Classification of Immature and Mature Coffee Beans Using HSV Features and Machine Learning Algorithms

Larry B. de Guzman

College of Computing Science, Information and Communication Technology, Isabela State University,
Echague, Isabela , Philippines, deguzman.larry.b@gmail.com

ABSTRACT

This paper separates immature from mature coffee beans using the hue, saturation and brightness values (HSV). It was proven in another paper that red, green, and blue (RGB) values differs in mature and immature coffee beans so the author thought of using HSV values to separate the two bean groups. The HSV values were extracted using image processing and machine learning algorithms were used in classification. The MATLAB Classification Learner App containing 23 different machine learning algorithms was used in discriminating between mature and immature coffee beans. The medium Gaussian support vector machine achieved the highest classification accuracy of 94 % at 0.14624 seconds.

Key words : Immature coffee beans; Average Hue, Average Saturation, Average Values; Medium Gaussian SVM

I. INTRODUCTION

Coffee is a beverage obtained from cherry, the fruit of coffee plant. The coffee plant refers to several species of the genus coffee of the Rubicaecea family, which is actually a tropical evergreen shrub that has the potential to grow to a height of 100 feet [1]. Due to its idiosyncratic smell and taste, the coffee plant has been represented as an essential part of special moments. Coffee , is known to have approximately 70 types is a well received, in demand drink and is a very significant commodity worldwide. Millions of small-scale farmers in developing countries are dependent in coffee farming as their major source of income. All over the world , there are 2.5 billion cups of coffee being consumed in a daily basis. While farming is done in developing countries, it is the industrialized countries that does most of the consumption. Green coffee beans is one of the most traded agricultural product in the world second to oil. Coffee sustained the economies of developing countries as it is considered as a major source of their foreign exchange [2]–[4]. Coffee cherries should only be harvested when they are fully red in the branches. However, coffee cherries do not ripen at the same time. Some cherries are fully red while other cherries are yellow and others remain green. Farmers should harvest the red cherries and wait for yellow and green beans to mature and become fully red. Because red, yellow and green coffee cherries when harvested and dried are indistinguishable from one another. Farmers opted to save money and harvest the cherries whether they are red, yellow or green[5]. While mature and immature beans are almost the same in the naked eye they are totally different in taste. The presence of immature coffee lowers the quality and also lowers the price of coffee beans[6, 7].

This study builds on the previous research done in immature and mature coffee beans [5] and answers the research question , "Which color features , RGB or HSV values is best in discriminating mature and immature coffee beans?" To compare which color feature is the best , this study also extracted the HSV using image processing and then subjected the extracted features for classification using the machine learning algorithms.

HSV is another format of how an RGB colors can be represented. Hue is the property of light that discriminates one color to other colors, for example the yellow color from brown or pink. Saturation is the level of white of a light source in an image. Values or brightness values is how intense the brightness of an image[8]. Image processing is a technique to carry out a particular procedures on an image so that a more refine, upgraded image would result or to extract useful features from an image[9-24].

The MATLAB Classification Learner App contains 23 machine learning classifiers. The extracted data thru image processing or any other type of data composed of numerical values are used as input. The App trains the data and validates the classification in terms of accuracy and training speed[24–27].

2. METHODOLOGY

2.1. Coffee Samples
The coffee beans used in this study is from the harvest season of coffee in the Philippines in the year 2019. These coffee beans were stored for a year with husks still on them and this March 2020, the husks were remove and used as samples in this study. There are two groups namely the mature coffee beans and the immature coffee beans from the Robusta specie. A total of 100 mature coffee beans and 100 immature coffee beans were used.

2.2 Feature Extraction

Upon visual inspection even an untrained eyes can see that there are differences in color between mature and immature coffee beans. The problem is when the two groups were mixed. Separating one from the other becomes a problem. It was confirmed in the previous study that mature and immature coffee beans differ in terms of RGB color values[5]. This study is further exploiting more deeply if the HSV color features can also differentiate mature and immature coffee beans because HSV is based in RGB. The HSV values were derived from the RGB values as related by the formula shown in Figure 1.

$$H = arc\ cos \left\{ \frac{\left[\frac{(R-G)+(R-B)}{2} \right]}{(R-G) + \sqrt{(R-B)(G-B)}} \right\}$$

$$S = 1 - \frac{3}{(R+G+B)} [min(R,G,B)]$$

$$V = \frac{1}{3} (R+G+B)$$

Figure 1 : Relation of HSV to RGB

It can be seen in Figure 1 that HSV values were derived from RGB values.

The set-up for taking of images was made an exact duplicate of [5]. The coffee beans were pictured using an A4tech PK-835 G webcam at the height of 13.5 cm with 2 LED strips at the side of the camera that are 2 inches apart. The HSV features were extracted using the Python program.

3. RESULTS AND DISCUSSION

Table 1 shows the   values of HSV color , extracted through image processing.

Table 1 shows that values of hues, saturation and brightness for mature and immature coffee beans are overlapping to one another. By using a classifier the author discriminated which one is mature and which one is an immature coffee bean.

 The classifier used are the 23 machine learning algorithms of MATLAB Classification Learner App. The 100 beans were used as inputs at once and the settings was chosen to be 5 folds. This means that 80 percent will be used as training samples. Five folds is the total number of samples per class divided 5, that is 100

divided 5 equals 20. A total of 80 samples were used as training samples and 20 samples were used as test samples. The result of classification is shown in Table 2.

Table 1: Extracted HSV values for Mature and Immature Coffee beans

Coffee Bean Type	Average Hue	Average Saturation	Average Value (Brightness)
Mature Coffee Beans	127 to 155	45 to 71	94 to 173
Immature Coffee Beans	100 to 147	41 to 71	110 to 177

Table 2: Classification Results

Classifier Type	Accuracy (%)	Training Time (Seconds)
Fine Tree	89	9.725
Medium Tree	89	0.16565
Coarse Tree	93	0.15127
Linear Discriminant	91.5	1.9528
Quadratic Discriminant	92	0.49839
Logistic Regression	92	7.0282
Linear Support Vector Machine (SVM)	93.5	2.2795
Quadratic SVM	94	0.20495
Cubic SVM	93	0.38023
Fine Gaussian SVM	91.5	0.17618
Medium Gaussian SVM	94	0.14624
Coarse Gaussian SVM	93	0.13773
Fine KNN	83	1.2831
Medium KNN	93.5	0.15308
Coarse KNN	87	0.14464
Cosine KNN	92	0.25553
Cubic KNN	93	0.17559
Weighted KNN	88.5	0.19314
Boosted Trees	87	3.6323
Bagged Trees	89	2.3021
Subspace Discriminant	91	1.7819
Subspace KNN	86.5	1.5994
RUS Boosted Trees	83.5	2.1379

Figure 2 shows the scatter plot of the Medium Gaussian SVM for mature and immature coffee beans. The mature coffee beans were represented by the orange colors and the immature coffee beans as the blue colors . It can be seen in Figure 2 that HSV values of the two groups are separable in the viewpoint of the Medium Gaussian SVM.
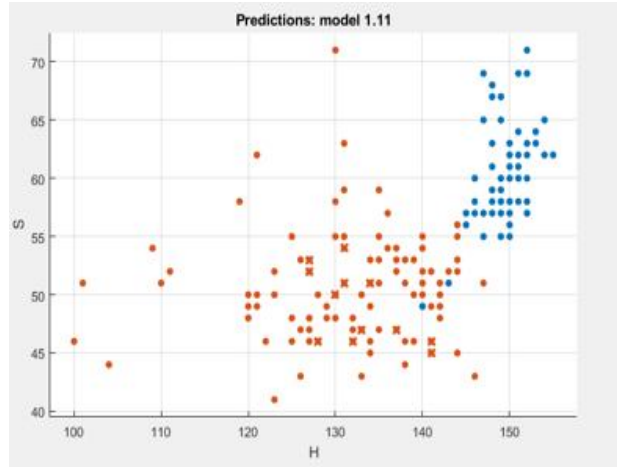

Figure 2 : Scatterplot of Medium Gaussian SVM

There are 2 classifiers that yielded the highest accuracy of 94 % namely the Quadratic SVM and the Medium Gaussian SVM. Among these 2 , the Medium Gaussian SVM has the fastest training time of 0.14624 seconds. Among 40 samples of mixed mature and immature coffee beans it will only take 0.14624 seconds to differentiate one class from the other. Even an expert human classifier will not be able to match that accuracy and speed.
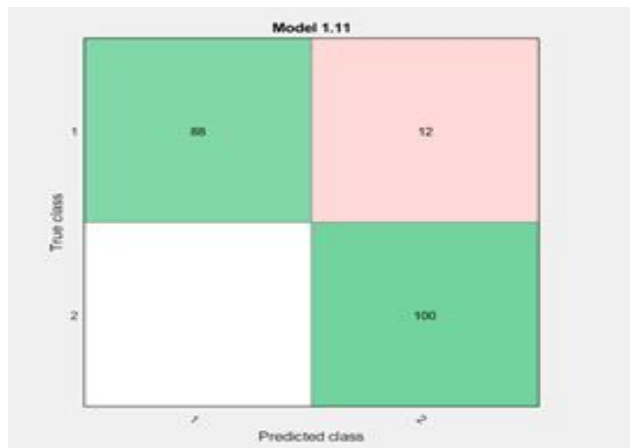

Figure 3: Confusion Matrix of Medium Gaussian SVM

It can be seen in Figure 3 that for the Medium Gaussian SVM has 12 wrong classifications for mature coffee beans and 100 percent classification for immature coffee beans. This means that the HSV values for immature coffee beans can be identified by Medium Gaussian SVM while the same algorithm has an error when it comes to identifying the HSV features of a mature coffee beans.

Figure 4 shows the ROC or the receiver operating characteristics of the Medium Gaussian SVM.
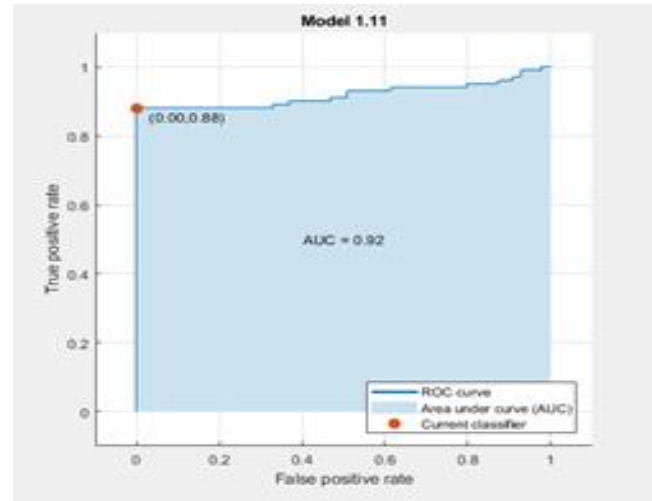
Figure 4: The ROC Curve of Medium Gaussian SVM

The receiver operating characteristic curve shows the relationship of the inputted features to the classifier. It can be seen in the Figure 4 that the area under the curve is 0.92, this indicates that there is a higher percentage of classification using the inputted features.

Figure 5 shows the parallel coordinates plot of the Medium Gaussian SVM. It can be seen in Figure 5 that HSV values for mature coffee beans (orange color) is below those of the immature coffee beans (blue color). Although the values were overlapping, the Medium Gaussian SVM was able to differentiate between the two classes.

Figure 5: Parallel Coordinates Plot of the Medium Gaussian SVM.

Table 3 compares this study from the previous study.

Table 3: Comparison of Results

Study	Fea-tures	Classifier	Accuracy	Training Time
Previous Study[5]	RGB	Quadratic SVM	94 %	0.62 seconds
This Study	HSV	Medium Gaussian SVM	94 %	0.14624seconds

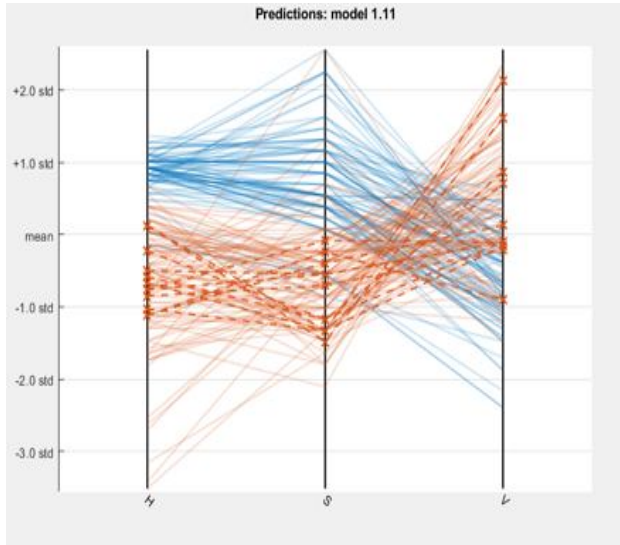Table 3 shows the comparison of the previous study [5] to this study . In the previous study that uses RGB as features for classification , an accuracy of 94 % was achieved by Quadratic SVM while in this study that uses HSV, an accuracy of 94 % was also achieved by Medium Gaussian SVM. The two differs in terms of speed, the previous study can achieved 94% accuracy at 0.62 seconds and this study can do it in 0.14624 seconds.

4. CONCLUSION

The research question being answered in this study is which color features, RGB or HSV is better to use for classification of immature and mature coffee beans. Although the classification accuracy of 94 % is the same

between [5] and this study, this study achieved faster training time. It can be concluded that in order to discriminate mature from immature coffee beans it is better to use the HSV features because it is 4 times faster compared to using the RGB features.

REFERENCES
[1]   M. Murphy and T. J. Dowding, "The Coffee Bean: A Value Chain and Sustainability Initiatives Analysis," pp. 1–19, 2017.
[2]   K. P. P. Nair, "6-Coffee," in *The Agronomy and Economy of Important Tree Crops of the Developing World*, Elsevier, 2010, pp. 181–208.
[3]   A. Farah and T. Ferreira, "The Coffee Plant and Beans: An Introduction," in *Coffee in Health and Disease Prevention*, V. R. Preedy, Ed. Academic Press, 2015, pp. 5–10.
[4]   M. Firmin, "From Bean to Cup : a history of coffee," 2014.
[5]   W. R. Eustaquio and J. L. Dioses Jr., "Classification of Immature and Mature Coffee Beans Using RGB Values and Machine Learning Algorithms," *Int. J. Emerg. Trends Eng. Res.*, 2020. https://doi.org/10.30534/ijeter/2020/22872020
[6]   E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "An Image Processing Technique for Coffee Black Beans Identification," in *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, 2018, no. May, pp. 1–5.
[7]   J. N. C. Sarino, M. M. Bayas, E. R. Arboleda, E. C. Guevarra, and R. M. Dellosa, "Classification Of Coffee Bean Degree Of Roast Using Image Processing And Neural Network," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 3231–3233, 2019.
[8]   E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "Green coffee beans feature extractor using image processing," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 18, no. 4, p. 2027, 2020.
[9]   J. S. Kumari and U. R. Nelakuditi, "Gestation Age Determination through Optimized Fuzzy Logic Based Edge Detection Algorithm in Fetal Ultrasound Images," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3057–3068, 2020. https://doi.org/10.30534/ijatcse/2020/87932020
[10]   A. E. Andaya, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa, "Meat Marbling Scoring Using Image Processing with Fuzzy Logic Based Classifier," *Int. J. Sci. Technol. Res.*, vol. 8, no. 08, pp. 1442–1445, 2019.
[11]   J. G. C. Rancapan, E. R. Arboleda, J. L. Dioses Jr, and R. M. Dellosa, "Egg Fertility Detection Using Image Processing And Fuzzy Logic," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 3228–3230, 2019.
[12]   R. A. Magbayao, E. R. Arboleda, and E. M. Galas, "Identification Of Asian Green Mussel Perna Viridis ' Sex Using Image Processing , Fuzzy Logic And K – Nearest Neighbor," *Int. J. Sci. Technol. Res.*, vol. 9, no. 01, 2020.
[13]   C. I. A. Baluran, E. R. Arboleda, M. G. Dizon, and

R. M. Dellosa, "Crab Gender Classification Using Image Processing , Fuzzy Logic And K Nearest Neighbor (KNN) Classifier," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 1458–1462, 2019.

[14] V. M. D. Manalo, E. R. Arboleda, J. L. D. Jr, and R. M. Dellosa, "Differentiation Among Lettuce (L . Sativa) Seed Varieties Grown In Gourmet Farms , Silang Cavite , Philippines Using Image Processing With Fuzzy Logic And Knn As Classifiers," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 8–11, 2019.

[15] R. E. T. Bae, E. R. Arboleda, A. Andilab, and R. M. Dellosa, "Implementation Of Template Matching , Fuzzy Logic And K Nearest Neighbor Classifier On Philippine Banknote Recognition System," *Int. J. Sci. Technol. Res.*, vol. 8, no. 08, pp. 1451–1453, 2019.

[16] V. A. C. Austria, E. R. Arboleda, and E. M. Galas, "Image Processing Of Clean And Dirty Dishes To Design And Construct A Fuzzy Logic Dishwasher," *Int. J. Sci. Technol. Res.*, vol. 8, no. 12, pp. 8–11, 2019.

[17] J. A. A. Garcia, E. R. Arboleda, and E. M. Galas, "Identification Of Visually Similar Vegetable Seeds Using Image Processing And Fuzzy Logic," *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, pp. 4925–4928, 2020.

[18] F. B. Vidad III, E. R. Arboleda, J. L. Dioses Jr, and R. M. Dellosa, "Morphological Based Seedling Comparison Of Natural And Hybrid Lemon," *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, 2020.

[19] E. J. Olaes, E. R. Arboleda, J. L. D. Jr, and R. M. Dellosa, "Bell Pepper And Chili Pepper Classification : An Application Of Image Processing And Fuzzy Logic," *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, pp. 4833–4839, 2020.

[20] C. L. Macalalad, E. R. Arboleda, A. A. Andilab, and R. M. Dellosa, "Morphological Based Grain Comparison of Three Rice Grain Variety," *Int. Jounal Sci. Technol. Res.*, vol. 8, no. 08, pp. 1446–1450, 2019.

[21] E. R. Arboleda, A. C. Fajardo, and R. P. Medina, "Classification of Coffee Bean Species Using Image Processing , Artificial Neural Network and K Nearest Neighbors," in *2018 IEEE International Conference on Innovative Research and Development (ICIRD)*, 2018, no. May, pp. 1–5.

[22] D. L. Bersabal, E. R. Arboleda, and E. M. Galas, "Coffee Bean Recognition Using Shape Features Using Decision Trees And Ensemble Classifiers," *Int. J. Sci. Technol. Res.*, vol. 9, no. 02, 2020.

[23] A. J. G. Gonzales, J. A. D. Sosa, E. R. Arboleda, and E. M. Galas, "Filtering Of Faded Coffee Beans Using Image Processing," *Int. J. Sci. Technol. Res.*, vol. 8, no. 12, pp. 3634–3637, 2019.

[24] A. M. Castillo, R. D. Aradanas, E. R. Arboleda, A. A. Dizon, and R. M. Dellosa, "Coffee Type Classification Using Gray Level Co- Occurrence Matrix Feature Extraction And The Artificial Neural Network Classifier," *Int. J. Sci. Technol. Res.*, vol. 8, no. 10, pp. 1463–1465, 2019.

[25] E. R. Arboleda, "Comparing Performances of Data Mining Algorithms for Classification of Green Coffee Beans," *Int. J. Eng. Adv. Technol.*, vol. 8, no. 5, pp. 1563–1567, 2019.

[26] E. R. Arboleda, "Discrimination of civet coffee using near infrared spectroscopy and artificial neural network," *Int. J. Adv. Comput. Res.*, vol. 8, no. 39, pp. 324–334, 2018.

[27] J. L. Dioses, "Discrimination of civet coffee using image processing and machine learning," *Int. J. Emerg. Trends Eng. Res.*, vol. 8, no. 4, pp. 1067–1072, 2020.
https://doi.org/10.30534/ijeter/2020/19842020

[28] J. L. Dioses, "Classification of pepper seeds using data mining algorithms," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1460–1464, 2020.
https://doi.org/10.30534/ijatcse/2020/84922020

APPENDIX

Data Set of the Study

Sample No.	H	S	V	Class
1	148	57	114	1
2	147	57	122	1
3	148	59	125	1
4	149	58	121	1
5	148	57	131	1
6	149	57	126	1
7	150	55	140	1
8	150	58	134	1
9	140	49	110	1
10	148	58	118	1
11	149	55	120	1
12	147	55	115	1
13	150	62	118	1
14	140	51	106	1
15	143	51	98	1
16	147	69	121	1
17	148	67	124	1
18	147	65	135	1
19	151	64	139	1
20	151	69	134	1
21	151	64	140	1
22	149	65	133	1
23	152	71	130	1
24	152	63	138	1
25	148	63	130	1
26	149	67	135	1
27	151	69	121	1
28	148	68	123	1
29	152	69	124	1

30	150	62	140	1	79	151	58	129	1
31	150	61	131	1	80	149	55	118	1
32	155	62	140	1	81	150	57	117	1
33	151	60	136	1	82	150	56	119	1
34	154	62	136	1	83	151	61	124	1
35	151	61	136	1	84	150	58	116	1
36	150	60	126	1	85	150	57	111	1
37	149	59	133	1	86	150	58	120	1
38	151	61	123	1	87	148	59	110	1
39	153	64	132	1	88	138	51	94	1
40	149	60	121	1	89	127	52	138	1
41	151	62	129	1	90	141	45	164	1
42	149	58	103	1	91	134	51	134	1
43	154	65	123	1	92	141	46	151	1
44	145	57	103	1	93	132	46	133	1
45	151	61	142	1	94	131	54	134	1
46	148	61	136	1	95	131	51	134	1
47	146	60	124	1	96	133	47	134	1
48	150	63	129	1	97	130	50	132	1
49	152	57	149	1	98	127	53	120	1
50	152	58	144	1	99	128	46	173	1
51	154	62	146	1	100	137	47	148	1
52	150	58	144	1	1	127	52	138	2
53	149	58	128	1	2	141	45	164	2
54	150	60	129	1	3	134	51	134	2
55	153	63	143	1	4	141	46	151	2
56	146	58	119	1	5	132	46	133	2
57	152	62	128	1	6	131	54	134	2
58	146	57	117	1	7	131	51	134	2
59	145	56	115	1	8	133	47	134	2
60	153	63	132	1	9	130	50	132	2
61	152	60	129	1	10	127	53	120	2
62	151	61	123	1	11	128	46	173	2
63	150	56	113	1	12	137	47	148	2
64	151	58	129	1	13	132	48	165	2
65	149	55	118	1	14	138	46	144	2
66	150	57	117	1	15	130	48	156	2
67	150	56	119	1	16	125	48	121	2
68	151	61	124	1	17	134	49	150	2
69	150	58	116	1	18	134	49	137	2
70	150	57	111	1	19	110	51	135	2
71	150	58	120	1	20	134	46	135	2
72	148	59	110	1	21	126	43	166	2
73	138	51	94	1	22	125	46	148	2
74	151	62	114	1	23	132	47	139	2
75	153	63	132	1	24	120	49	152	2
76	152	60	129	1	25	101	51	146	2
77	151	61	123	1	26	121	62	117	2
78	150	56	113	1	27	135	47	144	2

28	123	41	177	2	77	123	52	165	2
29	134	45	142	2	78	142	49	145	2
30	133	43	174	2	79	135	55	123	2
31	127	47	149	2	80	137	52	169	2
32	132	46	153	2	81	140	52	141	2
33	133	50	137	2	82	144	52	148	2
34	125	55	116	2	83	141	49	151	2
35	126	53	119	2	84	140	50	141	2
36	126	47	142	2	85	130	55	120	2
37	100	46	173	2	86	142	51	148	2
38	129	49	136	2	87	128	50	157	2
39	127	46	145	2	88	122	46	177	2
40	109	54	159	2	89	129	48	161	2
41	104	44	176	2	90	140	54	146	2
42	127	48	136	2	91	138	53	153	2
43	134	45	130	2	92	136	54	159	2
44	121	49	125	2	93	144	55	138	2
45	120	48	115	2	94	130	71	116	2
46	121	50	147	2	95	134	53	142	2
47	120	50	161	2	96	138	51	141	2
48	138	44	146	2	97	131	59	136	2
49	123	50	126	2	98	140	51	155	2
50	111	52	154	2	99	119	58	110	2
51	136	57	131	2	100	139	53	135	2
52	135	53	132	2					
53	144	53	159	2					
54	135	59	132	2					
55	143	52	155	2					
56	144	56	142	2					
57	131	55	121	2					
58	147	51	163	2					
59	142	48	168	2					
60	142	51	160	2					
61	140	52	152	2					
62	140	55	142	2					
63	139	50	135	2					
64	137	53	134	2					
65	130	58	122	2					
66	144	45	172	2					
67	135	51	160	2					
68	127	53	170	2					
69	142	50	149	2					
70	137	54	157	2					
71	138	53	134	2					
72	131	63	120	2					
73	144	52	151	2					
74	139	46	169	2					
75	141	52	151	2					
76	146	43	164	2					