

Bruno Severino Mascarenhas

Relatório das Práticas Laboratoriais

Visão Computacional e Reconhecimento
de Padrões

Salvador

2021

Sumário

1 Prática 4 – Análise de resultados.....	3
1.1 Objetivo.....	3
1.2 Descrição da prática.....	3
1.2.1 Roteiro I.....	3
1.2.2 Roteiro II.....	3
1.2.3 Roteiro III.....	3
1.3 Análise crítica dos resultados.....	4
1.3.1 Curva ROC e curva Precision-Recall.....	4
1.4 Referências.....	7

1 Prática 4 – Análise de resultados

1.1 Objetivo

A prática deve possibilitar um entendimento sobre as análises de resultados e suas formas de representação. Nesta etapa serão introduzidas as duas métricas de avaliação de resultado: ROC e Precision-Recall. Ao fim desta prática, espera-se reflexões sobre as análises e resultados obtidos.

1.2 Descrição da prática

1.2.1 Roteiro I

1. A partir das informações de predição dos classificadores da prática anterior (usando os dois modelos), calcule os valores da curva ROC;
2. Em seguida, gere um gráfico da curva ROC para cada modelo e salve as imagens. Estas imagens devem ser utilizadas no procedimento de análise no relatório;

1.2.2 Roteiro II

1. A partir das informações de predição dos classificadores da prática anterior, calcule os valores de Precision e Recall para cada classificador;
2. Em seguida, gere um gráfico de Precision-Recall para cada classificador e salve as imagens. Estas imagens devem ser utilizadas no procedimento de análise no relatório.

1.2.3 Roteiro III

1. A partir das informações de predição dos classificadores da prática anterior, calcule os valores médios da métrica F1-score para cada classificador (média dos dois modelos gerados no cross-validation).

1.3 Análise crítica dos resultados

Como discutido na prática anterior, os melhores resultados de ambos os classificadores foram obtidos com o conjunto de dados onde apenas os descritores de imagem GLCM – contraste e LBP foram utilizados, entretanto os resultados analisados nesta prática referem-se a todos os outros conjuntos de dados testado, porém com as melhores configurações – hiperparâmetros obtidos através das práticas anteriores. Os seguintes grupos foram analisados: apenas LBP, apenas GLCM, apenas o vetor de imagens (IMG), LBP + GLCM, LBP + IMG, GLCM + IMG, LBP + GLCM + IMG

As classificações foram divididas em quatro categorias: positivo verdadeiro, falso positivo, negativo verdadeiro e falso negativo. As métricas avaliadas utilizam como base essas quatro categorias.

1.3.1 Curva ROC e curva Precision-Recall

A curva ROC é um gráfico que representa a taxa de falsos positivos contra a taxa de positivos verdadeiros. É uma boa maneira de visualizar outros aspectos dos classificadores, além da acurácia média, muito usada para analisar a performance do classificador numa só figura (Bradley, 1997) e é comumente utilizada para problemas de classificação binária devido a sua natureza um para um, mas pode ser generalizado para problemas de várias classes como mostrado em Hand (2001).

A taxa de verdadeiros positivos é definida como a quantidade de verdadeiros positivos dividido pela soma dos verdadeiros positivos com os falsos negativos, enquanto a taxa de falso positivo é definida como a quantidade de falsos positivos dividida pela soma dos falsos negativos com os negativos verdadeiros. Analisando apenas as curvas com múltiplos conjuntos de dados pode tornar-se difícil, para isso a métrica área debaixo da curva (AUC) também foi calculada. Também conhecida como ROC – AUC, os seus valores podem variar de 0,0 a 1,0, significando a probabilidade do classificador prever uma instância positiva verdadeira em vez de uma positiva falsa.

Podemos definir o conceito de precision e de recall. Precision é definido como a quantidade de verdadeiros positivos dividido pela soma dos verdadeiros positivos com os falsos negativos, quantificando o número de predições positivas no total, este valor varia de 0,0 a 1,0. Recall é definido como a quantidade de verdadeiros positivos dividido pela soma dos verdadeiros positivos com os negativos falsos, quantificando a quantidade de verdadeiros positivos que deveriam ser preditos, variando de 0,0 a 1,0 também. Com a ideia de Precision e Recall podemos visualizar graficamente outro aspecto da performance do classificador, considerada mais realista comparada a curva ROC, varia de 0,0 a 1,0 e, para conjuntos de dados balanceados, é dito ruim um classificador constituinte de uma horizontal no 0,5 significando que a precisão é proporcional ao número de dados positivos no dataset e 1,0 um classificador perfeito. Pode-se interpretar também a AUC da curva precision-recall semelhantemente a AUC da curva ROC, onde um classificador perfeito possui 1,0 de cobertura do gráfico, podendo variar até 0,0.

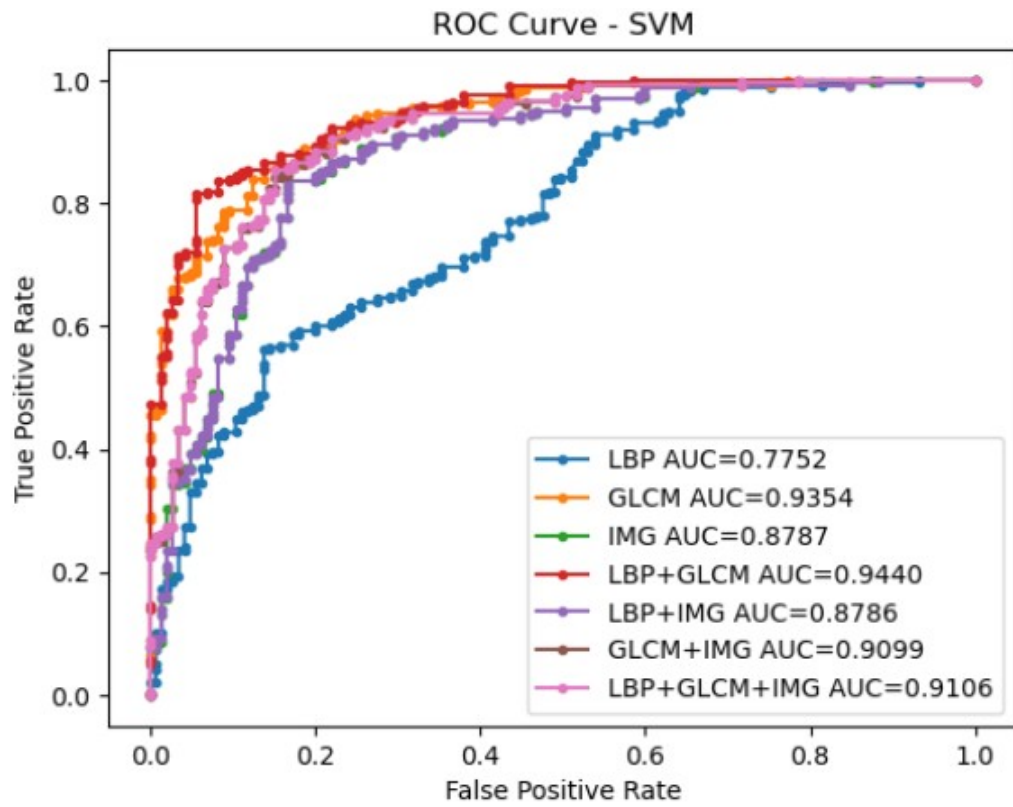


Figura 1 – Curva ROC para os conjuntos de dados testados com o classificador SVM.

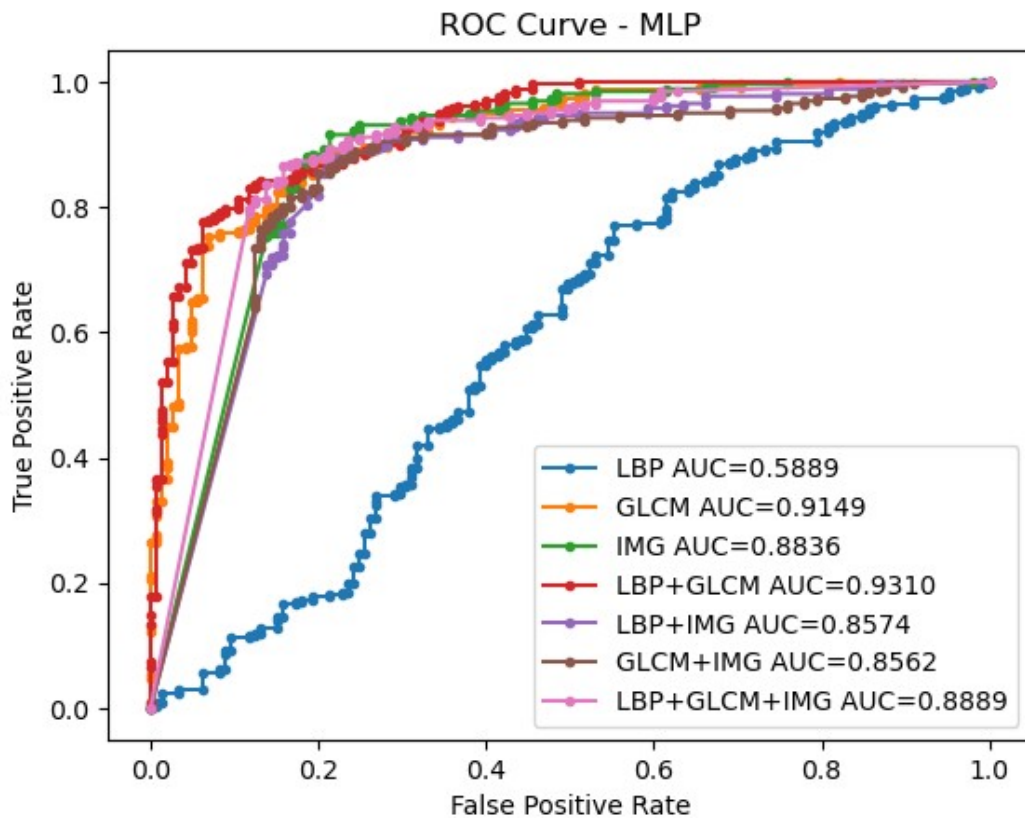


Figura 2 – Curva ROC para os conjuntos de dados testados com o classificador MLP.

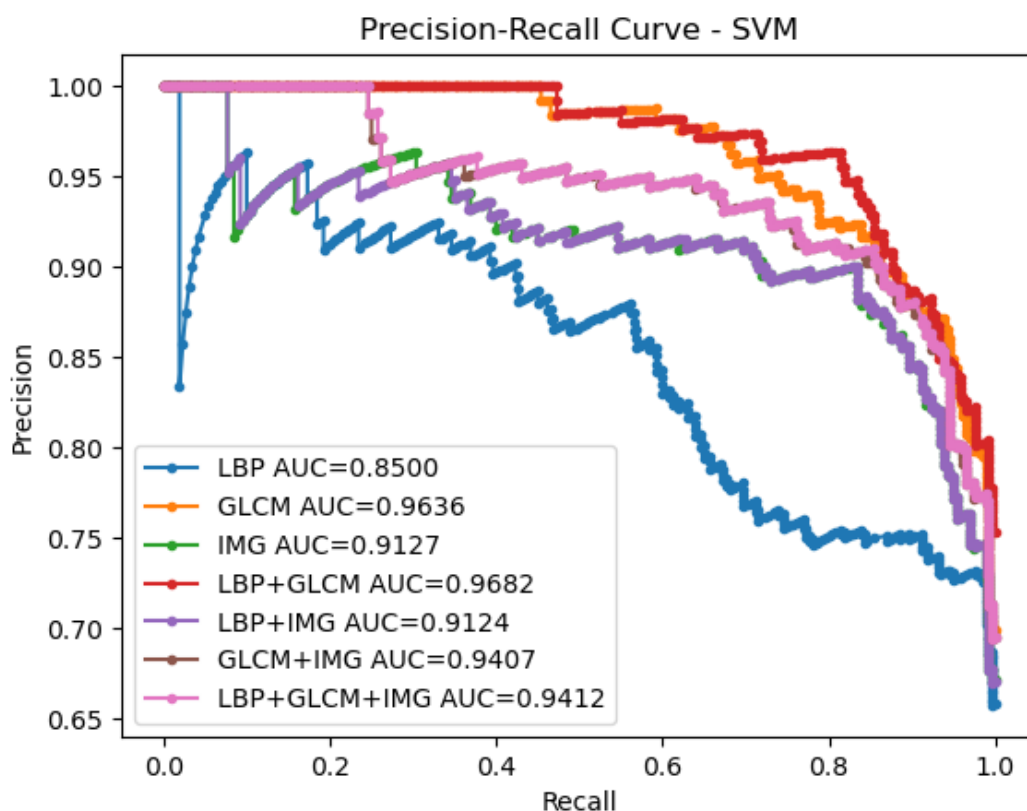


Figura 3 – Curva Precision-Recall para o classificador SVM.

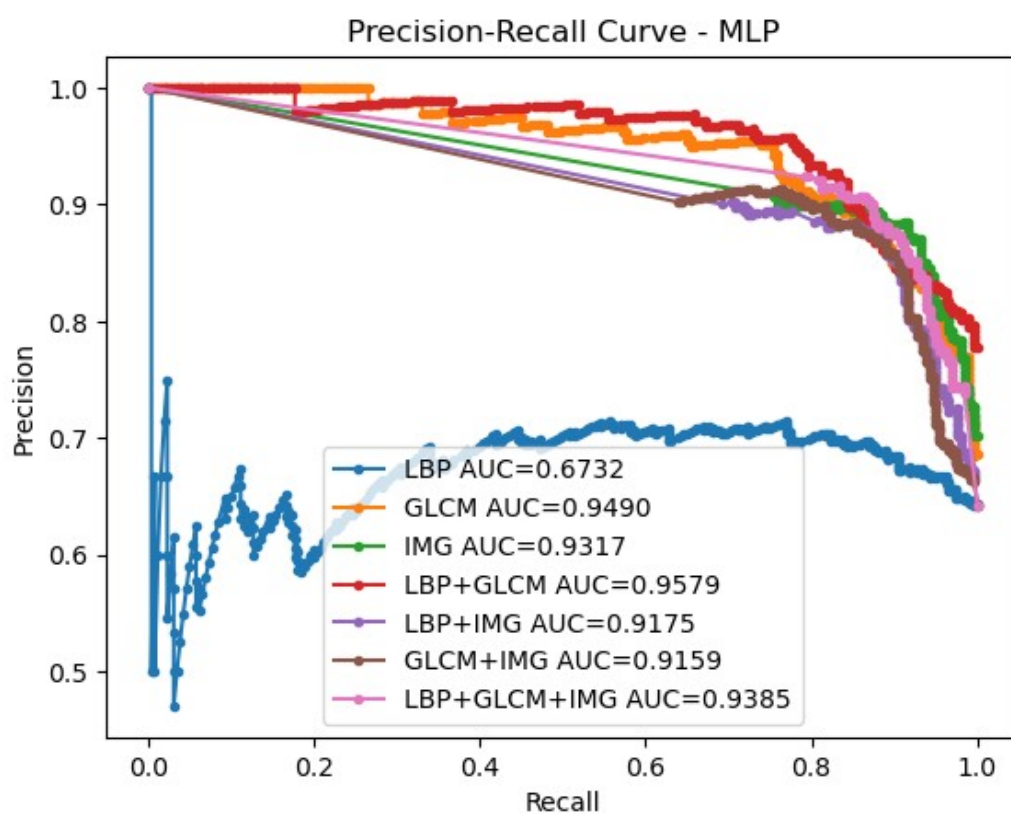


Figura 3 – Curva Precision-Recall para o classificador SVM.

Em ambos os casos o conjunto de dados composto pelas features LBP + GLCM performaram melhor que os demais, evidenciando a observação da prática anterior, onde apenas a acurácia média foi levada em consideração. O classificador SVM obteve métricas melhores que o classificador MLP em todos os casos. O conjunto de dados que considerou apenas o descritor LBP foi notavelmente inferior enquanto que o conjunto de dados considerando apenas o descritor GLCM performou quase tão bem quando o melhor conjunto avaliado (GLCM + LBP). Percebe-se que o descritor GLCM – contraste encontra boas features para o problema em questão, apesar dos dois descritores utilizados terem sido de textura.

Tabela 1 – Melhores resultados por configuração de dataset

Modelo x F1-Score Médio	GLCM	LBP	IMG	GLCM + LBP	GLCM + IMG	LBP + IMG	LBP + GLCM + IMG
SVM	0,85	0,72	0,83	0,89	0,84	0,83	0,84
MLP	0,83	0,71	0,83	0,83	0,81	0,77	0,85

É possível analisar a performance do classificador através da métrica F1-Score que é definida como a média harmônica entre o precision e o recall, no caso como são imagens médicas e o conjunto de dados é desbalanceado, avaliar a performance pelo F1-score tende a ser mais esclarecedor, tendo em vista que o modelo pode apenas prever tudo como positivo e obter uma boa acurácia devido a grande quantidade de dados positivos por exemplo. Na métrica de F1-score o conjunto de dados que obteve a melhor avaliação no classificador foi o GLCM + LBP como o esperado, entretanto para o classificador MLP o conjunto da dados LBP + GLCM + IMG obteve um F1-score melhor, generalizando melhor a sua classificação, um desempenho não tão esperado em razão das métricas anteriores.

Todos os códigos desenvolvidos estão disponíveis no repositório público: <https://github.com/Bruno-Mascarenhas/topicos-visao-computacional-1>.

1.4 Referências

- Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine learning*, 45(2), 171-186.
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145-1159.