative predictive value (PPV, NPV), and the area under the receiver operating characteristic curve (AUROC). Definition of perinatal asphyxia was based on clinical cut-off limits of pH < 7.20 and [lactate] > 4.8 mM, as proposed by Saling [79] and Kruger et al. [80], respectively.

Multivariate analysis

Multivariate analysis simultaneously evaluates various parameters and identifies existing correlations. Machine learning algorithms, based on multivariate analysis, are

characterized by their capacity to learn and adapt to new situations.

A 5-fold cross-validation was carried out for the development of different classification algorithms. Data was randomly stratified into 5 groups, where each of the groups was used for testing only once and one at a time. Predictive models were optimized by selecting the AUROC as scoring metric for evaluation. Thus, the separability between classes was maximized and consequently, the probability that a randomly chosen positive example was ranked higher than a randomly chosen negative example was also maximized. Despite random stratification of the two subsets BC ($n = 49$) and HI_p ($n = 49$), the variability of the HI_p group could have affected the model construction and hence, its performance. To avoid possible bias induced by this random designation of the folds, the cross validation was repeated 20 times for every classifier. The figures of merit from the cross validated prediction were then averaged and compared. The performance of different classifiers for a binary diagnosis of HI was evaluated. The correlation between the HI state and the clinical parameters was also determined by analyzing their contribution to the predictive models. Permutation importance technique calculates the decrease in a metric of a fitted model when a single feature is randomly shuffled. The difference reveals how decisive a physiological parameter is for generating a prediction.

A number of different classification algorithms – as single classifier, ensemble learning – were used for the detection of HI. Specifically, (i) partial least squares – discriminant analysis (PLS – DA) [71]; (ii) random forest (RF) [73]; (iii) support vector machine (SVM) [72]; (iv) adaptive boosting (AB) [74]; (v) gradient boosting (GB) [74, 75]; and (vi) extreme gradient boosting (XGB) [76] have been tested. Fundamental concepts of applied algorithms can be found in more detail in Section 2.6.

Finally, hybrid classifiers were tested. They combine partial least squares, as feature extractor for dimensionality reduction, with classification algorithms. Prior to feeding machine learning algorithms for training, redundant or irrelevant features were removed by projecting the data in a new reduced subset of latent vectors sorted by the amount of explained variance and extracted from combinations of the original features. The classifier was then constructed in this new latent space [81, 82].

To guarantee meaningful graphical representations with no loss of any relevant information, PLS was applied twice. First, the optimum number of latent vectors that

minimizes the prediction error was found by internal cross validation. PLS was performed for a second time to this new projected data such that most of the variance was captured and illustrated by the first two components. All data analysis was performed in Spyder, a Python open source integrated development environment. Sklearn, Scipy, Pandas and Numpy libraries were mainly used for the analysis and development of predictive models, while Seaborn and Matplotlib were applied to display plots and figures.

3.2 Results

3.2.1 Transient induction of hypoxia-ischemia

The effects on the brain during the transiently induced HI insult were documented by recording the aEEG trace (Figure 3.1). Abnormal aEEG background pattern or a sustained isoelectric EEG ($< 4 \mu\text{V}$) during the first postnatal hours is highly predictive of adverse outcome. Besides, electroencephalographic traces displayed the occurrence of seizures in 90% of the piglets. After unclamping both common carotid arteries (immediately postinjury), a reactive and transient hyperaemia to the brain was shown by the normalization of the carotid blood flow, remaining normal afterwards. In relation to the respiratory function, animals were similarly ventilated at same tidal volume.

3.2.2 Physiological parameters from blood gas analysis

Inducing HI causes a shift in several physiological parameters from BC, as shown in Figure 3.2. The Mann-Whitney U-test quantifies this alteration by unveiling a statistically significant difference for all measured parameters except for Na^+ , MetHb and Hct, (Table 3.2). As expected, pH and lactate present significant variation between BC and HI_p groups. Specifically, pH decreases from 7.37 ± 0.05 in BC to 7.2 ± 0.10 after HI, and more significant, lactate increases from 1.37 ± 0.46 in BC to 5.56 ± 2.50 after HI.

The analysis of individual parameters in early and late HI_p phase is illustrated in Figure 3.3. HI_p data, split into HI_{p1} - HI_{p3} subsets, show that the parameters related to blood oxygenation (SO_2 , pO_2 , O_2Hb) exhibit rapid recovery to initial values when ventilation is restored, making them unreliable for clinical diagnosis (Table 3.3, Table 3.4, Table 3.5). In contrast, base excess, pH, and lactate show a much slower recovery of

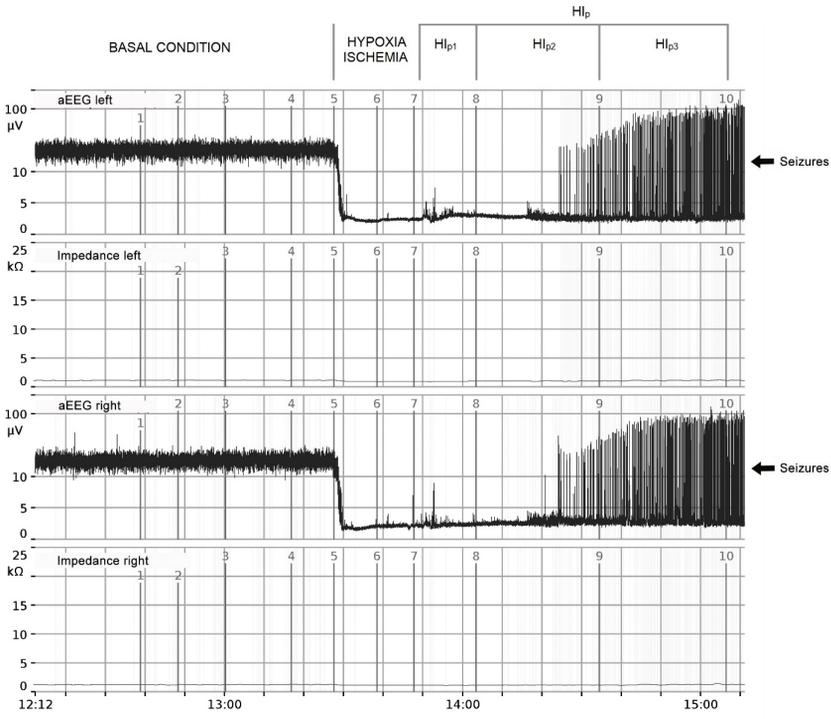


Figure 3.1: Representative brain electroencephalogram and impedance from a hypoxic-ischemic animal. Amplitude-integrated EEG (aEEG) and impedance during basal condition (BC), immediate post HI event (HI_{p1}), at the interval 15-30 minutes after HI event (HI_{p2}), and 60 minutes after HI event (HI_{p3}). Arrows at the right indicate the appearance of seizures.

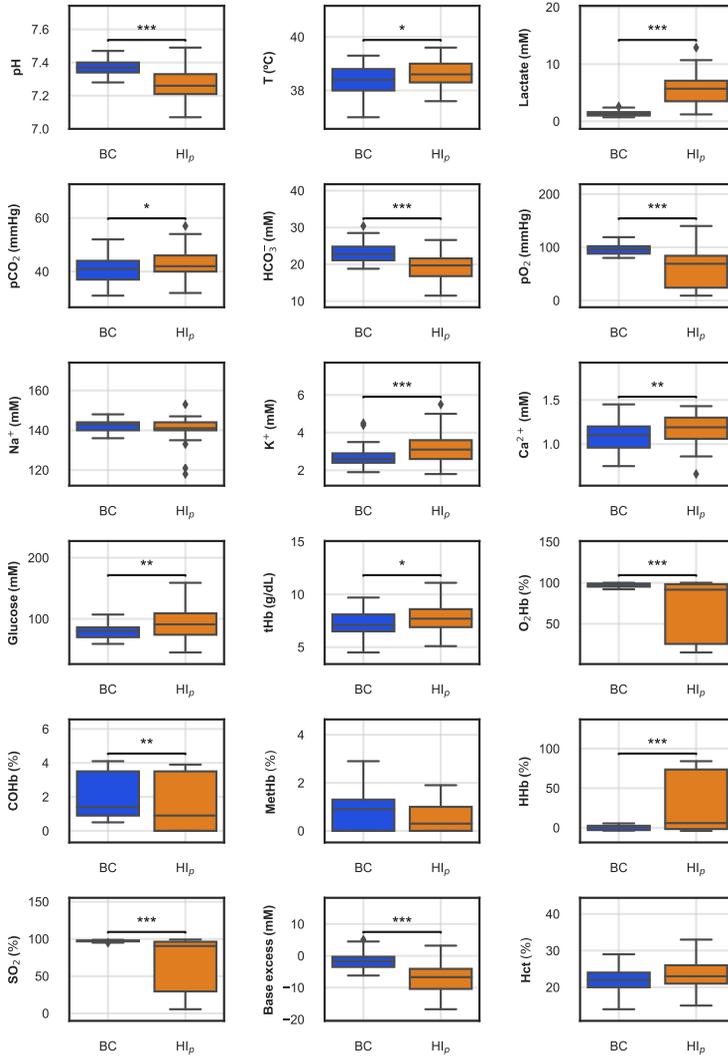


Figure 3.2: Exploratory analysis of basal condition (BC) and post-hypoxia-ischemia (HI_p) groups for different physiological parameters. Box-and-whisker plot facilitates the comparison between BC and HI_p groups by displaying the distribution and skewness of the data.

Mann-Whitney U-test
 * 0.01 < p-value ≤ 0.05
 ** 0.001 < p-value ≤ 0.01
 *** p-value ≤ 0.001

Table 3.2: Statistical significance of physiological parameters between basal condition (BC) and post-hypoxia-ischemia (HI_p) determined by Mann-Whitney U-test.

Parameter	U-statistic	p-value
Lactate	57.5	<0.001*
SO ₂	338.5	<0.001*
pH	383.5	<0.001*
Base excess	397.0	<0.001*
pO ₂	399.0	<0.001*
HCO ₃ ⁻	448.0	<0.001*
HHb	636.5	<0.001*
O ₂ Hb	642.5	<0.001*
K ⁺	746.5	<0.001*
Glucose	800.5	<0.01*
COHb	856.0	<0.01*
Ca ²⁺	863.5	<0.01*
T	927.0	0.03*
tHb	940.5	0.03*
pCO ₂	958.5	0.04*
Hct	971.5	0.051
MetHb	992.0	0.06
Na ⁺	1100.5	0.24

Mann-Whitney U-test

* p-value ≤ 0.05

preinjury levels, with strongest evidence in lactate even at late HI_p phases. Lactate levels did not even reach basal state levels in HI_{p3}, which is 1 hour after the HI event, while pH demonstrated already normal conditions (Table 3.6). Strongest relationship is presented between lactate and clinical diagnosis by a Spearman correlation coefficient of 0.82; other parameters that show good inverse correlation with an HI event are pH, base excess, SO₂, pO₂, HCO₃⁻ and O₂Hb, as displayed in Table 3.7.

The predictive power of current pH and lactate cut-off limits to identify HI is surprisingly limited (Figure 3.4). The poor outcome is due to the little capacity to identify positive cases resulting in an insufficient sensitivity of 24.5% for pH and 59.2% for lactate (Table 3.8). However, separability between groups, as quantified by the AUROC values (84% for pH, 97.6% for lactate), shows that lactate provides a much better classification performance than pH.

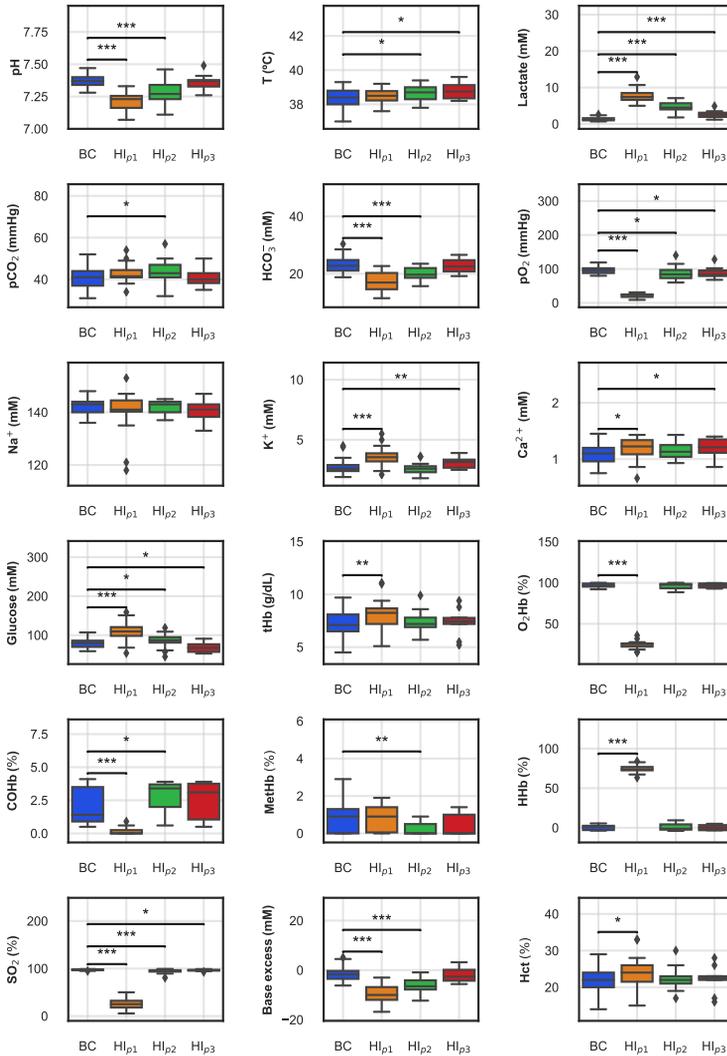


Figure 3.3: Exploratory analysis of basal condition (BC) and post-hypoxia-ischemia subsets (HI_{p1} , HI_{p2} , HI_{p3}) for different physiological parameters. Box-and-whisker plot facilitates the comparison between the groups by displaying the distribution and skewness of the data.

Mann-Whitney U-test
 * $0.01 < p\text{-value} \leq 0.05$
 ** $0.001 < p\text{-value} \leq 0.01$
 *** $p\text{-value} \leq 0.001$

Table 3.3: Statistical significance of physiological parameters between basal condition (BC) and immediate post-hypoxia-ischemia (HI_{p1}) determined by Mann-Whitney U-test.

Parameter	U-statistic	p-value
SO ₂	0.0	<0.001*
HHb	0.0	<0.001*
Lactate	0.0	<0.001*
O ₂ Hb	0.0	<0.001*
pO ₂	0.0	<0.001*
COHb	15.5	<0.001*
pH	27.0	<0.001*
Base excess	43.0	<0.001*
HCO ₃ ⁻	80.0	<0.001*
Glucose	147.0	<0.001*
K ⁺	160.5	<0.001*
tHb	345.5	<0.01*
Hct	360.5	0.01*
Ca ²⁺	373.5	0.02*
pCO ₂	430.0	0.09
T	481.0	0.24
MetHb	483.0	0.24
Na ⁺	485.5	0.25

Mann-Whitney U-test

* p-value ≤ 0.05

Table 3.4: Statistical significance of physiological parameters between basal condition (BC) and 15-30 minutes post-hypoxia-ischemia (HI_{p2}) determined by Mann-Whitney U-test.

Parameter	U-statistic	p-value
Lactate	10.0	<0.001*
Base excess	124.5	<0.001*
HCO ₃ ⁻	136.5	<0.001*
pH	140.0	<0.001*
SO ₂	182.5	<0.001*
MetHb	214.0	<0.01*
pO ₂	261.5	0.01*
COHb	281.0	0.02*
pCO ₂	281.0	0.02*
Glucose	294.5	0.04*
T	303.5	0.05*
Ca ²⁺	332.5	0.11
K ⁺	365.0	0.23
tHb	379.5	0.30
Hct	388.0	0.34
O ₂ Hb	407.5	0.45
HHb	408.5	0.46
Na ⁺	415.0	0.49

Mann-Whitney U-test

* p-value ≤ 0.05

Table 3.5: Statistical significance of physiological parameters between basal condition (BC) and 60 minutes post-hypoxia-ischemia (HI_{p3}) determined by Mann-Whitney U-test.

Parameter	U-statistic	p-value
Lactate	47.5	<0.001*
K ⁺	118.0	<0.01*
Glucose	131.0	0.01*
pO ₂	137.5	0.02*
T	142.5	0.02*
SO ₂	156.0	0.04*
Ca ₂₊	157.5	0.04*
MetHb	183.0	0.10
Na ⁺	199.5	0.18
COHb	201.5	0.19
HHb	212.0	0.26
tHb	215.5	0.28
pH	216.5	0.29
O ₂ Hb	217.0	0.29
Hct	223.0	0.33
Base excess	229.5	0.38
HCO ₃ ⁻	231.5	0.40
pCO ₂	242.5	0.48

Mann-Whitney U-test

* p-value ≤ 0.05

Table 3.6: Mean ± standard deviation of pH and lactate of basal condition (BC) and post-hypoxia-ischemia (HI_p) subsets (HI_{p1}, HI_{p2}, HI_{p3}).

	BC	HI _p	HI _{p1}	HI _{p2}	HI _{p3}
pH	7.369 ± 0.046	7.262 ± 0.096*	7.210 ± 0.072*	7.271 ± 0.091*	7.361 ± 0.061
[Lactate]	1.37 ± 0.46	5.6 ± 2.5*	7.6 ± 1.8*	4.6 ± 0.4*	2.70 ± 0.99*

Mann-Whitney U-test

* p-value ≤ 0.001

Table 3.7: Spearman's correlation between physiological parameters and a hypoxia-ischemia event.

Parameter	Coefficient	p-value
Lactate	0.83	<0.001*
HHb	0.41	<0.001*
K ⁺	0.33	<0.001*
Glucose	0.29	<0.01*
Ca ²⁺	0.24	0.02
T	0.20	0.051
tHb	0.19	0.06
pCO ₂	0.18	0.08
Hct	0.17	0.10
Na ⁺	-0.07	0.48
MetHb	-0.15	0.13
COHb	-0.25	0.01*
O ₂ Hb	-0.40	<0.001*
HCO ₃ ⁻	-0.54	<0.001*
pO ₂	-0.58	<0.001*
Base excess	-0.58	<0.001*
pH	-0.59	<0.001*
SO ₂	-0.62	<0.001*

Spearman's correlation

* p-value ≤ 0.05

Table 3.8: Classification performance of pH and lactate standard cut-off limits for detection of hypoxia-ischemia (HI) events; all figures of merit are in percentage (%).

	pH < 7.2	[Lactate] > 4.8 mM
Sensitivity	24.5	59.2
Specificity	100	100
PPV	100	100
NPV	57.0	71.0
Accuracy	62.2	78.0
AUROC	84.0	97.6

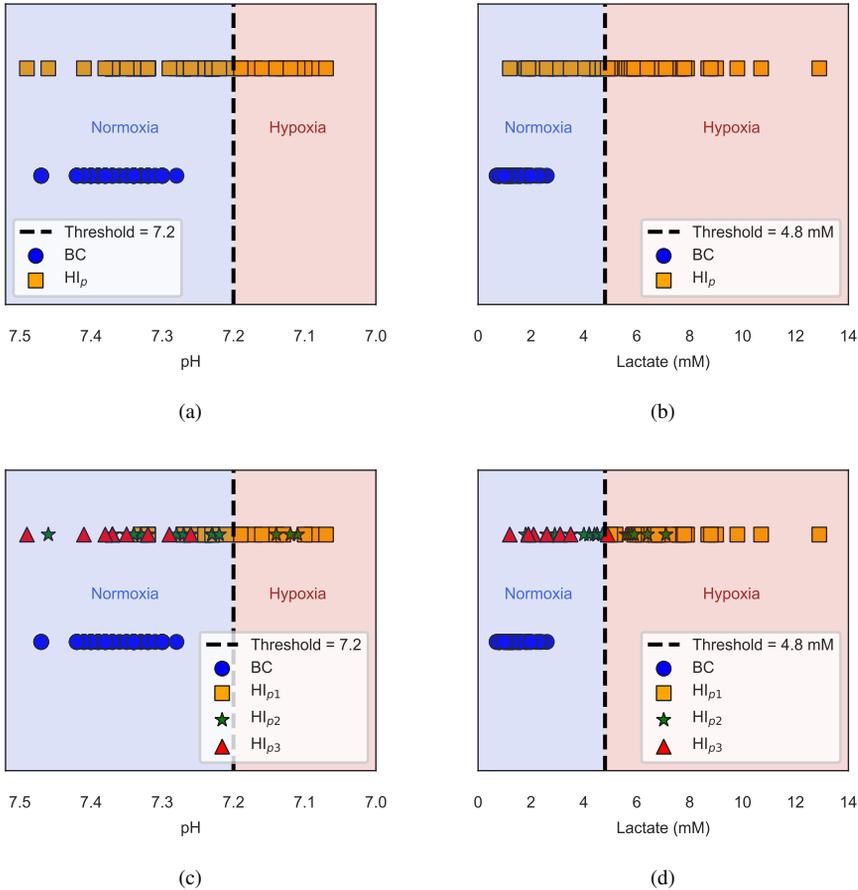


Figure 3.4: 1-D visual class separability between basal condition (BC) and post-hypoxia-ischemia (HI_p) based on standard clinical cut-off limits. **(a)** Decision region of binary classification (BC vs. HI_p) based on pH threshold = 7.2. **(b)** Decision region of binary classification (BC vs. HI_p) based on [lactate] threshold = 4.8 mM. **(c)** **(d)** Disaggregation of post-hypoxia-ischemia (HI_p) in a) and b) into time-dependent subsets HI_{p1}, HI_{p2} and HI_{p3}.

3.2.3 Development of multiparametric predictive models

Hct, MetHb and Na⁺, which were previously identified as non-significant by both Mann Whitney U test and Spearman's correlation, were excluded for multivariate analysis. To overcome limitations in predictive power of current gold standard method, different machine learning algorithms have been tested. As previously stated, due to the limited available data and the variability inside the HI_p group, 5-fold cross-validation was repeated 20 times for every classifier. Averaged figures of merit and standard deviations are displayed in Table 3.9. All algorithms present a prediction accuracy > 92.8%. Both SVM and PLS-DA classifiers exhibit a specificity of around 99%, demonstrating impressive performance on identification of the BC group, while their sensitivity is reduced to 87.2% and 88.6%, respectively. It is worth recalling the characteristics of these two classifiers to understand their behavior. SVM seeks a hyperplane that maximizes the margin between the hyperplane and data points. Due to higher variability of the HI_p group in comparison to BC group, it is reasonable to think that the margin favors the true negative rate. Similarly affected, PLS-DA combines maximization of covariance with classical discriminant analysis, which is no more than determining the class membership based on an arbitrary cut-off point. This behavior is slightly reduced with tree-based ensemble learners such as AB, GB, RF or XGB that have a non-parametric approach of the data resulting in both sensitivity and specificity > 90%.

Table 3.9: Classification performance of different machine learning algorithms for detection of hypoxia-ischemia (HI) events; all figures of merit are in percentage (%).

	PLS-DA	RF	SVM	AB	GB	XGB
Sensitivity	88.6 ± 1.9	90.3 ± 3.7	87.2 ± 2.4	90.5 ± 2.6	92.2 ± 2.4	92.0 ± 2.4
Specificity	99.4 ± 1.0	96.5 ± 2.0	98.7 ± 2.0	95.1 ± 2.1	95.9 ± 2.6	96.8 ± 1.8
PPV	99.3 ± 1.0	96.3 ± 2.0	98.6 ± 2.2	94.9 ± 2.1	95.8 ± 2.7	96.7 ± 1.8
NPV	89.7 ± 1.5	91.0 ± 3.1	88.6 ± 1.9	91.0 ± 2.2	92.5 ± 2.2	92.4 ± 2.1
Accuracy	93.98 ± 0.87	93.4 ± 1.9	93.0 ± 1.5	92.8 ± 1.5	94.1 ± 2.1	94.4 ± 1.7
AUROC	98.25 ± 0.44	97.85 ± 0.37	98.5 ± 1.2	96.2 ± 1.4	97.3 ± 1.1	98.10 ± 0.55

Once the predictive models are optimized, permutation feature importance is applied to interpret the predictive power of the physiological parameters regardless of the constructed model. Here, variables are randomly shuffled, one at a time, and one observes how accuracy is affected. Lactate is definitely the most relevant feature as it presents the greatest drop in terms of accuracy, while permutation of the rest of parameters is reasonably withstood by the distinct predictive models. Averaged decrease

of accuracy for SVM and XGB models is shown in Figure 3.5.

Projecting the data into new latent vectors, PLS maximizes the covariance with respect to the clinical state (BC or HI_p). To facilitate the graphical representation of the decision region of different classifiers, only two PLS components are considered. As previously mentioned, to avoid loss of relevant features, PLS is performed twice. PLS-based hybrid classifiers yield similar accuracies when compared to non-hybrid models. Accuracies of non hybrid classifiers lie within the range of 92.8 - 94.4% (Table 3.9), while hybrid classifiers reach values between 93.2% and 94.5% (Table 3.10). Even though the outcome of boosting algorithms GB and XGB, is slightly reduced in its performance when combined with PLS, all hybrid models achieve accuracies $> 93.2\%$. It is worth mentioning that the combination of RF and SVM with PLS delivers improved accuracies of 94.4% and 94.5%, respectively, compared to their usage as non-hybrid classifiers. An example of the decision region determined by SVM and GB in a reduced latent space of two components is displayed in Figure 3.6a and Figure 3.6b, respectively. The same two decision regions of PLS – SVM and PLS – GB hybrid classifiers with a more detailed disaggregation of the HI_p into subsets HI_{p1} - HI_{p3} (see Figure 3.6c and Figure 3.6d) reveal that controversial samples belong to latest HI phases.

Table 3.10: Classification performance of different hybrid machine learning algorithms for detection of hypoxia-ischemia (HI) events; all figures of merit are in percentage (%).

	PLS-RF	PLS-SVM	PLS-AB	PLS-GB	PLS-XGB
Sensitivity	92.6 ± 2.3	91.6 ± 1.0	91.1 ± 2.8	92.8 ± 2.8	92.1 ± 3.1
Specificity	96.2 ± 1.4	97.4 ± 1.1	95.2 ± 2.2	94.9 ± 1.9	94.6 ± 2.6
PPV	96.1 ± 1.3	97.3 ± 1.2	95.1 ± 2.1	94.8 ± 1.9	94.5 ± 2.6
NPV	92.9 ± 2.1	92.0 ± 0.9	91.5 ± 2.5	93.0 ± 2.6	92.4 ± 2.9
Accuracy	94.4 ± 1.1	94.5 ± 0.8	93.2 ± 1.6	93.8 ± 1.8	93.4 ± 2.5
AUROC	97.6 ± 1.2	98.25 ± 0.64	95.6 ± 1.2	97.5 ± 1.4	96.6 ± 1.4

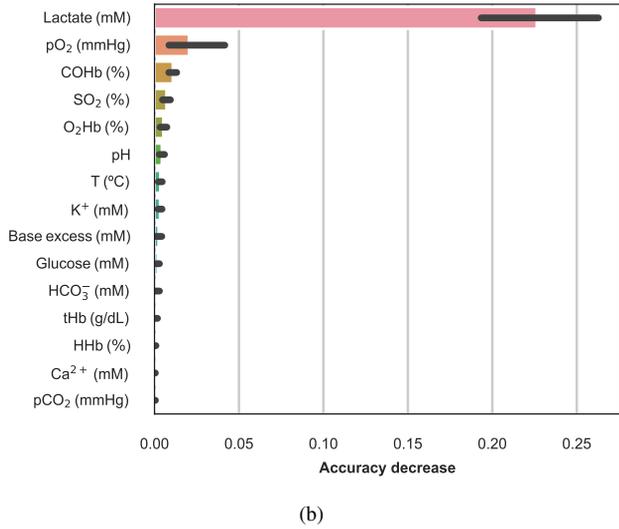
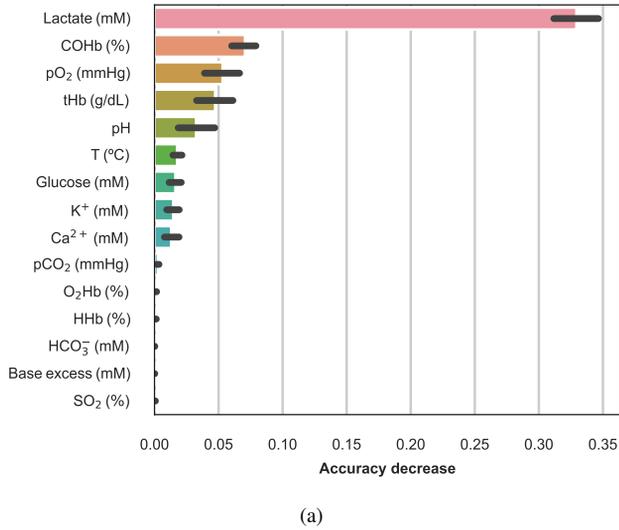


Figure 3.5: Permutation feature importance defined as accuracy decrease of fitted predictive models. **(a)** Averaged accuracy decrease of fitted Support Vector Machine (SVM) predictive models due to single parameter permutation and its standard deviation. **(b)** Averaged accuracy decrease of fitted Extreme Gradient Boosting (XGB) predictive models due to single parameter permutation and its standard deviation.

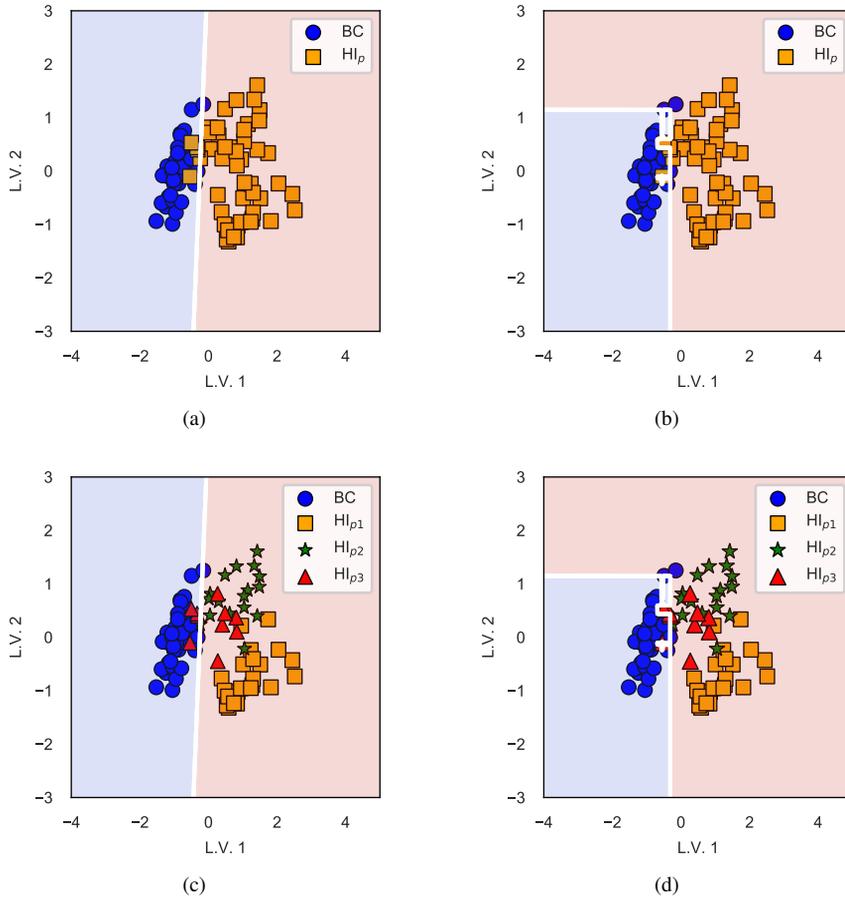


Figure 3.6: Example of 2-D visual class separability between basal condition (BC) and post-hypoxia-ischemia (HI_p) by hybrid classification models optimized by a 5-fold cross-validation. **(a)** Binary classification (BC vs. HI_p) based on the decision region defined by an SVM classifier in a 2-PLS component dataset. **(b)** Binary classification (BC vs. HI_p) based on the decision region defined by a GB classifier in a 2-PLS component dataset. **(c)** **(d)** Disaggregation of post-hypoxia-ischemia (HI_p) in a) and b) into time-dependent subsets HI_{p1}, HI_{p2} and HI_{p3}.

3.3 Discussion

Diagnosis of fetal asphyxia relies on the interpretation of fetal heart rate by cardiotocography and subsequent intermittent fetal scalp blood sampling with intradermal pH levels or alternatively lactate as decision-making parameter. This testing regime presents relevant false positive rates, resulting in unnecessary medical interventions as well as low negative predictive value that may result in prolonged fetal suffering. The development of new strategies for data interpretation, and hence, higher diagnostic confidence would naturally deliver added value in obstetric care.

Univariate analysis confirms lactate, pH and base excess as relevant biomarkers for hypoxic-ischemic events. This outcome is in good agreement with clinical practice for decision-making. The progressive restoration of these three parameters to basal levels in comparison with variables directly linked to oxygen, such as pO_2 , sO_2 or O_2Hb , convert them into more reliable indicators. However, it has been demonstrated that neither established pH nor lactate cut-off limits fulfill the expected performance. Although lactate proved better long-term evidence than pH, none of these parameters exhibits reliable predictive power, leading us to question the performance of current standards in obstetric care. AUROC values (84.0% for pH, 97.6% for lactate), which are independent of chosen cut-off limits and represent the ability to distinguish between HI_p cases and BC, suggest that other thresholds might provide better statistical metrics, particularly for post-detection of HI.

One has to emphasize at this point that the results are based on well-controlled experimental settings compared to the clinical situation with unknown conditions. This accentuates the weakness and vulnerability of a prediction based on univariate analysis that is prone to outliers and where relevant characteristic patterns can be overlooked. Moreover, lactate and pH are not disease-specific and are also used to evaluate the severity of other pathologies, such as sepsis or some types of cancer [43, 83]. Hence, physiological changes might also reflect other abnormal behavior rather than HI, making it even more complicated to define a specific threshold for classification. As a result of the univariate analysis of biochemical data from HI, it is recommended using lactate as the superior predictive diagnostic parameter rather than pH or base excess.

Even though, there exists a number of animal models of hypoxic-ischemic encephalopathy (HIE), as for example rodents, lambs, or pigs, the newborn pig animal

model has a number of advantages compared to other animal models: it has a greater similarity to a human newborn both in size and brain maturity; further, it resembles the cardiovascular system at the time of delivery; and in addition, the monitoring, instrumentation and evaluation of results in the pig model is similar to that used in the clinical care of babies with HIE. These points imply that the results obtained with utilized animal model have high potential for knowledge transfer to clinical settings [84].

Human metabolism entails a multiparametric space that can be much better understood by machine learning. Developing multiparametric models provides predictions that are based on a systemic clinical picture. Moreover, these models give insight into inner correlations between physiological parameters and quantify their significance. The adaptive abilities of machine learning algorithms mitigate the impact of outliers and allow for better modeling of the variability of the data.

To a great extent, machine learning algorithms do correctly classify samples in comparison with univariate analysis methods, which completely fail in detecting severe hypoxic-ischemic events after 30 minutes using pH and after 60 minutes using lactate as indicator. Furthermore, by different algorithms lactate has been identified as the main contributor, demonstrating to be a more reliable and long-term biomarker of HI. Multiparametric analysis of biochemical data derived from blood gas analysis enables diagnosis of severe HI even 30 to 60 minutes after the occurrence. The detection will help minimize risks and implications derived from adverse neonatal outcome.

The objective of hybrid classifiers is to reduce the dimensionality, and hence, facilitate the processing of different classification algorithms. The datasets of current study do not feature large sizes or complexity. Nevertheless, PLS – RF and PLS – SVM show better performance than just using them as non-hybrid classifiers. It is an undeniable advantage that boosting algorithms as AB, GB or XGB give special attention to misclassified data by an iterative development procedure of the model. As a consequence, the combination with PLS does not improve their performance in this study. Analysis by hybrid classifiers has demonstrated that dimensionality reduction firstly allows for graphical representation and subsequent better interpretation of classification by end users, and secondly represents a useful and necessary preprocessing step for complex data.

3.4 Conclusion

Analysis of pH from fetal scalp blood sampling is common practice in obstetric care for detection of metabolic acidosis and correlated perinatal asphyxia. The limitations of pH, being a reference parameter for correct medical classification of hypoxic-ischemic events, have been unveiled. As an alternative, the predictive capacity of lactate has been evaluated. Despite its potential for long-term evidence of hypoxia-ischemia, its performance is still limited by just using univariate analysis methods. As a consequence, predictive models based on multiparametric datasets of physiological parameters have been developed. By various machine learning methods, lactate has unambiguously been identified as the main contributor for the detection of hypoxia-ischemia. Moreover, it has also been validated that the systemic physiological picture, which includes the entirety of biochemical parameters, yields a substantial improvement in the diagnosis over the gold-standard techniques. This study represents a first step in the development of a new diagnostic tool for monitoring perinatal asphyxia.

pH and lactate sensing in body fluids

In the present chapter, the applicability of Raman spectroscopy together with machine learning is considered for quantitative and qualitative analysis of both pH and lactate in complex body fluids, such as blood and plasma. This approach represents a significant step toward diagnosis of perinatal asphyxia, where [lactate] > 4.8 mM [80] and pH < 7.20 [79] are defined as standards for intervention. To exclude secondary effects of complex media, the technology is firstly validated *in vitro* with a set of aqueous samples. A protocol has been developed for the determination of physiological pH and lactate, considering the entire process chain from sample preparation over Raman measurements to data evaluation. Following the protocol, the variation of pH and lactate *ex vivo* in blood and plasma samples from domestic pigs is then investigated.*†

*Published as: Ion Olaetxea, Ana Valero, Eneko Lopez, Hector Lafuente, Ander Izeta, Ibon Jaunarena, and Andreas Seifert. Machine learning-assisted Raman spectroscopy for pH and lactate sensing in body fluids. *Analytical Chemistry*, 92(20):13888–13895, 2020

†Presented as: I. Olaetxea, E. Lopez, A. Valero, and A. Seifert. Determination of physiological lactate and pH by Raman spectroscopy. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 475–481, 2019

4.1 Material & Methods

4.1.1 Customized Raman spectroscopy system

A customized free-space Raman spectroscopy system, which is shown in Figure 4.1a, has been designed for investigation of pH and lactate in fluidic samples. The idea is to combine two different laser excitation wavelengths, 785 nm and 1064 nm, which can be used interchangeably depending on the characteristics of the sample to be analyzed. As described in Section 2.4, near-infrared light sources are ideal for biological studies. Apart from a higher penetration depth, the use of longer wavelengths allows for a reduction of the fluorescent background and sample photodamage.

The Raman system is built in an inverted microscope configuration composed of two mirrored optical pathways. Each of the pathways is described in Figure 4.1b and consists of a continuous wave diode laser, emitting at 785 nm or 1064 nm with adjustable emission power, as Raman excitation source; adequate optical components for excitation of the sample and collection of Raman scattering; and a grating spectrometer – EAGLE Raman-S (Ibsen Photonics) with an Andor iVac 316 detector and a spectral resolution of 4 cm^{-1} for the pathway of 785 nm, and Kymera 193i with an iDus InGaAs 490 CCD detector and a variable spectral resolution with an adjustable slit for the pathway of 1064 nm – to collect the backscattered Raman light.

The free-space mirrored configuration allows for convenient replacement and modification of the optical components. Moreover, liquid samples are excited from below such that the effect of evaporation is minimized and unstable focusing and excitation is avoided. The Raman system emitting at 1064 nm could have not been optimized for measurements of biological samples so far. Thus, only measurements carried out with the laser wavelength of 785 nm have been considered. For this chapter, a laser power of 130 mW and a focusing lens with an NA of 0.23 were used.

4.1.2 Sample preparation and acquisition of Raman spectra

In the current chapter both *in vitro* and *ex vivo* samples were analyzed. First, spectroscopic measurements were carried out with aqueous solutions, such as milliQ water and phosphate-buffered saline (PBS, 10010015, Thermo Fisher Scientific), to demon-

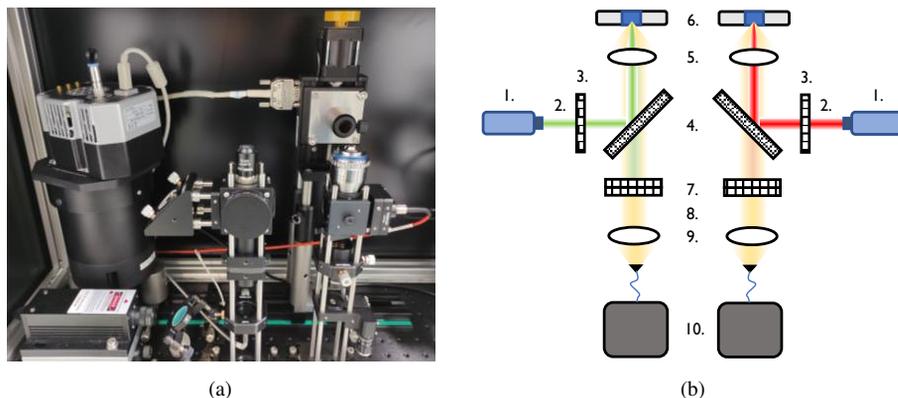


Figure 4.1: **(a)** Free-space customized Raman system for 785 nm and 1064 nm excitation wavelengths. **(b)** Schematic of the customized Raman system. 1. Laser source, 2. illumination path, 3. bandpass laser line filter, 4. dichroic mirror, 5. focusing lens, 6. motorized sample holder, 7. two longpass filter assembly, 8. filtered Raman scattered light, 9. refocusing lens, 10. spectrometer.

strate proof-of-principle. For a total number of 50 samples, pH was varied from 6.8 to 7.6, by adding 1 M sodium hydroxide (NaOH, S5881, Sigma-Aldrich) or hydrochloric acid (HCl, H1758, Sigma-Aldrich) diluted at 10%. For each sample, three consecutive spectra were collected; each spectrum was formed by accumulation of three measurements with an integration time of 10 seconds. Similarly, Sodium L-Lactate (71718, Sigma-Aldrich) was used to vary the concentration of lactate, firstly for the range of 0 - 30 mM (70 samples) and secondly for the physiologically relevant range of 0 - 10 mM (60 samples).

In a second step, blood from domestic pigs was analyzed. Blood, stored at 4°C, was provided by Biodonostia Health Research Institute. pH values from 30 samples were studied in the range from 6.97 to 7.53. The initial concentration of lactate in blood was varied by addition of sodium lactate. Due to the higher complexity in sample handling, a slightly broader range than the physiological range has been chosen (2 - 8 mM) and 32 samples prepared. For achieving a better signal-to-noise ratio (SNR) in blood, four consecutive spectra were collected per sample and each spectrum was obtained by accumulating five measurements with an integration time of 10 seconds.

Additionally to blood samples, the variation of lactate was also studied in plasma

samples, as a second example for analyzing complex media. Plasma was extracted from blood samples by centrifugation at 1000 g for 30 minutes and collection of the supernatant. Measurement specifications were maintained except the integration time, which was reduced to 6 seconds to avoid saturation of the detector.

For each sample, 250 μL were used and all measurements were performed at room temperature (23 $^{\circ}\text{C}$). Experimental procedures were conducted in strict compliance with European and Spanish regulations on the protection of animals used for scientific purposes (European Directive 2010/63/EU and Spanish Royal Legislative Decree 53/2013).

4.1.3 Preprocessing of Raman spectra

Analysis of Raman data in biological samples is extremely challenging due to their heterogeneous nature. Unwanted contributions caused by different background and noise sources, might hide weak Raman signals from the sample under test. Therefore, prior to develop any predictive model, preprocessing of raw Raman data is essential for the correct interpretation of the Raman spectra and for a reliable quantification of the samples. Hence, all Raman spectra were first trimmed to the range from 300 cm^{-1} to 1700 cm^{-1} and preprocessed in MATLAB (MathWorks[®]). An open toolbox [86] was used to apply a 6th order Extended Multiplicative Signal Correction (EMSC) to the raw spectra. Interfering additive and multiplicative artifacts were removed by scaling all the Raman spectra to the mean spectrum [61]. The median of consecutive spectra from each sample was taken to eliminate spike artifacts sporadically produced by cosmic rays. All Raman spectra were smoothed using a Savitzky-Golay filter with a window of 15 points and 3rd order polynomial at a sampling of 4 cm^{-1} . The asymmetric least squares (ALS) method was used to subtract a smooth background produced by the intrinsic fluorescence of the molecules [60]. The integrity of the Raman bands is retained by giving higher relevance to positive residuals [57]. Finally, to avoid uncertainties of the variance of the mean, the origin of the coordinate system is translated to the center of gravity of the dataset. Combination of all data preprocessing has been optimized based on maximization of Raman features and performance of the predictive models.

4.1.4 Multivariate analysis

Simultaneous analysis and combination of different variables enables the development of efficient predictive models. However, the correct extraction of significant features within thousands of misleading, correlated and redundant variables to be handled is a major challenge [57]. The high number of variables, or pixels in this case, obtained by Raman spectroscopy can lead to overfitting, impeding the model to generalize. Machine learning is used to reliably identify and model the spectral variability from preprocessed spectra [57, 35]. Both partial least squares (PLS) and principal component analysis (PCA) are techniques for dimensionality reduction used in machine learning. PCA was used to reveal patterns and facilitate visualization and interpretation of the spectral data. The discrimination of spectra as a function of their pH or concentration of lactate was evaluated by projecting the data onto the most relevant principal components [57]. Moreover, PLS was used to develop predictive regression models that define the correlation between the spectral variability and the pH and the concentration of lactate of the samples [69, 70]. Fundamental concepts of applied algorithms can be found in more detail in Section 2.6.

To evaluate the performance of predictive models, such as those resulting from PLS, machine learning algorithms consist of two stages. First, the prediction model is built and calibrated with a *training set* of measurement data and validated with a new *testing set* of unknown samples afterwards. For model calibration, leave-one-out cross-validation (LOOCV) method was used to determine an optimum set of latent vectors, and hence, avoiding overfitting. The root mean squared error of prediction (RMSEP) and the coefficient of determination R^2 were used as indicators of the quality of the predictive model. Whereas RMSEP indicates the standard deviation of the predictions from the observed values, R^2 reflects the amount of variance in the dependent variable that is explained by the model, or in other words, how well the model fits the data (the goodness of fit). All the analysis was performed in MATLAB environment.

4.2 Results

4.2.1 pH and lactate monitoring in aqueous solutions

pH analysis

Due to purity of milliQ water and its subsequent low buffering capacity, PBS has been used to avoid unwanted additional pH variations when exposed to air. As explained previously, Raman spectroscopy provides specific information about the vibrational modes of the molecules. Considering that pH is a metric to determine the acidity of a solution by measuring the concentration of hydrogen ions, the variation of pH, unlike lactate, is not directly quantifiable by Raman spectroscopy. However, the variation of pH can indeed induce changes either in the structure, the concentration or the chemical bonds of the molecules in the sample being measured, which are eventually detectable by Raman spectroscopy and can be quantified and correlated with pH. In the case of PBS, which is a buffered solution, pH is expected to provoke changes to the salts included. In Figure 4.2, Raman bands at 876 cm^{-1} and 1076 cm^{-1} , and the peak at 989 cm^{-1} , which have previously been related to H_2PO_4 and HPO_4 by Fontana et al. [87], have shown to be pH-dependent.

These spectral variations are highlighted by applying PCA. Figure 4.3a clearly separates the different samples as a function of their pH. The loading plot in Figure 4.3b substantiates that most of the spectral variation captured by the first two components (85% of the variance) is indeed caused by the effect of pH in the vibrational modes of both H_2PO_4 and HPO_4 .

A predictive PLS regression model has been developed for the prediction of pH in PBS samples. The total of 50 samples were divided into two sets, 40 training samples for calibration and 10 testing samples. In Figure 4.3c, the small RMSEP of 0.04 and R^2 of 0.98 obtained by the PLS predictive model for unknown samples demonstrates the quality of the model and the feasibility of the method to detect pH variations in aqueous solutions. Figure 4.3d shows the main contribution of H_2PO_4 and HPO_4 in the predictive model.

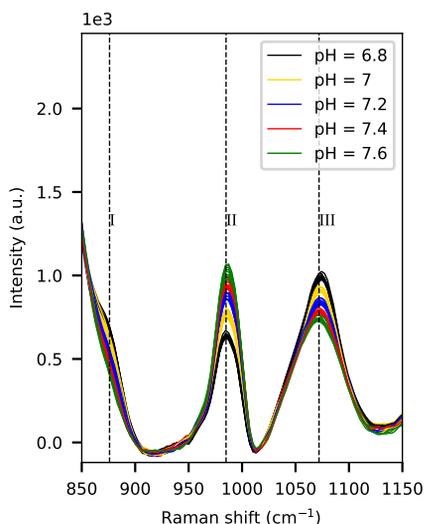


Figure 4.2: Preprocessed Raman spectra of PBS solutions with different pH values from 6.8 to 7.6. Only the range from 850 to 1180 cm^{-1} is shown for better visualization. (Peaks: I. 876 cm^{-1} , II. 989 cm^{-1} , III. 1076 cm^{-1})

Lactate analysis

The fingerprint region of pure sodium-lactate powder is shown in Figure 4.4. Some of the characteristic Raman peaks of lactate, previously reported [88, 89], are indicated: a sharp predominant peak at 853 cm^{-1} associated with a C–C single bond vibration, at 543 cm^{-1} from CO_2 wagging, at 1040 cm^{-1} and 1053 cm^{-1} due to C– CH_3 stretching vibrations, at 1081 cm^{-1} caused by C–O vibrations, and at 1459 cm^{-1} from asymmetric CH_3 deformation modes. Some other visible Raman bands in Figure 4.4 have not been addressed, but are characteristic of sodium-lactate.

As mentioned before, the first step was to identify specific Raman features of lactate in milliQ water solutions. Figure 4.5a and Figure 4.5b present the raw data and its corresponding preprocessed spectra from the set of samples with varying concentrations of lactate in a range between 0 and 30 mM. In raw data, it is extremely difficult to recognize any pattern. However, a completely different picture emerges after applying the described preprocessing, which demonstrates the importance of this step. Several regions of the spectra undergo fluctuations in the intensity of the Raman features.

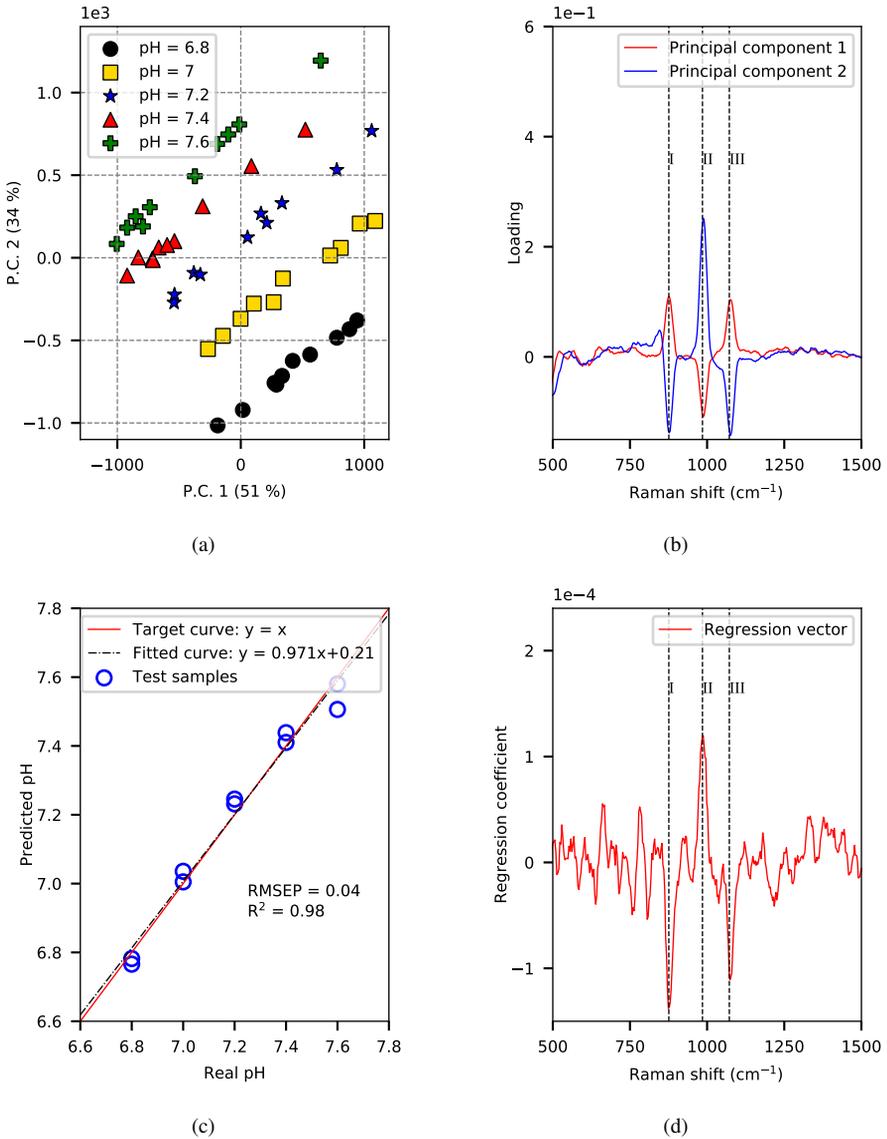


Figure 4.3: PBS solutions with varying values of pH in the physiological range of 6.8–7.6. **(a)** PCA score plot (PC-1 vs PC-2); **(b)** PCA loading plot; **(c)** prediction of the PLS regression model with an optimum number of latent vectors of $n=3$; the values of RMSEP and R^2 are characteristic of the model and are calculated with respect to the target curve; and **(d)** regression coefficient plot of the PLS regression model. (Peaks: I. 876 cm⁻¹, II. 989 cm⁻¹, III. 1076 cm⁻¹)

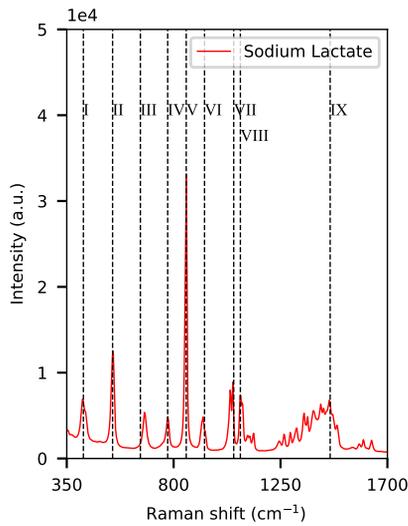


Figure 4.4: Raman spectrum of Sodium Lactate. (Peaks: I. 420 cm⁻¹, II. 543 cm⁻¹, III. 660 cm⁻¹, IV. 775 cm⁻¹, V. 853 cm⁻¹, VI. 930 cm⁻¹, VII. 1053 cm⁻¹, VIII. 1081 cm⁻¹, IX. 1459 cm⁻¹)

In Figure 4.5b, Raman bands from pure sodium lactate as well as some additionally reported peaks [90], such as 930 cm⁻¹ or 1125 cm⁻¹, are observed. Figure 4.5c is a magnification of the predominant peak at 853 cm⁻¹, which clearly shows how the intensity of the peak varies in correlation with the concentration of lactate; the higher the concentration of lactate, the higher the intensity of the peak. When the concentration of lactate is reduced to the common physiological range between 0 and 10 mM, lactate Raman features are still evident despite the little differences between the spectra, as shown in Figure 4.5d. The great advantage of using multivariate analysis is the possibility to take into account many features of the spectra instead of single peaks, thereby unveiling hidden information. Moreover, multivariate analysis omits redundant information so that reliable and accurate predictive models are developed.

To have a better visualization of the spectra, preprocessed spectra are projected onto the first two principal components, that explain 78% of the spectral variation (Figure 4.6a). The loading plot of PC-1 and PC-2 unveils the influence of lactate in the spectral variation captured by these components (Figure 4.6b). Spectral features associated with lactate constitute a dominant part. Even Raman peaks, such as 660 cm⁻¹,

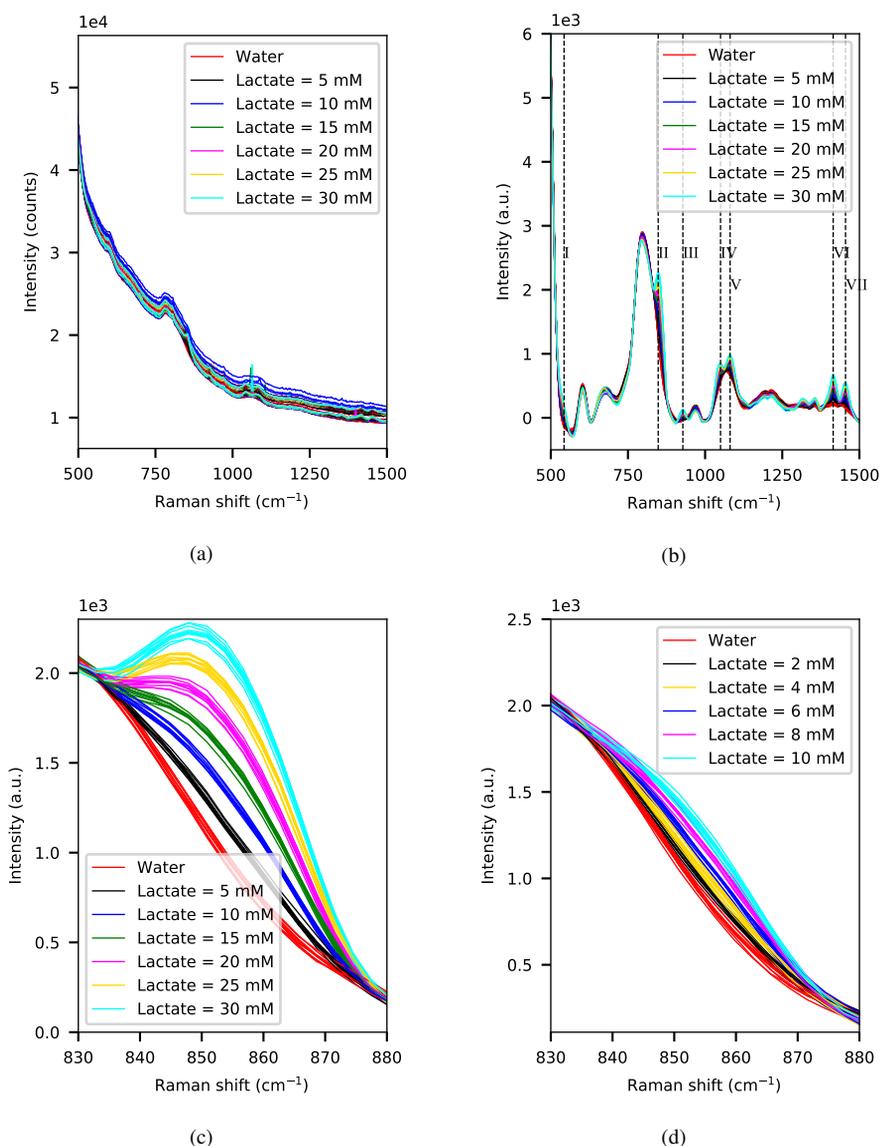


Figure 4.5: Aqueous solutions with different concentrations of sodium lactate. **(a)** Raw Raman spectra in the region from 500 to 1500 cm^{-1} with varying concentrations of lactate in the range of 0–30 mM and **(b)** its corresponding preprocessed spectra. **(c)** Peak at 853 cm^{-1} , associated with a C–C single bond vibration, for different concentrations of lactate in the range of 0–30 mM and **(d)** in the physiological range of 0–10 mM. (Peaks: I. 543 cm^{-1} , II. 853 cm^{-1} , III. 930 cm^{-1} , IV. 1053 cm^{-1} , V. 1081 cm^{-1} , VI. 1420 cm^{-1} , VII. 1459 cm^{-1})

775 cm^{-1} , 1125 cm^{-1} or 1315 cm^{-1} , that were not perceptible in Figure 4.5b by eye, are now gathered by PCA.

A predictive PLS regression model has been developed based on preprocessed spectra. A total of 60 samples were divided into two sets, 48 training samples for calibration and 12 testing samples. The model was calibrated using the LOOCV method and the optimum number of latent vectors was determined by Wold's criterion [91]. The predictive model has been successfully tested with unknown samples providing a very small RMSEP of 0.32 mM and R^2 of 0.99, which corroborates the quality of the model. Figure 4.6c compares the predicted and real values of lactate concentration. To understand how strongly each Raman band determines the model, the regression coefficients of the calibrated PLS model are shown in Figure 4.6d. The major contributing Raman features are associated directly with lactate.

4.2.2 pH and lactate monitoring in blood

pH analysis

To further study the efficacy of our method, Raman spectra from blood samples from domestic pigs, as representatives for real physiological samples, have been analyzed at different pH values (Figure 4.7). As mentioned before, the objective is to identify and indirectly quantify variations of pH by detecting Raman bands from molecules that are sensitive to pH. Vibrational modes of hemoglobin have actually been recognized as the main responsible ones for the Raman spectrum of blood [92].

In Figure 4.8a, the projection of spectra onto the first three principal components shows a clear clustering of the spectra with respect to their pH value. The loading plot in Figure 4.8b unveils some spectral variation in Raman bands, such as 420 cm^{-1} , 567 cm^{-1} , 754 cm^{-1} , 789 cm^{-1} , 1002 cm^{-1} , 1207 cm^{-1} , 1222 cm^{-1} , 1357 cm^{-1} , 1547 cm^{-1} and 1639 cm^{-1} , which have already been assigned to hemoglobin [93]. It has been reported that hemoglobin also exhibits characteristic active Raman bands in the region from 1300 cm^{-1} to 1650 cm^{-1} , however, it is interesting to notice that many of these reported peaks could not be precisely identified in Figure 4.8b. Nevertheless, the high variability of the spectra in this region indicates pH-induced modifications in the vibrational modes of hemoglobin.

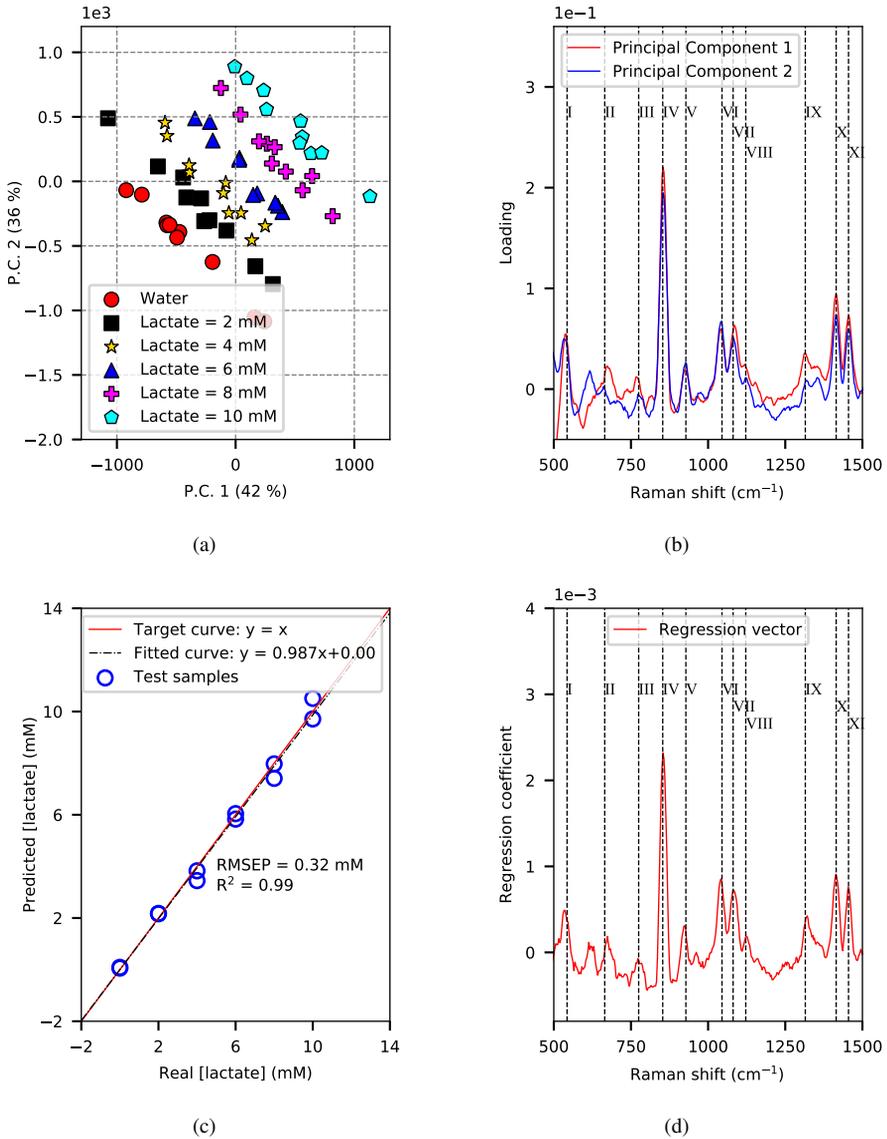


Figure 4.6: Aqueous solutions with varying concentrations of lactate in the physiological range of 0–10 mM. **(a)** PCA score plot (PC-1 vs PC-2); **(b)** PCA loading plot; **(c)** prediction of the PLS regression model with an optimum number of latent vectors of $n=3$; RMSEP and R^2 defined as in Figure 4.3; and **(d)** regression coefficient plot of the PLS regression model. (Peaks: I. 543 cm^{-1} , II. 660 cm^{-1} , III. 775 cm^{-1} , IV. 853 cm^{-1} , V. 930 cm^{-1} , VI. 1053 cm^{-1} , VII. 1081 cm^{-1} , VIII. 1125 cm^{-1} , IX. 1315 cm^{-1} , X. 1420 cm^{-1} , XI. 1459 cm^{-1})

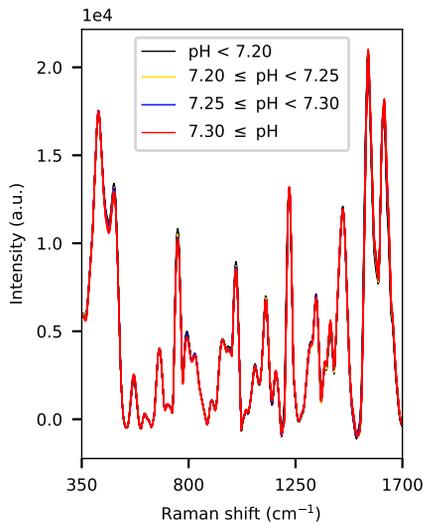


Figure 4.7: Preprocessed Raman spectra of pig blood with varying values of pH in the range of 6.97–7.53.

A PLS model has been implemented for the prediction of pH in blood from preprocessed spectra. The model was calibrated with 22 samples for future validation with 8 new samples. Figure 4.8c displays the excellent predictive ability of the model for unknown samples, which is reflected in the small RMSEP of 0.05 and the good R^2 of 0.92, when the reference and the predicted values of pH are compared. Each Raman band associated with hemoglobin is affected differently by the variation of pH, and hence, contributes individually to the model. For that reason, it is not straightforward to identify exact positions of the characteristic Raman peaks of hemoglobin in the regression vector in Figure 4.8d. Nonetheless, stronger contributions from regions around 420 cm^{-1} , 1000 cm^{-1} , 1200 cm^{-1} and 1600 cm^{-1} are observed.

Lactate analysis

Unlike in the case of pH, lactate is quantified by detecting characteristic Raman peaks in the preprocessed spectra from blood, which are displayed in Figure 4.9. The discriminating capability of PCA is demonstrated in Figure 4.10a, where the projection of preprocessed spectra in the first three principal components allows for visual separation of the samples into distinct groups. Looking at the loading plot in Figure 4.10b, the

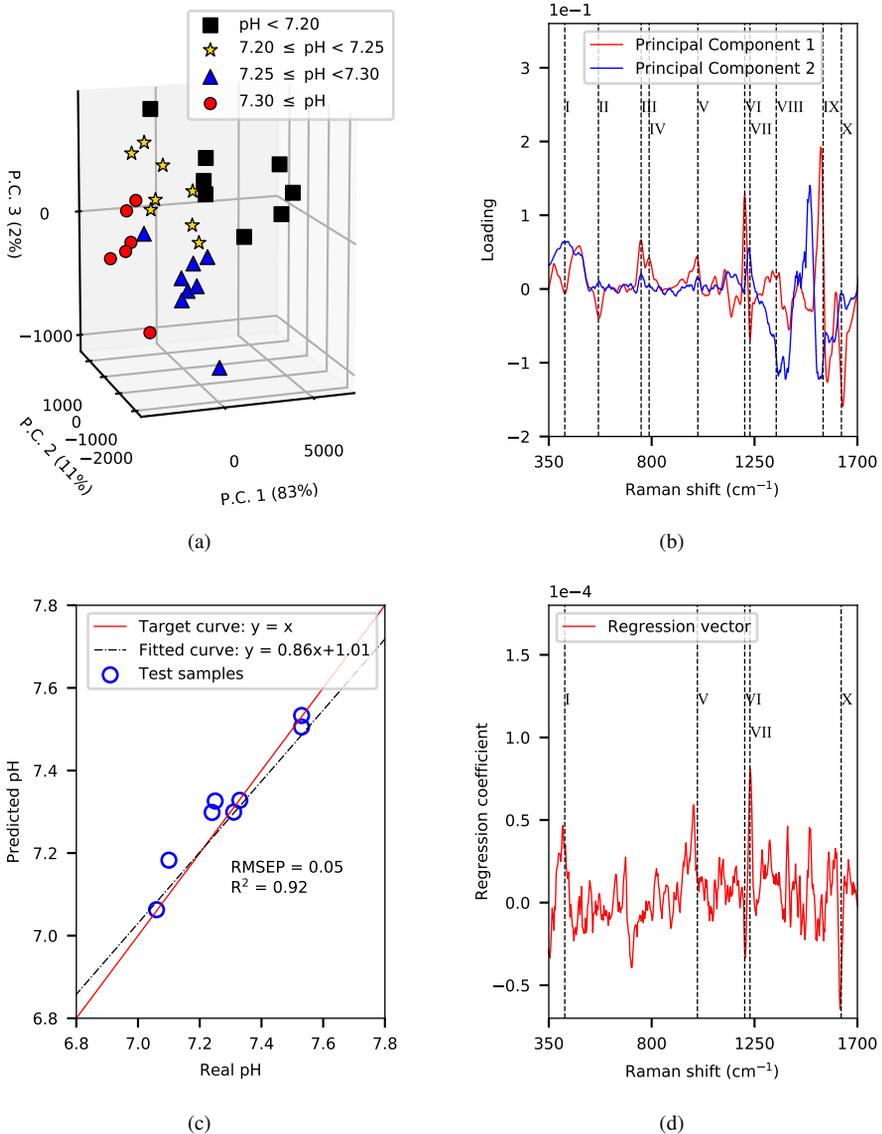


Figure 4.8: Pig blood with varying values of pH in the range of 6.97 – 7.53. **(a)** PCA score plot for the first three principal components; **(b)** PCA loading plot; **(c)** prediction of the PLS regression model with an optimum number of latent vectors of $n=6$; RMSEP and R^2 defined as in Figure 4.3; and **(d)** regression coefficient plot of the PLS regression model. (Peaks: I. 420 cm⁻¹, II. 567 cm⁻¹, III. 754 cm⁻¹, IV. 789 cm⁻¹, V. 1002 cm⁻¹, VI. 1207 cm⁻¹, VII. 1222 cm⁻¹, VIII. 1357 cm⁻¹, IX. 1547 cm⁻¹, X. 1639 cm⁻¹)

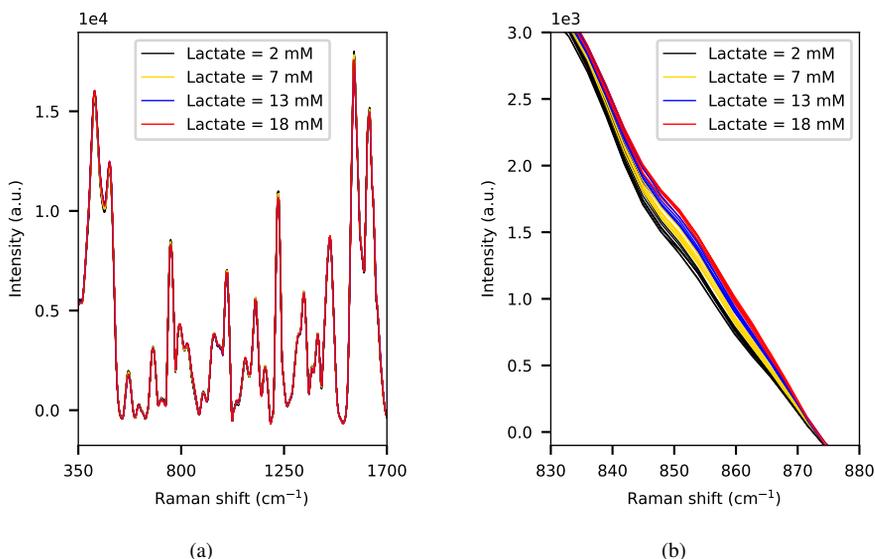


Figure 4.9: Preprocessed Raman spectra of pig blood with varying concentrations of lactate in the range of 2 – 18 mM. (a) Full fingerprint region from 300 to 1700 cm⁻¹. (b) Characteristic peak of lactate at 853 cm⁻¹.

spectral variation captured by the first two components (85% of the variance) reveals some of the characteristic features of lactate, but also many other bands that are not related to it.

For developing a PLS regression model, preprocessed spectra of 32 samples have been used, 24 for training and 8 for testing. The prediction, when tested with unknown samples, demonstrates again the quality of the model with an RMSEP of 1.25 mM and $R^2 = 0.96$, as shown in Figure 4.10c. Two main messages can be drawn from the regression coefficients of the predictive model in Figure 4.10d: first, most of the spectral variation present in the data is in accordance with different concentrations of lactate, since most of its characteristic Raman bands are identified; and second, this spectral variation is not only due to lactate since other Raman bands, which are not associated with lactate, are also contributing to the model.

To study where other contributions apart from lactate are coming from, a gold standard method, based on electrochemical analysis, has been used to take a reference measurement of the main parameters commonly measured in clinical settings (i-STAT,

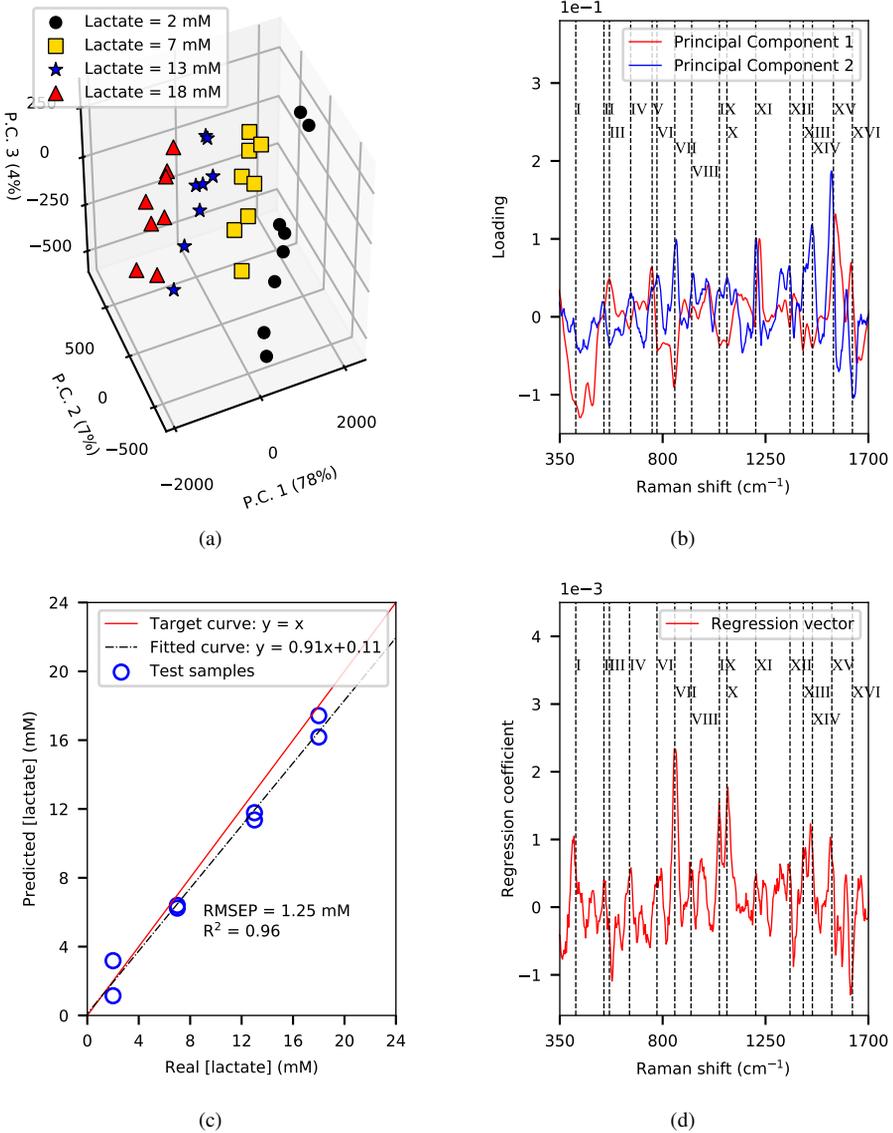


Figure 4.10: Pig blood with varying concentrations of lactate in the range of 2–18 mM. **(a)** PCA score plot for the first three principal components; **(b)** PCA loading plot; **(c)** prediction of the PLS regression model with an optimum number of latent vectors of $n=4$; RMSEP and R^2 defined as in Figure 4.3; and **(d)** regression coefficient plot of the PLS regression model. (Peaks: I. 420 cm^{-1} , II. 540 cm^{-1} , III. 567 cm^{-1} , IV. 660 cm^{-1} , V. 754 cm^{-1} , VI. 775 cm^{-1} , VII. 853 cm^{-1} , VIII. 930 cm^{-1} , IX. 1053 cm^{-1} , X. 1081 cm^{-1} , XI. 1207 cm^{-1} , XII. 1357 cm^{-1} , XIII. 1420 cm^{-1} , XIV. 1459 cm^{-1} , XV. 1547 cm^{-1} , XVI. 1639 cm^{-1})

Abbott Technologies). The values of four blood samples with different concentrations of lactate at two time intervals are displayed in Table 4.1. As expected, the concentration of lactate changes from sample to sample, but it can also be observed that many of these parameters are changing with time, meaning that sample preparation and therefore exposure of the samples to air is extremely determinant. Furthermore, Lemler et al. [94] already mentioned that hemoglobin experiences some chemical changes as soon as blood is extracted. Considering the fact that sample preparation was performed by groups at different times, the variation of the different parameters shown in Table 4.1 could explain the unexpected contributions present in the model (Figure 4.10c, Figure 4.10d). Moreover, it has been demonstrated that prolonged exposure to laser light and high incident power are responsible for hemoglobin denaturation and heme aggregates [94]. This unwanted effect might hinder possible correlations induced by the addition of lactate.

Table 4.1: Clinical parameters, measured with i-STAT (Abbott Technologies), from pig blood with varying concentrations of lactate at two different periods of time.

Clinical parameters	Sample 1		Sample 2		Sample 3		Sample 4	
	<i>t1</i>	<i>t2</i>	<i>t1</i>	<i>t2</i>	<i>t1</i>	<i>t2</i>	<i>t1</i>	<i>t2</i>
pH	6.92	7.02	6.99	7.06	6.98	7.08	6.99	7.06
pCO ₂ (mmHg)	125	95	103	84	101	76	95	79
pO ₂ (mmHg)	175	186	178	166	150	157	133	145
BE (mM)	-7	-6	-7	-7	-8	-7	-8	-8
HCO ₂ (mM)	25.6	24.7	24.4	23.5	23.8	22.6	23.0	22.7
TCO ₂	29	27	27	26	27	25	26	25
sO ₂ (%)	98	99	98	98	97	98	96	98
Lactate (mM)	1.78	2.07	7.51	7.68	13.08	13.42	18.67	18.98

4.2.3 Lactate monitoring in blood plasma

To minimize the effect induced by sample preparation for the prediction of lactate, blood cells had been removed and the experiment was repeated in plasma samples (Figure 4.11). Visual inspection of PCA reveals that plasma spectra still form clusters according to their concentration of lactate (Figure 4.12a) and that lactate Raman features are still captured by the first two components, which is reflected in the loading plot in Figure 4.12b.

In the case of plasma, a more robust and accurate PLS regression model is obtained. In comparison with blood samples, the prediction error has been reduced to a value of

RMSEP = 0.51 mM (Figure 4.12c), which demonstrates a much stronger contribution of lactate (Figure 4.12d).

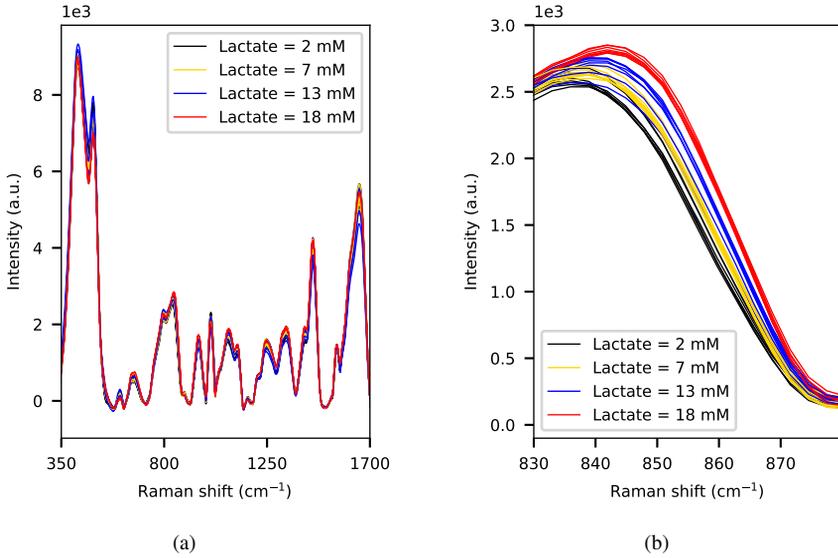


Figure 4.11: Preprocessed Raman spectra of pig blood plasma with varying concentrations of lactate in the range of 2–18 mM. (a) Full fingerprint region from 300 to 1700 cm^{-1} . (b) Characteristic peak of lactate at 853 cm^{-1} .

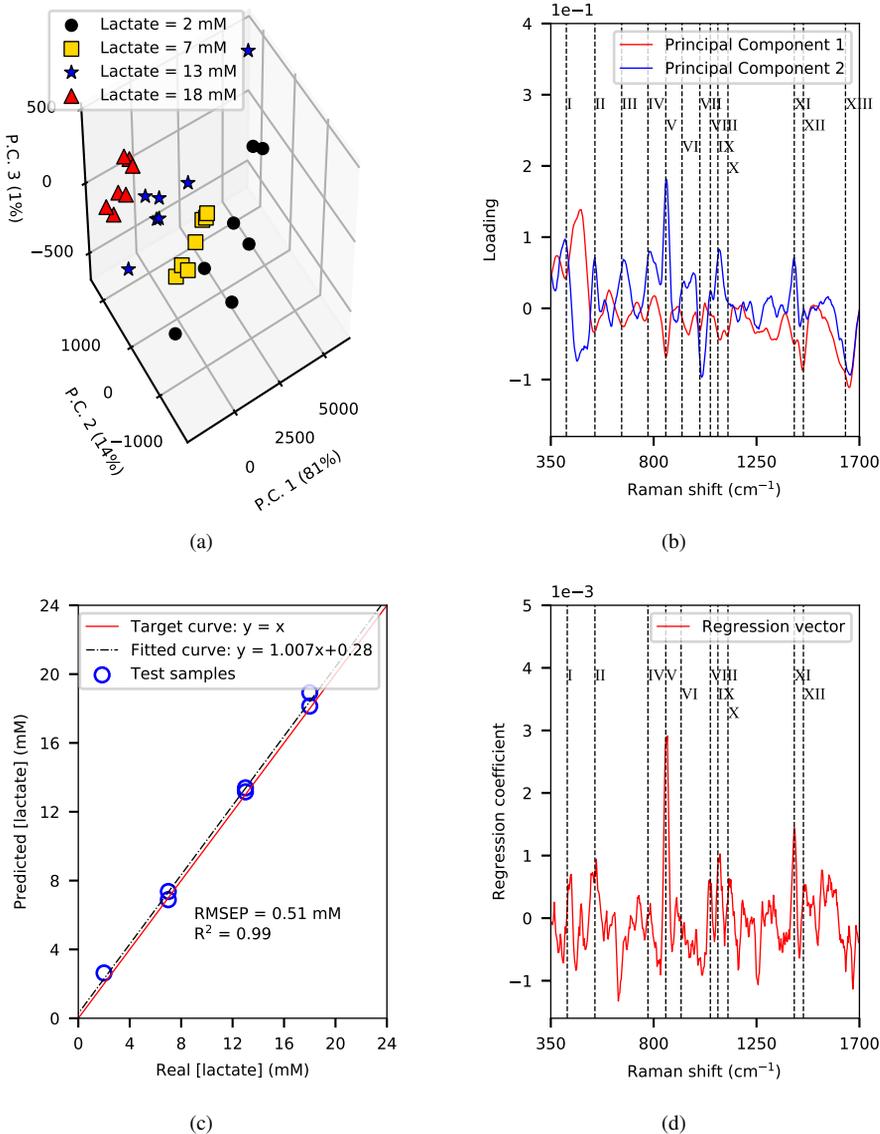


Figure 4.12: Pig blood plasma with varying concentrations of lactate in the range of 2 – 18 mM. (a) PCA score plot for the first three principal components; (b) PCA loading plot; (c) prediction of the PLS regression model with an optimum number of latent vectors of $n=5$; RMSEP and R^2 defined as in Figure 4.3; and (d) regression coefficient plot of the PLS regression model. (Peaks: I. 420 cm^{-1} , II. 540 cm^{-1} , III. 660 cm^{-1} , IV. 775 cm^{-1} , V. 853 cm^{-1} , VI. 930 cm^{-1} , VII. 1002 cm^{-1} , VIII. 1053 cm^{-1} , IX. 1081 cm^{-1} , X. 1125 cm^{-1} , XI. 1420 cm^{-1} , XII. 1459 cm^{-1} , XIII. 1639 cm^{-1})

4.2.4 Evaluation of predictive models

PLS regression models have been implemented and used for evaluating the predictive power of pH and lactate in aqueous solutions and body fluids by Raman spectroscopy. Quantification errors are summarized in Table 4.2. It has been shown that data preprocessing plays a crucial role in the development of models with good performance. The capability of our method for sensing lactate and pH variations has been demonstrated by this study. Moreover, it has been shown that blood, especially hemoglobin, is extremely sensitive to measurement conditions and affects the model performance for lactate prediction. External circumstances, such as the exposure of the sample to air or an excessive power of the laser, can introduce uncontrolled additional variability to the predictive model that should be considered.

Table 4.2: Prediction errors of pH and lactate in aqueous solutions and body fluids.

	PBS/Water	Blood	Plasma
pH	0.04	0.05	-
Lactate (mM)	0.32	1.25	0.51

4.3 Conclusion

The present study has demonstrated that the combination of Raman spectroscopy with machine learning represents a suitable tool for monitoring lactate and pH values from body fluids. Principal component analysis has proved clear discrimination of the spectra as a function of the pH value and concentration of lactate. For future clinical applications, this constitutes a powerful tool for both binary and multiclass classification of perinatal asphyxia, as well as for other fields, such as fatigue analysis or sepsis classification. In addition, partial least squares predictive models have been successfully tested with unknown samples, providing clinically promising errors that verify the reliability, the accuracy and hence, the quality of predictive models. In conclusion, it has been demonstrated that developing a new diagnostic method to monitor perinatal asphyxia *ex vivo* could be feasible.

Optimization of a Raman probe system

Prototyping and development of hardware components of a specific Raman probe for accurate fetal monitoring represents a challenge and plays a major role in the future commercialization of the technology. Given the existence of a mature and wide market of optical probes, valuable practical and technical expertise is already available. Therefore, as part of the validation of the technology and prior to conducting a preclinical study, an evaluation of different commercial Raman probes has been performed to give insight into essential characteristics our customized Raman probe will require. Particularly, this chapter describes the comparison of the "EmVision Raman Lensed probe" (EmVision LLC) [95] and the "RamanProbeTM" (RPS785, Inphotonics) [96].

5.1 Comparison of commercial probes

5.1.1 Optical and mechanical characteristics

EmVision Raman Lensed probe consists of a single excitation fiber surrounded by seven collection fibers. All fibers are composed of silica with low-hydroxyl (OH) content and have a core diameter of $300\ \mu\text{m}$ with a numerical aperture (NA) of 0.22. Designed to be utilized both for immersion and direct contact measurements, the performance of the probe at the contact surface is maximized by a specific geometry of the lens combination at its distal end. A bandpass laser line filter, centered at 785 nm, is placed in front of the excitation fiber to selectively transmit a narrow laser band. A donut-shaped longpass notched filter avoids the entrance of Rayleigh scattered light in the collection fibers. On top of the filters, a two-component converging lens overlaps the excited and collected volumes from the contact surface. Approximately, an area of $0.8\ \text{mm}^2$ is irradiated. All components are encased in a stainless steel needle tube of 2.1 mm. Unlike the donut shape at the tip, seven collection fibers are linearly arranged for the probe-spectrometer coupling. An FC-PC and an SMA-SMA connector are used for laser and spectrometer coupling, respectively [95].

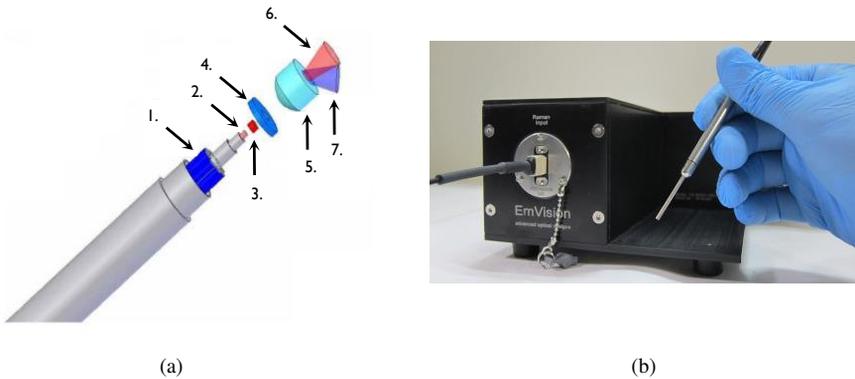


Figure 5.1: **(a)** Exploded view of the tip of the EmVision Raman Lensed probe [97] consisting of: 1. seven collection fibers, 2. single excitation fiber, 3. bandpass laser line filter centered at 785 nm, 4. donut-shaped longpass notched filter, 5. two-component converging lens, 6. excited cone, 7. collected cone of a single collection fiber. **(b)** Picture of the EmVision Raman Lensed probe [95].

RamanProbe™ follows a more standard geometry. The lens combination, similar to our customized free-space Raman system described in Section 4.1.1, includes a dichroic filter which allows for collection of the backscattered Raman signal. The probe consists of a single excitation and a single collection fiber with a core diameter of $105\ \mu\text{m}$ and $200\ \mu\text{m}$, respectively. A $785\ \text{nm}$ laser line clean-up filter is placed in front of the excitation fiber to selectively transmit a narrow laser band. The dichroic filter allows for transmission of the laser while longer wavelengths, generated as a result of the Raman scattering, are reflected through the collection path. Maximum throughput is achieved with a complete overlap of excitation and collection volumes. Nonetheless, the RamanProbe™ is not designed to work in contact with the surface. The focus spot has a theoretical area of $0.08\ \text{mm}^2$ and is dependent of the working distance, which has experimentally been determined at $5.6\ \text{mm}$. A longpass filter assembly avoids the entrance of Rayleigh scattered light in the collection fiber. All components are embedded in a protective jacket. An FC-PC and an SMA-SMA connector are used for laser and spectrometer coupling, respectively [96].

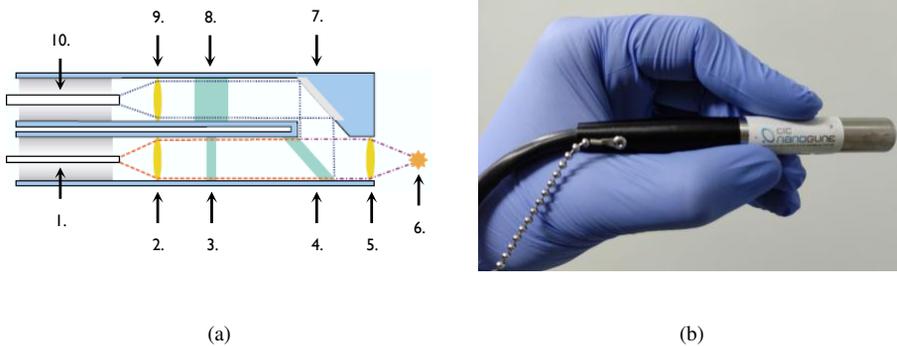


Figure 5.2: **(a)** Cross-sectional view of the tip of the RamanProbe™ consisting of: 1. single excitation fiber, 2. collimating lens, 3. bandpass laser line filter centered at $785\ \text{nm}$, 4. dichroic mirror, 5. focusing lens, 6. measurement spot, 7. reflective mirror, 8. longpass filter assembly, 9. refocusing lens, 10. single collection fiber. **(b)** Picture of the RamanProbe™.

5.1.2 Measurement performance of the Raman probes

Polypropylene has been examined as standard reference material due to its strong Raman signal and broad Raman-active spectral range. Figure 5.3a shows the spectrum from polypropylene measured with the EmVision Raman Lensed probe and the RamanProbe™ for an excitation wavelength of 785 nm. Moreover, polypropylene spectrum obtained by our customized free-space Raman spectroscopy system was used for comparison.

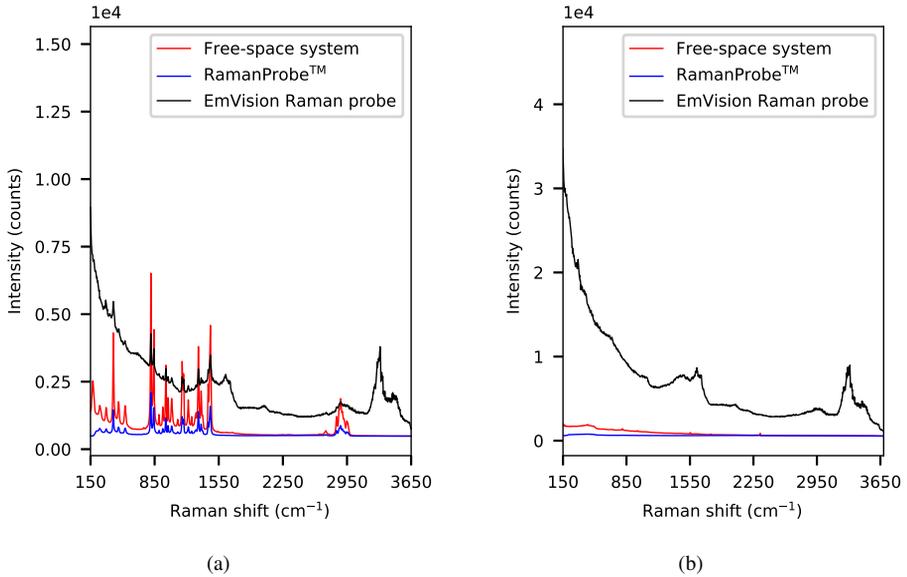


Figure 5.3: Performance comparison of the free-space customized Raman spectroscopy system, the EmVision Raman Lensed probe, and the RamanProbe™ for an excitation wavelength of 785 nm. **(a)** Polypropylene Raman spectrum collected for a power of 25 mW and an integration time of 1 second. **(b)** Background signal without sample excitation for a power of 25 mW and an integration time of 20 seconds.

The EmVision Raman Lensed probe exhibited a strong fluorescent background. This spectral residual was also observed when no sample was irradiated and the probe was placed inside a black anodized aluminium box with the laser switched on, as seen in Figure 5.3b. The two-component converging lens at the tip of the probe and the silica component of the collection fiber emit an intrinsic fluorescence that reaches the detector and disturbs the Raman signal of the polypropylene. The spectrometer from

EmVision integrates an extra laser blocking filter, critical according to the developers, at the collimating section to avoid the effect of fluorescence. However, how a longpass filter, placed at the entrance of the spectrometer, would help to reduce the fluorescence is not obvious. If such a filter is placed after the light has been transmitted through the fiber, fluorescence generated by the silica component of the collection fiber is never avoided. A potential justification could be that fluorescence is mainly generated by internal components of the spectrometer. The RamanProbeTM mitigates this interference by mounting a longpass filter assembly with optical density above eight ($OD > 8$) before entering the collection fiber. Such a filter blocks transmission of the laser wavelength with high efficiency. Similarly, our customized free-space Raman spectroscopy system includes a combination of two longpass filters with the same purpose.

The coupling between the probes and the Eagle Raman-S spectrometer (Ibsen) also affects their performance. The spectrometer has an integrated fiber bundle composed of eight fibers with a core diameter of $400\ \mu\text{m}$ and NA of 0.3, geometrically arranged such that the total diameter is around 1.3 mm (Figure 5.4a). However, the linear arrangement of the collection fiber bundle in the EmVision probe results in a total length of at least 2.1 mm (Figure 5.4b), clearly out of the effective margin of the spectrometer's integrated fiber. For the RamanProbeTM, which has a single collection fiber of $200\ \mu\text{m}$ (Figure 5.4c), the coupling effect hardly affects its performance. However, due to the small core size of its fiber, Raman signal from a much smaller collection cone is recorded in comparison with the free-space Raman spectroscopy system. Reduction of the Raman signal in liquids or thick transparent samples is counteracted by its confocal characteristic.

The EmVision Raman Lensed probe has already demonstrated fantastic performance and great potential for different applications, such as tumor malign delineation [98] or characterization during prostatectomy procedures [99], but coupling of the probe to their own customized Raman system has proven to be crucial. Even though its optical characteristics and design will be considered for future development of an optimized Raman probe for fetal monitoring, the RamanProbeTM is used for further preclinical validation of the technology. Moreover, Figure 5.5 supports the election, considering the great performance of the RamanProbeTM to distinguish different lactate concentrations in liquid samples.

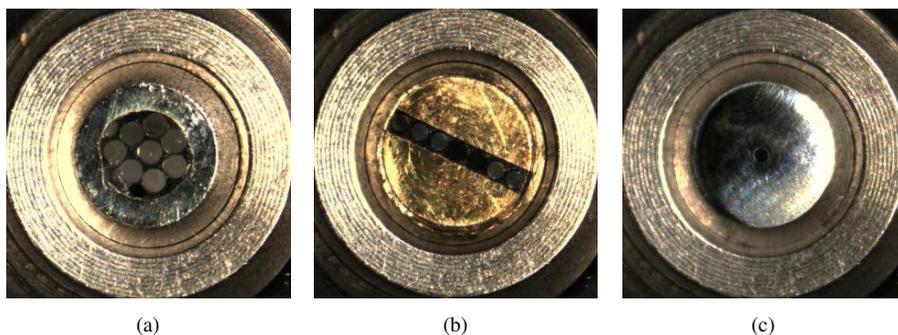


Figure 5.4: **(a)** Integrated fiber bundle of the Eagle Raman-S spectrometer. 8 fibers with a core diameter of $400\ \mu\text{m}$ and a numerical aperture of 0.3. **(b)** Linear arrangement of the collection fibre bundle at the proximal end of the EmVision Raman Lensed probe. 7 fibers with a core diameter of $300\ \mu\text{m}$ and a numerical aperture of 0.22. **(c)** Single collection fiber of the RamanProbeTM with a core diameter of $200\ \mu\text{m}$ and a numerical aperture of 0.22.

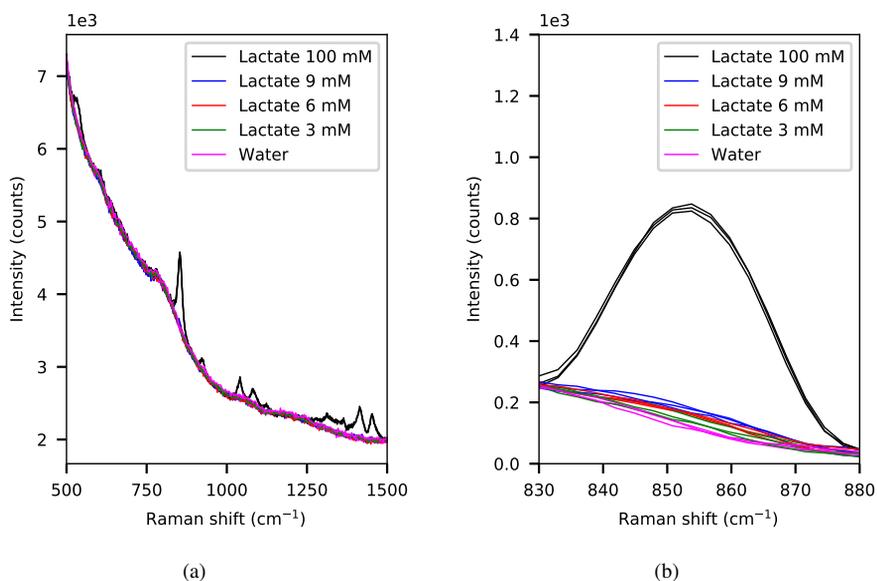


Figure 5.5: The RamanProbeTM in aqueous solutions. **(a)** Raw Raman spectra of aqueous solutions with varying concentrations of lactate in the region of $500\ \text{cm}^{-1}$ to $1500\ \text{cm}^{-1}$. **(b)** Characteristic peak of lactate at $853\ \text{cm}^{-1}$.

5.2 *In vivo* validation of the probe system

Prior to evaluating the capacity of the technology to detect hypoxic-ischemic events, measurement conditions were optimized. The mechanical protective jacket of the commercial RamanProbe™ has also been adapted to allow for contact measurements.

5.2.1 Optimization of the focal position

As previously described, the RamanProbe™ is not designed to be used in contact with the surface to be measured. The working distance of 5.6 mm is not ideal for *in vivo* measurements through the skin. The thickness of the human skin ranges from 0.5 mm to 4 mm. Thus, when the probe is directly placed onto the skin, the signal comes from somewhere below as the working distance is larger. Moreover, light absorption, transmission and scattering through the skin, as well as the depth of field and the energy density of the incident beam, affect the Raman intensity, which originates from an excitation cone rather than a unique point. Figure 5.6 shows the module, with adjustable working distance, that has been designed and implemented to carry out contact measurements. The module consists of two threaded pieces with a very fine pitch and a window-end. Such a configuration allows us to determine a precise focal position that maximizes the Raman signal at the contact surface.



Figure 5.6: Implementation of an opto-mechanical module to the RamanProbe™ that allows for contact measurements.

The variability of the Raman peaks has been analyzed for different tissues, such as the abdomen or the tongue of a pig, as a function of the position of the focus. From the

innermost position, the working distance was shifted by $500\ \mu\text{m}$ at each step for a total of 13 different positions and a range of 6.5 mm to guarantee that the analysis was done beyond the surface. Figure 5.7a and Figure 5.7b show the Raman intensity profile of different tissues which allows for identification of the position of major Raman intensity,

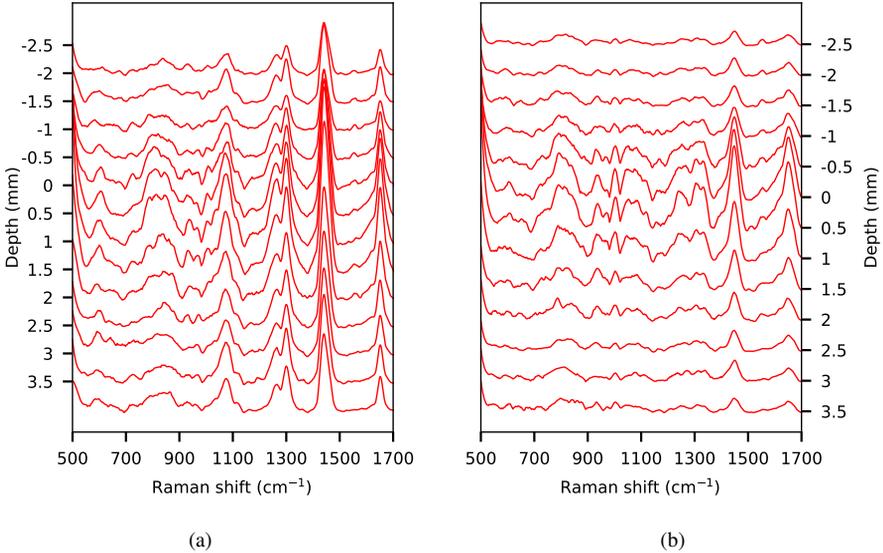


Figure 5.7: Raman intensity profile as a function of the focal position in the tissue, where 0 represents the surface, while positive and negative values correspond to inner and outer part of the skin, respectively. (a) Abdomen of a pig. (b) Tongue of a pig.

5.2.2 Tissue differentiation

Depending on their location, tissues of a living being show differences in their functionality and accordingly in their composition. Raman spectra from the abdomen, snout, ear and tongue of a piglet should exhibit distinct patterns according to the chemical structure of their components. As part of the evaluation process, our ability to distinguish different tissues by using the RamanProbeTM was investigated. Apart from the potential capacity of our technology, this approach gave us insight into the most appropriate tissue for further validation. The scalp was excluded due to the high fluorescence emitted by the skull. Possible interference of the bone components needs to be addressed in the development of a customized probe.

A total of 110 measurements (30 at the abdomen, snout and tongue; 20 at the ear) from three piglets were collected. An integration time of 60 s and a laser power of 30 mW were used. The same preprocessing procedure described in Section 4.1.3 was used here. Preprocessed Raman spectra are displayed in Figure 5.8, where different patterns of the Raman peaks are clearly distinguished among tissues. Even though the origin of the Raman peaks of each type of tissue is out of the scope of this study, Figure 5.9 demonstrates that the projection of the data in its first latent vectors (Partial Least Squares) increases the difference and makes the ability to distinguish the type of tissue more evident. Data from the ear was omitted to facilitate the visualization of the clusters. The tongue is selected for further experiments due its suitable Raman signal and high level of irrigation.

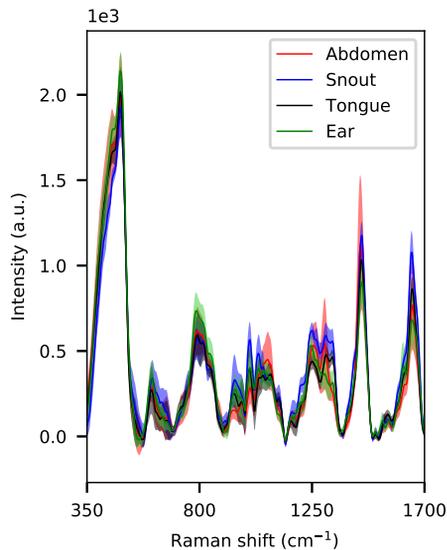


Figure 5.8: Averaged preprocessed Raman spectra from different piglet tissues \pm their standard deviation represented by the shaded area.

5.2.3 Pre- & post-mortem differentiation

In two of those piglets, our ability to distinguish completely extreme physiological situations was investigated again by using the RamanProbeTM. As an interesting example, Raman spectra from pre-mortem and post-mortem state were compared. In fact, a post-

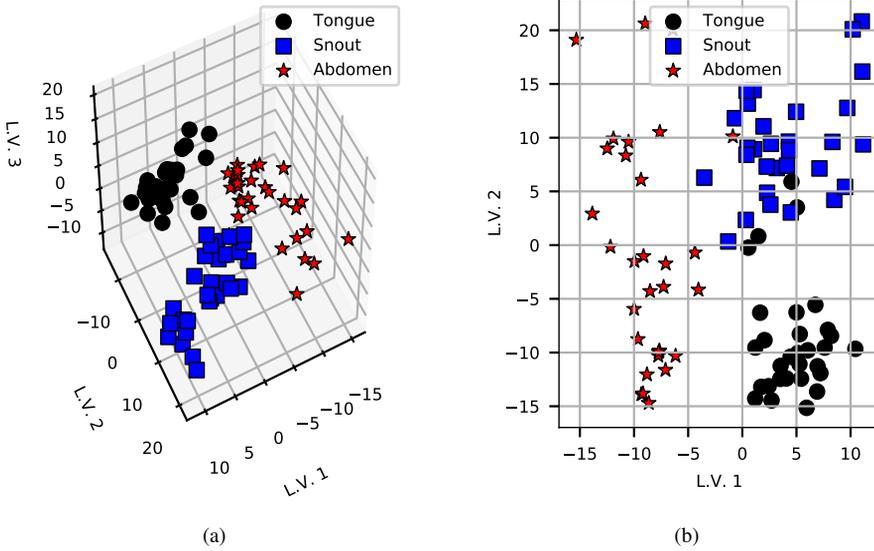


Figure 5.9: Partial Least Squares score plot of different piglet tissues. (a) 3-D projection of the Raman data into the first three latent vectors. (b) 2-D projection of the Raman data into the first two latent vectors.

mortem clinical state implies nothing else than an extreme hypoxic-ischemic situation, and such analysis would represent a real estimate of the potential of our technology to identify hypoxic-ischemic events.

A total of 65 measurements (32 pre-mortem and 33 post-mortem) were collected at the tongue. An integration time of 60 s and a laser power of 30 mw were used. Raman spectra, displayed in Figure 5.10, followed the same preprocessing procedure defined in Section 4.1.3. Despite the difficulty to identify clear characteristic patterns in the Raman curves, projection of the data into its first latent vectors in Figure 5.11 reveals two distinct clusters associated to the clinical state. This result represents a fantastic achievement in the process of identifying hypoxic-ischemic events.

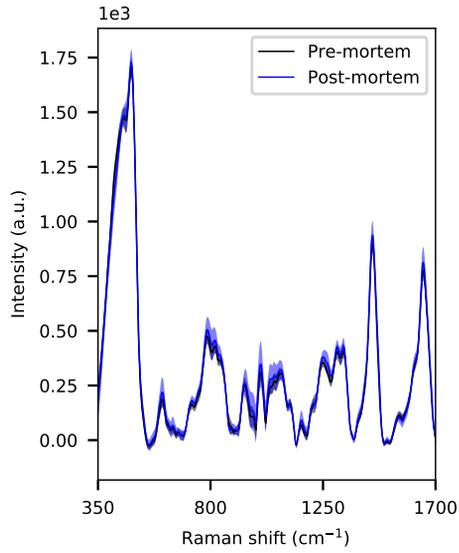


Figure 5.10: Averaged preprocessed Raman spectra from piglets at pre- and post-mortem clinical states \pm their standard deviation represented by the shaded area.

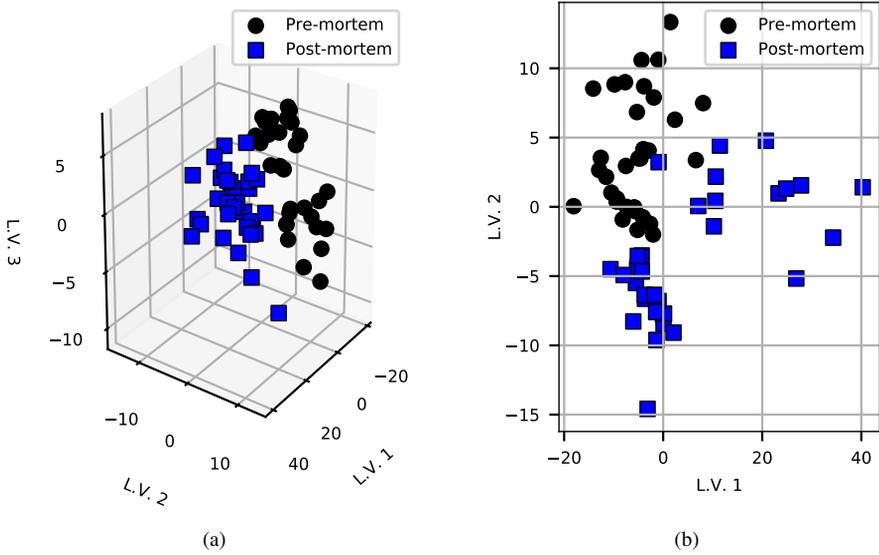


Figure 5.11: Partial Least Squares score plot of piglets at pre- and post-mortem clinical states. **(a)** 3-D projection of the Raman data into the first three latent vectors. **(b)** 2-D projection of the Raman data into the first two latent vectors.

In vivo non-invasive monitoring of hypoxia-ischemia

In this chapter, the new diagnostic tool, which combines Raman spectroscopy with machine learning, is validated *in vivo*. In this pre-clinical phase, the RamanProbe™ [96] described in Chapter 5 is tested for identification of hypoxic-ischemic events. The technology is evaluated as a genuine alternative for supporting in an innovative way immediate medical decision-making by non-invasive, continuous and real-time monitoring of perinatal asphyxia.

6.1 Material & Methods

6.1.1 Experimental procedure

9 piglets, 1 to 3-d old were anesthetized with sevoflurane (Induction 5% and maintenance 2.5%) and fentanyl (0.0002 – 0.0005 mg/kg/h), and then paralyzed by a perfusion of vecuronium (1.8 mg/kg/h) administered through the ear vein. Animals were then tracheotomized and ventilated with positive pressure. The femoral artery was cannulated to monitor blood pressure and to obtain blood samples. Blood oxygen saturation was monitored by transcutaneous pulse oximetry.

After stabilization, the animals were kept under normoxic conditions for 1 hour. The induction of hypoxic-ischemic damage was carried out by reducing the inspiratory oxygen fraction between 8 – 10% until reaching a base excess (BE) ≤ 15 mM and/or a pH ≤ 7 and/or a lactate ≥ 12 mM (moderate-severe hypoxic-ischemic damage). In addition, a mean arterial pressure < 20 mmHg was established as an end-point criterion for hypoxia, since in these cases the life of the animal is seriously compromised. To reduce the inspiratory fraction of oxygen below 21%, the ventilation system was modified, replacing the medical air supply with nitrogen and readjusting oxygen concentrations with an oximeter. In addition, CO₂ was administered in the inspiratory fraction (PaCO₂ 8.0 – 9.5 kPa), with the aim to mimic the perinatal asphyxia. After hypoxic damage, the inspired oxygen fraction was restored to baseline levels and kept for 1 hour. Thus, in total three phases were established throughout the experiment, as shown in Figure 6.1.



Figure 6.1: Experiment timeline divided in three phases, basal condition phase (60 minutes before the damage phase), hypoxic-ischemic damage phase (variable time until reaching one or more of the established thresholds), post-hypoxia-ischemia phase (60 minutes after the damage phase).

At the end of the experiment, the piglets were euthanized with an intravenous injection of potassium chloride. All experimental procedures and euthanasia of the

animals were conducted in strict compliance with European and Spanish regulations on the protection of animals used for scientific purposes (European Directive 2010/63/EU and Spanish Royal Legislative Decree 53/2013). The protocols were approved by the Committees on the Ethics of Laboratory Animal Welfare of Biodonostia Health Research Institute (Permit Numbers: OH 18_22 and OH 20_36) and performed in its experimental surgical theater.

6.1.2 Data acquisition: blood gas analysis and Raman spectra

In the current chapter both Raman spectra and physiological parameters were simultaneously acquired. Raman measurements were performed with the RamanProbe™ (RPS785, InPhotonics), which had been adapted to be used in contact with the skin. The RamanProbe™, with an excitation fiber of 105 nm core and a collection fiber of 200 nm, is connected to a continuous wave diode laser, emitting at 785 nm, and to the EAGLE Raman-S grating spectrometer with an Andor iVac 316 detector (Ibsen Photonics) and a spectral resolution of 4 cm^{-1} . Based on the previous work described in Section 5.2, the tongue was used as excitation tissue for identification of hypoxic-ischemic events (Figure 6.2). Even though Raman spectra were continuously collected, for each measurement the probe was replaced at a different position to avoid any possible bias or sample damage caused by long and permanent exposures. An integration time of 60 seconds and a power of 60 mW were used. In total, 1033 Raman spectra were analyzed. The database consists in three subsets, normoxia or basal condition phase (BC, $n = 346$), hypoxia-ischemia phase (HI, $n = 323$) and post-hypoxia-ischemia phase (HI_p, $n = 364$).

Throughout the experiment, blood samples were extracted from femoral artery for gas analysis with the i-STAT 1 handheld blood analyzer (Abbott Laboratories). Table 6.1 shows the 8 physiological parameters that were provided by the gas analyzer. Due to the amount of time required for the whole extraction and analysis procedure, blood samples were extracted every ten Raman measurements. Nonetheless, in order to match each Raman measurement with a blood gas analysis, a linear equation was fitted every two points for each physiological parameter, and values in between assigned to their corresponding spectrum. Thus, a parallel database with 1033 blood gas analyses has also been generated. Likewise, this second database is also divided in three same subsets, normoxia or basal condition phase (BC, $n = 346$), hypoxia-ischemia phase (HI, $n = 323$) and post-hypoxia-ischemia phase (HI_p, $n = 364$).



Figure 6.2: Experimental setup for *in vivo* non-invasive monitoring of hypoxia-ischemia (HI) by Raman spectroscopy in piglets. HI damage is induced by reducing the inspired oxygen fraction. The RamanProbe™ is placed at the tongue.

Table 6.1: Physiological parameters measured by blood gas analysis.

pH (6.88 - 7.55)	pCO ₂ (mmHg) (30.3 - 85.0)	pO ₂ (mmHg) (14 - 608)	[Base excess] (mM) (-20 - 13)
[HCO ₃ ⁻] (mmHg) (11.9 - 37.2)	TCO ₂ (mmHg) (13 - 39)	SO ₂ (mmHg) (13 - 100)	[Lactate] (mM) (0.41 - 14.2)

6.1.3 Preprocessing of Raman spectra

As explained in Section 4.1.3, analysis of Raman data in biological samples is extremely challenging due to their heterogeneous nature. Moreover, such a challenge is magnified when dealing with *in vivo* data. Therefore, prior to perform any statistical analysis or develop any predictive model, preprocessing of raw Raman data is essential for the correct interpretation of the Raman spectra and for a reliable classification of the samples. Even though the preprocessing methods described in Section 4.1.3 have been similarly applied to *in vivo* data, slight modifications were performed to optimize the Raman signal.

Accordingly, all Raman spectra were first trimmed to the range from 535 cm^{-1} to 1717 cm^{-1} , the so called fingerprint region. Then, a 6th order Extended Multiplicative Signal Correction (EMSC) was applied to the raw spectra to remove interfering additive and multiplicative artifacts by scaling all the Raman spectra to the mean spectrum [61]. An asymmetric least squares (ALS) method, based on Whittaker filter, was used to subtract a smoothed background produced by the intrinsic fluorescence of the molecules [60]. The integrity of the Raman bands is retained by giving higher relevance to positive residuals [57]. All Raman spectra were smoothed using a Savitzky-Golay filter with a window of 15 points and 3rd order polynomial at a sampling of 4 cm^{-1} . Finally, to avoid uncertainties of the variance of the mean, the origin of the coordinate system is translated to the center of gravity of the dataset. Some machine learning algorithms are built upon the variance of specific features of data, and then, it is important to center the data relative to a reference point, which in this case is the mean. Combination of all data preprocessing has been optimized based on maximization of Raman features and performance of the predictive models.

Even though the experiments were carried out in a well-controlled experimental context, *in vivo* measurements imply a considerable degree of uncertainty. In order to avoid the presence of outliers in the database, a principal component analysis (PCA) is individually applied to the Raman data of each animal. Outliers are identified based on a Q residuals vs. Hotelling's T^2 plot. Such a plot represents the lack-of-fit and the captured variation of each sample within the model. Samples out of the confidence limit of 95% were not considered and were discarded for the proceeding data analysis. All data preprocessing was carried out with the PLS_Toolbox (Eigenvector) in MATLAB

environment (MathWorks®).

6.1.4 Data analysis

Univariate analysis

Similar to Section 3.1.2, physiological parameters depicted in Table 6.1 have individually been assessed. Box-and-whisker plots were used to depict data distribution. Non-parametric Mann-Whitney U-test was applied to determine statistical significance among groups, while Spearman's correlation is used to measure the monotonic relationship between physiological parameters and the clinical state. Finally, predictive power of lactate and pH for classification of HI was examined by means of statistical metrics derived from a confusion matrix, including sensitivity, specificity, accuracy, positive and negative predictive value (PPV, NPV), and the area under the receiver operating characteristic curve (AUROC). Again, definition of perinatal asphyxia was based on clinical cut-off limits of $\text{pH} < 7.20$ and $[\text{lactate}] > 4.8 \text{ mM}$, as proposed by Saling [79] and Kruger et al. [80], respectively.

Multivariate analysis

It is well-known that a hypoxic-ischemic event induces a physiological response, which has been demonstrated in Chapter 3 to cause a generalized systemic reaction rather than the variation of a single biomarker. Such a physiological response is reflected in distinct spectral features of Raman bands. However, identification and assignment of peaks, and consequently, differentiation among groups constitutes a serious challenge, which is even more evident when considering *in vivo* data classification. A number of different classification algorithms – as single classifier, ensemble learning and hybrid – were used for the detection of HI. Specifically, (i) partial least squares – Discriminant analysis (PLS – DA) [71]; (ii) random forest (RF) [73]; (iii) support vector machine (SVM) [72]; (iv) adaptive boosting (AB) [74]; (v) gradient boosting (GB) [74, 75]; and (vi) extreme gradient boosting (XGB) [76] have been tested. Fundamental concepts of applied algorithms can be found in more detail in Section 2.6.

Relevant spectral bands for identification and classification of the clinical state, associated to defined phases in Section 6.1.1, were determined. Their contribution to

different predictive models was calculated by the permutation importance technique. The accuracy decrease of each optimized predictive model was calculated when a single wavenumber was randomly shuffled. The difference reveals how decisive that spectral feature is for generating a prediction.

Additionally, hybrid classifiers were tested. Considering the large number of variables related to wavenumbers obtained by Raman spectroscopy, feature extraction and dimensionality reduction prior to feeding machine learning algorithms could result in a substantial improvement of their performance. First, Raman data was projected into the optimum number of latent vectors. Redundant or irrelevant features were removed by projecting the data in a new reduced subset of latent vectors sorted by the amount of explained variance and extracted from combinations of the original features. A second PLS was applied to this new projected data such that most of the variance was captured and illustrated by the first two components. The classifier is then constructed in this new latent space [81, 82]. This type of hybrid classifiers allows for meaningful graphical representations with no loss of any relevant information.

Due to the limited number of animals ($n = 9$), the procedure of data splitting into calibration and validation datasets was avoided. Instead, a leave-one-subject-out cross-validation (LOSOCV) method was implemented for predictive model construction and optimization. Similar to k-fold cross-validation, data is divided in groups. The characteristic of LOSOCV is that groups are formed with data from a unique subject. Therefore, each subject is tested only once and one at a time, while the remaining subjects are used for training. Avoiding data from a same pig to be in different groups prevents the model from overfitting and better simulates a real predictive situation of unknown data. Parameters were tuned according to the best averaged cross-validated performance by selecting balanced accuracy as scoring metric for evaluation.

The following three classification problems have been considered:

1. Basal condition vs. hypoxia-ischemia,
2. Basal condition vs. post-hypoxia-ischemia,
3. Basal condition vs. hypoxia-ischemia vs. post-hypoxia-ischemia.

All data analysis was performed in Spyder, a Python open source integrated devel-

opment environment. Sklearn, Scipy, Pandas and Numpy libraries were mainly used for the analysis and development of predictive models, while Seaborn and Matplotlib were applied to display plots and figures.

6.2 Results

6.2.1 Induction of hypoxia-ischemia

While BC and HI_p phases were defined to last 60 minutes, the duration of the HI phase was variable as the time of reaching the end-point criterion was different for each experiment. The mean duration of the HI damage was 49 ± 15 minutes. For seven of the piglets, experimental procedure was carried out in accordance with the protocol defined in Section 6.1.1 and no difficulties were reported. The piglet with replicate ID n° 6 suffered a cardiac arrest during the HI phase after 19 minutes of damage and fulfillment of end-point criterion. Despite this unexpected setback, experiment could be completed after resuscitation. Similarly, the piglet with replicate ID n° 9 also suffered a cardiac arrest. In this case, the heart failure happened when normal oxygenation was restored. Experiment could not be completed since we did not succeed in the resuscitation. Therefore, there is no available data from HI_p phase for the piglet with replicate ID n° 9.

6.2.2 pH and lactate as gold standard parameters

Figure 6.3 reveals that induction of an HI event is reflected in all measured physiological parameters, with a more or less pronounced shift from their basal values. Such a result supports conclusions obtained in Chapter 3 and confirms our initial hypothesis that a prediction based on systemic variations would provide a much better performance. The Mann-Whitney U-test unveils statistical significance for all parameters, including O₂ and pO₂, both for HI (Table 6.2) and HI_p (Table 6.3) conditions when compared to BC. However, as the sample size is relatively large, p-value is not enough to assume a significant difference [100]. For large datasets, the p-value of a statistical test is very likely to be significant even for minor variations. Hence, the effect size describes the real magnitude of that difference and is reported by the rank-biserial correlation. The simplified coefficient proposed by Wendt et al. [101] and defined as,

$$\text{Effect size} = 1 - \frac{2U}{n_1 n_2}, \quad (6.1)$$

takes values between 0, for no correlation, and 1, for maximum correlation, where U is the test statistic, and n_1 and n_2 are the group sizes. Even though the magnitude of the effect size is considerable for all parameters during HI (Table 6.2), the effect size for SO_2 and pO_2 is negligible when HI_p is considered (Table 6.2).

Accordingly, parameters such as lactate, pH, base excess, TCO_2 or HCO_3^- show a much slower recovery to initial values, when ventilation is restored after a severe HI event. Indeed, these parameters show strong correlation with the clinical state (BC, HI, HI_p), resulting in an absolute Spearman correlation coefficient close to 0.8. While lactate exhibits a positive correlation with the occurrence of an HI event, the rest of relevant parameters show a negative correlation.

Table 6.2: Statistical significance of physiological parameters determined by Mann-Whitney U-test between basal condition (BC) and hypoxia-ischemia (HI) condition.

Parameter	U-statistic	p-value	Effect size
SO_2	0.0	<0.001*	1.0
pO_2	0.0	<0.001*	1.0
Lactate	1862.0	<0.001*	0.96
Base excess	6118.0	<0.001*	0.89
pH	6171.5	<0.001*	0.89
HCO_3^-	8218.5	<0.001*	0.85
TCO_2	9317.5	<0.001*	0.83
pCO_2	25688.0	<0.001*	0.54

Mann-Whitney U-test

* p-value ≤ 0.05

Binary classification of BC and HI conditions for clinical pH and lactate cut-off limits is displayed in Figure 6.4b. The figures of merit in Table 6.5 summarize their classification performance. Similar to the results achieved in Chapter 3, lactate demonstrates to be a reasonably better indicator of an HI event than pH. Likewise, the little capacity of current standards to detect positive cases provides a poor sensitivity of 55.7% for pH and 72.1% for lactate. The time lag between the beginning of the HI event and predefined cut-off limits being exceeded is responsible for this poor outcome. Unlike in Chapter 3, the effect of this delay in their classification performance is slightly mitigated as hypoxia-ischemia is prolonged beyond 20 minutes, giving additional time to both pH and lactate to exceed those thresholds in most of the blood gas analyses

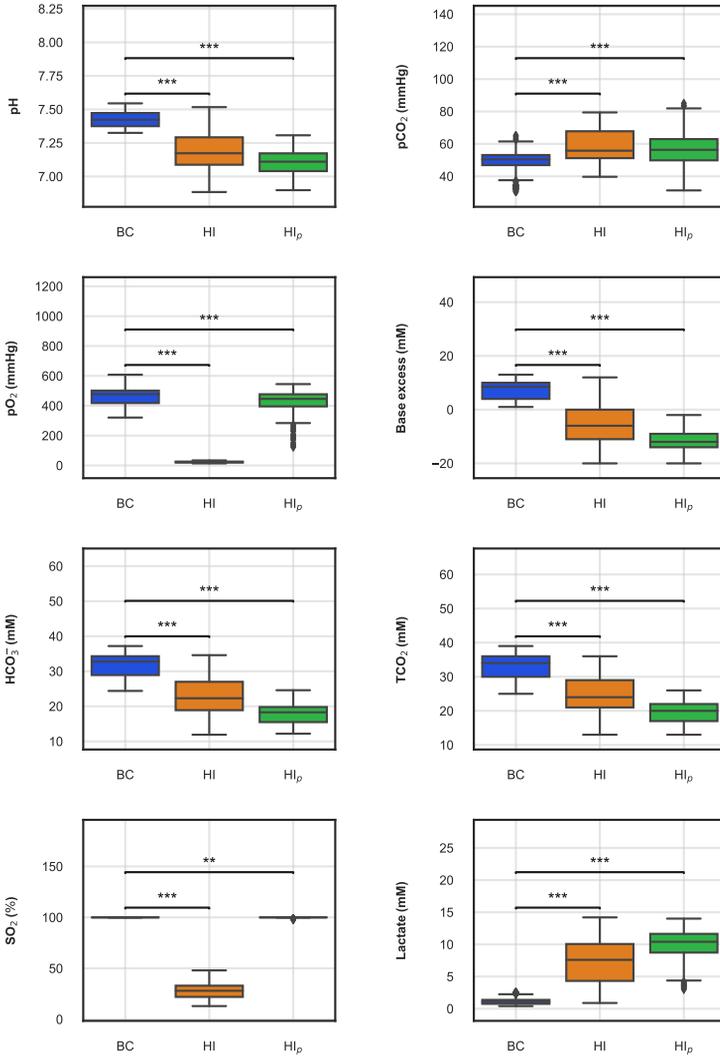


Figure 6.3: Exploratory analysis of basal condition (BC), hypoxia-ischemia event (HI) and post-hypoxia-ischemia (HI_p) groups for different physiological parameters. Box-and-whisker plot facilitates the comparison between the groups by displaying the distribution and skewness of the data.

Mann-Whitney U-test
 * 0.01 < p-value ≤ 0.05
 ** 0.001 < p-value ≤ 0.01
 *** p-value ≤ 0.001

Table 6.3: Statistical significance of physiological parameters determined by Mann-Whitney U-test between basal condition (BC) and post-hypoxia-ischemia (HI_p) condition.

Parameter	U-statistic	p-value	Effect size
Lactate	0.0	<0.001*	1.0
Base excess	0.0	<0.001*	1.0
pH	0.0	<0.001*	1.0
HCO ₃ ⁻	2.5	<0.001*	1.0
TCO ₂	38.0	<0.001*	1.0
pCO ₂	33258.0	<0.001*	0.47
pO ₂	47549.5	<0.001*	0.24
SO ₂	61415.0	<0.01*	0.02

Mann-Whitney U-test

* p-value ≤ 0.05

Table 6.4: Spearman's correlation between physiological parameters and the clinical state associated to established phases (basal condition, hypoxia-ischemia, post-hypoxia-ischemia).

Parameter	Coefficient	p-value
Lactate	0.81	<0.001*
pCO ₂	0.33	<0.001*
SO ₂	0.01	0.84
pO ₂	-0.10	<0.001*
pH	-0.76	<0.001*
TCO ₂	-0.81	<0.001*
Base excess	-0.81	<0.001*
HCO ₃ ⁻	-0.81	<0.001*

Spearman's correlation

* p-value ≤ 0.05

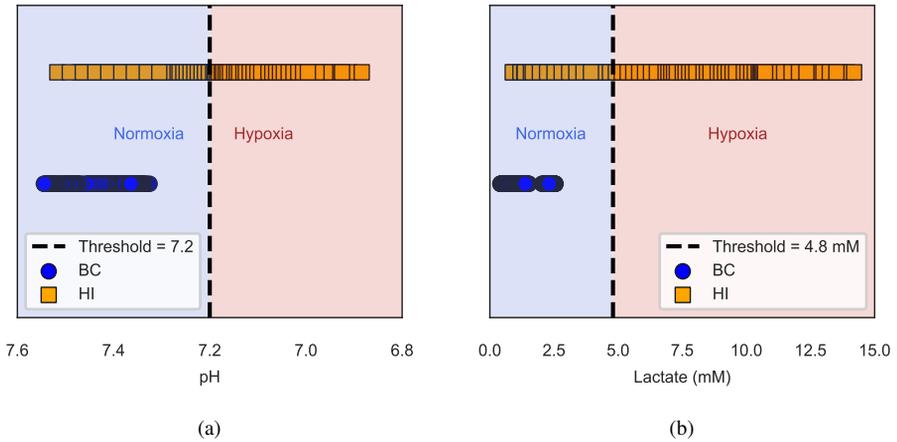


Figure 6.4: 1-D visual class separability between basal condition (BC) and hypoxia-ischemia (HI) based on standard clinical cut-off limits. **(a)** Decision region of binary classification (BC vs. HI) based on pH threshold = 7.2. **(b)** Decision region of binary classification (BC vs. HI) based on [lactate] threshold = 4.8 mM.

Table 6.5: Classification performance of pH and lactate standard cut-off limits for differentiation of basal condition (BC) and hypoxia-ischemia (HI) condition; all figures of merit are in percentage (%).

	pH < 7.2	[Lactate] > 4.8 mM
Sensitivity	55.7	72.1
Specificity	100	100
PPV	100	100
NPV	70.8	79.4
Accuracy	78.6	86.5
AUROC	77.9	86.1

carried out at that phase.

Performance of pH and lactate cut-off limits improves for binary classification between BC and HI_p conditions due to a considerable increase of their sensitivity, as shown in Table 6.6. As the HI phase is not considered, the delay between the beginning of the HI and the exceeding of the cut-off limits is not reflected in the biochemical data from blood gas analyses. Instead, values corresponding to HI_p condition already start far above established pH and lactate thresholds. Moreover, induction of a severe HI results in a slow recovery to basal condition, which contributes to a greater class separability in terms of AUROC and, consequently, a better classification performance (Figure 6.5).

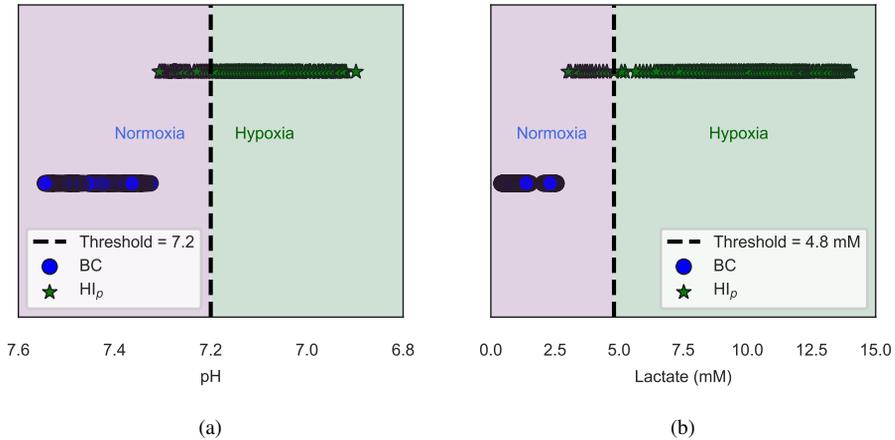


Figure 6.5: 1-D visual class separability between basal condition (BC) and post-hypoxia-ischemia (HI_p) based on standard clinical cut-off limits. (a) Decision region of binary classification (BC vs. HI_p) based on pH threshold = 7.2. (b) Decision region of binary classification (BC vs. HI_p) based on [lactate] threshold = 4.8 mM.

Table 6.6: Classification performance of pH and lactate standard cut-off limits for differentiation of basal condition (BC) and post-hypoxia-ischemia (HI_p) condition; all figures of merit are in percentage (%).

	pH < 7.2	[Lactate] > 4.8 mM
Sensitivity	79.7	95.9
Specificity	100	100
PPV	100	100
NPV	82.3	95.8
Accuracy	89.6	97.9
AUROC	89.8	97.9

When all three phases are considered, prediction based on univariate analysis of a single parameter such as pH or lactate, is not feasible since additional valuable information for differentiation cannot be provided. Apart from established cut-off limits, no further distinction is done among the groups. As a result, an overlap between hypoxia-ischemia and post-hypoxia-ischemia is observed in Figure 6.6b.

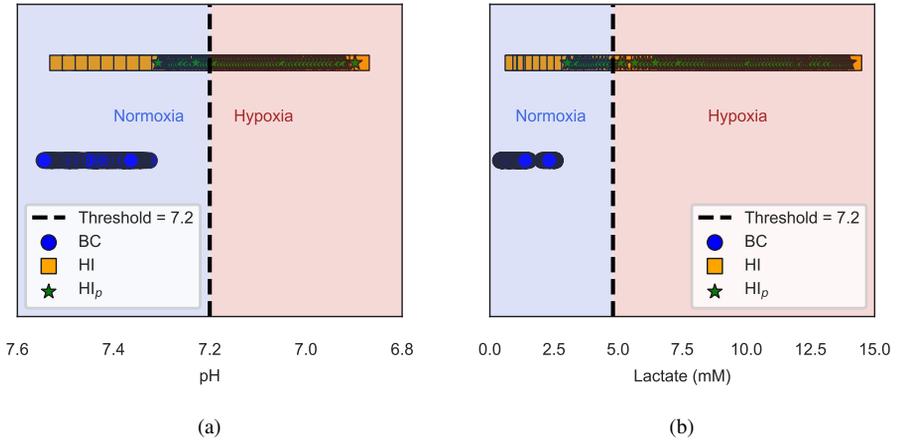


Figure 6.6: 1-D visual multiclass separability between basal condition (BC), hypoxia-ischemia (HI) and post-hypoxia-ischemia (HI_p) based on standard clinical cut-off limits. (a) Decision region of binary classification (BC vs. HI vs. HI_p) based on pH threshold = 7.2. (b) Decision region of binary classification (BC vs. HI vs. HI_p) based on [lactate] threshold = 4.8 mM.

6.2.3 Predictive models based on Raman spectra

The combination of Raman spectroscopy with machine learning algorithms is tested as an alternative to current gold standard method. The performance of different predictive models is evaluated for the classification problems stated above. Processed Raman spectra are shown in Figure 6.7.

Basal condition vs. hypoxia-ischemia

Figures of merit of different classifiers for differentiation of BC and HI phases are displayed in Table 6.7. Apart from SVM, which exhibits an accuracy of 85.4%, all predictive models present an accuracy > 87.4%, which is slightly superior to the per-

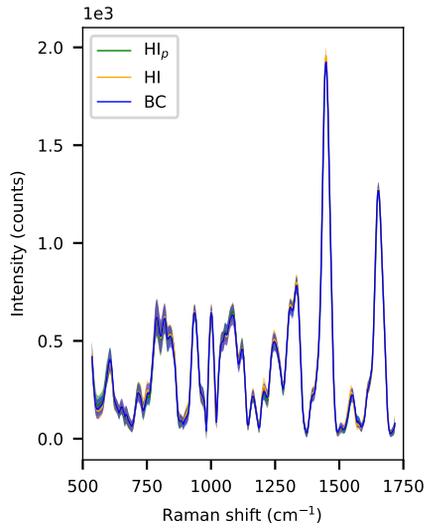


Figure 6.7: Averaged preprocessed Raman spectra from different defined phases \pm their standard deviation represented by the shaded area.

formance achieved by the cut-off limit of lactate; even greater is the difference with respect to pH. The substantial improvement of the sensitivity ($> 88\%$) achieved by the predictive models constitutes the major variation in the classification of BC and HI phases.

Table 6.7: Classification performance of different machine learning algorithms for differentiation of basal condition (BC) and hypoxia-ischemia (HI) phases; all figures of merit are in percentage (%).

	PLS-DA	RF	SVM	AB	GB	XGB
Sensitivity	89.5	88.9	88.5	88.2	89.5	90.4
Specificity	86.4	86.1	82.4	87.6	89.0	89.0
PPV	86.0	85.7	82.4	86.9	88.4	88.5
NPV	89.8	89.2	88.5	88.9	90.1	90.9
Accuracy	87.9	87.4	85.4	87.9	89.2	89.7
AUROC	94.7	93.7	93.4	92.5	94.9	95.4

To understand how differentiation of BC and HI phases is done by different algorithms, the accuracy decrease, obtained by feature permutation of PLS-DA and XGB predictive models, is displayed in Figure 6.8a and Figure 6.8b, respectively. Significant Raman features are determined as the prediction error increases when these features

are shuffled. While PLS–DA, which maximizes the covariance between the matrix of observation and the matrix of outcomes, gets more or less affected by every feature permutation, XGB algorithm demonstrates to be much more robust to variations.

According to Figure 6.8a, the PLS–DA predictive model identifies a number of relevant Raman features in the region from 1100 cm^{-1} to 1650 cm^{-1} that arises from Raman active blood components such as proteins or heme groups [93]. Prominent peaks at 1222 cm^{-1} and 1547 cm^{-1} have already been assigned to oxygenated and deoxygenated red blood cells, respectively [93, 102, 103]. Interestingly, bands associated to lactate at 853 cm^{-1} from a C–C aliphatic stretching, 1053 cm^{-1} from C–CH₃ vibration, or 1420 cm^{-1} from COO⁻ stretching are also observed. Raman band around 543 cm^{-1} is probably the overlap of bands associated to the CO₂⁻ wagging of lactate and Fe–O₂ stretching of the hemoglobin.

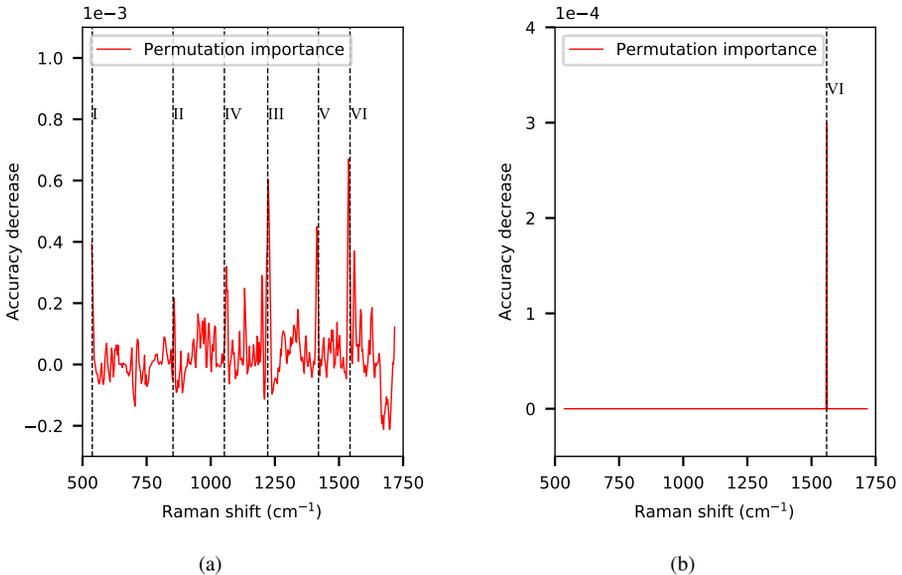


Figure 6.8: Permutation feature importance defined as accuracy decrease of fitted predictive models for BC vs. HI classification. (a) Averaged decrease of accuracy and standard deviation of fitted Partial Least Squares-Discriminant Analysis (PLS–DA) predictive models due to single parameter permutation. (b) Averaged decrease of accuracy and standard deviation of fitted Extreme Gradient Boosting (XGB) predictive models due to single parameter permutation. (Peaks: I. 543 cm^{-1} , II. 853 cm^{-1} , III. 1053 cm^{-1} , IV. 1222 cm^{-1} , V. 1420 cm^{-1} , VI. 1547 cm^{-1})

In the case of PLS-based hybrid classifiers, similar predictive power is achieved. Accuracies of non-hybrid classifiers reach values of 85.4 - 89.7%, while hybrid classifiers lie within the range of 85.2 - 87.7%. Interestingly, figures of merit of PLS – SVM even improve compared to its usage alone. As previously mentioned, PLS is applied twice such that the information from relevant features is represented by only two PLS latent vectors. To guarantee that applying PLS a second time does not affect the data and exclude relevant information, performance has first been compared to the single use of PLS, where the classifier is built in a latent space formed by the optimum number of latent vectors. Even though it is not shown, it has been observed that applying PLS twice does not result in any significant variation in the performance. Figure 6.9a and Figure 6.9b illustrate the decision region of SVM and RF classifiers in one step of the LOSOCV, respectively.

Table 6.8: Classification performance of different hybrid machine learning algorithms for differentiation of basal condition (BC) and hypoxia-ischemia (HI) phases; all figures of merit are in percentage (%).

	PLS – RF	PLS – SVM	PLS – AB	PLS – GB	PLS – XGB
Sensitivity	88.9	89.5	87.0	87.3	88.2
Specificity	83.5	86.1	85.8	83.2	83.2
PPV	83.4	85.8	85.2	82.9	83.1
NPV	88.9	89.8	87.6	87.6	88.3
Accuracy	86.1	87.7	86.4	85.2	85.7
AUROC	92.5	94.7	93.9	92.6	93.6

Basal condition vs. post-hypoxia-ischemia

When ventilation is restored and blood oxygenation returns to its basal levels, identifying that an HI event has occurred becomes more challenging. This is reflected in the figures of merit obtained by different classifiers for differentiation of BC and HI_p phases, which are depicted in Table 6.9. As time passes, the difficulty to detect physiological changes caused by the HI event leads to a general decrease in the predictive power of the classifiers, whose accuracies lie within the range of 68.9 - 74.4.

Raman feature importance of PLS – DA and XGB, which additionally have the best performance among the classifiers (Table 6.9), is calculated by permutation importance technique and displayed in Figure 6.10a and Figure 6.10b, respectively. In the PLS – DA model, the contribution of the region from 1100 cm⁻¹ to 1650 cm⁻¹ is reduced for identification of the HI_p phase. Contributions from 1222 cm⁻¹ and 1547 cm⁻¹ associated

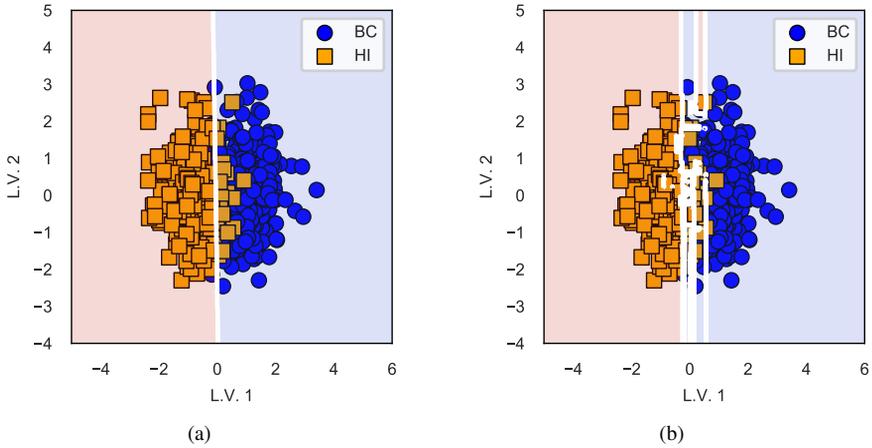


Figure 6.9: Example of 2-D visual class separability between basal condition (BC) and hypoxia-ischemia (HI) phases by hybrid classification models optimized by a leave-one-subject-out cross-validation. **(a)** Binary classification (BC vs. HI) based on the decision region defined by an SVM classifier in a 2-PLS component dataset. **(b)** Binary classification (BC vs. HI) based on the decision region defined by an RF classifier in a 2-PLS component dataset.

Table 6.9: Classification performance of different machine learning algorithms for differentiation of basal condition (BC) and post-hypoxia-ischemia (HI_p) phases; all figures of merit are in percentage (%).

	PLS-DA	RF	SVM	AB	GB	XGB
Sensitivity	80.8	70.1	77.5	76.6	76.9	78.6
Specificity	67.6	67.6	68.8	66.2	67.9	67.6
PPV	72.4	69.5	72.3	70.5	71.6	71.9
NPV	77.0	68.2	74.4	72.9	73.7	75.0
Accuracy	74.4	68.9	73.2	71.5	72.5	73.2
AUROC	82.5	75.3	81.6	77.5	78.8	79.6

with blood oxygenation are still present. New bands around 620 cm^{-1} or 1120 cm^{-1} are assigned to protein denaturation or heme aggregates caused by the HI event, while the peak at 1639 cm^{-1} is a marker for hemoglobin oxygenation. Prominent contribution of phenylalanine band from proteins around 1005 cm^{-1} is also observed [93, 102, 103]. Moreover, contribution of peaks associated to lactate at 543 cm^{-1} , but mainly at 853 cm^{-1} , is evident. Contribution of lactate is even more obvious in the case of XGB model, where only permutations at 853 cm^{-1} and 1459 cm^{-1} produce a decrease in the prediction accuracy.

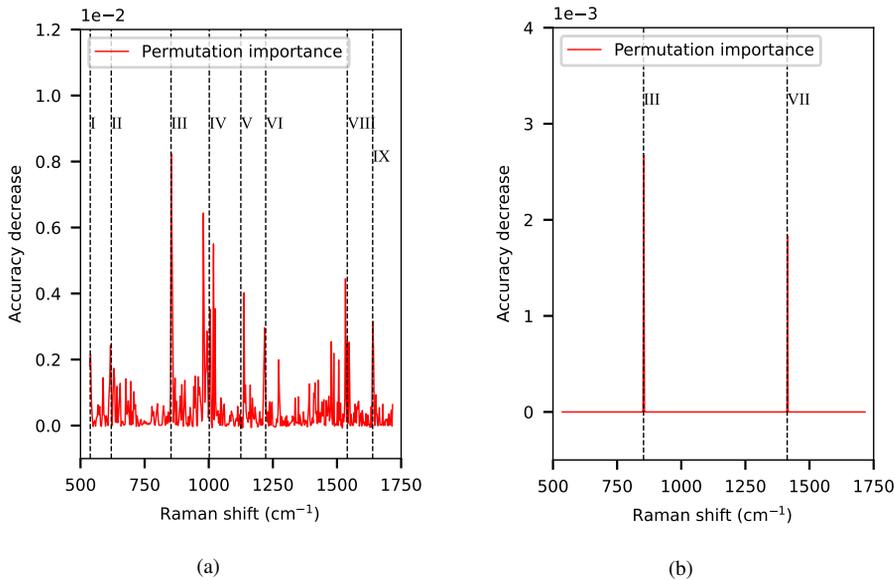


Figure 6.10: Permutation feature importance defined as accuracy decrease of fitted predictive models for BC vs. HI_p classification. **(a)** Averaged decrease of accuracy of fitted Partial Least Squares-Discriminant Analysis (PLS-DA) predictive models due to single parameter permutation. **(b)** Averaged decrease of accuracy of fitted Extreme Gradient Boosting (XGB) predictive models due to single parameter permutation. (Peaks: I. 543 cm^{-1} , II. 620 cm^{-1} , III. 853 cm^{-1} , IV. 1002 cm^{-1} , V. 1125 cm^{-1} , VI. 1222 cm^{-1} , VII. 142 cm^{-1} , VIII. 1547 cm^{-1} , IX. 1639 cm^{-1})

Unlike for the differentiation of HI phase, when it comes to HI_p , all hybrid predictive models show improved predictive power in comparison to non-hybrid models. Identifying an HI event after the ventilation has been restored poses additional difficulty, given the recovery of the physiological parameters to basal levels and, consequently, the greater similarity of the BC and HI_p Raman curves. In this case, dimensionality

reduction represents a useful preprocessing step for complex data. An example of the decision region of SVM and RF classifiers for differentiation of BC and HI_p is shown in Figure 6.11a and Figure 6.11b, respectively.

Table 6.10: Classification performance of different hybrid machine learning algorithms for differentiation of basal condition (BC) and post-hypoxia-ischemia (HI_p) phases; all figures of merit are in percentage (%).

	PLS-RF	PLS-SVM	PLS-AB	PLS-GB	PLS-XGB
Sensitivity	80.2	81.3	81.9	81.9	81.3
Specificity	69.9	67.9	68.8	69.7	69.1
PPV	73.7	72.7	73.4	73.9	73.4
NPV	77.1	77.6	78.3	78.5	77.9
Accuracy	75.2	74.8	75.5	75.9	75.4
AUROC	82.1	82.4	79.8	80.1	80.0

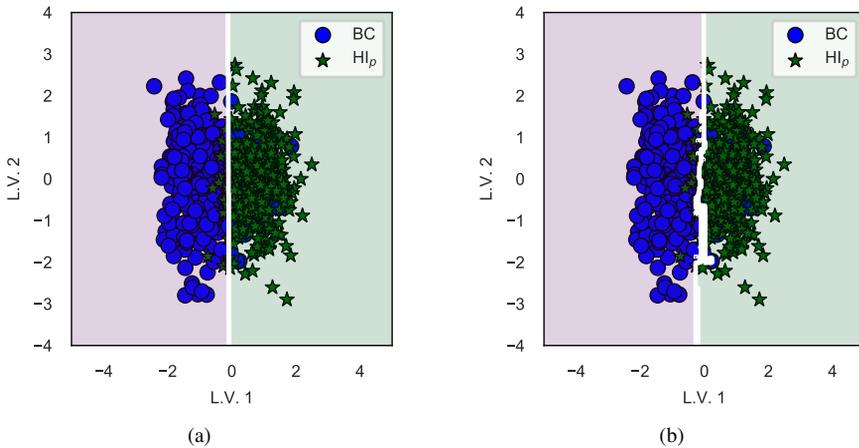


Figure 6.11: Example of 2-D visual class separability between basal condition (BC) and post-hypoxia-ischemia (HI_p) phases by hybrid classification models optimized by a leave-one-subject-out cross-validation. **(a)** Binary classification (BC vs. HI_p) based on the decision region defined by an SVM classifier in a 2-PLS component dataset. **(b)** Binary classification (BC vs. HI_p) based on the decision region defined by an RF classifier in a 2-PLS component dataset.

Basal condition vs. hypoxia-ischemia vs. post-hypoxia-ischemia

Predictive models based on multiparametric data such as Raman spectra lead to a better representation of the clinical picture. Raman bands provide additional information associated to systemic physiological variations, which allows us to classify all different

phases regarding an HI event. Figures of merit of different classifiers for differentiation of BC, HI and HI_p phases are displayed in Table 6.11. All of them present a total accuracy around 80%, similar to that obtained for differentiation of only BC and HI_p; a reasonable outcome considering the similarities between groups.

Table 6.11: Classification performance of different machine learning algorithms for multiclass differentiation of basal condition (BC), hypoxia-ischemia (HI) and post-hypoxia-ischemia (HI_p) phases; all figures of merit are in percentage (%).

	PLS-DA	RF	SVM	AB	GB	XGB
Sensitivity	73.0	68.6	72.3	68.1	71.6	71.8
Specificity	86.3	84.0	85.9	83.7	85.6	85.6
PPV	72.7	68.1	72.1	68.5	71.6	71.7
NPV	86.5	84.2	86.0	83.7	85.7	85.7
Accuracy	81.9	78.8	81.3	78.5	80.9	81.0

Consequently, the permutation feature importance of PLS-DA in Figure 6.12a is a combination of those obtained for differentiation of HI (Figure 6.8a) and HI_p (Figure 6.10a) independently. Thus, Raman bands at 543 cm⁻¹, 571 cm⁻¹, 620 cm⁻¹, 853 cm⁻¹, 1005 cm⁻¹, 1053 cm⁻¹, 1125 cm⁻¹, 1222 cm⁻¹, 1420 cm⁻¹, 1459 cm⁻¹, 1547 cm⁻¹ and 1639 cm⁻¹ are identified as relevant for detection of an HI event and classification of BC, HI and HI_p phases. XGB is affected only by permutations at 1547 cm⁻¹.

Hybrid classifiers have demonstrated to yield a predictive power similar, or even slightly better in some cases, to non-hybrid ones; corresponding figures of merit are depicted in Table 6.12. As stated before, a big advantage of hybrid predictive models is the possibility to graphically illustrate the decision region defined by different classifiers. Figure 6.13a and Figure 6.13b are explicit meaningful representations of a multiclass classification, where BC, HI and HI_p phases are differentiated according to their Raman spectra by PLS-SVM and PLS-RF classifiers, respectively.

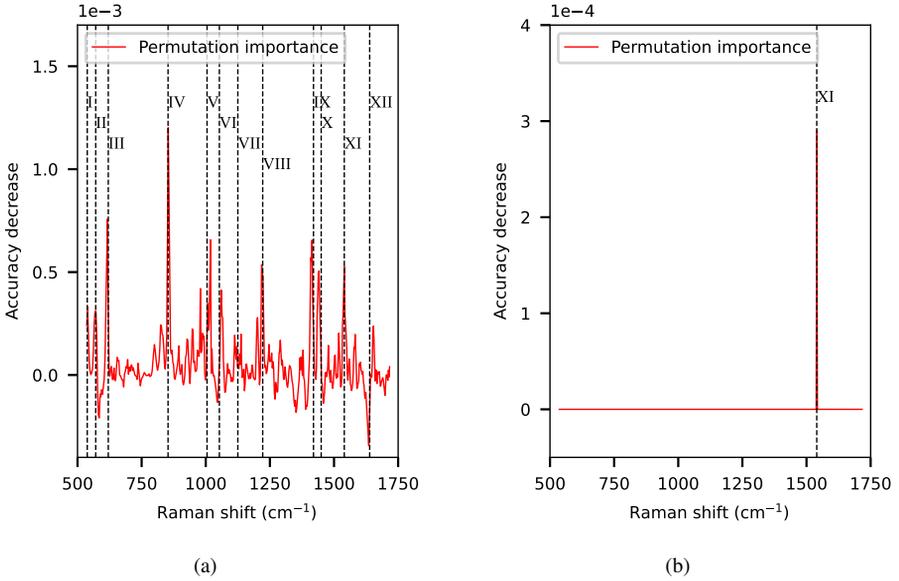


Figure 6.12: Permutation feature importance defined as accuracy decrease of fitted predictive models for BC vs. HI vs. HI_p classification. **(a)** Averaged decrease of accuracy of fitted Partial Least Squares-Discriminant Analysis (PLS-DA) predictive models due to single parameter permutation. **(b)** Averaged decrease of accuracy of fitted Extreme Gradient Boosting (XGB) predictive models due to single parameter permutation. (Peaks: I. 543 cm⁻¹, II. 571 cm⁻¹, III. 620 cm⁻¹, IV. 853 cm⁻¹, V. 1002 cm⁻¹, VI. 1053 cm⁻¹, VII. 1125 cm⁻¹, VIII. 1222 cm⁻¹, IX. 1420 cm⁻¹, X. 1459 cm⁻¹, XI. 1547 cm⁻¹, XII. 1639 cm⁻¹)

Table 6.12: Classification performance of different hybrid machine learning algorithms for multiclass differentiation of basal condition (BC), hypoxia-ischemia (HI) and post-hypoxia-ischemia (HI_p) phases; all figures of merit are in percentage (%).

	PLS-RF	PLS-SVM	PLS-AB	PLS-GB	PPLS-XGB
Sensitivity	72.2	73.2	72.8	72.2	71.7
Specificity	85.9	86.4	86.2	85.9	85.6
PPV	72.1	73.1	72.6	72.2	71.7
NPV	85.9	86.5	86.2	86.0	85.7
Accuracy	81.3	82.0	81.7	81.3	81.0

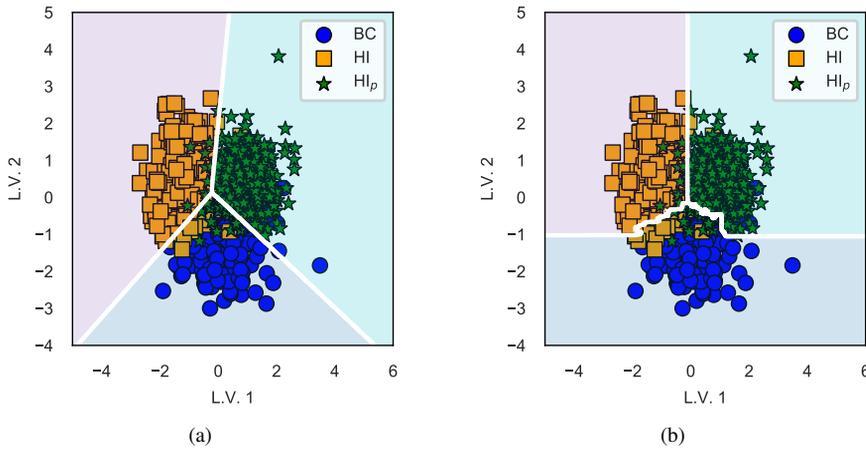


Figure 6.13: Example of 2-D visual multiclass separability between basal condition (BC), hypoxia-ischemia (HI) and post-hypoxia-ischemia (HI_p) phases by hybrid classification models optimized by a leave-one-subject-out cross-validation. **(a)** Multiclass classification (BC vs. HI vs. HI_p) based on the decision region defined by an SVM classifier in a 2-PLS component dataset. **(b)** Multiclass classification (BC vs. HI vs. HI_p) based on the decision region defined by a RF classifier in a 2-PLS component dataset.

6.3 Discussion

Apart from operating in an invasive manner, fetal scalp blood sampling (FSBS) has proven limited evidence for accurate identification of perinatal asphyxia. Therefore, the need of a new diagnostic tool that overcomes these limitations is clear. Our technology, which combines Raman spectroscopy with machine learning, exhibits great potential for accurate and robust asphyxia monitoring.

Statistical tests, such as Mann-Whitney test or Spearman's correlation, have demonstrated that inducing HI event disturbs the basal state of all measured physiological parameters. Following the gold standard technique, pH and lactate have been used as decision-making parameter. Both of them present difficulties to detect an HI event closely after its occurrence, due to the delay that exists from the insult until exceeding the cut-off limits. However, determining if irreversible damage has already occurred by then, it is not guaranteed. Prolonged HI results in extreme pH and lactate values and consequently, in a slow recovery to basal state. Therefore, the predictive power consid-

erably increases for differentiation of HI_p phase, especially in the case of lactate, which overall demonstrates to be a better indicator than pH. It is important to point out that this outcome is obtained as a result of a severe HI event extended over almost 50 minutes. In Chapter 3, where the duration of the HI event was much shorter (20 minutes), their performance was considerably reduced. Another big issue of the diagnosis based on a single parameter is the lack of valuable information to differentiate HI and HI_p phases. Distinction between these two stages can be of great help for medical assessment.

Alternatively, machine learning algorithms have been applied to Raman spectra, which considers the entirety of physiological parameters that are accessible by the laser beam. Predictive models have demonstrated great performance for identification of an HI event. Even though their predictive capacity is reduced at HI_p, all models achieve excellent discriminating power during the HI phase, even better to that obtained by lactate or pH from blood gas analysis. No significant differences have been found between different algorithms, but overall, PLS–DA and XGB have shown slightly better performance. Efficiency of PLS has widely been confirmed for functional data (i.e. curve or spectrum type problems), where the dimension of the observation space (broad spectral range) includes multicollinearity of features. The potential of XGB for classification problems is beyond doubt, since it made itself known after winning different classification competitions. Nonetheless, each model makes its own interpretation of the data. The permutation feature importance technique allows us to understand how the model is constructed and, consequently, guarantees a meaningful classification. Raman bands associated to red blood cell oxygenation, such as heme groups or protein denaturation, have been identified as main contributors for classification of HI. In line with Chapter 3, lactate is also recognized as a suitable biomarker, especially at HI_p, when normal oxygenated ventilation is restored.

As stated above, PLS-based hybrid classifiers allow for graphical representation, as Raman data is projected to a new space of latent vectors. As a result, decision regions defined by different classifiers can be compared. Moreover, hybrid classifiers have demonstrated to achieve similar, or in some cases even better, performance than non-hybrid ones. Given the ability of PLS to deal with multicollinearity, applying classifiers after extraction of relevant features can be of great help for success.

Despite the limited number of experiments, these preliminary results already exhibit the great potential of the technology for HI monitoring. This is supported by the

decreasing trend of the learning curve of the PLS–DA model for all proposed classification problems (BC vs. HI, Figure 6.14; BC vs. HI_p, Figure 6.15; BC vs. HI vs. HI_p, Figure 6.16). The learning curve is calculated as the average of individual learning curves, and is represented by the classification error rate as a function of the number of training replicates [104]. In other words, for a given number of training replicates, a piglet is selected for validation while the rest are used to form all possible permutations. Classifiers that have been constructed from each of those training sets are then tested in the validation piglet and the error rate is averaged. This procedure is repeated for each piglet replicate (Figure 6.14a, Figure 6.15a, Figure 6.16a). The mean and the standard error (error of the mean) of individual learning curves are then depicted in Figure 6.14b, Figure 6.15b and Figure 6.16a. It is important to emphasize that this study has been carried out with a commercial Raman probe (the RamanProbeTM, InPhotonics) [96], which despite of an application-specific modification, it has not been designed for this specific task. Development of an optimized probe will certainly result in a considerable improvement of the performance.

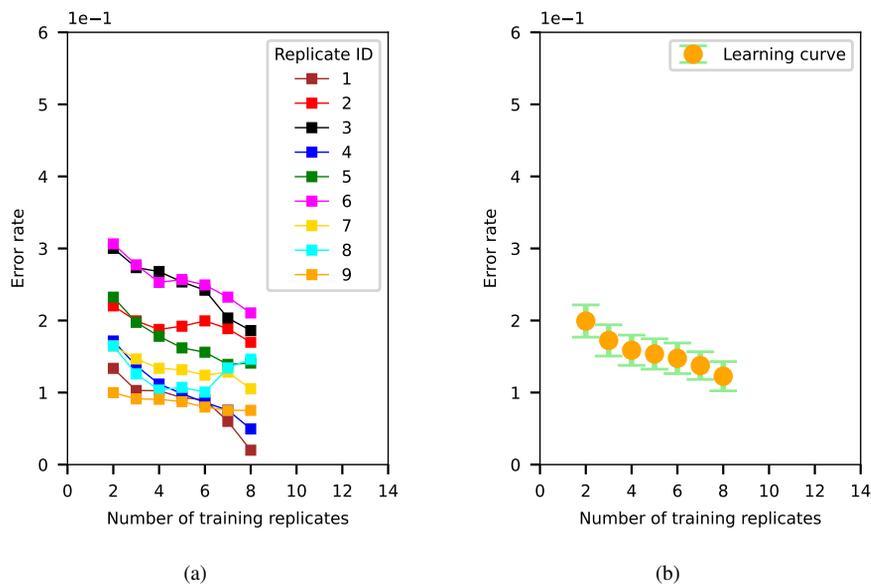


Figure 6.14: Trend of the error rate for BC vs. HI classification as a function of the training sample size. (a) Learning curve for each piglet replicate. (b) Averaged learning curve and its standard error (error of the mean).

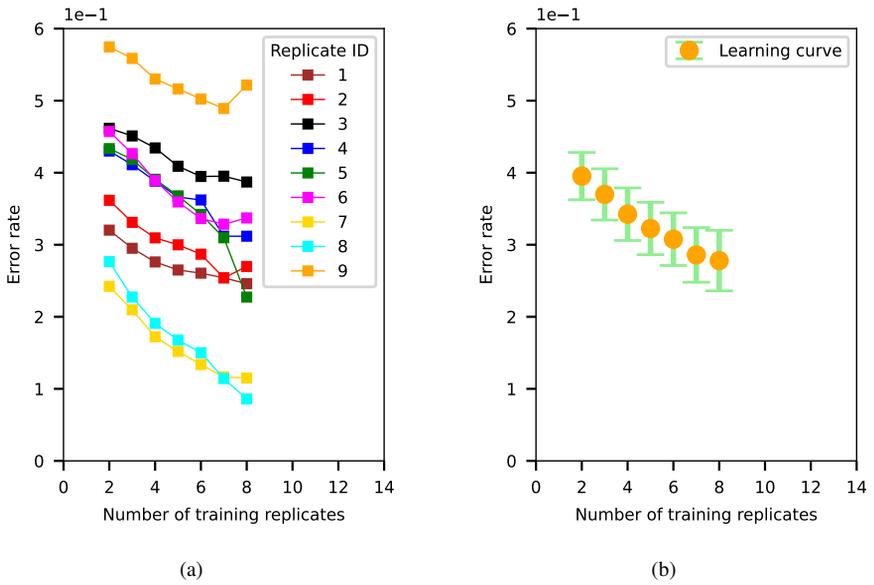


Figure 6.15: Trend of the error rate for BC vs. HI_p classification as a function of the training sample size. **(a)** Learning curve for each piglet replicate. **(b)** Averaged learning curve and its standard error (error of the mean).

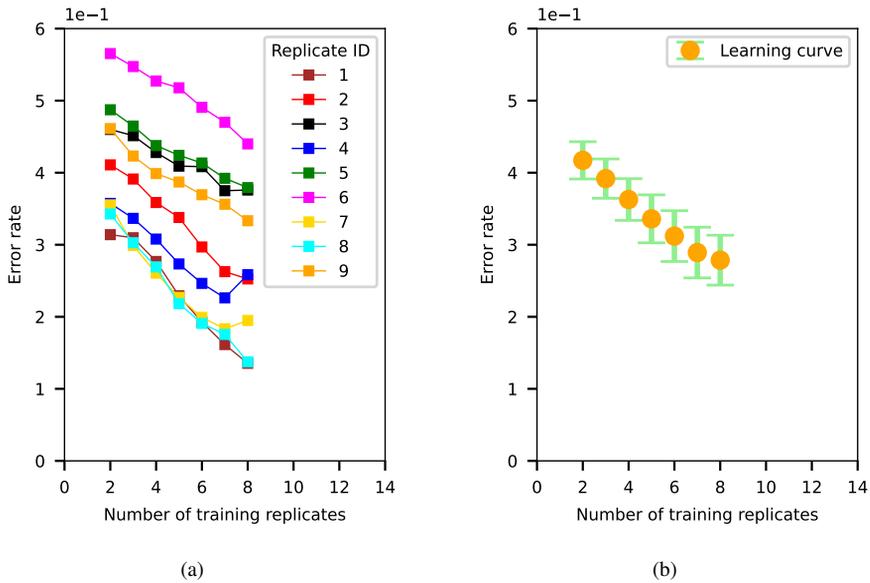


Figure 6.16: Trend of the error rate for BC vs. HI vs. HI_p classification as a function of the training sample size. **(a)** Learning curve for each piglet replicate. **(b)** Averaged learning curve and its standard error (error of the mean).

6.4 Conclusion

For many years obstetric care is relying on pH from fetal scalp blood sampling as a standard biomarker for fetal surveillance. The limited capacity to avoid adverse neonatal outcomes and cesarean deliveries, gave lactate, in vain, the opportunity to be considered as a more suitable alternative. It has been demonstrated that hypoxia-ischemia is responsible for an overall impact in the metabolism, which is reflected in systemic variations of different physiological parameters. The ability to non-invasively provide high molecular selectivity makes Raman spectroscopy an ideal method for clinical assessment. This study presents the combination of Raman spectroscopy with machine learning as a potential diagnostic tool for monitoring of perinatal asphyxia. Apart from exhibiting great performance for identification of a hypoxia-ischemia event, predictive models based on Raman spectra perform relevant distinction between damage (hypoxia-ischemia) and recovery (post-hypoxia-ischemia) phases. In conclusion, the technology represents an innovative approach that will support immediate medical decision-making by non-invasive, continuous and real-time monitoring of perinatal asphyxia.

CHAPTER 7

Technology transfer

This doctoral thesis is embedded in a research project with high innovative power, as it originates from real clinical requirements and includes from the beginning market-driven needs and real-life applications into its research. Starting at lowest technology readiness level (TRL), the technology was elevated to TRL6 within the framework of this project. Apart from its origin, this chapter describes the innovative aspects, highlights the socio-economic importance, and gives insight into market analysis and strategic planning.

7.1 Origin of the project

This project was born as a result of discussions with the Donostia University Hospital, where obstetricians are facing in their daily work during labor the shortcomings of state-of-the-art perinatal monitoring. The novelty a new technology should provide is a non-invasive method that delivers continual data in real-time; none of these three conditions is fulfilled by today's technology. In 2017, a multidisciplinary team from CIC nanoGUNE and Biodonostia Health Research Institute, including physicists, chemists, biologists, engineers, and a physician, established various consecutive collaboration projects to elaborate and develop technical solutions for the joint vision, "*Reduce global neonatal morbidity and mortality rate, as well as complications that may derive from perinatal asphyxia*". To achieve such an ambitious objective, our mission consisted in developing a new disruptive diagnostic method for monitoring perinatal asphyxia.

7.2 Proposed diagnostic procedure

When the rupture of amniotic membranes occurs, a customized Raman probe is introduced in the birth canal and externally placed at the scalp of the baby, as described in Figure 7.1. Unlike current procedure, neither incision nor blood extraction is required for further analysis. Raman spectroscopy provides a clinical picture of the physiological status of the baby through the interaction of the incident photons with the irradiated tissue volume. Raman spectra are then analyzed by a predictive model that has been previously designed and trained with representative known data. Based on characteristic patterns of the data, the model predicts with a certain probability adverse physiological conditions and conveys the information to the obstetrician as a digital display.

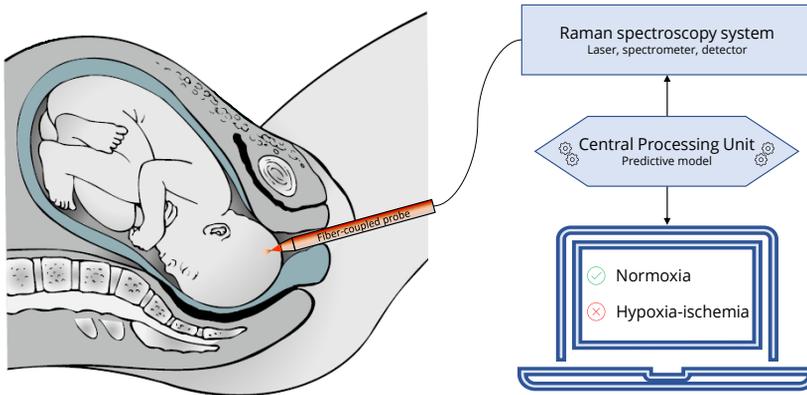


Figure 7.1: Schematic of the complete integrated Raman system, including the hardware (Raman probe, laser, spectrometer, detector, central processing unit, and a display interface) and software part (source code of predictive models).

7.3 Innovation & impact

7.3.1 Major innovation

The proposed diagnostic method stands out for its innovative nature, enabling a continuous, non-invasive and real-time fetal monitoring during labor. The combination of a customized optical probe, a Raman spectroscopy system and the use of machine learning results in a powerful diagnostic tool for identification of hypoxic-ischemic insults. The technology is not constrained to perform the examination of a unique parameter like pH or lactate, as the current standard methods do. Instead, it considers all Raman-active molecules of the entire physiology irradiated by the incident laser, delivering a systemic picture at molecular level. Thus, apart from providing quantitative analysis of pH or lactate if required, our technology offers a more accurate and reliable diagnosis that supports decision-making in obstetric care.

7.3.2 Social impact

Improvement of fetal monitoring during the intrapartum period can considerably reduce the number of neonatal morbidity and mortality rates. Thus, development of a new technology for such a purpose is aligned with the targets and goals for sustainable development proposed by the United Nations (UN) as part of the 2030 agenda [105]. In particular, our technology primarily addresses SDG 3 (sustainable development goal), which pursues a development that *"ensures healthy lives and promotes well-being for all at all ages"*, and its Target 3.2 which states that *"by 2030, end preventable deaths of newborns and children under 5 years of age, with all countries aiming to reduce neonatal mortality to 12 per 1000 live births"*. The compact size of the system provides a portable technology that can be easily used in different scenarios. Moreover, availability for all countries is guaranteed with a fixed and reduced cost given the reusable characteristic of the technology without need of further investment. As stated by Target 3.D and 10.3, it is crucial to *"strengthen the capacity of developing countries for early warning, risk reduction and management of health risks"*, as well as *"ensure equal opportunity and reduce inequalities of outcome"*, to achieve development sustainability.

Beyond the direct effect our technology can have on neonatal morbidity and mortality rates, associated medical, social or financial implications will have considerable significance. Consequences of birth asphyxia are not limited to the fetal outcome at the exact moment of birth but entail further mental and physiological disorders both for the baby and the mother. Required medical examinations, treatments or prolonged hospital stays represent substantial administrative management and additional costs. Cesarean section is a common clinical intervention in high-risk births.

Even though the reasons for performing a cesarean section are diverse, its overuse is also recognized as a consequence of a preventive, but often unnecessary, procedure to avoid intrapartum asphyxia. It has been demonstrated that the use of current standard methods for fetal monitoring has resulted in an excessive use of cesareans and operative deliveries. The World Health Organization (WHO) assumes an occurrence of 15% of high-risk births, where lives are at stake and cesarean section is justified. However, in 2008, global unnecessary interventions represented an estimated cost of approximately US\$ 2.3 billions. A reliable and accurate fetal monitoring provided by our technology could reduce the average cost by 50%.

7.4 Market opportunity

The existence of a real clinical need, which has still not been fulfilled, is the better demonstration of a clear market opportunity for the development of our technology. Ibon Jaunarena, obstetrician at Hospital Universitario Donostia, Édouard Lecarpantier, obstetrician at Intercommunal Hospital of Creteil (Paris), and Xavier Durrmeyer, neonatologist also at Intercommunal Hospital of Creteil, have expressed their enthusiasm and emphasized the clear need of our technology. Other entities, such as the European Space Agency and French Defense Innovation Agency have also shown their interest in the development of the technology. The available products in the market are based on arterial blood gas analysis and therefore require blood sampling. Some of the major players in clinical practice are Abbott, Medtronic, Clinical Innovatios, Philips, Johnson & Johnson, General Electric, Roche, or Nova Biotechnology. Even though the entry barriers for a new player are moderate, the competitive advantages of our technology over the state-of-the-art provides us a great unique selling point.

The market of medical devices/equipment is characterized by its great responsiveness to innovation. Technological advances, prevalence of chronic diseases and increase of aging population is prompting the demand of medical devices for early and fast diagnosis. The global medical devices market is expected to grow from \$505 billion in 2020 to almost \$630 billion in 2026 [106]. European market accounts for around one third of the global market. Considering exclusively the fetal monitoring market, the size is estimated to be \$3.73 billion in 2021 and projected to reach \$5.26 billion by 2026 [107]. Given the great size of the market, 140 million births per year, the intrinsic success of the technology would result in business scalability and product expansion. Moreover, the ability to rapidly adjust the technology both in design and software allows for diversification of the portfolio to other pathologies. Only in France, more than 3000 potential clients have been identified, distributed over 471 obstetric units, 283 neonatal intensive care units, 510 reanimation units, 657 emergency rooms and 1200 pneumonology units.

7.5 Roadmap

At a quite early stage of the project, specialized consultancy agents had been employed to help us develop a business strategy, sort out the market opportunities, and to identify unique selling points. In this context, the company SILO (Science & Innovation Link) was involved and financed by BIC GIPUZKOA (facilitator in the process of creating new, innovative businesses, as well as active agent in the support and promotion of entrepreneurial culture in Gipuzkoa). Financed by nanoGUNE, the private consultancy company Inventik (Santander) elaborated a Technology Dashboard for advancing technology transfer.

A European patent application was filed on the 18th of August 2020 to protect the technology. The application n° EP20382756.3, with the title "Method and device for the determination of hypoxia", is currently under evaluation. In February 2021, the start-up company Optec4Life SAS (Optical Technologies for Life and Health) was founded in Paris, whose business model is based on our technology with exclusive rights for patent exploitation and presented subsequently.

Business Model Canvas

<p>Key partners</p> <ul style="list-style-type: none"> - Collaborators <ul style="list-style-type: none"> o Biodonostia o Intercommunal Hospital of Creteil o Leartiker o Tecnalia o University of Eastern Finland o Finnish research institution VTT o COMATE Engineering & Design - Providers <ul style="list-style-type: none"> o Consumables (ex: Optical components) o Equipment (ex: Spectrometer) o Software licenses 	<p>Key activities</p> <ul style="list-style-type: none"> - Pre-clinical validation (animals) - Clinical validation (humans) - Development of a relevant database - Development of robust and high-performance predictive models - Design & development of an optimized Raman probe - Evaluate the social impact of the technology 	<p>Value propositions</p> <ul style="list-style-type: none"> - Innovative detection method <ul style="list-style-type: none"> o Non-invasive o Continuous o Real-time - More reliable and high-performance diagnosis - Reduction of risk associated to intrapartum period <ul style="list-style-type: none"> o Reduction of neonatal adverse outcomes (morbidity/mortality) o Reduction of cesarean sections - Reduction of associated implications <ul style="list-style-type: none"> o Medical interventions o Financial costs o Administrative management - Reusable method/probe 	<p>Customer relationship</p> <ul style="list-style-type: none"> - Iterative development of the technology with end-users - Initial training course to customers - Personal customer service 	<p>Customer segments</p> <ul style="list-style-type: none"> - Hospitals/Clinics - Public Health Service - Industry of medical devices
<p>Cost structure</p> <ul style="list-style-type: none"> - Consumables - Equipment 	<p>Key resources</p> <ul style="list-style-type: none"> - Qualified team - Experience and knowledge sharing between collaborators - Funding for product development - Intellectual property 	<p>Channels</p> <ul style="list-style-type: none"> - Website (Social networks) - Dissemination <ul style="list-style-type: none"> o Seminars o Conferences o Publications o Fairs 	<p>Revenue streams</p> <ul style="list-style-type: none"> - Complete integrated system <ul style="list-style-type: none"> o Hardware: customized Raman probe, laser, spectrometer, detector, central processing unit o Software: Source code from predictive models - Recurring revenue: software update, customer service 	<p>Revenue streams</p> <ul style="list-style-type: none"> - Patent licensing

According to the European definition for the level of technological maturity, the product development of our technology is at TRL 6. As presented, the technology has been validated in vitro and in a collaboration with Biodonostia, a preclinical study with pigs is in progress. Given its position of strategic partner, Optec4Life will subcontract CIC nanoGUNE for the scientific development of both the spectroscopy system and the specific Raman probe for fetal monitoring. Moreover, contacts have been established with the obstetric and pediatric units of the Intercommunal Hospital of Creteil in Paris. Additional collaborations are planned with Leartiker (Markina-Xemen), Tecnalía (San Sebastián) and COMATE Engineering & Design (Leuven) for engineering guidance in technical aspects, which guarantee the CE marking and the classification as a medical device.

Optec4Life was chosen within the European program ACTPHAST4.0 for indirect financial support towards prototyping. ACTPHAST (ACceleraTing PHotonics innovAtion for SME's: a one STop-shop-incubator) is a unique one-stop-shop rapid prototyping incubator for supporting photonics innovation by European companies, which is financially supported by the European Commission under Horizon2020. The University of Eastern Finland together with the Finnish research institution VTT have been granted to further develop specific hardware components of the spectroscopy system for Optec4Life.

The participation in the following acceleration programs has complemented the scientific activity within the technology transfer towards the market.

Healthtech Translation Advisory Board (TAB) is a European Commission-funded instrument that was created to accelerate the translation of promising healthtech projects towards the market.

BBK Venture Philanthropy is a BBK Foundation program that aims to encourage entrepreneurship and social and environmental impact investment in Bizkaia.

European Technology Platform in Nanomedicine is an initiative led by industry since 2005 and set up together with the European Commission, to address the application of nanotechnology in healthcare.

XX premio Manuel Laborde Werlinden is part of the Entreprenari program which gives support and boosting to entrepreneurship based on innovative ideas.

Final summary

Within the framework of such an ambitious project, this doctoral thesis has demonstrated the capabilities of our technology for monitoring perinatal asphyxia. From very basic principles, the feasibility and viability of the initial proof-of-concept has progressively been examined.

First, the limitations of pH and lactate, being reference parameters for correct medical classification of hypoxic-ischemic events, have been unveiled. Alternatively, it has been shown that predictive models based on the systemic physiological picture, which includes the entirety of biochemical parameters, yields a substantial improvement in the diagnosis over the gold-standard techniques.

In a next step, the technology has been validated *in vitro* and *ex vivo*. The combination of Raman spectroscopy with machine learning algorithms as suitable tool for lactate and pH monitoring in body fluids has been demonstrated. Predictive models have been successfully tested providing clinically promising errors that verify their reliability, their accuracy and hence, their quality. In conclusion, it has been demonstrated that developing a new a diagnostic method to monitor perinatal asphyxia *ex vivo* could be feasible.

Finally, the combination of Raman spectroscopy with machine learning as a potential diagnostic tool for *in vivo* monitoring of perinatal asphyxia has been tested in a pre-clinical phase. Apart from exhibiting great performance for identification of a

hypoxia-ischemia event, predictive models based on Raman spectra perform relevant distinction between damage (hypoxia-ischemia) and recovery (post-hypoxia-ischemia) phases. In conclusion, the technology represents an innovative approach that will support immediate medical decision-making by non-invasive, continuous and real-time monitoring of perinatal asphyxia.

Publications

1. Ion Olaetxea, Ana Valero, Eneko Lopez, Hector Lafuente, Ander Izeta, Ibon Jaunarena, and Andreas Seifert. Machine learning-assisted Raman spectroscopy for pH and lactate sensing in body fluids. *Analytical Chemistry*, 92(20):13888–13895, 2020
2. Hector Lafuente*, Ion Olaetxea*, Ana Valero, Francisco José Alvarez, Ander Izeta, Ibon Jaunarena, and Andreas Seifert. Identification of hypoxia-ischemia by chemometrics considering systemic changes of the physiology. *IEEE Journal of Biomedical and Health Informatics*, 2022, DOI:10.1109/JBHI.2022.3142190
*Authors contributed equally
3. Jaione Etxebarria-Elezgarai, Miriam Mowat, Eneko Lopez, Carlos Rodríguez, Ion Olaetxea, and Andreas Seifert. Gaussian beam shaping and multivariate analysis in plasmonic sensing. *Analytical Chemistry*, 92(24):16236–16244, 2020
4. Ion Olaetxea et al. Photonic technology for in vivo monitoring of hypoxia-ischemia. (to be submitted)

Proceedings

1. I. Olaetxea, E. Lopez, A. Valero, and A. Seifert. Determination of physiological lactate and pH by Raman spectroscopy. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 475–481, 2019
2. Jaione Etxebarria-Elezgarai, Carlos Rodríguez Molinuevo, Marina Echave del Carmen, Eneko Lopez Corriero, Ion Olaetxea Azkarate-Askatsua, Maria Carmen Morant-Miñana, and Andreas Seifert. Improved plasmonic resonance by gaussian beam shaping. In *2019 International Conference on Optical MEMS and Nanophotonics (OMN)*, pages 208–209, 2019

Acknowledgment

After four years of intense effort, it is time to thank all of those, that in way or another, have contributed in this thesis.

First of all, I would like to thank my director Andreas for having trusted in me for this challenging project. I still remember our first video call, with very bad connection, from a wild place somewhere in the other side of the world. At that time, I did not know what the future was holding for my professional career, but for sure I was not expecting that 4-5 years later I would be a PhD, Thank you for that, and for guiding and supporting me during the process.

Thank you also to Joseba Zubia, my tutor/director at the university, Even though he has not actively participated in my thesis, he has always been willing to help.

I would like to give specials thanks to Ana, who was my supervisor during the first two years. Apart from her help, advice and affection, because she became a good friend of mine, which is the most valuable thing at the end.

To Eneko for all this work in the background, which is often unseen but crucial for making things happen. After these 4 years, we have come to the following conclusions: 1. No matter how much we try, experiments turn out well only on Fridays, Do not try to find an explanation, That's how it is. 2. You are the perfect definition of "chaquetero", You change your football shirt more than the skin of a chameleon. 3. Although you are

Basque and turn 31 this year, we all know you have an Andalusian soul and a metabolic age of... Estás casacau tio...

To Jaione for those good moments and discussions about "Konkis" or whatever. No matter what the topic is, always against me, but with a large dose of humor. And this is how good friendships are forged.

To other present and former nanoengineering group members: Mathias, our first crazy wine/cheese/surf-lover PhD student; Laura, always hard-working, thoughtful and kind; Stephen, our dear smiley Irish guy who seems to have found his place in the Basque Country; Harun, the next doctor of the group, full of joy and motivation; Olga, Gajendra, Maica, Marina, José, Mourad... Thank you to all who were part of the group at some point. I wish you the best for the future.

To all nanogune people who are not part of the group but are sometimes even more important in the daily day. I mean all the people with whom I shared coffee and lunch breaks, Nerea, Juanma, Roger, Andoni, Charly, Carlos, Cesar, Iker, Francesco, Gabrielle, Inge, María, Beatriz, Sarai, Leire, David, Mikel, Julene, Leire, Nerea, Monica... and many others.

I would also like to thank our project collaborators in Biodonostia Ibon, Ander and Hector. To Ibon, who has been part of the project from the very beginning, and Ander, who I already knew beforehand, for introducing me to Andreas and for sharing your knowledge with me. And to Hector, for all these hours we shared together in the darkness of the operating room.

And finally and most important, to my family and lifelong friends. For nothing in particular but for everything at the same time. As the saying goes, "you will be judged by the company you keep" ("dime con quien andas y te diré como eres"). I have been blessed to be surrounded by the best.

Eskerrik asko!

Bibliography

- [1] Roberto Antonucci, Annalisa Porcella, and Maria Dolores Pilloni. Perinatal asphyxia in the term newborn. *Journal of Pediatric and Neonatal Individualized Medicine (JPNIM)*, 3(2):e030269, 2014.
- [2] Anne CC Lee, Naoko Kozuki, Hannah Blencowe, Theo Vos, Adil Bahalim, Gary L Darmstadt, Susan Niermeyer, Matthew Ellis, Nicola J Robertson, Simon Cousens, et al. Intrapartum-related neonatal encephalopathy incidence and impairment at regional and global levels for 2010 with trends from 1990. *Pediatric research*, 74(S1):50–72, 2013.
- [3] Lucia Hug, Monica Alexander, Danzhen You, Leontine Alkema, and UN Inter-agency Group for Child. National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis. *The Lancet Global Health*, 7(6):e710–e720, 2019.
- [4] Jennifer J Kurinczuk, Melanie White-Koning, and Nadia Badawi. Epidemiology of neonatal encephalopathy and hypoxic–ischaemic encephalopathy. *Early human development*, 86(6):329–338, 2010.
- [5] Juan Arnaez, Alfredo García-Alix, Gemma Arca, Sonia Caserío, Eva Valverde, M Teresa Moral, Isabel Benavente-Fernández, and Simón Lubián-López. Population-based study of the national implementation of therapeutic hypother-

mia in infants with hypoxic-ischemic encephalopathy. *Therapeutic hypothermia and temperature management*, 8(1):24–29, 2018.

- [6] Ellen Blix, Robyn Maude, Elisabeth Hals, Sezer Kisa, Elisabeth Karlsen, Ellen Aagaard Nohr, Ank de Jonge, Helena Lindgren, Soo Downe, Liv Merete Reinart, et al. Intermittent auscultation fetal monitoring during labour: A systematic scoping review to identify methods, effects, and accuracy. *PloS one*, 14(7):e0219573, 2019.
- [7] Ana Pinas and Edwin Chandrachan. Continuous cardiotocography during labour: Analysis, classification and management. *Best practice & research Clinical obstetrics & gynaecology*, 30:33–47, 2016.
- [8] Deitra Leonard Lowdermilk, Mary Catherine Cashion, Shannon E Perry, Kathy Rhodes Alden, and Ellen Olshansky. *Maternity and Women's Health Care E-Book*. Elsevier Health Sciences, 2019.
- [9] Zarko Alfirevic, Declan Devane, and Gillian ML Gyte. Continuous cardiotocography (ctg) as a form of electronic fetal monitoring (efm) for fetal assessment during labour. *Cochrane database of systematic reviews*, (5), 2013.
- [10] Christine E East, Leo R Leader, Penelope Sheehan, Naomi E Henshall, Paul B Colditz, and Rosalind Lau. Intrapartum fetal scalp lactate sampling for fetal assessment in the presence of a non-reassuring fetal heart rate trace. *Cochrane Database of Systematic Reviews*, (5), 2015.
- [11] Christina Rørbye, Anette Perslev, and Carsten Nickelsen. Lactate versus pH levels in fetal scalp blood during labor—using the lactate scout system. *The Journal of Maternal-Fetal & Neonatal Medicine*, 29(8):1200–1204, 2016.
- [12] Ana Pilar Betrán, Jianfeng Ye, Anne-Beth Moller, Jun Zhang, A Metin Gülmezoglu, and Maria Regina Torloni. The increasing trend in caesarean section rates: global, regional and national estimates: 1990-2014. *PloS one*, 11(2):e0148343, 2016.
- [13] Ugo Indraccolo, Margherita Pace, Giovanna Corona, Marco Bonito, Salvatore Renato Indraccolo, and Romolo Di Iorio. Cesarean section in the absence of labor and risk of respiratory complications in newborns: a case-control study. *The Journal of Maternal-Fetal & Neonatal Medicine*, 32(7):1160–1166, 2019.

-
- [14] Yan Li, Caixia Zhang, and Dongfeng Zhang. Cesarean section and the risk of neonatal respiratory distress syndrome: a meta-analysis. *Archives of gynecology and obstetrics*, pages 1–15, 2019.
- [15] Samantha C Lean, Hayley Derricott, Rebecca L Jones, and Alexander EP Heazell. Advanced maternal age and adverse pregnancy outcomes: A systematic review and meta-analysis. *PloS one*, 12(10):e0186287, 2017.
- [16] Eva Rydahl, Eugene Declercq, Mette Juhl, and Rikke Damkjær Maimburg. Cesarean section on a rise - does advanced maternal age explain the increase? a population register-based study. *PloS one*, 14(1):e0210655, 2019.
- [17] Bruno Carbonne, Kelly Pons, and Emeline Maisonneuve. Foetal scalp blood sampling during labour for pH and lactate measurements. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 30:62–67, 2016.
- [18] Ingrid Stål, Ulla-Britt Wennerholm, Lennart Nordstrom, Lars Ladfors, and Eva Wiberg-Itzel. Fetal scalp blood sampling during second stage of labor—analyzing lactate or ph? a secondary analysis of a randomized controlled trial. *The Journal of Maternal-Fetal & Neonatal Medicine*, pages 1–8, 2020.
- [19] Ion Olaetxea, Ana Valero, Eneko Lopez, Hector Lafuente, Ander Izeta, Ibon Jaunarena, and Andreas Seifert. Machine learning-assisted Raman spectroscopy for pH and lactate sensing in body fluids. *Analytical Chemistry*, 92(20):13888–13895, 2020.
- [20] Pradeep Chowriappa, Sumeet Dua, and Yavor Todorov. Introduction to machine learning in healthcare informatics. In *Machine Learning in Healthcare Informatics*, pages 1–23. Springer, 2014.
- [21] Jenna Wiens and Erica S Shenoy. Machine learning for healthcare: on the verge of a major shift in healthcare epidemiology. *Clinical Infectious Diseases*, 66(1):149–153, 2018.
- [22] Patricia C Toomey and Lawrence Oppenheimer. Prediction of hypoxic acidemia in last 2 hours of labour in low-risk women. *Journal of Obstetrics and Gynaecology Canada*, 41(11):1564–1570, 2019.
-

-
- [23] Zafer Cömert, Abdulkadir Şengür, Ümit Budak, and Adnan Fatih Kocamaz. Prediction of intrapartum fetal hypoxia considering feature selection algorithms and machine learning models. *Health information science and systems*, 7(1):17, 2019.
- [24] Zhidong Zhao, Yang Zhang, Zafer Comert, and Yanjun Deng. Computer-aided diagnosis system of fetal hypoxia incorporating recurrence plot with convolutional neural network. *Frontiers in physiology*, 10:255, 2019.
- [25] Hamid Abbasi, Paul P Drury, Christopher A Lear, Alistair J Gunn, Joanne O Davidson, Laura Bennet, and Charles P Unsworth. Eeg sharp waves are a biomarker of striatal neuronal survival after hypoxia-ischemia in preterm fetal sheep. *Scientific reports*, 8(1):1–12, 2018.
- [26] Takuji Ube, Yasuko Yoneyama, and Takashi Ishiguro. In situ measurement of the pH-dependent transmission infrared spectra of aqueous lactic acid solutions. *Analytical Sciences*, 33(12):1395–1400, 2017.
- [27] Laura Marcu, Stephen A Boppart, Mark R Hutchinson, Jürgen Popp, and Brian C Wilson. Biophotonics: the big picture. *Journal of biomedical optics*, 23(2):021103, 2017.
- [28] Waseem K Jerjes, Tahwinder Upile, Brian J Wong, Christian S Betz, Henricus J Sterenborg, Max J Witjes, Kristian Berg, Robert Van Veen, Merrill A Biel, Adel K El-Naggar, et al. The future of medical diagnostics. *Head & Neck Oncology*, 3(1):38, 2011.
- [29] Alex Mason, Olga Korostynska, Julien Louis, Luis Eduardo Cordova-Lopez, Badr Abdullah, Jacob Greene, Rob Connell, and John Hopkins. Noninvasive in-situ measurement of blood lactate using microwave sensors. *IEEE Transactions on Biomedical Engineering*, 65(3):698–705, 2018.
- [30] GEC Ellerby, CP Smith, F Zou, P Scott, and BR Soller. Validation of a spectroscopic sensor for the continuous, noninvasive measurement of muscle oxygen saturation and ph. *Physiological measurement*, 34(8):859, 2013.
- [31] Holly J Butler, Lorna Ashton, Benjamin Bird, Gianfelice Cinque, Kelly Curtis, Jennifer Dorney, Karen Esmonde-White, Nigel J Fullwood, Benjamin Gard-

-
- ner, Pierre L Martin-Hirsch, et al. Using Raman spectroscopy to characterize biological materials. *Nature protocols*, 11(4):664, 2016.
- [32] Praveen C Ashok, Mario E Giardini, Kishan Dholakia, and Wilson Sibbett. A Raman spectroscopy bio-sensor for tissue discrimination in surgical robotics. *Journal of biophotonics*, 7(1-2):103–109, 2014.
- [33] Saranjam Khan, Rahat Ullah, Shaheen Shahzad, Nasira Anbreen, Muhammad Bilal, and Asifullah Khan. Analysis of tuberculosis disease through Raman spectroscopy and machine learning. *Photodiagnosis and Photodynamic Therapy*, 24:286–291, 2018.
- [34] Rishikesh Pandey, Santosh Kumar Paidi, Tulio A Valdez, Chi Zhang, Nicolas Spegazzini, Ramachandra Rao Dasari, and Ishan Barman. Noninvasive monitoring of blood glucose with Raman spectroscopy. *Accounts of chemical research*, 50(2):264–272, 2017.
- [35] Félix Lussier, Vincent Thibault, Benjamin Charron, Gregory Q Wallace, and Jean-Francois Masson. Deep learning and artificial intelligence methods for Raman and surface-enhanced Raman scattering. *TrAC Trends in Analytical Chemistry*, 124:115796, 2020.
- [36] Marius Nache, Rico Scheier, Heinar Schmidt, and Bernd Hitzmann. Non-invasive lactate-and pH-monitoring in porcine meat using Raman spectroscopy and chemometrics. *Chemometrics and Intelligent Laboratory Systems*, 142:197–205, 2015.
- [37] Min Ren and Mark A Arnold. Comparison of multivariate calibration models for glucose, urea, and lactate from near-infrared and Raman spectra. *Analytical and bioanalytical chemistry*, 387(3):879–888, 2007.
- [38] Nilam C Shah, Olga Lyandres, Joseph T Walsh, Matthew R Glucksberg, and Richard P Van Duyne. Lactate and sequential lactate- glucose sensing using surface-enhanced Raman spectroscopy. *Analytical chemistry*, 79(18):6927–6932, 2007.
- [39] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander D Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential cell biology*. Garland Science, 2015.
-

-
- [40] Uwe Schlattner, Malgorzata Tokarska-Schlattner, and Theo Wallimann. Mitochondrial creatine kinase in human health and disease. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease*, 1762(2):164–180, 2006.
- [41] Wafae El Otmani and Susie Davies. Starvation, exercise and the stress response. *Anaesthesia & Intensive Care Medicine*, 2020.
- [42] Deborah Anne Burton, Keith Stokes, and George M Hall. Physiological effects of exercise. *Continuing Education in Anaesthesia Critical Care & Pain*, 4(6):185–188, 2004.
- [43] Bandarn Suetrong and Keith R Walley. Lactic acidosis in sepsis: it’s not all anaerobic: implications for diagnosis and management. *Chest*, 149(1):252–261, 2016.
- [44] Avital Schurr. Lactate, not pyruvate, is the end product of glucose metabolism via glycolysis. *Carbohydrate*, 2017.
- [45] Ana Arias-Oliveras. Neonatal blood gas interpretation. *Newborn and Infant Nursing Reviews*, 16(3):119–121, 2016.
- [46] L Lee Hamm, Nazih Nakhoul, and Kathleen S Hering-Smith. Acid-base homeostasis. *Clinical Journal of the American Society of Nephrology*, 10(12):2232–2242, 2015.
- [47] Lindsay M Biga, Sierra Dawson, Amy Harwell, Robin Hopkins, Joel Kaufmann, Mike LeMaster, Philip Matern, Katie Morrison-Graham, Devon Quick, and Jon Runyeon. *Anatomy & physiology*. OpenStax & Oregon State University, 2020.
- [48] Chester B Martin Jr. Normal fetal physiology and behavior, and adaptive responses with hypoxemia. In *Seminars in perinatology*, volume 32, pages 239–242. Elsevier, 2008.
- [49] Dominique Singer. The human fetus and metabolic adaptations to hypoxia. *Hypoxic Respiratory Failure in the Newborn: From Origins to Clinical Management*; Dakshinamurti, S., Ed, pages 6–11, 2021.
- [50] Jeffrey M Perlman. Summary proceedings from the neurology group on hypoxic-ischemic encephalopathy. *Pediatrics*, 117(Supplement 1):S28–S33, 2006.

-
- [51] P Greco, G Nencini, I Piva, M Scioscia, CA Volta, S Spadaro, M Neri, G Bonaccorsi, F Greco, I Cocco, et al. Pathophysiology of hypoxic–ischemic encephalopathy: a review of the past and a view on the future. *Acta Neurologica Belgica*, 120(2):277–288, 2020.
- [52] Jayasree Nair and Vasantha HS Kumar. Current and emerging therapies in the management of hypoxic ischemic encephalopathy in neonates. *Children*, 5(7):99, 2018.
- [53] Peter Larkin. *Infrared and Raman spectroscopy: principles and spectral interpretation*. Elsevier, 2017.
- [54] Abdullah Chandra Sekhar Talari, Zanyar Movasaghi, Shazza Rehman, and Ihtesham Ur Rehman. Raman spectroscopy of biological tissues. *Applied spectroscopy reviews*, 50(1):46–111, 2015.
- [55] Sara Mosca, Claudia Conti, Nick Stone, and Pavel Matousek. Spatially offset Raman spectroscopy. *Nature Reviews Methods Primers*, 1(1):1–16, 2021.
- [56] David Tuschel. Selecting an excitation wavelength for Raman spectroscopy. *Spectroscopy*, 31(3):14–23, 2016.
- [57] Rekha Gautam, Sandeep Vanga, Freek Ariese, and Siva Umopathy. Review of multidimensional data processing approaches for Raman and infrared spectroscopy. *EPJ Techniques and Instrumentation*, 2(1):1–38, 2015.
- [58] Sinead J Barton and Bryan M Hennelly. Signal to noise ratio of Raman spectra of biological samples. In *Biophotonics: Photonic Solutions for Better Health Care VI*, volume 10685, page 106854F. International Society for Optics and Photonics, 2018.
- [59] Jianwen Luo, Kui Ying, and Jing Bai. Savitzky–golay smoothing and differentiation filter for even number data. *Signal processing*, 85(7):1429–1434, 2005.
- [60] Paul HC Eilers and Hans FM Boelens. Baseline correction with asymmetric least squares smoothing. *Leiden University Medical Centre Report*, 1(1):5, 2005.
- [61] Tom Fearn. Extended multiplicative scatter correction. *NIR news*, 16(4):3–5, 2005.
-

-
- [62] Barry M Wise, NB Gallagher, R Bro, J Shaver, W Windig, and R Scott Koch. PLS toolbox 4.0. *Eigenvector Research Incorporated: Wenatchee, WA, USA*, 2007.
- [63] Richard G Brereton. *Chemometrics for pattern recognition*. John Wiley & Sons, 2009.
- [64] Gangadhar Shobha and Shanta Rangaswamy. Machine learning. In *Handbook of statistics*, volume 38, pages 197–228. Elsevier, 2018.
- [65] Issam El Naqa and Martin J Murphy. What is machine learning? In *machine learning in radiation oncology*, pages 3–11. Springer, 2015.
- [66] Payam Refaeilzadeh, Lei Tang, and Huan Liu. Cross-validation. *Encyclopedia of database systems*, 5:532–538, 2009.
- [67] Alaa Tharwat. Classification assessment methods. *Applied Computing and Informatics*, 2020.
- [68] Mauno Vihinen. How to evaluate performance of prediction methods? measures and their interpretation in variation effect analysis. In *BMC genomics*, volume 13, pages 1–10. BioMed Central, 2012.
- [69] Randall D Tobias et al. An introduction to partial least squares regression. In *Proceedings of the twentieth annual SAS users group international conference*, pages 1250–1257. SAS Institute Inc Cary, 1995.
- [70] Hervé Abdi. Partial least square regression (pls regression). *Encyclopedia for research methods for the social sciences*, 6(4):792–795, 2003.
- [71] Davide Ballabio and Roberto Todeschini. *Multivariate Classification for Qualitative Analysis*. Elsevier, 2009.
- [72] William S Noble. What is a support vector machine? *Nature biotechnology*, 24(12):1565–1567, 2006.
- [73] Gérard Biau. Analysis of a random forests model. *The Journal of Machine Learning Research*, 13(1):1063–1095, 2012.
- [74] Christopher M Bishop. *Combining Models*. springer, 2006.

-
- [75] Jerome H Friedman. Stochastic gradient boosting. *Computational statistics & data analysis*, 38(4):367–378, 2002.
- [76] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.
- [77] Hector Lafuente Echevarria. *Efecto neuroprotector del cannabidiol y la hipotermia en la encefalopatía hipóxico-isquémica neonatal*. PhD thesis, Universidad del País Vasco-Euskal Herriko Unibertsitatea, 2014.
- [78] Hector Lafuente, Ion Olaetxea, Ana Valero, Francisco Jose Alvarez, Ander Izeta, Ibon Jaunarena, and Andreas Seifert. Identification of hypoxia-ischemia by chemometrics considering systemic changes of the physiology. *IEEE Journal of Biomedical and Health Informatics*, 2022.
- [79] E Saling. Fetal scalp blood analysis. *Journal of perinatal medicine*, 9(4):165, 1981.
- [80] Kerstin Kruger, B Hallberg, M Blennow, M Kublickas, and M Westgren. Predictive value of fetal scalp lactate concentration and pH as markers of neurologic disability. *American journal of obstetrics and gynecology*, 181(5):1072–1078, 1999.
- [81] Micheal O Arowolo, Sulaiman O Abdulsalam, Rafiu M Isiaka, and Kazeem A Gbolagade. A hybrid dimensionality reduction model for classification of microarray dataset. *International Journal of Information Technology and Computer Science (IJITCS)*, 9(11):57–63, 2017.
- [82] Bijan Yeganeh, M Shafie Pour Motlagh, Yousef Rashidi, and Hamidreza Kamalan. Prediction of co concentrations based on a hybrid partial least square and support vector machine model. *Atmospheric Environment*, 55:357–365, 2012.
- [83] Lars W Andersen, Julie Mackenhauer, Jonathan C Roberts, Katherine M Berg, Michael N Cocchi, and Michael W Donnino. Etiology and therapeutic approach to elevated lactate levels. In *Mayo Clinic Proceedings*, volume 88, pages 1127–1140. Elsevier, 2013.

-
- [84] Raymond C Koehler, Zeng-Jin Yang, Jennifer K Lee, and Lee J Martin. Perinatal hypoxic-ischemic brain injury in large animal models: Relevance to human neonatal encephalopathy. *Journal of Cerebral Blood Flow & Metabolism*, 38(12):2092–2111, 2018.
- [85] I. Olaetxea, E. Lopez, A. Valero, and A. Seifert. Determination of physiological lactate and pH by Raman spectroscopy. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 475–481, 2019.
- [86] Kristian Hovde Liland, Achim Kohler, and Nils Kristian Afseth. Model-based pre-processing in Raman spectroscopy of biological samples. *Journal of Raman Spectroscopy*, 47(6):643–650, 2016.
- [87] Marc D Fontana, K Ben Mabrouk, and Thomas H Kauffmann. Raman spectroscopic sensors for inorganic salts. *Spectrosc Prop Inorg Organomet Compd*, 44:40–67, 2013.
- [88] Chad G Atkins, Kevin Buckley, Deborah Chen, H Georg Schulze, Dana V Devine, Michael W Blades, and Robin FB Turner. Raman spectroscopy as a novel tool for monitoring biochemical changes and inter-donor variability in stored red blood cell units. *Analyst*, 141(11):3319–3327, 2016.
- [89] Immanuel Valpapuram, Patrizio Candeloro, Maria Laura Coluccio, Elvira Immacolata Parrotta, Andrea Giugni, Gobind Das, Gianni Cuda, Enzo Di Fabrizio, and Gerardo Perozziello. Waveguiding and SERS simplified Raman spectroscopy on biological samples. *Biosensors*, 9(1):37, 2019.
- [90] G Cassanas, M Morssli, E Fabregue, and L Bardet. Vibrational spectra of lactic acid and lactates. *Journal of Raman spectroscopy*, 22(7):409–413, 1991.
- [91] MP Gómez-Carracedo, JM Andrade, DN Rutledge, and NM Faber. Selecting the optimum number of partial least squares components for the calibration of attenuated total reflectance-mid-infrared spectra of undesigned kerosene samples. *Analytica chimica acta*, 585(2):253–265, 2007.
- [92] Kyle C Doty and Igor K Lednev. Differentiation of human blood from animal blood using Raman spectroscopy: A survey of forensically relevant species. *Forensic science international*, 282:204–210, 2018.

-
- [93] Chad G Atkins, Kevin Buckley, Michael W Blades, and Robin FB Turner. Raman spectroscopy of blood and blood components. *Applied spectroscopy*, 71(5):767–793, 2017.
- [94] P Lemler, WR Premasiri, A DelMonaco, and LD Ziegler. NIR raman spectra of whole human blood: effects of laser-induced and in vitro hemoglobin denaturation. *Analytical and bioanalytical chemistry*, 406(1):193–200, 2014.
- [95] EmVision LLC. Raman lensed probe. <https://emvisionllc.com/probes/>.
- [96] InPhotonics. The RamanProbe™. <https://www.inphotonics.com/probes.htm>.
- [97] Michelle Agenant, Matthijs Grimbergen, Ronald Draga, Eric Marple, Ruud Bosch, and Christiaan van Swol. Clinical superficial Raman probe aimed for epithelial tumor detection: phantom model results. *Biomedical optics express*, 5(4):1203–1216, 2014.
- [98] Conor C Horgan, Mads S Bergholt, May Zaw Thin, Anika Nagelkerke, Robert Kennedy, Tammy L Kalber, Daniel J Stuckey, and Molly M Stevens. Image-guided Raman spectroscopy probe-tracking for tumor margin delineation. *Journal of biomedical optics*, 26(3):036002, 2021.
- [99] Michael Pinto, Kevin C Zorn, Jean-Philippe Tremblay, Joannie Desroches, Frédérick Dallaire, Kelly Aubertin, Eric T Marple, Chris Kent, Frederic Leblond, Dominique Trudel, et al. Integration of a Raman spectroscopy system to a robotic-assisted surgical system for real-time tissue characterization during radical prostatectomy procedures. *Journal of biomedical optics*, 24(2):025001, 2019.
- [100] Gail M Sullivan and Richard Feinn. Using effect size—or why the p value is not enough. *Journal of graduate medical education*, 4(3):279–282, 2012.
- [101] Hans W Wendt. Dealing with a common problem in social science: A simplified rank-biserial coefficient of correlation based on the statistic. *European Journal of Social Psychology*, 1972.
- [102] Bayden R Wood, Peter Caspers, Gerwin J Puppels, Shveta Pandiancherri, and Don McNaughton. Resonance Raman spectroscopy of red blood cells using near-infrared laser excitation. *Analytical and bioanalytical chemistry*, 387(5):1691–1703, 2007.
-

-
- [103] Michele Casella, Andrea Lucotti, Matteo Tommasini, Marzia Bedoni, Elena Forvi, Furio Gramatica, and Giuseppe Zerbi. Raman and SERS recognition of β -carotene and haemoglobin fingerprints in human whole blood. *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 79(5):915–919, 2011.
- [104] Nairveen Ali, Sophie Girnus, Petra Rösch, Jürgen Popp, and Thomas Bocklitz. Sample-size planning for multivariate data: a Raman-spectroscopy-based example. *Analytical chemistry*, 90(21):12485–12492, 2018.
- [105] United Nations. *The Sustainable Development Goals Report 2020*, 2020.
- [106] Mordor Intelligence. Medical devices market - growth, trends, covid-19 impact, and forecasts (2021-2026). <https://www.mordorintelligence.com/industry-reports/global-medical-device-technologies-market-industry>. Accessed: 2022-01-10.
- [107] Markets and Markets. Fetal monitoring market - global forecast to 2026. <https://www.marketsandmarkets.com/Market-Reports/fetal-monitoring-market-35700261.html>, March 2021. Accessed: 2022-01-10.
- [108] Jaione Etxebarria-Elezgarai, Miriam Mowat, Eneko Lopez, Carlos Rodríguez, Ion Olaetxea, and Andreas Seifert. Gaussian beam shaping and multivariate analysis in plasmonic sensing. *Analytical Chemistry*, 92(24):16236–16244, 2020.
- [109] Jaione Etxebarria-Elezgarai, Carlos Rodríguez Molinuevo, Marina Echave del Carmen, Eneko Lopez Corriero, Ion Olaetxea Azkarate-Askatsua, Maria Carmen Morant-Miñana, and Andreas Seifert. Improved plasmonic resonance by gaussian beam shaping. In *2019 International Conference on Optical MEMS and Nanophotonics (OMN)*, pages 208–209, 2019.