ISSN 0974 - 9373

Vol. 27 No. 1 (2023) **J. Int. Acad. Phys. Sci.** pp. 23-33

# Pumpkin Seeds Classification: Artificial Neural Network and Machine Learning Methods

**K. V. Prasad, Hanumesh Vaidya**
Department of Studies in Mathematics,
Vijayanagara Sri Krishnadevaraya University, Ballari, Karnataka, India
Email: prasadkv2007@gmail.com, hanumeshvaidya@gmail.com

**Kumar Swamy K* and Renuka S**
Department of Studies in Computer Science,
Vijayanagara Sri Krishnadevaraya University, Ballari, Karnataka, India
Email: kswamy1483@gmail.com, 24renusali@gmail.com

(Received January 23, 2023, Accepted March 10, 2023)

**Abstract:** Pumpkin seeds may be small, but they're packed with valuable nutrients. Eating just a small amount can provide a significant amount of healthy fats, magnesium and zinc. For this reason, pumpkin seeds are associated with various health benefits. These include improving heart health, prostate health, and protection against certain cancers. In addition, these seeds are easy to incorporate into the diet. The main objective of this research is to design and implement an artificial neural network model, and the machine learning models, i.e., LR, SVM, DT, Naive Bayes, and K-NN, were investigated for their accuracy, precision, and recall on 2500 samples obtained from Kaggle. The average accuracy of LR, SVM, DT, Naive Bayes and K-NN were 99.81%, 48.00%, 100.00%, 95.80%, 46.60% and 77.20%, respectively. Thus, it was found that the decision tree model performed better than the other models. The proposed model can be effectively used for characterization, classification and identification of pumpkin seeds.

**Keywords:** Artificial Neural Network (ANN), Logistic Regression (LR), Support Vector Machine (SVM), Decision Tree (DT), K-Nearest Neighbor's (K-NN).

## 1. Introduction

Pumpkin seeds are a popular snack product in several countries, including Turkey and Greece. They are eaten either raw or roasted (salted or unsalted) and used in cooking and baking as an ingredient in bread, cereals, salads and cakes. The consumption of pumpkin seed products has increased recently, and it has been suggested that the inclusion of this functional food in the daily diet could help improve consumer health. However, a better knowledge of the composition of these products is essential. Most pumpkins and pumpkin seeds are now grown in China and India, but Russia, Ukraine, the U.S., and Mexico are also large producers. 95% of U.S. pumpkins are grown in Illinois and are primarily for pumpkin pie. According to the Food and Agriculture Organization of the United Nations, pumpkins were grown in 117 different countries in 2017.

Pumpkin seeds are rich in antioxidants, iron, zinc, magnesium and many other nutrients such as carotenoids and vitamin E to determine the oxidizability of lipids and their inhibition by endogenous antioxidants, the authors have developed a simple fluorescence technique based on solubilisation of oils in aqueous buffer, labelling of the resulting emulsions with a suitable reporter fluorophore reflecting lipid oxidation, and continuous monitoring of the decomposition process as described by Fruhwirth et al[1]. Zenoozian et at[2], the aim of their study was to predict the effects of different drying conditions (temperature, air velocity, drying time, sample thickness) and different osmotic treatments using neural networks. In addition, pumpkin seed oil is now widely accepted not only as edible oil, but also as a dietary supplement. Pumpkin seeds and pumpkin seed oil are believed to have numerous health benefits due to their macro and micro composition. A diet rich in pumpkin seeds is associated with a lower risk of stomach, breast, lung, prostate, and colon cancer. Xanthopoulou et al[3], their study reports the antioxidants capacity and inhibitory activity of pumpkin seed extracts against DPPH free radical formation and soybean lipoxygenase, respectively. In order to provide environmental and economic benefits to farmers who grow pumpkin and pumpkin seeds and ensure their livelihood, there is an urgent need to refine various pumpkin seeds growing technologies to meet our specific requirements. Jamuna et al[4] developed a model using machine learning techniques such as Naive Bayes classifier, decision tree classifier, and multilayer perceptron to classify seed cotton yield. Their study shows that the results obtained using multilayer perceptron and decision tree classifier have the same accuracy, but the time required to build the model using multilayer perceptron is higher compared to decision tree classifier. Their research focuses on an automatic method for identifying different rice seeds using machine vision technology was investigated, and a recognition system consisting of an automatic inspection machine and an image processing unit was developed as demonstrated by OuYang et al[5]. In their research, they looked at eight different similar looking food grains, to recognize and classify them; they created an artificial neural network model as mentioned by

Dayanand[6]. Ahamed et al[7] have examined the traits of various pumpkin germplasm collections in their investigation. Selvi et al[8] conducted the study using 15 pumpkin genotypes to assess the features of growth and yield. According to Zaineddin et al[9], in a large observational study, consumption of pumpkin seeds was found to be associated with a lower risk of breast cancer in postmenopausal women. Sharma et al[10], the goal of their research was to better understand the nutritional value of the important but underutilised fruit pumpkin at various developmental stages. Pandey et al[11] studies show that using the Content Based Image Retrieval (CBIR) technique to identify seeds, e.g. wheat, rice, chickpea, etc., based on their characteristics. CBIR is a technique to identify or recognize an image based on the features present in the image.

In order to address the crop selection problem, maximise net yield rates across seasons, and ultimately achieve the greatest possible economic growth for the nation, the Crop Selection Method (CSM) was presented by Kumar at al[12]. Bulut[13] has developed a classification model for ensemble learners on imbalanced dataset to classify accurately. Shaban et al[14] review focuses on the characteristics, uses, and pathophysiologies of numerous different pumpkin species from pumpkin seed oils against illness. They have a strong conviction that their review will give chemists, biologists, and researchers a broad understanding of the functions of pumpkin seed oil extracts, which have potential biological properties. In lab settings, the Demir et al[15] calculated the tridimensional physical characteristics of the pumpkin seeds, including volume, surface, projected area, geometric mean diameter, phericity, and linear dimensions. Utilizing neural networks, these parameters were predicted. Ramjan et al[16], the goal of the study was to assess the average performance of different pumpkin genotypes in terms of growth, yield, and quality parameters. Seymen et al[17] concludes that pumpkins are divided into many types and further explains that one of these species is known as Urgup Sivrisi, the other type of pumpkin seeds is Cercevelik. It is a special species grown in Turkey, in Nevsehir, Karacaoren, and known in Turkey as Topak. With the increasing importance of the use of geographic information systems (GIS) and technological developments, Artificial Neural Network and Machine learning methods have been used in agriculture and other areas of the world. Many researchers have used artificial neural networks and machine learning models to evaluate the characterization, classification, and identification of different varieties of seeds.

The study was conducted on the two main types of pumpkin seeds, Urgup Sivrisi and Cercevelik, which are usually grown in the Urgup and Karacaoren regions of Turkey. The dataset is from Kaggle and includes 2500 samples containing 12 parameters, namely Area, Perimeter, Major axis length, Minor axis length, Eccentricity, Convex area, Extent, Equiv. diameter, Solidity, Aspect ratio, Roundness, and Compactness. The artificial neural network and machine learning model were developed to compare the accuracy, precision and recall. Finally, a conclusion is drawn from all the machine learning models, and ANN shows which model performs best. This paper is organised as follows: In Section 2, an insight into the methodology is given, followed by a detailed explanation of ANN and the machine learning models. In section 3, dataset and proposed model. Section 4, experimental results. Section 5, Conclusion.

## 2. Methodology

### 2.1 Artificial Neural Network

Based on the information about the function of the brain, a new technology of artificial neural networks has been developed. A neural network is a network of interconnected neurons inspired by the studies of the biological nervous system. Artificial neural networks are biologically inspired, meaning that they are composed of elements that function in a way that corresponds to the most basic functions of the biological neuron. These elements are then organised in a manner related to the anatomy of the brain. Despite this superficial similarity, ANNs exhibit a surprising number of brain characteristics. For example, they learn from experience, generalize from previous examples to new ones, and abstract essential features from inputs that contain irrelevant data. Some areas where neural networks are used are classification, signal processing, speech recognition, medicine, intelligent control, and more. There are generally two types of neural networks, with the most rudimentary and simplest ANN paradigm being the feed-forward neural network (FFNN), a multilayer composite of perceptrons in which the output layer does not form a loop for feedback connections or recurrent networks, but rather in a unidirectional forward flow as addressed by Saikia et al[18].

### 2.2 Machine Learning

A computer program that learns from experience is called a machine learning program or simply a learning program. Such a program is also called as learner. Machine learning is about using the right features to create the right models to perform the right tasks. Machine learning is a scientific technique in which computers learn how to solve a problem without being explicitly programmed. In machine learning, binary classification is a supervised learning algorithm that classifies new observations into one of two classes. The application of supervised machine learning methods to solve complicated problems has become increasingly important in many industries, including the agriculture. Most of these complex problems hindered critical decision making and progress in the agriculture, so researchers gradually moved away from using empirical correlations and linear regression models as mentioned by Otchere et al[19].

## 3. Dataset and Proposed model

The dataset is from Kaggle and consists of two types of pumpkin seeds such as Urgup Sivrisi and Cercevelik grown in Turkey. The pumpkin seed database consists of 12 input features and 2500 samples. The characteristics of the dataset are summarised in Table 1. Area (A), Perimeter (p), Major axis length (Maj.AL), Minor axis length (Min.AL), Eccentricity (e), Convex area (CA), Extent (E), Equiv. diameter (ED), Solidity (s), Aspect ratio (AR), Roundness (r), and Compactness (C). The result is a class or dependent variable, while the other 12 characteristics are properties or independent variables. The class property can take two values, 0 for Cercevelik and 1 for Urgup Sivrisi.

**Table I.** Sample Dataset of Pumpkin seeds

| Area | Perimeter | Major_Axis _Length | Minor_Axis _Length | Convex _Area | Equiv _Diameter | Eccentricity | Solidity | Extent | Roundness | Aspect _Ration | Compact ness | Class |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 56276 | 888.242 | 326.1485 | 220.2388 | 56831 | 267.6805 | 0.7376 | 0.9902 | 0.7453 | 0.8963 | 1.4809 | 0.8207 | Cercevelik |
| 76631 | 1068.146 | 417.1932 | 234.2289 | 77280 | 312.3614 | 0.8275 | 0.9916 | 0.7151 | 0.844 | 1.7811 | 0.7487 | Cercevelik |
| 71623 | 1082.987 | 435.8328 | 211.0457 | 72663 | 301.9822 | 0.8749 | 0.9857 | 0.74 | 0.7674 | 2.0651 | 0.6929 | Cercevelik |
| 66458 | 992.051 | 381.5638 | 222.5322 | 67118 | 290.8899 | 0.8123 | 0.9902 | 0.7396 | 0.8486 | 1.7146 | 0.7624 | Cercevelik |
| 66107 | 998.146 | 383.8883 | 220.4545 | 67117 | 290.1207 | 0.8187 | 0.985 | 0.6752 | 0.8338 | 1.7413 | 0.7557 | Cercevelik |
| 73191 | 1041.46 | 405.8132 | 231.4261 | 73969 | 305.2698 | 0.8215 | 0.9895 | 0.7165 | 0.848 | 1.7535 | 0.7522 | Cercevelik |
| 76910 | 1146.88 | 483.5875 | 203.5562 | 77745 | 312.9295 | 0.9071 | 0.9893 | 0.7629 | 0.7348 | 2.3757 | 0.6471 | Urgup Sivrisi |
| 118751 | 1468.224 | 629.723 | 240.9782 | 120036 | 388.8425 | 0.9239 | 0.9893 | 0.744 | 0.6922 | 2.6132 | 0.6175 | Urgup Sivrisi |
| 86565 | 1215.697 | 508.1608 | 218.3103 | 87292 | 331.9909 | 0.903 | 0.9917 | 0.7189 | 0.736 | 2.3277 | 0.6533 | Urgup Sivrisi |
| 72106 | 1131.138 | 484.9191 | 193.0989 | 72800 | 302.9987 | 0.9173 | 0.9905 | 0.6187 | 0.7082 | 2.5112 | 0.6248 | Urgup Sivrisi |
| 86686 | 1208.535 | 509.7207 | 216.9252 | 87352 | 332.2229 | 0.9049 | 0.9924 | 0.7411 | 0.7458 | 2.3498 | 0.6518 | Urgup Sivrisi |
| 91703 | 1279.029 | 549.6899 | 213.0754 | 92721 | 341.7015 | 0.9218 | 0.989 | 0.7226 | 0.7044 | 2.5798 | 0.6216 | Urgup Sivrisi |

From the literature review, decision tree approaches performed well compared to other approaches such as Artificial Neural Networks, Support Vector Machine, Naive Bayes, Logistic Regression, and K-Nearest Neighbour for pumpkin seed classification. The decision tree model provided more effective and acceptable results. In the proposed technique, artificial neural networks (ANNs) and machine learning models were used to achieve better accuracy, precision, and recall in pumpkin seed classification. In the modeling phase, the classification of pumpkin seeds was performed using four machine learning methods, and a model ANN was also created. The architecture of the proposed model is shown in Figure 1. The performances of these models used in the last step were evaluated.

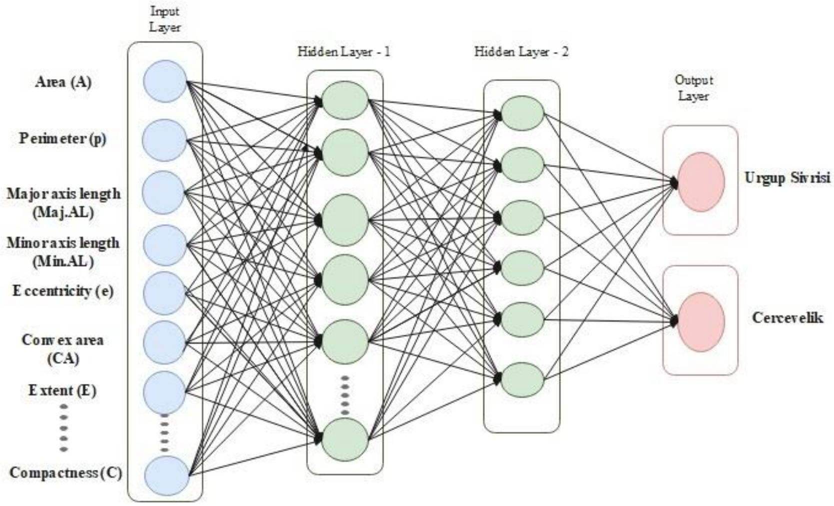


**Figure 1.** Architecture of ANN Model

## 4. Experimental Results

Experiments were performed with pumpkin seed data sets. Several machine learning methods are used to analyze the performance of pumpkin seeds: Logistic Regression, Support Vector Machine, Decision Tree, Naive Bayes, K-Nearest Neighbor and the method ANN. Therefore, researchers introduce different machine learning approaches and ANN to classify pumpkin seeds.

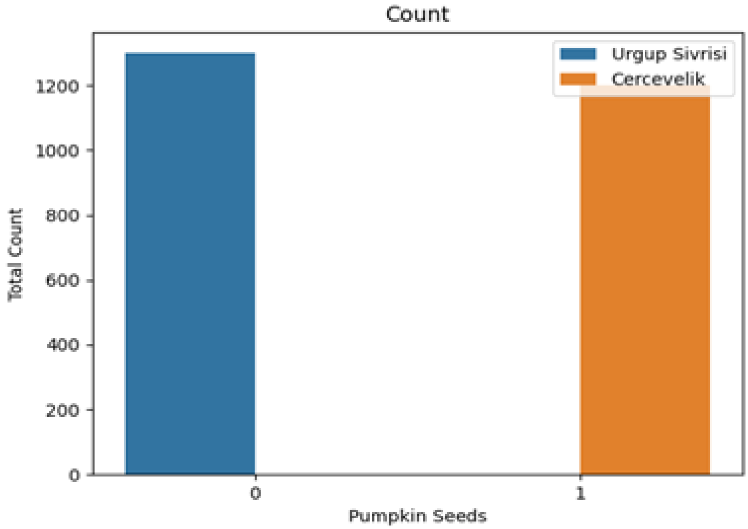


**Figure 2.** Bar plot visualising the number of classes Urgup Sivrisi(0) and Cercevelik(1) in the pumpkin seed data set

The aim of this paper is to develop a computational model for accurate classification of pumpkin seeds. To evaluate the capabilities of the ANN model and machine learning models, the dataset should ideally be divided into two subsets for the training phase and the testing phase. The pumpkin seed dataset consists of 2500 instance tuples with 12 features. From the Figure 2. It was found that the pumpkin seed record contains 1200 Cercevelik and 1300 Urgup Sivrisi. The pumpkin seed dataset is suitable for binary classification models because it contains only binary values such as 0 and 1, *i.e.* 0 for Cercevelik and 1 for Urgup Sivrisi. The sample data of the pumpkin seed data set are shown in Table I. Where Area - Specifies the number of pixels within the boundaries of a pumpkin seed, Perimeter - Specifies the perimeter in pixels of a pumpkin seed, Major axis length - Specifies the perimeter in pixels of a pumpkin seed, Minor axis length - Specifies the minor axis distance of a pumpkin seed, Eccentricity - Specifies the eccentricity of a pumpkin seed, Convex area - Specifies the number of pixels of the smallest convex shell in the area of the pumpkin seed, Extent - Specifies the ratio of the area of the pumpkin seed to the pixels of the bounding box, Equivalent diameter - It was formed by multiplying the area of the pumpkin seed by four and dividing by the number pi and taking the square root, Compactness - It put the area of the pumpkin seed in proportion to the area of the circle with the same circumference, Firmness - He considered the convex and convex nature of the pumpkin seeds, Roundness - He measured the ovality of the pumpkin seeds without considering the distortion of the edges, and Aspect Ratio - He gave the aspect ratio of the pumpkin seeds method used by Koklu et al[20]. The statistical distribution of the pumpkin seed data set is shown in Table II.

**Table II.** Statistical Distribution of Pumpkin Dataset

| No | Features | Min | Mean | Max | Std. Dev |
|---|---|---|---|---|---|
| 1 | Area | 47,939.0 | 80,658.220 | 136,574.0 | 13,664.510 |
| 2 | Perimeter | 868.485 | 1,130.279 | 1,559.45 | 109.256 |
| 3 | Major axis length | 320.844 | 456.601 | 661.911 | 56.235 |
| 4 | Minor axis length | 152.171 | 225.794 | 305.818 | 23.297 |
| 5 | Eccentricity | 0.492 | 0.860 | 0.948 | 0.045 |
| 6 | Convex area) | 48,366.0 | 81,508.084 | 138,384.0 | 13,764.092 |
| 7 | Extent | 0.467 | 0.693 | 0.829 | 0.060 |
| 8 | Equiv. diameter | 247.058 | 319.334 | 417.002 | 26.891 |
| 9 | Solidity | 0.918 | 0.989 | 0.994 | 0.003 |
| 10 | Aspect ratio | 1.148 | 2.041 | 3.144 | 0.315 |
| 11 | Roundness | 0.554 | 0.791 | 0.939 | 0.055 |
| 12 | Compactness | 0.560 | 0.704 | 0.904 | 0.053 |

The present work is for accurate classification of pumpkin seeds using ANN model and machine learning methods such as LR, SVM, DT, Naive Bayes and K-NN. The proposed model is measured by evaluation metrics such as accuracy, precision, and recognition, as shown in (4.1), (4.2) and (4.3).

(4.1) $$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

(4.2) $$Precision = \frac{TP}{TP+FP}$$

(4.3) $$Recall = \frac{TP}{TP+FN}$$

Table III summarizes the accuracy performance of the models. As can be seen in Table III, the accuracy rates of the models created were over 85 percent. The accuracy value of the DT model was 100.00%, which proved this model to be the most accurate.

**Table III.** Results of the performance measurement of the models

| Models | Accuracy | Precision | Recall |
|---|---|---|---|
| Logistic Regression | 0.998 | 1.000000 | 0.995851 |
| Support Vector Machines | 0.522 | 0.004167 | 1.000000 |
| Decision Tree | 1.000 | 1.000000 | 1.000000 |
| Naive Bayes | 0.520 | 0.000000 | 0.000000 |
| Artificial Neural Network | 0.958 | 0.950000 | 0.962025 |
| K-Nearest Neighbor | 0.772 | 0.729167 | 0.781250 |

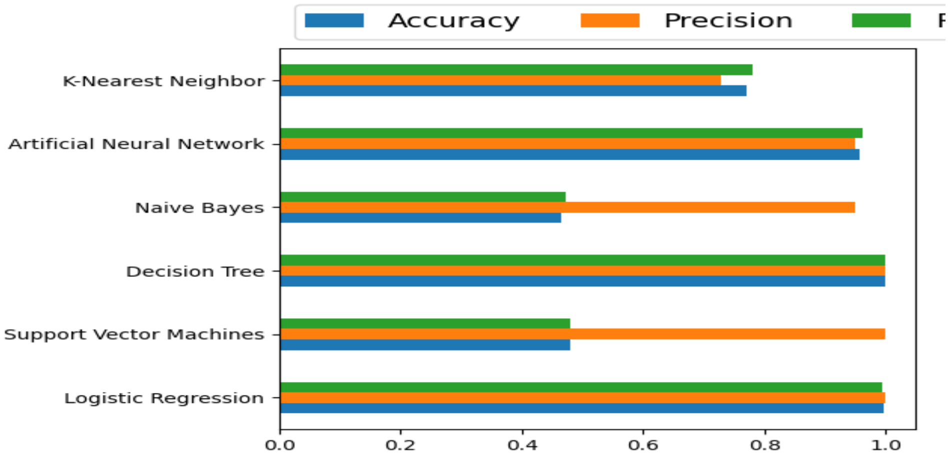


**Figure 3.** Performance chart of the models for two types of pumpkin seeds.

In terms of precision, the success rate of the SVM model was 100.00%, while that of the ANN model was 96.20%. According to this evaluation, the

DT model provided more accurate results than the ANN model. However, in terms of accuracy, precision and recall evaluation, the results showed that the DT model was more efficient than the other models. All the performances of the models can be understood from the diagram in Figure 3. According to this diagram, the models achieved a success rate of over 85 percent for all performance values.

## 5. Conclusion

The classification models using ANN and classical machine learning methods to classify the two types of pumpkin seeds like Cercevelik and Urgup Sivrisi in the pumpkin seed dataset. In this paper, a DT model of a machine learning algorithm was used to classify the pumpkin seeds. This approach effectively improves accuracy, precision, and recall. By comparison, the DT model provides a method for classifying the pumpkin seeds.

In the future, we plan to develop a deep learning model for pumpkin seed images, which are challenging to identify and classify. We also plan to improve the performance of deep learning classification and recognition models for other agricultural applications. Finally, we expect that this morphological dataset will be useful to study how to classify the two types of pumpkin seeds.

## References

1. Fruhwirth, Gilbert Otto, Thomas Wenzl, Rosemarie El- Toukhy, Franz Siegfried Wagner, and Albin Hermetter; Fluorescence screening of antioxidant capacity in pumpkin seed oils and other natural oils, *European Journal of Lipid Science and Technology* 105(6) (2003), 266-274.
2. Zenoozian, M. Shafafi, S. Devahastin, M. A. Razavi, F. Shahidi, and H. R. Poreza; Use of artificial neural network and image analysis to predict physical properties of osmotically dehydrated pumpkin, *Drying Technology* 26(1) (2007), 132-144.
3. Xanthopoulou, Marianna N., Tzortzis Nomikos, Elizabeth Fragopoulou, and Smaragdi Antonopoulou; Antioxidant and lipoxygenase inhibitory activities of pumpkin seed extracts, *Food Research International* 42(5-6) (2009), 641-646.
4. Jamuna, K. S., S. Karpagavalli, M. S. Vijaya, P. Revathi, S. Gokilavani, and E. Madhiya, Classification of seed cotton yield based on the growth stages of cotton crop using machine learning techniques, In *2010 International Conference on Advances in Computer Engineering*, (2010), 312-315.

5. OuYang, Ai-Guo, Rong-jie Gao, Xu-dong Sun, Yuan-yuan Pan, and Xiao-ling Dong; An automatic method for identifying different variety of rice seeds using machine vision technology, In *2010 Sixth International Conference on Natural Computation*, 1 (2010), 84-88.
6. S. A. V. A. K. A. R. Dayanand; Recognition and Classification of similar looking food grain images using Artificial Neural Networks, *Journal of Applied Computer Science & Mathematics* 6(2) (2011), 61-65.
7. K. U. Ahamed, B. Akhter, M. R. Islam, N. Ara, and M. R. Humauan; An assessment of morphology and yield characteristics of pumpkin (Cucurbita moschata) genotypes in northern Bangladesh, *Tropical Agricultural Research and Extension* 14(1) (2012).
8. N. A. Selvi, P. Jansirani, L. Pugalendhi, and A. Nirmalakumari; Per se performance of genotypes and correlation analysis in Pumpkin (Cucurbita moschata Duch. ex Poir), *Electronic journal of plant breeding* 3(4) (2012), 987-994.
9. Zaineddin, Aida Karina, Katharina Buck, Alina Vrieling, Judith Heinz, Dieter Flesch-Janys, Jakob Linseisen, and Jenny Chang-Claude; The association between dietary lignans, phytoestrogen-rich foods, and fiber intake and postmenopausal breast cancer risk: a German case-control study, *Nutrition and cancer* 64(5) (2012), 652-665.
10. Sharma, Sonu and T. V. Ramana Rao, Nutritional quality characteristics of pumpkin fruit as revealed by its biochemical analysis, *International Food Research Journal* 20(5) (2013).
11. Naveen Pandey, Satyanarayan Krishna, and Shanu Sharma, Automatic seed classification by shape and color features using machine vision technology, *International Journal of Computer Applications Technology and Research* 2(2) (2013), 208-213.
12. Rakesh Kumar, M. P. Singh, Prabhat Kumar, and J. P. Singh, Crop Selection Method to maximize crop yield rate using machine learning technique, In *2015 international conference on smart technologies and management for computing, communication, controls, energy and materials (ICSTM)*, (2015) 138-145.
13. F. Bulut; Performance Analysis of Ensemble Methods on Imbalanced Datasets, *Bilişim Teknolojileri Dergisi* 9(2) (2016), 153-159.
14. Shaban, Ahmed, and Ravi P. Sahu, Pumpkin seed oil: an alternative medicine, *International journal of pharmacognosy and phytochemical research* 9(2) (2017).
15. Demir, Bunyamin, E. S. K. I. Ikbal, and Sezai ERCISLI, Prediction of physical parameters of pumpkin seeds using neural network, *Notulae Botanicae Horti Agrobotanici Cluj-Napoca* 45(1) (2017), 22-27.

16. Ramjan, Md, A. K. Pandey, Megha Raghavan, Hau Ngaih Lian, and Mohd Talha Ansari, Variations in pumpkin (Cucurbita moschata Duch. ex Poir.) landraces for earliness, yield and quality attributes in North East India, *Int. J. Curr. Microbiol. App. Sci* 7(7) (2018), 741-748.
17. Seymen, Musa, Duran Yavuz, Atilla Dursun, Ertan Sait Kurtar, and Önder Türkmen; Identification of drought-tolerant pumpkin (Cucurbita pepo L.) genotypes associated with certain fruit characteristics, seed yield, and quality, *Agricultural Water Management* 221 (2019), 150-159.
18. Saikia, Pallabi, Rashmi Dutta Baruah, Sanjay Kumar Singh, and Pradip Kumar Chaudhuri; Artificial Neural Networks in the domain of reservoir characterization: A review from shallow to deep models, *Computers & Geosciences* 135 (2020), 104357.
19. Otchere, Daniel Asante, Tarek Omar Arbi Ganat, Raoof Gholami, and Syahrir Ridha; Application of supervised machine learning paradigms in the prediction of petroleum reservoir properties: Comparative analysis of ANN and SVM models, *Journal of Petroleum Science and Engineering* 200 (2021), 108182.
20. Koklu, Murat, Seyma Sarigil, and Osman Ozbek; The use of machine learning methods in classification of pumpkin seeds (Cucurbita pepo L.). *Genetic Resources and Crop Evolution* 68(7) (2021), 2713-2726.