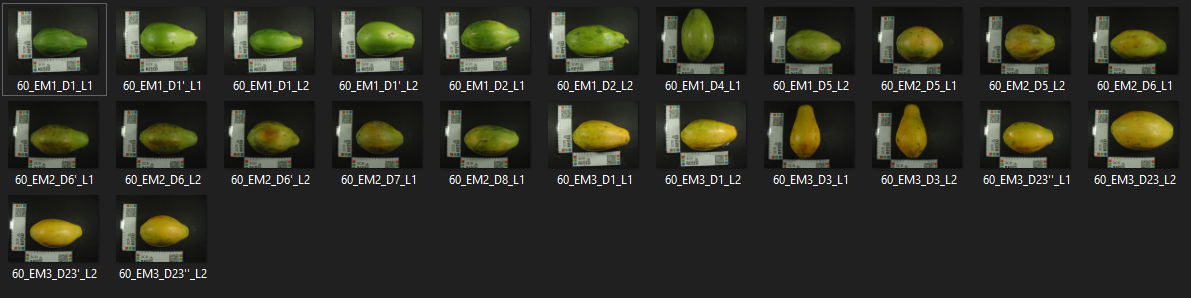
# Pré-processamento das Amostras

Foram disponibilidas 24 imagens de mamãos. Cada imagem possui entre 1 MB e 1.7 MB de tamanho, portanto, o primeiro passo foi criar uma pasta chamada “mamao-menor” e utilizar um algoritmo para reduzir as imagens em 60%, para facilitar o processamento.

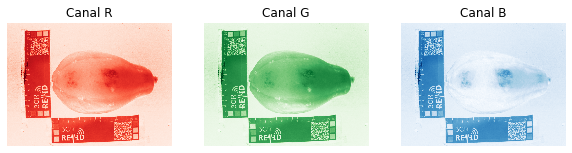
Cada novo arquivo teve o número 60 adicionado ao início do seu nome, para identificar o percentual de redução aplicado.

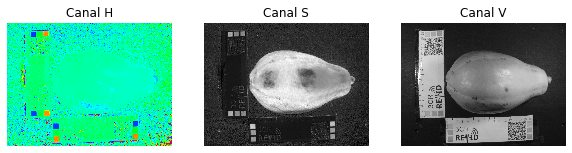


# Desenvolvendo o Algoritmo de Segmentação

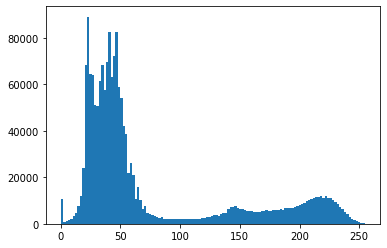
A estratégia selecionada para o desenvolvimento do algoritmo foi a de selecionar 1 amostra, desenvolver o algoritmo e depois aplicar para todas as outras e aplicar o *tuning* conforme necessário.

O Primeiro passo foi plotar as amostras nas camadas RGB e HSV. Após analisar os resultados, foi decidido utilizar o canal S da camada HSV, pois apresentou uma boa definição do objeto na imagem.

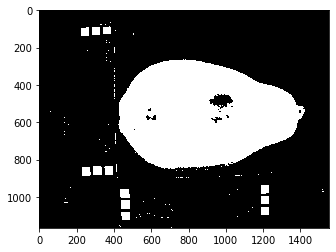




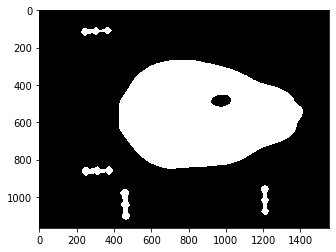
Após isolar o canal S da imagem, foi analisado o histograma do canal S para definir o limiar a ser utilizado. Foi decidido então por utilizar o limiar entre 100 e 260, pois os bins menores representam o fundo e outros artefatos na imagem.



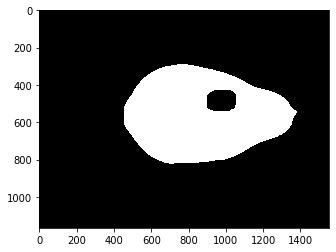
Ao aplicar o limiar na imagem, utilizando o arquivo iMFerramentas, o resultado foi que o fundo foi eliminado, sobrando apenas alguns ruiídos causados provavelmente pelo flash da câmera utilizada e pelo marcador utilizado.



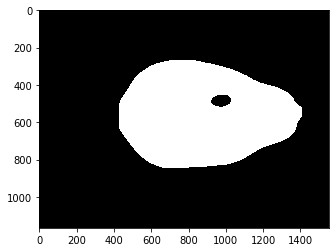
Em seguida foi aplicado o filtro Passa-Baixa para suavizar as imperfeições e eliminar os ruídos menores.



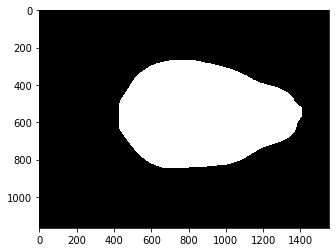
Após aplicar o filtro blur, foi utilizada a técnica de erosão para remover as informações do marcador. Essa etapa teve que ser reajustada no final, pois o número de iterações definidos no início não trouxe resultados bons para outras amostras, após verificar o plot de todas as imagens, foi decidido por utilizar 3 iterações.



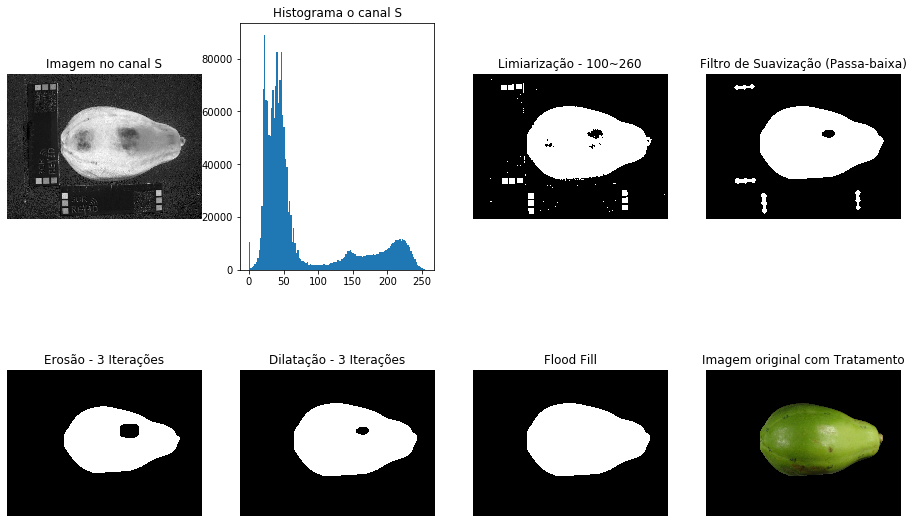
Com a erosão, algumas informações (como as bordas) foram perdidas. Então foi utilizada a técnica de dilatação para recuperar as informações perdidas na Erosão.

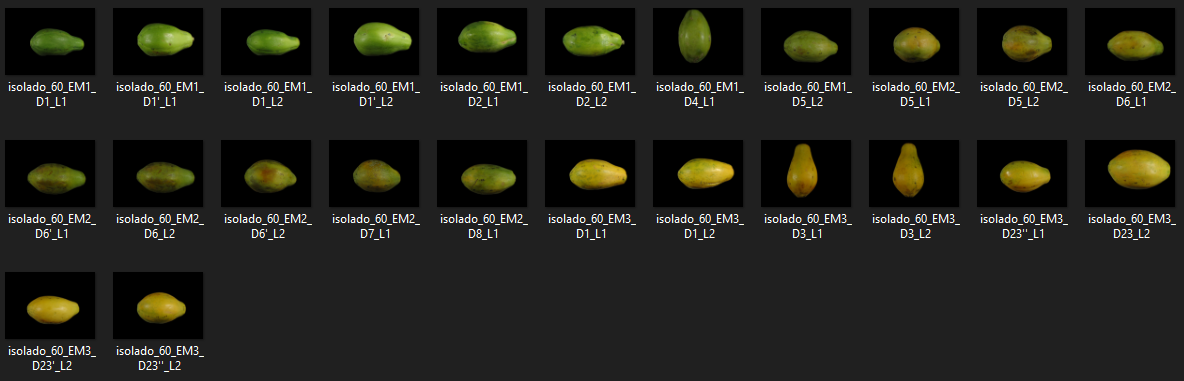


Como a amostra continuava com falhas internas no ROI (*region of interest*), e para esse estudo é interessante obter toda a imagem da amostra, foi decidido aplicar o método Flood Fill para preencher os buracos que ficaram em algumas amostras.



Por fim, a máscara foi aplicada em todas as imagens para validação. Após o *tuning* do algoritmo de segmentação, todas as imagens resultantes foram salvas na pasta “mamao-tratado”.



****

# Extraindo e analisando os descritores de cor

Sabendo que as amostras foram classificadas de acordo com o grau de maturação obtido pelo exame laboratorial, as imagens foram classificadas utilizando essa informação, sendo:

**Classe EM1** – Estágio de Maturação 1

**Classe EM2** – Estágio de Maturação 2

**Classe EM3** – Estágio de Maturação 3

Em seguida, das imagens tratadas após a aplicação do algoritmo de segmentação, foram extraídas as seguintes features:

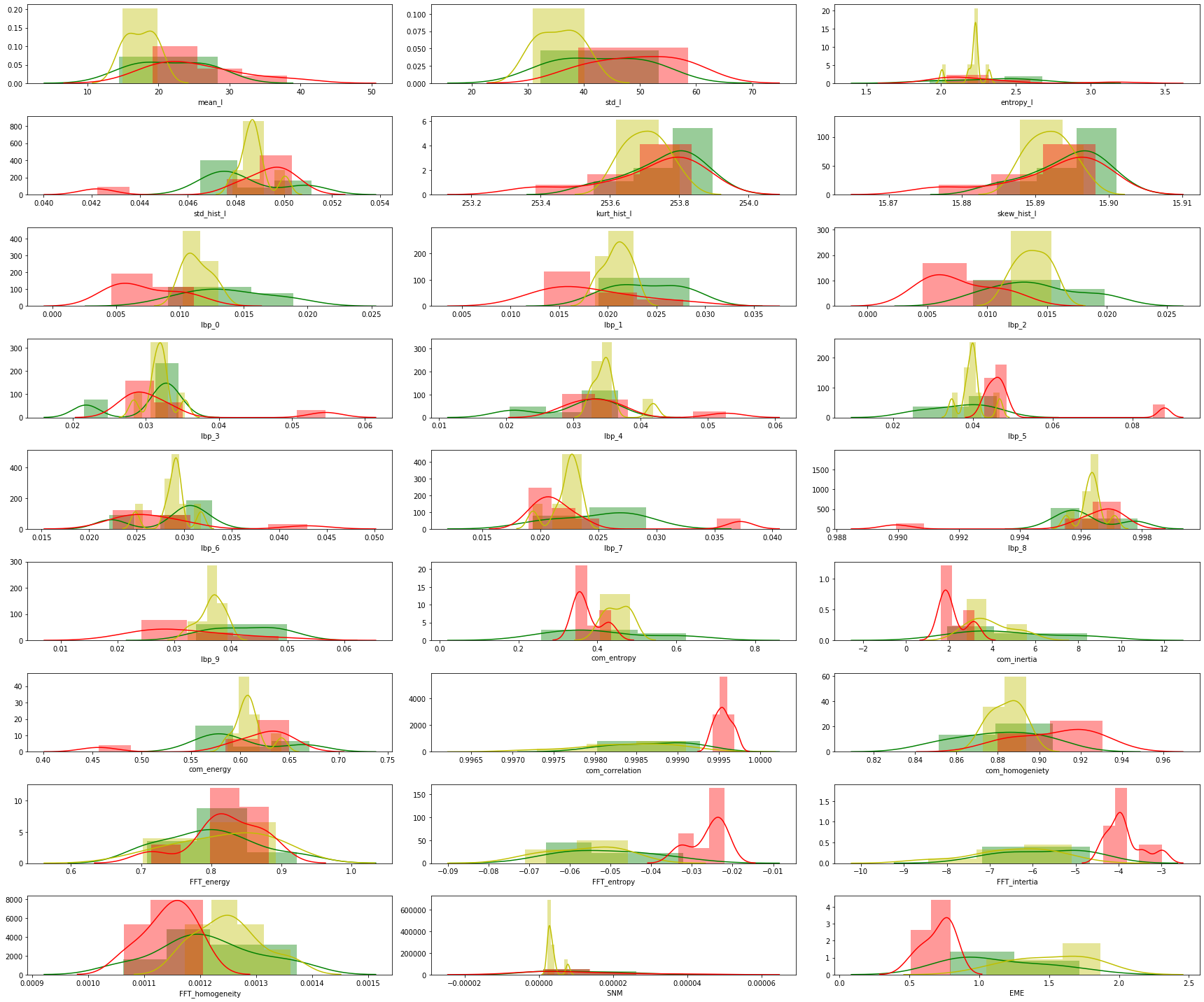
**mean\_I:** média do canal I (HSI)  
**std\_I:** desvio padrão do canal I (HSI)  
**entropy\_I:** entropia do canal de intensidade  
**SNM:** Métrica de qualidade de imagem  
**EME:** Métrica de melhoria  
**hu\_sobel[1–7] e hu\_canny[1–7]:** são métricas de borda.  
**std\_hist\_I:** desvio padrão do histograma  
**kurt\_hist\_I:** kurtosis do histograma  
**skew\_hist\_I:** Skewness do histograma  
**lbp [0–9]:** Textura via LBP  
**com\_entropy:** Textura via entropia  
**com\_inertia:** Textura via inércia  
**com\_energy:** Textura via energia  
**com\_correlation:** Textura via correlação  
**com\_homogeniety:** Textura via homogeneidade  
**FFT\_energy:** Textura via energia no domínio de Fourier  
**FFT\_entropy:** Textura via entropia no domínio de Fourier  
**FFT\_intertia:**Textura via inércia no domínio de Fourier  
**FFT\_homogeneity:**Textura via homogeneidade no domínio de Fourier

Essas informações foram salvas em um dataset em formato CSV (Dataset\_mamao.csv), com a adição do nome do arquivo na primeira coluna e da classe (grau de maturação) na última coluna.

O primeiro passo da análise foi plotar cada feature em formato de gráfico de distribuição (distplot), utilizando a biblioteca *seaborn*.

Os gráficos abaixo representam todas as amostras separadas por classes, sendo EM1 na cor Verde, EM2 na cor Amarela e EM3 na cor Vermelha

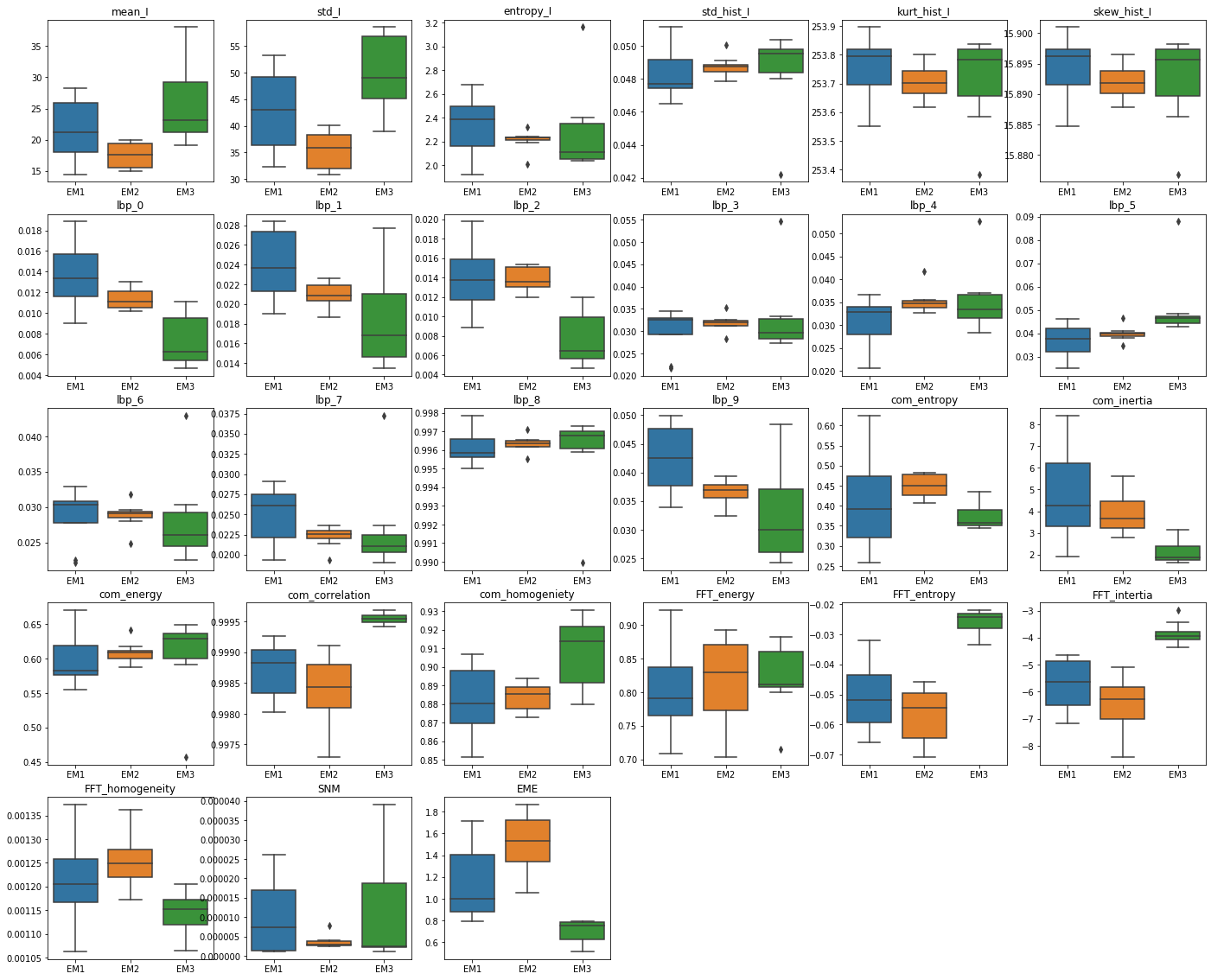
Cada gráfico representa uma das features extraídas das imagens.



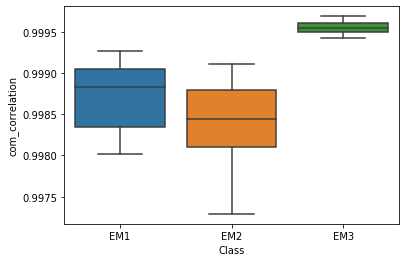
**Resultado da análise**: Foi inconclusivo, pois utilizando essa visualização não ficou muito claro quais seriam as melhores features para diferenciar as amostras.

Para uma melhor análise, cada feature for então plotada seguindo a mesma forma, porém utilizando *Bloxplots*, também da biblioteca seaborn.

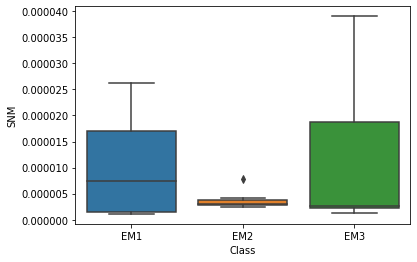
Neste caso, o EM1 está representado pela cor Azul, o EM2 pela cor Laranja e o EM3 pela cor Verde.



**Resultado da análise:** Com a análise de bloxplots, ficou mais visível a segmentação das classes. Por exemplo, a feature **com\_correlation** segrega claramente todas as amostras de classe EM3, pois todas possuem valor superior a 0.9994, enquanto as amostras de outras classes possuem valores sempre inferiores.



Das amostras que não se encaixarem neste critério, foi observado que a feature **SNM** apresenta uma distinção aceitávem entre as 2 classes, sendo considerada **SNM** todas as que não se encaixam no critério anterior utilizado para a EM3, e que possuem o valor de **SNM** dentro do range 0.000002 à 0.000008.



# Comparação dos Resultados

O algoritmo então comparou as amostras e atribuiu o grau de maturação predito de acordo com os parâmetros acima mencionados. O resultado foi salvo em um arquivo CSV chamado mamao-resultado.csv e ficou da seguinte forma:

|  |  |  |
| --- | --- | --- |
| **File** | **Class** | **Result** |
| isolado\_60\_EM1\_D1'\_L1.JPG | EM1 | EM1 |
| isolado\_60\_EM1\_D1'\_L2.JPG | EM1 | EM1 |
| isolado\_60\_EM1\_D1\_L1.JPG | EM1 | EM1 |
| isolado\_60\_EM1\_D1\_L2.JPG | EM1 | EM1 |
| isolado\_60\_EM1\_D2\_L1.JPG | EM1 | EM1 |
| isolado\_60\_EM1\_D2\_L2.JPG | EM1 | EM1 |
| isolado\_60\_EM1\_D4\_L1.JPG | EM1 | EM1 |
| isolado\_60\_EM1\_D5\_L2.JPG | EM1 | EM2 |
| isolado\_60\_EM2\_D5\_L1.JPG | EM2 | EM2 |
| isolado\_60\_EM2\_D5\_L2.JPG | EM2 | EM2 |
| isolado\_60\_EM2\_D6'\_L1.JPG | EM2 | EM2 |
| isolado\_60\_EM2\_D6'\_L2.JPG | EM2 | EM2 |
| isolado\_60\_EM2\_D6\_L1.JPG | EM2 | EM2 |
| isolado\_60\_EM2\_D6\_L2.JPG | EM2 | EM2 |
| isolado\_60\_EM2\_D7\_L1.JPG | EM2 | EM2 |
| isolado\_60\_EM2\_D8\_L1.JPG | EM2 | EM2 |
| isolado\_60\_EM3\_D1\_L1.JPG | EM3 | EM3 |
| isolado\_60\_EM3\_D1\_L2.JPG | EM3 | EM3 |
| isolado\_60\_EM3\_D23''\_L1.JPG | EM3 | EM3 |
| isolado\_60\_EM3\_D23''\_L2.JPG | EM3 | EM3 |
| isolado\_60\_EM3\_D23'\_L2.JPG | EM3 | EM3 |
| isolado\_60\_EM3\_D23\_L2.JPG | EM3 | EM3 |
| isolado\_60\_EM3\_D3\_L1.JPG | EM3 | EM3 |
| isolado\_60\_EM3\_D3\_L2.JPG | EM3 | EM3 |

O campo ***File*** representa o nome do arquivo verificado, o campo ***Class*** refere-se ao grau de maturação obtido por meio de análise laboratorial das amostras e por fim o campo ***Result*** refere-se à classificação que o algoritmo atribuiu à amostra utilizando os parâmetros **com\_correlation** e **SNM.**

Das 24 amostras verificadas, 1 foi predita de forma errada pelo algoritmo.