



Dry Beans Research Paper

CPSC 4310

SeattleU

The sole purpose for this research paper is to classify and define all given features corresponding to the dataset which in this case is (Dry Beans). The dataset provided 17 different features, 16 of those being of numerical type and 1 being categorical. We can start by listing all 17 features in their correct type and defining each feature.

For type numerical we have the following:

- I. Area (A) – The area of a bean zone and the number of pixels within the boundaries. From the dataset provided we can see throughout time the number starts increasing.
- II. Perimeter (P) – Bean circumference is defined as the length of its border.
- III. Major axis length (L) – The distance between the ends of the longest line that can be drawn from a bean.
- IV. Minor axis length (l) – The longest line that can be drawn from the bean while standing perpendicular to the main axis.
- V. Aspect ratio (K) – Defines the relationship between (L) and (l).
- VI. Eccentricity (Ec) – Eccentricity of the ellipse having the same moments as the region.
- VII. Convex area (C) – Number of pixels in the smallest convex polygon that can contain the area of a bean seed.
- VIII. Equivalent diameter (Ed) – The diameter if a circle having the same area as a bean seed area
- IX. Extent (Ex) – The ratio of the pixels in the bounding box to the bean area.
- X. Solidity (S) – Also known as convexity. The ratio of the pixels in the convex shell to those found beans.
- XI. Roundness (R) – Calculated with the following formula $(4\pi A)/(P^2)$
- XII. Compactness (CO) – Measures the roundness of an object: Ed/L
- XIII. ShapFactor1 (SF1)
- XIV. ShapFactor2 (SF2)
- XV. ShapFactor3 (SF3)
- XVI. ShapFactor4 (SF4)

For categorical we have:

- XVII. Class (Seker, Barbunya, Bombay, Cali, Dermosan, Horoz and Sira)

For our attribute types we have categorical, integer and real numbers. Our dataset does not contain missing values. The number of instances we have our 13611 and 17 attributes.

Some relevant information:

Seven different types of dry beans were used in this research, considering the features such as form, shape, type, and structure by the market situation. A computer vision system was developed to distinguish seven different registered varieties of dry beans with similar features to obtain uniform seed classification. For the classification model, images of 13,611 grains of 7 different registered dry beans were taken with a high-resolution camera. Bean images obtained by computer vision system were subjected to segmentation and feature extraction stages, and a total of 16 features; 12 dimensions and 4 shape forms, were obtained from the grains.

Figure 1a shows each feature calculation for the corresponding index. Five number summaries.

Index	Area	Perimeter	MajorAxisLength	MinorAxisLength	AspectRatio	Eccentricity	ConvexArea	EquivDiameter	Extent	Solidity	roundness	Compactness	ShapeFactor1	ShapeFactor2	ShapeFactor3	ShapeFactor4
count	13611	13611	13611	13611	13611	13611	13611	13611	13611	13611	13611	13611	13611	13611	13611	13611
mean	53048.3	855.283	320.142	202.271	1.58324	0.758095	53768.2	253.064	0.749733	0.987143	0.873282	0.799864	0.00656361	0.00171595	0.64359	0.995063
std	29324.1	214.29	85.6942	44.9701	0.246678	0.0920018	29774.9	59.1771	0.0490864	0.00466038	0.0595199	0.0617135	0.001128	0.000595875	0.0909962	0.00436646
min	20420	524.736	183.601	122.513	1.02487	0.218951	20684	161.244	0.555315	0.919246	0.408618	0.640577	0.00277801	0.000564169	0.410339	0.947687
25%	36328	703.523	253.304	175.848	1.43231	0.715928	36714.5	215.068	0.710634	0.98567	0.832096	0.762469	0.00509992	0.00115352	0.581359	0.993703
50%	44652	794.941	296.883	192.432	1.55112	0.764441	45178	238.438	0.759859	0.988283	0.883157	0.801277	0.00664517	0.00169353	0.642044	0.996386
75%	61332	977.213	376.495	217.832	1.70711	0.810466	62294	279.446	0.786851	0.990013	0.916869	0.83427	0.00727142	0.00217027	0.696006	0.997883
max	254616	1985.37	738.86	460.198	2.43031	0.911423	263261	569.374	0.866195	0.994677	0.990685	0.987303	0.0104512	0.00366497	0.974767	0.999733

Figure 1b shows all features plotted on a histogram. From the given histograms we have a mix of skewed and normally distribution.

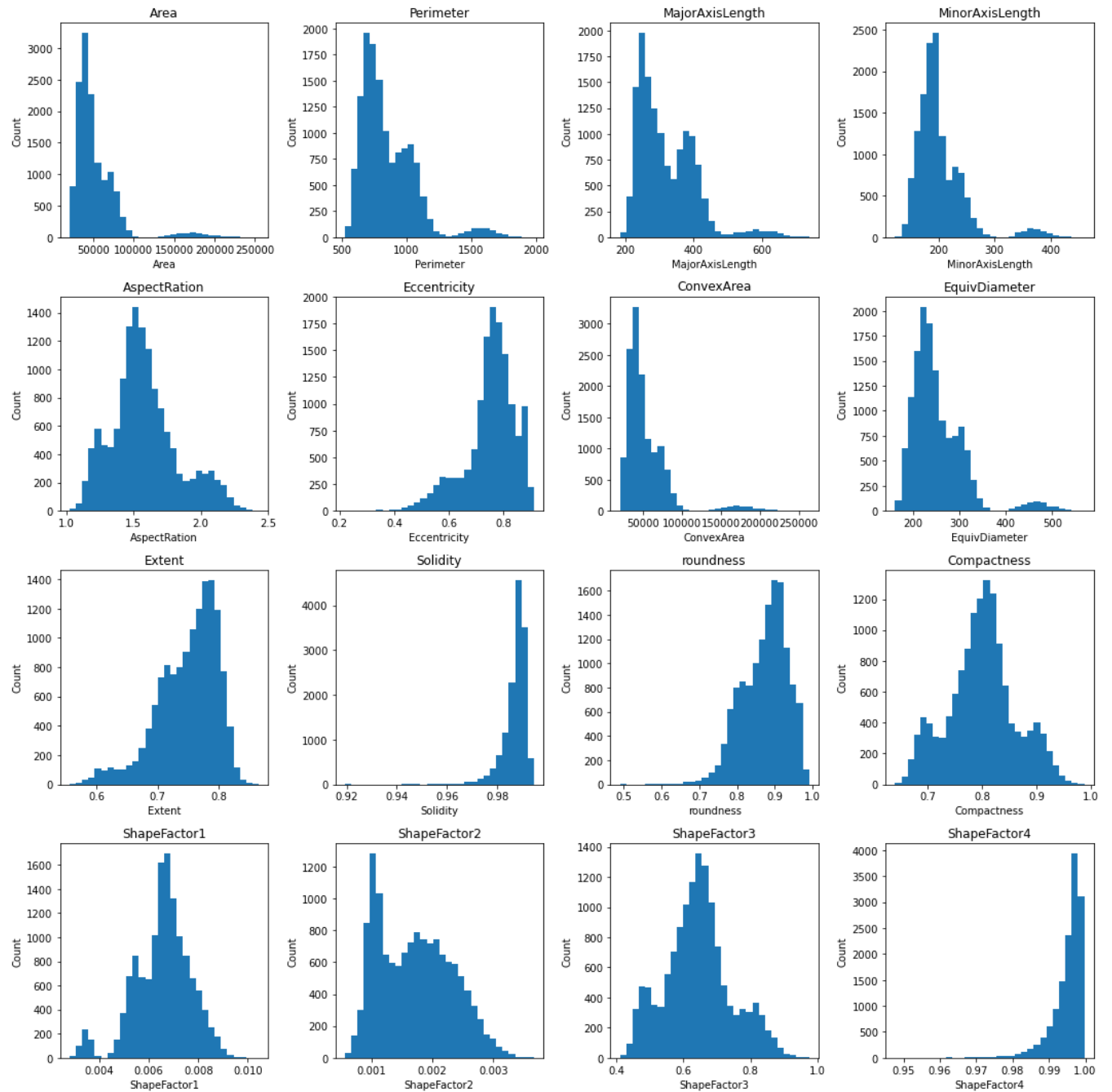


Figure 1c shows the categorical type plotted on a bar chart. From the chart we can see that majority of the beans belong to Dermason and Sira.

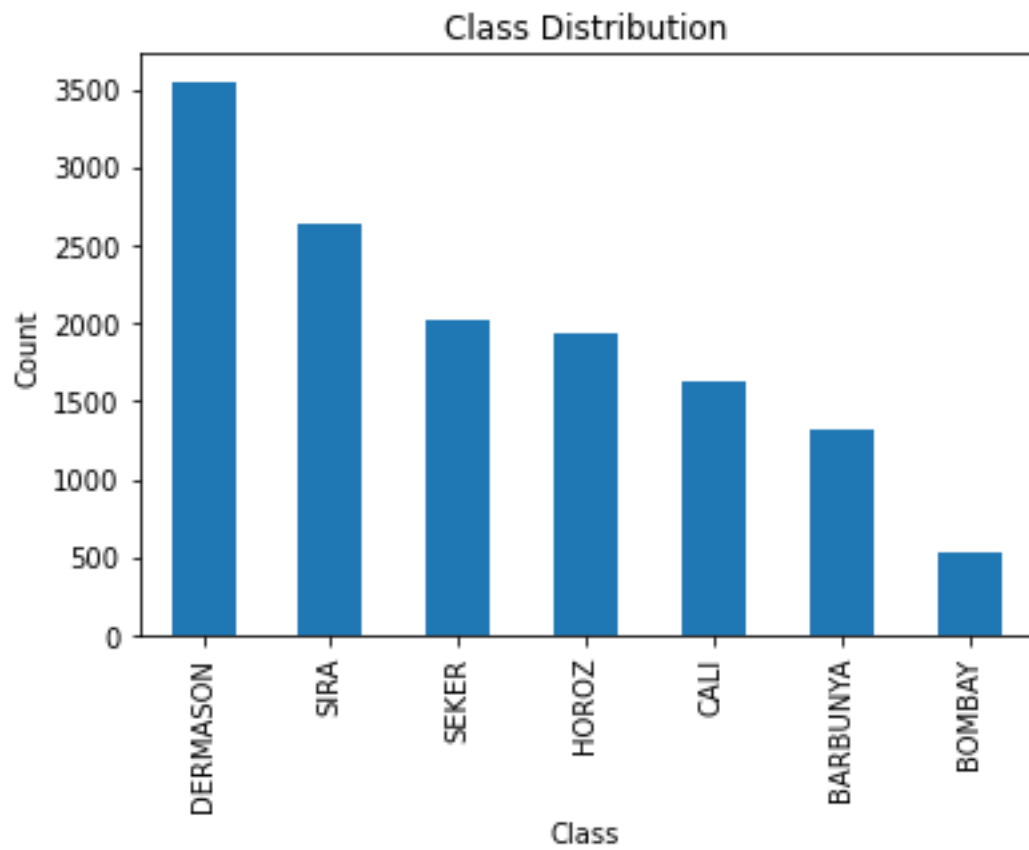
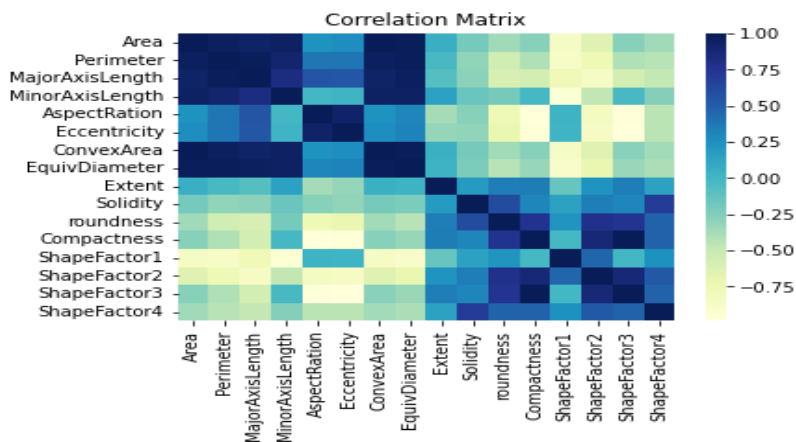


Figure 1d shows a heatmap of the correlation matrix.



Citation Requests / Acknowledgements:

KOKLU, M. and OZKAN, I.A., (2020), “Multiclass Classification of Dry Beans Using Computer Vision and Machine Learning Techniques.” Computers and Electronics in Agriculture, 174, 105507.

DOI: <https://doi.org/10.1016/j.compag.2020.105507>

UCI Machine Learning Repository. (n.d.). Dry Bean Dataset. Retrieved May 7, 2023, from <https://archive.ics.uci.edu/ml/datasets/Dry+Bean+Dataset>.