Wheat Kernel Classification Using Machine Learning

Kaustubh Ramesh Dubey, Adityaraj Hemant Chaudhari, Karthik Kishan Gatla, Prof. Gayatri Bhandari

Computer Department

JSPM's Bhivarabai Sawant Institute of Technology and Research,

Wagholi, Pune, Maharashtra

Abstract: Machine learning is mainly divided into two sections, specifically supervised and unsupervised learning. Supervised learning is additionally divided into two fundamental parts, that is, the regression type and classification type. Unsupervised learning additionally has its types like clustering, association, PCA, and so on. Here, we consider the use case from the supervised machine learning approach. We consider the dataset of the wheat kernel, used to classify 3 types of wheat, specifically Kama, Rosa, and Canadian, consisting of properties of the wheat kernel like area, compactness, perimeter, and so on. So, as this dataset comprises 3 classes or 3 distinct results, it is called multi-class data, which is multiple classes.

Various machine learning algorithms can assist us in solving this multi-class classification problem. Some of the algorithms from Bagging [Ensemble Technique] are Random Forest, Boosting [Ensemble Technique] are LightGBM, XGBoost, Gradient Boosting, Support Vector Machine Classifier, Decision Tree Classifier, Logistic Regression, and so on. The above-mentioned algorithms work in different ways to find the solution to multi-class classification problems. In a large number of problems, we use an SVM classifier to solve the wheat kernel identifier use case. The SVM classifier is essential in solving binary classification problems. Nonetheless, here, in the wheat kernel identification case, multiple classes are three types of classes of items in the original dataset. This work focuses on learning the approaches for improving the results by using the support vector classifier and applying legitimate hyper-parameter tuning to get generalized output.

M9-9-9

1. INTRODUCTION

Cereals are important in our lives. Food is a necessity that has its diversity. Various types of people with various cultures need different types of cereals. Here, our focus is specifically on wheat grains and their types. High-quality seeds of wheat seeds are essential for quick, natural, and innovative advancement. Ideal seeds are guaranteed by their variousness and the degree of damage to the kernels. Firstly, the seeds ought to be of a variety described by high efficiency and appropriate for the climatic conditions where they are grown. Also, the seeds ought to have an elevated degree of germination capacity and be free from the damage of biotic and abiotic character. To fulfill the above factors, it takes at least a decade or two. The possibility for variety on average diminishes by 20% per year, and a couple of years after the fact it is supplanted. To maintain the variety and its conciseness, we need to take some drastic measures. Transitions to hybrids reduce the time of getting new seeds to five years, yet the agricultural economy is subject to breeding centers. The grains that are already harvested are not useable as seeds. This situation with seeds makes us hunt for new ways to deal with the improvement in their quality.

In this assignment, we propose to talk about one methodology for the acquisition of seeds with further refined properties. This method gives us a definite picture of the review, the identification of wheat, and its categories. There are three types of wheat, for example, Rosa, Kama, and Canadian, which we have proposed to distinguish. It is often said that, in an organism, maybe it be in one form or another, "all is reflected in everything". In practical breeding determination, the financially valuable properties of plants are most frequently connected with their morphological features. These features must be in a kernel and have them as a seed. These ideas were suggested quite frequently in the past, yet there were no sources for the validations of these connections, nor strategies for the mathematical depiction of the morphology of kernels, permitting to identify of the characteristic attributes or pointers allocating specific batches of kernels to the morphological kind, diversity, or cultivar. In today's world, the improvement of such strategies helps us to solve problems that have not yet been determined and understood before by anyone.

In the following part of this project, the expected definitions essential for the determination of morphological characters, and the obtained results are depicted. Until now, there exist mature technical potential outcomes of the rapid Identification of the type of kernels and high-speed separation of grain flow in light of chosen elements or properties. The choice of

kernels, whose shape adjusts to the particular type of variousness, will guarantee long-lasting supervision of that variety.

2. MOTIVATION

This project aims to identify wheat type based on its kernel properties which would help an individual to discover the type of wheat.

In the agricultural sector, agricultural produce market committees and corporate sectors that buy wheat in bulk need to identify the types of wheat they buy. Here wheat type identifier plays a major role to identify the type of wheat.

Also, in the field of Agricultural studies, this application can help the students to distinguish the wheat types based on their properties from which they can learn more about wheat and its types.

3. PROBLEM STATEMENT

To develop an application that can be used to classify wheat seed types named Canadian, Rosa, and Kama from each other i.e [Multi-Class Classification] according to their Kernel features/properties using Support Vector Machine Classifier.

4. PROBLEM DEFINITION

India is one of the fundamental makers of wheat in the entire world. There are different sorts of wheat like Triticum Vulgare or aestivum), T. -durum, T. -compactum, and so on.

These kinds of wheat can be distinguished by their kernel properties, so by utilizing machine algorithms, we can recognize different sorts of cereals by their size like tallness, width, area, etc.

To advance an application that can be used to gather such sorts of wheat grain, specifically, Canadian, Rosa, and the Kama from one another, for instance, Multi-Class Classification as shown by their Kernel features/properties using Support Vector Machine Classifier.

5. OBJECTIVES

The principal objective is to classify the types of wheat based on their kernel properties into 3 distinct types. Here Support Vector Machine Classifier is utilized for Multiclass Classification. SVM Classifier acts great with the multiclass problem, so after the user enters appropriate values in the frontend in the separate fields, he/she will obtain a result i.e., type of Wheat-based on kernel properties entered by a user. India is one of the leading makers of wheat in the entire world. There are different types of wheat. These types of wheat can be classified based on their kernel properties, so with the help of machine
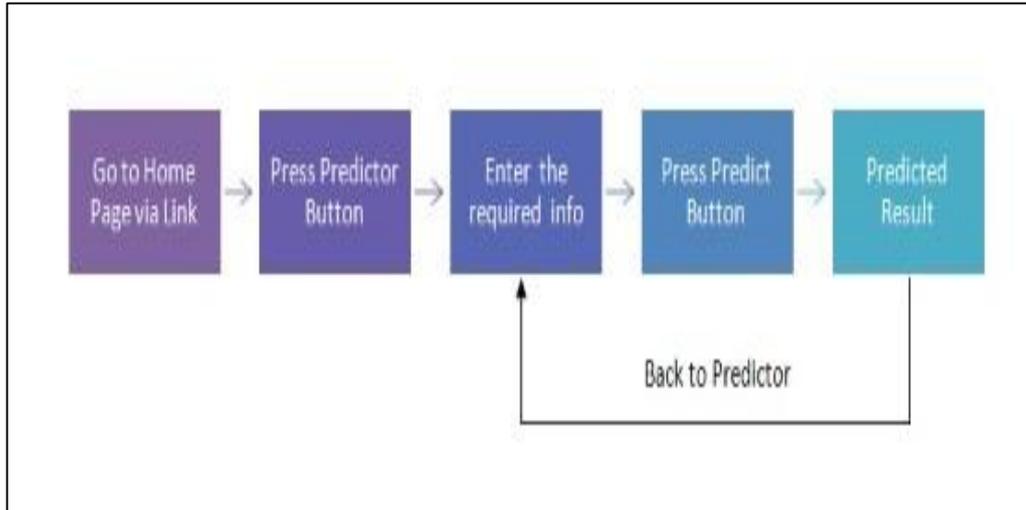
learning algorithms, we can classify the various types of wheat based on their kernel properties like area, perimeter, etc. So, to classify the wheat based on its kernel properties, we utilize a designed application that makes the framework more lovely and fundamentally more secure.

6. FIGURES AND TABLES

Fig.1 Architecture design.

7. USE CASE DIAGRAM

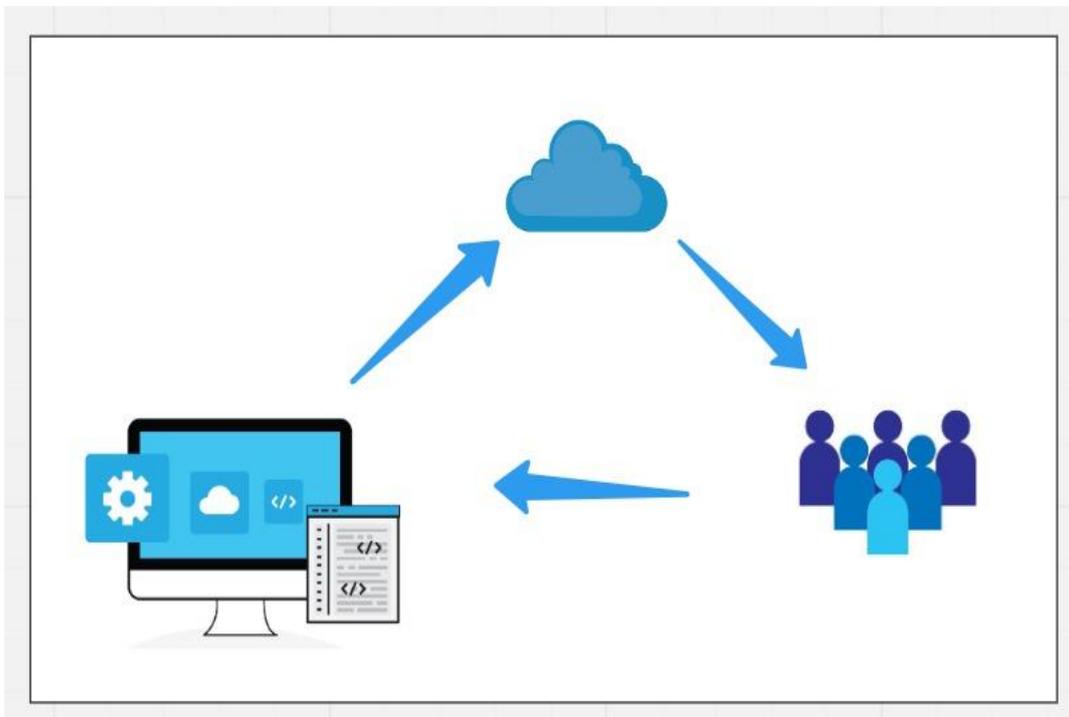*Fig.2 Use Case Diagram.*

8. FLOW OF APPLICATION

Fig.3 Flow of Application.

M9-9-9

9. PROPOSED SYSTEM

9.2 METHODOLOGY

• University of California and Irvine (UCI) Machine learning repository provides data sets of 3 types of wheat, mainly, Rosa, Kama, and Canadian.

• Libraries used for the development of the application are, Pandas for loading the dataset, feature engineering, and data pre-processing. NumPy for numerical values and arrays, sci-kit learn is used for statistical modeling and classification, and clustering for our project.

• To build a successful model, we have tried various machine learning algorithms such as Random Forest, Decision Tree, and Support Vector Classifier (SVM) Classifier.After running our models through various algorithms, we found that in Random Forest the data pre-processing takes more time as there are thousands of decision trees and hence it takes longer to generate the output. By using a decision tree, our model gets overfitted compared to SVM.

• After analyzing all the results from various algorithms, we found that SVM provides us with better results and optimal accuracy.

• To get a generalized model hyperparameter tuning is done.

• We store the data in a serialized format to get constant accuracy. This can be achieved by storing the data in the format of a pickle file.

• For the proper user, the experience application is developed using HTML and CSS.Psycopg2 is used for Database connectivity and the Postgres database is used for storing the input values with results.

• We can also display the result set in statistical and graphical format using the matplotlib library.

• Heroku Cloud platform, provided by salesforce is used for the deployment of the project which is completely free. Heroku runs the required files in its own environment.

10. ALGORITHM

Step 1: First, we need to open the designed application to distinguish the wheat into different types.

Step 2: There are various links like info, about the developer, and how-to on the home page to get insights about the developer and the application.

Step 3: There is a predictor button on the home page that the client can use to navigate to the main application.

Step 4: After Navigating to the main page user is expected to enter the appropriate values in the respected fields like area, compactness, perimeter, etc.

Step 5: If the values entered are validated properly, then the user can press the predict button to get the desired results.

Step 6: There is also a reset button that gives a client a free hand to re-enter the values.

Step 7: If values are not validated, go to step 4.

Step 8: Press Predict button to obtain the results.

Step 9: After getting the result there is a back button on the final page through which the user can navigate back to the application page. To continue to follow from step 4.

Step 10: Stop.

11. ABOUT WHEAT GRAIN

- What are the dimensions of wheat grain?

Across all genotypes and environments, individual grain size (projected area) ranged from 2.67 to 29.98 mm2. This illustrates the big selection of variations in grain size.

- How do you measure wheat?

To begin to estimate wheat yield, willy-nilly choose a minimum of five spots during a field to be averaged for taking measurements. At every spot, count the number of harvestable heads from five feet of row. From that very same spot, count the number of kernels on a minimum of five heads and take the typical.

- What is the compactness of wheat grain?

Unit volume weight shows the density and compactness of a given volume of grain; a base test weight prerequisite is for the most part one of the essential particulars utilized in wheat reviewing and characterization. Test weight or density in the wheat kernel is an actual

quality trademark thought about principally by flour and semolina mill operators. By and large, high weight (as per wheat class) may demonstrate a grain test's solid and ideal appearance, while low weight can happen because of at least one antagonistic occasion, for example, bug harm, heat pressure, or deferred reaping.

- Area and Perimeter of the wheat kernel?

Area in the range 14-18 mm2, perimeter 14-15 mm, major axis length 4.2-5.1 mm, and minor axis length 1.8-2.2 mm.

- What is the length width and thickness of the wheat grain?

In a wheat kernel, three chief aspects are ordinarily estimated length, width, and thickness, which regularly are resolved to utilize a Micrometer or Caliper are revealed in millimeters. The principal axial aspects of grains are helpful in choosing strainer separators and for the estimation of extraction rate during size decrease. These estimations can likewise be utilized to work out the volume of kernels, which are significant during the demonstration of grain drying, air circulation, warming, and cooling. The impacts of size and surface region on drying paces of particulate materials can likewise be described by utilizing the surface-to-volume proportion. The portions at the spikelet had different individual masses. The components of the wheat kernels inside a plant fluctuated fundamentally and the improvement rates and aspects of pieces are unique.
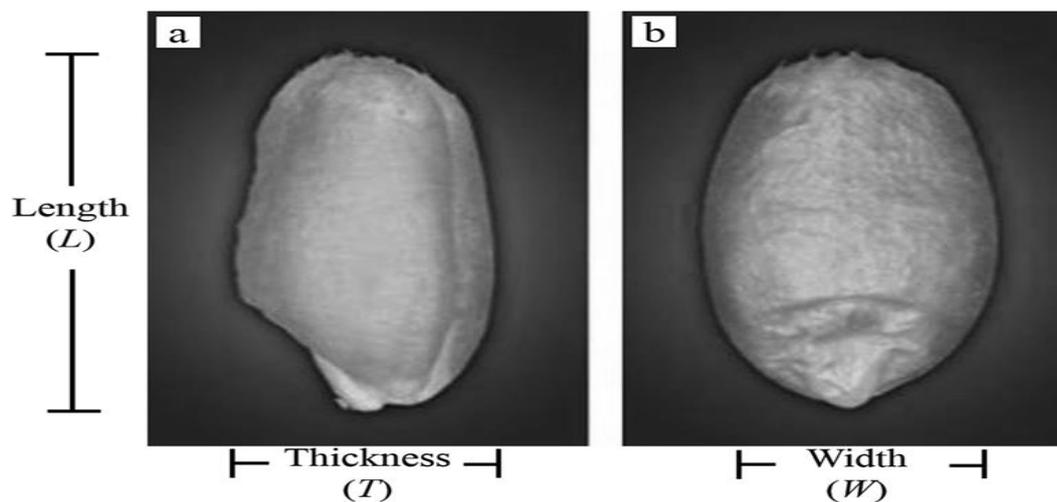

Fig.4 Axial dimensions of the wheat kernel (a) Length (L) and Thickness (T); (b) Width (W)

Image Source- https://www.intechopen.com/chapters/53879

12. RESULT

After the user enters the values, the result is predicted according to the parameters entered. The parameters and the results are stored in the database as given below: -

Area	Perimeter	Compactness	Kernel-length	Kernel-width	Asymmetrical Coefficient	Kernel Coefficient	Wheat Type
11.5	15.6	0.87	5.24	3.32	6.45	5.21	Canadian
11.5	18.65	0.845	5.24	3.21	1.44	5.29	Kama
11	20	0.91	5.9	4	6.39	6.6	Rosa
11.5	14.89	0.798	5.1	2.75	1.56	5.52	Canadian
10.059	21.83	0.845	5.24	3.761	6.99	4.5	Kama
11.05	15.25	0.849	5.094	3.519	2.55	5.75	Canadian
11.8	20.5	0.845	6.62	3.9	4.55	6.52	Rosa

13. ADVANTAGES, LIMITATIONS & APPLICATIONS

ADVANTAGES

1. It is secure and simple to use.
2. It tends to be useful in the agricultural sector and agricultural field where research and development is a fundamental factor.
3. It is open effectively and is educational.
4. Various types of wheat types can be identified using simple parameters.

LIMITATIONS

1. To use this application, client/user needs a stable and smooth internet connectivity.
2. The Client is required to insert proper valid parameters; else the output would be irrational or irrelevant.
3. A pre-requisite knowledge regarding wheat is required, hence to perform a valid and desired prediction.

M9-9-9

APPLICATIONS

1. It is useful for an individual, who doesn't have any knowledge regarding wheat types, and wants to explore its uses and types.
2. Agricultural Committees and Corporate Sectors can make extensive use and the application might come in handy to them.
3. The people who actually have their business in this sector can make proper use of the application.

11. PROJECT SCOPE

This application is planned to mark different kinds of wheat in light of variables like compactness, asymmetric coefficient, etc. Using Machine Learning, we can separate the wheat into different sorts of given variables. So, the client is supposed to open the application using the interface. To view application information and data or various content, please check the link. Then, at that point, using the predictor button, the client proceeds to the standard application. Then, the client enters the evaluations in individual fields. After the assessments are arranged and precisely approved, the customer has the choice to press the submit button to see the outcome. On the occasion, that the client needs to change all of the gauge values, there is a decision of the reset button on the screen that can help the client with resetting the gauge values entered ahead of time. So finally, after the button is squeezed, the client can sort the wheat into specific sorts like Rosa, Kama, and Canadian, in light of the assessed values entered by the client. To get back to the principal application there is a button on the screen to return to the application. So, in the wake of returning, the client is free to again enter the quantities of his choice to obtain the ideal result. The assessments and the results entered by the client are kept in the Postgres database.

12. CONCLUSION & FUTURE SCOPE

FUTURE SCOPE

Currently, we have considered the dataset which has three types of wheat, namely, the Kama, Rosa & Canadian, but, in real-time there are many types so we can add more wheat types.

There can be some more factors of wheat that can contribute to the model, so, by doing some extensive research we can add some more things that can help us to do so.

M9-9-9

Here, we have only considered wheat other than that there are several grain types like jowar, bajra, rice, corn, barley, millet, etc. which we can add in the app.

CONCLUSION

Thus, we can use the Support Vector Machine Classifier to solve the Multi-Class Classification problem. In our case, the Wheat Kernel Identification problem has three classes, namely Rosa, Kama & Canadian.

By using SVM and doing proper Hyper Parameter Tuning, we can classify types of wheat (Rosa, Kama & Canadian) based on their kernel properties.

ACKNOWLEGEMENT

REFERENCES

[1] *Multi-class Support Vector Machine Application in the Field of Agriculture and Poultry: A Review Nurhanna, A. A. ∗ and Othman, M. F. Center for Artificial Intelligence & Robotics, Malaysia-Japanese International Institute of Technology, Universiti Teknologi Malaysia, Malaysia.*

[2] *Review on Classification of Wheat Grain Using Machine Algorithms Meesha Punn, Nidhi Bhalla*

[3] *Li, Q. Y., and Hui, Y. H. (2019). Kinetic models of peroxidase activity in potato leaves infected with late blight based on hyperspectral data. Int. J. Agric. Biol. Eng. 12, 160–165. doi: 10.25165/j.ijabe.20191202.4574*