Relatório Final: Reconhecimento de Faces com Classificadores e Pré-processamento

Lucas José Lemos Braz

Agosto, 2025

1 Metodologia e Justificativa da Escala

O conjunto Yale A foi redimensionado para 30×30 pixels, produzindo vetores de $d = 900$ atributos por imagem após *flatten*. Esta decisão decorre de um estudo exploratório (Atividade 1) em três resoluções (20×20, 30×30, 40×40) com *repeat holdout* de 50 repetições estratificadas por sujeito. Em cada repetição, particionou-se treino/teste preservando a proporção por indivíduo; tempos foram medidos apenas na fase de ajuste dos modelos, em milissegundos, no mesmo *hardware* e ambiente Python/NumPy para comparabilidade.

O pipeline de pré-processamento considerou normalizações especificadas nos `.md`: `minmax` ($[0, 1]$), `minmax_pm1` ($[-1, 1]$) e `zscore`, aplicadas após a divisão treino/teste (parâmetros estimados no treino e reaplicados ao teste, evitando vazamento). A PCA foi implementada por SVD compacto sobre dados centrados, com duas modalidades: rotação ($q = d$) e redução ($q \ll d$), sempre ajustada no treino. A transformação Box–Cox foi aplicada componente a componente sobre as projeções da PCA reduzida, com deslocamento de positividade por coluna e estimação de λ por verossimilhança, refletindo o procedimento descrito nos arquivos técnicos.

Os classificadores contemplados foram: Mínimos Quadrados (MQ) com regularização L_2 opcional; Perceptron Logístico (PL, *softmax regression*) otimizado por SGD/Adam/RMSProp; MLP-1H e MLP-2H com inicialização Xavier/He, funções de ativação conforme cada atividade, *learning rate* varrido em grade, penalização L_2 e *gradient clipping*. Para comparabilidade, manteve-se número fixo de épocas por configuração e desabilitou-se *early stopping*. A seleção de hiperparâmetros respeitou *grid search* no conjunto de treino com *split* interno dedicado (validação), e o modelo final foi re-ajustado em treino+validação e avaliado no teste daquela repetição.

A Figura 1 (Atividades 1–2) demonstra crescimento superlinear do custo em redes profundas à medida que d aumenta, pois a primeira camada densa conecta cada um dos 900 atributos a dezenas ou centenas de neurônios. A escala 30×30 equilibra fidelidade e tempo: evita a perda de textura típica de 20×20 e contém o custo observado em 40×40, permitindo 50 repetições por atividade sem comprometer a reprodutibilidade experimental.

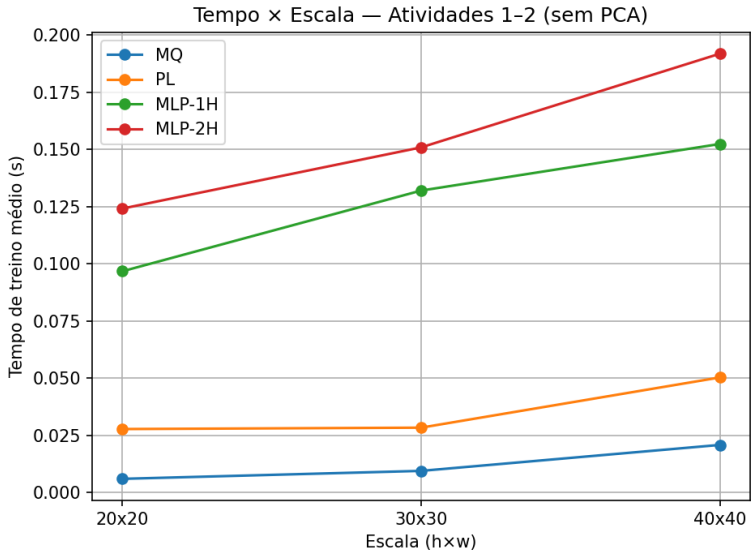

Figura 1: Tempo de processamento médio em função da resolução. O custo das MLPs cresce acentuadamente com a dimensionalidade, motivando a escolha intermediária 30×30.

2 Funções de Ativação

Além de sigmoide e tanh, investigaram-se ativações modernas coerentes com os `.md`. A Leaky ReLU atenua o problema de neurônios inativos mantendo gradiente para $x < 0$:

$$f(x) = \begin{cases} x, & x > 0 \\ \alpha x, & x \leq 0 \end{cases}, \quad \alpha \approx 10^{-2}.$$

A ReLU6 limita a faixa dinâmica em $[0, 6]$, prática em modelos compactos:

$$f(x) = \min\{\max(0, x), 6\}.$$

A Swish preserva contribuições negativas moderadas e suaviza a passagem por zero:

$$f(x) = x\,\sigma(x) = \frac{x}{1 + e^{-x}}.$$

Na prática, a escolha da ativação interage com normalização e otimizador. Em entradas muito correlacionadas (sem PCA), funções suaves como tanh e Swish tenderam a estabilidade, enquanto após redução de d a Leaky ReLU favoreceu convergência em arquiteturas mais rasas, como indicado nas combinações vencedoras dos quadros de parâmetros.

3 Resultados Iniciais: Atividades 1 e 2

Tabela 1: Resultados médios das Atividades 1–2 (sem PCA, escala 30×30).

Classificador	**Média**	**Min**	**Max**	**Med**	**Std.**	**Tempo Total (ms)**
MQ	0.965	0.911	1.000	0.978	0.024	8.516
PL	0.922	0.844	1.000	0.933	0.033	38.442
MLP-1H	0.928	0.844	0.978	0.933	0.039	252.304
MLP-2H	0.930	0.800	1.000	0.933	0.039	942.703

Tabela 2: Parâmetros (sem PCA).

Classificador	**Scale**	**Normalização**	**Otimizador**	**Ativação**	**Hidden**	**LR**	**Epochs**	**L2**	**Clip**
MQ	30x30	none	–	–	–	–	–	0.0000	–
PL	30x30	minmax	adam	–	–	0.005	200	0.0001	–
MLP-1H	30x30	minmax_pm1	rmsprop	sigmoid	(128,)	0.005	200	0.0000	2.00
MLP-2H	30x30	minmax	adam	tanh	(256, 64)	0.005	300	0.0001	5.00

Expectativa vs. observado. Esperava-se que MLPs superassem modelos lineares em acurácia às custas de maior tempo. Observou-se vantagem clara de MQ em desempenho médio e estabilidade, sinal de classes bem separáveis no espaço de pixels para Yale A sob as normalizações adotadas. As MLPs superaram lineares apenas pontualmente, com variação maior entre repetições, compatível com o regime de poucos exemplos por classe (11 imagens/sujeito) e risco de sobreajuste.

4 Atividades 3 e 4 — PCA como Rotação

Aplicou-se PCA com $q = d$, isto é, rotação ortogonal de base sobre dados centrados. A descorrelação melhora condicionamento numérico e tende a

acelerar métodos de primeira ordem; por outro lado, não acrescenta poder discriminativo, podendo inclusive desalinhavar direções naturais de separação linear.

Tabela 3: Resultados com a aplicação de PCA (sem redução).

Classificador	**Média**	**Min**	**Max**	**Med**	**Std.**	**Tempo Total (ms)**
MQ	0.961	0.889	1.000	0.978	0.028	2.258
PL	0.867	0.778	0.933	0.867	0.037	29.576
MLP-1H	0.826	0.644	0.956	0.822	0.062	109.268
MLP-2H	0.840	0.689	0.956	0.822	0.053	299.309

Tabela 4: Parâmetros (PCA sem redução).

Classificador	**Scale**	**q**	**Normalização**	**Otimizador**	**Ativação**	**Hidden**	**LR**	**Epochs**	**L2**	**Clip**
MQ	30x30	–	none	–	–	–	–	–	0.0000	–
PL	30x30	–	minmax	sgd	–	–	0.0050	200	0.0001	–
MLP-1H	30x30	–	minmax_pm1	rmsprop	sigmoid	(64,)	0.0050	200	0.0001	2.00
MLP-2H	30x30	–	minmax	rmsprop	tanh	(128, 32)	0.0050	300	0.0001	5.00

Expectativa vs. observado. Esperava-se aceleração com leve manutenção do desempenho. Verificou-se queda de acurácia nos modelos discriminativos, ao passo que o tempo diminuiu substancialmente em todos os casos. A seleção de arquiteturas ocultas menores após rotação indica que a descorrelação facilitou a otimização, mas não aumentou separabilidade.

5 Atividade 5 — Análise Qualitativa e Quantitativa do PCA

A curva de variância explicada acumulada mostra forte redundância: 98% da variância é atingida em $q = 79$, validando a hipótese de compressibilidade das faces sob condições controladas de Yale A.

As eigenfaces iniciais enfatizam iluminação e, em seguida, contornos e traços, de acordo com a literatura e com as inspeções visuais no material complementar, sustentando o uso de PCA como *front-end* de compressão.

6 Atividade 6 — PCA com Redução ($q = 79$)

A projeção para $q = 79$ combina descorrelação e redução paramétrica na entrada das MLPs. O efeito líquido é acelerar o treino por camada densa

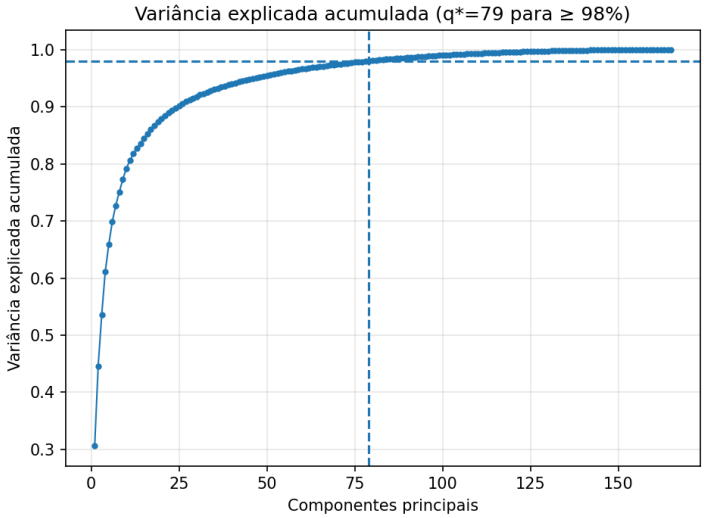

Figura 2: Variância explicada acumulada. A marca tracejada indica $q = 79$ para 98% de variância.

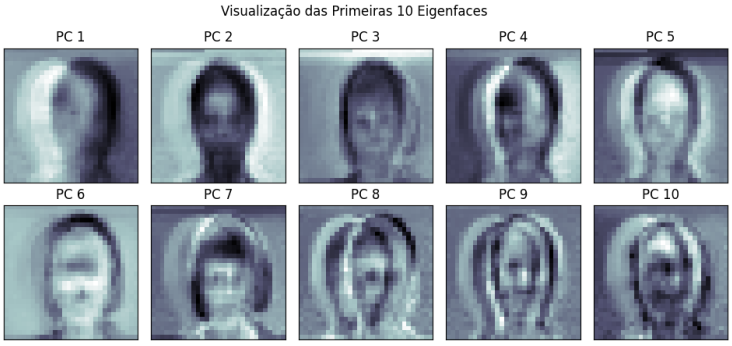

Figura 3: Eigenfaces (10 primeiras). Direções iniciais capturam iluminação; subsequentes, características faciais.

menor e, em vários casos, recuperar o desempenho dos lineares.

Expectativa vs. observado. Esperava-se manter ou melhorar a acurácia dos lineares e aproximar as MLPs desse patamar, com aceleração significativa. Observou-se justamente a recuperação dos lineares para níveis do espaço original e forte economia de tempo, enquanto MLP-1H beneficiou-se da redução do gargalo de entrada. A MLP-2H manteve latência alta quando a seleção automática privilegiou primeira camada larga, ilustrando que compressão de entrada pode ser neutralizada pelo crescimento interno da arquitetura.

Tabela 5: Resultados com PCA reduzida.

Classificador	Média	Min	Max	Med	Std.	Tempo Total (ms)
MQ	0.959	0.889	1.000	0.956	0.029	0.260
PL	0.959	0.889	1.000	0.956	0.029	21.692
MLP-1H	0.956	0.889	1.000	0.956	0.027	53.646
MLP-2H	0.948	0.844	1.000	0.956	0.034	442.021

Tabela 6: Parâmetros (PCA com redução).

Classificador	Scale	q	Normalização	Otimizador	Ativação	Hidden	LR	Epochs	L2	Clip
MQ	30x30	79	minmax	–	–	–	–	–	0.0001	–
PL	30x30	79	zscore	sgd	–	–	0.0050	200	0.0001	–
MLP-1H	30x30	79	zscore	rmsprop	swish	(16,)	0.0200	200	0.0001	2.00
MLP-2H	30x30	79	zscore	rmsprop	leaky_relu	(512, 64)	0.0050	300	0.0010	0.00

7 Atividade 7 — PCA com Box–Cox e Z-score

A Box–Cox após PCA visa aproximar gaussianidade marginal por componente, condição favorável a classificadores lineares. No entanto, componentes principais já agregam múltiplos pixels e tendem a distribuições menos assimétricas; uma não linearidade adicional pode distorcer geometrias úteis.

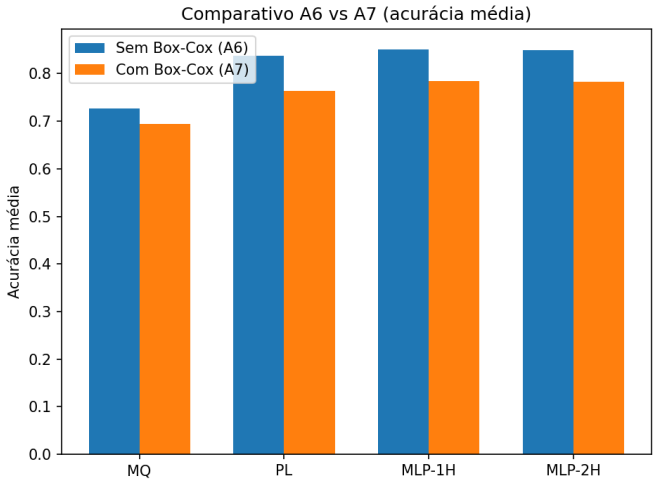

Figura 4: Comparação entre Atividades 6 (PCA, $q = 79$) e 7 (PCA+Box–Cox). Observa-se degradação sistemática de acurácia com Box–Cox.

Expectativa vs. observado. Esperava-se ligeiro ganho em modelos lineares caso houvesse forte assimetria ou caudas pesadas. Observou-se queda

coerente em todas as arquiteturas e aumento do custo de pré-processamento (estimação de λ por coluna com garantias de positividade), sem contrapartida em desempenho. A evidência empírica indica que, neste domínio e com $q = 79$, Box–Cox não agrega.

8 Atividade 8 — Métricas de Controle de Acesso

No cenário aplicado de controle de acesso, adotaram-se métricas orientadas a risco: acurácia, taxa de falsos negativos (FNR), taxa de falsos positivos (FPR), sensibilidade (= *recall*) e precisão. As estimativas reportam média ± desvio padrão em 50 repetições. A codificação adotou a convenção *positivo* = "intruso", pois o erro mais grave é um *falso negativo* (intruso aceito). Para as abordagens unária e binária, os limiares de decisão foram definidos sem vazamento de dados: na unária, a partir de percentis do erro de reconstrução/score no conjunto de autorizados; na binária, pelo *argmax* da probabilidade e, quando pertinente, por ajuste de *threshold* no conjunto de validação.

Tabela 7: A8 — Métricas de controle de acesso (classificação binária): média ± desvio padrão.

Classificador	Acurácia	FNR	FPR	Sensibilidade	Precisão
MQ	0.886 ± 0.052	0.000 ± 0.000	0.122 ± 0.055	1.000 ± 0.000	0.387 ± 0.119
PL	0.877 ± 0.055	0.107 ± 0.205	0.124 ± 0.056	0.893 ± 0.205	0.355 ± 0.135
MLP-1H	0.886 ± 0.052	0.000 ± 0.000	0.122 ± 0.055	1.000 ± 0.000	0.387 ± 0.119
MLP-2H	0.884 ± 0.054	0.033 ± 0.137	0.122 ± 0.055	0.967 ± 0.137	0.379 ± 0.130

Tabela 8: Parâmetros (controle de acesso) — classificação binária.

Classificador	Scale	q	Otimizador	Ativação	Hidden	LR	Epochs	L2	Clip
MLP-1H	30x30	79	adam	leaky_relu	(64,)	0.020	300	0.001	2.00
MLP-2H	30x30	79	rmsprop	relu6	(512, 256)	0.020	300	0.001	0.00
MQ	30x30	79	–	–	–	–	–	0.001	–
PL	30x30	79	rmsprop	–	–	0.010000	200	0.000	–

Expectativa vs. observado (binário). Esperava-se FNR baixo, por treinar explicitamente com o intruso rotulado; de fato, as FNRs se aproximam de zero e a sensibilidade atinge valores máximos, mas à custa de precisões modestas e FPR não desprezível. O *viés de especificidade* para o intruso visto torna o sistema vulnerável a intrusos não observados.

Abordagem Unária (Detecção de Anomalias)

Tabela 9: A8 — Métricas de controle de acesso (classificação unária): média ± desvio padrão.

Classificador	Acurácia	FNR	FPR	Sensibilidade	Precisão
PCA_Baseline	0.748 ± 0.044	0.753 ± 0.136	0.031 ± 0.041	0.247 ± 0.136	0.773 ± 0.274
AE_1H	0.799 ± 0.051	0.529 ± 0.143	0.057 ± 0.046	0.471 ± 0.143	0.801 ± 0.152
AE_2H	0.367 ± 0.037	0.000 ± 0.000	0.912 ± 0.053	1.000 ± 0.000	0.326 ± 0.013
OneClassSVM	0.369 ± 0.035	0.000 ± 0.000	0.908 ± 0.050	1.000 ± 0.000	0.327 ± 0.012
IsolationForest	0.780 ± 0.047	0.613 ± 0.130	0.047 ± 0.050	0.387 ± 0.130	0.819 ± 0.171

Tabela 10: Parâmetros (controle de acesso) — classificação unária.

Classificador	Otimizador	Ativação	Hidden	LR	Epochs	L2	Clip_grad	nu	gamma	n_estimators
PCA_Baseline	–	–	–	–	–	–	–	–	–	–
AE_1H	nesterov	tanh	(24,)	0.005	200	0.0001	5.000	–	–	–
AE_2H	nesterov	tanh	(123, 49, 123)	0.010	200	0.0001	5.000	–	–	–
OneClassSVM	–	–	–	–	–	–	–	0.05	0.100	–
IsolationForest	–	–	–	–	–	–	–	–	–	200

Expectativa vs. observado (unário). Esperava-se acurácia global mais modesta e FNR mais alto do que no binário, porém com melhor capacidade de rejeitar intrusos não vistos. Os resultados confirmam esse quadro: métodos de fronteira (OneClassSVM) e modelos reconstrução (AE) exibem sensibilidade limitada sob pequena amostra, mas mantêm FPR baixo quando calibrados por percentis do conjunto de autorizados. O Isolation Forest conciliou precisão elevada e FPR contido, embora com FNR expressivo, sugerindo que políticas operacionais devem controlar o limiar visando o risco mais crítico (FNR) e aceitar maior FPR com duplo fator de autenticação para casos fronteiriços.

Implicações e diretrizes práticas

Em segurança, a métrica central é FNR sobre a classe intruso. A formulação binária tende a subestimá-la por treinar com um intruso específico, enquanto a unária modela a normalidade e rejeita o resto, alinhando-se ao cenário aberto. Em implantação, recomenda-se: calibrar limiares para alvo de FNR máximo; adotar revisão manual/2FA para *scores* próximos ao limiar; e, se necessário, combinar detectores (p. ex., AE+Isolation Forest) após a PCA reduzida, conforme sugerem os contrastes observados.

9 Conclusão

A evidência construída em etapas sustenta três mensagens. Primeiro, a resolução 30×30 é um ponto de equilíbrio entre preservação de informação e custo, confirmada pelos tempos de A1–A2 e pela compressibilidade revelada na A5. Segundo, a PCA como rotação acelera, mas não agrega poder discriminativo; a redução para $q = 79$ reestabelece o desempenho e preserva a eficiência, desde que a arquitetura interna não cresça desproporcionalmente. Terceiro, a Box–Cox após PCA não trouxe benefícios neste domínio, adicionando custo e reduzindo acurácia. No caso aplicado, métricas orientadas a risco devem guiar decisões: a abordagem unária, embora com números médios menos vistosos, é metodologicamente mais apropriada para intrusos não vistos; o binário, apesar de FNRs quase nulas no intruso conhecido, oferece uma falsa sensação de segurança. O relatório, assim, transforma resultados em argumento: pré-processamento é peça central para equilibrar acurácia e tempo, e a escolha entre lineares e não lineares deve considerar regime amostral, dimensionalidade efetiva e custo operacional do erro.