Non-invasive Soya Bean Seed Analysis Using Machine Learning

Amrutha Patil, Shashikumar G. Totad

Abstract: ***The soya bean is economically the most important legume in the world. Therefore, it is important to grow good quality seeds for a better yield. Identifying the right set of seeds is a difficult task when done manually since, there are no definite external characteristics of soya bean that correlate with its germination potential. Therefore, in this work an attempt is made at correlating the physical properties of soya bean with its germination potential using the concepts of machine learning and image processing. The input here being images, there are different methods to take images of soya bean, that is by using digital camera or radiography. The pros and cons of these methods are discussed. Since, using radiography images is not cost-efficient and its local availability for research purpose is scarce, a digital camera is used to take soya bean images. Once the image dataset is available, different classification methods are employed to classify the images into 'germinating' and 'non-germinating' seeds. The classifiers used are CNN, KNN and SVM and the average accuracy of the classifiers is 66.17%. The performance of different classifiers is analyzed to find the most suitable classifier. It is observed that most of the 'germinating' seeds have intact seed coat, elongated spherical shape, smooth texture and are evenly colored. Whereas, the other half has damaged seed coat, flat shape or not completely spherical, are unevenly textured and discolored at parts. Finally, the suggestions are made to improvise the results.***

Keywords– CNN, germination potential, KNN, machine learning, non-invasive, radiography, seed analysis, soya bean, SVM.

I. INTRODUCTION

Agriculture is known to be one of the most significant economic activities in India and farmers are the backbone of it. Soya bean is a staple in the diets of many people in India. It provides vegetable protein for millions of people and ingredients for hundreds of chemical products. Hence, it is important for both agriculturalists and consumers that good quality soya bean are grown. There are no definite external characteristics, which assure that a seed will germinate or not.Thus, it is generally difficult to predict the germination potential of the soya bean based on the external look of it. Usually a germination test of sample seeds is done before sowing, that determines the germination potential of that seed lot. This test takes about a week, is tedious, requires laboratory equipment and demands precision; therefore, experts usually do it. Hence, a more generalized, fast and convenient method is required to analyse the soya bean quality.

Many parameters determine the health of Soya bean and having those parameters in a good condition is important. Few of the significant seed parameters are its Embryo, Moisture content, Temperature, Dimensions, Mass and density, Colour, texture etc. (Embryo is the part of the seed, which further grows into root and shoot of the plant. Cotyledons serve as a stored for the growth of embryo). The embryo of Soya bean is small and fragile in nature. So, sometimes even when the seed is fit externally, if the embryo is harmed then there is a high probability that it won't germinate.

Based on different seed parameters there are many existing methods that test the germination potential of a seed, but most of them are invasive in nature. Meaning that, the seed's structural integrity is disturbed in some way (say, by cutting it), which make is unusable for further growing into plant. Thus, in this research we aim at taking a non-invasive approach to identify the germination potential of a seed. In this work, the seeds with good germination potential are addressed as 'germinating' and 'non-germinating' otherwise. The scope of the term agriculture is broader than it is generally anticipated to be. Since this is an era of technology, it is not new to incorporate programming knowledge in agricultural field. Machine learning – ML is a highly growing domain that can be utilized in bringing profit to the agriculture sector. Because ML exploits image, it serves exceptionally well to the cause of non-invasiveness. With this, the process of identifying seeds with good seed quality will be simpler and more fruitful. Convolutional neural network, Support vector machine and K nearest neighbors are some of the ML classifiers that are used in this work.

II. RELATED WORK

One of the basic technique to take images for this research would be by using a digital camera like DSLR. The camera captures the external appearance of the seeds and gives details about characteristics like the dimensions - length, breadth, width, sphericity (to define the how spherical the seed is), color and texture of the seeds. These physical properties of the soya bean are used to classify and determine the germination potential of the seed. For this, it is necessary to understand the correlation to other physical properties like moisture content and temperature, which play a major role in its germination potential.

From the previous work, it is know that the shape of grain and characteristic dimensional properties such as length, width, thickness, the arithmetic average diameter and the geometric average diameter, increased linearly depending on the increase of moisture content [1], [3]. The mass,

Revised Manuscript Received on Janurary 25, 2019.

Amrutha Patil, School of computer Science and Engineering, KLE Technological University, Hubballi, Banglore, India

Dr. Shashikumar G. Totad, School of computer Science and Engineering, KLE Technological University, Hubballi, Banglore, India

roundness, sphericity and terminal velocity of the soybean seeds also increased with increase in moisture content [1], [4]. Further, the degree of sphericity (N), was determined using the equation from Moshesin [1], [4] where length, L; weight, W; and thickness, T are the three linear dimensions [1].

$$N = (LWT)^{1/3}/L$$

Other important conclusion from the previous work were that the True density, bulk density and porosity decreased linearly depending on the increase of moisture content [2], [3], [4]. Hence, it is necessary to have the right and consistent moisture content to acquire accurate results.

Another prominent parameter of soya bean is temperature. It primarily effects the seed's moisture, its enzyme activity, its hormone production and in determining its dormancy period. Therefore, classification can be done using the temperature property of the seed too. Infrared Thermography – IRT can diagnose the developmental stage of a germinating seed, noninvasively and in real time [5]. The conduction of this experiment require special equipment (Thermal cameras) and precise monitoring from experts. Thermal images are highly reliable in identifying the healthy, dead and infected seeds. Thereupon it can contribute to a better seed analysis.

Embryo is the single most important part of a seed and since it is covered with seed coat and cotyledons, it is difficult to image it with conventional methods. So, X ray was employed to obtain the internal image of the embryo. The shape and intactness of the seed embryo can tell a lot about its germination potential. This requires the usage of special devises like X ray machine, CT scanner etc. There are essentially two types, namely 'hard X rays' and 'soft X ray' [7]. The X ray machine uses soft X rays.

The basic approach to take the radiography images is using X ray machine. From the referred papers it is now known that, X ray can give an insight on the other properties like endosperm content, free area etc. [6], which can be prove beneficial in classification. Also, it gives information on seed irregularities, malformations, diseases and improved seed component positioning [8].

Some of the key downsides of using X rays were that they were suitable for only flat seeds like chili, cucumber, watermelon [6], [7]. Other problems faced with this were that normal/wet seeds have high density and high moisture content, therefore dry seeds were used [7].

Another way is to use three-dimensional X rays, which would be suitable for spherical seeds [8] and also would provide Better detection of seed irregularities, diseases and improved seed component positioning. However, this includes high price, and more time for structure building of seed [8].

Non-destructive imaging of seeds was demonstrated using the synchrotron-based X-ray imaging technique, diffraction enhanced imaging – DEI [9]. The three principle ways X-rays interact with matter in DEI are absorption, refraction, and scatter. The seed images obtained had good contrast and definition, allowing anatomical structures and physiological events to be observed [9]. Novel anatomical and physiological observations were made that would have been difficult to make using other techniques [9]. Furthermore, as high-energy X-rays are used, little X-ray absorption occurs, resulting in low levels of radiation damage [9]. The challenge faced here is the availability of synchrotron (only two synchrotrons: Indus-1 and Indus-2 in Indore, India).

III. IMPLEMENTATION METHODS

Proposed methodology

One of the most realistic approach to this problem is by using digital camera images to classify the soya bean seeds. It works on the external physical property of the seeds and takes into consideration the dimensions and color factors of the seeds. In this research, we will use the term 'external properties/characteristics' to signify color and dimensions of soya bean.

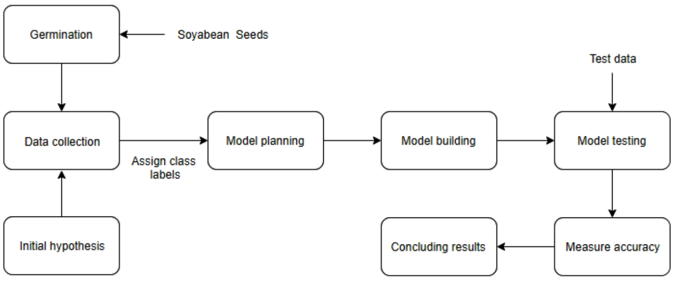

Figure 1: Flow model of proposed methodology

We begin with an initial hypothesis in our mind that – 'A correlation exists between the germination potential of the Soya bean seed and its external properties'. Nonetheless, the probability of any soya bean seed's germination is one-half even when guessed by humans who are well versed in this domain. Knowing this, based on the results of an automated classification, the hypothesis can either be accepted or rejected.

Firstly, the soya bean seeds of variety 'DSP 21' were procured from the Karnataka Rajya Beeja Nigama in Dharwad. At the time of research, the moisture content of the seed lot was 11.5-12.5 percentage. The weight of the seeds was 1 Kg and the number of seeds utilized was 1950. The seeds were indexed and they were imaged using a Canon Camera accordingly. These images of soya bean seeds serve as our primary dataset.

Then the germination test was conducted on those indexed seed. (1) The germination sheets were soaked in water for few seconds and then the excess water was drained. (2) The seeds were carefully placed on the wet germination sheets according to their labelled indices. (3) After this, another germination sheet was laid on top of first sheet. (4) Henceforth, the sheets were rolled and placed in a polythene covering. (5) This process was repeated until all seeds were used. (6) The rolled sheets were watered every alternate days. (7) After 5 days, the sheets were uncovered and results were recorded. (8) Based on whether the individual seed has germinated or not, labels of 'germinated' or 'not germinated' were assigned respectively against their indices.

Ultimately, the final dataset of labelled images was obtained. This dataset is divided into train and test dataset

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

(the labels from test dataset was removed). This dataset is first pre-processed and then used for classification.

The model planning was done for selecting appropriate classifier models. It involved pre-processing the dataset by resizing, renaming and cropping the images to a correct proportion to maintain uniformity. One of the important step is to do feature extraction on the dataset to get a common vector form of images for processing it with different classifiers. Thereafter, selection of appropriate hyper parameters was done by trying out some different combinations of parameters. Once the right parameters were found, the models were ready to process.

Hereupon, model building was done where the classifier models were built using python code and run on the training dataset. Once the training was complete, the model was then tested for its accuracy. Lastly, performance of different models were analyzed and the best amongst them was selected as our classifier for prediction. The classifiers used are (i) Convolution Neural Network – CNN, (ii) Support Vector Machine – SVM and (iii) K - Nearest Neighbors – KNN.

Alternate methodology

One of the other approaches that can be employed is by using radiography images instead of regular camera images. Radiography images give a better insight at the internal property of the seed like embryo size, tissue density, amount of endosperm, free area inside the seed, irregularity in shape etc. This indeed will provide a better shot at accuracy and better understanding of seeds behavior.

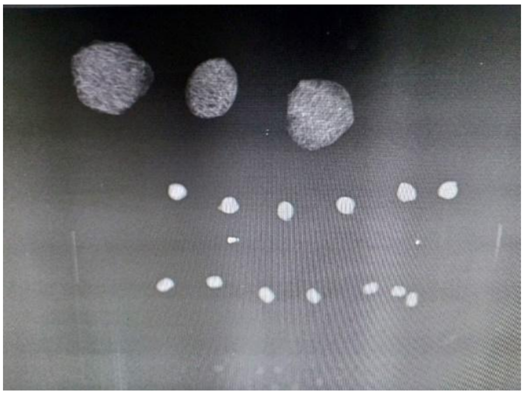

Figure 2: X ray image of soya chunks, chickpea and soya bean

To work on the radiography of seed, a sample of soya bean and other seeds (like chili, chickpea, and soya chunks) were taken. The X-ray device from KIMS, Hubballi was utilized for this purpose, which made use of 'soft x-rays'. The X ray image of the seeds acquired was as in Figure 2. This approach was rather challenging implementation wise since, such work demands special seed conditioning. Seeds are required to be dry since high moisture in the seeds could results in blurriness/smudginess in X-ray images. Penetration of X rays is difficult in high-density seeds hence this should be taken account of. Priming and Image contrasting are few of the methods to avoid such problems. Additionally, the radiation of soft X ray are low in energy (up to 12 kV) and are not powerful enough to explicitly differentiate the parts of soya bean. Furthermore, the X ray devices are to be operated by experts only and the usage of such rays can be harmful and damaging for the seeds.

Another way to have radiography images is using Computed Tomography – CT. It gives the advantage of having a 3D imaging of seeds. The CT device revolves around the object and captures multiple dimensions of it. Hence, this would work wonderfully for spherical or irregular seeds like cabbage, soya bean etc. The drawback here is that it takes around 20 minutes for imaging one object therefore; it would be not suitable to image huge number of seeds.Magnetic Resonance Imaging – MRI is also good option for considering radiography images but given the cost of a single MRI, this method would be expensive and not cost efficient of large imaging size. The availability of these devices for the purpose of research is rare. In addition, if this approach is adopted at classification in a real time situation - every time someone wants to test the seed quality of soya bean, they need to take the X ray image of soya bean sample first; which is not feasible. Hence, working on this approach may be challenging.

IV. RESULTS AND DISCUSSIONS

The results of the classifier are tabulated in Table 1. From the table we can say that the CNN and SVM bearing the highest accuracy proves to be better classifiers. It is important to note that KNN and SVM's accuracy are close to the CNN's accuracy and have done a good job in classification of the images.

Name of the Classifier	Accuracy
CNN	66.85%
KNN	65.37%
SVM	66.30%
Average accuracy	66.17%

TABLE I. Accuracies of different classifiers

Hence, we can say that an automated prediction of germination potential is possible using machine learning with the highest accuracy being 66.85%, which is better than the guessed prediction of a human (50% accuracy). With this we accept the initial hypothesis and state that, - 'A correlation exists between the germination potential of the Soya bean seed and its external properties'.

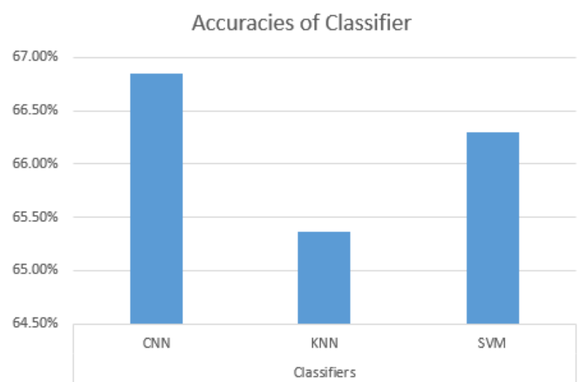

Figure 3: Performance comparison of seed classifiers

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

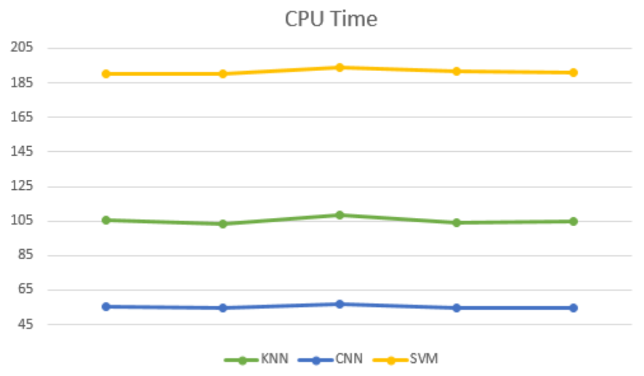

Figure 4: CPU time (speed) of seed classifiers

Though all three classifiers have similar performance, some of the reasons for obtaining such results as above could be (1) Because of differences in appearance of seeds, there is large variation in dataset images. (2) CNN is a state-of-the-art classifier in the machine learning domain. The use of neural networks and backward propagation makes its training stronger and a good classifier. (3) CNN takes lesser time for classification than other two. It is relatively faster classifier. (4) SVM works by creating hyperplanes that separate the data cluster and are excellent for binary classification but large variations in data makes it difficult to find the right hyperplane. (5) CNN and SVM have relatively better feature space than KNN.(6) KNN unable to form reliable neighbors when there is a large sparsity in data. Thereupon it can be stated that CNN has better performance than SVM and KNN and is most suitable for seed classification.

From observing the images of dataset, seed quality analysis of germinating and non-germinating seeds can be done to identify few of the correlating external characteristics of soya bean. It is observed that most of the 'germinating' seeds (1) have an intact seed coat, (2) are elongated spherical in shape, (3) smooth in texture and (4) are brighter and evenly colored. Whereas, the 'non-germinating' seeds have (1) damaged seed coat – peeled off, cuts or scratches, (2) flat shape or not completely elongated spherical or in shape and tiny in size, (3) are unevenly or roughly textured and (4) are darker in color or discolored at some parts.

V. CONCLUSION AND FUTURE SCOPE

Conclusion

Machine learning is an innovate way to determine the germination potential of soya bean non-invasively; by using classification. Images from digital camera can be used to capture the physical properties of soya bean. CNN outperformed SVM and KNN with an accuracy of 66.85% while the others being 66.30% and 65.37% respectively. Additionally, the speed of CNN was higher than both SVM and KNN. Even though KNN is a lazy learner, it performed considerably faster than SVM. A correlation is possible between the germination potential and external properties of soya bean. The germinated seeds seem to have an intact seed coat, were elongated spherical, were smooth in texture and had even bright color. Having radiographic images is a clever alternate approach but unavailability of radiography devices, their high cost and large imaging time makes it unsuitable for seed classification.

Future Scope

More research is needed in this interdisciplinary field of machine learning and agriculture. The major concern during this research was the unavailability of dataset therefore, a more professional approach can be employed to image the soya bean seeds. Additionally, different classifiers can be explored which may perform better than CNN, SVM and KNN. Further, it would be advantageous if tools like CT and MRI are recognized and made accessible for research purpose. More non-invasive methods like DEI should be encouraged that minimizes the loss and help attain productive results.

REFERENCES

1. T.Y. Tunde-Akintunde, J.O. Olajide, B.O. Akintunde. (2016). Mass-Volume-Area Related and MechanicalProperties of Soybean as a Function of Moisture and Variety
2. Engr. Onu John Chigbo. (2008). Selected Physical Properties of Soybean In Relation To Storage Design
3. H. Kibar, T. Öztürk. (2005). Physical and mechanical properties of soybean
4. Sachin Vilas Wandkar, Pravin Dhangopal Ukey, Dilip Ananda Pawar. (2012). Determination of physical properties of soybean at different moisture levels
5. Ilse Krannera, Gerald Kastbergerb, Manfred Hartbauerb, and Hugh W. Pritcharda. (2010). Noninvasive diagnosis of seed viability using infrared thermography
6. Dr. Henry Bruggink. (2012). X-ray based seed analysis and sorting
7. Bruggink, Henry & Van Duijn, Bert. (2017). X-ray based seed analysis. Seed Testing International. 45-50.
8. G. Gomes jr, Francisco & Van Duijn, Bert. (2017). Three-dimensional (3-D) X-ray imaging for seed analysis. Seed Testing International. 154. 48-52.
9. Lester W. Young, Christopher Parham, Zhong Zhong, Dean Chapman, Martin J. T. Reaney; Non-destructive diffraction enhanced imaging of seeds, *Journal of Experimental Botany*, Volume 58, Issue 10, 1 July 2007, Pages 2513–2523