sensors

Article

Photothermal Radiometry Data Analysis by Using Machine Learning

Perry Xiao* and Daqing Chen

School of Engineering, London South Bank University; xiaop@lsbu.ac.uk
* Correspondence: xiaop@lsbu.ac.uk; Tel.: +44 (0) 2078157569

Abstract: Photothermal techniques are infrared remote sensing techniques that have been used for biomedical applications as well as industrial non-destructive testing (NDT). Machine Learning is a branch of artificial intelligence, which includes a set of algorithms for learning from past data and analyzing new data without being explicitly programmed to do so. In this paper, we first review the latest development of Machine Learning and its applications in photothermal techniques. Next, we present our latest work on Machine Learning for data analysis in Opto-Thermal Transient Emission Radiometry (OTTER), which is a type of photothermal techniques that has been extensively used in skin hydration, skin hydration depth profiles, skin pigments, as well as topically applied substances skin penetration measurements. We have investigated different algorithms such as Random Forest Regression, Gradient Boosting Regression, Support Vector Machine (SVM) Regression, Partial Least Squares Regression, as well as Deep Learning Neural Networks Regression. We first introduce the theoretical background, then illustrate its applications with experimental results.

Keywords: photothermal techniques, skin hydration, machine learning, deep learning, regression, classification;

1. Introduction

Photothermal techniques [1] are infrared remote sensing techniques that have been used for biomedical applications as well as industrial non-destructive testing (NDT). They can be dated back to the 1970s [2,3]. Photothermal techniques have since developed into different approaches, such as photothermal radiometry [4-7], photothermal tomography [8], photothermal imaging [9], photothermal radar [10], photothermal lens [11,12], photothermal cytometry [13] and so on. The main advantages of photothermal techniques lie in their non-invasive, remote-sensing, most importantly spectroscopic nature, which make photothermal techniques a potentially powerful tool in many industrial, agricultural, environmental and biomedical applications. Pawlak has highlighted the advantages of spectrally resolved photothermal radiometry measurements on semiconductor samples [14].

Machine learning [15,16] is a branch of artificial intelligence, which includes a set of algorithms for learning from the past data and analyzing the new data without being explicitly programmed to do so. Machine Learning can be generally divided into Supervised Learning, Un-supervised Learning, Semi-supervised Learning and Reinforcement Learning. Machine Learning has also been used in photothermal techniques recently. Verdel et al have developed a predictive model for the quantitative analysis of human skin using photothermal radiometry and diffuse reflectance spectroscopy [17,18], as well as a hybrid technique for characterization of human skin by combining Machine Learning and inverse Monte Carlo approach [19], and they made their Machine Learning model publically available through GitHub platform [20]. Ahmadi et al have developed a customized deep unfolding neural network, called Photothermal-SR-Net, for enabling super resolution (SR) imaging in photothermal radiometry [21]. Their model was based on an original

Citation: Lastname, F.; Lastname, F.; Lastname, F. Title. *Sensors* **2022**, *22*, x. https://doi.org/10.3390/xxxxx

Academic Editor: Firstname Lastname

Received: date
Accepted: date
Published: date

deep unfolding neural network (USRNet) [22]. Jawa et al have used Machine Learning and statistical methods for studying voids and photothermal effects of a semiconductor rotational medium with thermal relaxation time [23]. Kovács et al [24] have investigated Deep Learning approaches, based on U-net [25], for recovering initial temperature profiles from thermographic images in non-destructive material testing. There are also several studies using Deep Learning neural networks on infrared thermal images for machine health monitoring [26,27], as well as for pavement defect detection and pavement condition classification [28]. Qu et al have developed a low-cost thermal imaging with Machine Learning for non-invasive diagnosis and therapeutic monitoring of pneumonia [29]. Gajjela et al have leveraged mid-infrared spectroscopic imaging and deep learning for tissue subtype classification in ovarian cancer [30]. Li Voti et al have developed photothermal depth profiling by Genetic Algorithms [31]. Xiao et al have conducted a review of the field including photothermal depth profiling techniques [32,33].

In this paper, we use Machine Learning for analyzing our own measurement data by using Opto-thermal transient emission radiometry (OTTER), which is a type of photothermal radiometry technique that has been used in skin hydration, hydration depth profiling, skin pigments and trans-dermal drug delivery studies [32-39]. Compared with other technologies, OTTER has the advantages of non-contact, non-destructive, quick to make a measurement (a few seconds), and being spectroscopic in nature. It is also color blind, and can work on any arbitrary sample surfaces. It has a unique depth profiling capability on a sample surface (typically the top 20 μm)[33], which makes it particularly suitable for skin measurements. OTTER is information rich, however to analyze the signal and get the information is often difficult. To solve this problem, we proposed using Machine Learning for data analysis. Comparing conventional mathematical analysis, the main advantage of Machine Learning is that it can study and learn to analyze the data automatically, without the need of building complex mathematical models. We have investigated different algorithms such as Random Forest Regression, Gradient Boosting Regression, Support Vector Machine (SVM) Regression, Partial Least Squares Regression, as well as Deep Learning Neural Networks Regression. We first introduce the theoretical background, then illustrate its applications with experimental results.

2. Materials and Methods

This section describes the OTTER apparatus used, the machine learning algorithms developed, the volunteer information and the measurement procedures.

2.1. OTTER Apparatus

Figure 1 shows the schematic diagram of Opto-thermal transient emission radiometry (OTTER). It uses a pulsed laser (Er:YAG laser, 2.94μm, a few milli joules per pulse) as a heat source to heat the sample, an ellipsoidal mirror, and a fast infrared MCT (mercury cadmium telluride, InfraRed Associates, Inc., USA) detector to measure the consequent blackbody radiation increase of the sample [31,32]. The MCT detector used is the most sensitive infrared detector on the market. It is liquid nitrogen cooled and has a wide sensitivity spectrum range (3-15μm), high bandwidth (10MHz), and a purposely designed amplifier. A narrow band interference filter is also used in front of the MCT detector to select different detection wavelengths. By analyzing the OTTER signals, we can get the optical properties, thermal properties, and layered structure information from the sample. The selection of detection wavelength is achieved by using narrow bandpass mid-infrared interference filters. By selecting different detection wavelengths using different narrow band interference filters, we can measure different properties of the sample, for example, the water concentration information in skin (13.1 μm) or solvent concentration information within skin (9.5 μm). The OTTER detection depth is about 20μm. No other techniques can do depth-profiling in this range on in-vivo samples [32]. The OTTER skin measurements therefore should only be confined within Stratum Corneum, which is the outmost skin layer.

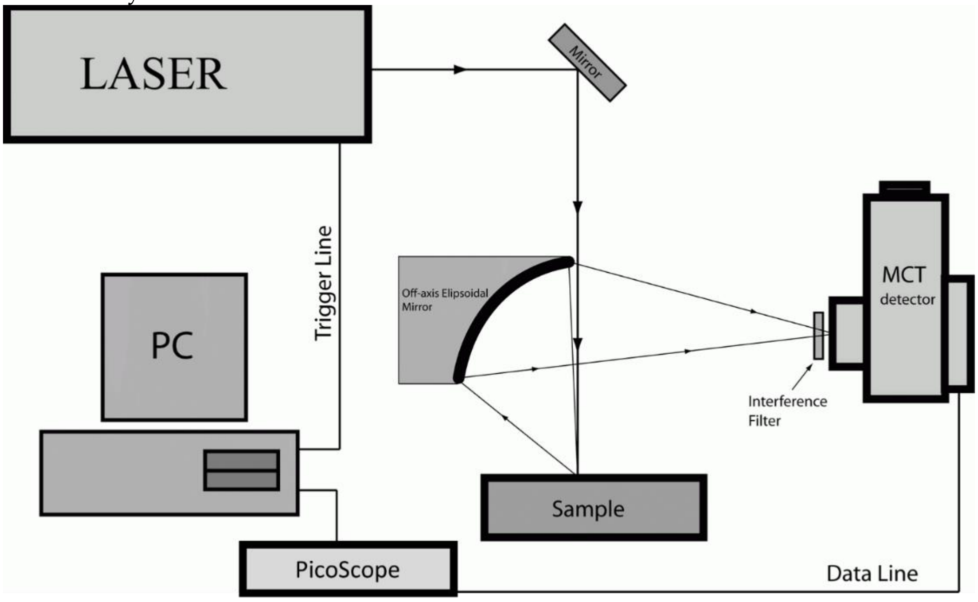


Figure 1. The schematic diagram of OTTER measurements [33].

For most OTTER measurements, it can be simplified as one dimensional semi-infinite problem [31]. For a semi-infinite, optically homogenous material, the OTTER signal can be generally expressed as [5-7],

$$S(t) = Ae^{t/\tau}\mathrm{erfc}\sqrt{t/\tau} \tag{1}$$

Where A is the amplitude of the signal, $\tau=1/(\beta^2 D)$ is the signal decay lifetime, β is the sample's emission absorption coefficient, and D is the sample's thermal diffusivity. By fitting the OTTER signal using Eq.(1), we can get the best fit β, and from β we can get the water content H in the sample, i.e. skin, hair, or nail [32].

$$H = \frac{\beta_w - \beta}{\beta_w - \beta_d} \tag{2}$$

Where β_w is the emission absorption coefficient of water, β_d is the emission absorption coefficient of dry sample. By using segmented least square (SLS) fitting, we can also get the water content at different depth, details are available elsewhere [33-35].

For a semi-infinite, optically non-homogenous material, the first assumption is that β is a linear function of depth [32],

$$\beta(z) = \beta_0 + w_\beta z \tag{3}$$

where β_0 is the absorption coefficient of the surface of the skin, and w_β is the gradient of the absorption coefficient. Then, the corresponding OTTER signal can be calculated as:

$$S(t) = A\left(\frac{2W\sqrt{t\tau}}{\sqrt{\pi}(2Wt+1)} + \frac{1}{\sqrt{2Wt+1}}e^{\frac{t/\tau}{2t/\tau+1}}\mathrm{erfc}\left(\frac{\sqrt{t/\tau}}{\sqrt{2Wt+1}}\right)\right) \tag{4}$$

Where $W = w_\beta D$ is the effective gradient, and $\tau=1/(\beta^2 D)$ is the signal decay lifetime. By fitting the OTTER signal with Eq. (4) we can get the skin surface absorption coefficient β_0 and the effective gradient W.

For most complex materials, where β is not a linear function of depth, we can use the enhanced segmented least squares (SLS) fitting algorithm [33], to get the skin hydration depth profiles in the following steps:

1. Load the OTTER signal
2. Find the starting point and end point of the signal

3. Fit the entire signal with Eq.(1) to get an average sample's emission absorption coefficient β.
4. Divide signal into 10 slices.
5. Fit the first slice of the signal with Eq.(1) to get the first β, then calculate the corresponding detection depth z.
6. Fit the first and the second slice of the signal with Eq.(1) to get the second β, then calculate the corresponding detection depth z.
7. Repeat step 6 until all the slices are used.

With the above algorithm, we can then plot β against depth z to get a depth resolved emission absorption coefficient. With Eq.(2) we can also interpret the plot as skin hydration levels at different depth (in micron meters), as shown in Figure 2.

As the we can see, the skin water hydration levels depth profiles are not linear, to simplify the problem, we fit the skin hydration depth profiles results in Figure 2 with Eq(3), to get simplified linear distribution of skin water content, as shown in Figure 3.

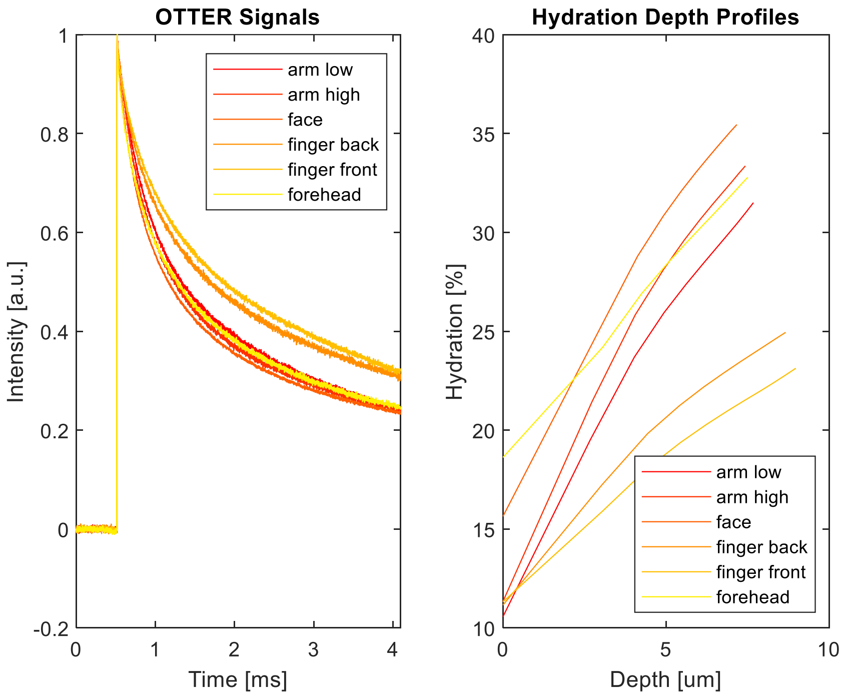


Figure 2. The typical OTTER measurement signals (left) and the corresponding hydration depth profiles (right) analyzed by using enhanced segmented least squares (SLS) fitting algorithm, of skin site at arm low, arm high, face, finger back, finger front and forehead.

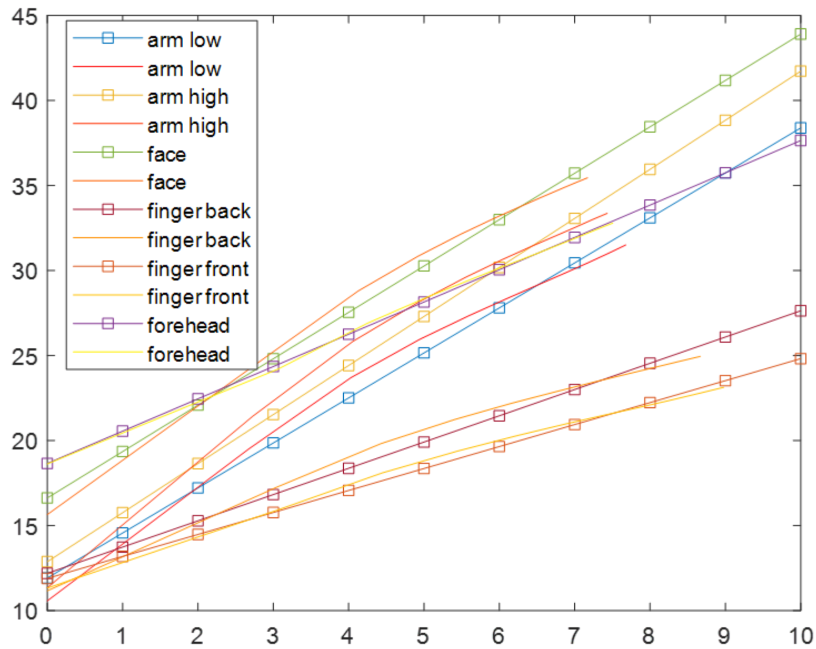


Figure 3. The simplified linear skin hydration distribution by fitting the skin hydration profiles in Figure 2 with Eq(3). The smooth curves are original profiles, the curves with squared markers are fitted straight line profiles.

2.2. Machine Learning Algorithms

From the history of Artificial Intelligence (AI) development [39], it can be roughly divided into three stages, artificial neural networks (1950s – 1970s), Machine Learning (1980s – 2010s) and Deep Learning (2010s – present). Generally speaking, Machine Learning is considered as a subset of AI, and Deep Learning is considered as a subset of Machine Learning. Machine Learning was originally developed in 1980s and consists a set of mathematical algorithms that can automatically analyze the data without being specifically programmed to do so. Machine Learning can be divided into Supervised Learning, Unsupervised Learning, Semi-supervised Learning and Reinforcement Learning [40]. In this paper, we will mainly focus on Supervised Learning, for the purpose of Regression and Classification. For Regression, we have investigated different algorithms such as Lasso (least absolute shrinkage and selection operator) [41], ElasticNet [42], Decision Tree [43], Support Vector Machine [44], Gradient Boosting [45], Linear Regression [46], Random Forest [47], K Nearest Neighbors [48], Extreme Gradient Boosting [49], Partial Least Squares(PLS) Regression [50], Voting Regression [51], Ridge regression with built-in cross-validation (RidgeCV) [52], as well as Deep Learning Neural Networks [53,54], to analyze the OTTER data. For Classification, we have investigated different Supervised Learning algorithms for classifying OTTER data.

Lasso Regression and Ridge Regress can be viewed as improved versions of Linear regression [55]. For linear regression, the cost function RSS (Residual Sum of Squares) can be written as:

$$RSS\ (W) = \sum_{i=1}^{N}(y_i - \hat{y})^2 = \sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{M}(w_j x_{ij})\right)^2 \quad (4)$$

Where y_i is the individual y values, N is total number of y values, w_j is the corresponding weight for the x_{ij}, M is the total number of x values. In order to minimize this cost, we generally use an algorithm called "gradient descent" [56]. Gradient descent means to calculate the partial differentiation of the above equation against weight w_j, and adjust weight in each iteration until it reaches the optimum stage. However, when the gradient is close to zero, the gradient descent algorithm will stop to work. This is commonly known as vanishing gradient [57].

Ridge Regression calculate the cost function RSS as the following, with sum of weight squares:

$$RSS\ (W) = \sum_{i=1}^{N}(y_i - \hat{y})^2 = \sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{M}(w_j x_{ij})\right)^2 + \lambda \sum_{j=1}^{M}\left(w_j\right)^2 \quad (4)$$

The λ is the calculation parameter. When we do the partial differentiation of the above equation, it is equivalent reduce the effect of weight, and can help in the event vanishing gradient problem.

Lasso Regression calculate the cost function RSS as the following, with the sum absolute value of the magnitude of weights:

$$RSS\ (W) = \sum_{i=1}^{N}(y_i - \hat{y})^2 = \sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{M}(w_j x_{ij})\right)^2 + \lambda \sum_{j=1}^{M}\left|w_j\right|^2 \quad (4)$$

Ridge Regression includes all (or none) of the features in the model, hence has the advantage of coefficient shrinkage and reducing model complexity.

Lasso Regression also has several benefits, apart from shrinking coefficients, it also performs feature selection. This is equivalent to exclude certain features from the model.

Elastic_Net Regression uses the linear combination of the penalty functions of Ridge Regression and Lasso Regression. By using this approach Elastic_Net can help on overfitting and underfitting problems.

Decision Tree and Random Forest are very popular Machine Learning algorithms. They are commonly used for classification. For Regression, the tree predicted outcome can be considered a real number, and it can contain different levels of depth, not enough layers of depth can result to underfit, and too many layers of depth can lead to overfit.

Support Vector Machine (SVM) is another popular Machine Learning algorithm, that is commonly used in Classification. For Regression, Support Vector Regression (SVR)'s goal is to find a function that approximates the relationship between the input variables and an output variable, with minimum error. SVR can handle non-linear relationships between the input variables and the target variable and makes it a powerful tool for analyzing complex problems.

Gradient boosting is a relatively new Machine Learning algorithm that is particularly suitable for tabular datasets. Gradient boosting is a type of ensemble methods where you create multiple weak models and in order to get better performance as a whole. It can find any nonlinear relationship between your model target and features and has great usability. It can also effectively deal with missing values, outliers, and high cardinality categorical values on your features. There are different versions of gradient boosting trees such as XGBoost or LightGBM.

Partial least squares regression (PLS regression) is a popular regression technique that is commonly used in spectral data analysis. It first projects the input data into a new space, then tries to fit the data by using a linear regression model in the new space. It is a quick, efficient and optimal regression technique. PLS regression is recommended in cases of regression where the number of explanatory variables is high, and likely multicollinearity among the variables [58,59].

Voting Regressions [60] belongs to the family of Ensemble Learning [61], which combines the predictions from multiple individual regression models to improve the performance. Voting Regressor can use simple averaging or weighted averaging to decide the final outcome.

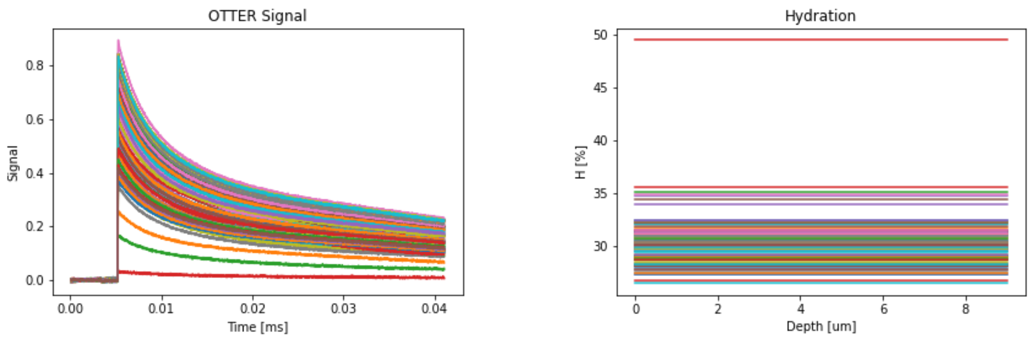
2.3. Measurement Procedure

All the measurements were performed on healthy volunteer (male and female, age 25 - 55), under normal ambient laboratory conditions of 20-21°C and 40-50% RH. The volunteer was instructed avoid excess water intake and the measurements were perform in the morning. The volar forearm skin sites used were initially wiped clean with ETOH/H2O (95/5) solution. The volunteer was then acclimatized in the laboratory for 20 minutes prior to the experiments.

3. Results and Discussions

3.1. Regression - Homogenous Model

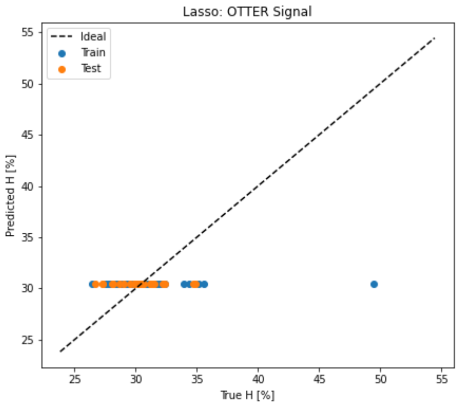
All the OTTER measurements are done and analyzed using the steps described in section 2.1. OTTER signals are analyzed by using Eq.(1) and the skin hydration are calculated by using Eq.(2). Figure 4 shows 97 OTTER skin measurement signals and the corresponding skin hydration levels in percentages calculated by using Eq.(1) and Eq.(2). These OTTER signals are were measured from the volar forearm of healthy volunteers, 20-30 years old, understand the standard laboratory condition (21°C, 40%).



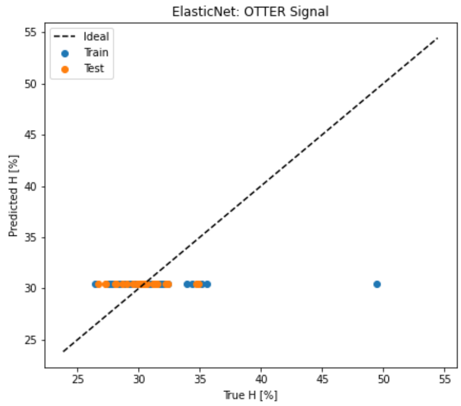
(**a**) (**b**)

Figure 4. The OTTER skin measurement signals **(a)** and corresponding skin hydration levels in percentages **(b)**.

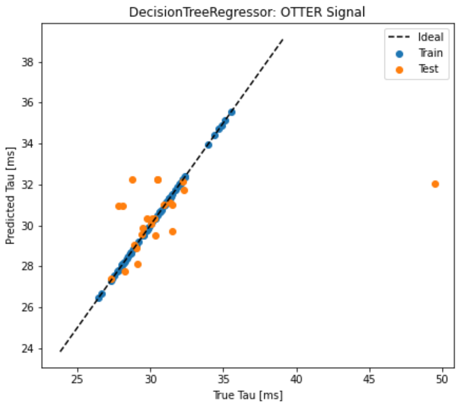
We randomly divided the above set of 97 measurement data into 75% as training dataset, and 25% as testing dataset and fed them into different Machine Learning algorithms models. Figure 5 shows the different Machine Learning Regression results. The results show that Lasso, Elasticnet, and Support Vector Machine Regressor (SVR) are almost completely not working in this case. Gradient Boosting, Extreme Gradient Boosting, as well as Decision Tree, work fine for the training data, but not very well for the testing data. Linear Regression gives the best results, followed by K Nearest Neighbors, Partial Least Squares Regression(PLS) and Random Forest. Deep Learning Neural Network, see Figure 6 for the architecture, was also used. It works fine for the training data, but not very well for the testing data.


Lasso: OTTER Signal
Ideal
Train
Test
Predicted H [%]
True H [%]

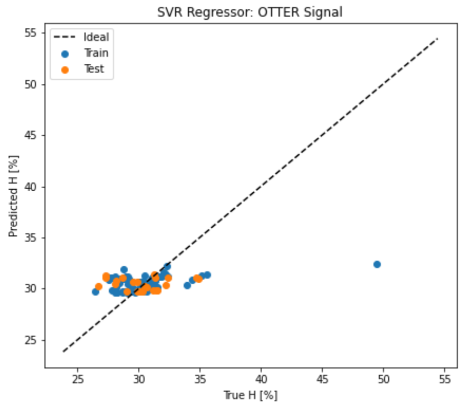
(A)


ElasticNet: OTTER Signal
Ideal
Train
Test
Predicted H [%]
True H [%]

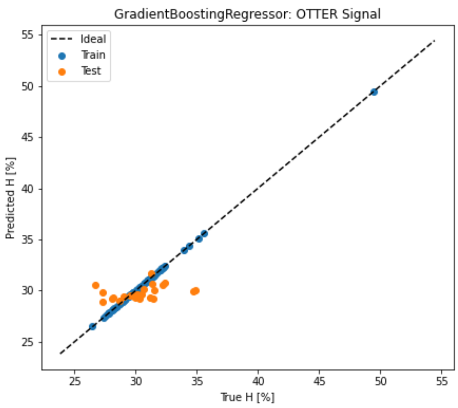
(B)


DecisionTreeRegressor: OTTER Signal
Ideal
Train
Test
Predicted Tau [ms]
True Tau [ms]

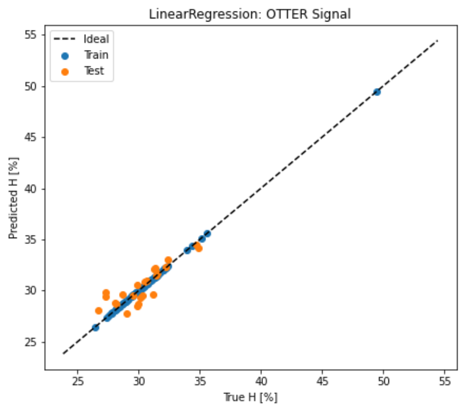
(C)


SVR Regressor: OTTER Signal
Ideal
Train
Test
Predicted H [%]
True H [%]

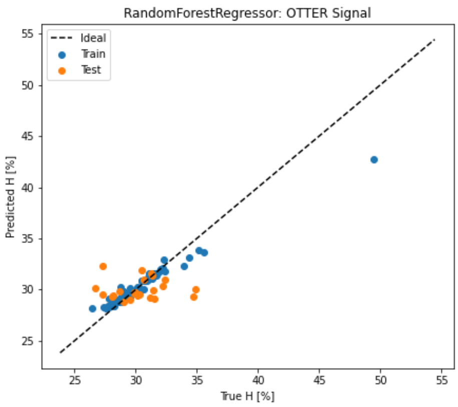
(D)


GradientBoostingRegressor: OTTER Signal
Ideal
Train
Test
Predicted H [%]
True H [%]

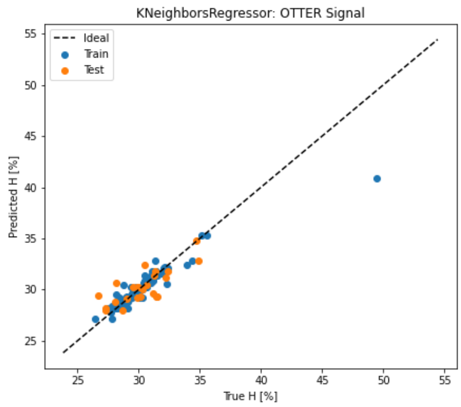
(E)


LinearRegression: OTTER Signal
Ideal
Train
Test
Predicted H [%]
True H [%]

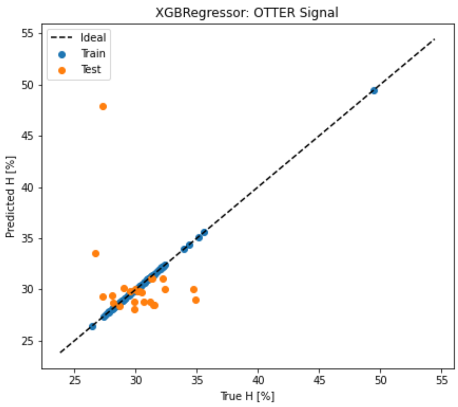
(F)



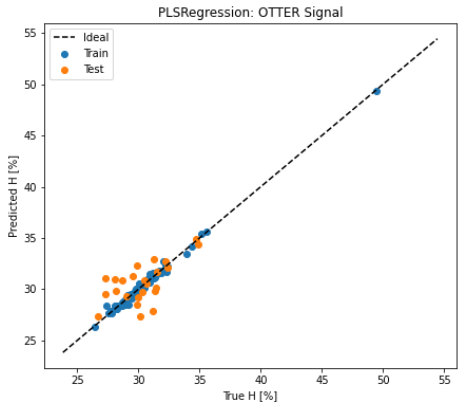
(G)



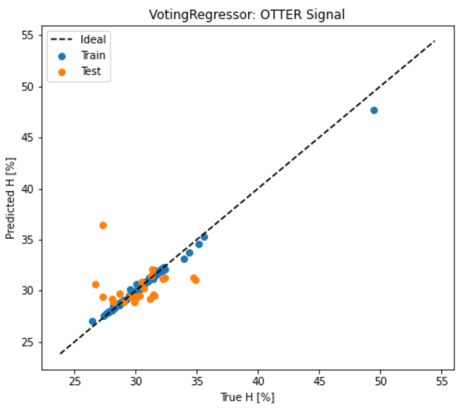
(H)



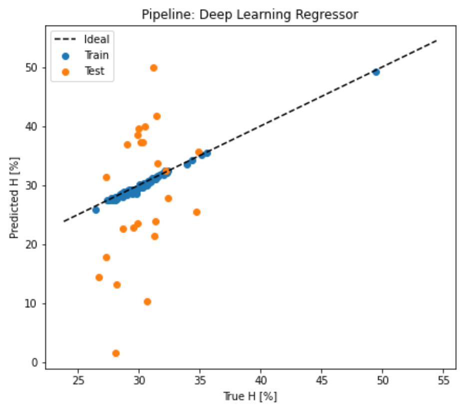
(I)



(J)



(K)



(L)

Figure 5. The Regression results of different Machine Learning algorithms models, (A) Lasso, (B) ElasticNet, (C) Decision Tree, (D) Support Vector Machine, (E) Gradient Boosting, (F) Linear Regression, (G) Random Forest, (H) K Nearest Neighbours, (I) Extreme Gradient Boosting, (J) Partial Least Squares(PLS) Regression, (K) Voting Regression, (L) Deep Learning.

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 256)	1048832
dense_1 (Dense)	(None, 256)	65792
dense_2 (Dense)	(None, 32)	8224
dense_3 (Dense)	(None, 11)	363

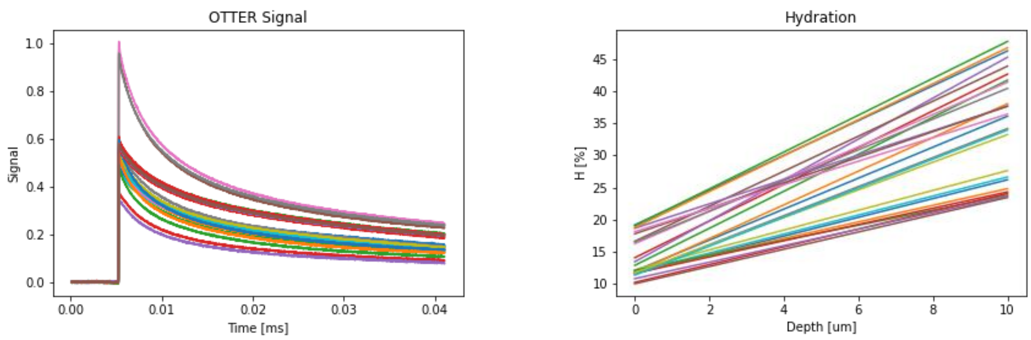
Total params: 1,123,211
Trainable params: 1,123,211
Non-trainable params: 0

Figure 6. The Deep Learning model architecture.

3.2. Regression - None-Homogenous Model

Figure 7 shows the same 97 OTTER skin measurement signals and the corresponding skin hydration depth distributions analyzed by using enhanced segmented least squares (SLS) fitting algorithm, then fitted with Eq(3).

Figure 8 shows the different Machine Learning Regression results. As you can see, again, Linear Regression gives the best result, it works well for both training data and testing data. RidgeCaV also gives a very good result, followed by PLS regression and K Nearest Neighbor. Deep Learning Neural Networks with the same architecture shown in Figure 6 was also used, again, it does not work very well.



(**a**) (**b**)

Figure 7. The OTTER skin measurement signals **(a)** and corresponding skin hydration [%] linear distribution depth profiles **(b)**.

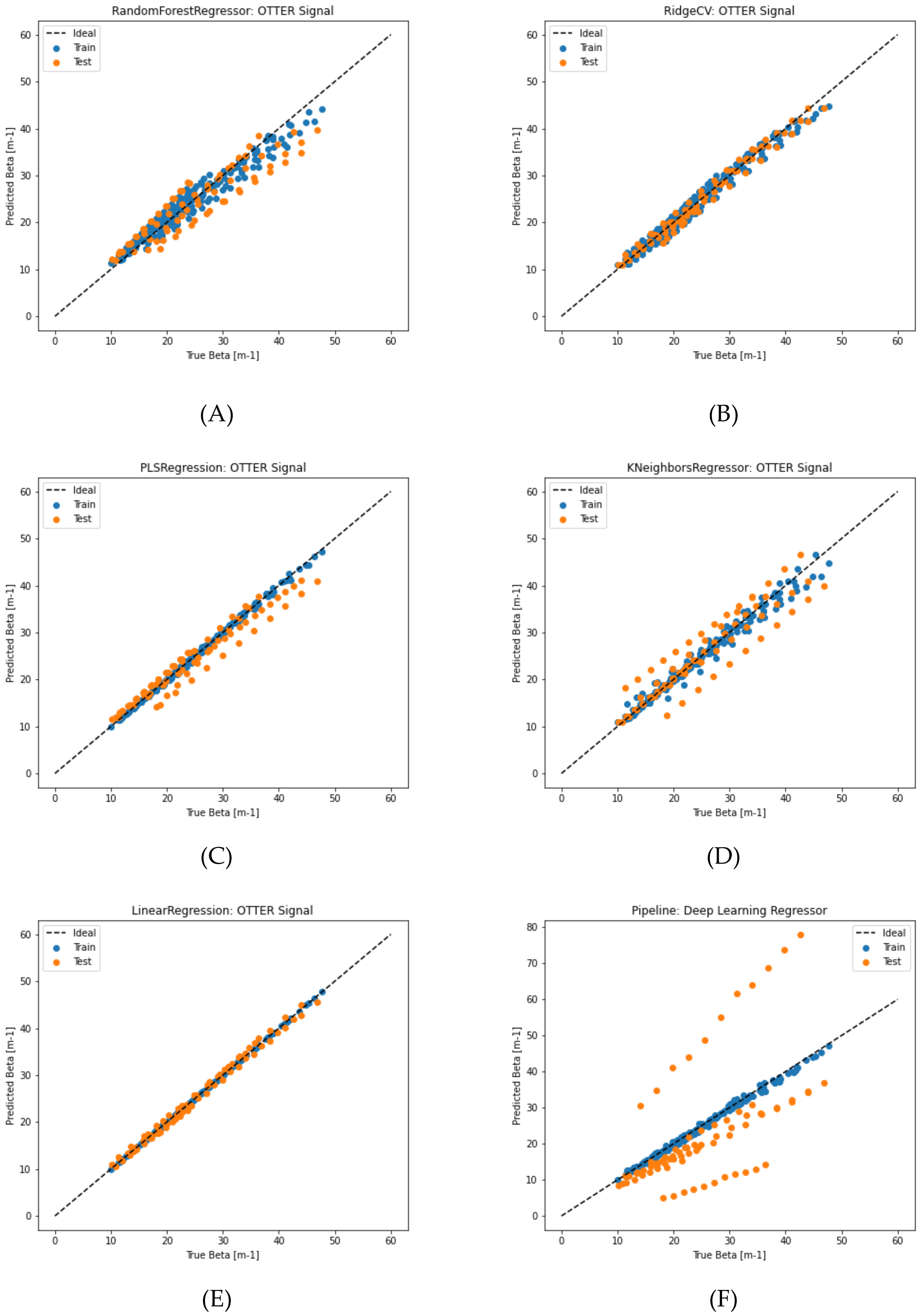
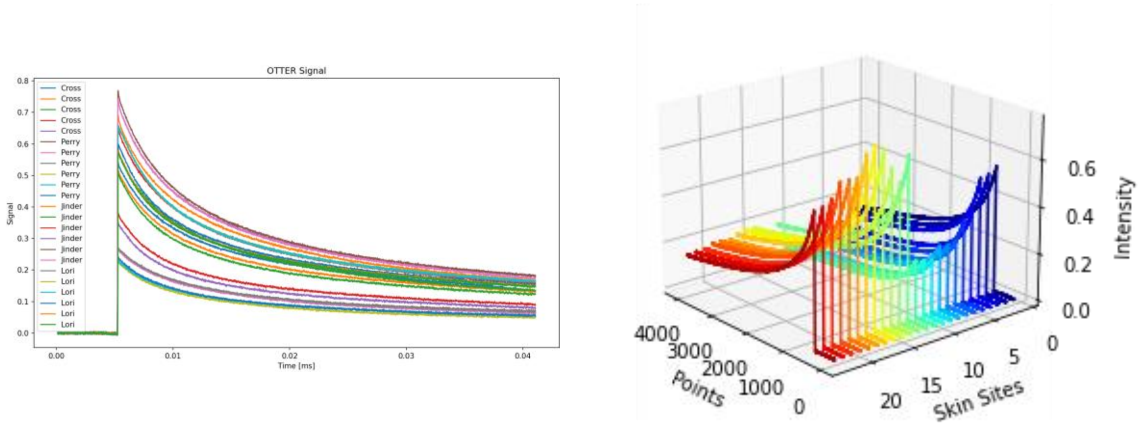


Figure 8. The Regression results of different Machine Learning algorithms models, (A) Random Forest, (B) RidgeCV, (C) Partial Least Squares(PLS) Regression, (D) K Nearest Neighbours, (E) Linear Regression, (F) Deep Learning Neural Networks..

3.3. Classification - Real OTTER Data

Figure 9 shows 20 OTTER signals of 4 different healthy volunteers (male and female, aged 25 - 55 years old) on the volar forearm, each volunteer has 5 measurement signals and volunteers are classified as 1, 2, 3, and 4.



(**A**) (**B**)

Figure 9. The 20 OTTER signals of 4 different volunteers on the volar forearm **(A)** and the corresponding 3D presentation **(B)**.

The 20 OTTER signals were then randomly divided into a 75% training dataset and a 25% testing dataset. The training dataset was used to train Machine Learning models, and trained Machine Learning models were then tested on the testing dataset. The following are classification results, as shown in Table 1. Accuracy means how many percentage of data that a model predicted correctly. Logistic, Ada Boost, and Gradient Boost give the best results, which achieved 100% accuracy for training data and 100% accuracy for testing data. The Deep Learning Neural Networks model based on the architecture shown in Figure 7, also performs well and reached 88.2% for training data and 83.3% for testing data.

Table 1. The classification accuracy results for Logistic, Naïve Bayes, SVC, Random Forest, Bagging Classifier, Ada Boost Classifier and Gradient Boosting Classifier.

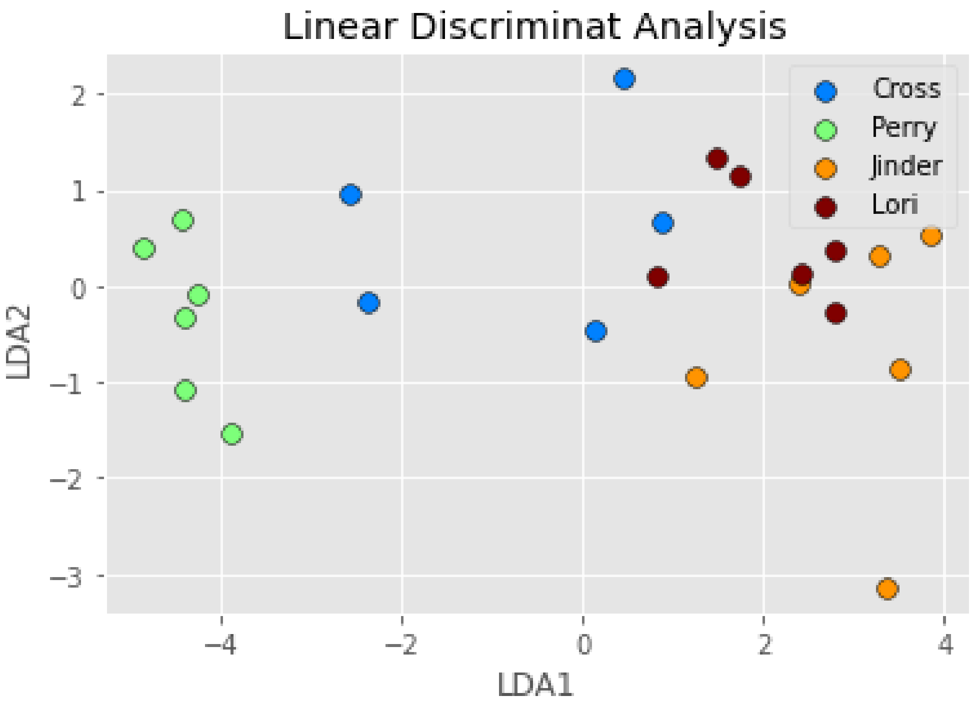
Models	Accuracy (Training) [%]	Accuracy (Test) [%]
Logistic	100.0%	100.0%
Naive Bayes	100.0%	83.3%
SVC	82.4%	83.3%
Random Forest	100.0%	83.3%
Bagging	70.6%	66.7%
Ada Boost	100.0%	100.0%
Gradient Boost	100.0%	100.0%
Deep Learning	88.2%	83.3%
LDA	82.4%	83.3%

Linear Discriminant Analysis (LDA) [62] and Principal Component Analysis (PCA) [63] are two related Machine Learning Algorithms for dimensionality reduction before later classification. LDA projects the data into a lower dimensioned space to separate the data better into different classes and to reduce computational costs, whilst PCA aims to project the data into new axis (called components), to maximize the variance. LDA first calculates the mean and covariance matrix for each class in the data, then calculates the scatter matrix between classes and that of within each class. The goal is to find a projection that can maximize the ratio of the scatter matrix between classes and that of within each class. PCA first centers the data around its mean, then finds the eigenvectors and eigenvalues of the covariance matrix, which are then used to project the data onto a lower-dimensional space. The eigenvectors specify the directions of maximum variance, and eigenvalues specify the corresponding amount of variance. The number of principal components represents the amount of variance we want to retain. Typically, we choose the number of principal components that is enough to explain a certain percentage of the total variance in the data.

Figure 10 shows the LDA plot of the first two components of the 20 OTTER signals of 4 different volunteers on the volar forearm. The results show that LDA can reasonably separate the OTTER signal from different volunteers effectively, the classification results show that LDA can reach 82.4% accuracy on training data and 83.3% accuracy on testing data.

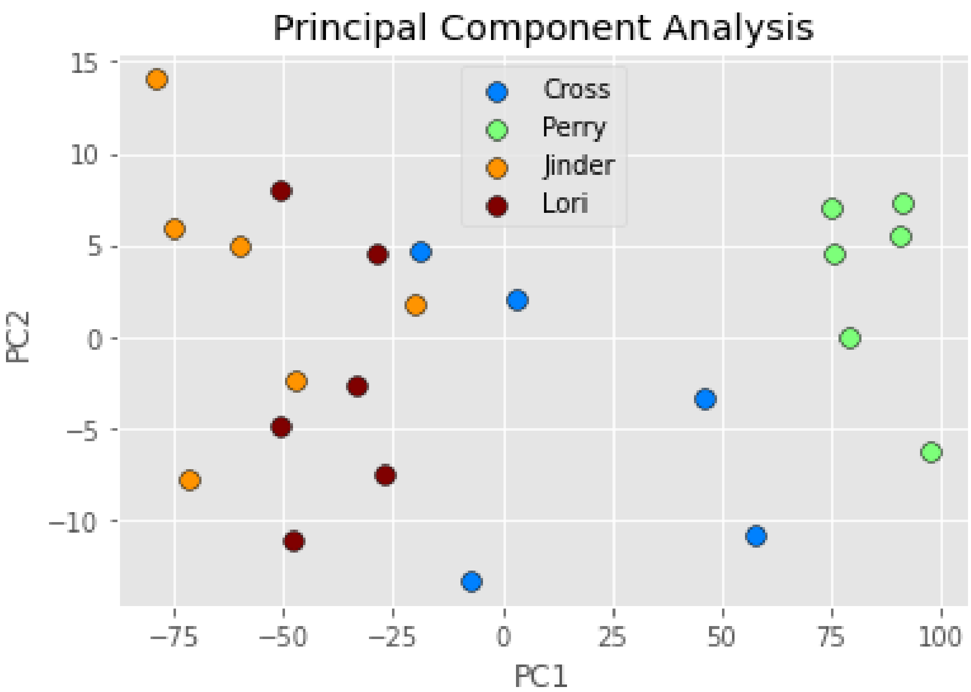
Figure 11 shows the PCA plot of the first two components of the 20 OTTER signals of 4 different volunteers on the volar forearm. The results show that PCA can also reasonably separate the OTTER signal from different volunteers effectively. By applying Random Forest Classifier on PCA results, we can also achieve 100% accuracy was achieved on classifying training data and 100% accuracy on classifying testing data.

With SHAP (SHapley Additive exPlanations) [63] values we can also evaluate the importance of each feature, and how it affects each final prediction. SHAP is originally a game theoretic approach that measures each player's contribution to the final outcome, and now has been widely using in Machine Learning to analyze the feature importance. In Machine Learning, each feature is assigned an important value representing its contribution to the model's output. By plotting the features according to their importance values, we can understand which are the most important features and which are the least important features. SHAP values can be used to interpret any machine learning model, such as Linear regression, Decision trees, Random forests, Gradient boosting models, and Neural networks and so on. Figure 12 shows the important features for OTTER data classification. As we are using OTTER signal data values as features, features 0, 1, 2, 3, 4 are the first four data points of the OTTER signal. This means that for classification, the early part of the signal is more important than the later part of the signal.



355

Figure 10. The LDA plot of the first two components of the 20 OTTER signals of 4 different volunteers on the volar forearm. 356 357



358

Figure 11. The PCA plot of the first two components of the 20 OTTER signals of 4 different volunteers on the volar forearm. 359 360

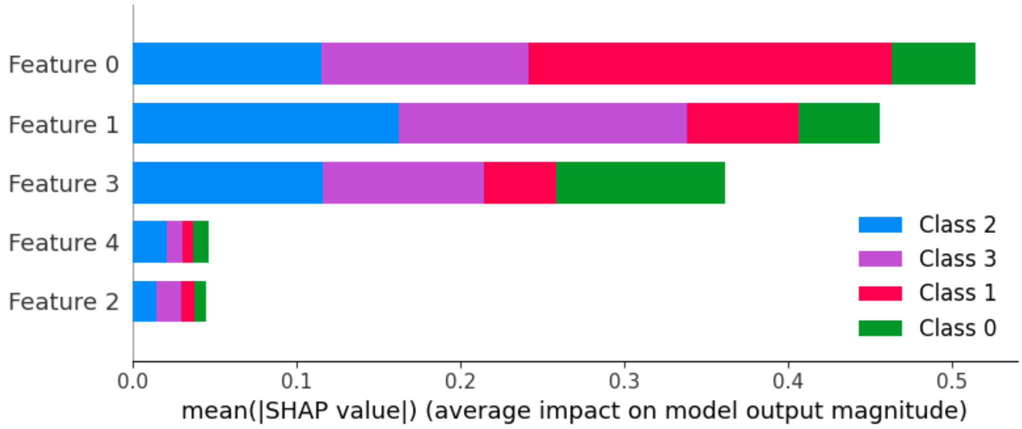


Figure 12. The most important features according to SHAP values.

As for the future work, we can further improve the classification accuracy in two ways, fine tuning model hyper parameters [65] and using Voting Classifier [66].

For most Machine Learning models, they have many hyper-parameters, and choosing the correct values for the hyper-parameters can have a good impact for the prediction accuracy. Take SVM (Support Vector Machine) for example, it can have the following hyper-parameters, C: the regularization parameter, kernel: the kernel type ('linear', 'poly', 'rbf', 'sigmoid', 'precomputed', or a callable) to be used in the algorithm, degree: the degree of the polynomial kernel function ('poly') and ignored by all other kernels, the default degree value is 3, gamma: the kernel coefficient for 'rbf', 'poly', and 'sigmoid'. If gamma is 'auto', then 1/n_features will be used instead. There can be several ways to find the best hyper-parameter values. The simplest one is exhaustive grid search, i.e. search all possible combinations. As you can see, this touch is comprehensive, but could be very time-consuming. An alternative approach is randomized parameter optimization, in which you first randomized the hyper-parameter values, then perform searching for the optimized values.

A voting classifier is a machine learning model that improves the classification accuracy by using a collection of models and predicts the results based on the largest majority of votes. It averages each classifier's results into the voting classifier. There are two different types of voting classifiers: Hard Voting and Soft Voting. Hard Voting predicts output with the highest majority of votes. Soft Voting averages the probabilities of the classes determine which one will be the final prediction.

5. Conclusions

We have investigated a range of Machine Learning algorithms for analysing our opto-thermal transient emission radiometry (OTTER) signals. For regression, we have investigated the OTTER signals using both homogenous model and non-homogenous model. For homogeneous model, the results show that Lasso, Elasticnet, and Support Vector Machine Regressor (SVR) are not working at all. Linear Regression gives the best results, followed by K Nearest Neighbors and Random Forest. For non-homogeneous model, Linear Regression gives the best result, followed by RidgeCV, PLS regressor and K Nearest Neighbors. In both cases, Deep Learning Neural Network model does not work well. For classification, Logistic, Ada Boost, and Gradient Boost give the best results, which achieved 100% accuracy for both training data and testing data. LDA and PCA can effectively separately the OTTER signals from different volunteers. By applying Random Forest Classifier on PCA results, we can also achieve 100% accuracy on classifying both training data and testing data. With SHAP values we can understand the importance of the different features. The results show that for classification, the early part of the OTTER signal is more important than the later part of the signal. For the future work, we can further improve the classification accuracy by using fine tuning model hyper parameters and Voting Classifier.

The main advantage of Machine Learning algorithms is that it can learn through training data and once trained, it can automatically analyze any unseen data, without the needing of complex mathematical models. The main disadvantage of Machine Learning algorithms is that many works like a blackbox, more work is needed for explainable Machine Learning algorithms.

Author Contributions: Conceptualization, P.X. and D.C.; methodology, P.X.; software, P.X.; validation, D.C.; formal analysis, P.X. and D.C.; investigation, P.X. and D.C.; resources, P.X.; data curation, P.X. and D.C.; writing—original draft preparation, P.X.; writing—review and editing, P.X. and D.C.; visualization, P.X.; supervision, P.X.; project administration, P.X.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: All the data generated during the study are available upon requests.

Acknowledgments: We thank London South Bank University and Biox Systems Ltd for the research support.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Tessier, G. (2009). Photothermal Techniques. In: Volz, S. (eds) Thermal Nanosystems and Nanomaterials. Topics in Applied Physics, vol 118. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-04258-4_13
2. Rosencwaig, A., and Gersho, A., Journal of Applied Physics 47, 64 (1976)
3. Rosencwaig, A., J. Opsal, W.L. Smith, et al.: Applied Physics Letters 46, 1013 (1985)
4. Tam, A.C., Sullivan, B., "Remote sensing applications of pulsed photothermal radiometry", Appl. Phys. Lett., 43, 333-335, 1983.
5. Imhof R.E., Birch D.J.S., Thornley F.R., Gilchrist J.R. and Strivens T.A., "Opto-thermal transient emission radiometry", J. Phys. E: Sci. Instrum., 17, 521-525, (1984).
6. Imhof, R.E., Zhang, B., and Birch, D.J.S., "Photothermal Radiometry for NDE", in Mandelis A (ed): Progress in Photothermal and Photoacoustic Science and Technology. PTR Prentice Hall, Englewood Cliffs (USA), vol. II, pp 185-236, 1994.
7. Imhof, R.E., McKendrick, A.D., and Xiao, P., "Thermal emission decay Fourier transform infrared spectroscopy", Rev. Sci. Instrum., 66, 5203-5213, 1995.
8. Thapa D, Welch R, Dabas RP, Salimi M, Tavakolian P, Sivagurunathan K, Ngai K, Huang B, Finer Y, Abrams S, Mandelis A, Tabatabaei N. Comparison of Long-Wave and Mid-Wave Infrared Imaging Modalities for Photothermal Coherence Tomography of Human Teeth. Ieee Transactions On Bio-Medical Engineering. PMID 35196221 DOI: 10.1109/TBME.2022.3153209
9. Tavakolian P, Mandelis A. Perspective: Principles and specifications of photothermal imaging methodologies and their applications to non-invasive biomedical and non-destructive materials imaging Journal of Applied Physics. 124: 160903. DOI: 10.1063/1.5044748
10. Sreekumar K, Mandelis A. Ultra-Deep Bone Diagnostics with Fat–Skin Overlayers Using New Pulsed Photothermal Radar International Journal of Thermophysics. 34: 1481-1488. DOI: 10.1007/S10765-013-1399-X
11. Folorunsho, O.G., Oloketuyi, S.F., Mazzega, E. et al. Nanobody-Dependent Detection of Microcystis aeruginosa by ELISA and Thermal Lens Spectrometry. Appl Biochem Biotechnol 193, 2729–2741 (2021). https://doi.org/10.1007/s12010-021-03552-6
12. Humberto Cabrera, Leja Goljat, Dorota Korte, Ernesto Marín, Mladen Franko, A multi-thermal-lens approach to evaluation of multi-pass probe beam configuration in thermal lens spectrometry, Analytica Chimica Acta, Volume 1100, 2020, Pages 182-190, ISSN 0003-2670, https://doi.org/10.1016/j.aca.2019.12.009.
13. Mikhail A. Proskurnin,1 Tatyana V. Zhidkova,2 Dmitry S. Volkov,1 Mustafa Sarimollaoglu,3 Ekaterina I. Galanzha,3 Donald Mock,4 Dmitry A. Nedosekin,3 Vladimir P. Zharov3 *, In Vivo Multispectral Photoacoustic and Photothermal Flow Cytometry with Multicolor Dyes: A Potential for Real-Time Assessment of Circulation, Dye-Cell Interaction, and Blood Volume, Cytometry Part A, 79A: 834847, 2011
14. Pawlak, M., Photothermal, photocarrier, and photoluminescence phenomena in semiconductors studied using spectrally resolved modulated infrared radiometry: Physics and applications J. Appl. Phys. 126, 150902 (2019), https://doi.org/10.1063/1.5114719
15. Machine Learning, https://en.wikipedia.org/wiki/Machine_learning (Accessed in September 2, 2023)

16. Xiao, P., Artificial Intelligence Programming with Python: From Zero to Hero, Wiley; 1st edition, ISBN-13 : 978-1119820864, 11 Mar. 2022.
17. Verdel N, Tanevski J, Džeroski S, Majaron B., Predictive model for the quantitative analysis of human skin using photothermal radiometry and diffuse reflectance spectroscopy. Biomed Opt Express. 2020 Feb 28;11(3):1679-1696. doi: 10.1364/BOE.384982. PMID: 32206435; PMCID: PMC7075612.
18. Verdel N, Tanevski J, Džeroski S, Majaron B., A machine-learning model for quantitative characterization of human skin using photothermal radiometry and diffuse reflectance spectroscopy, Proc. SPIE 10851, Photonics in Dermatology and Plastic Surgery 2019, 1085107 (26 February 2019); https://doi.org/10.1117/12.2509691.
19. Verdel N, Tanevski J, Džeroski S, Majaron B., Hybrid technique for characterization of human skin by combining machine learning and inverse Monte Carlo approach, in Clinical and Preclinical Optical Diagnostics II, Vol. EB101 of SPIE Proceedings (Optica Publishing Group, 2019), paper 11075_54.
20. SkinModel , https://github.com/jtanevski/SkinModel (Accessed in September 2, 2023).
21. Samim Ahmadi, Jan Christian Hauffen, Mathias Ziegler, Photothermal-SR-Net: A Customized Deep Unfolding Neural Network for Photothermal Super Resolution Imaging, https://arxiv.org/pdf/2104.10563v1.pdf (Accessed in September 2, 2023).
22. Deep unfolding network for image super-resolution, https://github.com/cszn/USRNet (Accessed in September 2, 2022).
23. Taghreed M. Jawa, Azhari A. Elhag, Tahani A. Aloafi, Neveen Sayed-Ahmed, F. S. Bayones, Jamel Bouslim, "Machine Learning and Statistical Methods for Studying Voids and Photothermal Effects of a Semiconductor Rotational Medium with Thermal Relaxation Time", Mathematical Problems in Engineering, vol. 2022, Article ID 7205380, 18 pages, 2022. https://doi.org/10.1155/2022/7205380.
24. Péter Kovács, Bernhard Lehner, Gregor Thummerer, Günther Mayr, Peter Burgholzer, and Mario Huemer, Deep learning approaches for thermographic imaging, Journal of Applied Physics 128, 155103 (2020); https://doi.org/10.1063/5.0020404
25. Pytorch-unet, https://github.com/jvanvugt/pytorch-unet (Accessed in September 2, 2023).
26. O. Janssens, R. Van de Walle, M. Loccufier and S. Van Hoecke, "Deep Learning for Infrared Thermal Image Based Machine Health Monitoring," in IEEE/ASME Transactions on Mechatronics, vol. 23, no. 1, pp. 151-159, Feb. 2018, doi: 10.1109/TMECH.2017.2722479.
27. M. Keerthi, R. Rajavignesh, Machine Health Monitoring Using Infrared Thermal Image by Convolution Neural Network, https://www.ijert.org/research/machine-health-monitoring-using-infrared-thermal-image-by-convolution-neural-network-IJERTCONV6IS07026.pdf (Accessed in September 2, 2023).
28. Chen, C.; Chandra, S.; Han, Y.; Seo, H. Deep Learning-Based Thermal Image Analysis for Pavement Defect Detection and Classification Considering Complex Pavement Conditions. Remote Sens. 2022, 14, 106. https://doi.org/10.3390/rs14010106
29. Qu Y, Meng Y, Fan H, Xu RX. Low-cost thermal imaging with machine learning for non-invasive diagnosis and therapeutic monitoring of pneumonia. Infrared Phys Technol. 2022 Jun;123:104201. doi: 10.1016/j.infrared.2022.104201. Epub 2022 May 14. PMID: 35599723; PMCID: PMC9106596.
30. Chalapathi Charan Gajjela, Matthew Brun, Rupali Mankar, Sara Corvigno, Noah Kennedy, Yanping Zhong, Jinsong Liu, Anil K. Sood, David Mayerich, Sebastian Berisha, Rohith Reddy, Leveraging mid-infrared spectroscopic imaging and deep learning for tissue subtype classification in ovarian cancer, https://arxiv.org/abs/2205.09285 (Accessed in September 2, 2023).
31. R. Li Voti, C.Sibilia, M.Bertolotti, Photothermal depth profiling by Genetic Algorithms and Thermal Wave Backscattering, International Journal of Thermophysics 26 1833-1848 (2005).
32. Xiao, P., The Opto-thermal Mathematical Modelling and Data Analysis in Skin Measurements, PhD thesis, London South Bank University, November, 1997.
33. Xiao P., "Photothermal Radiometry for Skin Research", Cosmetics, 3(1), 10, doi:10.3390/cosmetics3010010, 29 Feb 2016. 20.
34. X Zhang, C Bontozoglou, P Xiao, In Vivo Skin Characterizations by Using Opto-Thermal Depth-Resolved Detection Spectra, Cosmetics 6 (3), 54, 2020
35. P. Xiao, J.A. Cowen and R.E. Imhof, "In-Vivo Transdermal Drug Diffusion Depth Profiling - A New Approach to Opto-Thermal Signal Analysis", Analytical Sciences, Vol 17 Special Issue, pp s349-s352, 2001.
36. Xiao, P., Ou, X., Ciortea, L.I., Berg, E.P., Imhof, R.E., "In Vivo Skin Solvent Penetration Measurements Using Opto-thermal Radiometry and Fingerprint Sensor", Int J Thermophys (2012) 33:1787–1794. DOI 10.1007/s10765-012-1318-6.
37. Xiao, P., and Imhof, R.E., "Data Analysis Technique for Pulsed Opto-Thermal Measurements", UK Patent Application 0004374.5, 2000.
38. Xiao, P., and Imhof, R.E., "Apparatus for in-vivo Skin Characterization", UK Patent Application GB1014212.3, 2010.
39. Deep Learning, https://developer.nvidia.com/deep-learning (Accessed in September 2, 2023).
40. Geron, A., Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems，OReilly ISBN: 1492032646, 978-1492032649, 14 Oct. 2019.
41. Lasso (statistics), https://en.wikipedia.org/wiki/Lasso_(statistics) (Accessed in September 2, 2023).
42. Elastic net regularization, https://en.wikipedia.org/wiki/Elastic_net_regularization (Accessed in September 2, 2023).
43. Decision Tree, https://en.wikipedia.org/wiki/Decision_tree (Accessed in September 2, 2023).
44. Support Vector Machine, https://en.wikipedia.org/wiki/Support-vector_machine (Accessed in September 2, 2023).
45. Gradient Boosting, https://en.wikipedia.org/wiki/Gradient_boosting (Accessed in September 2, 2023).

46. Linear Regression, https://en.wikipedia.org/wiki/Linear_regression (Accessed in September 2, 2023).
47. Random Forest, https://en.wikipedia.org/wiki/Random_forest (Accessed in September 2, 2023).
48. K Nearest Neighbors, https://en.wikipedia.org/wiki/K-nearest_neighbors_algorithm (Accessed in September 2, 2023).
49. Extreme Gradient Boosting, https://en.wikipedia.org/wiki/XGBoost (Accessed in September 2, 2023).
50. Partial Least Squares Regression, https://en.wikipedia.org/wiki/Partial_Least_Squares_Regression (Accessed in September 2, 2023).
51. Voting Regression, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingRegressor.html (Accessed in September 2, 2023).
52. RidgeCV, https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.RidgeCV.html (Accessed in September 2, 2023).
53. Deep Learning Neural Networks, https://en.wikipedia.org/wiki/Deep_learning (Accessed in September 2, 2023).
54. Goodfellow, I., Bengio, Y., Courville, A., Deep Learning，MIT Press, 2016, ISBN: 0262035618, 978-0262035613.
55. Lasso and Ridge Regression in Python, https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial (Accessed in January, 25, 2024).
56. gradient descent, https://en.wikipedia.org/wiki/Gradient_descent (Accessed in January, 25, 2024).
57. Vanishing gradient problem, https://en.wikipedia.org/wiki/Vanishing_gradient_problem (Accessed in January, 25, 2024).
58. PARTIAL LEAST SQUARES REGRESSION (PLS), https://www.xlstat.com/en/solutions/features/partial-least-squares-regression (Accessed in January, 25, 2024).
59. What is partial least squares regression? https://support.minitab.com/en-us/minitab/21/help-and-how-to/statistical-modeling/regression/supporting-topics/partial-least-squares-regression/what-is-partial-least-squares-regression/ (Accessed in January, 25, 2024).
60. Voting Regressor, https://www.geeksforgeeks.org/voting-regressor/, (Accessed in January, 25, 2024).
61. Ensemble Learning, https://en.wikipedia.org/wiki/Ensemble_learning, (Accessed in January, 25, 2024).
62. Linear discriminant analysis, https://en.wikipedia.org/wiki/Linear_discriminant_analysis (Accessed in September 2, 2023).
63. Principal component analysis, https://en.wikipedia.org/wiki/Principal_component_analysis (Accessed in September 2, 2023).
64. SHAP document, https://shap.readthedocs.io/en/latest/, (Accessed in January, 25, 2024).
65. Tuning the hyper-parameters of an estimator, https://scikit-learn.org/stable/modules/grid_search.html, (Accessed in January, 25, 2024).
66. Voting Classifier, https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.VotingClassifier.html, (Accessed in January, 25, 2024).