Diffuse Reflectance Spectroscopy

Using Multivariate analysis method for determination of tissue optical properties

Hasti Yavari

Johan Axelsson   Stefan Andersson-Engels

Department of Physics

Lund University

This dissertation is submitted for the degree of

Master of Science

Acknowledgements

Foremost, I would like to express my deepest gratitude to my main supervisor, Johan Axelsson, for his unwavering support and patience throughout the duration of this thesis. I am very thankful for his generous share of knowledge and helpful insights in our regular meetings. I would like to genuinely thank my co-supervisor, Stefan Andersson-Engels, for giving me the opportunity to take this project and trusting me with additional academical tasks within the group. Special thanks to Lisa Kobayashi Frisk for accompanying me through the early stages of this work. Lastly, I would like to thank my parents, sister and Erik Wik for their unconditional moral support.

Abstract

Diffuse reflectance Spectroscopy is a non-invasive and real-time technique used both in research and clinical studies for purposes such as identifying tumors and monitoring their response to therapy. Here, a compact, cost-effective and portable experimental setup is used in order to acquire the diffuse reflectance spectra from tissue-like liquid phantoms. Two fiber optic probes with different source-detector separations are used for collecting the diffuse light. A phantom preparation protocol is proposed in order to construct a dataset of diffuse reflectance spectra from phantoms with different tissue chromophores compositions. Nonlinear least-squares support vector machines (LS-SVM) regression technique within the Multivariate analysis (MVA) framework is employed in order to extract the optical properties of the tissue-like phantoms. Validation measurements of the liquid phantoms demonstrate a higher prediction accuracy for larger number of training samples. Percentage error of $<2\%$ is observed when testing 1 sample in both reduced scattering coefficient and blood volume fraction models. The reduced scattering coefficient can be estimated with higher accuracy in the models constructed with the data collected using both probes. Ways to improve the regression models' performance are proposed along with suggestions for future work.

Table of contents

Symbols and Abbreviations

ε	Margin of tolerance
ε_i	Extinction coefficient of chromophore i
γ	Regularization or scaling parameter
λ_0	Normalization wavelength
μ_{eff}	Effective attenuation coefficient
μ_a	Absorption coefficient
μ_s	Scattering coefficient
μ_s'	Reduced scattering coefficient
ν	Volume fraction
ω	Regression coefficient
Φ	Fluence rate
ρ	Photon density
σ	Radial basis function kernel width
C_i	Molar concentration of chromophore i
HbO_2	Oxy-hemoglobin
SO_2	Oxygen saturation
ξ_i	Slack variable
b_{Mie}	Mie scattering power

D	Diffusion coefficient
f_{Ray}	Rayleigh scattering fraction
CF	Calibration factor
CV	Cross-validation
DR	Diffuse reflectance
DRS	Diffuse reflectance spectroscopy
FEM	Finite element method
g	anisotropy factor
Hb	Hemoglobin
IL	Intralipid
J	Photon flux vector
L	Radiance
LIF	Laser-induced fluorescence spectroscopy
LOOCV	Leave-one-out cross-validation
LS-SVM	Least-squares support vector machines
LUT	Lookup table
MC	Monte Carlo
MVA	Multivariate analysis
PCA	Principal component analysis
PDT	Photodynamic therapy
PLS	Partial least-squares regression
PpIX	protoporphyrin IX
RTE	Radiative transport equation
SDS	Source-detector separation

SNR   Signal to noise ratio

SSE   Sum-squared error

SVM   Support vector machines

Chapter 1

Introduction

1.1 Diffuse reflectance spectroscopy as an optical diagnostic tool

Diffuse reflectance spectroscopy (DRS), also known as elastic scattering spectroscopy, is a noninvasive spectroscopic technique used in quantitative optical characterization of tissue. The basic principle behind DRS is depicted in Figure 1.1. Here a broadband light source with wavelength range from UV to NIR is irradiating the sample, followed by recording the reflected light after propagating through the sample. Using model-based techniques or statistical approaches such as multivariate analysis, optical properties of the sample can be extracted from light distribution in tissue that further leads to valuable diagnostics information such as the health state of tissue. The two most important of these optical properties include scattering (depending on size, density and refractive index variation within a tissue type) and absorption (depending on tissue chromophore composition). Scattering, absorption and other optical properties of the tissue are referred to in more details in section 2.1.
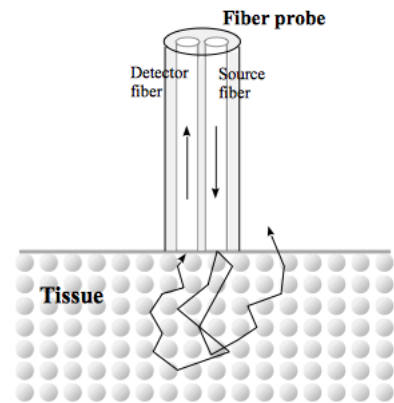

Fig. 1.1: Basic concept behind diffuse reflectance spectroscopy.

The DRS technique has been extensively used both in research and clinical studies for purposes such as identifying tumors and monitoring their response to photodynamic therapy (PDT) or radiotherapy for different cancer types such as breast [1–4], colon [5], prostate [6], cervical [7] and lung [8, 9].

1.2   Purpose and outline

Having precise knowledge of the optical properties of biological tissues leads to valuable diagnostic information about the health state of the particular sample. For instance, the amount of oxygen in cancerous cells is noticeably less than its healthy counterparts. In this way, the response of a tumor to therapy can be evaluated that could form important information for better prognosis of the treatment. The main goal of this thesis is investigating methods that can extract the optical properties of tissues with an improved accuracy. Here a DRS spectrum is obtained from liquid phantoms consisting of major chromophores that mimic real biological tissues. The main aims of this work are:

- To present a literature study of the current state-of-art in diffuse reflectance spectroscopy for estimation of tissue chromophores based on regression analysis.

- To establish phantom preparation protocols for intralipid blood phantoms.

- To develop evaluation protocols for the estimation of tissue chromophores and scattering parameters using LS-SVM technique.

To this end, an overview of the tissue diagnostics field as well as the methodology and obtained results are presented in this thesis in the following order:

Chapter 2 provides an overview of the theoretical aspects of this work required for better understanding of the proceeding chapters. First, the fundamental light and tissue interactions are explained followed by presenting the main methods used for simulating the light propagation. The following discussion then introduces ways to extract the optical properties of a medium by analysing the recorded signal. Lastly, data extraction within the multivariate framework using both linear and nonlinear models and the motivation for the approach chosen in this work are given.

Chapter 3 includes the experimental setup, the calibration technique and the phantom preparation protocol used in this work. Next, implementation of the multivariate analysis protocol based on LS-SVM is presented.

Chapter 4 provides the results obtained in evaluation of the simulated data as well experimental measurements using LS-SVM regression analysis. Next, the errors between the extracted optical properties of the prepared phantom and theoretically expected values are elucidated for each particular model and evaluation technique.

Chapter 5 includes the discussion about the obtained results and suggests ways to improve the regression models performances.

Chapter 6 concludes this thesis by stating the most important aspects and findings with respect to the original purpose of the work. In addition, suggestions for future studies are given.

Chapter 2

Theoretical Background

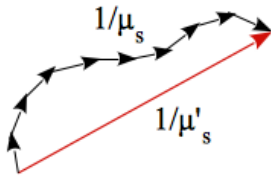## 2.1 Tissue optical properties

2.1.1 Scattering

Fig. 2.1: Graphical representation of the relation between μ_s' and μ_s.

Refractive index variation for different chromophores distributions results in scattering of the propagating light in the tissue [10]. Scattering can be described using two parameters; namely, scattering coefficient (μ_s $[cm^{-1}]$) and anisotropy factor (g). The scattering coefficient is defined as the probability per unit length for a scattering photon and the anisotropy factor is defined as the average scattering direction. In turbid media, i.e. when the light propagation can be treated as diffuse, these two parameters can be combined to form a new parameter called the reduced scattering coefficient (μ_s' $[cm^{-1}]$) as:

$$\mu_s' = (1-g)\mu_s. \tag{2.1}$$

Under the assumption that the tissue is composed of spherical particles of various sizes, the reduced scattering coefficient can alternatively be defined based on Mie and Rayleigh scatterings [11] as expressed in Equation 2.2:

$$
\begin{aligned}
\mu_s'(\lambda) &= \mu_s'(\text{Mie}) + \mu_s'(\text{Rayleigh}) \\
&= a\left((1-f_{\text{Ray}})\left(\frac{\lambda}{\lambda_0}\right)^{-b_{\text{Mie}}} + f_{\text{Ray}}\left(\frac{\lambda}{\lambda_0}\right)^{-4} \right),
\end{aligned}
\tag{2.2}
$$

where a is the scaling factor, b_{Mie} is the Mie scattering power, λ_0 is the normalization wavelength and f_{Ray} is the fraction of Rayleigh scattering. b_{Mie} is related to the particle size of the tissue, with approximate mean values of 0.45 in fatty tissues and 1.09 in brain tissues [12]. The wavelength of the propagating light and the assumed spheres sizes are normally comparable and therefore making Mie the dominant form of scattering; however, in the visible wavelength range Rayleigh scattering has a more considerable contribution.

In a liquid phantom consisting of diluted emulsion of intralipid as the scatterer, the reduced scattering parameter has been estimated by Staveren et al. [13] as:

$$\mu_s'(\lambda) = C \cdot [0.58(\lambda/1\mu\text{m}) - 0.1] \cdot 0.32(\lambda/1\mu\text{m})^{-2.4}, \tag{2.3}$$

where C [ml/l] denotes the concentration of 20%-intralipid in the phantom and λ refers to the light wavelength in μm. This expression is particularly valid in liquid phantoms with low concentrations of intralipid, where a more linear behavior exists between the intralipid concentration and the reduced scattering coefficient [14].

2.1.2 Absorption

The absorption coefficient ($\mu_a(\lambda)$ $[cm^{-1}]$) of a tissue is defined as the absorption probability of light per unit path length. It can be described as $\mu_a(\lambda) = \sum_i \varepsilon_i C_i$, where ε_i $[M^{-1}cm^{-1}]$ is the extinction coefficient, C_i $[M]$ is the molar concentration of chromophore i. The main absorbing chromophores in biological tissues at visible and near infrared regions are blood, water and lipid. For this reason, the absorption coefficient can be described as the sum of the absorption coefficients of each of these chromophores weighted by their respective concentrations as shown in Equation 2.4:

$$\mu_a(\lambda) = v_{\text{blood}}\,\mu_a^{\text{blood}}(\lambda) + v_{\text{water}}\,\mu_a^{\text{water}}(\lambda) + v_{\text{lipid}}\,\mu_a^{\text{lipid}}(\lambda), \tag{2.4}$$

where v_i is the volume fraction of the particular chromophore and μ_a^{blood}, μ_a^{water} and μ_a^{lipid} refer to the absorption coefficients of the particular chromophores for 100% blood (assuming 150g Hb/l), 100% water and 100% lipid, respectively. Whole blood consists of hemoglobin (Hb) and oxy-hemoglobin (HbO$_2$) that are connected to each other with the oxygen saturation parameter (SO$_2$) as SO$_2$ = [HbO$_2$]/[HbO$_2$] + [Hb]. SO$_2$ value can vary from approximately 97% in arterial blood to 75% in venous blood [15].

The absorption behavior of the main absorbing chromophores in tissue for wavelength range of 500 nm to 1200 nm is depicted in Figure 2.2. Optical window or NIR window can be seen in the wavelength range of 650 nm to 1000 nm. The propagating light penetrates the deepest at this window, making scattering more dominant than absorption. Water and lipid absorption coefficients slightly vary with temperature [16, 17], and therefore ensuring a stable temperature during experimental studies is necessary.
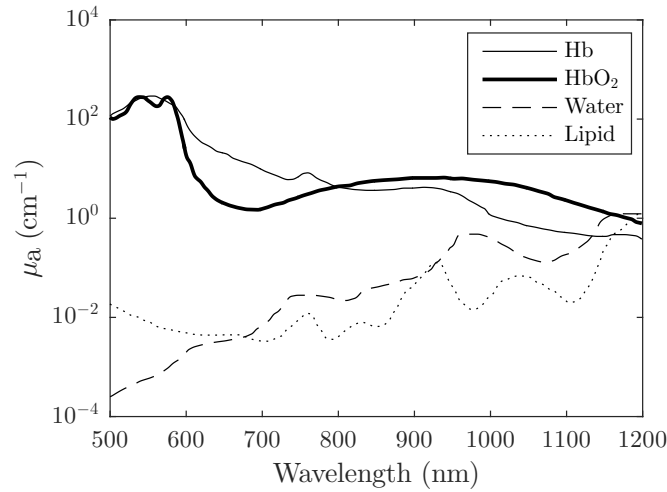

Fig. 2.2: Absorption coefficient of major blood chromophores with 100% volume fractions [18].

2.2   Simulating the diffuse light

2.2.1   Forward Problem

Generally a forward problem refers to the mathematical description of a given phenomenon using the laws of physics. In DRS, this consists of a model numerically describing the light propagating in tissue with known optical properties. Four primary physical quantities concerning formulating the diffuse light in a turbid media using a forward model are mentioned below:

- **Photon distribution** $N(\vec{r}, \hat{s}, t)$ [$1/m^3$sr] is the number of photons propagating in direction $\hat{s}$ per unit volume per unit solid angle at position $\vec{r}$ [m] and time t [s].

- **Radiance** $L(\vec{r}, \hat{s}, t) = h\nu c N(\vec{r}, \hat{s}, t)$ [W/m^2sr] is the radiant flux in direction $\hat{s}$ per unit area per unit solid angle at position $\vec{r}$ and time t, where h is the Planck's constant [m^2kg/s], ν is light's frequency [$1/s$] and c is the speed of light in tissue [m/s].

- **Fluence rate** $\Phi(\vec{r},t) = \int\limits_{4\pi} L(\vec{r},\hat{s},t)d\omega = h\nu c\rho(\vec{r},t)$ $[\mathrm{W/m^2}]$ is the power per unit area at position $\vec{r}$ and time t, where ρ denotes the photon density $[1/\mathrm{m^3}]$.

- **Photon flux vector** $\mathrm{J}(\vec{r},t) = \int\limits_{4\pi} L(\vec{r},\hat{s},t)\hat{s}d\omega$ $[\mathrm{W/m^2}]$ is the axial energy transfer of photons per unit area at position $\vec{r}$ and time t.

2.2.1.1   Radiative transport equation

The radiative transport equation (RTE), also known as the radiative transfer equation, is the fundamental model for formulating the light transport in turbid medium such as biological tissues [19, 20]. The basis of derivation of RTE equation lies in conservation of energy in a specific direction $\hat{s}$ within a small volume V. Figure 2.3 shows 5 possible events that are taken into account in derivation of RTE for a photon travelling in a small volume V. The numbers correspond to the particular event mentioned below:
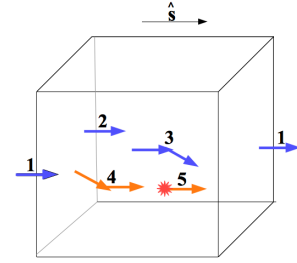

Fig. 2.3: Different scenarios for a photon travelling in a small volume V.

1. Photon in direction $\hat{s}$ passing through the boundaries.

2. Photon in direction $\hat{s}$ absorbed.

3. Scattering of a photon in direction $\hat{s}$ into any other direction.

4. Scattering of a photon from any direction into direction $\hat{s}$.

5. Photon emitted in direction $\hat{s}$ from the source.

Equation 2.5 takes into account these 5 events for describing the distribution of photons travelling in a certain direction $\hat{s}$ within a small volume V. In short, photon losses occur due to scattering into another direction and absorption, while photon gains are due to scattering from another directions into beam direction $\hat{s}$ and radiation sources. Photon losses and gains can be observed by the negative and positive signs in Equation 2.5, respectively. Each segment number refers to the particular event mentioned above. Nonlinear effects, electromagnetic wave properties (e.g. coherence, polarization) and particle characteristics (e.g. inelastic collisions) are not considered when deriving the radiative transport equation.

$$\int_V \frac{\partial N}{\partial t}dV = -\underbrace{\int_S cN\hat{s}.\hat{n}\mathrm{d}A}_{1} - \underbrace{\int_V c\mu_a NdV}_{2} - \underbrace{\int_V c\mu_s NdV}_{3} +$$

$$\underbrace{\int_V c\mu_s \int_{4\pi} p(\hat{s}',s')N(\hat{s}')\mathrm{d}\omega'\mathrm{d}V}_{4} + \underbrace{\int_V q\mathrm{d}V}_{5}, \qquad (2.5)$$

where $p(\hat{s}',s')$ is the scattering phase function characterizing the probability of scattering of the travelling light from the direction $\hat{s}'$ to $\hat{s}$ with a solid angle $\mathrm{d}\omega'$. $q(\vec{r},\hat{s},t)$ describes the light source at position $\vec{r}$ and time t.

Rewriting Equation 2.5 in terms of radiance gives the Equation 2.6, which describes the radiance variation in direction $\hat{s}$, time t and location $\vec{r}$ [19]:

$$\frac{1}{c}\frac{\partial L}{\partial t} = h\nu q + \mu_s \int_{4\pi} p(\hat{s}',\hat{s})L\mathrm{d}\omega' - \hat{s}\cdot\nabla L - \mu_s L - \mu_a L. \qquad (2.6)$$

Solving the Equation 2.6 for more advanced geometries is a non-trivial and computationally demanding task. In fact, analytical solutions are only obtained for isotropic medium [1] [21] and simple geometries [22, 23]. Methods such as Monte Carlo simulations (MC) [2] and Finite Element method (FEM)[3] are employed to solve the RTE.

Diffusion Equation

The diffusion equation (DE) is an approximation of the RTE commonly used for preliminarily analysis of light propagation in tissues and in applications such as dosimetry (e.g. in Photodynamic therapy), optical imaging and spectroscopy due to its high computational efficiency. The number of independent variables in DE are reduced as directional dependency is omitted during the course of derivation from RTE[4]. Here the radiance L is expanded into its first order spherical harmonics and hence ensuring a nearly isotropic source (anisotropy factor $g \ll 1$). The medium for which the diffusion approximation is valid, is assumed to be highly scattering

---

[1]Medium with uniform scattering of light in all directions.

[2]Probabilistic technique widely used in analysing scenarios where there are many possible outcomes with various contributing factors. More information about modeling of the light transport in tissue using MC techniques can be found in [24].

[3]Numerical technique used to solve for boundary value problems. More information can be found in [25].

[4]A hand waving explanation of DE derivation is presented in this work. The detailed derivation of DE can be found in Wang and Wu's book on biomedical imaging [26].

with $\mu'_s \gg \mu_a$. This ensures a sufficient number of scattering events before losing photons through a tissue boundary.

The diffusion theory is governed by the time-resolved Equation 2.7 that yields the fluence rate (rather than the radiance as in the case of RTE). Here D is the diffusion coefficient and $q_o(\vec{r},t)$ denotes an isotropic light source:

$$\frac{1}{c}\frac{\partial \Phi(\vec{r},t)}{\partial t} - \nabla.[D(\vec{r})\nabla\Phi(\vec{r},t)] + \mu_a \Phi(\vec{r},t) = q_o(\vec{r},t). \tag{2.7}$$

By applying appropriate boundary conditions and defining the light source, one can find unique solutions to Equation 2.7 for a specific geometry using the Green's function [27]. For a semi-infinite slab shaped homogeneous medium, an analytical solution is obtained by using a mirror-image approach. Where an isotropic point source (positive source) is placed at the distance z_0 equal to one mean free path length into the medium and the image source (negative source) is placed at the distance $-2z_b - z_0$. An extrapolated boundary is defined at $-z_b$ to account for the Fresnel reflections due to refractive index mismatch between the tissue and surrounding medium by ensuring zero fluence value at the boundary. Figure 2.4 shows the geometry of a semi-infinite slab shaped medium where the light is propagating along the z-axis.
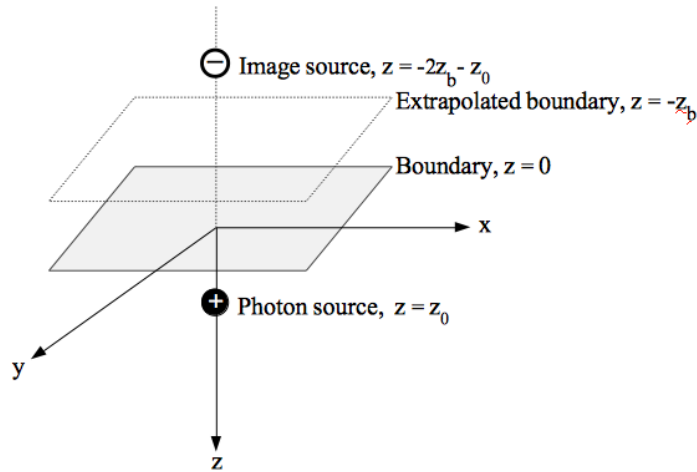

Fig. 2.4: The schematics of a semi-infinite slab shaped medium. Zero fluence is ensured at the extrapolated boundary (dotted line plane) by placing the isotropic point source of photons (black dot) and negative image source (white dot) at the height $z = z_0 + z_b$ above and below the extrapolated boundary.

The model for a point source in a semi-infinite geometry by Farrell et al. [28] is commonly used for simulating the diffuse light within tissue as given in Equation 2.8:

$$
R[\mu_a(\lambda), \mu_s'(\lambda), \rho] = \frac{\mu_s'}{4\pi(\mu_s' + \mu_a)} \Big[z_0 \Big(\mu_{\text{eff}} + \frac{1}{r_1} \Big) \frac{\exp(-\mu_{\text{eff}} r_1)}{r_1^2}
$$
$$
+ (z_0 + 2z_b) \Big(\mu_{\text{eff}} + \frac{1}{r_2} \Big) \frac{\exp(-\mu_{\text{eff}} r_2)}{r_2^2} \Big],
$$

(2.8)

where $\mu_{\text{eff}} = [3\mu_a(\mu_a + \mu_s')]^{1/2}$ is the effective attenuation coefficient, $z_0 = (\mu_a + \mu_s')^{-1}$ is the location of an artificial isotropic photon source, i.e. the depth where the light has first fully lost the direction at origin, $r_1 = (z_0^2 + \rho^2)^{1/2}$ is the distance between the photon source and the collecting fiber and $r_2 = [(z_0 + 2z_b)^2 + \rho^2]^{1/2}$ is the distance between the image source and the collecting fiber. The parameter $z_b = 2AD$ denotes the extrapolated boundary position, where $D = [3(\mu_a + \mu_s')]^{-1}$ is the diffusion coefficient and A is the internal reflection parameter that varies with the refractive index of the tissue and surrounding medium described in Groenhuis et al. paper [29]. For a matched boundary, the internal reflection parameter is equal to 1.

Despite the simplicity of applying the DE, there are some considerations that need to be made prior to its application. The fluence rate can be best measured deep into the tissue away from the light source, however in order for the DE to be valid an adequately large distance between observation point and the source is required [30]. Additionally, the assumption of having isotropic media only allows obtaining the analytical solutions for simple probe-sample geometries (e.g. a slab) with homogeneous optical properties. Lastly, neglecting the absorption effects can be problematic in the case of highly vascularized tumors and for applications such as bioluminescent and fluorescent imaging (due to the relatively high absorption of the bioluminscence markers and fluorescent proteins in the wavelength range of 400 to 600 nm).

2.3   Quantification of chromophores

Accurate extraction of tissue optical properties from the experimentally obtained DRS spectra is not a trivial task. Model-based approaches are commonly used for this purpose where the light distribution within the sample is mathematically modeled (e.g. using DE) and then fitted to the experimental data. However, as explained before, there are certain limitations involved with these models. More accurate and versatile methods have been investigated to overcome these limitations both empirically and experimentally. Examples include Inverse-MC models [31] and more sophisticated probe configurations [32–34]. An alternative approach is using

a lookup table-based inverse model (LUT) constructed experimentally from calibration standard phantoms where optical properties are extracted by iterative fitting of the reflectance signals [35, 36]. LUT has the advantage of independency from light propagation models without the need of altering the conventional measurement probe geometries; however, it is very computationally demanding. A more recent and less explored technique is using the multivariate analysis (MVA) tool that enables both regression and classification studies. In this work, optical data extraction within the MVA framework is investigated.

2.3.1   Inverse problem

Generally, an inverse problem describes situations where the solution is known but not the question. In DRS this refers to the process of extracting attributes of the tissue properties from diffuse reflectance spectra. The performance of an inverse model is evaluated by how accurate the tissue properties are assessed with minimum amount of residuals between the spectra and the forward model. The relation between a model and measurement data with respect to forward and inverse models are depicted in Figure 2.5.
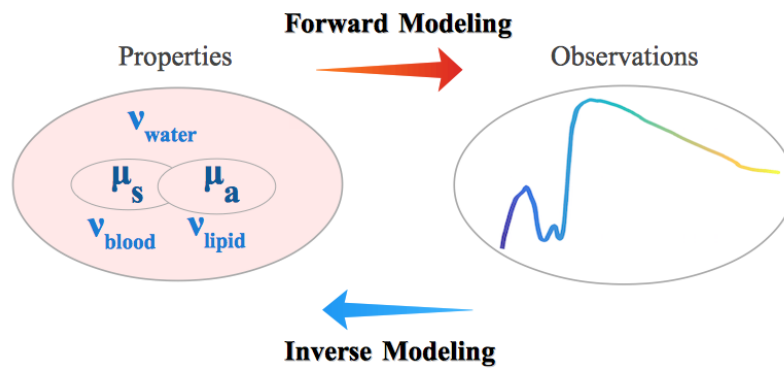

Fig. 2.5: The relation between the model properties and the measurement data in forward and inverse modelling.

2.3.1.1 Model-based inverse problem

In a model-based inverse problem, the diffuse reflectance is first calculated using a forward model such as the diffusion equation or Monte Carlo simulations. Figure 2.6 shows the steps involved in a general model-based inverse problem. A forward model describing the obtained data d for parameter m being sought, e.g. a particular chromophore concentration, can be denoted by $G(m)$. The minimization for a linear problem is computed using the least-squares objective function. The function obtains the best fit for model parameter m that leads to Ω closest to zero in Equation 2.9:

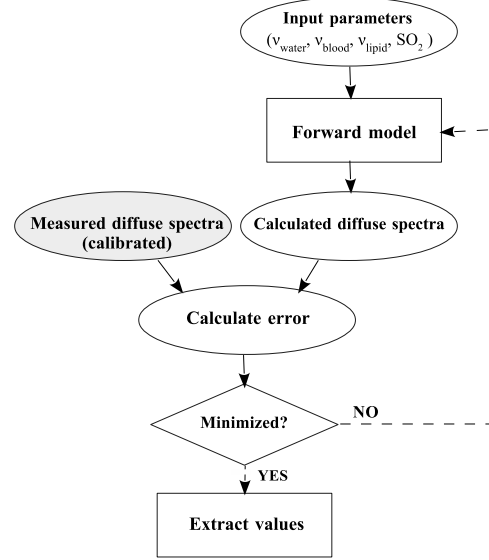$$\Omega = \|d - G(m)\|^2. \qquad (2.9)$$

Fig. 2.6: A model-based inverse problem flow chart.

2.4 Multivariate analysis

Multivariate analysis (MVA) is a statistical modern data analysis tool with the ability to perform multiple variable analysis at the same time. One distinguishing characteristic of MVA, when analysing systems comprising of several variables, is its ability to group a set of factors and assign them to their respective observation. Both regression (dependence) and classification (independence) analysis can be performed within the MVA framework. In regression analysis (also known as function estimation), there is a relationship between different variables of the system whereas the opposite holds for the classification case. In function estimation, a prediction model is constructed for a certain variable sought capable of evaluating unknown datasets. Different multivariate methodologies can be classified based on the linearity state of the input to predictor space transformation. Figure 2.7 shows an example of linear and nonlinear classification of the data in input space. Main techniques with linear transformation are principal component analysis (PCA) and partial least-squares regression (PLS); whereas, both linear and nonlinear transformation can be performed in support vector machines (SVM). One form of SVM is least-squares support vector machines (LS-SVM) that is the multivariate technique used in this work. In this section an overview of these techniques is given.
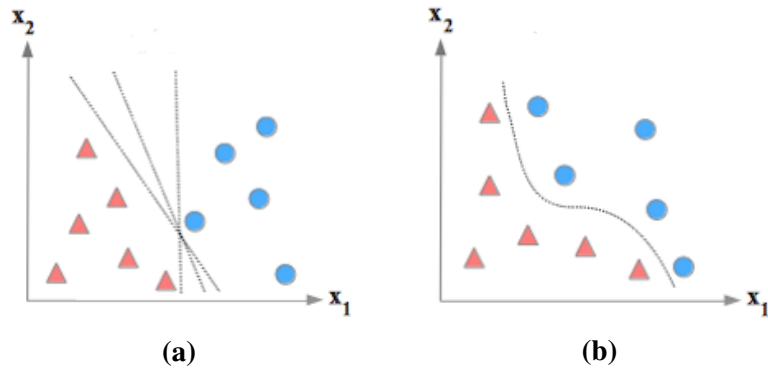
Fig. 2.7: Decision boundary for (a) Linearly separable data and (b) Nonlinearly separable data in 2-case classification problems.

2.4.1 Principal component analysis

Principal component analysis (PCA) is a statistical non-parametric method commonly used for data reduction. It operates on the basis of orthogonal linear transformation, where the observation data is converted into a set of linearly uncorrelated variables called *principal components* or loadings. These principle components are placed in decreasing order of variance with the condition of being orthogonal to the preceding component. PCA can only be used for classification and prediction purposes when combined with a discrimination algorithm.

2.4.2 Partial least-squares

Partial least-squares (PLS) regression is a statistical method that operates on the basis of data transformation into a new space named the predictor space that contains the linear combinations of the original data obtained. Unlike PCA, PLS takes the relation between data into account by finding the maximum correlation between spectral observations. These correlations are used to form a prediction model by finding the predictive relationship between variables. The two main goals here are minimizing the response predictor variation errors and prediction variation errors. In other words, the prediction model performs the transformation using a linear function capable of explaining the maximum number of variations in each response, as well as having the ability to be used for predicting a new set of data.

In a problem with the goal of examining the fundamental relations between observation data and a dependent variable, the input data is $X = \left[\{x_{ij}\}\right]_{m \times n}$ and the quantity to be predicted is $Y = \left[\{y_i\}\right]_{m \times 1}$. For DRS of a set of samples, n denotes the number of spectral observations for each sample and m denotes the number of samples.

2.4.3 Support vector machine

Support vector machine (SVM) is a powerful MVA tool first introduced by Vapnik [37] in 1995. It is comprised of a number of machine learning algorithms enabling nonlinear analysis of data. Both classification and regression studies can be performed using SVM on datasets with linear or nonlinear distributions. The relation between the input regressors (x) and the dependent variable (y) in SVM estimation is expressed in Equation 2.10:

$$y = \omega^T x_i + b, \tag{2.10}$$

where ω denotes the regression coefficient or weight vector from the training dataset and b denotes the bias. The training set consists of N data points ($\{x, y\}_{i=1}^N$), where $x_i \in \mathbb{R}^d$ and $y_i \in \mathbb{R}^d$ are the i-th input and output pattern, respectively. For a nonlinear dataset the general idea is to map the original feature space to a higher-dimensional feature space, where the training set is separable. This ability makes SVM ideal for analysing phenomena with intrinsic nonlinear properties, such as the light transport within diffuse media. Various researchers have examined the effectiveness of using SVM for tissue diagnostics purposes. It has been proven to be successful in differentiating lesion from healthy cells in breast cancer using diffuse reflectance spectroscopy and fluorescence spectra at multiple excitation wavelengths [38], throat cancer and oral cavity cancer using laser-induced fluorescence (LIF) spectroscopy [39, 40], colonic, head and neck tissues using near-infrared Raman spectroscopy [41, 42] and in detection of brain tumours using magnetic resonance spectroscopy [43, 44].

Linear SVM classification

The optimal classifier for a data distribution case with more than one possible solution can be determined using SVM (Figure 2.7a). This is achieved by creating a hyperplane that performs the classification linearly. The decision boundaries of the hyperplane consist of lines meeting an input data point x_i from each class as depicted in Figure 2.8. The points situated on the decision boundaries (marked with $\times$) are called *support vectors*. In a 2-class problem, -1 or $+1$ label is assigned to each class ($y \in \{+1, -1\}$) and the classifier is defined as $y_i = \text{sign}[\omega^T x_i + b]$. Depending on the separability of data points, SVM uses hard or soft margins. Figure 2.8a shows an example of data classification by defining hard margins. Here the data points are linearly separable and the decision boundaries are located at data points x_i that lead to $y = 0$. The optimal classifier is obtained by maximizing the margin width ρ that

is equal to $2/\|\omega\|$ [5]. The optimization problem can be made convex by minimizing $\frac{1}{2}\|\omega^2\|$ instead of maximizing $2/\|\omega\|$ as written in Equation 2.11 [45]:

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}\|\omega^2\| \\
\text{subject to} \quad & y_i(\omega^T x_i + b) \geq 1, \quad \forall A.
\end{aligned}
\tag{2.11}
$$

In the case of noisy datasets, the classifier defined by hard margins can mistake the real data with noise. In such situations, over-fitting occurs even though the classifier appears to be defined correctly. This problem is addressed in SVM by defining soft margins with a slack variable ξ_i assigned to the misclassified points as shown in Figure 2.8b. The new optimization function accounting for these points is defined in Equation 2.12 as:

$$
\begin{aligned}
\text{minimize} \quad & \frac{1}{2}\|\omega^2\| + \gamma \sum_{i=1}^{N} \xi_i \\
\text{subject to} \quad & y_i(\omega^T x_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0,
\end{aligned}
\tag{2.12}
$$

where γ denotes the regularization or scaling parameter representing the distance between the closest points within different classification classes. Parameter γ is a tuning parameter used for governing the balance between the training error minimization and smoothness. The slack parameter ξ_i defines the distance between the misclassified data points and their respective decision boundary.
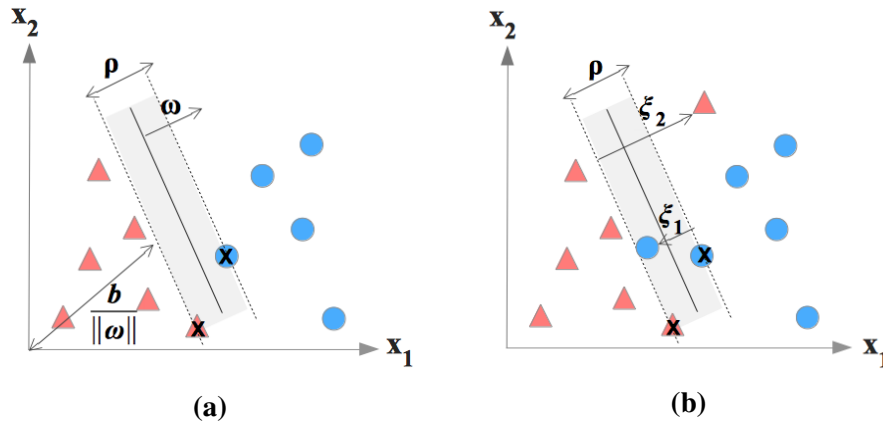

Fig. 2.8: Linear classifier defined in SVM by (a) hard margins and (b) soft margins.

The optimization problems defined in Equations 2.11 and 2.12 can be solved using the Lagrange dual problem. In this method, the optimization function is first written in its

---

[5] $\|\omega\| = \sqrt{\omega\omega^T}$.

Lagrangian form followed by construction of the dual optimization problem. In the dual optimization form, the problem is expressed with respect to only one variable (the Lagrange multiplier). In this way, the problem's complexity will be only dependant on the number of data points and not on the dimension of the feature space. The detailed mathematical procedure for the derivation of the Lagrange dual problem for Equation 2.12 is presented in Appendix A.

Nonlinear SVM classification

The principle of SVM classification for data that is not linearly separable in input space is depicted in Figure 2.9. Here a linear separation hyperplane is created that embeds the data in input space $(x_i \in \mathbb{R}^d)$ with $d = 2$ to a feature space of higher dimension $(\Phi(x_i) \in \mathbb{R}^{d'}, \quad d' > d)$ with $d' = 3$.

$$\Phi: \qquad \mathbb{R}^2 \to \mathbb{R}^3$$
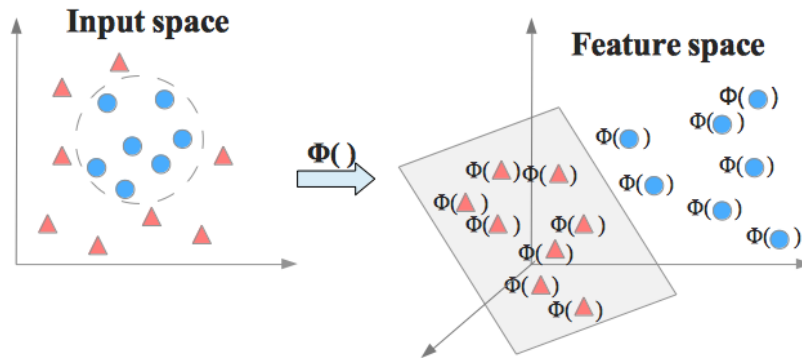$$(x_1, x_2) \mapsto (z_1, z_2, z_3)$$

Fig. 2.9: Transition from input space to feature space for SVM nonlinear classification.

The main advantage of creating this hyperplane is simplifying the solution since a nonlinear operation in input space is equivalent to a linear operation in feature space. The mapping is done using an appropriate kernel function. A kernel function performs the transformation by calculating the dot product of the input space data $(\mathbb{R}^d)$ in feature space $(\mathbb{R}^{d'})$ without having the input data transformed to higher dimensions first [46]. In this way the optimisation is achieved independently of the dimensionality of the feature space. The typical kernel functions used in nonlinear SVM are presented in Table 2.1, where parameter σ denotes the Radial basis function (RBF) kernel width.

Choosing the optimal kernel function is a nontrivial task. In practice this is a trial and error procedure with experimenting different kernel types and optimizing their pa-

Table 2.1: Main kernel types used in nonlinear SVM.

Kernel function	Formula	Remarks
Linear	$K(x_i, x_j) = \Phi(x_i)^T \Phi(x_j)$	
Polynomial	$K(x_i, x_j) = (1 + \Phi(x_i)^T \Phi(x_j))^d$	$d > 0$
Radial basis function (RBF) (i.e. Gaussian)	$K(x_i, x_j) = \exp(\gamma \|x_i - x_j\|^2)$	$\gamma = -1/2\sigma^2$

rameters. However, it is generally recommended to start with using a Gaussian kernel and fine-tuning the classification parameters [47]. The general form of the classifier is $y = \text{sign}[\sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j) + b]$, where $K(x_i, x_j)$ specifies the kernel function applied and α_i is the Lagrangian multiplier that denotes positive real constants called the support values. The data points corresponding to these support values are called support vectors with the same analogy as the case of linear SVM.

Nonlinear SVM regression

In the SVM regression task, the model is constructed by fitting model parameters to a training dataset. The model's task is to make a prediction at new points based on the training dataset. A similar analogy to SVM classification exists in regression problems with the main idea to define a hyperplane that maximizes the margin with the least possible error. The training algorithm in a regression task is relatively more complicated due to having many possible outcomes.

The optimization problem in nonlinear SVM regression as given in Equation 2.13 is defined similarly to the classification task with a new parameter ε [48]. The parameter ε denotes the margin of tolerance, i.e. the range of the true predictions, with a value equal to zero for a perfect prediction.

$$\text{minimize} \quad \frac{1}{2}\|\omega^2\| + \gamma \sum_{i=1}^{N} \xi_i$$
$$\text{subject to} \quad y_i(\omega^T x_i + b) \geq 1 - (\xi_i + \varepsilon_i), \quad \xi_i \geq 0$$

$$(2.13)$$

2.4.3.1 Least-squares support vector machines

The memory demand in support vector machines increases for solving the optimization tasks in larger scale problems. Least-squares support vector machines (LS-SVM) introduced by Suykens and Vandewalle in 1999 [49] addresses this issue by reformulating the optimization problem in such a way that it can be solved using linear equations. More specifically, the inequality constraint in Equation 2.12 is replaced with equality and the slack parameter ξ_i is replaced with a sum squared error cost function (SSE) e_i^2. The parameter e_i is the distance from classifier to the misclassified point as shown in Figure 2.10. The mathematical modifications in LS-SVM is presented in Equation 2.14 as:

Fig. 2.10: Graphical representation of error parameters ξ_i in SVM and e_i in LS-SVM.

$$\text{minimize} \quad \frac{1}{2}\|\omega^2\| + \frac{1}{2}\gamma\sum_{i=1}^{N} e_i^2$$
$$\text{subject to} \quad y_i(\omega^T x_i + b) = 1 - e_i. \tag{2.14}$$

In order to solve this problem, first the optimization function is written in it's Lagrangian form as in the case of nonlinear SVM. Whereas, in the next step a set of linear equations are constructed instead of defining the dual optimization problem. The detailed mathematical procedure used in the derivation of the linear equations is presented in Appendix A. The general form of a LS-SVM function estimator is $\sum_{i=1}^{N} \alpha_i y_i K(x_i, x_j) + b$, where $K(x_i, x_j)$ specifies the kernel function applied and α_i is the Lagrangian multiplier denoting the support values. These two values are proportionally related to each other. In LS-SVM, all the Lagrangian multipliers are nonzero, whereas in the standard SVM case α_i values are mostly equal to zero.

A LS-SVM regressor model is defined by determining the best combination of hyperparameters as in the SVM case. These hyperparameters include the regularization parameter γ assigned to minimize e_i, while avoiding overfitting, and the kernel parameters of the particular kernel function used (e.g. σ in RBF kernel). The process of choosing the right kernel in LS-SVM is the trial and error procedure similar to SVM. The input (X) and output data matrices (Y) in LS-SVM have the same structure as in PLS. The classical method for determining the hyperparameters giving the best model performance is the cross-validation method discussed in the last part of this chapter.
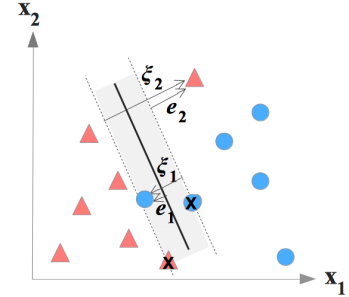
Utilizing LS-SVM for data analysis in tissue diagnostics is an approach recently explored in tissue analysis. Luts et al. [50] investigated the performance of LS-SVM based approach in identifying brain tumors using a combination of MRI and MRSI. The multivariate technique provided noticeably more accurate performance in comparison with linear discriminant analysis technique. In conjunction with DRS, Barman et al. studied using a LS-SVM regression algorithm for tissue optical properties extraction for the first time in 2011[51]. Two LS-SVM regression models μ_s' and μ_a were constructed from 24 tissue-like liquid phantoms comprised of polystyrene microspheres, ink and water. In addition, they compared three techniques of PLS, LUT and LS-SVM together. It was found that LS-SVM gives the best prediction accuracy with the fastest response. In 2014, Xie et al. used LS-SVM for performing validation measurements on 270 liquid phantoms made of different concentrations of water, intralipid, bovine blood and protoporphyrin IX (PpIX) [32]. Three LS-SVM regression models were constructed and tested with 2-fold cross-validation in order to predict μ_s'), blood and PpIX parameters in phantoms with mean prediction errors of 3%, <5% and 5%, respectively. In addition, LS-SVM classification was performed on in-vivo skin tumors with 100% accuracy in distinguishing the healthy cells from lesion.

Cross-validation methods

Cross-validation (CV) is a technique frequently used in performance evaluation of a regression model. The prediction accuracy of a model built from a training dataset is estimated using a testing dataset, i.e. validation dataset. Here instead of using the entire dataset for training the model, part of the dataset (known data) is used for training and the remaining part (unknown data) is used as a testing dataset to evaluate the model performance. In this way, problems such as overfitting or hypotheses suggested by the dataset consisting of various parameters are avoided. In a k-fold CV, the observation data is randomly partitioned into k subsets

1. Randomize the training points.
2. Partition the training points into k folds.
3. For $i = 1,..,k$:
   - Train the regressor using the data points in the remaining $k-1$ folds.
   - Test the regressor on the data points in fold i.
   - Determine the estimation error in fold i.
4. Return the combined estimation error.

Fig. 2.11: k-fold CV algorithm for tuning a LS-SVM regression model.

with equal sizes, having $k-1$ training subsets and retaining one as the testing subset. The cross-validation process is iterated until all partitions are used as the testing subset. The advantage of using a k-fold CV is that each data point is tested at least once regardless of how the dataset partitioning is done. For k equal to the number of samples, a special type of CV is constructed named leave-one-out cross-validation (LOOCV). In LOOCV all measurements

except one are included in the training subset with the remaining one being used as the testing subset. The cross-validation process is iterated until each measurement has been placed in the testing subset once.

In addition to using CV methods for evaluating the LS-SVM regression model performance, these methods are used for tuning the optimization parameters when constructing the model. The The algorithm of a k-fold CV technique used for defining the regression model is presented in Figure 2.11. The estimation error of each trial is computed and combined in order to determine the model accuracy.

Chapter 3

Methods

3.1 Instruments and software

3.1.1 Diffuse reflectance spectroscopy system

Optical system

The diffuse reflectance spectrum is obtained using the optical setup shown in Figure 3.1. The setup consists of a tungsten halogen broadband light source with an integrated shutter (Ocean Optics, HL-2000), a fiber optic probe, two spectrometers with CCD array detectors (Ocean Optics, USB4000) and a computer. Wavelength calibration is done by assigning a wavelength value to each pixel of the CCD array detectors. The fiber optics probe tip is
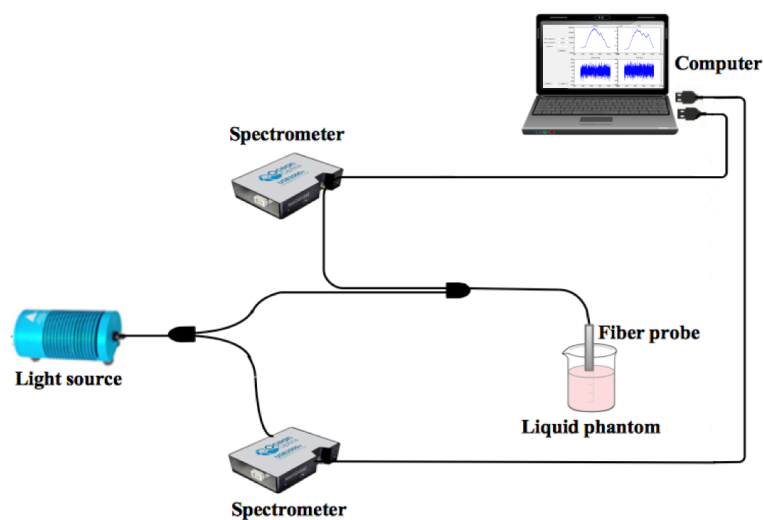

Fig. 3.1: Schematic diagram of the optical setup for the diffuse reflectance spectroscopy.

held perpendicularly in contact with the liquid phantom's surface. One of the spectrometers, referred to as the reference spectrometer, records the lamp spectrum and its background. The second spectrometer, referred to as the measurement spectrometer, records the reflected light's spectrum of the sample and its background using the fiber optic probe. The lamp operates on TTL mode so that the lamp/sample spectra are collected when the shutter is open, and the respective background spectra are collected when the shutter is closed. In this way, data calibration can be performed in such a way that the resulting spectrum is independent of the temporal fluctuations in the lamp intensity. Obtaining the background spectra is necessary in order to account for the dark current and electric offsets of the detectors. In order to minimize the background noise level, the ambient light is suppressed as much as possible by shielding the phantom container and taking measurements at a dark laboratory. Lastly, the DRS spectrum is obtained by calibrating the measured signal as explained in the Section 3.3.

Fiber optic probes

A fiber optic probe is a collection of optical fibers each used either for delivering or collecting the light. Depending on the arrangement of the fibers, different source detector distances can be achieved. Two fiber optic probes with different source-detector separations (SDS) are used in this work as depicted in Figure 3.2.
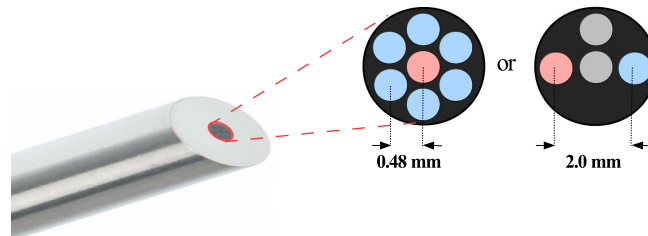

Fig. 3.2: Schematics of the cross-sections of the two fiber optic probes used in obtaining the diffuse signal. A red circle highlights the location of the illuminating fiber optic, a blue circle highlights the location of the collecting fiber optic and a gray circle indicates an unused fiber optic in the probe.

- The first probe used (Avantes, Reflection Probe (FCR-7IR400-2-ME)) consists of six collecting fibers in a circular arrangement and a single illumination fiber located at the center, bundled together into a steel tube. The collecting fiber is coupled to the measurement spectrometer and the illuminating fibers deliver the light from light

source. The distance between the centers of the illumination and collecting fibers is equal to 0.48 mm in this probe.

- The second probe used is a custom built probe consisting of four fibers, one placed at the center of the geometry and the other three located adjacent to it. The customized design of this probe enables achieving different source-detector separations depending on which fibers are used for illuminating and collecting the light. In the arrangement shown in Figure 3.2, source-detector separation of 2.0 mm is achieved by choosing the fibers located furthest apart from each other for illumination and collection purposes. Using this probe with a larger SDS makes it possible to look deeper into the tissue, where absorption phenomena are more pronounced.

Software

The broadband light source is controlled by a custom MATLAB program that acquires and plots the signals from the two reference and measurement spectrometers as well as their respective backgrounds. The exposure times of the two spectrometers and the number of signal acquisitions can be modified using the program's interface.

3.2 Phantom Preparation Protocol

Varying the blood volume and scattering

Phantom experiments are carried out in order to construct the dataset used for evaluating the performance of the multivariate regression models. To this end, tissue-like liquid phantoms are prepared by pipetting different amounts of the major chromophores existing in a biological tissue. Each phantom consists of a unique composition of water, bovine blood (purchased from a local supermarket) and intralipid (Sigma-Aldrich, I141; 20% emulsion). 81 phantoms consisting of (1%, 2%, 3%, 4%, 5%, 6%, 7%, 8% and 9% , v:v) intralipid and (1%, 2%, 3%, 4% and 5%, 6%, 7%, 8% and 9%, v:v) blood are prepared for the DRS measurements using the optical fiber probe with short SDS. 54 phantoms consisting of (1%, 2%, 3%, 4%, 5%, 6%, 7%, 8% and 9% , v:v) intralipid and (0.05%, 1%, 2%, 3%, 4% and 5%, v:v) blood are prepared for the DRS measurements using the optical fiber probe with long SDS. The volume fraction of water is then calculated for each sample in such a way that the sum of the volume fractions of blood, intralipid and water add up to 100%. Phantoms are prepared in glass vials covered in black tape with the purpose of reducing the background noise. Prior to the DR spectrum acquisition, each phantom is sonicated for 2 to 3 minutes in order to disperse the

solution. It can be assumed that there is no temperature variation between different phantoms as all measurements are carried out in room temperature.

Varying the scatterer

Mixtures of varying amounts of intralipid and water are prepared for calibration purposes as discussed in Section 3.3. Measurements were done using the two probes for total of 3 phantoms consisting of (2%, 4%, 6%, v:v) intralipid.

3.3   Data calibration

The data collected by the spectrometers need to be calibrated in order to obtain a diffuse reflectance spectrum independent of the background noise, light source intensity level, spectral sensitivity etc. A self-calibrating method similar to the one described by Yu et.al. [52] is used here, where the phantom spectrum is normalised with respect to another spectrum using a calibration factor (CF). In order to obtain the CF, DRS measurement is carried out on an intralipid-water solution, i.e. intralipid bath, or a reflectance standard, i.e. *puck*. Obtaining a reflectance measurement is necessary in order to account for the spectral shape of the light source and the spectral sensitivity of the detectors. For these reasons, scattering properties are usually measured with respect to a reference spectrum collected from an intralipid-water solution or a reference standard. The calibration factor is then calculated using Equation 3.1:

$$CF = \left(\frac{I^{IL/puck} - I_{bkg}^{IL/puck}}{I^{lamp} - I_{bkg}^{lamp}} \right) \cdot \left(\frac{t_{exp}^{lamp}}{t_{exp}^{IL/puck}} \right), \tag{3.1}$$

where $I^{IL/puck}$ and I_{bkg}^{Lamp} refer to the spectra collected by the measurement spectrometer from the calibration sample (intralipid or puck) and its background, respectively. I^{Lamp} and I_{bkg}^{Lamp} refer to the spectra collected by the reference spectrometer from the lamp and its background, respectively. $t_{exp}^{IL/puck}$ and t_{exp}^{lamp} denote the exposure times of the measurement and reference spectrometers, respectively. The diffuse reflectance spectrum acquired from the phantom is calculated in the same manner as:

$$R = \left(\frac{I^{phantom} - I_{bkg}^{phantom}}{I^{lamp} - I_{bkg}^{lamp}} \right) \cdot \left(\frac{t_{exp}^{lamp}}{t_{exp}^{phantom}} \right). \tag{3.2}$$

Finally the calibrated spectrum is calculated as:

$$R_{\text{calibrated}} = \frac{R}{CF}. \tag{3.3}$$

There are fundamental differences between using lipid-water solutions or reflectance standards as reference measurements for calculating the calibration factor. In this work, these two methods are compared with each other in order to achieve a more accurate and flexible calibration technique. For reasons explained below, the reference measurements from the reflectance standard is utilized for calibration of all the diffuse spectra.

Calibration using lipid-water solutions

Diffuse spectra are collected from intalipid-water solutions consisting of 2%, 4% and 6% lipid to be used as the reflectance spectra. In Figure 3.3a, the DR spectra of 3 phantoms
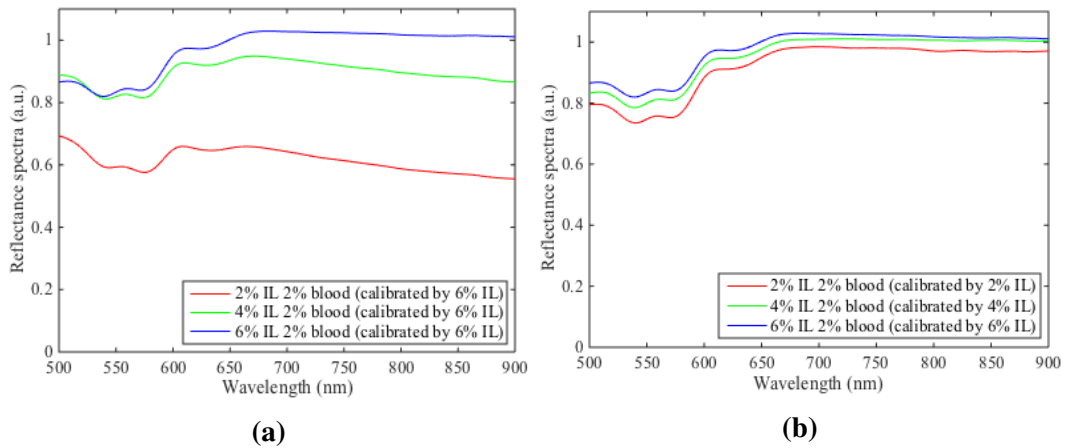

Fig. 3.3: Calibration of three phantoms using intralipid-water solution with (a) 6% intralipid and (b) 2%, 4% and 6% amounts of intralipid corresponding to the intralipid concentration present in the particular liquid phantom.

consisting of 2% blood plus 2% intralipid, 2% blood plus 4% intralipid and 2% blood plus 6% intralipid are calibrated using the spectrum of an intralipid-water solution with 6% intralipid. A decrease in the signal is observed at longer wavelengths due to the scattering behavior of intralipid. A more significant decay is present for larger differences between the amount of intralipids in the liquid phantoms and the intralipid-water solutions. In Figure 3.3b, the 3 phantoms consisting of 2% blood plus 2% intralipid, 2% blood plus 4% intralipid and 2% blood plus 6% intralipid are calibrated using 3 intralipid-water solutions consisting of 2%, 4% and 6% intralipid, respectively. The calibrated spectra have converged with a more subtle

decay at longer wavelengths. This can be explained by having matching amounts of intralipid in the liquid phantoms and the intralipid-water solutions.

In general, intalipid-water solutions with strong scattering and as low absorption as possible are preferred to be used as the reference spectra. The advantage of using an intralipid-water solution for calibrating the DR signals is having the same measurement geometry as the tissue-like phantoms, i.e. the reflected light is recorded after penetrating through the reference material. On the other hand, the reduction in intensity at higher wavelengths makes intralipid not an ideal scattering reference. This decay is less significant when the DR spectrum is calibrated using an intralipid-water solution with the same amount of intralipid as the liquid phantom. However, this is not a feasible solution specially with large number of unknown samples.

Calibration using reflectance standards

A white plastic with 5 cm diameter providing a highly Lambertian surface is used for the reflectance standard, as shown in Figure 3.4. The puck reflects the light uniformly by picking up the lamp spectrum with a reflectance value close to 1. The measurement geometry here is different compared to the DR spectra acquisition from the liquid phantoms, i.e. the light scattered from the puck surface is collected at different distances from the source. The DR spectra is acquired at various heights with the probe held perpendicularly at a distance above the reflectance standard. At 2 mm height, the signal with highest intensity is obtained.

Fig. 3.4: The puck used for data calibration.

Figure 3.5 shows the calibrated spectra of 3 phantoms with different scatterer concentrations using the puck reference spectrum at 2 mm height. It can be observed that the decay existing at higher wavelengths when using intralipid-water solutions is not present here. This is because the puck's reflectance behavior is not spectrally dependant. The advantage of using a puck for reference measurements is that all the phantoms regardless of their compositions can be calibrated using only one reference spectrum. Using one CF in order to calibrate the DR spectra of all the liquid phantoms is preferred for creating a more robust dataset for LS-SVM analysis. The important factor when constructing the dataset for LS-SVM regression is that both the training and testing subsets must have been calibrated in a similar way. For these reasons, all the DR spectra of the tissue-like phantoms are calibrated using the puck reflectance spectra at 2 mm.
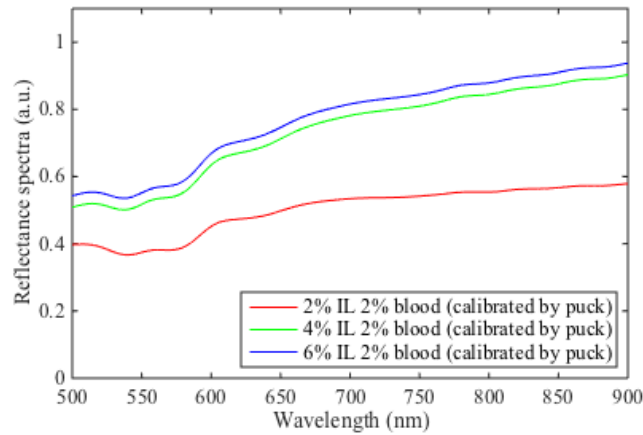
Fig. 3.5: Phantoms with different amount of intralipid and blood calibrated using the reflectance standard.

3.4 Extraction of the optical properties

3.4.1 LS-SVM

LS-SVM algorithm for the purpose of assessing the chromophore concentrations within the multivariate analysis framework is used in this work. Regression analysis is done with the diffuse reflectance spectra for different tissue-like phantoms as input regressors (X) and the reduced scattering coefficient at 700 nm ($\mu_s'(700)$) and blood volume fraction (ν_{blood}) of the particular phantoms as the dependent variables (Y). There is a need to construct an individual model for each dependent variable that is going to be investigated. The reason for not constructing a model based on the intralipid volume fraction is the rather constant absorption behavior of the intralipid in the wavelength range of 500 nm to 900 nm as shown in Figure 2.2.

Fig. 3.6: Different steps involved in LSSVM analysis.

The steps taken in order to predict new values using the constructed regression model are depicted in Figure 3.6. A model structure is formed by defining different parameters such as the input regressors (X) and their dependant variables (Y), the Kernel function type and
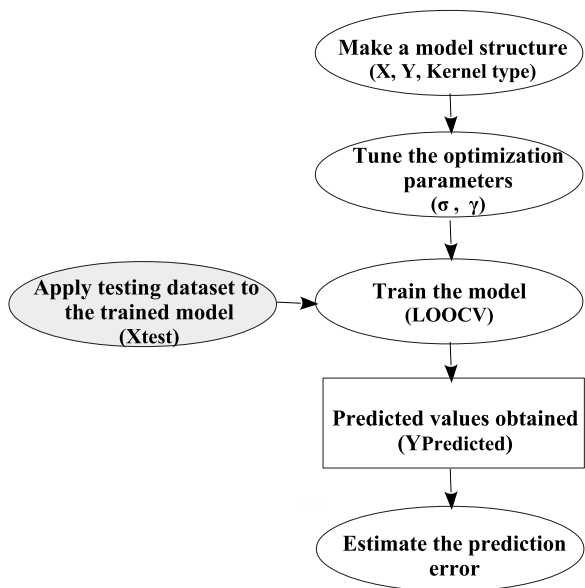
the optimization parameters σ and γ. The dimensionality of the model is defined by the number of columns in the matrix X. In order to evaluate the models' performance, the dataset is divided into two training and testing subsets. The training subset is used to build the regression model and the testing subset is used as validation data for evaluating the model by estimating the prediction error. The optimization of the model parameters is done using LOOCV method as in this way each datapoint is used both for building and testing the model. Two regression models for $\mu'_s(700)$ and v_{blood} are constructed using Gaussian (RBF) kernel function. The training and testing experimental datasets have been normalized with respected to the mean value of each individual spectrum. Three different cross validation techniques are used in order to test the performance of the prediction models, namely LOOCV, 9-fold/6-fold CV and 2-fold CV.

Datasets

- **Experimental dataset**

  - The diffuse reflectance spectra of 81 phantoms collected using the short SDS optical fiber probe.

  - The diffuse reflectance spectra of 54 phantoms collected using the long SDS optical fiber probe.

- **Simulation dataset** The diffuse reflectance spectra for the set of phantoms mixed in the experimental part for the short SDS probe is synthetically calculated using the diffusion equation for a semi-infinite geometry expressed in Equation 2.8. The reduced scattering coefficient of each phantom at 700 nm is calculated using the Equation 2.3. A total of 81 phantoms consisting of (1%, 2%, 3%, 4%, 5%, 6%, 7%, 8% and 9% , v:v) intralipid and (1%, 2%, 3%, 4% and 5%, 6%, 7%, 8% and 9%, v:v) blood are simulated. The water concentration and oxygenation values are kept constant at 80% and the SDS is set as 0.5 mm.

Software

Off-line data analysis for extraction of the optical properties using LS-SVM technique is performed in the Matlab® environment. The MATLAB toolbox StatLSSVM [53] is used for performing the regressing analysis and tuning the regression parameters kernel width σ and regularization parameter γ.

Chapter 4

Results

4.1 Datasets

The input dataset affects the performance of the constructed regression model. The general shape of the diffuse reflectance spectra varies depending on the type of the forward model used theoretically or the SDS of the fiber optic probe used experimentally. Figures 4.1, 4.2 and 4.3 show the DR spectra for added (a) absorber and (b) scatterer for the simulation, short SDS and long SDS datasets, respectively. The reduced scattering coefficient regression model is built using the DR spectra in the order of added absorber. Whereas, the blood volume fraction model is built using the DR spectra in the order of added scatterer.
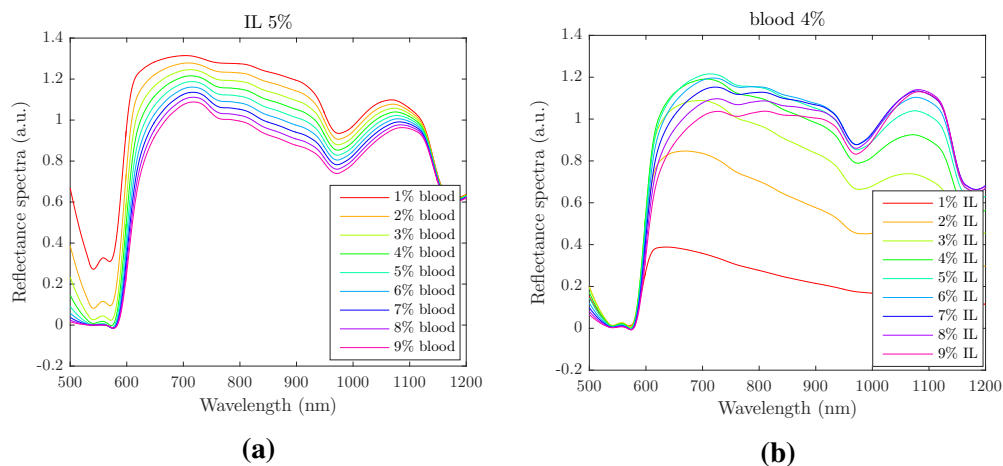

Fig. 4.1: Diffuse reflectance spectra simulated using the diffusion coefficient for (a) different amounts of blood and 5% intralipid and (b) different amounts of intralipid and 4% blood.
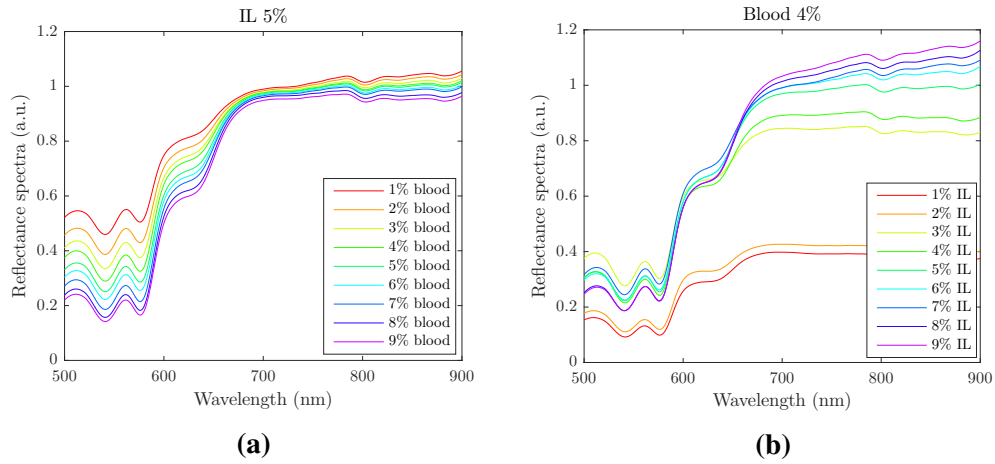
Fig. 4.2: Diffuse reflectance spectra collected using the short SDS fiber optic probe for (a) added absorber and (b) added scatterer.
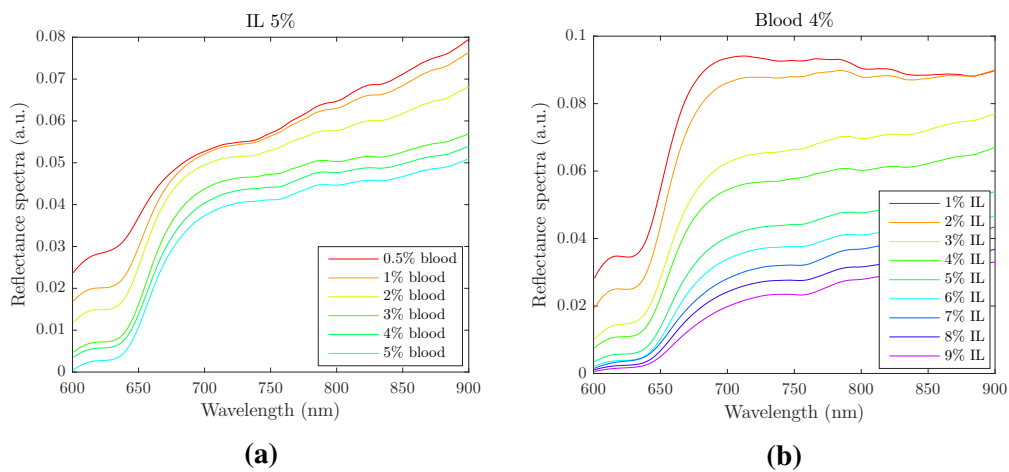

Fig. 4.3: Diffuse reflectance spectra collected using the long SDS fiber optic probe for (a) added absorber and (b) added scatterer.

4.2   LS-SVM data extraction

In this section the prediction model performances for the three datasets are evaluated and compared to eachother using cross-validation techniques of LOOCV, 9-fold CV (/6-fold CV) and 2-fold CV. In the simulation dataset, 2-fold CV corresponds to using 41 samples for training and 40 samples for validation. In the short SDS dataset, LOOCV corresponds to using 80 samples for training and 1 sample for validation, 9-fold CV corresponds to 72 samples for training and 9 samples for validation and lastly, 2-fold CV corresponds to using 41 samples for training and 40 samples for validation. In the long SDS dataset, LOOCV corresponds to using 53 samples for training and 1 sample for validation, 9-fold CV corresponds to 49 samples for training and 5 samples for validation, 6-fold CV corresponds to 45 samples for training and 9 samples for validation and lastly, 2-fold CV corresponds to using 27 samples for training and 27 samples for validation. Figures 4.4, 4.5 and 4.7 depict the difference between the true and predicted models for the simulation, short SDS and long SDS datasets, respectively. Figures 4.6 and 4.8 compare the percentage errors of the predicted parameters using different validation techniques for the experimental datasets.

4.2.1   Evaluation of simulation dataset

(a)   (b)

Fig. 4.4: Performance of the simulated prediction models for (a) reduced scattering coefficient at 700 nm and (b) blood volume fraction using 2-fold CV technique. The dashed line indicates the diagonal of best prediction, the red dots indicate the predicted points and the errorbars show the mean difference between the true and predicted values. R is the correlation coefficient between the true and predicted values.
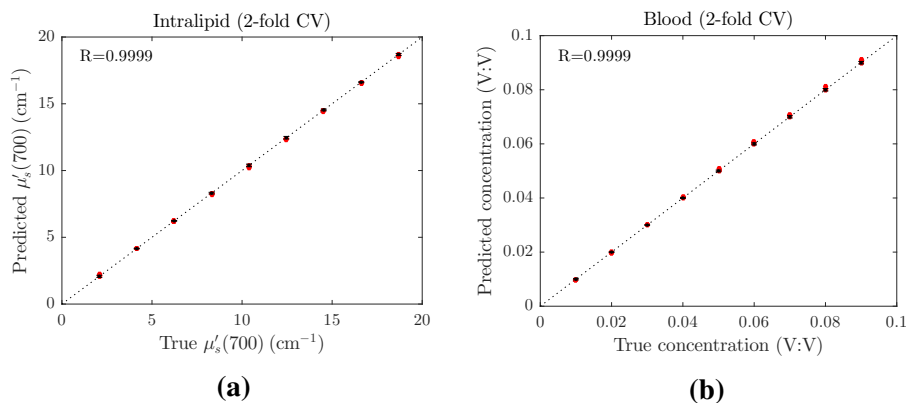
4.2.2   Evaluation of experimental datasets

Fig. 4.5: Performance of the short SDS prediction models for (a-c) reduced scattering coefficient at 700 nm and (d-f) blood volume fraction using different CV techniques. The dashed line indicates the diagonal of best prediction, the red dots indicate the predicted points and the errorbars show the mean difference between the true and predicted values. *R* is the correlation coefficient between the true and predicted values.

Fig. 4.6: Prediction error of the long SDS regression models for (a) reduced scattering coefficient at 700 nm and (b) blood volume fraction using different CV techniques. The bars indicate the mean percentage error ($100 \cdot |\text{True} - \text{Predicted}|/\text{True}$) of the predicted values for the respective model parameter and the applied cross-validation technique.
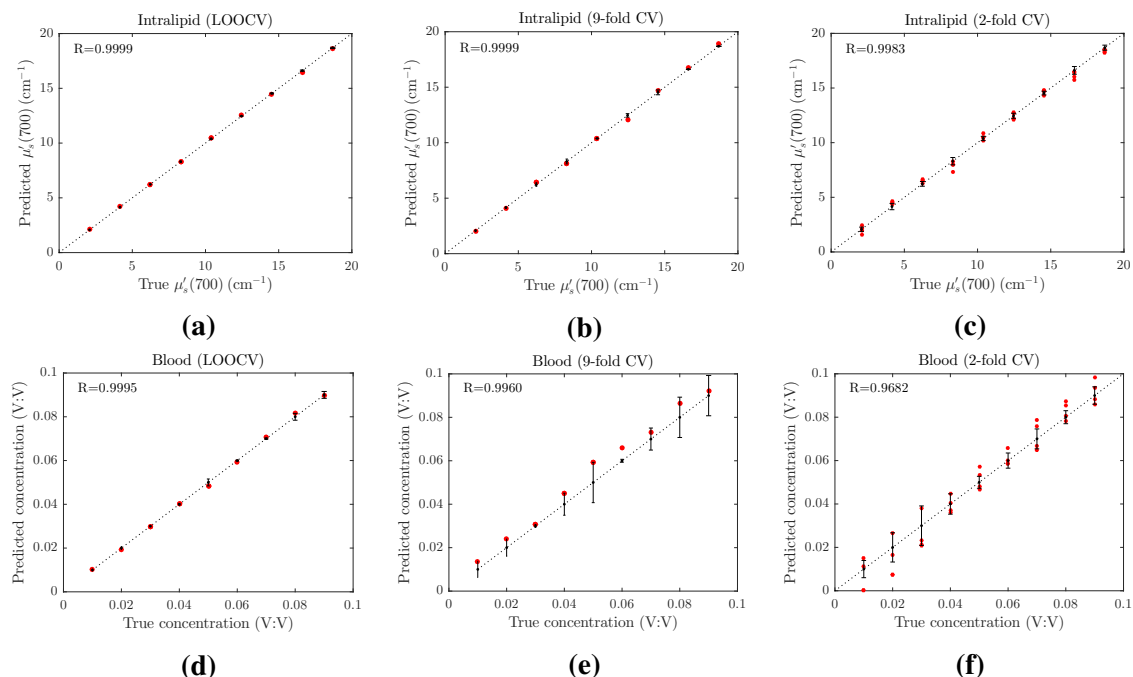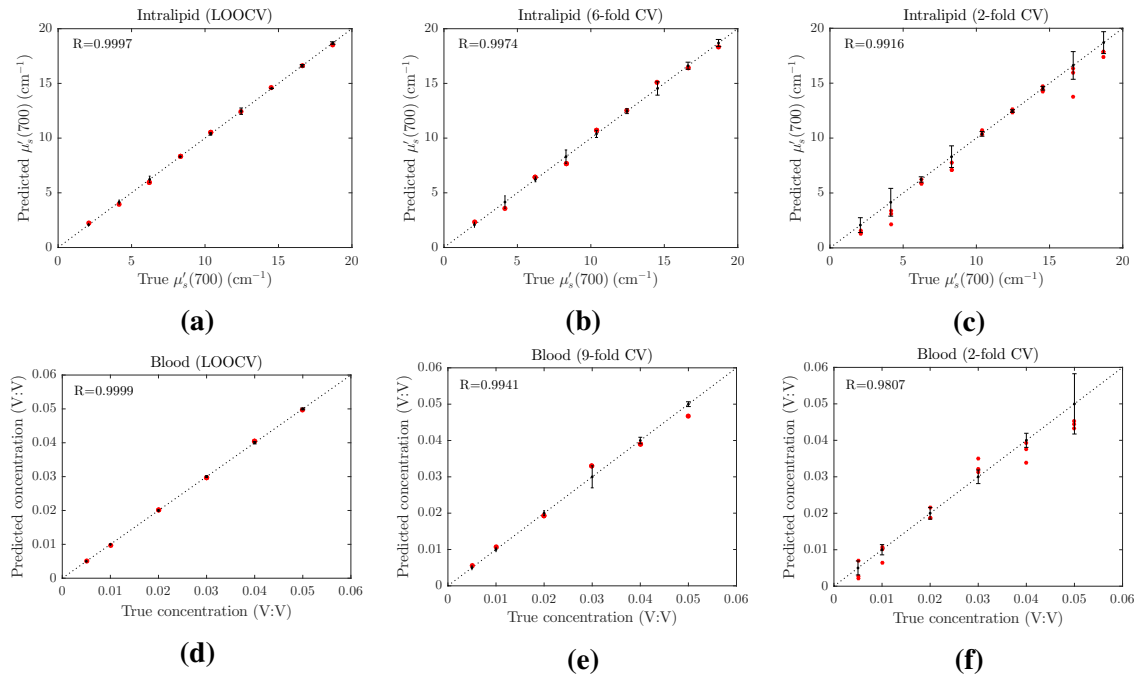
Fig. 4.7: Performance of the long SDS prediction models for (a-c) reduced scattering coefficient at 700 nm and (d-f) blood volume fraction using different CV techniques. The dashed line indicates the diagonal of best prediction, the red dots indicate the predicted points and the errorbars show the mean difference between the true and predicted values. R is the correlation coefficient between the true and predicted values.
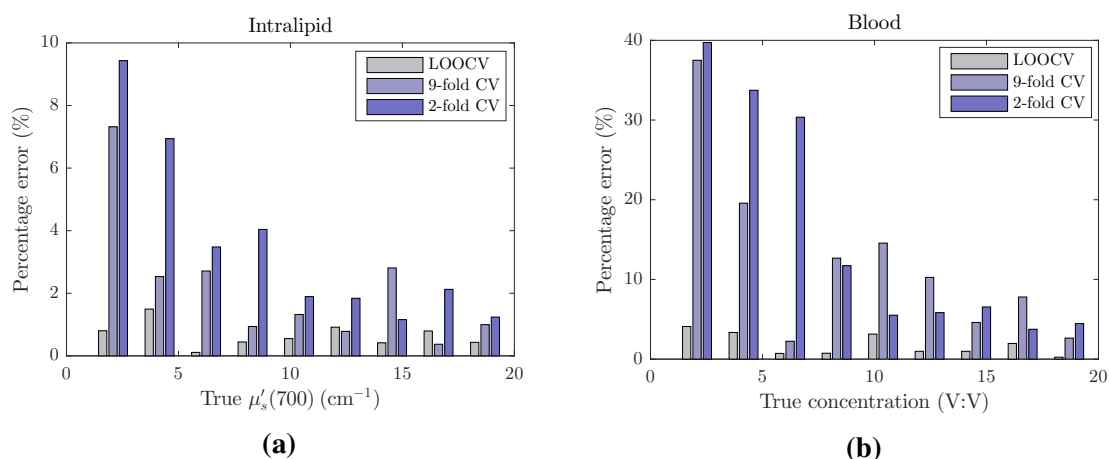

Fig. 4.8: Prediction error of the long SDS regression models for (a) reduced scattering coefficient at 700 nm and (b) blood volume fraction using different CV techniques. The bars indicate the mean percentage error ($100 \cdot |\text{True} - \text{Predicted}|/\text{True}$) of the predicted values for the respective model parameter and the applied cross-validation technique.
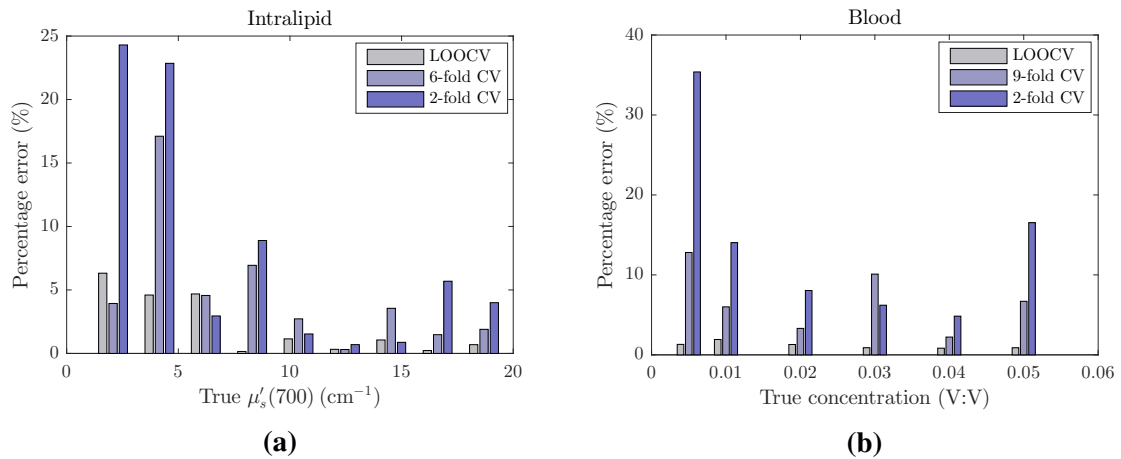
4.2.3 Summary of models' performances

The mean percentage errors of the models evaluated using the three different CV techniques are presented in Table 4.1. Table 4.2 contains the correlation coefficient of the models evaluated using the three different CV techniques presented tabularly for an easier comparison.

Table 4.1: The mean percentage error of the regression models for short SDS and long SDS datasets.

		Mean percentage error [%]		
	Regression model	LOOCV	9-fold/6-fold CV	2-fold CV
Short SDS	Intralipid	0.6626	2.0363	3.5719
	Blood	1.8023	12.872	15.736
Long SDS	Intralipid	2.1343	4.7242	7.9765
	Blood	1.1891	6.8550	14.173

Table 4.2: The correlation coefficient of the regression models for short SDS and long SDS datasets.

		Correlation coefficient		
	Regression model	LOOCV	9-fold/6-fold CV	2-fold CV
Short SDS	Intralipid	0.9999	0.9999	0.9983
	Blood	0.9995	0.9960	0.9682
Long SDS	Intralipid	0.9997	0.9974	0.9916
	Blood	0.9999	0.9941	0.9807

Chapter 5

Discussions

The SNR value of the signal at different wavelengths determine the wavelength range that can be used. In this way, wavelength range of 500 nm to 900 nm is selected for the short SDS probe dataset, 600 nm to 900 nm for the long SDS probe dataset and 500 nm to 1200 nm for the simulation dataset since there are no limitations imposed by the measurement instruments' sensitivities. In addition, the SNR value is used as a guidance in choosing the range which the chromophores can be varied. For instance, the signal taken with a long SDS probe becomes very noisy and loses its significance with higher amounts of blood concentration; hence, the range of 0.05% to 5% is chosen for blood in this dataset.

Figure 4.4 shows a prediction accuracy close to 100% for the blood volume fraction and reduced scattering coefficient regression models built using the simulated dataset and evaluated by 2-fold CV technique. The result suggests the LS-SVM as a very powerful tool, specially in cases where the spectra are obtained under identical conditions free from any sources of error. For the experimental datasets, the models performances vary noticeably depending on the validation technique used, i.e. the number of samples used for training compared to the number of samples used for testing. In this way, higher prediction accuracy is observed for LOOCV, 9-fold CV (/6-fold CV) and 2-fold CV in descending order (Table 4.1). Percentage error of <2% is observed when testing 1 sample in both reduced scattering coefficient and blood volume fraction models. In addition, the reduced scattering coefficient models show a better performance compared to the blood volume fraction models both for the short and long SDS datasets. This is explained by the more unique behavior of the DR spectra for different amounts of intralipid when comparing the spectral intensities of the phantoms with varying intralipid concentration (Figures 4.2b and 4.3b) with phantoms with varying blood concentration (Figures 4.2a and 4.3a).

Comparing the reduced scattering coefficient regression models for the two datasets, a relatively better performance is observed for the short SDS dataset. This can be due to having

a larger wavelength range (500 nm-900 nm compared to 600 nm-900 nm) and higher number of samples (81 to 54) compared to the long SDS dataset. There are prominent spectral features existing in the 500 nm to 600 nm wavelength range that can positively contribute to the model performance. In addition, each particular intralipid concentration is repeated 9 times for different amounts of blood in the short SDS dataset rather than 6 times. On the other hand for the blood volume fraction regression models, a slightly better performance is observed for the longer SDS dataset despite having a smaller wavelength range. Here, having fewer number of samples is not a relevant factor as each particular blood concentration is repeated 9 times for different amounts of intralipid. This higher prediction accuracy might be explained by the fact that the spectral intensities for the phantoms are more separated for the long SDS (Figure 4.3a) compared to the short SDS datasets where only subtle differences are seen in the intensities (Figure 4.2a).

In Barman's study [51], mean percentage error of 0.8% was found for the reduced scattering coefficient model and 3.77% for the absorption coefficient regression model, by applying LOOCV technique to the dataset constructed using 24 DR spectra from liquid phantoms comprising of polystyrene microspheres, ink and water. In Xie' study [32], percentage error of 3% was found for the reduced scattering coefficient model and 4% the blood volume fraction model, by applying 2-fold CV technique to the dataset constructed using 270 DR spectra from liquid phantoms comprising of water, 9 different amounts of PpIX, 6 different amounts of intralipid and 5 different amounts of bovine blood. In both studies the model for the reduced scattering coefficient showed a better performance. The better prediction accuracy in Xie's work can be explained by using a much larger dataset and employing a higher performance detection unit in the DRS experimental setup.

Varying the oxygenation level of a phantom changes the spectral behavior of the DR signal. When the models were tested using an spectra with a different oxygenation level and known chromophore compositions, the regression models failed. This result confirmed that in a LS-SVM evaluation protocol both the training and testing datasets must have the same nature, i.e. obtained in a similar manner.

The robustness of the regression models can be improved by increasing the number of samples, using a larger wavelength range and ensuring that all measurements are carried out in an identical manner (e.g., the precision in pipetting, duration of sonicating, the probe's placement in the liquid phantom, the blood's age and the background noise). The experimental setup used in this study is using rather cost-effective spectrometers. Using spectrometers with higher dynamic range can improve the quality of measurement data. In this way, spectra with higher signal to noise ratio are obtained and a larger wavelength range

can be probed. This is specially true in the case of long SDS probe where the intensity is generally weaker and more prone to noise.

Chapter 6

Conclusions and Outlook

6.1   Conclusions

In summary, this work has encompassed the development and testing of an evaluation protocol that utilizes LS-SVM. Literature study of the current state-of-art in using diffuse reflectance spectroscopy for estimation of tissue chromophores based on regression analysis was presented. Phantom preparation protocols for liquid intralipid blood phantoms were established. Finally, evaluation protocols for the estimation of tissue chromophores and scattering parameters were established using LS-SVM technique. Two regression models were created for the reduced scattering coefficient and the blood volume fraction using three datasets: simulated and experimental, collected using a short and a long SDS probe. The simulated models demonstrated a very high prediction accuracy when evaluated with 2-fold CV method. For the experimental datasets, three cross-validation techniques of LOOCV, 9-fold CV (/6-fold CV) and 2-fold CV were applied to each model and the correlation coefficients between the true and predicted data with range of 0.96 to 0.99 were achieved. The level of percentage error decreases noticeably with using a higher number of samples for training the model. Percentage error of <2% is observed when testing 1 sample in both reduced scattering coefficient and blood volume fraction models. In addition, it was found that the reduced scattering coefficient regression models have a relatively better performance compared to the blood volume fraction regression model in both datasets. The blood volume fraction model for the long SDS gave a slightly better performance as opposed to the short SDS model. In general, the spectral shape of the initial dataset, the number of samples used for training and the robustness of the detection system used affect the performance of the constructed regression model.

6.2   Outlook

The obtained results suggest a promising outlook for in-vivo studies of using LS-SVM technique in order to extract the biological tissue optical properties. The final goal with this study is to create an evaluation protocol using a true prospective dataset from liquid phantoms mimicking the optical characteristics of the biological tissue with an as good as possible agreement. Ways to construct an improved dataset include varying the tissue chromophores with shorter intervals, selecting the variation range boundaries close to real life biological tissue values, using a higher quality blood and taking the measurements using a more robust setup. If the sought-after parameter is not known, regression LS-SVM would not be the appropriate evaluation method since the prediction accuracy will be unknown. Moreover, classification studies for differentiating different types of biological tissues from each other as well as differentiating the healthy from cancerous tissue using DRS can be done. Constructing a dataset with varying oxygenation will be specially valuable for in-vivo studies, since the amount of oxygen in a healthy tissue differs from a cancerous tissue as well as during different stages of therapy.

References

[1] M. Zurawska-Szczepaniak Y. Yang, A. Katz and R. Alfano. Optical spectroscopy of benign and malignant breast tissues. *Lasers Life Science*, 7:115–127, 1996.

[2] S. Naumov S. Pushkarev and S. Vovk. Application of laser fluorescence spectroscopy and diffuse reflection spectroscopy in diagnosing the states of mammary gland tissue. *Opto-electronics, Instrumentation and Data-Processing*, 2:71–76, 1999.

[3] S. Bown I. Bigio and G. Briggs. Diagnosis of breast cancer using elastic-scattering spectroscopy: Preliminary clinical results. *Biomedical Optics*, 5:221–228, 2000.

[4] F. Bevilacqua-N. Shah D. Hsiang J. Butler D. Jakubowski, A. Cerussi and B. Tromberg. Monitoring neoadjuvant chemotherapy in breast cancer using quantitative diffuse optical spectroscopy: a case study. *Journal of Biomedical Optics*, 9:230–238, 2004.

[5] V. Backman-R. Manoharan M. Fitzmaurice J. Van Dam G. Zonios, L. Perelman and M. Feld. Diffuse reflectance spectroscopy of human adenomatous colon polyps in vivo. *Applied Optics*, 38:6628–6637, 1999.

[6] Y. Liu-M. Lucia A. Bokhoven H. Sullivan E. David Crawford P. Maroni F. Kim J. Daily P. Werahera, E. Jasion and F. La Rosa. Diagnosis of high grade prostate cancer using diffuse reflectance spectroscopy. *Journal of urology*, 191:593–593, 2014.

[7] E. Atkinson-A. Malpica M. Follen Y. Mirabal, S. Chang and R. Richards-Kortum. Reflectance spectroscopy for in vivo detection of cervical precancer. *Journal of Biomedical Optics*, 7:587–594, 2002.

[8] H. Klomp-J. van Sandick M. Wouters G. Lucassen B. Hendriks J. Wesseling D. Evers, R. Nachabe and T. Ruers. Diffuse reflectance spectroscopy: A new guidance tool for improvement of biopsy procedures in lung malignancies. *Clinical Lung Cancer*, 13:424–431, 2012.

[9] D. Evers H. Klomp J. van Sandick M. Wouters R. Nachabe G. Lucassen B. Hendriks J. Wesseling Spliethoff, Jarich W.a and T. Ruers. Improved identification of peripheral lung tumors by using diffuse reflectance and fluorescence spectroscopy. *Lung Cancer*, 80:165–171, 2013.

[10] J. Boyer T. Johnson J. Mourant, T. Fuselier and I. Bigio. Predictions and measurements of scattering and absorption over broad wavelength ranges in tissue phantoms. *Applied Optics*, 36:949, 1997.

[11] S. Jacques I. Saidi and F. Tittel. Mie and rayleigh modeling of visible-light scattering in neonatal skin. *Applied Optics*, 34:7410–7418, 1995.

[12] SL. Jacques. Optical properties of biological tissues: a review. *Physics in Medicine and Biology*, 58:5007–8, 2013.

[13] B. Wilson W. Star F. Staveren, S. L. Jacques and M. van Gemert. Optical properties of intralipid: a phantom medium for light propagation studies. *Lasers in Surgery and Medicine*, 12:510–519, 1992.

[14] S. Del Bianco G. Zaccanti and F. Martelli. Measurements of optical properties of high-density media. *Applied Optics*, 42:4023–4030, 2003.

[15] M. Aalders T. van Leeuwen N. Bosschaart, G. Edelman and D. Faber. A literature review and novel theoretical approach on the optical properties of whole blood. *Lasers in Medical Science*, 39:1–27, 2013.

[16] A. Pifferi A. Torricelli RLP Van Veen, H. Sterenborg and R. Cubeddu. Determination of vis-nir absorption coefficients of mammalian fat, with time-and spatially resolved diffuse reflectance and transmission spectroscopy. *Proceedings of Biomedical Topical Meetings*, 2004.

[17] G. Hale and M. Querry. Optical constants of water in the 200-nm to 200-um wavelength region. *Applied Optics*, 12:555–563, 1973.

[18] A. Desjardins R. Nachabe, B. Hendriks and H. Sterenborg. Estimation of lipid and water concentrations in scattering media with diffuse optical spectroscopy from 900 to 1600 nm. *Biomedical Optics*, 15, 2010.

[19] A. Ishimaru. *Wave propagation and scattering in random media.*, volume 1. Academic Press, New York, 1978.

[20] A. Welch and M. Van Gemert. *Linear transport theory.* Addison-Wesley Educational Publishers Inc., US, 1967.

[21] A. Liemert and A. Kienle. Analytical solution of the radiative transfer equation for infinite-space fluence. *Physical Review A: Atomic, Molecular Optical Physics*, 83, 2011.

[22] J. Pontaza and J. Reddy. Least-squares finite element formulations for one-dimensional radiative transfer. *Journal of Quantitative Spectroscopy and Radiative Transfer*, 95:387–406, 2005.

[23] S. Arridge and J. Hebden. Optical imaging in medicine: Ii. modelling and reconstruction. *Physics in Medicine and Biology*, 42:841–853, 1997.

[24] S. Jacques L. Wang and L. Zheng. Monte carlo modeling of light transport in multi-layered tissues. *Computer Methods and Programs in Biomedicine*, 47:131–146, 1995.

[25] I. Babuska B. Szabo. *Finite Element Analysis.* John Wiley Sons, 1991.

[26] L. Wang and H. Wu. *Biomedical Optics: Principles and Imaging.* John Wiley Sons, 2007.

[27] M. Patterson T. Farrell and B. Wilson. A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in-vivo. *Medical Physics*, 19:879–888, 1992.

[28] M. Patterson T. Farrell and B. Wilson. A diffusion theory model of spatially resolved, steady-state diffuse reflectance for the noninvasive determination of tissue optical properties in vivo. *Medical Physics*, 19:879–888, 1992.

[29] H. Ferwerda R. Groenhuis and J. Bosch. Scattering and absorption of turbid materials determined from reflection measurements. 1: theory. *Medical Physics*, 22:2456–2462, 1983.

[30] A. Welch and M. Van Gemert. *Optical-thermal response of laser-irradiated tissue.*, volume 2. Springer, Berlin, 2011.

[31] M. Larsson I. Fredriksson and T. Stromberg. Inverse monte carlo method in a multilayered tissue model for diffuse reflectance spectroscopy. *Journal of Biomedical Optics*, 17(4), 2012.

[32] M. Mousavi N. Bendsoe M. Brydegaard J. Axelsson X. Haiyan, X. Zhiyuan Xie and S. Andersson-Engels. Design and validation of a fiber optic point probe instrument for therapy guidance and monitoring. *Applied Optics*, 11:1–11, 2014.

[33] O. Amar R. Reif and I. Bigio. Analytical model of light reflectance for extraction of the optical properties in small volumes of turbid media. *Applied Optics*, 46:7317–7328, 2007.

[34] Q. Fang A. Brightwell M. Carnohan G. Cottone R. Ross Russel T. Papaioannou, N. W. Preyer, L. R. Jones, and L. Marcu. Effects of fiber-optic probe design and probe-to-target distance on diffuse reflectance measurements of turbid media: an experimental and computational study at 337 nm. *Applied Optics*, 43(14):2846–2860, 2004.

[35] T. Nguyen N. Rajaram and J. Tunnell. Lookup table-based inverse model for determining optical properties of turbid media. *Journal of Biomedical Optics*, 13:050501, 2008.

[36] D. Cuccia A. Durkin T. Erickson, A. Mazhar and J. Tunnell. Lookup-table method for imaging optical properties with structured illumination beyond the diffusion theory regime. *Journal of Biomedical Optics*, 15:036013, 2010.

[37] V. Vapnik. *Statistical Learning Theory*. Wiley, New York, 1998.

[38] T. Breslin F. Xu K. Gilchrist G. Palmer, C. Zhu and N. Ramanujam. Comparison of multiexcitation fluorescence and diffuse reflectance spectroscopy for the diagnosis of breast cancer. *IEEE Transactions on Biomedical Engineering*, 50:1233–1242, 2003.

[39] P. Yuen W. Wei J. Sham P. Shi X. Lin, X. Yuan and J. Qu. Classification of in vivo autofluorescence spectra using support vector machines. *Journal of Biomedical Optics*, 9:180–186, 2004.

[40] N. Ghoshand P. Gupta S. Majumder. Support vector machine for optical diagnosis of cancer. *Journal of Biomedical Optics*, 10:24–34, 2005.

[41] W. Zheng E. Widjaja and Z. Huang. Classification of ent tissue using near-infrared raman spectroscopy and support vector machines. *SPIE Proceedings*, 5862:25–30, 2005.

[42] W. Zheng E. Widjaja and Z. Huang. Classification of colonic tissues using near-infrared raman spectroscopy and support vector machines. *International Journal Of Oncology*, 32:653–662, 2008.

[43] M. Julia-Sape P. Krooshof S. Tortajada J. Robledo W. Melssen E. Fuster-Garca I. Olier G. Postma D. Monleon A. Moreno-Torres J. Pujol A. Candiota M. Martanez-Bisbal J. Suykens L. Buydens B. Celda S. Hufiel C. Arus J. Garcia-Gomez, J. Luts and M. Robles. Multiproject-multicenter evaluation of automatic brain tumor classification by magnetic resonance spectroscopy. *Magnetic Resonance Materials in Physics, Biology and Medicine*, 22:5–18, 2009.

[44] J.a-Gomez J. Vicente S. Tortajada J. Luts D. Dupplaw S. Van Hufiel C. Saez, J. M. Garcia-Gomez and M. Robles. A generic and extensible automatic classification framework applied to brain tumour diagnosis in healthagents. *The Knowledge Engineering Review*, 26:283–301, 2011.

[45] C. Cortes and V. Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

[46] S. Mendelson and A. Smlal. *Advanced Lectures on Machine Learning. Machine Learning Summer School 2002, Canberra, Australia*, volume 2600. Springer, Berlin, 2003.

[47] U. Fayyad and R. Uthurusamy, editors. *First International Conference on Knowledge Discovery and Data Mining*. AAAI, 1995.

[48] A. Smola and V. Vapnik. Support vector regression machines. *Advances in neural information processing systems*, 9:155–161, 1997.

[49] J. Suykens and J. Vandewalle. Least squares support vector machine classifiers. *Neural Processing Letters*, 9:293–300, 1999.

[50] J. Suykens J. Luts, A. Heerschap and S. Van Hufiel. A combined mri and mrsi based multiclass system for brain tumour recognition using ls-svms with class probabilities and feature selection. *Artificial Intelligence in Medicine*, 40:87–102, 2007.

[51] N. Rajaram J. Tunnell R. Dasari I. Barman, N. Dingari and M. Feld. Rapid and accurate determination of tissue optical properties using least-squares support vector machines. *Biomedical Optics Express*, 2:592–599, 2011.

[52] H.L. Fu B. Yu and N. Ramanujam. Instrument independent diffuse reflectance spectroscopy. *Journal of Biomedical Optics*, 16, 2011.

[53] J. Suykens K. De Brabanter and B. De Moor. Nonparametric regression via statlssvm. *Journal of Statistical Software*, 55:1–21, 2013.

Appendix A

Solving the optimization problem

A.1   Nonlinear support vector machines (SVM)

The optimization problem for non-separable data distribution in the SVM is defined in Equation A.1. This optimization problem can be solved using the Lagrange dual problem. In this method first the optimization function is written in it's Lagrangian form $\mathfrak{L}$. Next the conditions for $\mathfrak{L}$'s optimality are examined in order to construct the dual optimization problem (quadratic programming problem) where the optimization is done with respect to only one variable (The Lagrange multiplier).

$$
\begin{aligned}
\min_{\omega,b,\xi} \ \mathfrak{I}(\omega,b,\xi) &= \ \frac{1}{2}\|\omega^2\| + \gamma \sum_{i=1}^{N} \xi_i \\
\text{subject to} \qquad & y_i(\omega^T x_i + b) \geq 1 - \xi_i \\
& \xi_i \geq 0.
\end{aligned}
\tag{A.1}
$$

The Lagrangian of the optimization problem is defined in Equation A.2 with Lagrange multipliers α_i and β_i $(i = 1,...,N)$, where $\alpha_i \geq 0$ due the inequality constraint

$$
\mathfrak{L}(\omega,b,\xi,\alpha,\beta) = \mathfrak{I}(\omega,b,\xi) - \sum_{i=1}^{N} \alpha_i \{y_i(\omega^T \Phi(x_i) + b) - 1 + \xi_i\} - \sum_{i=1}^{N} \beta_i \xi_i.
\tag{A.2}
$$

In order to solve for ω, b and ξ parameters, the partial derivatives of Lagrangian are found and set to zero. In other words, $\mathfrak{L}$ is minimized with respect to ω, b and ξ.

$$
\begin{cases}
\frac{\partial \mathfrak{L}}{\partial \omega} = 0 & \rightarrow \quad \omega = \sum_{i=1}^{N} \alpha_i y_i \Phi(x_i) \\
\frac{\partial \mathfrak{L}}{\partial b} = 0 & \rightarrow \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \\
\frac{\partial \mathfrak{L}}{\partial \xi_i} = 0 & \rightarrow \quad 0 \leq \alpha_i \leq +\infty, \quad i = 1,...,N
\end{cases}
$$

After substituting the above relations back into the Lagrangian Equation A.2 and simplifying it, we obtain

$$
\mathfrak{L}(\omega, b, \xi, \alpha, \beta) = -\frac{1}{2} \sum_{i,j=1}^{N} y_i y_j \langle x_i, x_j \rangle \alpha_i \alpha_j + \sum_{i,j=1}^{N} \alpha_i.
$$

The dual optimization problem can then be constructed, where the original minimization problem with respect to ω, b and ξ parameters in Equation A.1 is rewritten in the form of a maximization problem with respect to the Lagrange multiplier a_i. Here The dot product between two inputs x_i and x_j is replaced with an appropriate Kernel function $K(x_i, x_j)$ using the "Kernel trick".

$$
\begin{aligned}
\max_{\alpha} \ \mathfrak{Q}(\alpha) = \quad & -\frac{1}{2} \sum_{i,j=1}^{N} y_i y_j K(x_i, x_j) \alpha_i \alpha_j + \sum_{i,j=1}^{N} \alpha_i \\
\text{subject to} \quad & \sum_{i,j=1}^{N} \alpha_i y_i = 0 \\
& 0 \leq \alpha_i \leq +\infty, \quad i = 1,...,N.
\end{aligned}
\tag{A.3}
$$

A.2 Least-squares support vector machines (LS-SVM)

The optimization problem in LS-SVM is defined in Equation A.1. In order to solve this problem first the optimization function is written in it's Lagrangian form $\mathfrak{L}$ as in the case of nonlinear SVM. Whereas, in the next step a set of linear equations are constructed instead of defining the dual optimization problem.

$$
\begin{aligned}
\min_{\omega, b, e} \ \mathfrak{I}_{LS}(\omega, b, e) = \quad & \frac{1}{2} \|\omega^2\| + \frac{1}{2} \gamma \sum_{i=1}^{N} e_i^2 \\
\text{subject to} \quad & y_i(\omega^T x_i + b) = 1 - e_i.
\end{aligned}
\tag{A.4}
$$

The Lagrangian for Equation A.4 is constructed in the same way as the nonlinear SVM. The parameter α_i is the Lagrange multipliers that does not necessarily need to be positive due to

the equality constraint

$$\mathcal{L}(\omega, b, e; \alpha) = \mathfrak{J}_{LS}(\omega, b, e) - \sum_{i=1}^{N} \alpha_i \{y_i(\omega^T \Phi(x_i) + b) - 1 + e_i\}. \tag{A.5}$$

Next, the partial derivatives of $\mathcal{L}$ are found and set to zero in order to solve for ω, b, e_i and α_i:

$$\begin{cases} \frac{\partial \mathcal{L}}{\partial \omega} = 0 \quad \rightarrow \quad \omega = \sum_{i=1}^{N} \alpha_i y_i \Phi(x_i) \\ \frac{\partial \mathcal{L}}{\partial b} = 0 \quad \rightarrow \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \\ \frac{\partial \mathcal{L}}{\partial e_i} = 0 \quad \rightarrow \quad \alpha_i = \gamma e_i, \quad i = 1, ..., N \\ \frac{\partial \mathcal{L}}{\partial \alpha_i} = 0 \quad \rightarrow \quad y_i[\omega^T \Phi(x_i) + b] - 1 + e_k = 0, \quad i = 1, ..., N \end{cases}$$

A set of linear equations are constructed as:

$$\left[\begin{array}{ccc|c} I & 0 & 0 & -Z^T \\ 0 & 0 & 0 & -Y^T \\ 0 & 0 & \gamma I & -I \\ \hline Z & Y & I & 0 \end{array}\right] \left[\begin{array}{c} \omega \\ b \\ e \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ 0 \\ 0 \\ \hline \vec{1} \end{array}\right], \tag{A.6}$$

with
$$Z = [\Phi(x_1)^T y_1; ...; \Phi(x_N)^T y_N]$$
$$Y = [y_1; ...; y_N]$$
$$\vec{1} = [1; ...; 1]$$
$$e = [e_1; ...; e_N]$$
$$\alpha = [\alpha_1; ...; \alpha_N].$$

After eliminating ω and e in Equation set A.6, the following is obtained:

$$\left[\begin{array}{c|c} 0 & -Y^T \\ \hline Y & ZZ^T + \gamma^{-1}I \end{array}\right] \left[\begin{array}{c} b \\ \hline \alpha \end{array}\right] = \left[\begin{array}{c} 0 \\ \hline \vec{1} \end{array}\right]. \tag{A.7}$$