IDENTIFICATION OF PROTEIN CONTENT IN MILK POWDER USING SPECTROSCOPIC TECHNIQUES

Abstract

There are many applications of remote sensing available for material analysis. It is important for the dairy food industry to figure out what makes up milk. This helps determine its value, inform consumers, and ensure quality control. The fat and protein levels in milk are important for the economy because many countries rely on these components, along with somatic cell counts, for milk trade. In this study, ASD FieldSpec4 spectroradiometer is used for capturing the spectral data of various types of Milk powder samples purchased from local shops. For measuring and analysis purpose RS3 and ViewSpecPro software were used. Hyperspectral data collected through ASDFieldSpec4 Spectroradiometer and Partial Least Square Regression (PLSR) Machine Learning technique was used for analysis of Milk powder and estimates the protein contents from various types of milk powder.

Index Terms: Milk Powder, Protein, VIS-NIR, Partial Least Square Regression, Spectral data.

Authors

Aparna Magare
Department of Computer Science & Information Technology
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra, India.
aparnam240@gmail.com,

Chetan Pattebahadur
Department of Computer Science & Information Technology
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra, India.
chetu358@gmail.com

Dr. Ratnadeep Deshmukh
Department of Computer Science & Information Technology
Dr. Babasaheb Ambedkar Marathwada University
Aurangabad, Maharashtra, India.
rrdeshmukh.csit@bamu.ac.in

I. INTRODUCTION

Milk is really important for people throughout their lives. Milk from farm animals, including cows, goats, and sheep, is a common food source for people throughout their lives. This milk is used to make various products like cheese through fermentation and other processes. Milk and dairy products are considered important in the diets of people from all over the world, whether their country is developed or still developing. Milk powder is used in bakeries, confectioneries, baby formula, and other types of food that provide nutrients. Because milk powder is used a lot in daily life, it is a big problem worldwide when it is mixed with harmful things. Milk powder is made up of proteins, lactose, carbohydrates, vitamins, moisture, and other things. It is difficult to check if this component is in milk powder using regular methods because they have drawbacks. Chemical methods, such as HPLC and GCMS, take a long time, are expensive, and can destructive technique [1]. Instead of using old-fashioned methods, spectroscopic techniques are faster, more dependable, less expensive, and don't harm the sample. With spectroscopy, the sample can be used again, but with destructive techniques, it can't be reused. In this study, we used an ASDFieldSpec4 Spectroradiometer to find out how much protein is in milk powder [2].

II. STUDY AREA & SPECTRAL MEASUREMENTS USING FIELDSPEC4

We bought ten different type of milk powder from a local market in Aurangabad. All the glass containers, measuring 65mm in width and 14mm in height, were completely filled with milk P We used a Fieldspec4 to collect signatures of milk powder samples. We measured the wavelengths from 350nm to 2500nm. We performed the spectrum scanning method in a dark room to minimize the effect of light from outside. The instrument measures things at specific times and can see small details. It can measure things between 350 and 100 nanometers with a distance of 1. 4 nanometers between each measurement. It can also measure things between 1000 and 2500 nanometers with a distance of 2 nanometers between each measurement.

We used 100 grams of milk powder. This device uses a halogen light as a light source to illuminate the samples using a fibre-optic cable with an 8* field-of-view. Reflectance spectra of milk powder samples were obtained using reflectance light [3]. The RS3 spectral acquisition system was used to capture reflectance spectra of milk powder samples with wavelengths ranging from 350nm to 2500nm. A standardized white spectral on a panel with 100% reflectance was used to increase the signal and calibrate the accuracy. We used a method called progressive scan to analyse each milk powder sample.

There are 10 milk powder samples and each one was checked 10 times in total. So we looked at 100 milk powder samples using the PLSR algorithm. The average spectra of these scans were created using a software called ViewSpecPro. This recorded the reflectance of the milk powder samples and represented them as a spectrum.

1. **Proposed Methodology:** The milk powder samples are collected from the local market. The FieldSpec4 Spectroradiometer is used to collect the spectral data for these samples. 10 scan for each sample is captured through a gun of fieldspec4. Then mean spectral signatures for each sample is built using ViewSpec4 software. The wavelengths collected using fieldspec4 were pre-processed and then the PLSR regression technique can be used

for finding the protein content in each sample. The steps required in the proposed methodology are shown in below figure1.

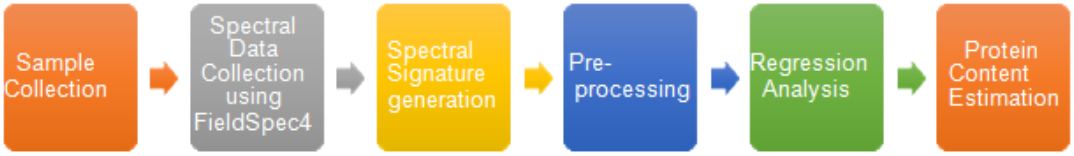


Figure 1: Proposed Methodology

- **Step 1:** In this step, we measured the spectrum of samples that were collected earlier. We used the Fieldspec4 Spectroradiometer to take measurements of the different samples. We used a special device called fieldspec4 to capture spectra. Then, we connected the device to a computer and used RS3 software to show the spectra on the computer screen. To make sure everything is working correctly, two groups of samples were made for testing and confirming.[4] Out of 10 milk powder samples, 8 were used to set the calibration, and the others were used to test the validation model. Model calibration means figuring out what the model's outputs should be based on past information. Model validation means checking if the calibrated model works correctly by using different information that wasn't used during the calibration.

- **Step 2:** After spectral measurements, spectra generated for each sample were pre-processed using different pre-processing techniques, such as First Derivative Pre-processing and Savitzky Golay Pre-processing. The data pre-processing was done using the View Spec Pro software. Generally, pre-treatment is carried out on data in order to remove inconsistent, inappropriate, noisy data. To achieve better accuracy pre-processed data is helpful. The following are pre-treatment methods that are applied in this study:

  - **First Derivative:**The first derivative is the simplest kind of derivative. In this way, each point is considered The variable in the sample is taken away from its close variable. The signal that is the same between
    two variables is removed, and only the different parts are considered using this pre-treatment method. This pre-processing removes the starting signal when used on all the information.

  - **Savitzky-Golay:** Noise has occurred, as derivatives emphasize lower frequency signals and emphasize higher frequency signals. Hence, for smoothing the data, Savitzky-Golay pre-processing method is utilized to smooth the data, because it improves the usefulness of derivatives [4,5].

- **Step 3:** In this step, the method of Partial Least Squares Regression (PLSR) is applied to the data that was previously processed.

  - **PLSR Method:** The Partial Least Square Regression in this study predicts the accuracy of our estimate. For estimation of better accuracy, Partial Least Square

Regression is a widely utilized regression method to relate spectral data captured by the fieldspec4 device with the estimations of laboratory iron oxide. If predictors are majorly collinear with each other and if you have more predictors than observations, then partial least square regression is particularly useful. The outcomes can differ significantly from those calculated individually for the response variables since PLSR models the response variables in a multivariate manner [6].

In spectral analysis, Partial Least Square Regression has been confirmed to be a vigorous and credible procedure, because of its advantage in reducing dimensions and figuring out co- linearity difficulties among independent variables [7-8].

- **Step 4:** In this step, it creates a model that is accurate by using the PLSR method. The accuracy of the model is measured using the RMSE and R2 value. When the RMSE value is lower, it means that the accuracy of the prediction is better.

The RMSE was used to check how accurate the predictions were. The R2 values were calculated to see how reliable the predictions were. The model that had the lowest root mean square error (RMSE) and the highest determination coefficient (R2) was selected as the best model.

In simpler terms, Ym, represents the value of milk powder Protein that was actually measured. Yp, represents the value that was predicted. Ymean, is the average of all the measured values. Lastly, n represents the number of measured or predicted values, starting from 1 and increasing by 1 each time.

III.PLSR METHOD AND RESULT

The PLSR model helps determine the protein content by analyzing the reflection of light. This method considers all the different features of light and simplifies them to a few basic parts. These parts are then used to figure out the amount of protein. After using the first derivative pre-processing on the data, we found that the calibration had an R2 value of 0. 91 and an RMSE value of 0. 008711For the validation, the R2 value was 0. 92 and the RMSE value was 0. 001624This was achieved by utilizing Python programming to establish a connection between milk powder protein and hyperspectral data. The below figure shows the regression plots and R2 value.

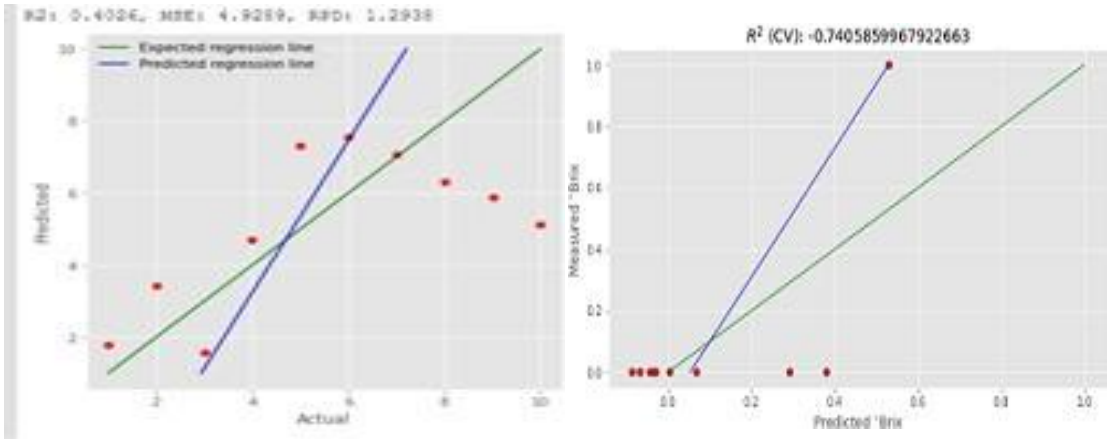


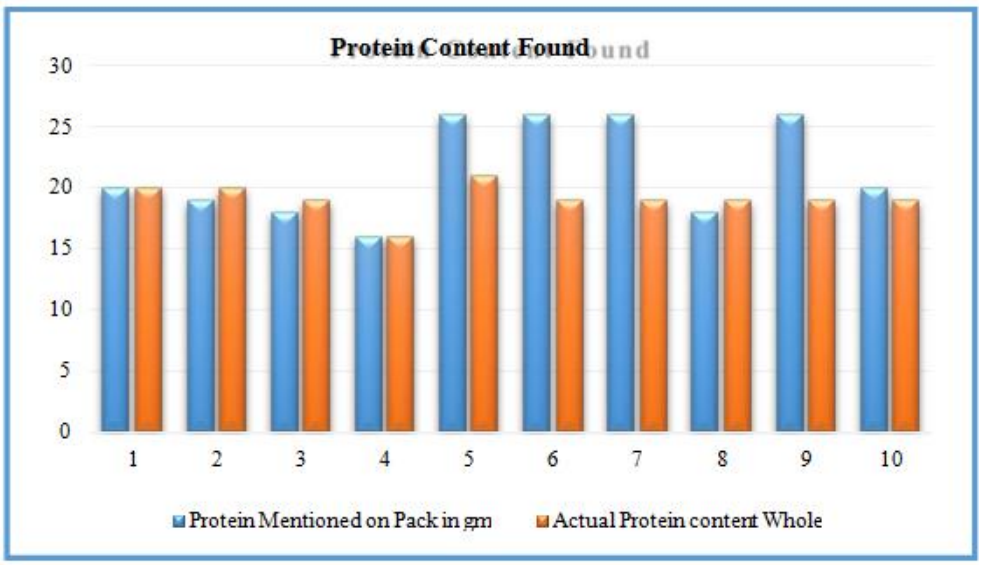
Figure 2: Regression Plots

Figure 2 displays the results of the PLSR models. It shows graphs of the actual and predicted protein concentrations for the calibration and validation data sets. This was achieved through PLSR Analysis. The R2 values of the validation set are higher and the RMSE values are lower compared to the calibration set. The Graphical User Interface of the developed model for the estimation of protein contents from milk powder samples developed in Python language using machine learning libraries is shown in the below figure.

The GUI showing Protein estimation is shown in figure 3

Figure 3: Protein Estimation from Sample 1

The graph of the table showed Table showed the prediction accuracy of the model. The Protein contents mentioned in the milk powder pack and the model show how much actually present in it.



IV. CONCLUSION AND FUTURE SCOPE

This research found that scientists have used methods that don't involve cutting into something to analyse and measure the protein content of Milk Powder. The study discovered that using a method called spectroscopy is helpful in analysing milk and identifying its components. This approach is more effective than the traditional method used for milk powder and dairy products. This study aims to measure the amount of protein in milk powder using a special machine called ASD Fieldspec4 Spectroradiometer that collects hyperspectral reflectance data. You can figure out how much Protein is in milk powder samples. In this research, the Savitzy Golay filter was used along with the first derivative and PLSR regression method. The findings showed that it provided accurate results. This allows evaluating the amount of protein in milk powder without damaging it. It gives the opportunity to estimate Protein contents from milk powder in a non-destructive way.

V. CONTRIBUTION AND SIGNIFICANCE

- Developed Spectral database of ten different of Milk Powder.
- The model is developed in python that predict the Protein contents from any type of Milk Powder.

Future Scope

- Study of the various types of Milk Powder like Goat, Camel, etc.
- Production procedure of Milk Powder can be considered.
- Spectral bands can be identified for different types of contents in milk powder.
- Developed a model for the various application of Dairy Product content finding.

REFERENCES

[1] Navale, D. and Gupta, S., Detection of Lactose and Casein in Different Milk Samples, IJLTEMAS, Vol. 5 (6), ISSN 2278-2540, June 2016.

[2] Wu, T., Chen, H., Lin, Z. and Tan, C., 2016. Identification and quantitation of melamine in milk by near-infrared spectroscopy and chemometrics. Journal of Spectroscopy, 2016.
[3] Wu, D., He, Y. and Feng, S., 2008. Short-wave near-infrared spectroscopy analysis of major compounds in milk powder and wavelength assignment. Analytica chimica acta, 610(2), pp.232-242.
[4] Luo, A., Zheng, Y., Chen, X. and Cong, F., 2018. Undergraduate laboratory experiment on determination of total protein content in milk powder by moving reaction boundary titration. Biochemistry and Molecular Biology Education, 46(6), pp.644-651
[5] McGoverin, C.M., Clark, A.S.S., Holroyd, S.E. and Gordon, K.C., 2010. Raman spectroscopic quantification of milk powder constituents. Analytica Chimica Acta, 673(1), pp.26-32.
[6] Mehmood, Tahir, Kristian Hovde Liland, Lars Snipen, and Solve Saebo, "A review of variable selection methods in partial least squares regression", Chemometrics and Intelligent Laboratory Systems 118, 62-69, 2012.
[7] Xie X., Pan X.Z., Sun B., "Visible and near-infrared diffuse reflectance spectroscopy for prediction of soil properties near a Copper smelter", Pedosphere, 22: 351–366, 2012.
[8] Garfagnoli, Francesca, Gianluca Martelloni, Andrea Ciampalini, Luca Innocenti, and Sandro Moretti, "Two GUIs-based analysis tool for spectroradiometer data pre-processing", Earth Science Informatics 6, no. 4, 227-240, 2013.
[9] Ranjitha G.,Srinivasan M.R ,Abburi Rajesh ,Detection and estimation of damage caused by aphid, aphis gossypii (glover) in cotton using hyperspectral radiometry, International Journal of Recent Scientific Research, Vol.5(4) ,PP.750-753, 2014
[10] Ruffin, C., King, R.L. and Younan, N.H., 2008. A combined derivative spectroscopy and Savitzky-Golay filtering method for the analysis of hyperspectral data. GIScience & Remote Sensing, 45(1), pp.1-15

