***PISTACHIO NUT CLASSIFICATION USING ARTIFICIAL NEURAL NETWORKS***

*Ankur Jain, Rishabh Gupta, Uddhav Davey*

*School of Computer Science and Engineering, VIT Bhopal University, Bhopal, India*

**Abstract**

Pistachio is a shelled fruit from the anacardiaceae family. The Kirmizi pistachios and Siirt pistachios are two of the major types of pistachios grown. This study aims to identify these two types of pistachios by classifying them using the machine learning technique of artificial neural networks. The dataset consists of different physical parameters like area, perimeter, and eccentricity, which are important to our classification process. An advanced classifier model was designed on this dataset, and a multi-level system including feature extraction, encoding, and scaling has been proposed. The presented high-performance classification model provides an important need for the separation of pistachio species and increases the economic value of species. In addition, the developed model is important in terms of its application to similar studies.

**Introduction**

Pistachio is an agricultural product native to the Middle East and Central Asia. The world’s major pistachio producers, contribute close to 90% of the total production worldwide. Pistachio production in Turkey covers a large number of types in different names. Among these pistachio types, the most preferred ones are Kirmizi and Siirt. In order to keep the economic value of pistachio nuts, which have an important place in the agricultural economy, the efficiency of post-harvest industrial processes is very important. Pistachio kernel, which is a good source of fat (50%–60%), contains unsaturated fatty acids necessary for the nutrition of humans. It is widely used in the manufacture of confectionery and snack foods. Due to the dark green color of its kernel, pistachio is highly preferred in ice cream and pastry industries. The shell (endocarp) of the pistachio is hulled along its seams. This is desirable since pistachios are often marketed in their shells to be eaten by hand as a kind of snack food. Different pistachio species address different markets, which increases the need for the classification of pistachio species.

**Motivation**

Pistachio nut can be counted as one of the most significant commodities. Artificial neural network models have been shown to achieve better performance than conventional statistical classification techniques in many applications. There have been several successful applications of artificial neural networks in agriculture fields. The neural network applications involve Multilayer Perceptron architecture. As MLPs instructed with the standard back propagation algorithm are effective pattern classifiers researchers have used this system in many applications.

Our MLP architecture has the ability to classify the non-linearly separable feature data measured from the pistachio samples using the non-linear processing elements in the hidden and output layers of the network. It could be trained to produce a correct target output when presented with the corresponding input pattern. The specific objectives of this research are to develop a neural network classifier, which could classify the two major varieties of pistachio nuts and to evaluate the performance of this classifier using various performance metrics and visual analysis techniques.

**Focus and Purpose**

Pistachio nuts are commonly classified based on their size, color, and shape. The primary reasons for classifying pistachios are:

* Quality control: Classifying pistachios helps ensure that they meet certain quality standards. This is important for both producers and consumers. Producers want to ensure that their pistachios are of high quality, while consumers want to know that they are buying a product that meets certain standards.
* Marketing: Classifying pistachios also helps with marketing efforts. By classifying pistachios based on their size and appearance, producers can target different markets and customers. For example, larger pistachios may be marketed to gourmet food stores, while smaller ones may be marketed to snack food companies.
* Pricing: Pistachios are often priced based on their size and quality. By classifying pistachios, producers can set prices appropriate for each size and quality category.
* Processing: Classifying pistachios also helps with processing. Different size and quality categories may require different processing methods, such as roasting or salting.

Overall, classifying pistachio nuts is an important part of the production and marketing process, helping to ensure quality and consistency for both producers and consumers

**Broad Problem**

In order to keep the economic value of pistachio nuts, which have an important place in the agricultural economy, the efficiency of post-harvest industrial processes is very important. To provide this efficiency, new methods and technologies are needed for the separation and classification of pistachio nuts. Different pistachio species address different markets, which increases the need for their classification. In this study, we have aimed to develop a classification model different from traditional separation methods, based on artificial intelligence and neural networks which are capable of providing the required classification.

A number of different observations were taken, in which the values of many different kinds of features for these two kinds of pistachios were considered, like area, perimeter, eccentricity, skewness, and kurtosis, amongst many others. All in all, the Pistachio dataset had 28 feature columns to help in the classification. The presented high-performance classification model provides an important need for the separation of pistachio species and increases the economic value of the species. In addition, the developed model is important in terms of its application to similar studies.

**Applications**

Pistachio Classification model can be used to classify different types of Pistachio’s varieties using various kinds of models like the VGG-16, VGG-19 and AlexNet.   
The results can be used to identify the more consumed type of pistachio.

**Previous Work**

**OBJECTIVE**

The physical parameters that were identified for grading pistachios were size, shape and external appearance. These parameters were characterized by nut length, diameter, weight, volume and color. So, a pistachio classification system based on an artificial neural network classifier with an external appearance was created.

Ahmad\_aghayee, Akbari, Badami, Fandoghi and Kaleh ghoochi were the types of Pistachios used in this study.

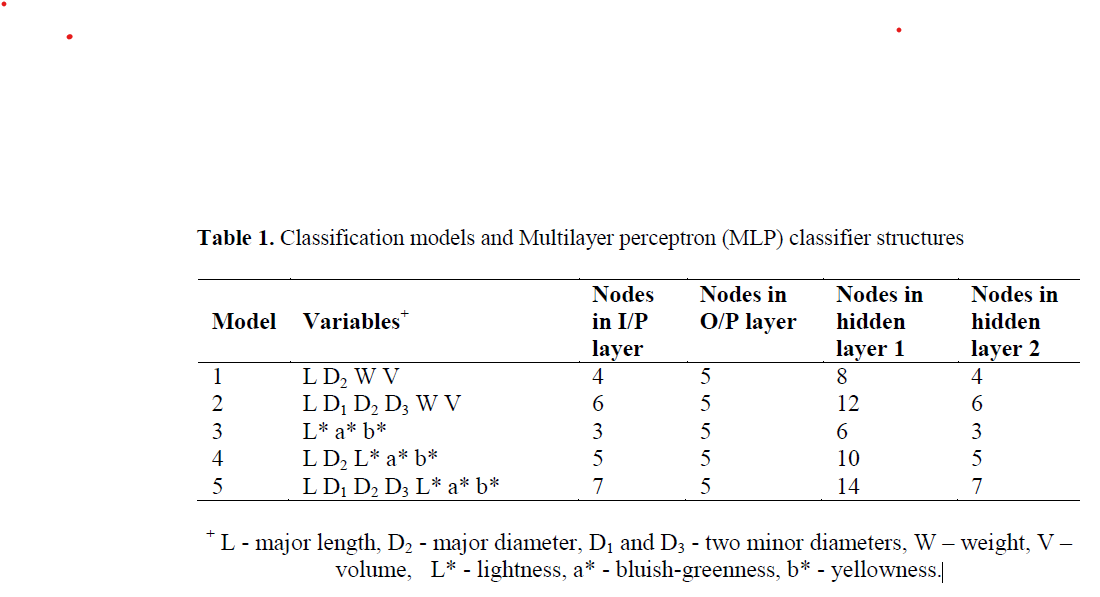
****

**MATERIAL AND METHODS**

In the classification analysis, each sub-sample of the varieties was considered as a separate class with the aforementioned attributes.

* **Classification Models**

From the sub-samples of five pistachio varieties, the data of nine physical parameters were pooled and five classification models were formed with the physical attributes measured from the sub-samples of five varieties (Table 1). These five models were composed of all physical parameters.



* **Neural Network Classifier**

The basic structure of the neural network chosen for classifying the pistachio varieties is a feed forward MLP as shown in Figure:

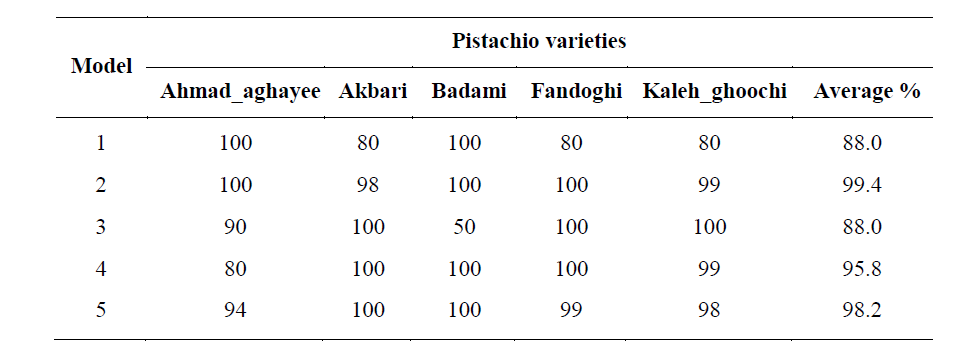


Feed forward MLPs trained with the back propagation algorithm were used in this model. The transfer function and learning rule adopted for the two hidden layers were hyperbolic tangent non linearity and momentum respectively. The transfer function gave the network the computational ability to learn the problem, but the learning rule, which in this case was gradient search, was used to calculate the weight update. Training stopped when the set epochs of 10,000 was reached.

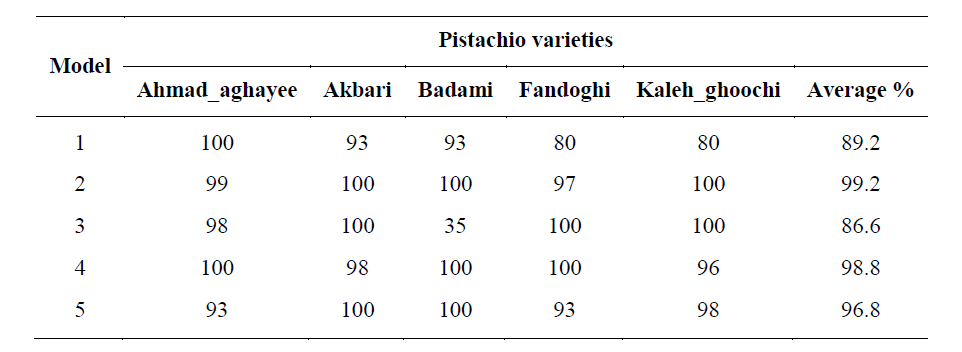
**RESULTS AND DISCUSSION**

Table 1 shows the summarized results of the classification analysis by the MLP classifier, while Table 2 shows classification accuracy by statistical methods.

**TABLE 1**



**TABLE 2**



Out of the five classification models, models 1 gave higher classification accuracies with its accuracy of 99.4%. Thus, the procedure using model 1 features on a neural network classifier was considered for classifying the sub-samples of the five pistachio varieties.

**CONCLUSIONS**

From the classification results, it can be concluded that the performance of an MLP neural network classifier is better than a statistical classifier in classifying varieties of pistachios using selected physical attributes.

**Novelty**

Traditionally, pistachio nuts are inspected/classified via visual inspection of workers, manually. As a result, classification process is subjected to poor efficiency in terms of time and cost. Moreover, visual inspection and classification by hand is a tedious process and may contain various health risks. Siirt pistachio nuts differ from other pistachio species such as Kirmizi pistachio according to their shape, size and taste properties. Our developed Artificial Neural Networks Classification system aims to classify pistachio nuts as Siirt or Kirmizi in a fully automated manner.

**Advantages**

Quality Control: Classifying pistachio nuts helps in maintaining the quality of the product. It ensures that the nuts are sorted and graded based on their size, color, and other characteristics. This helps in ensuring that the nuts meet the required standards of the market.

High Accuracy: ANNs can classify pistachio nuts with high accuracy, especially when trained with large amounts of data. This can lead to more precise sorting of pistachios based on their size, shape, and color, which can improve quality control in the production process.

Efficiency and Speed: Classifying pistachio nuts using automated machines can be more efficient than manual sorting. This can help reduce labor costs and increase productivity.

Consistency: ANNs can classify pistachio nuts with consistency, without being influenced by factors such as fatigue or personal bias, which can improve the reliability of the sorting process. This helps in creating a consistent product that customers can rely on.

Marketing: Classifying pistachio nuts can help in marketing the product. Grading the nuts based on their size and quality can help in creating different product lines that cater to different customer segments.

**Disadvantages**

Data Availability: ANNs require a large amount of data to be trained effectively. If there is not enough data available or if the data is not representative of the pistachio population, then it can result in errors, which can affect the quality of the product by lowering the accuracy of the classification.

Initial Cost and Inaccuracies: Implementing ANNs can require a significant initial investment in hardware and software. This may not be feasible for smaller businesses or producers, and may make it less competitive in the market in the initial stages.

Maintenance: ANNs require ongoing maintenance and updates to ensure they continue to perform well. This can be costly and time-consuming.

Interpretability: ANNs are often considered "black box" models, meaning that it can be difficult to understand how the model is making its classifications. This lack of interpretability can make it challenging to troubleshoot issues as well as improve the model over time.

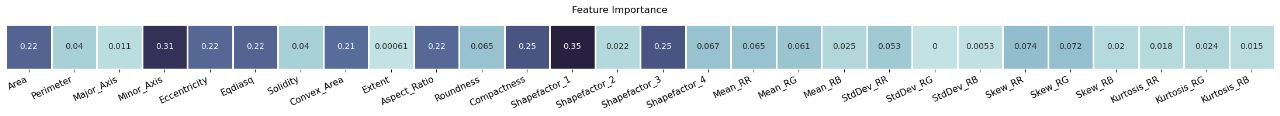
In summary, using ANNs to classify pistachio nuts can provide a fast, accurate, and consistent sorting process. However, it can also require significant initial investment and ongoing maintenance, and may be limited by the availability and representativeness of data.

## METHODOLOGY

**Analysis**

The entire dataset is divided into a Feature Matrix and a Label Vector, which is the output we aim to find. All features are converted to Floats, which is important for many Machine Learning Algorithms. A Training Set and a Test Set are then created, in the ratio of 80:20.

Using our Training Set, we calculate the "mutual information" between each feature in the input training dataset and the target variable for each observation. We do this with the help of K-Nearest Neighbours algorithm. This basically returns an array of mutual information scores for each feature in the dataset, essentially marking the features with the most correlation as well as the least correlation with the target variable. This is represented in the form of a Heatmap containing all of the given features.



We then proceed to consider only those features that have strong correlation with the target output, and perform Robust Scaling on them. This scale features using statistics that are robust to outliers, by removing the median and scaling the data according to the quantile range. We then perform Train-Test Split again using this updated feature matrix, that is scaled robustly.

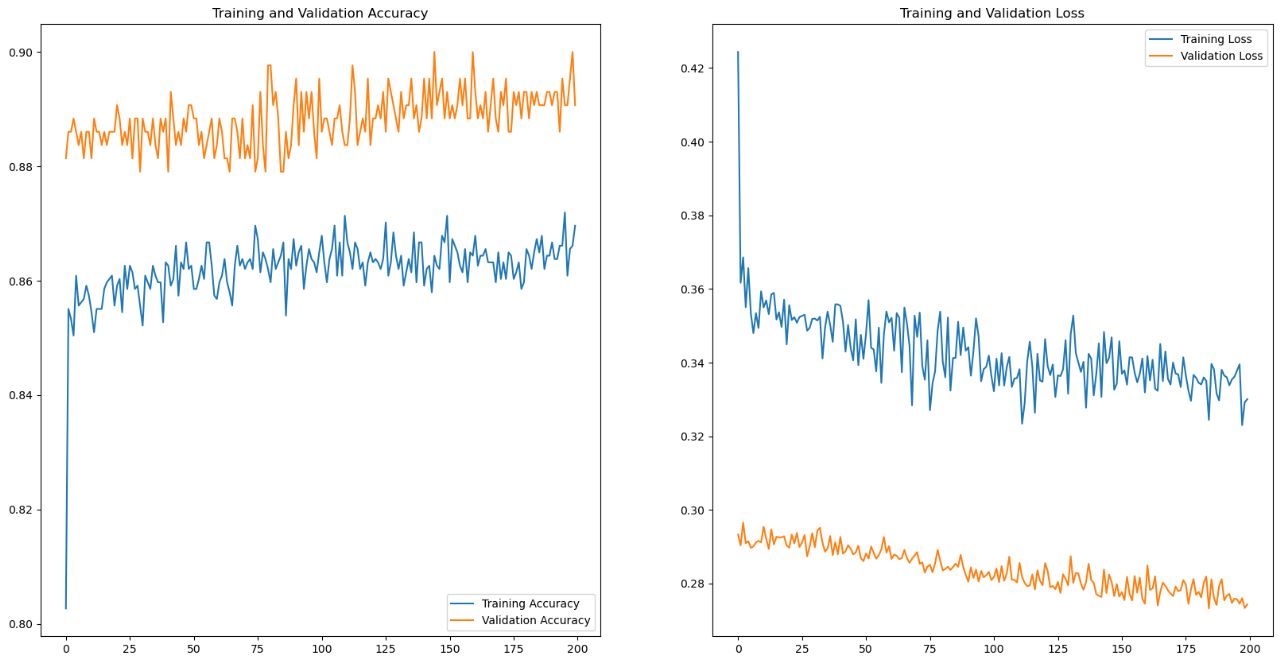
We then finally proceed to build the neural network model using the Keras API in TensorFlow.

## RESULTS

### Comparative study: While defining the Input Layer as well as the Hidden Layer, we use the Scaled Exponential Linear Unit (SELU) Activation Function. SELUs enable deep neural networks since there is no problem with vanishing gradients. Furthermore, in contrast to ReLUs, SELUs cannot die. SELUs on their own learn faster and better than other activation functions, even when they are combined with batch normalization.

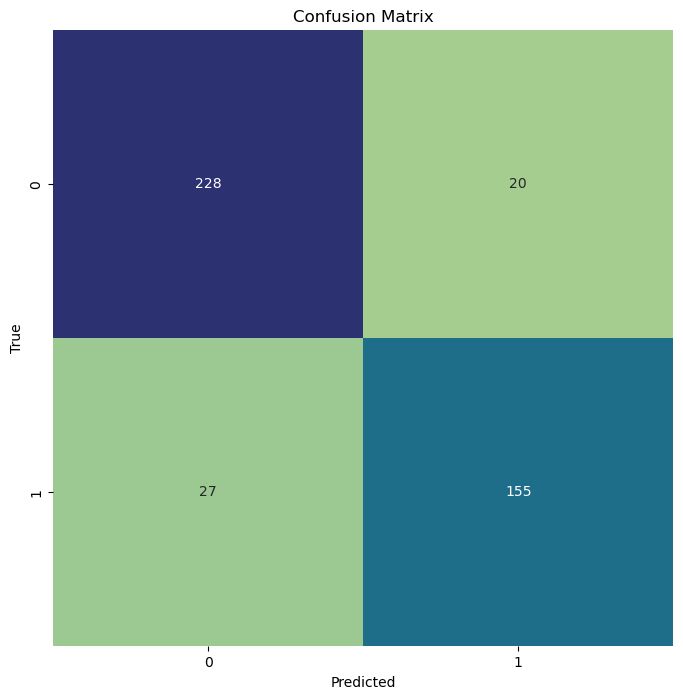
### We also add a Dropout Layer in order to prevent overfitting, before defining the Output Layer with the number of nodes set to the output shape, and the Activation Function set to Sigmoid. This model is compiled using the binary crossentropy loss function and the Adam optimizer. We then proceed to set the number of epochs and train the Keras Model using the Robustly Scaled training data.

### This model achieves an experimental accuracy of 86.55%. The training and validation accuracy and loss of the deep learning model can be observed below:



Our model also generates a confusion matrix to evaluate the performance of the neural networks model on a test set. We predict the target variable for the test set using our model, and round the predicted values to either 0 or 1. This is because the model is likely to predict probabilities rather than binary values, and rounding them gives us binary predictions.

Finally, the confusion matrix is calculated using the predicted and actual target value, using Seaborn's heatmap function.



To understand the confusion matrix, we first need to understand the following terms: true positive, true negative, false negative, and false positive.

True Positive (TP) — model correctly predicts the positive class (prediction and actual both are positive).

True Negative (TN) — model correctly predicts the negative class (prediction and actual both are negative).

False Positive (FP) — model gives the wrong prediction of the negative class (predicted-positive, actual-negative). FP is also called a TYPE I error.

False Negative (FN) — model wrongly predicts the positive class (predicted-negative, actual-positive). FN is also called a TYPE II error.

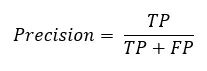
## CONCLUSIONS

We will look at the conclusion for this project with the help of a report. This is a classification report for our ANN model that has been trained to classify data into different classes. This report summarizes the performance of the model on a test set. We use of the list of class names that was created earlier to ensure that the report includes information about each and every class.

The classification report includes performance metrics such as precision, recall, F1-score, and support for each class, as well as overall accuracy and a weighted average of the metrics across all classes. With the help of TP, TN, FN, and FP, these performance metrics can be comprehended and calculated.

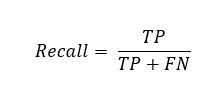
Precision

It represents the percentage which is truly positive out of all the positive observations predicted. The precision value lies between 0 and 1.



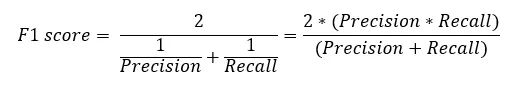
Recall

It represents the percentage that is predicted as positive out of the total positive observations.



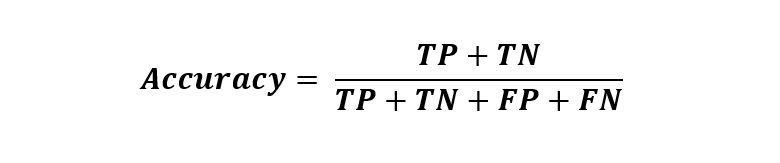
F-1 Score

It is the harmonic mean of Precision and Recall. It takes both false positive and false negatives into account, and gives the same weightage to both Recall and Precision. Therefore, it performs well on an imbalanced dataset.



Accuracy

Accuracy is defined as the number of all correct predictions divided by the total number of observations in the dataset. The best accuracy is 1, whereas the worst is 0.



## REFERENCES

1. Casasent, D. A., Sipe, M. A., Schatzki, T. F., Keagy, P. W., Lee, L. C. Neural net classification of X-ray pistachio nut data. Lebensmittel Wissenschaft und Technologie 1998; 31(2), 122–128.
2. Kouchakzadeh, A. Moisture diffusivity of five major varieties of Iranian pistachios. American Journal of Food Technology 2011. 6 (3): 253-259.
3. Omid, M., Mahmoudi, A., Omid, H. M. An intelligent system for sorting pistachio nut varieties. Expert Systems with Applications 2009; 36: 11528–11535