

Resultados:

1. Médias:

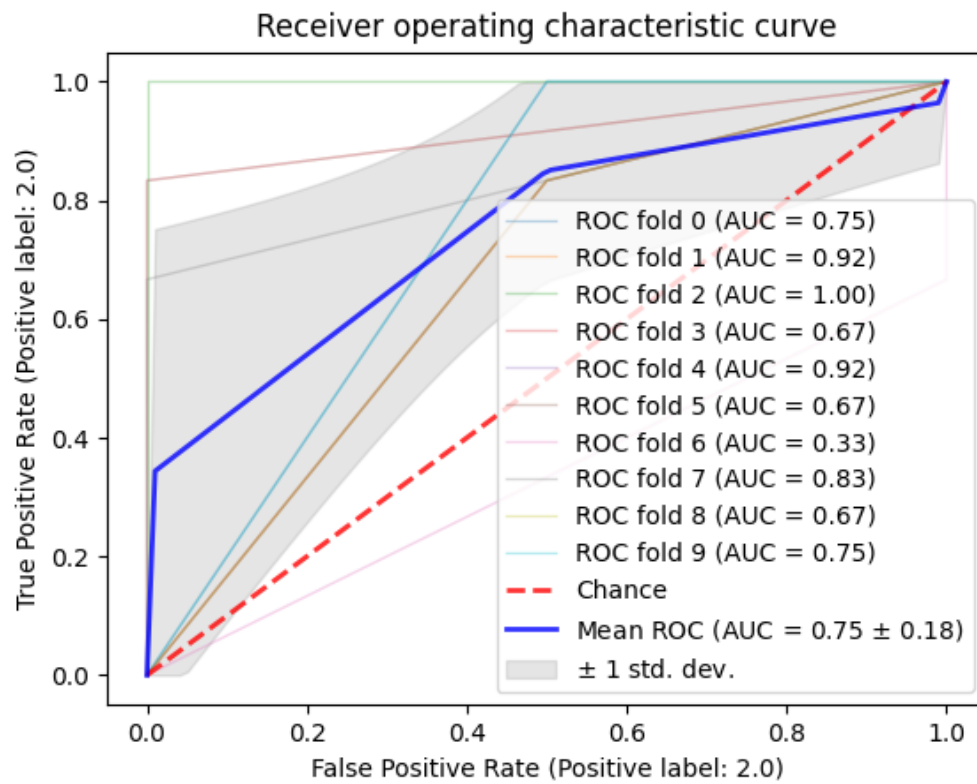
SEX	1.112500
STEROID	1.500000
ANTIVIRALS	1.875000
FATIGUE	1.329114
MALAISE	1.607595
ANOREXIA	1.784810
LIVER_BIG	1.786667
LIVER_FIRM	1.540541
SPLEEN_PALPABLE	1.753247
SPIDERS	1.649351
ASCITES	1.857143
VARICES	1.844156
BILIRUBIN	1.455844
ALK_PHOSPHATE	114.095238
SGOT	98.657895
ALBUMIN	3.828571
PROTIME	59.952381
HISTOLOGY	1.537500

Desvios:

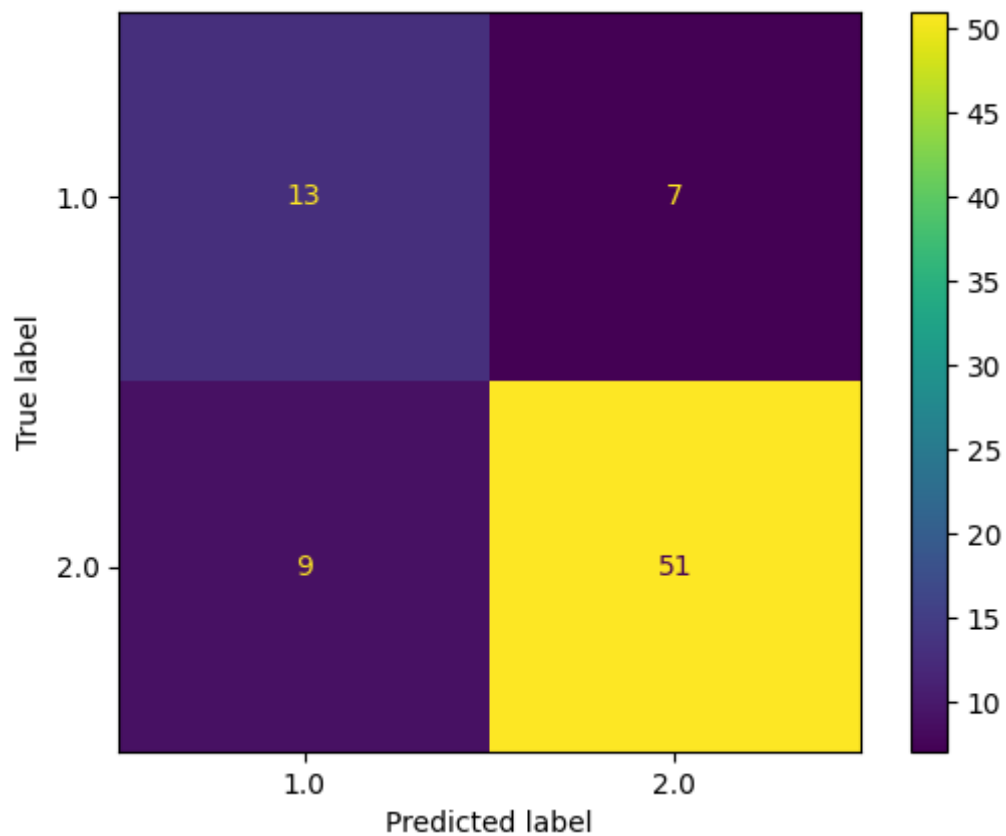
SEX	0.317974
STEROID	0.503155
ANTIVIRALS	0.332805
FATIGUE	0.472894
MALAISE	0.491406
ANOREXIA	0.413580
LIVER_BIG	0.412420
LIVER_FIRM	0.501756
SPLEEN_PALPABLE	0.433949
SPIDERS	0.480302
ASCITES	0.352222
VARICES	0.365086
BILIRUBIN	1.351655
ALK_PHOSPHATE	58.586499
SGOT	110.588312
ALBUMIN	0.729133
PROTIME	22.971830
HISTOLOGY	0.501737

2. Árvore de Decisão:

Curvas Roc:

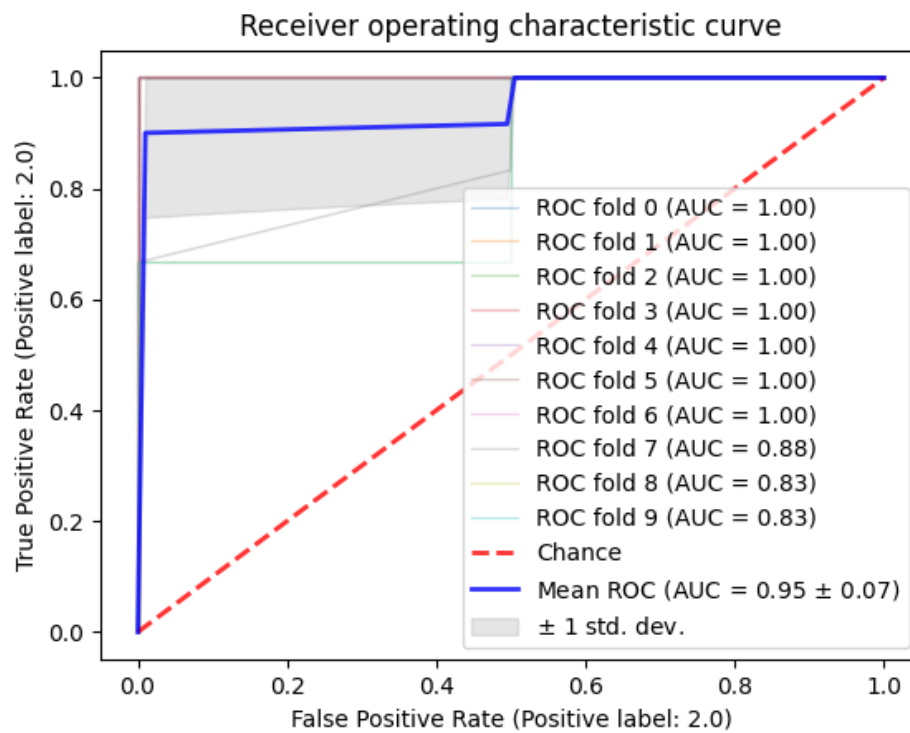


Matriz de confusão (com soma de todas as matrizes da validação cruzada):

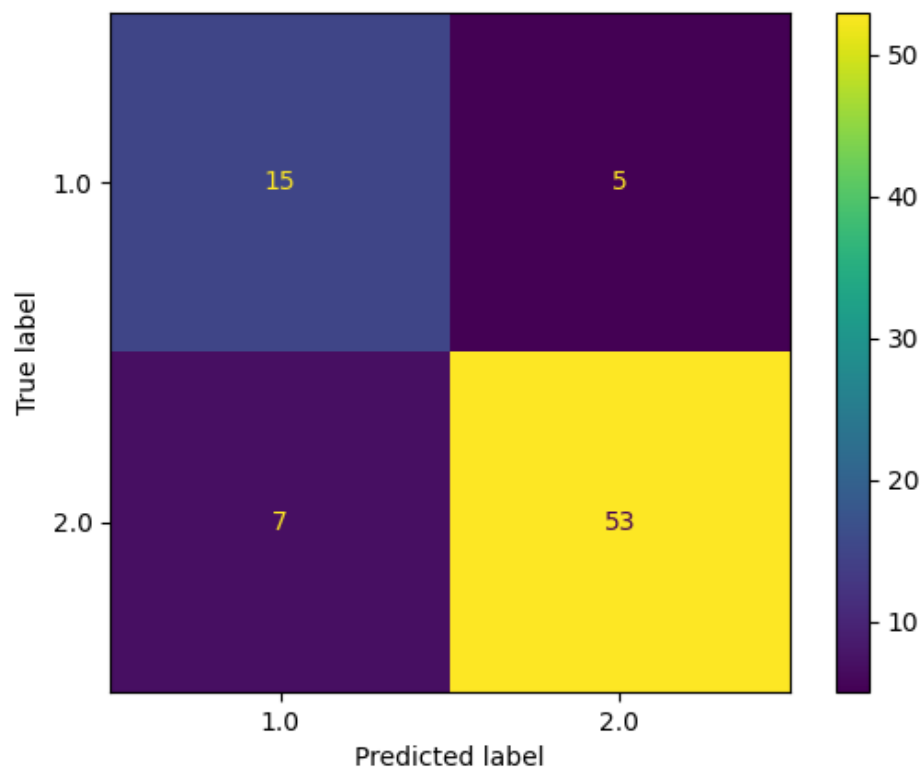


3. Floresta Randômica (m = 19):

Curvas Roc:

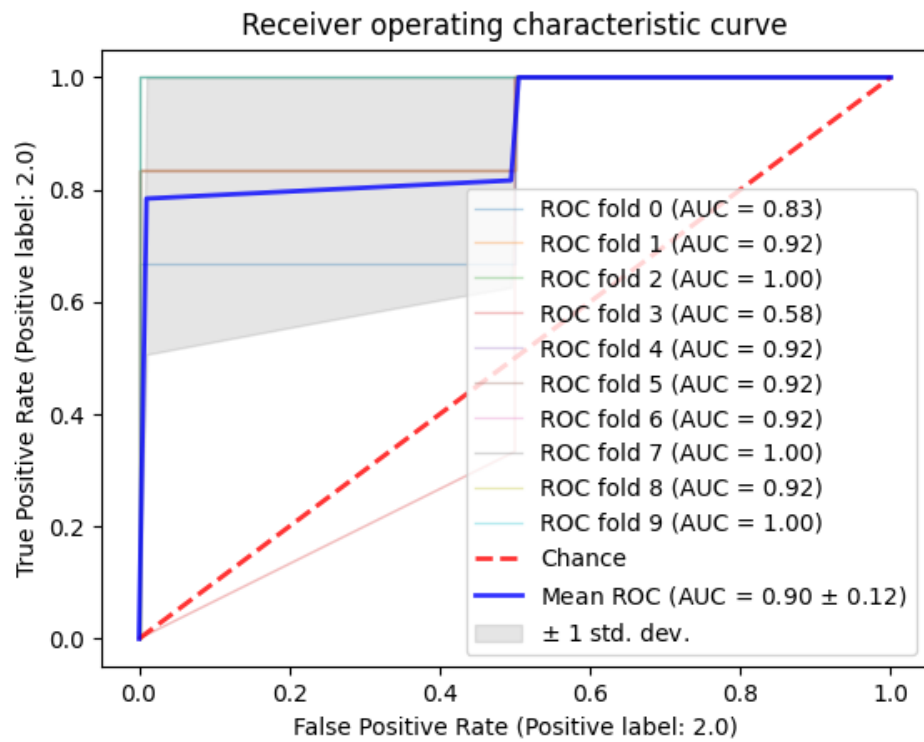


Matriz de confusão (com soma de todas as matrizes da validação cruzada):

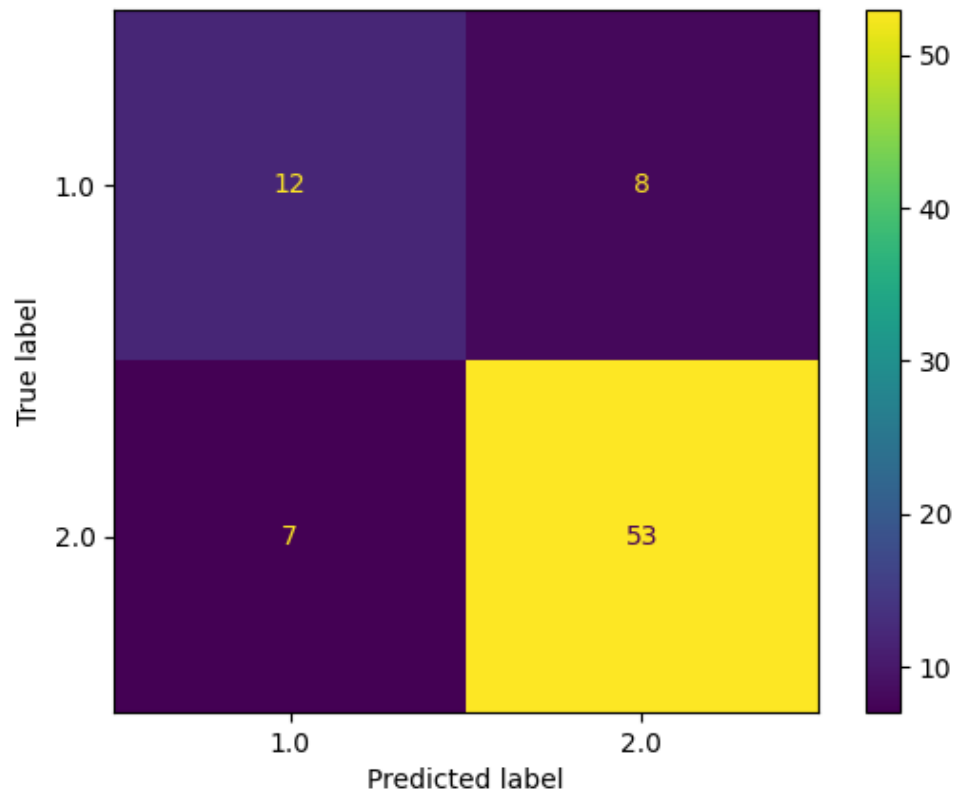


4. Floresta Randômica ($m = \sqrt{19}$):

Curvas Roc:

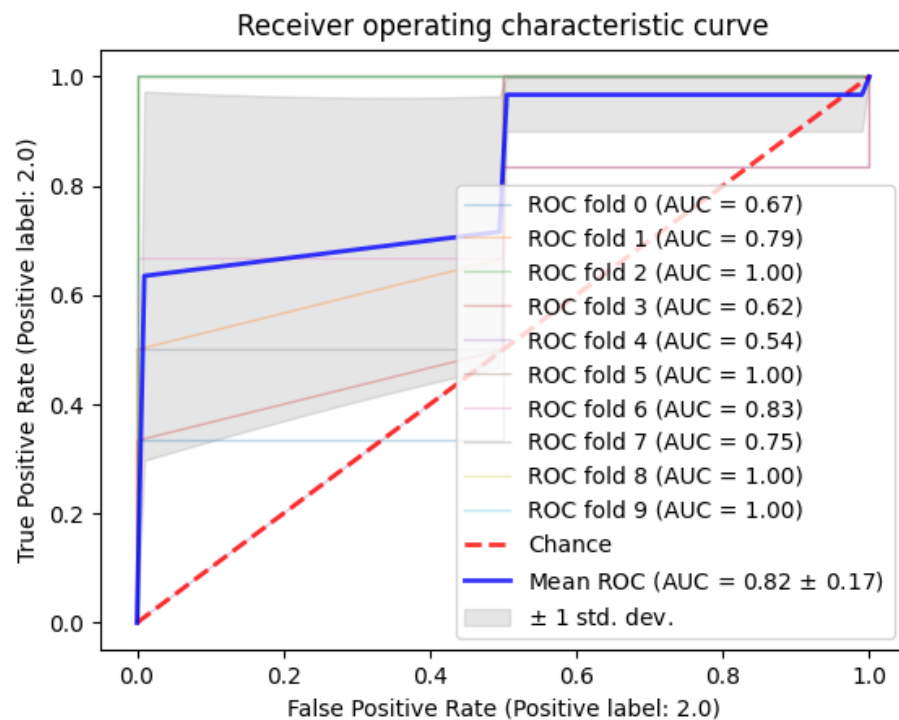


Matriz de confusão (com soma de todas as matrizes da validação cruzada):

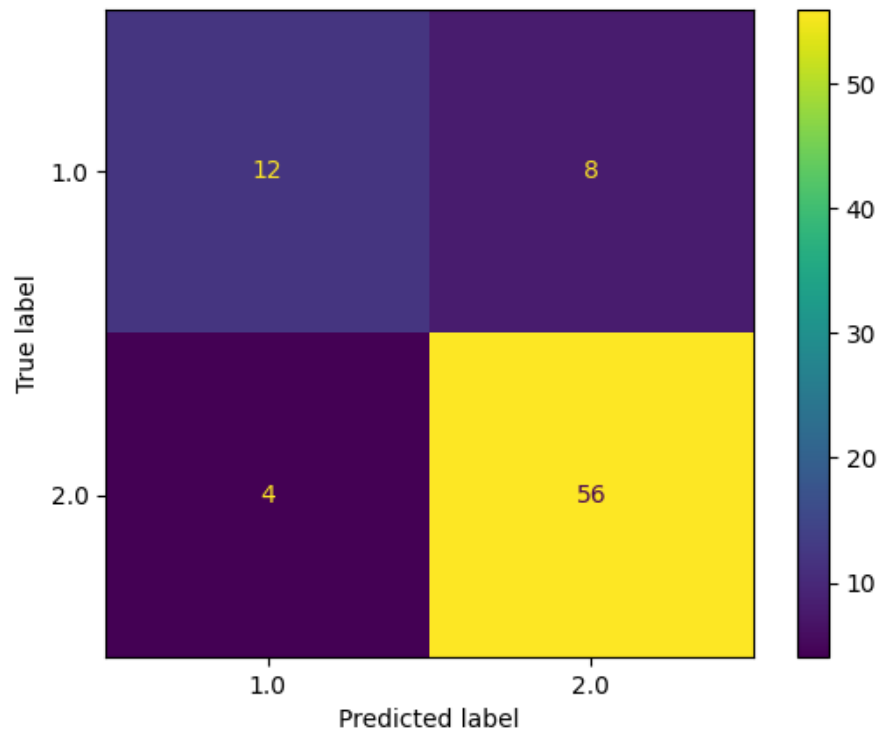


5. O melhor resultado foi do algoritmo de floresta randômica com $m = 19$, modifiquei o código para exibir uma lista da importância de cada feature e rodei o programa de novo, obtive os seguintes resultados:

Curva Roc:



Matriz de confusão (com soma de todas as matrizes da validação cruzada):



Média da importância de cada feature nas 10 validações:

SEX :	0.027902419737711837
STEROID :	0.005633104652458138
ANTIVIRALS :	0.002245603766410047
FATIGUE :	0.023911901570906988
MALAISE :	0.010986580690711128
ANOREXIA :	0.00887798569210119
LIVER_BIG :	0.008701925515982317
LIVER_FIRM :	0.013699080365087657
SPLEEN_PALPABLE :	0.01969606151859001
SPIDERS :	0.05373187352056487
ASCITES :	0.1669516979533462
VARICES :	0.0072288395705310355
BILIRUBIN :	0.17962528630863348
ALK_PHOSPHATE :	0.11006070102903995
SGOT :	0.05775738999406246
ALBUMIN :	0.24822044472790603
PROTIME :	0.041180184749598316
HISTOLOGY :	0.013588918636358402

Como pode-se observar as duas features de maior importância são:
ALBUMIN e BILIRUBIN nessa ordem.