MAESTROS II, a multispectral Time-Domain NIRS system for non-invasive placental monitoring

Frédéric Lange[1], Uzair Hakim[1], Jack Highton[1], Niccole Ranaei-Zamani[2], Olayinka Kowobari[2], Sarah Johnson[2], Olivia Newth[2], Subhabrata Mitra[2] and Ilias Tachtsidis[1]

1. *Department of Medical Physics and Biomedical Engineering, University College London, United Kingdom*
2. *Elizabeth Garrett Anderson Institute for Women's Health, University College London Hospital, United Kingdom*

Abstract: *Objective:* To enable bedside, depth-resolved monitoring of placental oxygenation and metabolism, we redesigned the first-generation MAESTROS platform into MAESTROS II, a multispectral time-domain near-infrared spectroscopy (TD-NIRS) system. *Methods:* MAESTROS II combines a 6.5 W supercontinuum laser and acousto-optic tuneable filters with a four-channel detection block built around hybrid photomultiplier tubes and eight-channel TCSPC electronics. Hardware optimisation, including collimated fibre coupling and an inline reference arm, boosted responsivity and narrowed the instrument-response-function width by more than 40%. A two-step analysis pipeline, comprising spectrally constrained two-layer modelling followed by gated Beer–Lambert estimation, yields absolute chromophore concentrations and dynamic changes in superficial and placental compartments. *Results:* Basic Instrument Protocol (BIP) and MEDPHOT testing showed improved light harvesting and a narrower instrument response function (IRF), translating into mean absolute errors below 10% in the retrieval of μ_a and μ_s' in homogeneous media. Monte Carlo simulations across 356 patient-informed anatomies predicted an averaged placental sensitivity of 7.30 ± 12.31% across all anatomies. Two-layer blood phantoms demonstrated separation of superficial and deep oxygenation with the current method. In preliminary clinical use (n = 185 monitoring sessions), a support vector machine (SVM) using both optical features and available metadata achieved 92.1% accuracy for risk prediction and 78% for outcome prediction. A separate model estimated placental depth directly from optical data with mean error 0.42 cm. Restricting analysis to cases with ultrasound or optically measured placental depth below 3.5 cm increased accuracy for both risk and outcome to 93.3% and 78.9%, respectively. *Conclusion:* MAESTROS II improves photon harvest, temporal response, and clinical usability, enabling longitudinal, radiation-free assessment of in-vivo placental physiology. *Significance:* These first results are encouraging and outline a practical path toward early detection of placental insufficiency; ongoing work will tighten absolute accuracy and generalisability through refined phantoms, algorithmic updates, and expanded clinical validation.

INTRODUCTION

The placenta is a highly specialised organ which develops throughout pregnancy and fulfils the role of all major organ systems while they differentiate and mature in the foetus. The primary function of the placenta is to provide support to the developing foetus, through the provision of water, nutrients, gases and to regulate maternal-foetal interactions through hormone production, as well as removing waste materials [1], [2]. Thus, the importance of the placenta in supporting healthy development of the foetus cannot be overstated. Reductions in oxygen and nutrient transfer from the mother to the foetus has been linked with numerous adverse birth outcomes including foetal growth restriction which is the most prevalent cause of perinatal death after premature birth [3]. Placental dysfunction is also the underlying pathophysiology in pre-eclampsia which is a major cause of prematurity and subsequent sequalae [4] .

Despite the important role of placental dysfunction in these conditions, little is known about its aetiology, and its mechanistic link to these conditions. Historically, placental insufficiency has been examined post-birth, with ex-vivo perfusion studies [5], [6], however, ex-vivo studies can be challenging to conduct, and the longitudinal assessment of placental function throughout the gestation period cannot be evaluated.

In-vivo placental function assessment has not been widely used in clinical settings due to methodological limitations in imaging modalities [7]. At the time of writing, the primary imaging modality

to image the placenta, both structurally and functionally, is through ultrasound imaging. Structurally, ultrasounds are often used to assess placental location, cord insertion, and foetal growth. Furthermore, the placental blood flow to and from the foetus can also be monitored using Doppler ultrasound, providing functional information from various foetal anatomical regions including the umbilical cord, middle-cerebral artery and ductus venosus as well as the maternal uterine arteries [7]. However, ultrasound offers a static assessment of fetal wellbeing and is highly resource dependent. Additionally, ultrasound-derived estimated fetal weight has a 10-15% margin of error with variable reproducibility.

The use of MRI to monitor the placenta has not seen wide clinical uptake, however developments in MRI techniques in research setting promise to provide quantitative measurements of placental function. Of particular interest is the ability of MRI to monitor the oxygen usage due to the change in magnetic properties of haemoglobin depending on its oxygenation state, the same underlying principle that fMRI has been used to assess brain function. Dynamic Contrast Enhanced MRI has also been used to monitor maternal perfusion typically using gadolinium as a contrast agent. However, the safety profile of gadolinium in pregnancy is unknown, limiting the wide use of this technique [8]. Promising avenues of MRI applied to the quantification of placental function, include the application of Arterial Spin Labelled, Diffusion-Weighted, and Diffusion-tensor MRI. However, the use of MRI during pregnancy is limited due to the accessibility issues surrounding MRI scanners, the high cost of usage and the discomfort that can be experienced during the scan.

In that context, Near-Infrared Spectroscopy (NIRS) is an appealing modality in order to evaluate the placental oxygenation. Indeed, this modality has been extensively used to non-invasively monitor the oxygenation of multiple tissues such as the brain, breast or muscle [9], [10]. Briefly, NIRS is an optical technology that uses NIR light (650 – 900 nm) to probe the tissue. A light source is placed on the tissue and a detector is positioned a few centimetres away from the source in order to detect the reflected light. Indeed, in the NIR range, light can penetrate deep into the tissue, because the absorption is low and the scattering is the dominant interaction process. Moreover, each absorber has a specific extinction coefficient which allows to quantify them if multiple wavelength are used. One of the main absorbers in the NIR is the haemoglobin in its oxygenated (HbO_2) and deoxygenated (HHb) form. Therefore, by quantifying these 2 species, NIRS can extract information about tissue oxygenation non-invasively, and in real-time.

NIRS have already been used to try to gain insight into placental oxygenation. Kakogawa et al. [11] used a commercial NIRS system (i.e. NIRO-300) to monitor the placental oxygenation in pregnant women with uncomplicated pregnancies, threatened preterm delivery, and intrauterine growth restriction. The authors concluded that higher tissue oxygen saturation (StO_2) suggests decreased placental function and that StO_2 monitoring may be clinically useful for prenatal management regardless of gestational age at measurement. Razem et al. [12] used another commercial system (i.e. Nonin Equanox 7600) during labour and found that the number and velocity of placental deoxygenation episodes were higher in cases with neonatal acidosis, and that NIRS outperformed cardiotocography (CTG) in predicting acidosis. The authors concluded that placental NIRS monitoring is associated with fetal/neonatal acidosis and offers superior predictive value to CTG during labour.

These studies demonstrate potential for placental NIRS, but they used systems not optimised for this application. Indeed, one of the key limitations of NIRS is the limited penetration depth of light into the tissue. Depending on the specific NIRS technology and algorithm used, this penetration depth can change, and the absolute oxygenation values extracted will be highly dependent on these parameters. Both instruments above are continuous-wave NIRS (CW-NIRS), the simplest form in which continuous light is emitted and attenuated light is detected at centimetre-scale separations. In CW-NIRS the source–detector distance strongly influences depth sensitivity. Adapting CW-NIRS for placental monitoring can leverage this, and wearable devices have been piloted. For example, Nguyen et al [13] developed a wearable, transabdominal CW-NIRS device and demonstrated feasibility in 12 women with anterior placentas, reporting lower placental oxygenation in pregnancies with maternal complications. Interestingly, this contrasts with Kakogawa et al. [11], underscoring the need to standardise measurements and obtain reproducible, instrument-independent oxygenation values.

More advanced forms of NIRS may address these limitations. Wang et al [14] have developed a system combining frequency-domain NIRS (FD-NIRS) with ultrasound and reconstruction algorithms that integrate ultrasound-derived tissue-layer morphology with optical haemodynamic. FD-NIRS uses intensity-modulated NIR light and measures both amplitude and phase of the detected light; the phase information enables simultaneous estimation of absorption and scattering, unlike CW-NIRS which relies on a priori scattering assumptions. Moreover, the specific algorithm developed in this work, based

multilayer fitting, also increases the reliability of the measurements. Indeed, standard NIRS system relies on assuming an homogeneous tissue. This is not the case in placental monitoring where the light penetrates multiple layers of tissue (skin, adipose, muscle, placenta). Therefore, using a layered model is likely to increase the reliability of the measurement. Using this system in a third-trimester pilot study of normal and abnormal placentas, the authors reported that baseline haemoglobin metrics did not differ, but that haemodynamic responses to maternal hyperoxia were significantly different, suggesting utility for early detection of placenta-related adverse outcomes.

The variability of reported oxygenation results highlights the importance of establishing reference measurements for placental oxygenation to enable fair evaluation of new instruments. To that end, Time-domain NIRS (TD-NIRS), would provide the best chances of establishing this standard. Indeed, TD-NIRS displays the greatest deep tissue sensitivity. This comes from the fact that the time-of-flight dimensions inherently encodes depth in the measurements [10]. Indeed TD-NIRS uses picosecond pulses of NIR light into tissue and records the histogram of photon arrival times at a detector. Moreover, fitting these time-of-flight data with light transport models allows to recover the absorption and scattering information about the tissue. The use of TD-NIRS is thus the most appealing in order to get reliable information of tissue oxygenation in depth.

However, being able to achieve recording depth above 15 mm is extremely challenging [15]. Indeed, the ability to record information about deep tissues is limited by a combination of factors between the actual optical properties of the tissues, the instruments, and the algorithm used to extract the information [16]. For example, Mora et al [15] showed that the upper limit of the detection of a change in absorption generated by a $1cm^3$ fully absorbing material embedded in a semi-infinite medium could be up to 6 cm using an ideal system. However, this capability largely depends on the characteristic of the instruments, with the main parameters at play being (i) the responsivity of the detector, (ii) the source power, (iii) the shape of the instrument response function (IRF), and (iv) the collection time allowed for the recording. A level of acceptable noise also needs to be set in order to fix a threshold to get acceptable signals.

Given this information, the first step to consider in order to be able to develop a TD-NIRS system optimized for deep tissue monitoring is to optimize these parameters in an instrument. To do this, one needs to maximise the available power to be able to deliver to the tissue, within the limits of the maximum permitted exposure (MPE) [17]. Then, the responsivity of the detection scheme of the instrument needs to be maximized. To do so, the main components to look at are the photodetector used (i.e., photomultiplier tube (PMT), Silicon photomultiplier (SiPM), etc.) and the interface between the tissue and the photosensitive area. Most of the systems are fibres based, and some optical components needs to be arranged in order to minimize photon losses at the various interface between the detector plan and the tissues. A recent strategy developed is to put the detector directly in contact with the tissues to avoid any extra optical component and maximizing detection efficient, by using advantages of the SiPM detectors [18]. However, when a high number of wavelengths is needed, the application of this method can reveal challenging. Therefore, the design of the detection scheme of the instrument needs to be design with other constraints than pure responsivity.

The photon counting electronic is also a key factor in the capacity of the system to harvest useful photons. Indeed, in the recent years, the sustained count rates of time counting electronics has drastically [19]. This allows to increase the number of detected photons, without 'blinding' the system.

Finally, the shape of the IRF also plays a major role in order to avoid rejection of early photons to later times, reducing the depth specificity. For example, the use of hybrid PMT with reduced afterpulsing allows acquisition of TD-NIRS signals over a larger dynamic and temporal range [16] . As the shape of the IRF will be influence by all of the components of the instruments (i.e. source, detectors, photons counting electronics, optical interfaces), its optimisation will have to be a trade-off between the optimal characteristics of all the other components of the system, in order to get a system that allows good responsivity, and narrow IRF.

In this paper, we will report the hardware upgrades of a TD-NIRS system named MAESTROS, that had been developed for brain monitoring applications [20]. This instrument had been developed with now outdated technologies. Here we will report the redesign of the system, which uses new components, and their effects on the improvements of the key characteristics of a TD-NIRS system. This evolution enables (i) better responsivity, lower crosstalk and improve depth selectivity, and (ii) to improve the overall ease of use in the clinic, via a redesign of the mechanical hardware, and addition of a UPS.

This paper will be structured as followed. Firstly, we will report on the redesign of the system and its description. Then we will describe the algorithm developed in order to enhance the sensitivity to the

placenta of our measurements. Secondly, we will report the precise characterisation of the system, both from the point of view of basic characterisation of TD-NIRS systems, using well known protocols like the BIP [21] and the Medphot protocols [22], and of the specific characterisation of the sensitivity to deep tissues in the context of an inhomogeneous tissues, via Monte-Carlo simulations of the light transport in tissues, and phantoms measurements on 2-layer blood phantoms. Finally, we will present preliminary clinical use and a machine-learning proof-of-concept that classifies risk and outcome from TD-NIRS-derived physiological features.

System description

Overview

Fig. 1 shows the completed upgraded system, MAESTROS II (Metabolism and hAemoglobin Evaluation via a Spectroscopic Time Resolved Optical System II) and schematic of the components. The general principle of this instrument is the same as the first version where a supercontinuum laser produces broadband light which passes through a dual acousto-optic tuneable filter (AOTF). The AOTFs allow the wavelength selection in the NIR range. The two spectrally filtered beams are then injected into two optical fibres that takes the light to the tissues. It is worth noting that only one source point is used here. The reflected light is then collected by up to four optical fibres, separated from the source by a few centimetres. The collected light is then detected by the detection scheme. It first passes through variable optical attenuators (VOAs), in order to adjust the signal dynamics before reaching the Hybrid photomultiplier tube (H-PMT). The four H-PMT outputs are connected to a multichannel event timer and Time-Correlated Single Photon Counting (TCSPC) device (MultiHarp 150, picoQuant). This optical scheme allows us to measure the four different distributions of arrival time of photons (DTOF) from each of the four H-PMTs, simultaneously. We will bellow briefly describe the emission stage (that is mainly unchanged) and detection stage, that has undergone significant redesign. Moreover, we will detail the addition of a reference arm, the upgraded hardware and software, and the design of the probe to interface with the abdomen, and of an IRF module enabling the measurement of IRF without altering the probe.

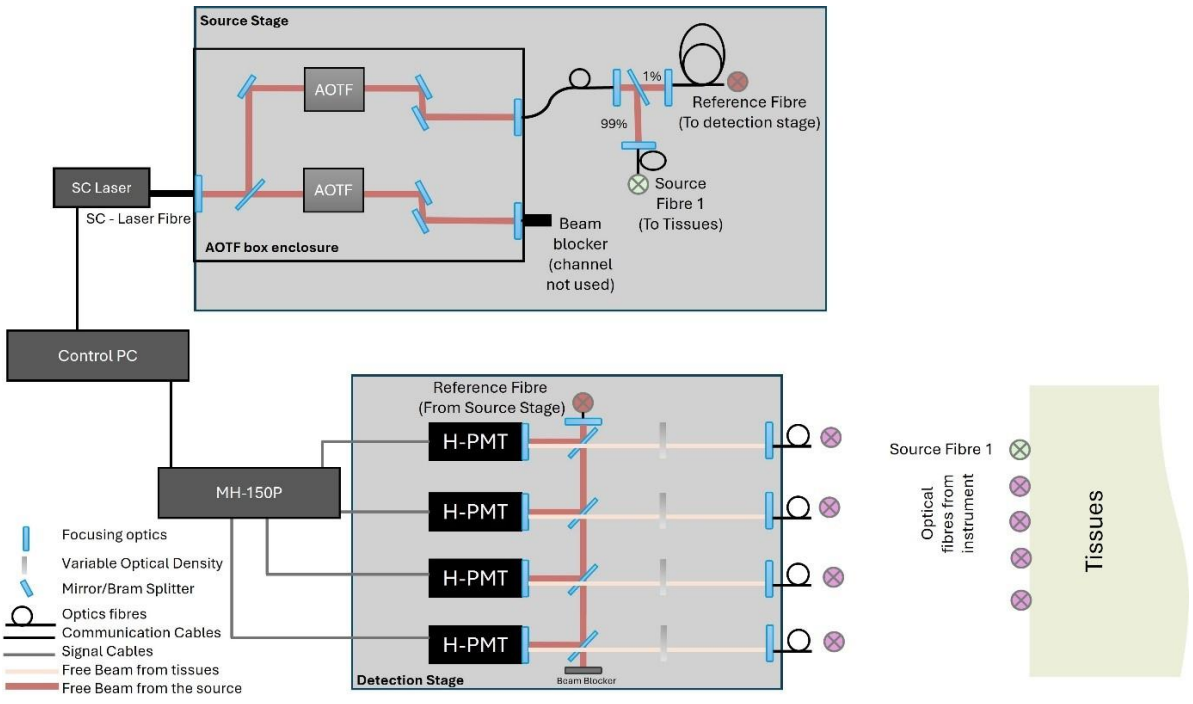


Fig. 1. Schematic illustration of the major system components.

Emission stage

The emission stage principle is the same as the one of the first version, based on supercontinum

laser coupled with 2 AOTF, allowing up to 2 source points on the tissue. Briefly, after passing through the two crystals, the two spectrally filtered beams are then injected into two optical fibres (see section VI - probes and IRF module section for technical details about the fibre used). We used the same AOTF assembly and only replaced the laser due to a fault on the previous one. The new laser is a supercontinuum FIR-20 (NKT, UK). It has a total power of 6.5 W and a spectral bandwidth ranging from 475 to 2300 nm. The laser produces pulses at a fixed repetition rate of 78 MHz.

Detection stage

The diffused light from the sample is then collected by four optical fibres (see section VI - probes and IRF module section for technical details about the fibre used). The collected light first passes through variable optical attenuators (VOAs). This VOAs have been modified compared to the first version of MAESTROS. The original version of the attenuation scheme could only change between 12 steps with predefined Optical Density (OD) levels from 3.5 to 0. This was done to accommodate various light levels, but in practice, when measuring on tissues, only 4 steps were used (between 2 and 0 OD). The steep changes between OD levels meant that the light level could be sub-optimum. Changes have been made to use continuously variable OD (NDC-25C-2 -Thorlabs) between 0 and 2 ODs coupled with a piezoelectric motor (ELL14K - Thorlabs) with minimum rotation steps of 0.002°. This resulted in the possibility of using up to 135000 steps, allowing a finer adjustment of light levels. The new motor is also more compact, reducing the footprint of this feature. Figure 2 displays the original and upgraded optical density range of MAESTROS.

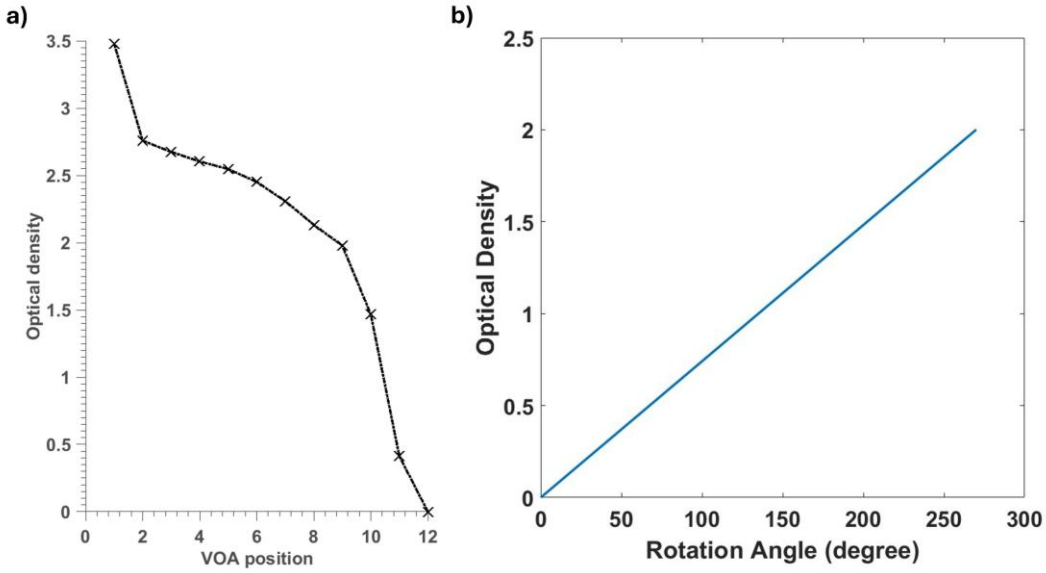


Fig. 2. Optical density range . a) MAESTROS I version. b) MAESTROS II version

The PMTs in the original system have been replaced by Hybrid PMTs (H-PMT, PMA Hybrid-50P picoquant). These detectors present more or less the same form factor and responsivity. The main difference is that H-PMT does not present after-pulsing. Indeed, an ideal single-photon detector generates one and only one electric pulse for every incident photon. However, in a real detector, it is possible that a single incident photon results in more than one electrical pulse per incident photon. This is known as after-pulsing. It is known that PMTs generate after-pulsing with a magnitude of about 1% of the main peak. This does not happen with H-PMT (see system's characterisation for comparison). In practice, after-pulsing causes a reduction of depth sensitivity. Indeed, as photons arriving at an early time are also producing signals at longer times, this will overlap with the photon that does only arrive at later times (i.e. emerging from the deeper tissues). This will create crosstalk between early and late photons, thus reducing the depth sensitivity. The use of H-PMT eliminates this problem, improving depth sensitivity.

Moreover, we added light collimator to couple the light from the detection fibres attached to the tissue, to the H-PMT. The original implementation of the system did not include any collimators to focus the light from the optics fibres to the detector. Thus, even if the fibre tip was positioned close to the detector, the absence of collimation resulted in some light loss between the fibre and the detector. Reflected

collimators (RC02SMA-P01-Thorlabs) are used to collimate the light from the fibre and mounted achromatic doublets (AC254-030-B-ML) are mounted to the H-PMT entrance to focus the light onto them.

Finally, the photon counting electronic was changed. As explained previously, the previous detection scheme was composed of a TCSPC card and a router, to be able to use a single card. The improvement in photon counting electronic has drastically reduced the cost of this element and enabled us to now have a unit with 8 independent channels (MultiHarp 150, picoQuant) with comparable characteristics. The main advantage of this upgrade is the suppression of any crosstalk between the channels, and the increase of the maximum possible photon counts. Indeed, these cards can typically accommodate a rate of about 10Mhz. The fact that this maximum rate is not shared between 4 channels can give us at least four times an increase in the maximum photon count rate.

Reference arm

A reference channel has also been added to the system. Indeed, in a clinical environment, the stability of the system can be at risk, as the environment is less controlled than in the laboratory. To monitor the system stability, (which will mainly vary with temperature changes), a reference channel has been added to the system. It monitors any fluctuation in the signal of the system independently from the signal coming from the tissue. To do so, a fraction of the light from the source (1%) is directly injected in the detection stage. This allows monitoring of any fluctuation of the temporal distribution of the photons due to the system itself. It is worth mentioning that optical fibres of different lengths are used in order to avoid any overlap between the light from the reference channel, and the light coming from the tissues. The 1% of the source light is collected after the AOTF. The output light of the AOTF is connected to a 50cm single core grade-index fibre (Ø62.5 µm Core, 0.27 NA, GIF625 - Thorlabs) and directed onto a broadband dichroic mirror (BB1-E03 -Thorlabs) that reflects 99% of the light. The reflected light is then injected onto the source fibre directed onto the tissue via a collimator (RC08SMA-P01 -Thorlabs), and the transmitted light (<1%) is injected into a 6m single core grade-index fibre (Ø62.5 µm Core, 0.27 NA, GIF625 - Thorlabs) via a collimator (RC08SMA-P01 - Thorlabs), to be direct onto the H-PMT. In order to merge the light coming from the tissue and the one from the reference arm, four D-Shape mirrors (BBD1-E03 – Thorlabs) are used. The light coming from the AOTF is collimated using an achromatic doublet (AC127-025-B-ML - Thorlabs) and the 4 D-Shape Mirror are slightly vertically shifted in order to let enough light through to the next stage (see figure 1). The light coming from the tissue passes above the D-Shape mirror, enabling to merge the 2 beams without attenuation from the reference arm.

Mechanical hardware and power supply

As for version one, all of the elements of the system are mounted into a 25U 19” rack mounted on castors, to make it transportable. The final dimensions of the system are 95 x 60 x 130 cm. A 1000 VA Reo medical isolation transformer which complies with the EN60601 standard (3rd edition) is used to achieve a low general earth leakage current to protect the system. Additionally, a 2U Uninterrupted Power Supply (UPS, APC Smart-UPS 1500VA) has been added to the system. This addition facilitates the clinical use by allowing the system to be powered and transported to other clinical rooms. This way, the system does not need to be warmed-up in the same room as it will be used, which can be a challenge in demanding clinical environment. It also enables to move it between clinical rooms without being turned off, facilitating greatly its use.

Probes and IRF module

Two sets of probes were developed. In the following of the paper, the use of these probes will be referred to as MAESTROS V.2.0 and MAESTROS V.2.1 for probe V1 and V2 respectively. The first version was a prototype to test the usability in the clinic and accommodate for the timeline of the project regarding clinical data collection. This was used in the first stage of the project. The second version was developed in partnership with pioNIRS, to refine the design and be able to use a IRF module enabling the collection of the IRF without undoing the fibres from the holder, which drastically improve the simplicity of use of the system in the clinical environment. This is currently the version used by the system.

Probe V1 (MAESTROS V2.0): A 3D-printed flexible probe hosts both injection and detection fibres. Optical fibres are coupled to the tissue by means of a prism, which bends light of 90° (Thorlabs -ref),

thus making it possible to keep the fibres parallel to the probed surface, for improved contact during movement of the probed tissue. The fibres used for the source is a single core grade-index fibre (Ø62.5 µm Core, 0.27 NA, GIF625 - Thorlabs). The fibres used to collect the backscattered light are single core multimode optical fibre (0.39 NA, Ø1000 µm, FT1000UMT – Thorlabs). The distance between injection and collection points on the sample surface are 2, 3,4 and 5 cm and the fibres length used are all 3 m. It is worth noting that these fibres were used due to their availability and time constraint of the project. Moreover, the length of 3m was used as, at the time of the redesign, some uncertainties were still present in terms of operation of the system. Indeed, the system being used in a clinical environment, longer fibres were used as the space restriction were still unclear, even though longer fibres degrade the performance of TD-NIRS, as it causes a broader IRF [23] .

Probe V2 (MAESTROS V2.1): Here the same configuration was used but the probe holder design was refined, to be more compact. Also, the optical fibres were changed to improve the optical performances of the system. The source fibres used is a 100um core glass, and the detector fibres are 0.9um core POF fibres. Also, the fibres length was reduced to 1.5m. Indeed, after a few months of operation, it was clear that shorter fibres could be used, which can improve the performances of the system. The new set of fibres are also more flexible than the previous set, which facilitated its use by the clinicians. The IRF module consist of a black box with white 2 opening for the source and the detectors. The detector opening is cover with white Teflon to allow the excite all the mode of the fibre to record accurate IRF. A reflective surface coats the bottom of the box for the light to be reflected to the detectors. A 3D rendering of the ProbeV2 and of the IRF box is presented in figure 3 below.

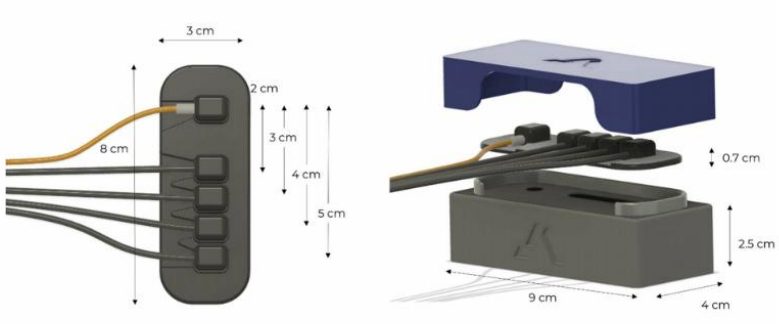


Fig. 3. Left - 3D rendering of the ProbeV2. Right – Rendering of the Probe V2 in the IRF Box.

Software

The software controlling the system is written in LabVIEW (National Instruments). An all-in-one user friendly VI is used to control (1) the MH 150 parameters, (2) the AOTF wavelengths used, (3) the VOA level, and (4) the data collection, by setting information of the subject. The data processing workflow is performed offline and detailed in the next section.

Algorithm for extraction of placental parameters

To analyse the data acquired with the MAESTROS II system, we designed an algorithm to optimize the placental sensitivity of the measurement. This algorithm is using a 2-step process including (i) the extraction of the absolute information regarding the tissue composition (for the superficial and the deep layer) and (ii) the assessment of the changes in oxygenation and metabolism parameters during each recording sessions. The following sections will detail these 2 steps. All the processing was done with Matlab R2022b.

Extraction of absolute information of the tissue composition

The extraction of the absolute information of the tissue composition is done via an optimisation procedure looking at minimising the differences between the average distribution of time-of-flight (DTOF) recorded in each session and a theoretical model, using the ‘fmincon’ function in Matlab.

The averaging was done to increase the SNR of the signal and provides information about the tissue composition which won't change during the session. Then, the variation of the parameters that can change during the session (i.e., tissue saturation (StO2) and [oxCCO]), are calculated for each time point in the second step (describe in the following section). This 2-step process is common in TD-NIRS as it is assumed that the tissue composition will remains stable over the course of the recording and only parameters related to oxygenation and metabolism will change. Thus, they can be extracted in a second step by looking at the changes in the signals [24].

For the extraction of the tissue composition, the parameters of the models are extracted by using a spectrally constraint method [25] fitting the data to a 2-layer model [26], [27] after convolution with the IRF. Indeed, the spectrally constraint method has been shown to be more robust to noise [25] , and as our system allows the recording of 16 wavelength, we wanted to take advantage of this. Moreover, as the tissue that we are targeting cannot be assumed to be homogeneous, using a 2-layer model can help to model it more accurately.

The parameters of the fit are the [HHb] and [HbO2] concentrations, the fat (set to zero for the second layer) and water content fraction, and the scattering power and amplitude for the two layers. This fit is performed on the average DTOF acquired during each session for the detector at 4cm from 80% of the rising edge, to 0.1% of the falling edge of the DTOF. The thickness of each layer is known from the ultrasound imaging, and we consider the first layer to be: skin + adipose tissue + muscle, and the second layer to be placenta (infinite medium).

Estimation of the changes in concentrations during the recording session

The changes in the chromophore's concertation ([HbO2], [HHb], [oxCCO]) is estimated using the microscopic, modified Beer-Lambert law using 2 gates [28] . Briefly, by gating the signal, i.e., summing all the photon arrived at certain times, we can select the depth sensitivity of the signals, photon arriving at later times having a greater probability of reaching deep tissues. The superficial tissues were selected using an early gate (50% to 95% of the rising edge of the DTOF) and the detector at 3cm to reenforce the selectivity of superficial tissue (as there is a depth dependency on source/detector distances). The deep tissues were selected using the same principle using the detector at 4 cm and a late gate (70% to 1% of the falling edge of the DTOF).

Finally, to get the variation in StO2 values, the changes in [HHb] and [HbO_2] are added to the absolute value of hemoglobins retrieved in the multilayer fit.

Performance assessment

Basic Phantom Protocols

BIP

The basic Instrument protocol (BIP) enables to evaluate the basic characteristics of an instrument. It is especially useful to quantify the improvements between several instrument iterations. The full description of the protocol can be found in [29].

In this paper, we will focus on two metrics that are key and expected to be enhanced by the upgraded component: (1) the responsivity and (2) the instrument response function (IRF) characteristics. The responsivity is characterised by the capacity of the detection chain to harvest photons. The measurement of the responsivity is done by shining a collimated light on a calibrated phantom. Then, by measuring the amount of transmitted light detected by the instrument, and by knowing the input light intensity, the capacity of the instrument to collect light is calculated and gives an absolute number in m2/Sr (see [18] for details).

The responsivity as function as wavelength of the initial and upgraded instrument is presented in figure 4.

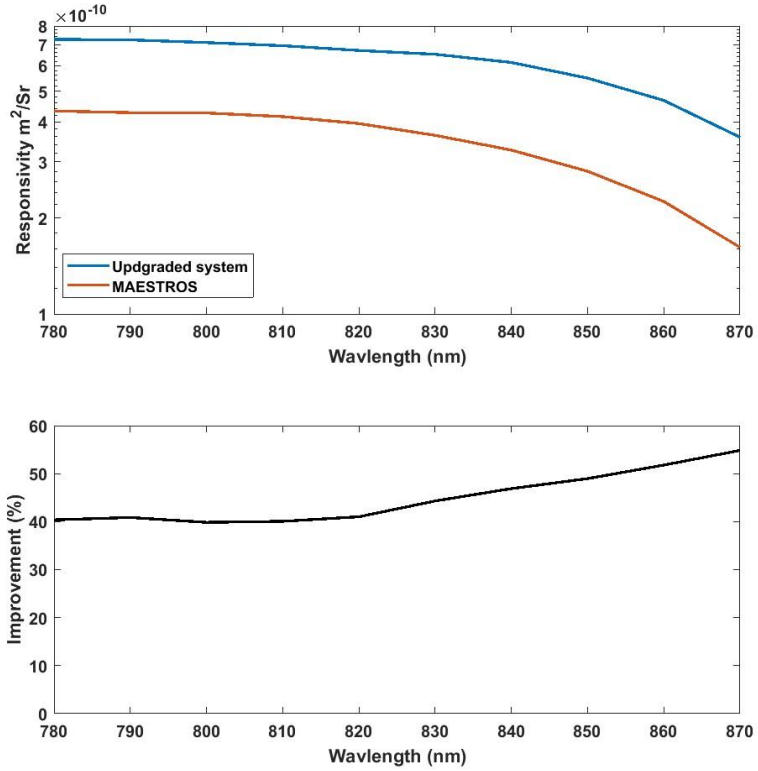


Fig. 4. (Top) Responsivity of MAESTROS and the upgraded instrument. (Bottom) Percentage of improvement with the upgraded system.

We can see that the upgraded instrument can harvest on average (over the entire wavelength range) 45% more light than the original MAESTROS system (See table 1). This improvement is likely due to the addition of the light collimation that have reduced light losses between the fibres and the detectors. It is also interesting to see that the gain is higher for longer wavelength. This is especially useful as we will receive less photon from the tissue at these wavelengths, so this boost will improve the SNR of our measurement in-vivo.

Table 1 Responsivity of systems for 3 selected wavelengths

Wavelength (nm)	Responsivity MAESTROS V1.0 (m^2/Sr)	Responsivity MAESTROS V 2.0 (m^2/Sr)	Improvement (%)
780	4.33e-10	7.26e-10	40.36
830	3.63e-10	6.52e-10	44.31
870	1.62e-10	3.58e-10	54.84

The IRF is one of the main characteristics of a TD-NIRS system. It governs the 'temporal blurring' of the DTOF which will affect the depth selectivity of the system. An ideal TD-NIRS system would have a very narrow IRF with no after-pulsing.

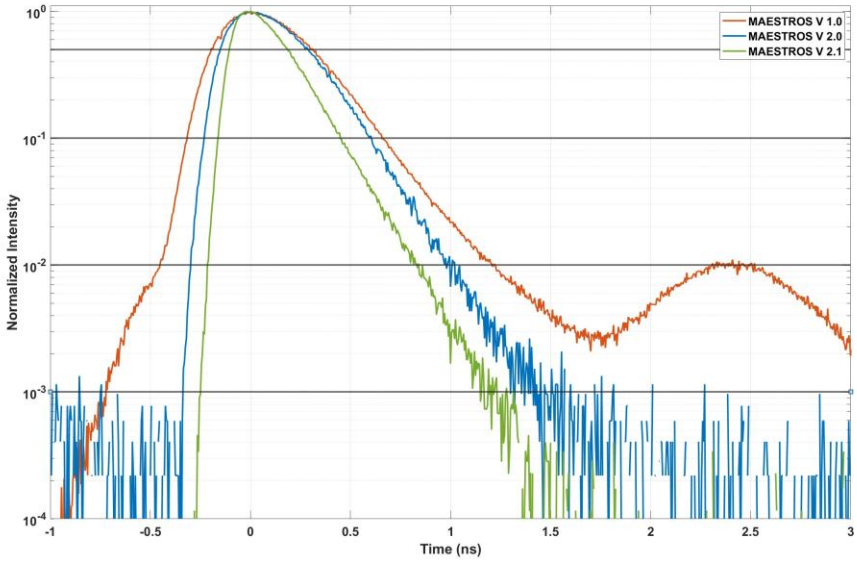


Fig. 5. IRF characteristic of MAESTROS V.1.0, V2.0 and V2.1. and the upgraded system. Note the sharper IRF (no after-pulsing typical of PMTs)

Figure 5 presents the typical IRF of the original MAESTROS system, and of the upgraded instrument. One can easily note that the after-pulsing have been removed with the upgraded detectors, which massively improved the temporal characteristics of the IRF. To quantify the improvement, the width at 50%,10% 1% an 0.1% of the maximum are compared before and after the upgrade. These values are reported in table 2. We can see a significant reduction in the widths at all thresholds, highlighting the narrower IRF of the upgraded system.

Table 2.IRF temporal characteristics of the original and upgraded versions of MAESTROS

Width	**V.1.0**	**V.2.0**		**V.2.1**	
		Width	Improvement	Width	Improvement
50%	496.35ps	439.69ps	11.41%	292.00ps	41.17%
10%	974.59ps	837.11ps	14.11%	612.00ps	37.20%
1%	1659.1ps	1277.5ps	23.00%	1046.7ps	36.91%
0.1%	3885.2ps	1637.7ps	57.85%	1398.7ps	64.00%

Medphot Protocol

The MEDPHOT protocol has been designed to test the accuracy of the measurement of the optical properties of tissues with TD-NIRS. It consists of a set of solid phantoms with different optical properties (i.e., absorption properties: μa, and reduced scattering properties: μ's). Details about this protocol can be found in reference [22].

In order to measure the optical properties of a phantom (or tissue), one needs to fit the measured DTOF to a model of the light transport in tissues. Indeed, the shape of the DTOF will be mainly dictated by the absorption and scattering properties of the tissues. An example of the DTOF measured on two phantoms is presented in figure 6. More information can be found in the literature about this topic 10. The standard way in the MEDPHOT protocol is to fit the DTOF to the solution of the diffusion equation for a semi-infinite media.

We validated the upgraded systems V.2.0 and V.2.1 using the same set of phantoms (BioPixS matrix, BioPixS Ltd, Ireland). This set of phantoms was not available during the characterization of the first version of MAESTROS (V.1.0), and therefore a direct comparison could not be performed. However, these measurements provide a robust validation of the accuracy of the newly developed systems and allow assessment of the effect of the probes.

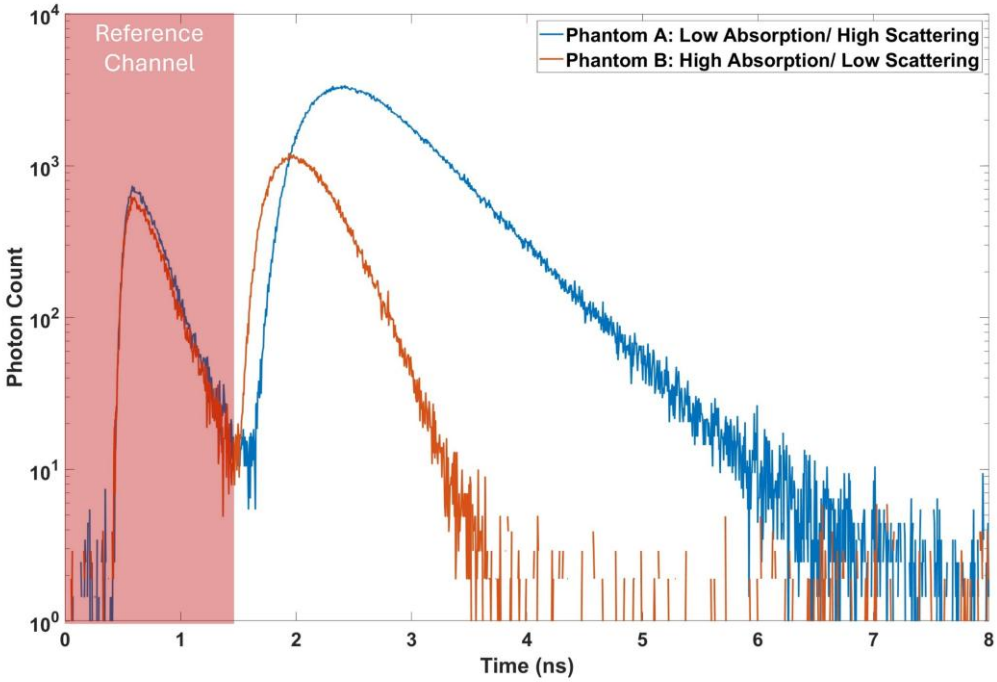


Fig. 6. Example of DTOF for 2 phantoms: (red) high absorption, low scattering, (blue) low absorption, high scattering. One can see the effect of the absorption (change of the slope of the falling edge of the DTOF), and scattering (change in the peak time of the DTOF) on the DTOF. V2.0 used here.

In figure 9, we report the retrieved optical properties of 9 phantoms measured with the MAESTROS v2.1 . The Phantom ID are A, B, C for scattering values (5, 9 and 13 cm^{-1} respectively), and 2,3,5 for absorption values (0.05, 0.10,and 0.21 cm^{-1} respectively). The measurements were repeated 20 times per phantom. The mean values and standard deviation of the results are reported with error bars. We can see that the linearity of the absorption and scattering property retrieve is preserved, in agreement with our previous results [30] .

In Table 3, we report the accuracy of optical property estimation for the two upgraded configurations, MAESTROS V.2.0 and V.2.1, expressed as the mean deviation from reference values at 828 nm. The results show that both μ_a and μ_s' retrievals are improved in V.2.1 compared to V.2.0, with mean errors reduced from 14.61% to 9.05% for μ_a (38.06% improvement) and from 11.40% to 7.55% for μ_s' (33.77% improvement). These values are within the previously reported range for other instruments [30].

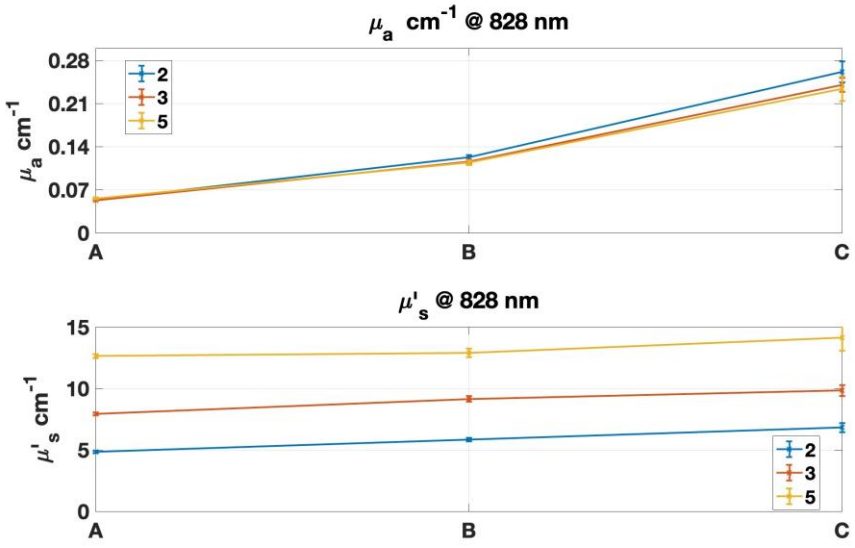


Fig. 7. Results of the MEDPHOT protocol for the MAESTROS V2.1 upgraded system.

Table 3.Accuracy of the optical properties' estimation for both MAESTROS upgrade configurations (V2.0 and V2.1).

	mean deviation from reference values at 828 nm		
	V.2.0	**V.2.1**	**Improvement**
μ_a	14.61%	9.05%	38.06%
μ'_s	11.40%	7.55.%	33.77%

Validation of system's sensitivity to placenta

Simulations

Computational assessment of the placental sensitivity to NIR light was evaluated using the Monte Carlo eXtreme (MCX) [31], [32]. All simulations were conducted on a local desktop running Ubuntu, using two Nvidia GTX 4090. For each simulation $15x10^9$ photons were used and released into a 4-layer square volume, not including air. The top layer was skin, followed by adipose, muscle and finally placenta. Each voxel of the volume was $1mm^3$ and each volume was 100 voxels in each dimension. The source was placed at the middle-point of the x-y plane of the top layer of the volume. Mimicking the MAESTROS system, four detectors were placed at 20, 30, 40 and 50mm away from the source. Each detector had a radius of 3mm, to ensure adequate SNR for further analysis. The duration of each simulation was set to 1e-8 seconds, with intermediary time-steps of 2e-11s. The time-step duration was chosen based on the requirement for adequate GPU memory and to approximate the time-gating of the MAESTROS system.

343 simulations were conducted. Each simulation utilised real-world ultrasound patient data to model the thickness of each layer. In particular, the skin, adipose and muscle thicknesses were modelled based on patient data, with the remainder of the volume being labelled as placenta. The total haemoglobin content in the muscle and placenta layers were fixed to 50μMol and the StO_2 to 60 and 80% for the muscle and placenta layers respectively.

Optical properties of the skin layer were computed using the fraction of melanosome associated to each real patient data. As these were only used to calculate the sensitivities and for computational efficiency, we computed the optical properties for one wavelength, 780nm. These were determined using the Fitzpatrick scale, where skin is given a number from 1-6 depending on light/darkness, with 1 being light. From these values, an associated fraction of melanosome was ascribed based on [33]. The volume fraction of melanosome relates how much melanosome is present for a skin type, shown in table 4, with 6 unique values.

Table 4. Fitzpatrick score link to volume fraction of melanosome

Fitzpatrick Score	Volume Fraction of Melanosome
1	0.0255
2	0.0255
3	0.0902
4	0.155
5	0.155
6	0.3055

The absorption coefficient of the skin layer is then computed as the absorption coefficient of the dermis, summed with the epidermis. The absorption coefficient for the dermis is computed as [34]:

$$(1 - F_b)\mu_{\{a,skinbaseline\}} + \log(10) \cdot HbT_{\{dermis\}} \cdot \left(StO2_{\{dermis\}} \cdot ext_{\{HbO2\}} + \left(1 - StO2_{\{dermis\}}\right) \cdot ext_{\{Hb\}} \right)$$

Where F_b is the volume fraction of blood in the dermis, $HbT_{\{dermis\}}$ is the total haemoglobin in the dermis layer of the skin, $StO2_{\{dermis\}}$ is the oxygen saturation in the dermis layer, all parameters obtained from [34], except from $\mu_{\{a,skinbaseline\}}$ which is taken from reference [35].

The computation of the epidermis absorption coefficient follows the same method as [34], following the formula:

$$F_{mel} * \mu_{\{a,mel\}} + (1 - F_{mel}) * \mu_{\{a,skinbaseline\}}$$

Where F_{mel} is the volume fraction of melanosome as described above, and the absorption coefficient of melanin is obtained from [34] and described as:

$$519 \left(\frac{\lambda}{500}\right)^{-3}$$

Following the computation of the absorption coefficients of the dermis and epidermis, these are combined to obtain the absorption coefficient of the skin. This was done with the consideration of the fractional contribution of each layer to the total skin. Using the formula from [36] the fraction of dermis and epidermis was computed based on the thicknesses of the dermis and epidermis, taken as 4550 micron and 127.2 micron respectively. Thus the final absorption coefficient for the skin is computed as:

$$F_{\{dermis\},\mu_{\{a,dermis\}}} + F_{\{epidermis\},\mu_{\{a,epidermis\}}}$$

Based on these calculations, the absorption coefficient of the skin layer is dependent on the volume fraction of melanosome, therefore for these simulations 4 different absorption coefficients were used, based on the specific Fitzpatrick score for each patient.

Table 5. Fitzpatrick score relation with absorption coefficient

Fitzpatrick score	Skin $\mu_{a\ (mm^{-1})}$
1-2	0.0406
3	0.0646
4-5	0.0887
6	0.1445

The remaining optical properties were constant for each Fitzpatrick score and are listed below.

Table 6. Optical properties used in the simulation

	Skin	Adipose	Muscle	Placenta
g	0.83	0.9	0.9	0.9
n	1.47	1.37	1.37	1.4
$\mu_{s\ (mm^{-1})}$	14.384	13.647	8.6121	8.8147
$\mu_{a\ (mm^{-1})}$	See table 5	0.0022	0.0119	0.113

Sensitivity Profile

Sensitivity profiles for each simulation were computed using the adjoint method [31]. To do so simulations for all patients are first acquired with source and detectors in their original positions. The

output of this is a volumetric flux distribution with units 1/mm^{2}s, for each time-gate. To convert these to a volumetric fluence distribution (1/mm^2), the volumetric flux at each time-gate is multiplied by the width of the time-step, for our simulations this was 2e-11.The simulations are then re-computed with the source placed at the position of each detector and the conversion from flux to fluence computed for each simulation.

Finally, the volumetric fluence output of the source to each detector is then convolved with the volumetric fluence output of the source placed at that detector location for each time-gate. The output of this is a volumetric Jacobian (Δfluence/Δμa), reflecting sensitivity of a given source-detector pair for a given time-gate. This process is repeated for each detector. For each simulation sensitivity profiles are obtained for each of the four detectors, and at each time-gate. Time-bins with 0.2ns widths were used to assess the sensitivity. First, the 1% of the rising-edge and falling-edge were calculated for each TPSF. These served as the bounds within which time-bins would be created. Using these points, the number of 0.2ns time-bins were calculated. The first time-bin would begin from the 1% of the rising-edge and shifted along for the appropriate number of time-bins.

To compute the sensitivity of each layer for a given time-gate, the sensitivity profiles at said time-gate was normalized with respect to the sum of the sensitivity profile at that time gate. The normalised values were then summed within said time-gate to provide the sensitivity profile for that simulation. To discretise the sensitivities to each tissue layer, the values of the sensitivity profile were then summed within each tissue layer and multiplied by 100%.

Because the placenta is the deepest layer, and following the general mechanism of TD-NIRS, it is expected that the latest time-bin will have the highest sensitivity to the placenta layer. As such, results shown in Fig 8 and listed in Table 7 are from the last time-bin.

Table 7. mean and standard-deviation of sensitivity to each tissues type for all simulations

	Skin	Adipose	Muscle	Placenta
Sensitivity (mean +/- STD)	3.68 ± 1.48 %	16.77 ± 20.56 %	54.06 ± 16.95 %	7.30 ± 12.31 %

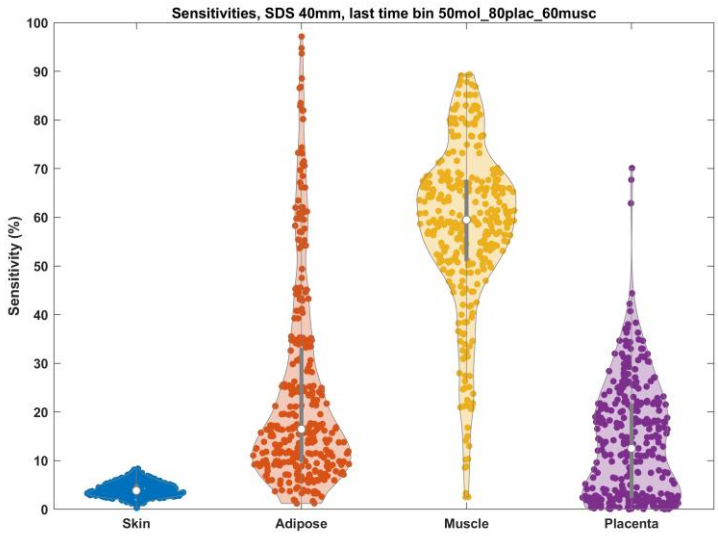


Fig. 8. Sensitivity to each tissue type for all 343 simulations.

Two-Layer Phantom

To experimentally validate the measurement of tissue oxygenation in layered media, we replicated the design of Sudakou et al [27] who developed a two-layer blood phantom. We conducted three experiments using this approach, varying the thickness of the superficial layer L (10, 16, and 20 mm, corresponding to Phantoms P1, P2, and P3, respectively). Each of the three experiments was carried out on a separate day. These three thicknesses approximate the 25th, 50th, and 75th percentiles of skin-to-placenta distances measured by clinicians via ultrasound in our study cohort.

The phantoms liquid of both compartments were prepared and continuously stirred using heating plates (one per compartment) equipped with 4 mm magnetic stirrers. The plates also included integrated thermometers, which maintained the liquid temperature at 37 °C. Blood oxygen saturation (StO_2) was manipulated separately in the superficial and deep compartments, one at a time. After filling both compartments with buffer, Intralipid, and blood, approximately 80 mL of the mixture was taken from the deep compartment, mixed with 10 g of dry bakery yeast (Dove, UK), and poured back in. The deep compartment was then fully reoxygenated by bubbling oxygen at 3 L/min through a submerged air stone until the dissolved oxygen (DO_2) reading reached 100%.

Following this initial cycle, yeast (5 g) was added to the superficial compartment in the same manner, after which the compartment was reoxygenated once. Finally, the deep compartment underwent an additional deoxygenation cycles. In total, the protocol consisted of three deoxygenation steps: one in the superficial compartment and two in the deep compartment. This protocol allowed us to investigate all combinations of oxygenation states in the superficial and deep compartments.

Data were acquired with an integration time of 100 ms per wavelength, yielding an effective acquisition rate of approximately 0.5 Hz for a full dataset (16 wavelengths from 780 to 870 nm at 6 nm steps, across 4 detectors), consistent with in-vivo acquisition settings. To increase SNR, DTOFs were averaged over the stable periods of each phantom (approximately 5 min per condition). StO_2 values were then estimated for each period using the algorithm described in Section III.
Table 8 reports the retrieved StO_2 values for the superficial and deep layers under the four oxygenation conditions across the three phantoms.

Table 8. Sto2 values retrieved for superficial and deep layer for the 4 conditions and 3 phantoms.

	Superficial Layer				Deep Layer			
ID	State	P1	P2	P3	State	P1	P2	P3
1	100%	61.81%	89.41%	77.22%	100%	70.42%	87.40%	68.55%
2	100%	75.73%	77.25%	54.60%	0%	38.60%	61.76%	50.38%
3	0%	18.78%	26.99%	15.83%	0%	12.15%	24.06%	24.46%
4	0%	15.93%	28.01%	17.34%	100%	31.45%	35.65%	34.65%

Clinical use of the system

In this section, we will report on the use of data recorded with MAESTROS II during monitoring sessions of pregnant participants at University College London Hospital. Specifically, we developed Machine learning algorithm to evaluate the possibility to classify the risk and outcomes of adverse pregnancy based on physiological data extracted from MAESTROS II. Additionally, we have also explored the possibility to use the raw data TD-NIRS data to extract information on the placental depth.

Clinical data set

The results reported here will focus on data acquired with the MAESTROS II system 158 monitoring sessions on pregnant participants at University College London Hospital between June 2023 and December 2024. The study was approved by the local ethics committee (protocol number: 156470).

The monitoring procedure was as followed: Following identification of placental location with Ultrasound Imaging, the MAESTROS probe was positioned on the maternal abdomen directly overlying the placenta and secured using a cardiotocography (CTG) belt and adhesive tape. All participants underwent TD-NIRS monitoring for a minimum of 30 minutes. Most were monitored in an upright semi-recumbent position (approximately 70° inclination). In cases of a posterior placenta, real-time ultrasound guidance was used to identify the most anterior or lateral placental region accessible from the maternal abdominal surface. In cases where skin-to-placenta depth was excessive in the upright position (defined as >4cm), left or right lateral maternal positioning was used to reduce the skin-placental distance to optimise signal penetration

The data recorded consisted thus of the optical data, the US imaged, used to measure the skin to placenta distance, and the CTG data. Moreover, the skin colour of the participants was recorded using the Fitzpatrick score.

The final clinical information gather was based on the clinicians classification of pregnancy as high or low risk based on pre-existing medical conditions, obstetric complications, lifestyle factors, age, and

previous pregnancy complications [37]. Indeed, high risk classification is strongly correlated with negative pregnancy outcomes. Pregnancy outcome was identified based on the Near Miss Criteria for Stillbirth , such as stillbirth or severe pre-eclampsia [38] . Table 9 summarizes the demographic data used.

Table 9. The number of processed cases with risk and outcome

Risk Classification		Mean Gestational Age	Pregnancy Outcome		Mean Gestational Age
Low Risk	158	35.2 ± 3.74	Positive	119	35.1 ± 3.80
			Negative	11	36.1 ± 2.99
High Risk	27	32.9 ± 3.45	Positive	13	35.9 ± 0.91
			Negative	14	31.8 ± 3.38

Machine learning algorithms

Deep learning model for placental depth estimation

As mentioned earlier, the variable depth of the placenta in the abdomen results in reliability issues for monitoring with NIRS. Here, a deep learning model is presented to estimate the placental depth using the DTOF measurements from the MAESTROS system positioned on the surface of the abdomen.

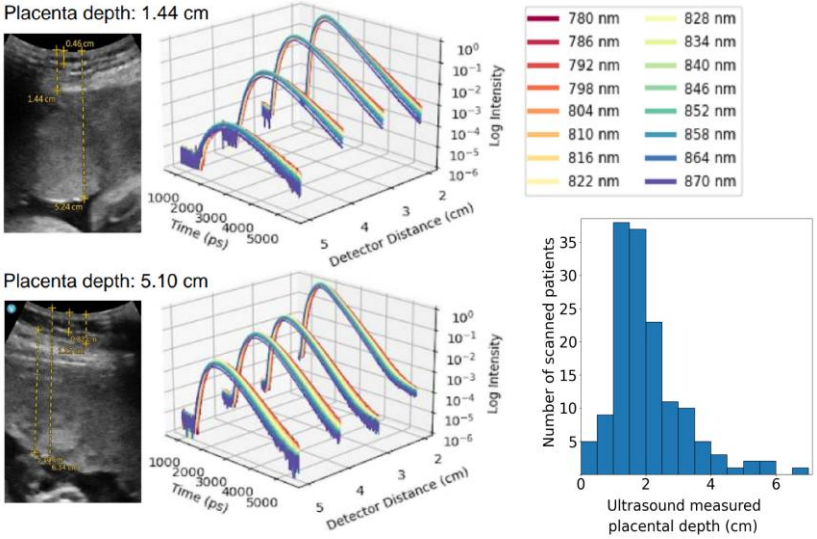


Fig. 9. Left: The average DTOF measured during a TD-NIRS scan, with 16 wavelengths (colour) and 4 source detector distances, next to ultrasound images demonstrating the placental depth in those cases. Bottom Right: The distribution of placental depths according to the ultrasound scans among the 128 participants.

A 500 MB neural network with two convolutional layers and two fully connected layers was constructed using PyTorch and trained to predict placental depth from the 64 data channels (see Figure 10). The input data was the average of a group of 30 consecutive DTOF time windows (1 to 5 ns), acquired in 60 seconds. Training data augmentation involved adding gaussian noise (s.d. 1% max intensity), random phase shifts in the time domain (s.d. 0.5 ns), and random dropout of 25% of the channels. The cost function was the mean square error between the placental depth measured from the ultrasound scans and the predicted value.

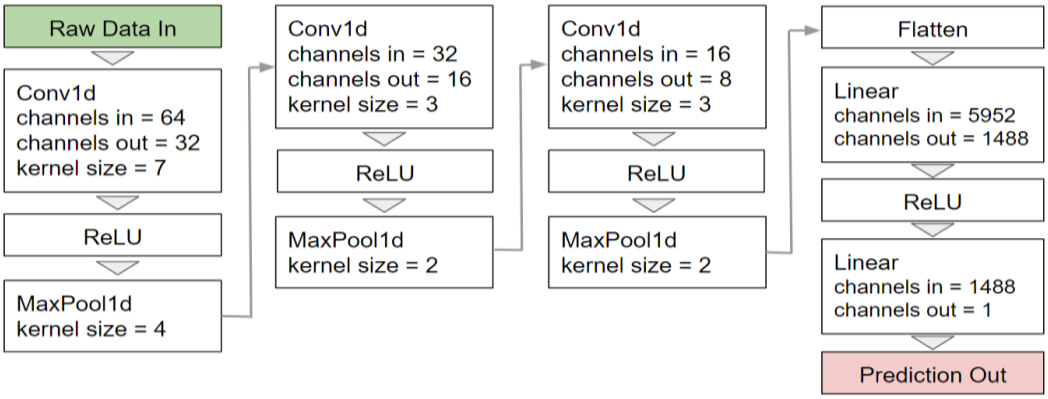


Fig. 10. The architecture of the convolutional neural network for regression of placental depth built using PyTorch.

Input gradient-based saliency [39] was applied to explain which features of the DTOF data the trained model uses to predict placental depth. This technique uses back projection of the trained neural network to quantify how much the output would be changed by variation in each element of each input case. Saliency was visualised by finding the mean absolute saliency for each timepoint and channel across all validation cases.

Risk and Outcome classification

Model Evaluation

The machine learning algorithm for binary classification of low or high risk therefore needs to produce good accuracy and sensitivity to high risk, despite the imbalanced dataset with a limited number of high-risk cases. Therefore, the metrics to evaluate the performance of the algorithm should also be robust to the class imbalance. Balanced Accuracy (BA) is the mean of the rates of correct classification for high risk and low risk cases in the dataset. This is required because the overall accuracy may appear high if the algorithm classifies every case as low risk. A second metric is the Area Under Receiver Operating Characteristic curve (AUROC), which tests a range of thresholds causing the algorithm to favour accuracy for each class to a different degree, and is also appropriate for datasets with class imbalance.

Usually, machine learning models are evaluated using test datasets, containing cases that were not used during training to guard against overfitting. However, the dataset is very small with a particularly limited number of high risk cases, so creating a single test dataset would undermine the size of the training dataset. Therefore, k-fold testing should be used, where a test dataset of 3 cases is randomly drawn from 165 cases and the model is trained with the remaining cases. This process can be repeated thousands of times with different combinations of training and test cases to evaluate how the model would perform with a large test dataset, like in real-world use.

Feature Extraction

The first step for building a machine learning model is to extract features from the raw training data, to produce a vector of numbers which can be fed into the model.

The processed MAESTROS data produces the physiological values that can be input into the machine learning model. Features can also be extracted from the time series data, by considering the standard deviation, the spectral centre of the Fourier transform, and six frequency windows within the Fourier transform. The semblance, i.e. a measure of the similarity between two time series, can also be used as further features.

Participant metadata can also be added as features. The model includes gestational age at the time of the scan. It also includes the Fitzpatrick scale. In total 158 features were extracted from the MAESTROS data, including: (i) 16 features obtained as a mean value during the scan. This includes superficial and deep values for: StO2, oxygenated haemoglobin concentration, deoxygenated haemoglobin concentration, total haemoglobin concentration, water percentage, fat percentage, optical scattering power, and optical scattering amplitude, (ii) 140 features extracted from 10 timeseries. The timeseries include superficial and deep values for: absolute oxygenation, oxygenated haemoglobin

concentration, deoxygenated haemoglobin concentration, total haemoglobin concentration, and CCO concentration. For each timeseries, 14 features were extracted. The standard deviation, 6 Fourier transform bands, the spectral centroid, semblance with CCO timeseries, semblance with absolute oxygenation time series, semblance with total haemoglobin concentration timeseries, and (iii) 2 features were extracted from the patient metadata: gestational age and Fitzpatrick scale.

Support Vector Machine

The Support Vector Machine (SVM) method [40] was chosen as model architecture. Indeed, a SVM is less likely to overfit with a small dataset, it effectively uses a small number of meaningful physiological features and their significance for classification is easily explained through model weights, and it trains very rapidly. The latter property benefits automatic retraining and revalidation in a cloud environment, as new data is frequently acquired.

Briefly, a SVM works by representing each case in the dataset as a point in a high dimensional space of features, before identifying three points (the support vectors) which best define a hyperplane to separate the clusters of points of each class. The robustness of SVMs to class imbalance can be tackled by assigning a higher penalty to misclassifications of the minority class, by a factor proportional to the relative frequency of the minority class. The SVM was created using the SVC application in the python library Scikit-Learn, with the “class_weight='balanced'” flag used to tackle class imbalance. By default SVMs are linear, meaning that the hyperplane is flat. However, a selection of non-linear kernels can optionally be used to warp feature space so that the hyperplane becomes curved, which can allow more accurate classification if the features have complex non-linear relationships with each other.

Results of machine learning classifications

Placental depth evaluation

The model trained with 108 cases predicted the placental depth in 20 test cases with a mean error of 0.42 cm and a strong statistical correlation between predicted values and the measurements from the ultrasound scans (see Figure 11). However, the correlation becomes much weaker if only the test cases with a placental depth less than 3 cm are considered, where the placenta is close enough to the skin for more reliable NIRS. Therefore, the model is likely most useful to identify the minority of cases where the placenta is deeper than 3 cm, which account for 20% of cases in this cohort (see Figure 5). When applied to this classification task, the model is 100% accurate with the test data.

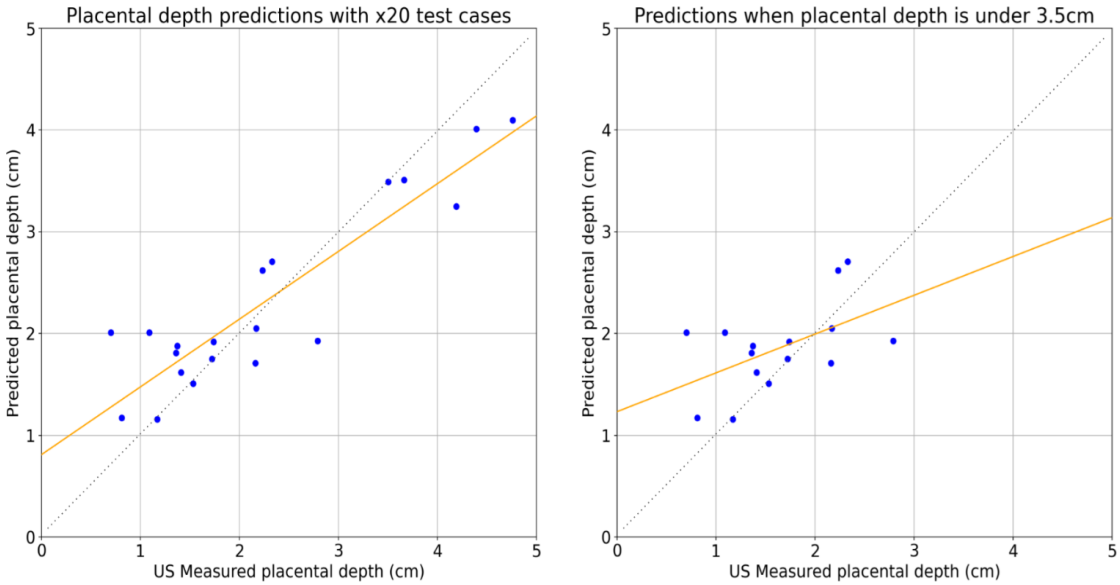


Fig. 11. Left: The result when the model was applied to unseen test cases data, where the positive correlation between the labeled placental depth and the model prediction shown by the linear fit (orange line) has high significance $p<0.0001$. Right: The significance of this relationship is weaker at $p=0.041$ if only cases with placental depth below 3cm are considered.

The model explanation analysis (Figure 12) shows that data from the shortest source detector distance (2 cm) has the least salience, which can be expected just because the photon diffusion trajectories have the least depth. The last part of the time window (4000-5000 ps) clearly has the strongest salience, representing the late arriving photons which have the deepest diffusion trajectories and so are most likely to interact with the placenta. The longest wavelength channels (840-870 nm) have the strongest saliency, which may be related to the high absorption coefficient of oxygenated haemoglobin in this range and the typically high oxygenation of the placenta compared to other abdominal tissue.

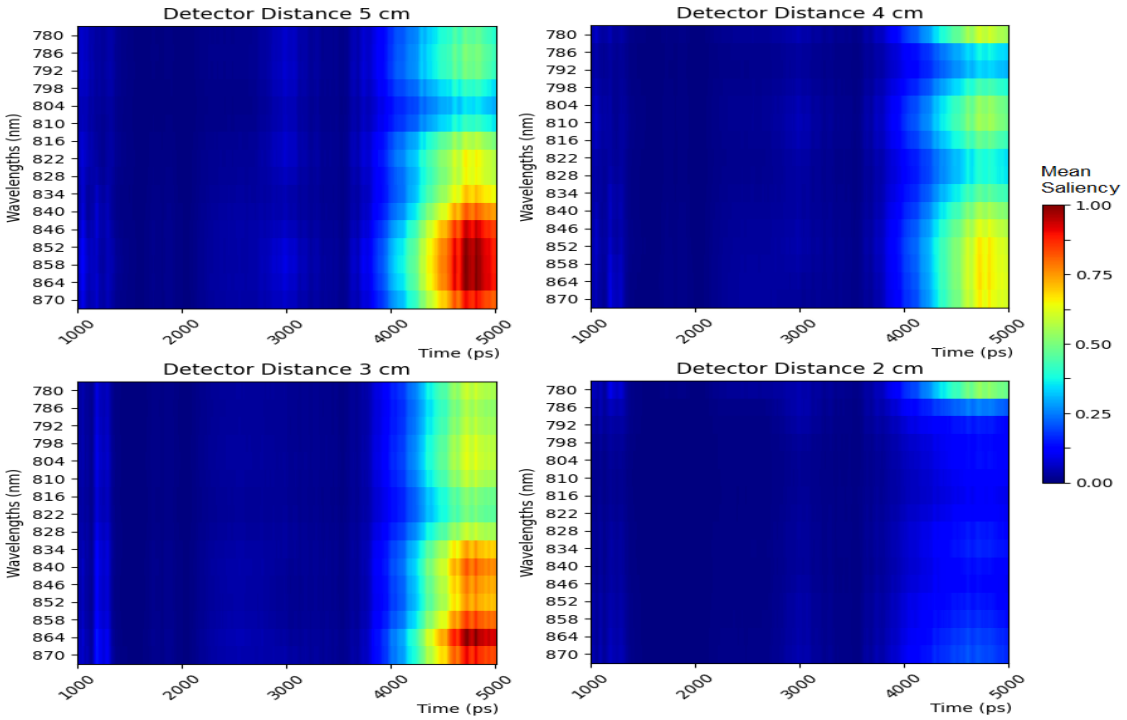


Fig. 12. The absolute input gradient-based saliency for the trained model, as the mean across all scans in the validation dataset, normalised from 0 to 1. The quadrants represent the four source-detector distances, each showing the salience of each time point in the input DTOF time window for each wavelength.

Risk and outcome classifications

The SVM architecture for risk classification performed best with both gestational age and Fitzpatrick scale metadata added as features alongside the 158 features extracted from the MAESTROS dataset, reaching a balanced accuracy of 92.1% (see table 10). The effect of adding the gestational age alone was to improve balanced accuracy by 2.3 percentage points, while adding the Fitzpatrick score alone improved balanced accuracy by 0.6 percentage points. The feature weights (see figure 13) also show that gestational age was the stronger metadata feature with a lower value strongly associated with high risk, while a higher Fitzpatrick score was slightly associated with high risk.

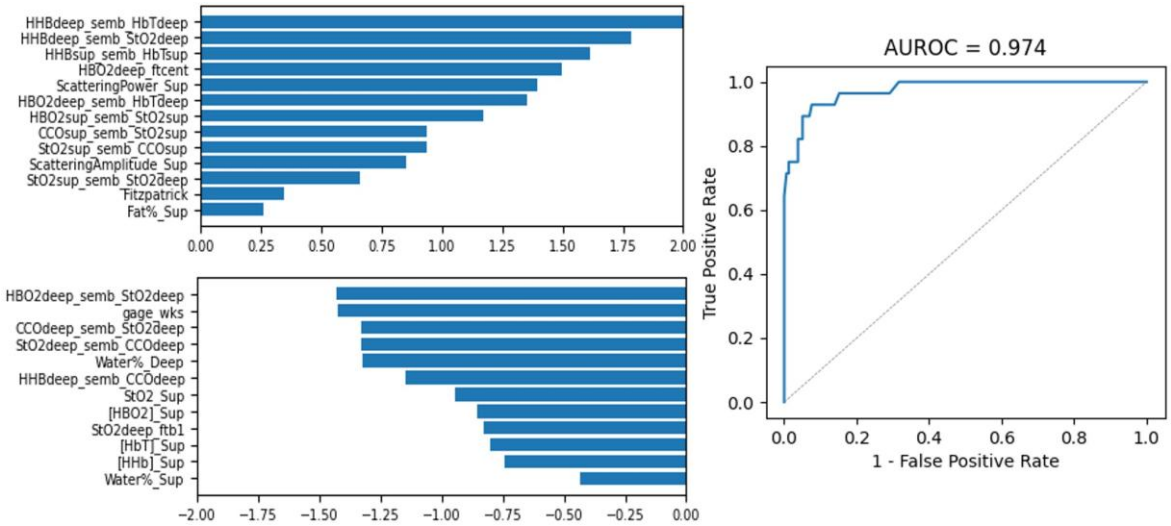


Fig. 13. Left: The SVM feature weights ordered by strength, where positive values mean larger feature values are associated with High Risk cases and visa-versa. Right: The AUROC curve for this model.

The SVM feature weights (see the 25 strongest weights in figure 12) show that high risk pregnancy classification is associated with: (i) lower oxygenation in both layer, reflected by negative weighting (St02_sup); (ii) higher noise in the timeseries signals. This is reflected by both the positive weighting of the Fourier transform centroid and the positive weighting of semblance of deoxygenated haemoglobin with oxygenated haemoglobin. These timeseries should instead have opposing trends if the data is of good quality, which is more common in low risk cases; (iii) higher superficial scattering power and amplitude.

Table 10. The risk classification models' performance results, comparing the effect of available metadata (gestational age and Fitzpatrick scale).

Gestational age	Fitzpatrick scale	Balanced Accuracy	AUROC
✓	✓	92.1%	97.4%
✓	×	90.8%	94.9%
×	✓	89.1%	93.7%
×	×	88.5%	90.1%

Table 11 shows the SVM architecture for outcome prediction performed best with both gestational age, Fitzpatrick scale, and the clinicians risk classification at the time of the scan was added as features alongside the 158 features extracted from the MAESTROS dataset, reaching a balanced accuracy of 78.0%. The clinicians' risk assessment was the strongest feature, removing this reduced the performance by 5.5%.

Table 11. The outcome prediction models' performance results, comparing the effect of available metadata (gestational age, Fitzpatrick scale and prior risk classification).

Gestational age	Fitzpatrick scale	Risk classification	Balanced Accuracy	AUROC
✓	✓	✓	78.0%	76.8%
✓	✓	×	72.5%	74.9%
✓	×	×	80.0%	73.9%
×	✓	×	67.9%	69.8%
×	×	×	68.4%	69.9%

The weights in figure 14 show superficial scattering power is linked to negative outcomes similarly to high risk classification, but more key features are associated with the placental layer. For example lower oxygenation in the placental layer (StO2_deep) is associated with negative outcomes, as are fluctuations in placental haemoglobin and oxygenation on the time scale of around 10 seconds extracted from the Fourier transform bands (HBTdeep_ftb2, St02deep_ftb3, St02deep_ftb4).

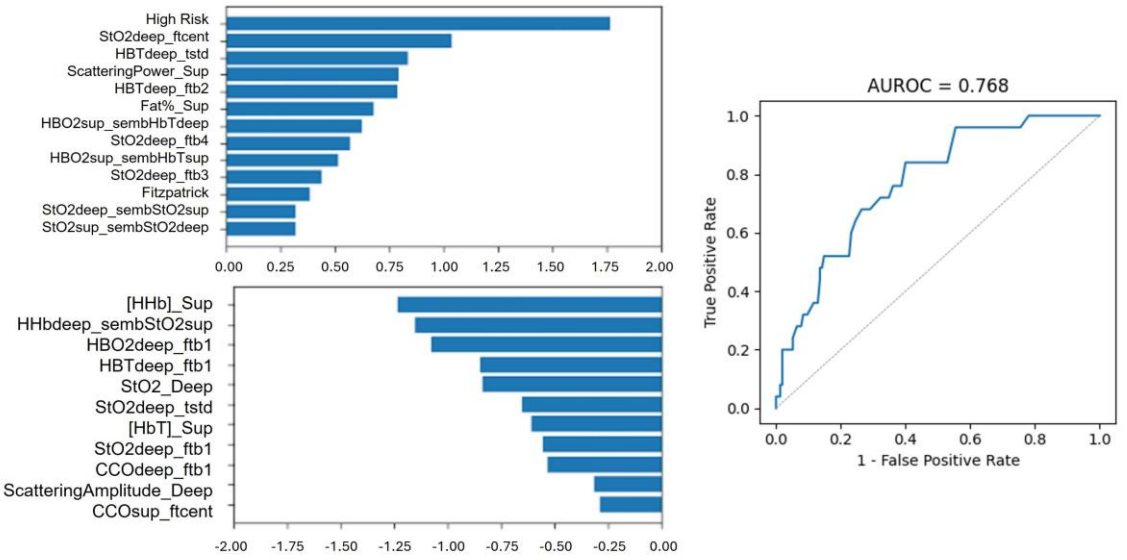


Fig. 14. Left: The SVM feature weights ordered by strength, where positive values mean larger feature values are associated with Negative Outcome cases and visa-versa. Right: The AUROC curve for this model.

Improvement of classification with depth measurements

When limiting the training and test data was restricted to cases with a maximum placental depth, the classification accuracy for both risk and outcome improved when the threshold of 3.5 cm was chosen (see tables 5 and 6). This was still the case when the depth values predicted by the deep learning model during its validation were used for this selection step, demonstrating its value for NIRS data cleaning.

Table 12. The risk classification models' performance (balanced accuracy) when data was restricted according to placental depth, using either the ultrasound data of the deep learning model for depth measurement.

Placental depth exclusion threshold	<2.5cm	<3.0cm	<3.5cm	<4cm	All data
Depth from ultrasound scan	89.2%	91.3%	93.3%	93.1%	92.1%
Depth from model prediction	88.1%	92.1%	93.2%	93.1%	

Table 13. The outcome classification models' performance (balanced accuracy) when data was restricted according to placental depth, using either the ultrasound data of the deep learning model for depth measurement.

Placental depth exclusion threshold	<2.5cm	<3.0cm	<3.5cm	<4cm	All data
Depth from ultrasound scan	69.0%	77.0%	78.9%	77.0%	78.0%
Depth from model prediction	67.3%	76.0%	78.2%	76.0%	

Discussion

We have described the design and evaluation of the second generation of the MAESTROS system. This new iteration provides hardware upgrades that translates in a gain in light harvesting capacity (nearly doubled responsivity compared to MAESTROS V.1.0) and IRF characteristics (FWHM = 274 ps), via an improved detection scheme, notably with the adoption of HPMT that exhibits negligible after pulse and the adoption of a new multichannel photon counting card, suppressing potential crosstalk between our channels. This allows us to boost the maximum photon counts and enables a high dynamic range between short and long source detector distance channels. Moreover, the robustness of use in a clinical context had also been improved via the addition of a reference channel and of the UPS, making the eventual shift in IRF shape due to difficult environment trackable and correctable, and allowing to move the system between clinical rooms without shutting it down.

The upgrades provided good optical-property retrieval with V.2.0, and further improved accuracy with V.2.1 owing to its sharper IRF, with accuracy increasing by 38% for μ_a and 34% for μ_s'.

Given the targeted application, we implemented validation steps to quantify sensitivity and accuracy for multilayer monitoring. MC simulations of layered tissue geometries parameterised by measured anatomy indicate that the time-domain measurements achieve a mean placental sensitivity up to 7.30 ± 12.31% across all 356 anatomies, with peak sensitivity of more than 60% for some cases. These values are comparable to typical CW-NIRS performance on the adult brain with a 3 cm source-detector separation, a setting in which NIRS has demonstrated reliability, supporting the plausibility of our estimates.

To complement the simulations and test depth selectivity under controlled conditions, we measured two-layer blood phantoms. These experiments show that the system can separate superficial and deep oxygenation signals. The absolute accuracy reported in these phantom measurements are limited for several reasons. Firstly, the phantom matrix is more than 98 percent water, which raises the absorption coefficient relative to tissue that typically contains 70/80% water. The effect is strongest at longer wavelengths and reduces SNR for this protocol, compare to in-vivo measurements. Secondly, the current reconstruction is a first iteration that can be strengthened with more advanced or complementary methods. Indeed, TD spatially enhanced NIRS [40] has been shown to improve parameter identifiability and could be incorporated, and recent multilayer models with explicit scattering estimation are promising; adopting the approach of Avanzi at al. [41] may help.

In addition, the phantom results reveal partial cross-talk between layers: when one layer was deoxygenated, the retrieved oxygenation of the other layer was also shifted. This behaviour reflects the overlap of photon sampling volumes and limitations of the current model. A dependence on superficial layer thickness was also observed, with retrieval accuracy decreasing as the superficial compartment became thicker, consistent with light transport theory. Importantly, however, the system consistently detected relative changes in the targeted layer, supporting the feasibility of tracking layered oxygenation

dynamics in vivo despite the noted limitations. Finally, it is worth noting that during internal QC of the two-layer phantom data we identified a systematic issue (i.e. reflections at the boundary due to imperfect contact of the probe with the Mylar Foil) that likely inflates the absolute recovery errors; the experiment will be repeated with the corrected procedure, and updated results will be included in a revised version. Despite these limitations, the present data indicate that the algorithm can distinguish superficial from deep oxygenation, although accuracy is currently limited. We are refining both the algorithm and the phantom recipe to better match tissue composition.

To assess real-world relevance, we applied the same gated and layer-specific features to our clinical recordings and trained a classifier to test whether they capture risk-related variation. In the current cohort (n = 185 sessions), the SVM model achieved balanced accuracy of 92.1 % and AUCROC of 97.4% when including gestational age and Fitzpatrick scale. Indeed, we evaluated the impact of non-optical metadata by training models with TD-NIRS features alone and models that combined the same features with the available metadata. The corresponding results are summarised in the classification tables for the risk and outcome tasks. Across both tasks, the combined models matched or exceeded the performance of the NIRS-only models, indicating that the metadata contribute complementary information rather than duplicating what is already contained in the optical signal.

A separate convolutional network trained on multi-distance, multi-wavelength DTOFs estimated placental depth based on optical data only, with mean error of 0.42 cm, and a strong statistical correlation between predicted values and the measurements from the ultrasound scans. Moreover, the saliency analysis highlighted late time windows and longer wavelengths as most informative for this procedure, aligning with TD-NIRS depth physics. These early results suggest an optical depth estimate could automate gate selection, provide a depth prior for the two-layer fit, and flag sessions that fall outside the instrument's sensitivity range. That said, this is preliminary and will require calibration and external validation before any clinical use.

Having access to depth sensitivity lets us stratify analyses by expected detectability and tailor processing accordingly. In our dataset, restricting to cases with ultrasound-measured placental depth below 3.5 cm improved balanced accuracy, and applying the same restriction using the DTOF-based predicted depth produced a similar gain, supporting its use for automated QC and gate selection.

These results constitute proof of concept rather than diagnostic performance. Indeed, the sample size is modest, labels may contain noise, and absolute recovery accuracy is currently limited. However, the classifiers performed well with feature importance dominated by deep-gated physiology and late time windows as expected from TD-NIRS physics. Taken together, these patterns indicate that the measurements already encode clinically relevant signal. These first results are encouraging, but establishing a gold standard for placental functional monitoring will still require coordinated work on reference phantoms, harmonised acquisition and analysis protocols, and multicentre validation against agreed clinical endpoints.

Conclusion

We presented MAESTROS II, a second-generation TD-NIRS platform purpose-built for non-invasive, depth-resolved assessment of placental oxygenation and metabolism at the bedside. Targeted hardware upgrades increased detected photons, narrowed the IRF, and improved stability; together with a two-step analysis (spectrally constrained two-layer fit + gated Beer–Lambert), this enables absolute composition estimates and layer-specific dynamics.

Validation with standard protocols (BIP, MEDPHOT), patient-informed simulations, and two-layer blood phantoms showed sensitivity to placental-depth ranges typical of late gestation and the ability to separate superficial from deep oxygenation signals. Early clinical analyses indicate that depth-gated features carry clinically relevant information when combined with routinely available metadata, and that DTOFs can support automated estimation of placental depth with physics-consistent saliency.

Limitations include preliminary absolute accuracy in layered phantoms, affected by a water-rich matrix and a probe–phantom contact artifact that will be corrected in a repeat experiment, and a modest, imbalanced clinical dataset. Next steps are to (i) repeat the phantom study with improved coupling, and (ii) improve the algorithm to increase the accuracy of the oxygenation estimation of the placenta.

These first results are encouraging, but establishing a gold-standard framework for placental functional monitoring will require coordinated reference phantoms, harmonised acquisition/analysis protocols, and multicentre validation against agreed clinical endpoints.

REFERENCES

[1] N. M. Gude, C. T. Roberts, B. Kalionis, and R. G. King, “Growth and function of the normal human placenta,” *Thromb. Res.*, vol. 114, no. 5–6, pp. 397–407, 2004, doi: 10.1016/j.thromres.2004.06.038.

[2] R. M. Roberts, J. A. Green, and L. C. Schulz, “The evolution of the placenta,” *Reproduction*, vol. 152, no. 5, pp. R179–R189, 2016, doi: 10.1530/rep-16-0325.

[3] R. Gagnon, “Placental insufficiency and its consequences,” *Eur. J. Obstet. Gynecol. Reprod. Biol.*, vol. 110, pp. S99–S107, 2003, doi: 10.1016/s0301-2115(03)00179-9.

[4] J. E. Wardinger and S. Ambati, “Placental Insufficiency.,” 2025.

[5] P. Calis, L. Vojtech, F. Hladik, and M. G. Gravett, “A review of ex vivo placental perfusion models: an underutilized but promising method to study maternal-fetal interactions,” *J. Matern.-Fetal Neonatal Med.*, vol. 35, no. 25, pp. 8823–8835, 2022, doi: 10.1080/14767058.2021.2005565.

[6] R. R. Zabel, R. R. Favaro, T. Groten, P. Brownbill, and S. Jones, “Ex vivo perfusion of the human placenta to investigate pregnancy pathologies,” *Placenta*, vol. 130, pp. 1–8, 2022, doi: 10.1016/j.placenta.2022.10.006.

[7] A. Clark *et al.*, “Developments in functional imaging of the placenta,” *Br. J. Radiol.*, vol. 96, no. 1147, p. 20211010, 2022, doi: 10.1259/bjr.20211010.

[8] J. G. Ray, M. J. Vermeulen, A. Bharatha, W. J. Montanera, and A. L. Park, “Association Between MRI Exposure During Pregnancy and Fetal and Childhood Outcomes,” *JAMA*, vol. 316, no. 9, pp. 952–961, 2016, doi: 10.1001/jama.2016.12126.

[9] R. Re *et al.*, “A review on time domain diffuse optics: principles and applications on human biological tissues,” *Riv. del Nuovo Cimento*, vol. 48, no. 3, pp. 157–239, 2025, doi: 10.1007/s40766-025-00067-2.

[10] F. Lange and I. Tachtsidis, “Clinical Brain Monitoring with Time Domain NIRS: A Review and Future Perspectives,” *Appl. Sci.*, vol. 9, no. 8, p. 1612, 2019, doi: 10.3390/app9081612.

[11] J. Kakogawa, K. Sumimoto, T. Kawamura, S. Minoura, and N. Kanayama, “Noninvasive Monitoring of Placental Oxygenation by Near-Infrared Spectroscopy,” *Am J Perinatol*, vol. 27, no. 06, pp. 463–468, 2010, doi: 10.1055/s-0030-1247600.

[12] K. Ražem, J. Kocijan, M. Podbregar, and M. Lučovnik, “Near-infrared spectroscopy of the placenta for monitoring fetal oxygenation during labour,” *PLoS ONE*, vol. 15, no. 4, p. e0231461, Apr. 2020, doi: 10.1371/journal.pone.0231461.

[13] T. Nguyen *et al.*, “Non-invasive transabdominal measurement of placental oxygenation: a step toward continuous monitoring,” *Biomed. Opt. Express*, vol. 12, no. 7, p. 4119, 2021, doi: 10.1364/boe.424969.

[14] L. Wang *et al.*, “Non-invasive monitoring of blood oxygenation in human placentas via concurrent diffuse optical spectroscopy and ultrasound imaging,” *Nat. Biomed. Eng.*, vol. 6, no. 9, pp. 1017–1030, Aug. 2022, doi: 10.1038/s41551-022-00913-2.

[15] A. D. Mora *et al.*, “Towards next-generation time-domain diffuse optics for extreme depth penetration and sensitivity,” *Biomed. Opt. Express*, vol. 6, no. 5, pp. 1749–1760, May 2015, doi: 10.1364/boe.6.001749.

[16] H. Wabnitz, D. Contini, L. Spinelli, A. Torricelli, and A. Liebert, “Depth-selective data analysis for time-domain fNIRS: moments vs. time windows,” *Biomed. Opt. Express*, vol. 11, no. 8, p. 4224, Aug. 2020, doi: 10.1364/boe.396585.

[17] “Safety of laser products. Equipment classification and requirements, BS EN 60825-1:2014, The British Standards Institution, 2014,” 2014.

[18] A. D. Mora *et al.*, “The SiPM revolution in time-domain diffuse optics,” *Nucl. Instrum. Methods Phys. Res. Sect. A: Accel., Spectrometers, Detect. Assoc. Equip.*, vol. 978, p. 164411, 2020, doi: 10.1016/j.nima.2020.164411.

[19] M. Wahl, T. Röhlicke, S. Kulisch, S. Rohilla, B. Krämer, and A. C. Hocke, “Photon arrival time tagging with many channels, sub-nanosecond deadtime, very high throughput, and fiber optic remote synchronization,” *Rev. Sci. Instrum.*, vol. 91, no. 1, p. 013108, 2020, doi: 10.1063/1.5121412.

[20] F. Lange, L. Dunne, L. Hale, and I. Tachtsidis, “MAESTROS: A Multiwavelength Time-Domain NIRS System to Monitor Changes in Oxygenation and Oxidation State of Cytochrome-C-Oxidase,” *IEEE J. Sel. Top. Quantum Electron.*, vol. 25, no. 1, pp. 1–12, Jan. 2019, doi: 10.1109/jstqe.2018.2833205.

[21] H. Wabnitz *et al.*, “Performance assessment of time-domain optical brain imagers, part 1: basic instrumental performance protocol,” *J. Biomed. Opt.*, vol. 19, no. 8, pp. 086010–086010, 2014, doi: 10.1117/1.jbo.19.8.086010.

[22] A. Pifferi *et al.*, “Performance assessment of photon migration instruments: the MEDPHOT protocol,” *Appl. Opt.*, vol. 44, no. 11, p. 2104, Apr. 2005, doi: 10.1364/ao.44.002104.

[23] A. Torricelli *et al.*, “Time domain functional NIRS imaging for human brain mapping,” *NeuroImage*, vol. 85, pp. 28–50, Jan. 2014, doi: 10.1016/j.neuroimage.2013.05.106.

[24] R. Re, I. Pirovano, D. Contini, L. Spinelli, and A. Torricelli, “Time Domain Near Infrared Spectroscopy Device for Monitoring Muscle Oxidative Metabolism: Custom Probe and In Vivo Applications,” *Sensors*, vol. 18, no. 1, p. 264, Jan. 2018, doi: 10.3390/s18010264.

[25] C. D’Andrea *et al.*, “Time-resolved spectrally constrained method for the quantification of chromophore concentrations and scattering parameters in diffusing media,” *Opt. Express*, vol. 14, no. 5, p. 1888, 2006, doi: 10.1364/oe.14.001888.

[26] A. Liemert and A. Kienle, “Light diffusion in N-layered turbid media: frequency and time domains,” *J. Biomed. Opt.*, vol. 15, no. 2, pp. 025002-025002–8, 2010, doi: 10.1117/1.3368682.

[27] A. Sudakou, H. Wabnitz, A. Liemert, M. Wolf, and A. Liebert, “Two-layered blood-lipid phantom and method to determine absorption and oxygenation employing changes in moments of DTOFs,” *Biomed. Opt. Express*, vol. 14, no. 7, pp. 3506–3531, 2023, doi: 10.1364/boe.492168.

[28] Y. Tsuchiya, “Photon path distribution and optical responses of turbid media: theoretical analysis based on the microscopic Beer-Lambert law,” *Phys. Med. Biol.*, vol. 46, no. 8, pp. 2067–2084, Aug. 2001, doi: 10.1088/0031-9155/46/8/303.

[29] P. Lanka *et al.*, “Multi-laboratory performance assessment of diffuse optics instruments: the BitMap exercise,” *J. Biomed. Opt.*, vol. 27, no. 7, pp. 074716–074716, June 2022, doi: 10.1117/1.jbo.27.7.074716.

[30] Q. Fang and D. A. Boas, “Monte Carlo simulation of photon migration in 3D turbid media accelerated by graphics processing units,” *Opt. Express*, vol. 17, no. 22, p. 20178, Oct. 2009, doi: 10.1364/oe.17.020178.

[31] L. Yu, F. Nina-Paravecino, D. Kaeli, and Q. Fang, “Scalable and massively parallel Monte Carlo photon transport simulations for heterogeneous computing platforms,” *J. Biomed. Opt.*, vol. 23, no. 1, pp. 010504–010504, 2018, doi: 10.1117/1.jbo.23.1.010504.

[32] A. E. Karsten and J. E. Smit, “Modeling and Verification of Melanin Concentration on Human Skin Type,” *Photochem. Photobiol.*, vol. 88, no. 2, pp. 469–474, 2012, doi: 10.1111/j.1751-1097.2011.01044.x.

[33] S. L. Jacques, “Optical properties of biological tissues: a review,” *Phys. Med. Biol.*, vol. 58, no. 11, pp. R37–R61, July 2013, doi: 10.1088/0031-9155/58/11/r37.

[34] N. Bosschaart, R. Mentink, J. H. Kok, T. G. van Leeuwen, and M. C. G. Aalders, “Optical properties of neonatal skin measured in vivo as a function of age and skin pigmentation,” *J. Biomed. Opt.*, vol. 16, no. 9, pp. 097003-097003–10, 2011, doi: 10.1117/1.3622629.

[35] P. Oltulu, B. Ince, N. Kokbudak, S. Findik, and F. Kilinc, “Measurement of epidermis, dermis, and total skin thicknesses from six different body regions with a new ethical histometric technique,” *Turk. J. Plast. Surg.*, vol. 26, no. 2, p. 56, 2018, doi: 10.4103/tjps.tjps_2_17.

[36] “Intrapartum care for women with existing medical conditions or obstetric complications and their babies”.

[37] S. J. Gordijn, A. T. Papageorghiou, A. L. David, S. Ali, and W. Ganzevoort, “Near-miss criteria for stillbirth in global research: the ‘In Utero’ consensus,” *Ultrasound Obstet. Gynecol.*, vol. 65, no. 1, pp. 5–8, 2025, doi: 10.1002/uog.29120.

[38] K. Simonyan, A. Vedaldi, and A. Zisserman, “Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps,” *arXiv*, 2013, doi: 10.48550/arxiv.1312.6034.

[39] C. Cortes and V. Vapnik, “Support-vector networks,” *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995, doi: 10.1007/bf00994018.

[40] L. Yang, H. Wabnitz, T. Gladytz, A. Sudakou, R. Macdonald, and D. Grosenick, “Space-enhanced time-domain diffuse optics for determination of tissue optical properties in two-layered structures,” *Biomed. Opt. Express*, vol. 11, no. 11, p. 6570, Nov. 2020, doi: 10.1364/boe.402181.

[41] E. Avanzi, L. D. Sieno, A. D. Mora, L. Spinelli, and A. Torricelli, “Evaluation of layered tissue scattering properties: a time-domain spatially resolved spectroscopy approach,” *J. Biomed. Opt.*, vol. 30, no. 7, pp. 075002–075002, 2025, doi: 10.1117/1.jbo.30.7.075002.