Spectrometer as an Ubiquitous Sensor for IoT Applications Targeting Food Quality

Dina Samak[1], Sarah Jad[1], Nour Nabil[1], Amr Wassal[2], Marwan Torki[1]
[1]*Computer and System Engineering Department, Alexandria University, Alexandria 11432, Egypt*
{es-dina.samak1217, es-sarah.jad1217, nour.nabil, mtorki}@alexu.edu.eg
[2]*Computer Engineering Department, Cairo University, Cairo 12613, Egypt*
wassal@eng.cu.edu.eg

***Abstract*—In this paper, we present an IoT application targeting food quality. We utilize the spectrometer as a ubiquitous sensor to provide a Near-infrared (NIR) spectrum that can be used to develop regression models targeting milk quality. We solve a regression problem for different milk components such as Fat, Protein, Lactose, and Solids-Non-Fat. We apply a variety of comparative configurations of regression models, pre-processing techniques, wavelength selection, and feature extraction methods. Our study shows that for the milk data, non-linear models (GPR and SVR) are better choices when compared to linear models.**

I. Introduction

Measuring and analyzing food quality became demanding nowadays. NIR spectroscopy is being widely used in various fields analyzing chemical and physical properties of the substances concerned. Our aim is to provide a simple means to test and analyze food quality and to be able to measure the concentration of a specific component in a specified food. We collected our data using MEMS-based spectrometers [1], which are compact and low-cost Fourier Transform Near Infrared (FT-NIR) spectral sensors that measure the spectral response of the transmitted or reflected light from material samples for quantification, qualification or identification. The sensors offer performance comparable to laboratory based spectrometers, but at dramatically smaller sizes and lower costs.

In this work, we are introducing a study that relies on five main aspects: pre-processing of spectral data to eliminate non-linear variations, selection of wavelengths that contribute with useful information, feature extraction and selection to choose best features, identification of suitable calibration models using linear/nonlinear regression.

We used NIR liquid milk data sets in our project with wavelength varies approximately from (1300 - 2600 nm) with four components for regression: Fats, Protein, Lactose and Solids non-fats. In figure 1 we show an example of a single reading spectrum.The small size and robustness of the FT-NIR spectrometer used can enable the collection and measurement of milk samples inline in the milking station without interrupting the milking process and saving the hassle of transporting the samples to the laboratory for measurement as conducted in other experiments [2] [3].

This process can be part of an IoT platform where the measured data can be subsequently transmitted over wires or wirelessly to different compute platform to perform the Machine Learning Regression and present the results.

We based our implementation on a comparative interaction between wavelength selection, pre-processing and linear/nonlinear regression methods inspired by the benchmark study [4]. We used Orthogonal Signal Correction (OSC), Scaling and Extended multiplicative signal correction (EMSC) for the pre-processing methods. Stepwise Forward Selection for Wavelength Selection. First and second derivative, first integral and their combinations with the original data, moreover, we implemented multi-level covariance with and without overlapping for feature selection. And for the regression models we implemented Principal Component Regression (PCR), partial least squares (PLS), Support Vector Regression (SVR), and Gaussian Process Regression (GPR). We preferred classical regression models rather than neural network models due to the lack of data availability.

We used this bank of pre-processing techniques, feature selection methods and regression models in our implementation to obtain best results on the Milk Data components and to report minimum root mean square error (RMSE) and mean absolute percentage error (MAPE) on each milk component. Based on our implementation we found that the **SVR** model showed best results in our validation set in the Proteins and Lactose components, using scaling as a pre-processing method and concatenation of the integration of data readings and the original data for Proteins and no pre-processing and original data for Lactose. While **GPR** gave best results in Fats and Solids non Fats when using scaling as a pre-processing method and concatenating the integration of the reading with its original reading as feature extraction technique.

II. Related Work

Detecting food components is considered a hot topic in research. Spectroscopy and hyper-spectral imaging (HSI) play a major role in providing applications for analyzing the components of materials. It is used to analyze chemicals, soil, pharmaceuticals, detecting food adulterants and many other materials.

Spectroscopic data is considered a challenging problem as the data samples are highly correlated and studies use sophisticated statistical methods to identify the presence or

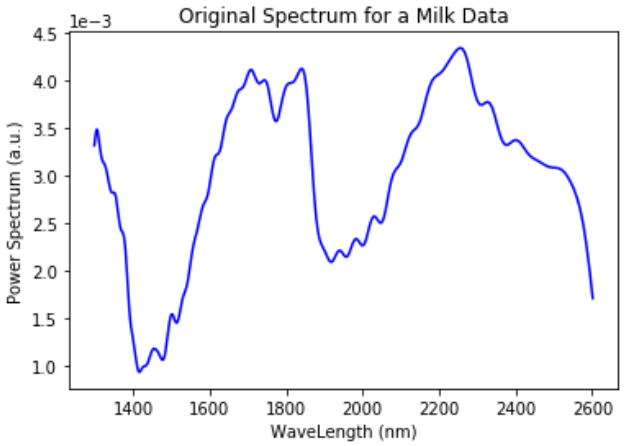

Fig. 1. Original spectrum of a sample milk reading.

the concentration of each component. In [19] they used NIR hyper-spectral imaging to detect peanut traces in wheat flour, they mainly used princioal component analysis (PCA) after normalizing and smoothing of the data as a pre-processing step to achieve remarkable results.

Another study [20] used visible and NIR spectroscopy for estimating soil organic carbon, soil nitrogen, enzymes and other soil properties. They applied in their experiments partial least squares regression (PLS) as a chemometric approach. Moreover, In [21] paper spray mass spectroscopy used partial least squares discriminant analysis to differentiate between beers brands.

A benchmark study [4], chose a bank of comparative studies with all combinations of pre-processing, wavelength selection, and regression models. for wavelength selection they used: step forward selection (SFS) and genetic algorithm optimization combined with partial least squares regression for spectral data (GAPLSSP). And for preprocessing techniques they used:orthogonal signal correction (OSC), extended multiplicative signal correction (EMSC) and optical path-length estimation and correction (OPLEC). And for regression models they used: partial least squares (PLS), least absolute shrinkage and selection operator (LASSO), least squares support vector machine (LS-SVM), and Gaussian process regression (GPR). They benchedmarked their findings on different materials including: pharmaceutical tablets, corn, sugarcane and a multi-component system consisting of water (H2O), deuterium oxide (D2O), ethanol (C2H5OH), and polystyrene particles. The study concluded that non-linear regressions with the addition of wavelength selection achieves superior performance over linear regressions without wavelength selection methods.

In this paper, we are working on liquid milk data, whereas it is not popular to find a similar dataset, however there is a study [7] that was conducted on fat content of milk using Raman. they used liquid milk, dried milk and) liquid milk contained in quartz cuvettes. In this study they also used PLS for modeling and gave good fat prediction with low root mean square errors and high correlation coefficients.

III. Machine Learning Techniques

A. Software

We implemented our *Machine Learning Library* using Python. We used scikit-learn [18] regression functions to build the GPR and SVR regression models and pychem library [14]to build the EMSC pre-processing method.

The offline experiments follow the sequential procedure of randomly splitting the whole data-set into the training set ,validation set and the testing set, then applying all the combinations of **wavelength selection pre-processing**, **feature selection methods**, and **regression models** on our training set. Firstly, we tune pre-processing hyper-parameters and models hyper-parameters on the training set to obtain best hyper-parameters that give least RMSE using k-fold cross validation algorithm, which is explained later, then fit the training set on the tuned pre-processing and model hyper-parameters. Afterwards, we obtain results by predicting the validation set on each model configuration, and compare the root mean square error (RMSE) and Mean Absolute Percentage Error (MAPE) of each configuration to choose the best model for each component. Then we apply the resulting pre-processing, wavelength selection, feature extracted and regression model configurations to predict the result of the test set. Finally, we report the RMSE and MAPE of the test set.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(P_i - O_i)^2} \tag{1}$$

$$MAPE = \frac{100\%}{n}\sum_{t=1}^{n}\frac{P_i - O_i}{O_i} \tag{2}$$

where P_i is the predicted value, while O_i is the original ground truth value.

B. Wavelength Selection

A technique used to decrease the number of NIR wavelengths used, this decreases the running time overhead and removes unwanted, redundant or noise data.

1) Step-Wise Forward Selection: It is used to compare regression models with and without a particular predictor, with a sequential addition of explanatory variables that continues until a stopping point based on the value of a certain criterion is reached [10].

Our implementation of SFS depends on the original least square regression model (OLS) to determine the least p-value of each wavelength in each iteration and append it to the selected wavelengths and continues until a stopping point based on the value of a threshold parameter that we adjusted to take only 10%-15% of the given wavelengths. We show the sample in figure 1 after SFS wavelength selection procedure in figure 2.

C. Pre-processing

We tried three different algorithms to pre-process our data to remove irrelevant and redundant information in the data-set, to try to improve the performance of the regression models afterwards.

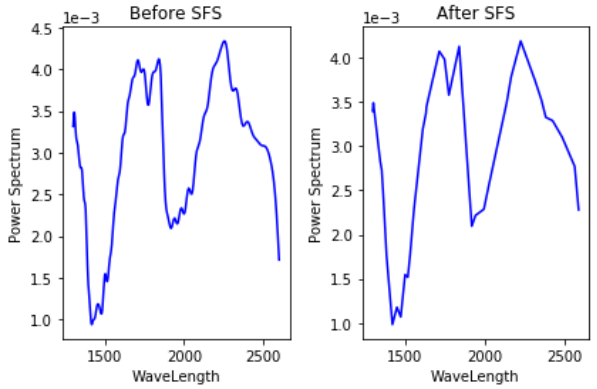

Fig. 2. SFS wavelength selection technique on a sample reading from milk data.

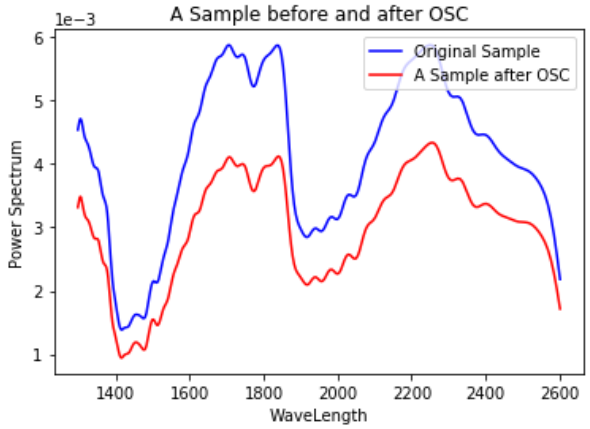

Fig. 3. OSC pre-processing technique on a sample reading from milk data.

1) Orthogonal Signal Correction: OSC [8] [9] was introduced in order to avoid removal of information that is important for prediction. The goal with OSC is to construct a filter that removes from the spectral matrix X only the part that is linearly unrelated (orthogonal) to the response matrix Y. OSC fearn algorithm is used. OSC fearn was obtained by solving the following constrained optimization problem:

$$\max_{||\mathbf{v}_1||=1} v_1^T X^T X v_1 \tag{3}$$

$$\text{subject to} \quad (Xv_1)^T Y = 0 \tag{4}$$

The solution of the optimization problem is an eigen vector of unit length corresponding to the largest eigen-value of the matrix Q(transpose(X)Y)transpose(X)X, where Q(transpose(X)Y) is given by :

$$I - (X^T Y)[(X^T Y)^T (X^T Y)]^{-1} (X^T Y)^T \tag{5}$$

$X^T Y$ is assumed to be of full column rank. This assumption is justified for most spectral data sets, as the number of columns of X is usually much larger than the number of columns of Y.

Typically Y is of the order of one or two. If (transpose(X)Y) is not of full column rank, then those constraints which are redundant can be deleted. The additional weight vectors v_j $(j = 2, ..., A)$ are given by eigenvectors of unit length corresponding to the jth largest eigenvalue of the matrix Q(transpose(X)Y)(transpose(X)X) where A is the maximum number of iterations (N-MAX).

The component and loading vectors of fearn's OSC are defined respectively as

$$r_j = X v_j \tag{6}$$

$$q_j = X^T r_j / (r_j^T r_j) \tag{7}$$

Clearly the optimization problem is a constrained PCA problem. In comparison with ordinary PCA, the extra constraint is enforced to ensure that the latent variables $r_j = X v_j$ are orthogonal to the target variable Y. fearn's OSC filter consists of the building block v_j, r_j, q_j and is given by:

$$\sum_{j=1}^{A} r_j q_j + E \tag{8}$$

We show the sample in figure 1 after OSC pre-processing in figure 3.

2) Scaling: For some data the range of its values is too small or too large which has a negative effect on regression models, so we need to transform the readings by scaling each reading to a given range. We used MinMax Scaler Scikit-learn [18] algorithm to scale the data-set between -1 and 1.

$$X_{std} = (X - X_{min})/(X_{max} - X_{min}) \tag{9}$$

$$X_{scaled} = X_{std} * (max - min) + min \tag{10}$$

where max and min is the range we want the data to fall in between.

3) Extended multiplicative signal correction: EMSC is used in elimination of uncontrollable path length or scattering effects in the data. As it isolates and removes complicated multiplicative effects caused by physical phenomena so that chemical effects can be more easily modeled. We used Pychem library [14] implementation and it showed improvement in error measurement especially in the PLS model. The EMSC model can be written as:

$$z_i \approx a_i + b_i z_{i,chem} + d_i \lambda + e_i \lambda^2 \tag{11}$$

where a_i baseline offset and b_i the path length, relative to the baseline offset and path length in a reference spectrum. Coefficients d_i and e_i allow for unknown, smoothly wavelength-dependent spectral variations from sample to sample. If the coefficients in the equation above had been known theoretically, or estimated perfectly, then the EMSC correction:

$$z_{i,corrected} = (z_i - a_i - d_i - e_i \lambda^2)/b_i \tag{12}$$

would remove the baseline and path-length variations as well as the wavelength-dependent spectral effects, yielding

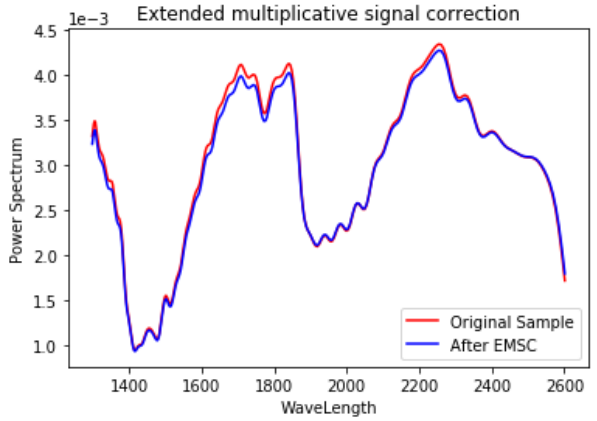

Fig. 4. EMSC pre-processing technique on a sample reading from milk data.

corrected spectra with only chemical absorbency information left: $z_{i,corrected} \approx z_{i,chem}$.

Function takes the each testing and training X-data sample with a polynomial order and return pre-processed testing and training X-data. The polynomial order is determined according to the pre-processing tuning parameters algorithm that will be mentioned afterwards, that chooses the polynomial that results in the least root mean square error of the validation data with the specified model, feature selection and pre-processing techniques. We show the sample in figure 1 after EMSC pre-processing in figure 4.

D. Features

We tried seven different feature extraction techniques shown in figure 5 and figure 6.

1) 1st Derivative: It calculates the first derivative of the data.

2) 2nd Derivative: It calculates the second derivative of the data.

3) 1st and 2nd Derivative Combined: It concatenates the first derivative and the second derivative of the data.

4) 1st Integral: It returns the area under curve from point 0 till point i, where i belongs to 1 : (length of data)-1 which is equivalent to the first integration of the X data-set.

5) 1st Derivative, 1st Integral and Original data Combined: It concatenates the first derivative, the first integration of the X data-set with the original data.

6) 1st Integral and Original data Combined: It concatenates the first integration of the X data-set with the original data to provide wider variety of attributes.

7) Covariance: For a given reading of the data, we calculate the first derivative, second derivative, first integral and second integral. Then we create a matrix of the first derivative, second derivative, original signal, first integral and second integral and calculate the covariance matrix of this matrix [15].

$$COV(X,Y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} \tag{13}$$

The resultant covariance matrix size is 25 5×5. Example of the covariance matrix is shown in figure 6.

We take only one triangle from the matrix so we get a vector of size 15 to remove redundancy. In one level covariance, the given data will be the reading (from 0 to 314) and the output vector size will be 15. In two levels covariance without overlap, the given data will be each half of the reading (from 0 to 156 and from 157 to 314) and the output vector will be 45 (15 of level 1 and 30 of level 2). In two levels covariance with overlap, we take the overlapped half in the middle of the reading (from 0 to 156, from 78 to 235 and from 157 to 314) and the output vector size will be 60 (15 of level 1 and 45 of level 2).

E. Regression Models

We include in our trials different techniques that covers the most commonly used models in similar problems with limited amount of training data. We used linear methods such as PCR and PLS. And we used non-linear methods such as GPR and SVR.

1) Linear Models: The idea is to find a few linear combinations of the original x-values and to use only these linear combinations in the regression equation. To discard any irrelevant and unstable information and keep only the most relevant part of the x-variation is used for regression [5]. The data matrices used as input to PCR and PLS are denoted by X and y. The model structure for both methods is given by the equation:

$$X = TP^T + E_y = T_q + f \tag{14}$$

The matrix P and vector q are called loadings and describe how the variables in T relate to the original data matrices X and y. The matrix E and vector f are called residuals and represent the noise or irrelevant variability in X and y, respectively. Unlike PCR, PLS uses both X and y in the estimation process directly.

a) Principal Component Regression: The PCR method is based on principal component analysis where the algorithm looks for the most dominant principal components of the data that are determined by their ability to account for the variability in the data. We need to determine the number of principal components that produces the least error in the training data. Then the model is re-trained using that number of components and tested on the validation set to capture accuracyof the results.

b) Partial Least Squares Regression: Development of PLS regression method was made to avoid the problem of deciding which components to use in the regression equation. PLS uses factors determined by employing both data and labels in the estimation of the component. PLS looks for components that maximize the covariance between the labels and all possible linear functions of the data, making these components more directly related to variability in labels than the principle components. During the training phase we initially have to determine the number of components to use for testing our model, an initial phase of training is done to determine the best number of components to be used for the current pre-processing and feature extraction operation with the goal of optimizing the error in prediction on the folds on

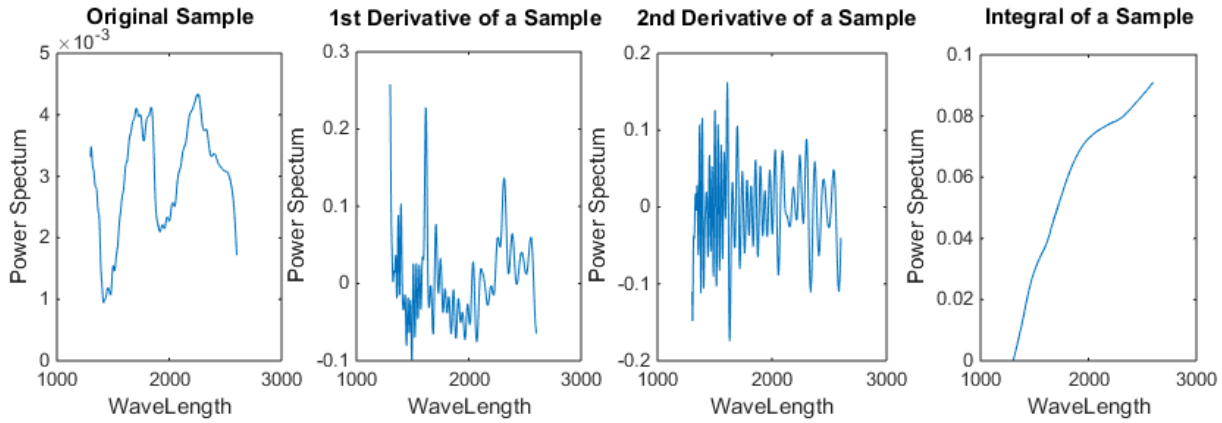

Fig. 5. Features extraction from the original Data

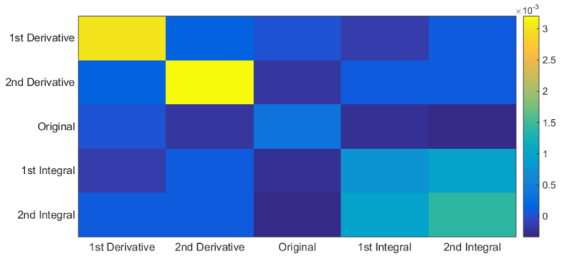

Fig. 6. 5x5 covariance matrix extracted from five variables: 1st, 2nd derivatives, 1st, 2nd integrals and the original signal.

the training data. Then after determining the optimal number of components. The models is re-trained and tested on the validation set to capture accuracy results.

2) non-linear methods:

a) Support Vector Regression: SVR [16] tries to solve the following optimization problem:

$$\text{minimize} \quad 0.5||w^2|| + C\sum_{i=1}^{l}(\varepsilon_i + \varepsilon_i^*) \tag{15}$$

$$\text{subject to} \begin{cases} y_i - < w, x_i > -b & \leq \epsilon + \varepsilon_i \\ < w, x_i > +b - y_i & \leq \epsilon + \varepsilon_i^* \\ \varepsilon_i, \varepsilon_i^* & \geq 0 \end{cases} \tag{16}$$

The kernel functions transform the data into a higher dimensional feature space to make it possible to perform the linear separation. We use the Radial basis function (RBF) kernel.

b) Gaussian Process Regression: GPR [17] is a probabilistic non parametric modeling technique, assuming that the joint distribution over any finite set of fixed test points is a multivariate Gaussian. The Gaussian process is fully specified using the mean function and the covariance function of this joint Gaussian distribution. The mean function is often assumed to be zero. The parameters y for the covariance function can be optimized to maximize the conditional probability

$$p(y_j(X, y)$$

using the training data. According to Bayes' rule, the related y for a new sample can then be predicted by extending the joint distribution with the optimized covariance function. The used kernel is Radial basis function kernel (RBF) and it is given by:

$$k(x_i, x_j) = exp\big(-\frac{1}{2}d(x_i/\lambda, x_j/\lambda)^2\big) \tag{17}$$

where λ is the length scale.

IV. Results and Discussion

A. Data Set

The data set is the NIR analysis of liquid MILK. We worked on 74 samples, where each sample is 5 readings. Samples are divided into 50 training set, 14 validation set and 10 testing set. The data set consists of NIR measurements of 314 wavelengths in the range of 1300 - 2600 nm . The NIR measurements is used to determine the percentage of each component in the Milk Samples, where the percentage of 0.7 to 4.4% for **fats**, 2.88 to 4.8% for **proteins**, 4.11 to 5.25% for **lactose** and 8.86 to 13.94% in **solids non-fats**.

B. Tuning

For each Regression Model and Pre-processing technique there are one or more hyper-parameter that needs tuning, our results is derived upon root mean square error (RMSE) between the predicted y and the referenced y data. We used 2 different approaches to determine best parameters using only the training data-set.

1) Leave one sample tuning:

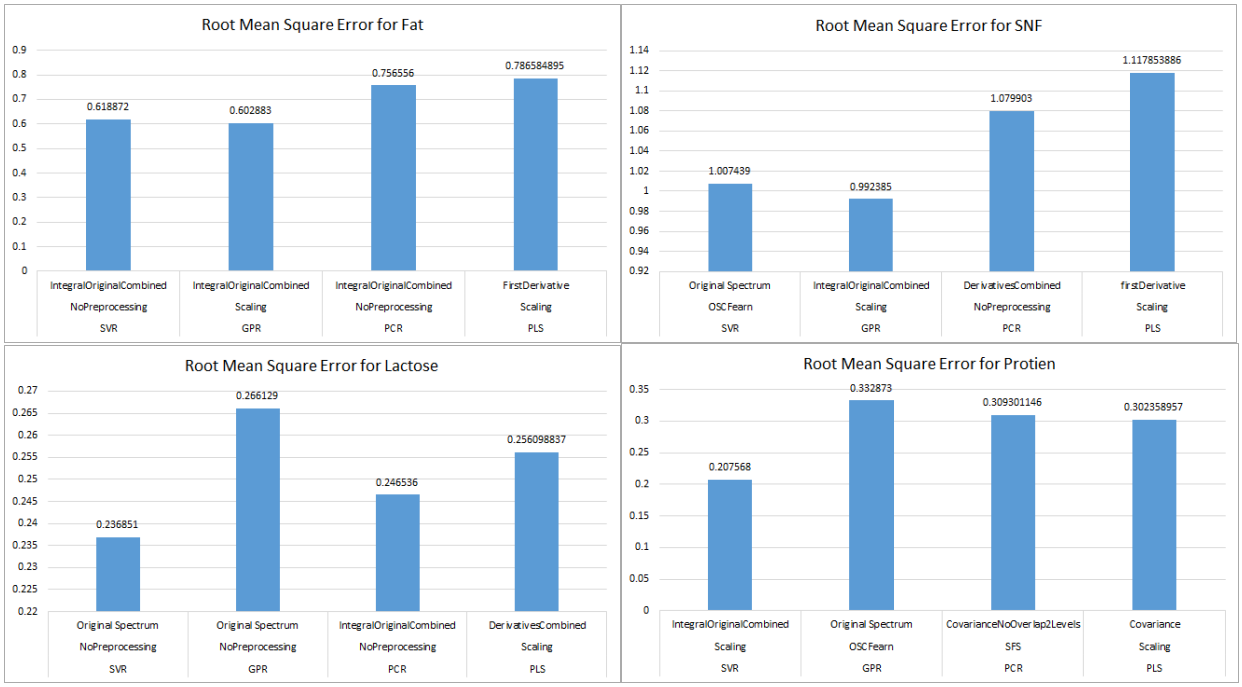

Fig. 7. Validation set results.

From all training data set, one sample as test data and the others are training.

2) Leave one fold tuning using k-folds:
Divide the training data-set to k-folds, one fold is for testing and k-1 folds for training.

For both methods, we try all the results of each value that could be given to the parameter being tuned, compare them and return best parameter value that have the least root mean square error for the average of all samples or folds.

C. Validation set results

We run tuning procedure over all combinations of 3 pre-processing techniques, 1 wavelength selection, 7 feature extraction methods and 4 regression models. This is done over four response variables. In total we had 1024 tuning experiments. Where we tune over a range for hyper-parameter values. We show the best results on the validation set for each of the response variables in figure 7.

D. Test set results

We used the validation set to tune our models. Next we used the best models on the test set. In table I, we see that the SVR and GPR regression models are doing the best across all response variables. Also we see that the pre-processing modules can be ignored in three variables. The best features were the original spectrum and the integral combined. We also average the values of each sample both testing y and the predicted y then calculate the RMSE and MAPE after the averaging.

V. Conclusion

Using variety of comparative configurations of regression models, pre-processing techniques, wavelength selection and feature extraction methods, our study concluded that for liquid milk data non-linear models (GPR and SVR) showed better results. Also, it was proved that despite the fact that pre-processing techniques as the OSC and EMSC showed great enhancement in previous studies for solid materials, it did not improve the results from our models.

Furthermore, We constructed 2 different libraries in our project, where both the online android application library and the offline machine learning library could accept the addition of extra pre-processing techniques, wavelength selection methods , feature extraction methods and regression models.

There is potential that in the future, our application would add food products beside milk. We can enhance our models by adding more samples to the training data-set and try deep neural networks regression models.

VI. Acknowledgement

We would like to express our appreciation to Si-Ware Systems for supporting this research with their state-of-the-art spectrometers as well as to Borg AlArab Innovation Cluster for funding and supporting the project.

References

[1] NeoSpectra.com. (2018). FT-NIR Spectral Sensors And Near Infrared Spectroscopy Sensors. [online] Available at: http://www.neospectra.com/

TABLE I
TEST SET RESULTS FOR DIFFERENT RESPONSE VARIABLES.

	Fats	Protiens	Lactose	Solid non fats
Pre-processing	Scaling	Scaling	No pre-processing	Scaling
Feature Extraction	IntegeralOriginalConcatenation	IntegeralOriginalConcatenation	Original	IntegeralOriginalConcatenation
Regression Model	GPR	SVR	SVR	GPR
RMSE (Per Reading)	0.927709	0.26908	0.235333	1.0745364
RMSE (Per Sample)	0.7798152564	0.2150748784	0.1988942704	0.9092644646
MAPE (Per Reading)	45.05551172	6.568899564	3.857070936	7.443633409
MAPE (Per Sample)	43.3081088	6.509300112	3.847015789	7.274388189

[2] Koops, Joop & Olieman, Kees. (1985). Routine Testing of Farm Tank Milk with the Milko-Scan 203. 3. Comparative Evaluation of Polarimetry, HPLC, Enzymatic Assay and Reductometry for the Determination of Lactose. Calibration for Infra-Red Analysis of Lactose. Calculation of Total Solids from Infra-Red Measurements. International Dairy Journal. 39-89.

[3] Van de Voort, F. R. (1980). An evaluation of the Milkoscan 104 infrared milk analyzer. J. Assoc. Off. Anal. Chem., 63, 973-80

[4] Wan, J., Chen, Y., Morris, A. and Thennadil, S. (2017). *A Comparative Investigation of the Combined Effects of Pre-Processing, Wavelength Selection, and Regression Methods on Near-Infrared Calibration Model Performance*. Applied Spectroscopy, 71(7), pp.1432-1446.

[5] Næs, T. (n.d.). A user-friendly guide to multivariate calibration and classification, NIR Publications, Chichester, UK, 2002.

[6] Myttenaere, A.D.; Golden, B.; Grand, B.L.; Rossi, F. Mean Absolute Percentage Error for regression models. Neurocomputing 2016, 192, 38–48.

[7] El-abassy, Rasha & Eravuchira, Pinkie & Donfack, Patrice & von der Kammer, B & Materny, Arnulf. (2011). Fast determination of milk fat content using Raman spectroscopy. Vibrational Spectroscopy - VIB SPECTROSC. 56. 3-8. 10.1016/j.vibspec.2010.07.001.

[8] Svensson, O., Kourti, T. and MacGregor, J. (2002). *An investigation of orthogonal signal correction algorithms and their characteristics*. Journal of Chemometrics, 16(4), pp.176-188.

[9] Li, A. J. Morris, and E. B. Martin, Orthogonal signal correction: algorithmic aspects and properties, Journal of Chemometrics, 16 (2002) 556

[10] B. Arguello, et al., A survey of feature selection methods: algorithms and software, Ph.D. thesis, 2015.

[11] Rinnan, A., Berg, F. and Engelsen, S. (2009). Review of the most common pre-processing techniques for near-infrared spectra. TrAC Trends in Analytical Chemistry, 28(10), pp.1201-1222.

[12] Zambom, Adriano & Kim, Jongwook. (2018). Consistent significance controlled variable selection in high-dimensional regression: Significance Controlled Variable Selection. Stat. 7. e210. 10.1002/sta4.210.

[13] Eugene Grechanovsky, IliaPinsker, Conditional p-values for the F-statistic in a forward selection procedure,Computational Statistics & Data Analysis, Volume 20, Issue 3,1995, Pages 239-263, ISSN 0167-9473.

[14] Arvis, R.M.; Broadhurst, D.; Johnson, H.E.; O'Boyle, N. & Goodacre, R. (2006) PyChem *multivariate analysis package for Python*. Bioinformatics 22(20):2565-2566.

[15] M. Hussein, M. Torki, M. Gowayyed, and M. El-Saban. Human Action Recognition Using a Temporal Hierarchy of Covariance Descriptors on 3D Joint Locations. In IJCAI, 2013.

[16] Nello Cristianini and John Shawe-Taylor. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press, Cambridge, U.K., 2000.

[17] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning*, the MIT Press, 2006, ISBN 026218253X. c 2006 Massachusetts Institute of Technology. www.GaussianProcess.org/gpml.

[18] Pedregosa, F. and Varoquaux, G. and Gramfort, A. and Michel, V. and Thirion, B. and Grisel, O. and Blondel, M. and Prettenhofer, P. and Weiss, R. and Dubourg, V. and Vanderplas, J. and Passos, A. and Cournapeau, D. and Brucher, M. and Perrot, M. and Duchesnay, E. (2011) Scikit-learn: Machine Learning in Python . Journal of Machine Learning Research 12 ,pp.2825–2830.

[19] Mishra, P., Herrero-Langreo, A., Barreiro, P., Roger, J., Diezma, B., Gorretta, N., Lleó, L., 2015. Detection and quantification of peanut traces in wheat flour by near infrared hyperspectral imaging spectroscopy using principal-component analysis. J. Near Infrared Spectrosc. 23 (1), 15–22. http://dx.doi.org/10.1255/ jnirs.1141.

[20] Ludwig, B., Vormstein, S., Niebuhr, J., Heinze, S., Marschner, B., Vohland, M., 2017. Estimation accuracies of near infrared spectroscopy for general soil properties and enzyme activities for two forest sites along three transects. Geoderma 288, 37-46.

[21] Pereira, H.V., Amador, V.S., Sena, M.M., Augusti, R., Piccin, E.: Paper spray mass spectrometry and PLS-DA improved by variable selection for the forensic discrimination of beers. Anal. Chim. Acta 940, 104–112 (2016).