OPTIMIZATION OF MACHINE LEARNING ALGORITHMS FOR DETECTION OF MONO – SODIUMGLUTAMATE

Dr V T Balamurugan[1], M Pravinkumar[2], S Saveetha[3], D Thenmozhi[4]

[1] *Professor, Department of Electronics and Instrumentation Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India*

[2,3,4] *Final Year Students, Department of Electronics and Instrumentation Engineering, Bannari Amman Institute of Technology, Sathyamangalam, India*

Abstract - Food is one of the basic needs for every living organism to survive on the planet. Food adulteration is becoming more common nowadays. Food adulteration is the act of adding one or more substances in order to increase the taste and quantity of the food. In recent times many people especially those in city areas prefer eating in food stalls and hotels due to its taste and to reduce timeconsumption for food preparation. On the other hand, they are not realizing the harmful effects which are caused by food adulteration. Adulterants have broughtabout various infections and unexpected losses. Food adulteration causes dangerous effects like burning sensation, neurological disorders, hypersensitivity, blood deficiency, malignant tumor etc... The main objective of our project is to detect the presence of Ajinomoto (Mono-sodium Glutamate) present in the food stuffs using Mid-Infrared (MIR) spectroscopy and using machine learning algorithms. In order to find the efficient algorithm to detect the Ajinomoto content in the food stuffs, we used different machine learning algorithms like principal component Analysis (PCA) algorithm, Partial Least Squares Regression (PLSR), K-means, Support Vector Machine (SVM). After comparing the efficiency of these algorithms, the efficient algorithm for Ajinomoto detection is identified andit is chosen for implementation.

Key Words: Ajinomoto, Adulteration, PCA, PLSR, K-Means, SVM, MIR

1.INTRODUCTION

In recent news, it is found that 80% of all unexpected losses is due to contaminated food and water. Food adulteration in India begins from the over usage of composts and pesticides. Yet, pesticide build-ups are by all account notthe only issue. Leafy foods once contained nutrients and minerals in them. Yet, presently they are loaded up with noxious synthetic compounds like endo-sulphite which ruin our wellbeing. Synthetics like carbide utilized for fast maturing of natural products quicker have made various wellbeing perils. Coming up next are a couple of wellbeing dangers because of contaminated food items.

- Addition of mineral oils to edible oil, can result in malignancies.
- Addition of Lead chromate to turmeric powder and flavors can lead to frailty, loss of motion, mind harm and premature births.
- Addition of lead with water, can result in lead harming. It causes sleep deprivation, clogging, sickliness, and psychological impediment.
- Addition of cobalt to H_2o and alcohol leads

to cardiovascular harm. Cu, Sn and Zn can result in retching and looseness of the bowels.

- Non-allowed shading or allowed food shading like metal yellow, past as far as possible in hued food can cause hypersensitivities, hyperactivity, liver harm, fruitlessness, paleness, malignancy and birth abandonment.
- In today's world, junk food capture's a great amount of attention among people, especially those in city areas. In order to increase the taste of the many adulterants are added to these junk foods. One among those adulterants is Ajinomoto (Monosodium Glutamate).
- Ajinomoto has many harmful effects in humans like burning sensation, neurological disorders, etc...
- So, an effort should be taken in order to detect the presence of Ajinomotoin food stuff.

2. LITERATURE SURVEY

In [1], Detection of adulteration in meat by bacteria is carried out using five different supervised and unsupervised machine learning algorithms. Random Forest is identified as an efficient algorithm to detect the adulteration in meat due to its accuracy of 97.57% and a web application is designed to show the detected bacteria.

In [2], Adulterant is detected using recent technologies like image processing and machine learning. Major sources include OpenCV and matplotlib.

In [3], Adulteration of milk is discussed. This paper gives the details of most essential researches that are carried out in the adulterated milk using cutting edge technology like machine learning and AI.

In [4], Adulteration in olive oil is identified and the adulterant isquantified using Spectroscopic data, supervised and unsupervised ANN's.

In [5], Fraudulent milk adulteration is detected using Fourier Transformed Infrared Spectroscopy (FTIR) in order to obtain the chemical composition of milk. The spectral data obtained by this technique is combined with machine learning methods like Neural Networks and Decision trees to develop a model that gives the feature of pure and adulterated milk.

In [6], Adulterants in paddy grains are recognized using image processing techniques and classified based on colour and text feature. In this, different paddy varieties are taken and the samples are obtained by mixing the different paddy grains together. Various machine learning algorithms are used and BPNN classification model with PCA is identified as a cost effective, harmless and an efficient algorithm with an accuracy of93.31%

In [7], Focus is made on Risks in Food adulteration. The risks on adulteration are direct, indirect, and technical. Public health gets worse every time, because the production department

is mostly not following good food practices. Examples of such chemicals includes melamine included milk, carcinogen colorants in foods, diluting fruit juices, species swapping of seafood, unauthorized repackaging, smuggling which avoids tax. This is tested not only from the result of the adulteration, but also the cause and motivation.

In [8], food adulteration is a menace, which people facing now. Detecting food adulteration is an important requirement for making sure safety of foods. Lab techniques are highly accurate, precise, and reliable and they are costly and time consuming. It is good to undergo reliable "quick screening tests" in which a common person is able to perform at the household level to have a broad picture of progress in adulteration in his food in case of doubt. Even though there are mighty chances for improvement and after development, some fast methods of adulterants detection, developed by several government and non-government agencies for household application, have been presented.

3.HARDWARE DESCRIPTION

3.1 SPECTROSCOPY

Spectroscopy is the study of structures of atoms and molecules in the matter. It is done by passing a large number of radiations into the matter. Spectroscopy is mainly used to recognize, determine or measure the molecular or structural constitution of a matter. Every molecule has its own characteristic way in emitting, absorbing and reflecting radiations like one molecule can absorb theradiations passed into it, another can emit the radiations while another can reflect the radiations. In simple terms, spectroscopy includes the use of radiation to gather information on the structure and properties of matter.

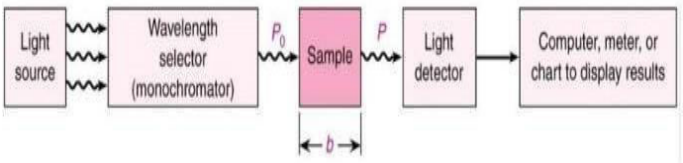

Figure 3.1shows the Process of Spectroscopy

Infrared spectroscopy involves the idea that molecules generally absorb specific frequencies of light that are normal for the corresponding structure of the atoms. When an infrared radiation is passed to the samples, it absorbs the radiation and excites the molecules present in it from lower energy level to higher energy level and produces an Infrared range of the energy consumed by an atom as an element of the recurrence or frequency of light. Various sorts of bonds react to the IR radiation in an unexpected way. There are different types of IR spectroscopy like Mid-Infrared, Near-Infrared, Far-Infrared, etc. depending on the range of electromagnetic radiation. Mid -Infrared region includes 400-4000cm^{-1}.

4.PROPOSED CONCEPT

4.1 BEER-LAMBERT'S LAW

When a monochromatic light of existing intensity I_0 is passed through an explicit vessel, some of its radiations are absorbed by the molecules present in the vessel so that the transmitted light intensity I is less than I_0.

The interconnection between I and I_0 rely on the length of the path l of the absorbing solution and concentration c of the absorbing solution.

$A=a_mcl$

A=Absorbance, a_m=Absorptivity, c=Concentration, L=Path Length

As per Beer-Lambert's law, absorbance and concentration of the medium are directly proportional.

3.2 TRANSMITTANCE

Transmittance is defined as the ratio of intensity while transmission to theinitial intensity.

$$T= I / I_0$$

The logarithm of transmittance gives the absorbance.

$$A=log(I/I_0)$$

4.3SAMPLE PREPARATION

- At first, pure sambar samples without Ajinomoto are taken and tested using the MIR Spectrometer by keeping the reference as distilled water and 30 readings are recorded.
- Sambar with 1gram Ajinomoto is tested and 30 readings are recorded using MIR Spectrometer by keeping the same reference as distilled water.
- Sambar with 2gram Ajinomoto is tested and 30 readings are recorded using MIR Spectrometer by keeping the same reference as distilled water.
- The raw data is then converted into transmittance data.
- The transmittance data is plotted using PCA, PLSR, K-Means and SVM algorithms.

5.BLOCK DIAGRAM

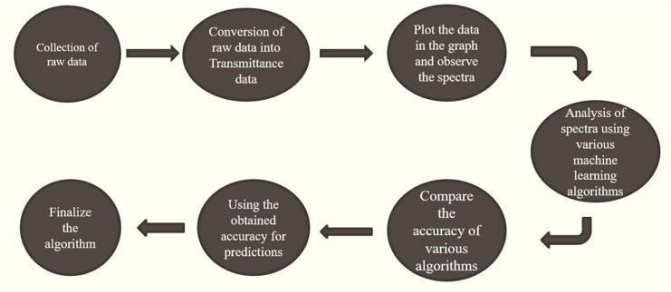

Figure 5.1 shows the Block diagram of Ajinomoto detection in food samples.

- At first, Collection of raw data is carried out using MIR spectrometer.
- After collecting raw data, it is converted into transmittance data.
- Transmittance data is then plotted in the graph and the spectra isobserved.
- Then analysis of the spectra is carried out using different kind of machine learning algorithms.
- By comparing the efficiency of each algorithms, the efficient algorithm is identified and chosen for implementation.

6.ALGORITHMS IMPLEMENTED

6.1 PCA(Principal Component Analysis)

Principal Component Analysis, or PCA, is a dimension-decrease strategy which is regularly helpful in lessening the dimension of huge information collections, by changing a huge number of data sets into compact datasets that actually includes the major information in the huge set. Lessening the quantity of factors of an information generally comes at the detriment of exactness. Since more compact informational collections are simpler to investigate and imagine and make analyzing information a lot simpler and quicker for AI calculations without irrelevant factors to measure. It involves the following steps: normalization, Divergence matrix computation, Calculation of eigenvalues and

eigenvectors, Characteristic variable, Recasting information.

Step-1

Normalization involves standardizing the range of continuous factors only then each factor contributes similarly to the analysis. When there are vast dissimilarities in the range of factors, the factors with higher ranges will overcome those with smaller ranges, which provides partial results. Along these lines, interchanging the information to practically identical ranges can prevent this problem.

Mathematically, this could be applicable by deducting the mean and isolating by the S.D for each estimation of every factor. When normalization is completed, every factor will be changed to a similar scale.

Step-2

The main objective of this step is to understand, in what way the factors of the input dataset vary from the average with relevance to one another. In other words, to check if there is any interconnection between them. For some instances, factors are extremely correlative in a manner that they are composed of unessential data. Hence, in order to spot these relations, we need to find the divergence array.

The divergence array could be a q*q consistency array, that has as elements as the divergences associated to whole potential couples of the initial factors. For instance, for a 3-d information set with three variables x, y, and z, the divergencearray could be a 3×3 matrix.

Since the divergence of a factor with itself is its variance (Cov(a,a)=Var(a)), in the principal diagonal. we really have the variances of each factor. Since the covariance is commutative, the elements of the covariance grid are consistent regarding the principal diagonal.

$$\begin{bmatrix} Cov(x,x) & Cov(x,y) & Cov(x,z) \\ Cov(y,x) & Cov(y,y) & Cov(y,z) \\ Cov(z,x) & Cov(z,y) & Cov(z,z) \end{bmatrix}$$

Covariance Matrix for 3-Dimensional Data

Step-3

Eigenvectors and eigenvalues are calculated from the divergence array to decide the main components of the information. Main components are obtained by calculating eigenvectors and arranging them by their eigenvalues in decreasing sequence. Main components are new variables that are built from the initial variables, that are unrelated and the vast majority of the information inside the initial variables is compacted into the principal components.

Ex: 100-d data will give 100 main components, in which PCA maintains the maximum information in the initial component, next maximum information in thesubsequent component and etc.

Step-4

This step decides either to have all the components or reject those with lower eigenvalues and creates a matrix of variables with the remaining components which is known as a characteristic variable. Thus, the characteristic variable is essentially an array that has as many variables as the eigenvectors of the components that we choose to keep. Hence this results in the initial move towards the decrease in dimension. Imagine we decide to have just q eigenvectors(parts) out of m, the end data set will contain just q dimensions.

Step-5

This step decides either to have all the components or reject those with lower eigenvalues and creates a matrix of variables with the remaining

components which is known as a characteristic variable. Thus, the characteristic variable is essentially an array that has as many variables as the eigenvectors of the components that we choose to keep. Hence this results in the initial move towards the decrease in dimension. Imagine we decide to have just q eigenvectors (parts) out of m, the end data set will contain just q dimensions.

6.2 PLSR (PARTIAL LEAST-SQUARES REGRESSION)

Partial Least Squares regression (PLS) is a quick, efficient and optimal regression method based totally on covariance. It is far encouraged in cases of regression where the wide variety of explanatory variables is high, and in which it is probable that the explanatory variables are correlated. This algorithm is used to find the relationship between dependent variable[x] and independent variable[y]. this algorithm uses line of best fit to define the relationship between the two variables in a graph. Simply, it is based on the idea that the square of the errors obtained must be minimized to the most possible extent.

In the case of the OLS and PCR methods, if fashions need to be computed for numerous dependent variables, the computation of the models is without a doubt a loop at the columns of the dependent variables desk Y. within the case of PLS regression, the covariance shape of Y also impacts the computations. The equation of the PLS regression model writes:

$$Y = ThC'h + Eh = XWh*C'h + Eh = XWh (P'hWh)\text{-}1\ C'h + Eh$$

in which Y is the matrix of the based variables, X is the matrix of the explanatory variables.Th, Ch, W*h , Wh and Ph, are the matrices generated with the aid of the PLS algorithm, and Eh is the matrix of the residuals. The matrix Bof the regression coefficients of Y on X, with h additives generated through the PLS regression algorithm is given through:

$$B = Wh(P'hWh)\text{-}1C'h$$

Note: the PLS regression ends in a linear model because the OLS and PCR do. PLS regression results: Correlation, observations charts and biplots A top notch benefit of PLS regression over conventional regression are the available charts that describe the statistics structure. thanks to the correlation and loading plots it is straightforward to study the relationship of some of the variables. it could be relationships among the explanatory variables or structured variables, in addition to among explanatory and structured variables. The score plot gives records approximately sample proximity and dataset structure. The biplot acquires these kinds of statistics in a single chart.

6.3 K-MEANS

K-means clustering is the simplest unsupervised machine learning algorithm which is used for clustering the data with similar features. It is an iterative algorithm that groups the data into a K number of clusters, in which each data belongs to only one group. It allocates data to the group such that the sum of thesquared distance between the data and the centroid is minimum. Centroid is nothing but the average sum of all the data.

STEPS

- First, initialize the K value which indicates the number of clusters.
- Initialize K number of centroids randomly from the given data.
- Considering centroids, Assign data to

the centroids such that the sum of squared distance between the data and centroid in less.

- Find the Centroid for the obtained clusters by finding the average sum ofall the data in the clusters.
- Re-assign the data with centroid which has least squared distance.
- Iterate this process until there is no change in the assignment of data to the centroid.

6.4 SVM(Support Vector Machine)

SVM is a supervised machine learning algorithm which is mainly used for classification related problems. In SVM, hyperplane plays a major role by classifying the data according to the features. Hyperplane is the line which separates the different classes. Data points falling on either side of the hyperplane can be attributed to different classes. Also, the dimension of the hyperplane depends upon the number of features. If the number of input features is 2, then the hyperplane is only one line.

The distance between hyperplane and the nearest point of both classes is known as margin. If there are more than 1 hyperplane, the best hyperplane is that which has the largest distance between the hyperplane and the nearest element in all the classes.

7. RESULT AND DISCUSSION

Datasets of raw food samples and food samples mixed with Ajinomoto are collected using MIR spectroscopy. The raw data is converted into transmittance data and the plot is obtained by using machine learning algorithms namely PCA, PLSR, and K Means algorithms. The output plot is obtained as follows:

7.1 PCA OUTCOME

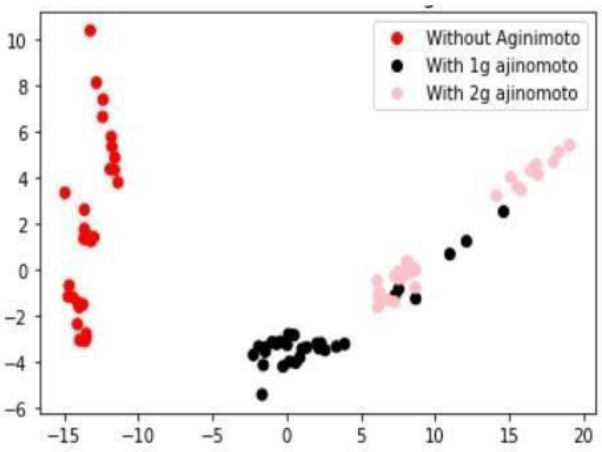

Figure 7.1shows the Plot obtained using PCA algorithm for sambar samples

7.2 PLSR OUTCOME

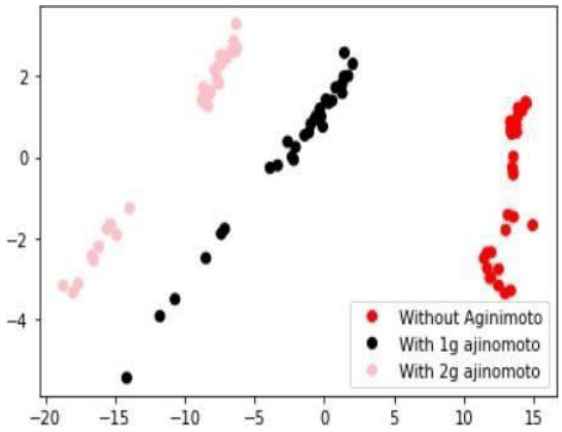

Figure 7.2 shows the plot obtained using PLSR algorithm for sambar samples

7.3 K-MEANS WITH PCA OUTCOME

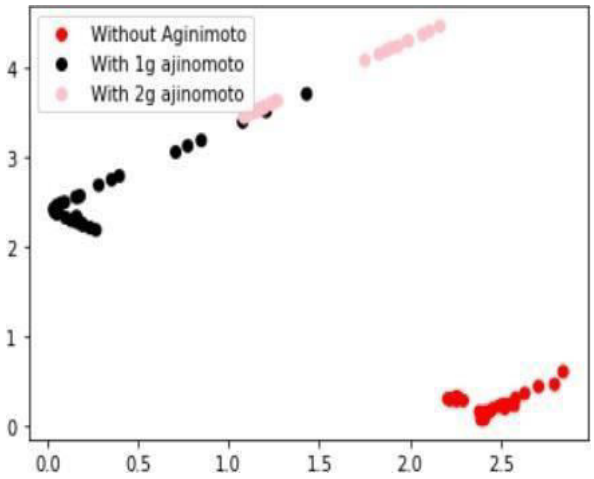

Figure 7.3 shows the plot obtained using K-means with PCA algorithm for sambar samples

Efficiency obtained using different algorithms is as follows:

Efficiency of PLSR algorithm is 0.96

Efficiency of PCA algorithm is 0.913

Efficiency of K-means with PCA algorithm is 0.94

Efficiency of SVM algorithm is 0.95

By comparing the efficiency of all the algorithms, PLSR(Partial Least Squares Regression) is identified as the efficient algorithm for Ajinomoto detection in sambar samples.

8.CONCLUSION

Ajinomoto causes many harmful effects in humans like burning sensation, neurological disorders, etc. By using the Mid Infrared spectroscopy, we obtained the data of the food products. Then we used different machine learning algorithms like PCA, PLSR, K-means, and SVM algorithms to plot the data and determined the efficient algorithm for detecting the presence of Ajinomoto. Thus, we prevent the consumers from eating harmful products. It provides a permanent solution for burning questions in food products with the help of the enhanced technologies. As referenced over, this model when fabricated into an application would serve each purchaser and seller in their everyday schedule of checking the merchandise they purchase.

REFERENCES

[1] Ma. Rica Bunuan; Relamae F. Chicote; Sheila May C. Espenida; Honeyleth L. Masangcay; Development of predictive models using machine learning algorithms for food adulterants Bacteria Detection.

[2] Anjana Purushothaman Maruveetil; Pradeepa K; Detecting food adulterants using the concepts of machine learning published on Conference: ICCIS-April 2019.

[3] Surya Vijayan; Senthilselvi Ayothi;A Qualitative Analysis of the Machine Learning Methods in Food Adultery: A Focus on Milk Adulteration Detection published on Journal of Advanced Research in Dynamical and Control systems, October 2020.

[4] John C Cancilla; Regina Aroca-Santos; Gemma Matute; Enrique S Pariente; Kacper W Wierzchos; and Jose S Torrecilla; Detecting, identifying, and quantifying adulterations in foods through computational artificial intelligence published on 9th Euro-Global Summit & Expo on Food & Beverages July 11-13, 2016 Cologne, Germany.

[5]Habib Asseiss Neto; Wanessa L.F. Tavares; Daniela C.S.Z. Ribeiro; Ronnie C.O. Alves; Leorges M. Fonseca and Sérgio V.A. Campos;On the utilization of deep and ensemble learning to detect milk adulteration published onJuly 8,2019.

[6] Basavaraj S. Anami; Naveen N. Malvade; Surendra Palaiah; Automated recognition and

classification of adulteration levels from bulk paddy grain samples published on December 6, 2018.

[7] Podstawka E; Swiatłowska M; Borowiec E; Proniewicz L; [2007]. Food additives characterization by infrared, Raman, and surface-enhanced Raman spectroscopies published in Journal of Raman Spectroscopy, 38(3),pp.356-363.

[8] Food Adulteration - an overview | ScienceDirectTopics",Sciencedirect.com,2018.[Online].Available:https://www.sciencedirect.com/topics/ foodscience/food-adulteration. [Accessed: 14- Aug- 2018].

[9] "FOOD ADULTERATION PART-II FRUITS/VEGETABLES All that shines is not pure! All India Ex-servicemen Joint Action Front (SanjhaMorcha)",Sanjhamorcha.com,2018.[Online].Available:http://www.sanjhamorcha.com/?p=4639. [Accessed: 14- Aug- 2018].

[10] "Adulteration and Harmful effects of Food Adulteration", India Study Channel,2018.[Online].Available:http://www.indiastudychannel.com/resources/ 172754-Adulterationand-Harmful-effects-of-Food-Adulteration.aspx. [Accessed: 14- Aug- 2018].

[11] Jiyu Peng; WeiyueXie; Jiandong Jiang; Zhang feng; Fei Zhou; Fei Liu; Fast Quantification of Honey Adulteration with Laser-Induced Breakdown Spectroscopy and Chemometric Methods Published on 14 march 2020.