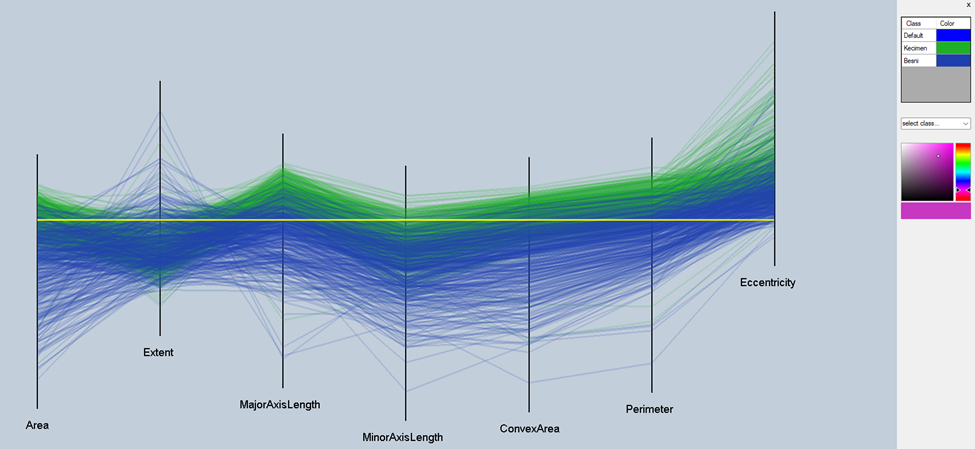
Chart, scatter chart

Description automatically generated

I chose the Raisin Dataset. This dataset consists of two classes of measurements from images of Kecimen and Besni raisin varieties. There are 900 total measurements of raisins with an equal distribution between the two varieties of raisins. The measures, or dimensions, of the data include a variety of morphological measurements. These measurements include area, major axis length, minor axis length, eccentricity, convex area, extent, perimeter, and then the classification into the two varieties. This makes the dataset 7-dimensional. The figure shown above is the default GLC-L from DV whereas the figure below is the parallel line coordinates from VisCanvas.



In comparison to the parallel line coordinates, GLC-L also suffers with difficulty from data occlusion due to density of data points. One thing additionally that this software does is that it allows for the user to perform classification from the data that is projected to the axis. This allows the user to gauge what class the data might belong to. Due to data that sometime overlaps between the classes in this dataset it is difficult to adequately classify this dataset much beyond the 86.0% that I received after fiddling around with projects and thresholds for quite some time. Overall this methodology is useful and quite distinct from parallel lines yet also deals with basic geometric features such as lines and angles with the caveat being connects at axes versus at some user-defined point in space.

I like this data visualization for more classification purposes yet since the thresholds and confusion matrix serve to give us a metric of classification and accuracy. However, where DV suffers is that I do not feel that I can get as much information “at-a-glance” as with VisCanvas.