CORA Cork Open Research Archive
Cartlann Taighde Oscailte Chorcaí

Title	AI-informed development for a Lactate measurement tool
Authors	Kiely, Cian;Rossberg, Nicola;Krishnamoorthy, Shree;Visentin, Andrea
Publication date	2024
Original Citation	Kiely, C., Rossberg, N., Krishnamoorth, S. and Visentin, A. (2024) 'AI-informed development for a Lactate measurement tool', 32nd Irish Conference on Artificial Intelligence and Cognitive Science, Dublin, Ireland, December 9-10, 2024, pp. 434-446.
Type of publication	Conference item
Rights	© 2024, the Authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0) - https://creativecommons.org/licenses/by/4.0/
Download date	2026-11-06 22:35:06
Item downloaded from	https://hdl.handle.net/10468/16872

UCC
University College Cork, Ireland
Coláiste na hOllscoile Corcaigh

AI-Informed Development for a Lactate Measurement Tool

Cian Kiely[1], Nicola Rossberg[1,2,*], Shree Krishnamoorthy[3] and Andrea Visentin[1,2,4,*]

[1]School of Computer Science & IT, University College Cork, Ireland
[2]SFI Center for Research Training in Artificial Intelligence, University College Cork, Ireland
[3]Biophotonics@Tyndall, IPIC, Tyndall National Institute, Ireland
[4]SFI Insight Centre for Data Analytics, University College Cork, Ireland

Abstract

Lactate has been identified as a key biomarker, with spikes co-occurring with high-risk medical conditions including sepsis and hypoxia. Despite its high medical value, current methods of Lactate measurement require repeated blood sampling from the patient, which is both costly and invasive, and consequently tends to be limited to intensive care units. Spectroscopy, a non-invasive light-based system, presents a cost-effective alternative to these traditional methods, which permits continuous measurement and improved patient monitoring. Through the use of machine learning, spectroscopic measurements can be used to estimate blood Lactate levels in an accessible and low-cost manner. In this study, machine learning models were trained on Near-infrared (NIR) spectroscopy data, to identify the best set-up for high-precision estimation of Lactate levels. The results of the analysis are used to determine the best path length for spectroscopic measurements. Feature selection is implemented to establish the most important wavelengths for prediction and inform on the most relevant spectral regions for the given task. Explainability is implemented to analyse feature contributions and allow inference of potentially interfering components that should be considered for further testing. The results showed that by using a random forest, R^2 values of 0.9986 can be achieved. Feature selection increased predictive performance considerably with R^2 values as high as 0.9996 and the implementation of explainability allowed the identification of important wavelength ranges.

Keywords

Lactate Measurement, Artificial Intelligence, Explainability, Chemometrics

1. Introduction

Lactate has been identified as an important marker of patient health and can function as a proxy alarm system for various severe health conditions including sepsis [1]. Lactate is produced during a process called 'Glycolysis', where glucose is broken down into Lactate or pyruvate and the energy released during this process is used to create high-energy molecules such as ATP [2]. As the key metabolite of the anaerobic pathway, Lactate is produced when aerobic respiration cannot meet tissues' energy demands. If Lactate is not adequately cleared due to illness or overproduction, lactic acidosis occurs, which has been identified as a precipitator to severe health complications. The current clinical method of measuring blood Lactate levels involves intermittent blood sampling using an arterial blood gas analyzer (ABG), which requires arterial blood samples [3, 4]. This method is usually restricted to use in intensive care units due to cost and high invasiveness. Consequentially, there is a need for an alternative method of monitoring Lactate levels in a hospital setting. Spectroscopy offers two key advantages over the traditional approach. First, spectroscopic measurements are comparatively cheaper, hence easing measurement and potentially allowing for more widespread implementation. Second, spectroscopic measurements are non-invasive and can therefore be implemented without additional blood samples, reducing patient burden. As such, spectroscopy may offer a feasible alternative to the traditional methods of estimating lactate concentrations.

In this study, these advantages of spectroscopy are leveraged to design a machine learning system

32nd Irish Conference on Artificial Intelligence and Cognitive Science, December 09–10, 2024, Dublin, Ireland
*Corresponding author.
120742001@umail.ucc.ie (C. Kiely); n.rossberg@cs.ucc.ie (N. Rossberg); shree.krishnamoorthy@gmail.com (S. Krishnamoorthy); andrea.visentin@ucc.ie (A. Visentin)
0009-0005-3883-5833 (N. Rossberg); 0000-0003-0653-599X (S. Krishnamoorthy); 0000-0003-3702-4826 (A. Visentin)

© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

that allows the prediction of Lactate levels through NIR spectroscopy. In previous works, machine learning was used as a post-hoc approach for the classification of spectroscopic signals [5]. This study now aims to use machine learning methodologies to inform the experimental design and measurement of spectroscopic signals and hence increase the predictive power of these models through intervention at the data collection stage. Three machine learning models are implemented to identify the best path length for predicting Lactate across a wide range of concentrations. Feature selection is conducted to identify important wavelength ranges and allow the reduction of the number of recorded signals. This feature selection enables the removal of redundancy in highly correlated data, increasing processing times and algorithm efficiency. After feature selection, SHapley Additive exPlanations (SHAP)[6] is used to analyse the contributions of each feature to the prediction of the outcome, allowing the identification of important features. This is advantageous, as certain wavelengths can be linked to biological components and by identifying which wavelengths lead to a given prediction, further testing for interfering components can be informed. Through these analyses, the best design and setup for spectral measurements and prediction of Lactate can be identified.

2. Literature Review

This literature review will first discuss the importance of Lactate as a biomarker for adult and infant health. It will then review current approaches to Lactate measurement and previous spectroscopic applications for Lactate measurement.

2.1. Lactate as Biomarker

Lactate has been identified as a key indicator of patient health in a series of medical conditions. While resting Lactate levels in healthy individuals vary between 1 and 2 mmol/L, increased resting levels are indicative of a wide range of serious health complications including hypoxia, sepsis, diabetes and toxin-related conditions [7]. Several previous studies have examined the predictive power of Lactate levels on patient development. A meta-analysis by Zhang and Xu [8] found that Lactate clearance was predictive of all-cause mortality in critically ill patients, with higher Lactate clearance predicting improved health outcomes. A second study by Mokline et al. [9] emphasised that plasma Lactate can function as a powerful predictive indicator of sepsis and mortality in patients suffering from burn wounds. However, while the importance of Lactate level monitoring as a marker for patient health and development has been established, the implementation of continuous monitoring in a hospital setting remains challenging. Traditionally, patient Lactate levels are established via blood samples. The problem with this methodology is the invasiveness and high cost of taking measurements. Additionally, this method does not lend itself to continuous measurement of Lactate levels. As such, alternative methods of establishing Lactate levels are desired for improved health monitoring.

2.2. Current Methods of Measuring Lactate

Several new methodologies have been proposed to allow the continuous measurement of Lactate levels. A comprehensive overview of the different approaches to Lactate monitoring can be found in Lafuente et al. [5]. Below two examples of such methods are reviewed and their advantages and shortcomings are detailed. Sughimoto et al. [10] designed a machine learning model to predict Lactate levels continuously between blood draws, using previous Lactate levels in combination with other diagnostic details. While this model allows for less frequent measurement and continuous estimation of Lactate levels, it does not account for 'black swan events', where Lactate levels may spike unexpectedly due to rapid, unexpected, negative developments in the patient's health. Ming et al. [11] proposed a method in which a microneedle patch could be worn for continuous Lactate measurement in a non-clinical context. The findings of the study were encouraging, with continuous measurements being successfully taken from the patch. However, the use of venous Lactate levels can only be used in place

of arterial Lactate levels accurately below 2 mmol/l, hence making it unsuitable for clinical settings where patients with critical illness may exceed this level [12].

2.3. Spectroscopy for Lactate Measurement

An alternative approach to the measurement of Lactate levels is through the use of spectroscopy. Here, light signals are used to infer Lactate levels, hence providing the opportunity for continuous non-invasive estimation based on a series of Lactate-correlates. Previous studies have employed Raman spectroscopy for the quantification of Lactate levels [13, 14, 15]. Raman spectroscopy involves illuminating a sample with a laser and analyzing the wavelength shifts in the scattered light. These shifts occur due to interactions with molecular vibrations within the sample, providing a "fingerprint" that can identify specific molecular structures. However, these results are limited to in-vitro and ex-vivo measurements, with in-vivo implementation remaining challenging. A different approach is through the use of NIR. NIR is an absorption spectroscopy method which records which parts of the light are transmitted through a sample. Two previous studies employing NIR for Lactate measurement were conducted by Budidha et al. [3] and Mamouei et al. [16]. The work by Budidha et al. [3] emphasised that the deep penetration of human tissue possible through NIR, as well as its high predictive values in in-vitro samples, recommends NIR for the continuous measurement of Lactate in an in-vivo setting. Furthermore Mamouei et al. [16] emphasises that NIR, in combination with Mid-Infrared visible and Ultraviolet optical spectroscopy has the potential to measure Lactate, albeit only indirectly for the in-vivo setting.

3. Methodology

This section first describes the problem at hand and then details the datasets and the necessary pre-processing. The machine learning employed models, explainability methods and performance metrics are then introduced. For the sake of reproducibility, the code has been made available[1].

3.1. Problem description

The goal of this study is to identify the optimal path length measurement to use for measuring Lactate concentration using NIR spectroscopy. Path length refers to the distance light travels through the sample inside a cuvette. As light absorbance varies with the concentration of a compound, the ideal path length depends on the compound's concentration. The relationship between absorbance and concentration is quantified by the Beer-Lambert law, where absorbance is calculated as the product of the molar absorbance of the sample, the concentration of the sample, and the path length [17]. If concentration becomes excessively large or small, the linearity of its relationship with absorbance is affected. Path length choice can rectify this, with shorter path lengths restoring linearity at higher concentrations and vice versa. The difference in the average response depending on path length can be seen in Figure 1. This demonstrates the importance of this choice, as the absorbance patterns vary considerably between path lengths. This study tests which path lengths are best suited for the estimation of Lactate at varying concentrations.

The dataset employed for this study contains the results of 1195 samples with varying concentrations of Lactate in a phosphate-buffered saline (PBS) solution, with the cuvette thickness as a variable quantifying path length. The three cuvette measures were 2 mm, 5 mm and 10 mm. Four different Lactate concentrations were measured at all path lengths alongside a clear PBS solution to use as a baseline reference. The four concentrations are 1.3 mmol/l, 13 mmol/l, 130 mmol/l and 1300 mmol/l. The spectra of the samples are measured across 350 wavelengths within the NIR range from 1014.08 nm to 2580.07 nm. To normalise the spectra, the baseline reflectance of the PBS solution is subtracted from all spectra to account for its absorption index. After pre-processing, the dataset is divided by path length and going forward, the respective subsets will be referred to as 'Path Length 2', 'Path Length 5',

[1] https://github.com/CianK99/Lactate-Detection

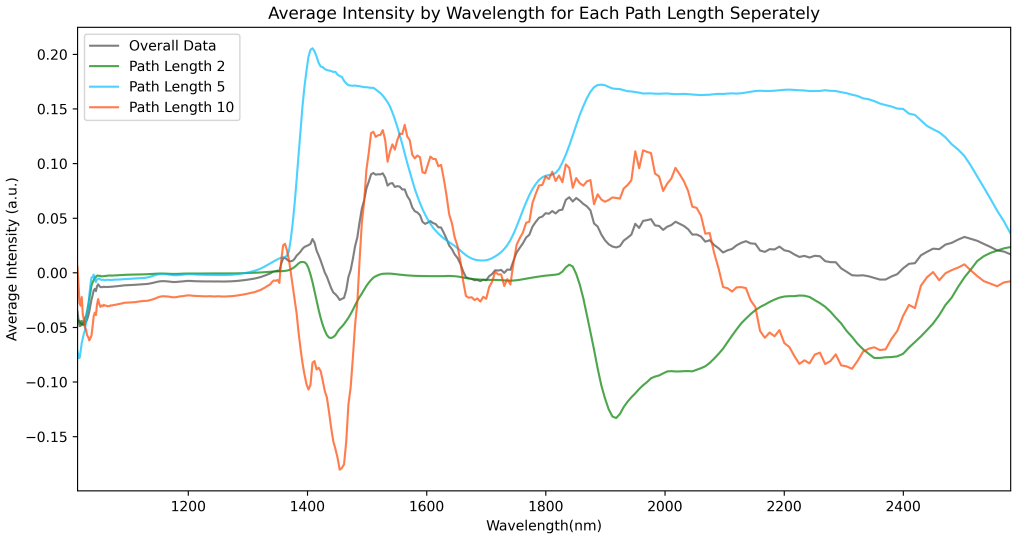


Figure 1: Plot of the average intensity by wavelength for each of the Path Length datasets

and 'Path Length 10'. Undersampling is implemented to account for data imbalance with respect to path length and concentration. After undersampling, a total of 1047 samples are retained.

3.2. Machine Learning Models

The machine learning models chosen for this project were Partial Least Squares (PLS), Least Absolute Shrinkage and Selection Operator (LASSO) and Random Forest (RF). Each will be described in turn below.

Partial Least Squares (PLS): PLS is a statistical method designed to model complex relationships by extracting latent structures from the data. It is particularly useful when predictors are highly collinear or when the number of predictors exceeds the number of observations, both of which are common problems with spectroscopic data. PLS creates latent variables as linear combinations of the original predictors and responses, aiming to maximise the covariance between these new variables while minimising the residual variance in the responses[18].

The basic form of the PLS model is:

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E}, \quad \mathbf{Y} = \mathbf{U}\mathbf{Q}^T + \mathbf{F}$$

Here, $\mathbf{X}$ represents the matrix of predictors and $\mathbf{Y}$ represents the matrix of responses. The matrices $\mathbf{T}$ and $\mathbf{U}$ contain the latent scores extracted from $\mathbf{X}$ and $\mathbf{Y}$, respectively. $\mathbf{P}$ and $\mathbf{Q}$ are matrices of loadings, which represent how the original variables relate to the latent components. $\mathbf{E}$ and $\mathbf{F}$ are residual matrices capturing the variability not explained by the model. PLS iteratively extracts latent variables from $\mathbf{X}$ and $\mathbf{Y}$ to optimise the shared variance. Once a latent component is extracted, the matrices $\mathbf{X}$ and $\mathbf{Y}$ are deflated to remove the explained variance. Finally, the response matrix $\mathbf{Y}$ is modelled as a linear combination of the latent variables:

$$\mathbf{Y} = \mathbf{T}\mathbf{C} + \mathbf{F}$$

where $\mathbf{C}$ is a matrix of regression coefficients for the latent variables. PLS is particularly suited for handling high-dimensional data, small sample sizes, and datasets with multicollinear predictors. The optimisation of shared variance between predictors and responses ensures the model is both predictive and interpretable.

Least Absolute Shrinkage and Selection Operator (LASSO): LASSO regression is a type of linear regression that includes a regularisation component. LASSO regression aims to enhance the prediction

performance and interpretability of the regression model by performing both variable selection and regularisation. The formula for LASSO is given by:

$$\text{minimise} \left\{ \frac{1}{2n} \sum_{i=1}^{n} (y_i - \beta_0 - \sum_{j=1}^{p} \beta_j x_{ij})^2 + \lambda \sum_{j=1}^{p} |\beta_j| \right\}$$

Here, the model aims to minimise the residual sum of squares subject to a penalty proportional to the absolute sum of the coefficients. The parameter λ controls the strength of the penalty; higher values lead to greater regularisation. This penalty term encourages the solution to have fewer non-zero coefficients, effectively conducting variable selection and promoting model simplicity and interpretability. This feature reduction is important as not all bands of light are needed for the detection of Lactate[19].

Random Forest (RF): RF is an ensemble model that enhances the performance of decision trees by combining multiple trees constructed from randomly selected subsets of data and features. Each tree in the forest operates independently, and their outputs are aggregated via averaging for regression and majority voting for classification to produce the final model prediction. This method improves generalisation and precision over individual decision trees and offers robust predictions and better handling of overfitting.

3.3. Evaluation

The evaluation metrics used to compare models and classifiers in this project are R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). For all metric evaluations, k-fold cross-validation is utilised, ensuring that every observation from the original dataset can appear in both the training and the test set. This method computes the model performance and aids in understanding model reliability and robustness across different subsets of data. For all the experiments in this project, 10 folds were used.

3.4. Explainability

Explainable Machine Learning (xAI) is essential in medical applications to ensure transparency and trust. Previous studies have shown that implementing explainability can increase practitioner trust and decrease model bias [20]. Furthermore, with the introduction of the EU AI Act, the implementation of explainability in the medical domain is a legal requirement to ensure system auditability and consequential accountability [21].

For explainability in this study, feature importance in the RF model using the Mean Decrease in Impurity (MDI) method was computed. The RF model is chosen for further analysis due to its strong predictive performance during initial testing on the whole dataset as seen in Section 4.1. This method assesses each feature's importance based on how much it reduces Gini impurity at each split in the decision trees. This reduction is summed across all trees and normalised to provide an overall importance score, with greater reductions in Gini Impurity leading to higher feature importance.

Feature Selection

Reducing the feature space through feature selection yields several advantages for spectroscopic data. First, it eases explainability as the high dimensionality and multicollinearity of data are decreased after feature selection. Second, the identification of important features becomes more straightforward, allowing the identification of links to underlying tissue components. Finally, the runtime and complexity of models may improve in a reduced feature space, allowing for faster data processing and improved medical implementation of developed algorithms. In this study, feature selection was implemented based on the feature importance scores computed by the random forest. Models were pruned by iteratively removing the least important features and recalculating importance after each reduction. The method involves removing the bottom 20% of features at each iteration up until it reaches the top 20 features, and from this point, one feature is removed at a time.

SHAP Analysis

To explain model prediction mechanisms and increase the auditability of the designed system, explainability is implemented through the use of SHAP. SHAP is a method for explaining the predictions of machine learning models based on concepts from cooperative game theory. It attributes the output of a model to its input features. SHAP values are additive, meaning the contributions of all features sum up to the difference between the average output and the actual prediction. This approach provides insights into how each feature influences the prediction, offering a powerful tool for understanding complex models. In the current study, this allows the identification of interfering features and the identification of candidate components for further testing can be identified.

4. Experimental Results

This section details the processes and results of this study's modelling. It first presents the results of initial testing on the full wavelength ranges in Section 4.1. The results of the feature selection, based on the best-performing model, are presented in Section 4.2 and a novel approach to creating a general wavelength range on their basis is presented in Section 4.3. Finally, the outcomes of the explainability analysis are presented in Section 4.4.

4.1. Initial Testing

Initial testing involved creating models on the full feature set for All Path Lengths, Path Length 2, Path Length 5 and Path Length 10. The initial prediction results of the implemented models are shown in Table 1. RF is found to perform the best on the All Path Lengths, Path Length 5 and Path Length 10 datasets, and is outperformed by LASSO and PLS in the Path Length 2 dataset.

Based on the initial testing results RF is identified as the best-performing model and selected for feature selection and additional testing. This selection is conducted on the basis of RF having the overall best performance as well as the most potential for improvement by feature selection. Feature selection is integrated into LASSO through L1 regularisation, effectively shrinking less important feature coefficients to zero, and in PLS through the calculation of latent variables. These built-in feature reductions limit the capacity of PLS and LASSO for further enhancement, as they have already optimised the feature space during training. As a result, RF holds more promise for iterative optimization and refinement and is used for all further testing in this paper.

Table 1
Comparison of model performance across the Path Length datasets

Dataset	Model	R2 Score	RMSE	MAE
All Path Lengths	**Random Forest**	**0.9778**	**76.38**	**18.84**
	LASSO	0.8746	196.88	147.47
	PLS	0.8730	198.16	148.21
Path Length 2	Random Forest	0.9697	65.05	14.23
	LASSO	0.9988	18.55	13.43
	PLS	**0.9993**	**14.05**	**10.81**
Path Length 5	**Random Forest**	**0.9977**	**18.96**	**4.87**
	LASSO	0.9948	35.12	25.51
	PLS	0.9938	34.47	23.00
Path Length 10	**Random Forest**	**0.9986**	**15.93**	**3.97**
	LASSO	0.9958	35.91	28.07
	PLS	0.9951	38.65	30.99

4.2. Modelling on Reduced Feature Sets

To explore model performance in a reduced feature set, feature selection is implemented based on the RF model and following the methodology specified in Section 3. At each stage of reduction, the train and test R^2, RMSE and MAE values are calculated. The set of features resulting in the highest test R^2 for each path-length dataset individually was selected and is referred to as the optimum feature set. As seen in Figure 2, the R^2 for the training set remains relatively high throughout the selection process, only decreasing when 3 or fewer features are maintained, which is likely a result of general overfitting in the training data. The testing R^2 is found to perform best when retaining 14 features, which was chosen as the final number. The optimum number of features for Path Length 2, Path Length 5 and Path Length 10 were 12, 14, and 26 features respectively. In Figure 3 the most important features of each of the path length datasets are plotted against the average response of all the samples. It is clear from this graph that the features most predictive of each respective path length dataset are distinctive, with some minor overlap between Path Length 10 and Path Length 5 between 1300 - 1400 nm. This also displays a high degree of clustering between the most important features. This gives promise to a more specified tool being possible.

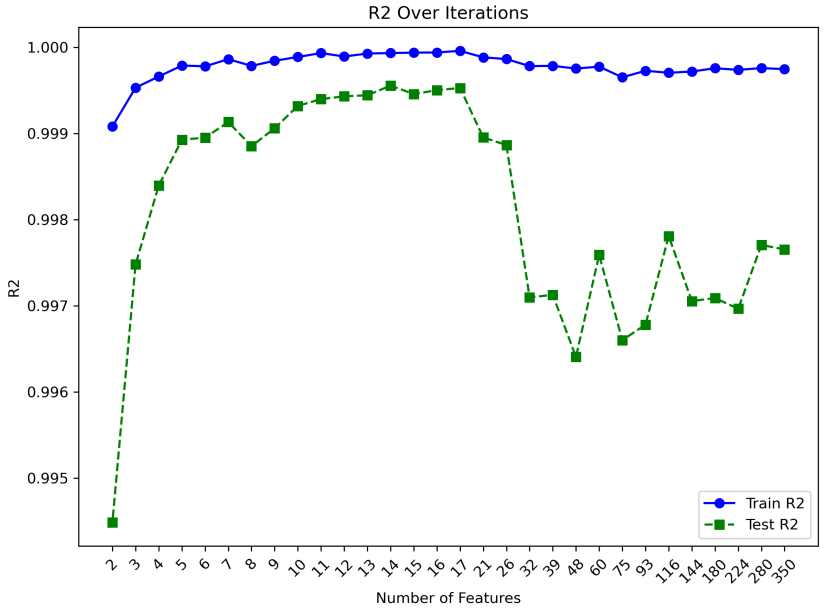


Figure 2: Train and Test R^2 Score for Path Length 5 Across Feature Count

The prediction results after feature selection are shown in Table 2. Performance increases considerably in the reduced feature set. Interestingly, Path Length 5 now surpasses the performance of Path Length 10 in all metrics. It is worth acknowledging these R^2 values are very high in all cases. While these high-performance values are encouraging and suggest a good fit for the data, they may be influenced by the lack of distinct Lactate levels in the dataset. This could result in an overestimation of model performance, as the model may be capturing patterns specific to the available data rather than generalising effectively to broader, more varied Lactate concentrations.

Table 2
Error Metrics Across All Path Lengths Based on the Optimum Set of Features

Dataset	R2 Score	RMSE	MAE
Path Length 2	0.9718	61.88	12.61
Path Length 5	**0.9996**	**6.21**	**1.33**
Path Length 10	0.9990	11.31	2.65

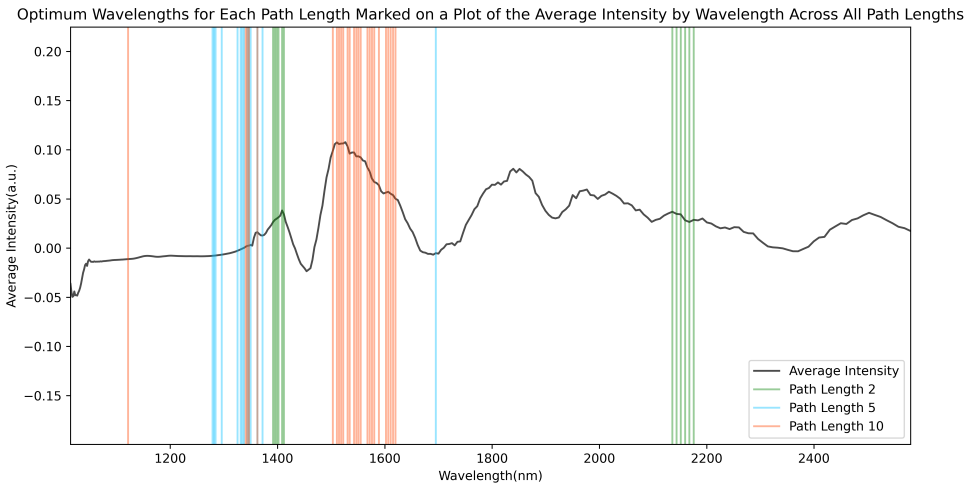


Figure 3: Optimum feature set for each path length plotted against the average response at each wavelength.

Table 3
RMSE at different levels of Lactate concentration using the optimal feature set.

Concentration	Path Length 2	Path Length 5	Path Length 10
0 mmol/l	25.34	**4.62**	41.74
1.3 mmol/l	9.21	11.72	**0.74**
13 mmol/l	0.79	**0.40**	0.82
130 mmol/l	39.42	**3.90**	4.90
1300 mmol/l	58.50	**2.07**	10.72

To investigate the concentrations of Lactate at which the model predicts with the least error, RMSE was computed for each concentration level. Table 3 presents the RMSE results for each concentration when using the optimal feature sets. Path Length 5 demonstrates the best performance across all concentrations, except at 1.3 mmol/l, where Path Length 10 outperforms it. It is important to note that the RMSE scores at zero concentration are poor for all path lengths. However, Lactate levels of zero are an artificial laboratory condition and do not reflect real-world scenarios, rendering this issue negligible. Additionally, while Path Length 5 performs best across the majority of the dataset, its lower performance at 1.3 mmol/l should be carefully considered, as this range is the most biologically significant for Lactate and should be prioritised in a medical context [22].

Table 4
Wavelength ranges and centres for each path length

Dataset	Wavelength Range (nm)	Centre (nm)	Feature Count
Path Length 2	1391 - 1412 2135 - 2176	1398 2160	13
Path Length 5	1278 - 1296 1325 - 1372 1695	1285 1347 1695	24
Path Length 10	1121 1340 - 1347 1362 1502 - 1620	1121 1343 1362 1565	34

4.3. Evaluating a General Wavelength Range

As the feature wavelengths used in the dataset are discrete and specific, a method to generalise this range for reproducibility and further experimentation was designed. By binning the wavelengths selected during feature selection, relevant sections of the spectrum can be identified and the size of the instruments needed for data collection could be reduced. The proposed approach takes the list of optimal features and defines a range around each identified feature. Preliminary testing determined that grouping wavelengths within a 15 nm window produced the best results. Any wavelengths within a 15nm window of each other were consolidated into the same range. A centre for each range was identified by computing the interpolated average of the wavelengths weighted by their feature importance. This means that if one of the edge wavelengths were most important for a defined range, the centre would be pulled closer to this important wavelength. These ranges and centres can be seen in Table 4. The new feature counts are presented alongside the new wavelength ranges.

Table 5
Error metrics across Path Lengths for the Optimum Generalised Wavelength Range.

Path Length	R2 Score	RMSE	MAE
Path Length 2	0.9689	65.14	13.29
Path Length 5	**0.9996**	**6.12**	**1.29**
Path Length 10	0.9992	9.98	2.31

When testing the RF model on these new feature ranges, the results presented in Table 5 are achieved. Path Length 5 and Path Length 10 perform better using the generalised range, with Path Length 2 performing marginally worse when compared to the results in Table 2. These ranges are more useful to inform the specifications of a tool than the wavelengths in the data. Using these feature ranges, the evaluation of RMSE by concentration yields a similar pattern to that observed with the optimal feature sets above. Specifically, Path Length 10 performs best at 1.3 mmol/l and Path Length 5 performs best for the others.

4.4. Implementing Explainability

As Path Length 5 has been found to perform best across the majority of Lactate concentrations, it was selected for further analysis. SHAP is used to analyse the feature usage of the Path Length 5 model to investigate the individual feature contributions to the prediction of the model. SHAP visualisations allow for easy identification of where mistakes occur in the model.

SHAP Beeswarm Plot Figure 4 shows the generated SHAP beeswarm plot with the most to least important wavelengths of the model being shown from top to bottom. The x-axis shows the degree to which a feature supports a prediction, with points left of the origin decreasing the prediction and vice versa. The colour range from blue to red shows the importance of the feature with red features contributing more. Most of the features have a positive relationship with the Lactate concentration, increasing the predicted value of Lactate. One exception is 1695.05, which tends to increase the prediction when its value is low. This implies that this wavelength differs from the remaining features included in the analysis. This is in keeping with expectations as all remaining wavelengths fall within 100 nm of each other (1278.72 nm - 1371.99 nm) and are likely modelling the same interaction in the data. As such it is reasonable to assume that the relative distance of the 1695.05 nm feature leads to its contrasting effect on the prediction.

SHAP Waterfall Plot Individual instances of both correct and incorrect predictions were further analysed to understand the cause of the mispredictions identified in the stacked force plot. An example of a correct and incorrect prediction, respectively, for a 1.3 mmol/l sample can be seen in Figures 5a and 5b. Figure 5a shows the correct prediction with all features detracting from the mean prediction of

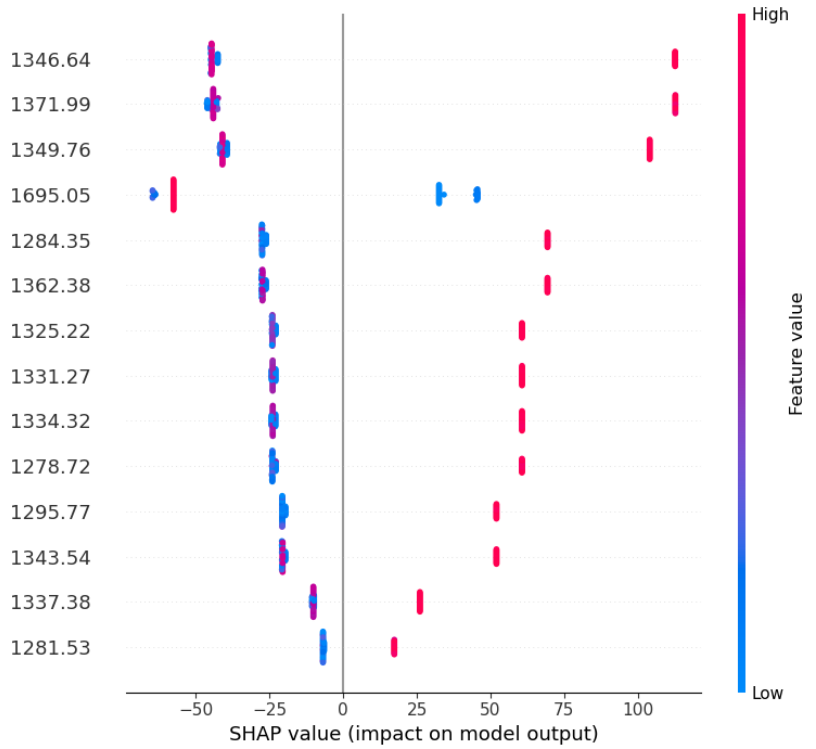
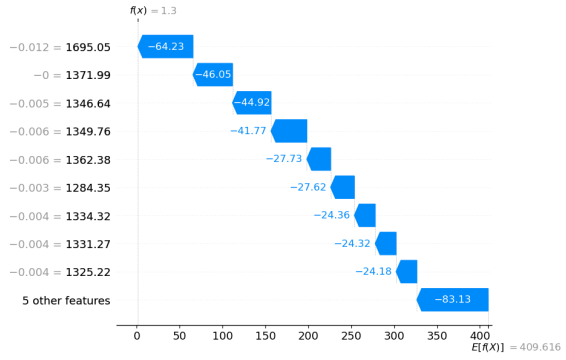
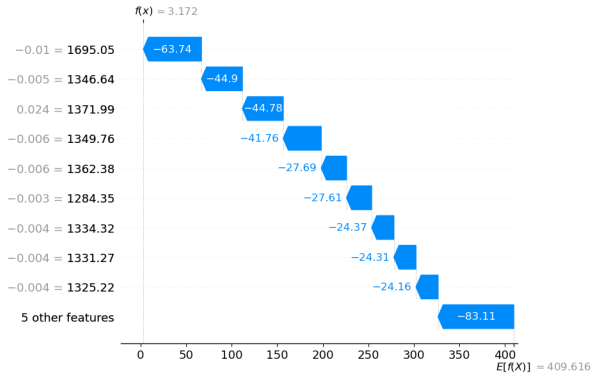


Figure 4: A SHAP beeswarm plot of the optimal features Path Length 5 model

409.616mmol/1. Figure 5b shows how the feature contributions differ for an incorrect prediction. It is clear from the plot that the influential feature is 1371.99 as this is the only feature differing from Figure 5a. This is a prime example of the importance of explainability as it permits the auditing of the system and the successful identification of potentially problematic features.



(a) A SHAP waterfall plot of a correct prediction.



(b) A SHAP waterfall plot of an incorrect prediction identified from the stacked force plot.

Figure 5: Comparison of SHAP waterfall plots for correct and incorrect predictions.

5. Conclusions

This study aimed to inform the design of a tool to measure blood Lactate levels non-invasively using NIR spectroscopy via machine learning. Through analysis of achieved prediction results, the best choice of path length for NIR measurements was identified. Feature selection was implemented to establish the best wavelengths and a general wavelength range was computed. Finally, through explainability

analysis, components interfering at given wavelengths can be identified and further testing for these compounds can be implemented accordingly. With RF, a 10 mm path length was found to perform best in the biologically relevant ranges of Lactate, whereas Path Length 5 performed best on average. Feature selection conducted based on the feature importance computed by the Random Forest model allowed for considerable improvement in algorithm performance. Explainability was implemented to audit the system and allowed the identification of important and potentially problematic features. In conclusion, this study demonstrates the ability of ML models to successfully estimate Lactate levels based on spectroscopic measurements and identify the best laboratory setup for such measurement. This contributes to the ongoing effort to establish non-invasive methods for the continuous measurements of Lactate in a hospital setting.

Acknowledgments

This work was conducted with the financial support of Science Foundation Ireland under Grant Nos. 12/RC/2289-P2 and 18/CRT/6223 which are co-funded under the European Regional Development Fund. This research was partially supported by the EU's Horizon Digital, Industry, and Space program under grant agreement ID 101092989-DATAMITE. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

References

[1] K. Rathee, V. Dhull, R. Dhull, S. Singh, Biosensors based on electrochemical lactate detection: A comprehensive review, Biochemistry and biophysics reports 5 (2016) 35–54.

[2] K. O. Alfarouk, D. Verduzco, C. Rauch, A. K. Muddathir, H. B. Adil, G. O. Elhassan, M. E. Ibrahim, J. D. P. Orozco, R. A. Cardone, S. J. Reshkin, et al., Glycolysis, tumor metabolism, cancer growth and dissemination. a new ph-based etiopathogenic perspective and therapeutic approach to an old cancer question, Oncoscience 1 (2014) 777.

[3] K. Budidha, M. Mamouei, N. Baishya, M. Qassem, P. Vadgama, P. A. Kyriacou, Identification and quantitative determination of lactate using optical spectroscopy—towards a noninvasive tool for early recognition of sepsis, Sensors 20 (2020) 5402.

[4] M. Mamouei, K. Budidha, N. Baishya, M. Qassem, P. Kyriacou, Comparison of wavelength selection methods for in-vitro estimation of lactate: a new unconstrained, genetic algorithm-based wavelength selection, Scientific Reports 10 (2020) 16905.

[5] J.-L. Lafuente, S. González, C. Aibar, D. Rivera, E. Avilés, J.-J. Beunza, Continuous and non-invasive lactate monitoring techniques in critical care patients, Biosensors 14 (2024) 148.

[6] S. M. Lundberg, S.-I. Lee, A Unified Approach to Interpreting Model Predictions, in: I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, volume 30, Curran Associates, Inc., 2017.

[7] A. Poscia, D. Messeri, D. Moscone, F. Ricci, F. Valgimigli, A novel continuous subcutaneous lactate monitoring system, Biosensors and bioelectronics 20 (2005) 2244–2250.

[8] Z. Zhang, X. Xu, Lactate clearance is a useful biomarker for the prediction of all-cause mortality in critically ill patients: a systematic review and meta-analysis, Critical care medicine 42 (2014) 2118–2125.

[9] A. Mokline, A. Abdenneji, I. Rahmani, L. Gharsallah, S. Tlaili, I. Harzallah, B. Gasri, R. Hamouda, A. Messadi, Lactate: prognostic biomarker in severely burned patients, Annals of burns and fire disasters 30 (2017) 35.

[10] K. Sughimoto, J. Levman, F. Baig, D. Berger, Y. Oshima, H. Kurosawa, K. Aoki, Y. Seino, T. Ueda, H. Liu, et al., Machine learning predicts blood lactate levels in children after cardiac surgery in paediatric icu, Cardiology in the Young 33 (2023) 388–395.

[11] D. K. Ming, S. Jangam, S. A. Gowers, R. Wilson, D. M. Freeman, M. G. Boutelle, A. E. Cass,

D. O'Hare, A. H. Holmes, Real-time continuous measurement of lactate through a minimally invasive microneedle patch: a phase i clinical study, BMJ Innovations 8 (2022).

[12] S. A. Samaraweera, B. Gibbons, A. Gour, P. Sedgwick, Arterial versus venous lactate: a measure of sepsis in children, European Journal of Pediatrics 176 (2017) 1055–1060.

[13] I. Olaetxea, E. Lopez, A. Valero, A. Seifert, Determination of physiological lactate and pH by Raman spectroscopy, in: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), 2019, pp. 475–481. Journal Abbreviation: 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC).

[14] I. Olaetxea, A. Valero, E. Lopez, H. Lafuente, A. Izeta, I. Jaunarena, A. Seifert, Machine Learning-Assisted Raman Spectroscopy for pH and Lactate Sensing in Body Fluids, Analytical Chemistry 92 (2020) 13888–13895. Publisher: American Chemical Society.

[15] N. C. Shah, O. Lyandres, J. T. Walsh, M. R. Glucksberg, R. P. Van Duyne, Lactate and Sequential LactateGlucose Sensing Using Surface-Enhanced Raman Spectroscopy, Analytical Chemistry 79 (2007) 6927–6932. Publisher: American Chemical Society.

[16] M. Mamouei, K. Budidha, N. Baishya, M. Qassem, P. A. Kyriacou, An empirical investigation of deviations from the Beer–Lambert law in optical estimation of lactate, Scientific Reports 11 (2021) 13734.

[17] D. F. Swinehart, The Beer-Lambert Law, Journal of Chemical Education 39 (1962) 333. Publisher: American Chemical Society.

[18] S. Wold, M. Sjöström, L. Eriksson, PLS-regression: a basic tool of chemometrics, Chemometrics and Intelligent Laboratory Systems 58 (2001) 109–130.

[19] D. Lafrance, L. C. Lands, D. H. Burns, In vivo lactate measurement in human tissue by near-infrared diffuse reflectance spectroscopy, The Second International Symposium on Two-dimensional Correlation Spectroscopy (2DCOS-II), University of Nottingham, UK, 21-23 August, 2003 36 (2004) 195–202.

[20] K. Rasheed, A. Qayyum, M. Ghaly, A. Al-Fuqaha, A. Razi, J. Qadir, Explainable, trustworthy, and ethical machine learning for healthcare: A survey, Computers in Biology and Medicine (2022) 106043.

[21] L. Edwards, The EU AI Act: a summary of its significance and scope, Artificial Intelligence (the EU AI Act) 1 (2021).

[22] D. Marikar, P. Babu, M. Fine-Goulden, How to interpret lactate, Archives of Disease in Childhood - Education and Practice 106 (2021) 167–171. Publisher: Royal College of Paediatrics and Child Health Section: Interpretations.