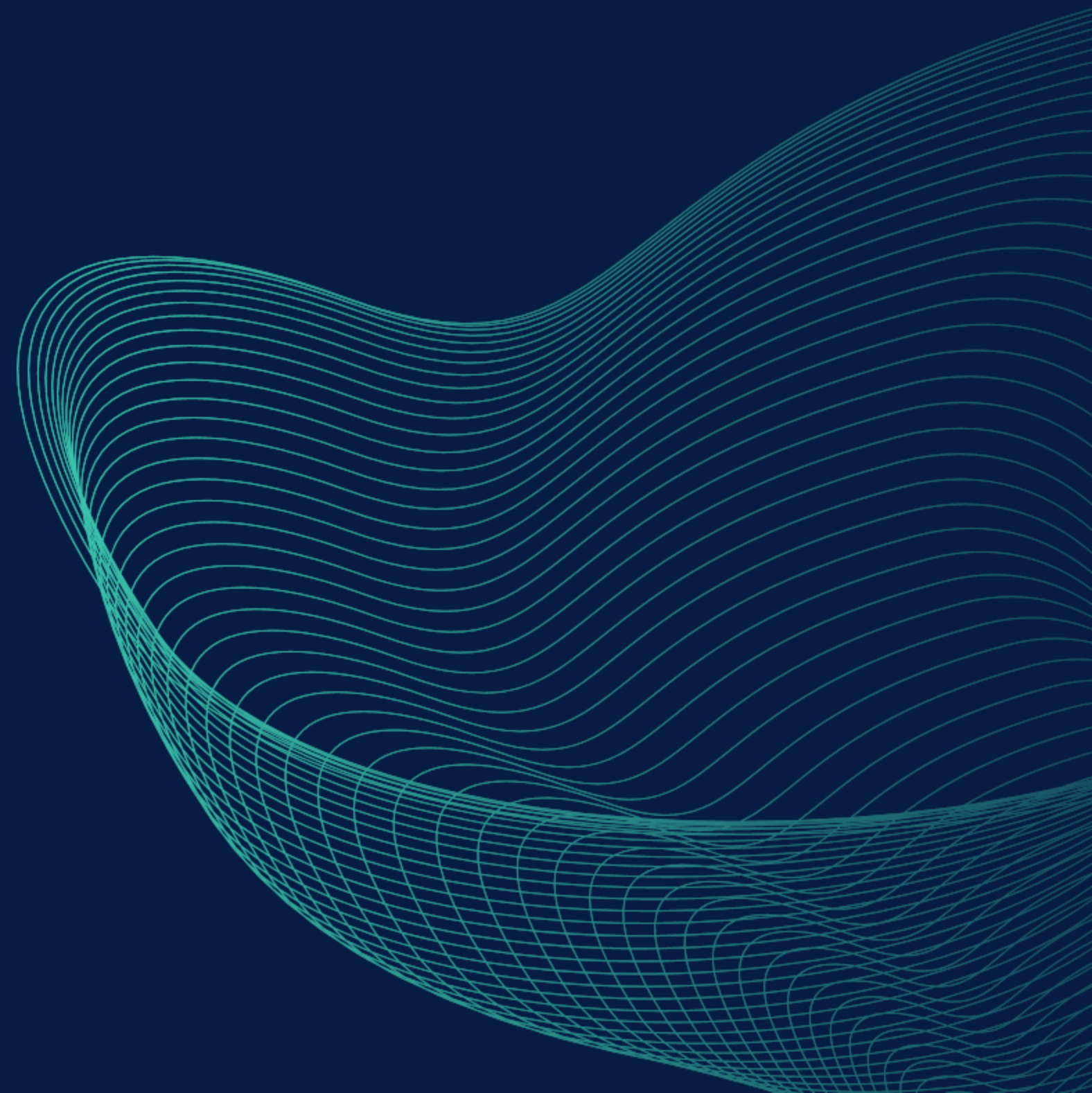


Projeto Final – Tópicos Avançados em Controle Processos de Sinais IV

**Estimativa da geração instantânea (nowcasting) de painéis
solares a partir de imagens do céu**

Gustavo Friol Bento
Prof. Danilo Silva



Introdução

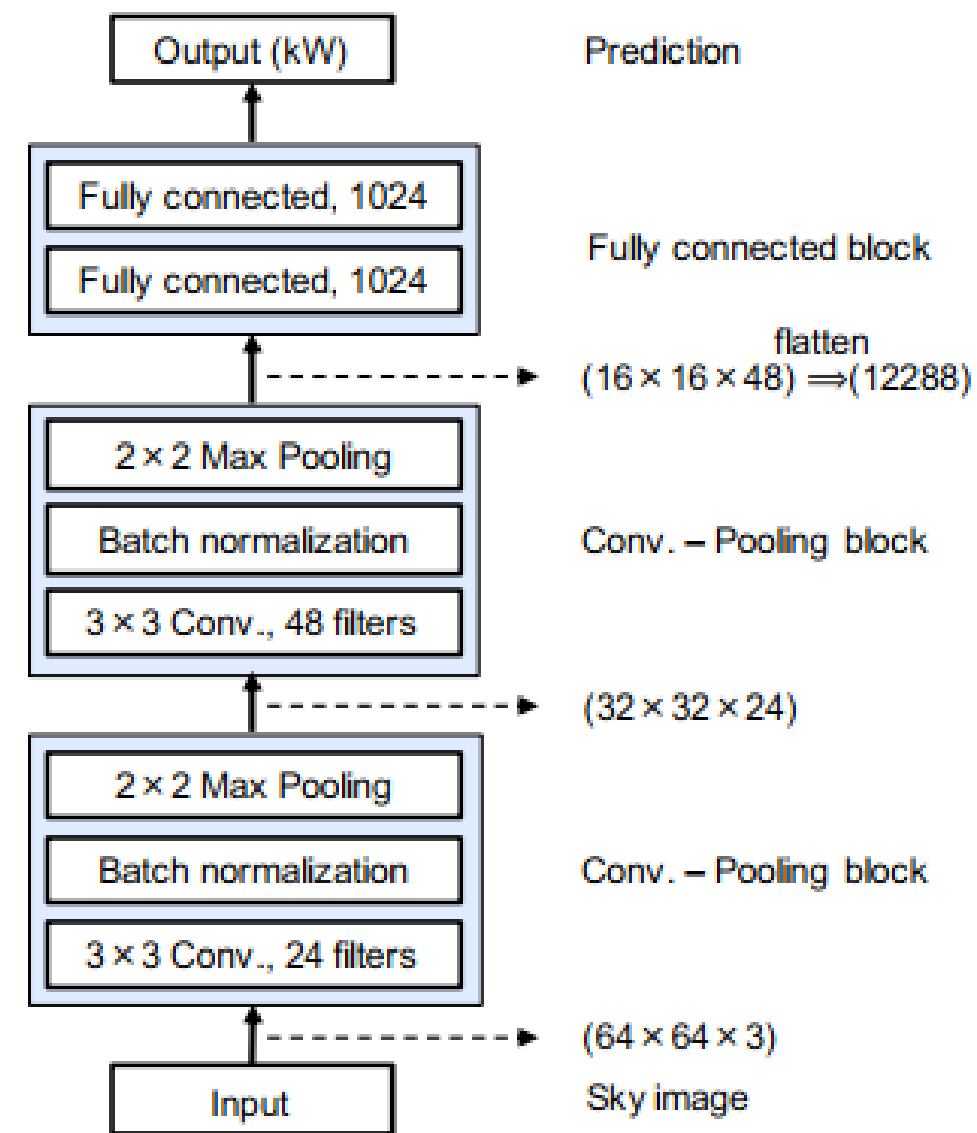
Motivação

- O Problema: A previsão de geração solar a curto prazo é extremamente desafiadora devido à volatilidade das condições meteorológicas, principalmente a cobertura de nuvens.
- O Impacto: Mudanças repentinas na nebulosidade podem reduzir a potência gerada a quase zero em minutos, o que representa um obstáculo para a integração da energia solar nas redes elétricas, dentre outros motivos.
- A Oportunidade: Imagens do céu são uma fonte rica de informação, contendo dados sobre a posição do sol, bem como a cor e transparência, que se correlacionam com a geração de energia.

Trabalhos Relacionados

O Modelo SUNSET

- O "Nowcast" (Previsão Instantânea): Em 2018, o grupo de pesquisa Environmental Assessment and Optimization (EAO) na Universidade de Stanford propôs uma CNN para a tarefa de nowcasting, que consiste em estimar a potência gerada no mesmo instante em que a imagem do céu é capturada.
- A Arquitetura: O modelo SUNSET utiliza uma estrutura clássica de CNN, com blocos de Convolução-Pooling que processam a imagem, seguidos por camadas totalmente conectadas que realizam a regressão para prever a potência em kW.



SUNSET – Nowcast

Dois blocos Conv-Pool seguidos de um bloco com duas camadas totalmente conectadas.

Conjunto de Dados

SKIPP'D – SKy Images and Photovoltaic Power Generation Dataset

- Dataset público e padronizado, criado pela Universidade de Stanford para facilitar a pesquisa e o benchmarking de modelos de previsão de energia solar baseados em imagens.
- Os dados foram coletados entre março de 2017 e dezembro de 2019, na própria universidade.
- Possui duas versões:
 - Dados brutos: Imagens de alta resolução (2048x2048).
 - Benchmark: Imagens redimensionadas (64x64).

Conjunto de Dados

SKIPP'D – Benchmark

- Conteúdo: Pares de imagens do céu e os respectivos dados de geração fotovoltaica (kW).
- As amostras são registradas com uma frequência de 1 minuto.
- Imagens: 64x64 pixels, redimensionadas a partir de imagens de alta resolução.
 - Capturadas por uma câmera fish-eye de 360 graus.
- Dados de Geração:
 - Os dados foram coletados de um sistema de painéis solares com capacidade de 30.1 kW-DC.
 - Os painéis estão localizados a aproximadamente 125 metros da câmera.

Conjunto de Dados

Pré-processamento

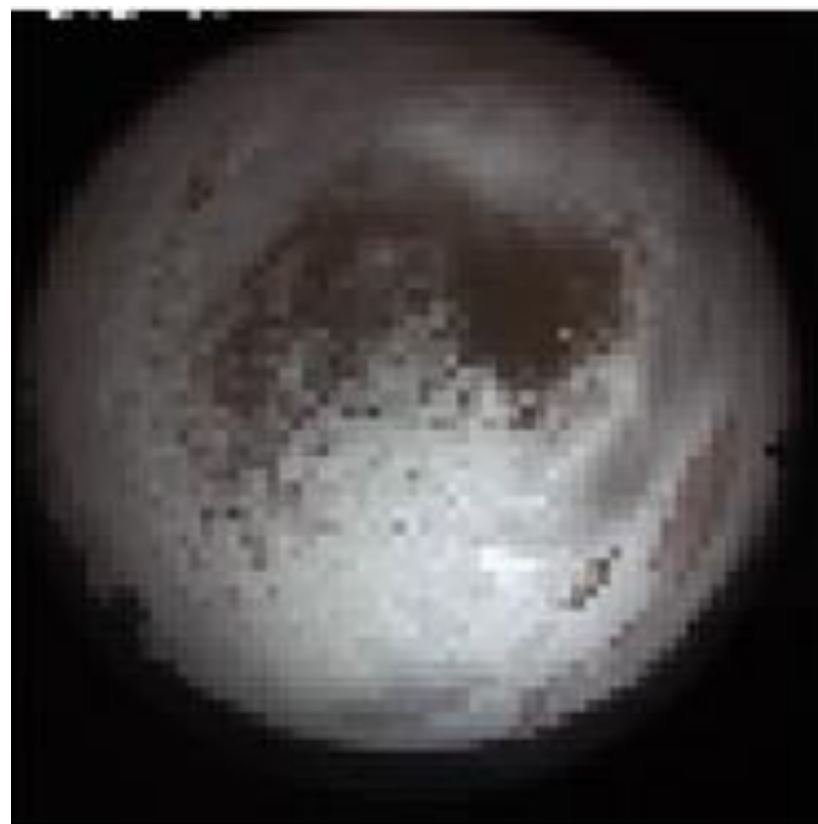
- Imagens:
 - Extraídas dos vídeos de alta resolução da câmera com uma frequência de 1 minuto.
 - Redimensionadas para 64x64.
- Dados de Geração:
 - Filtrados registros anormais (valores negativos ou falhas longas).

Ao final, as imagens processadas foram pareadas com os dados de geração correspondentes ao mesmo instante.

Conjunto de Dados

Particionamento

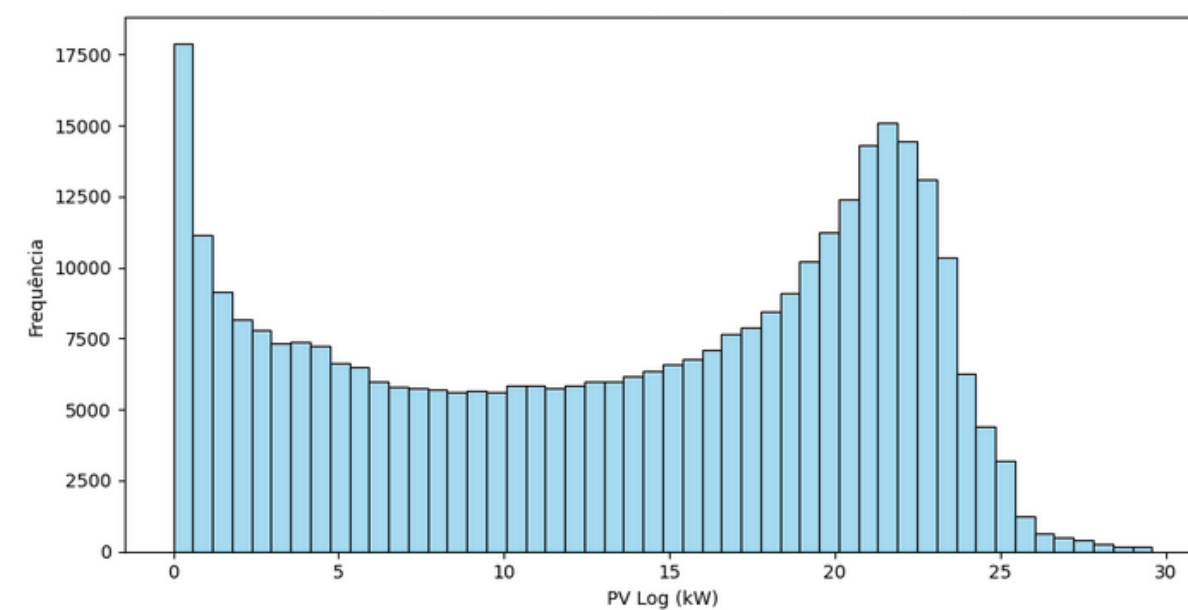
- O dataset é particionado em um conjunto de Desenvolvimento (para treino e validação) e um conjunto de Teste.
 - Critério do Teste: O conjunto de teste foi criado a partir da seleção manual de 10 dias ensolarados e 10 dias nublados, distribuídos ao longo dos anos.
- Tamanho:
 - Desenvolvimento: 349.372 amostras (~96% do total).
 - Teste: 14.003 amostras (~4% do total).



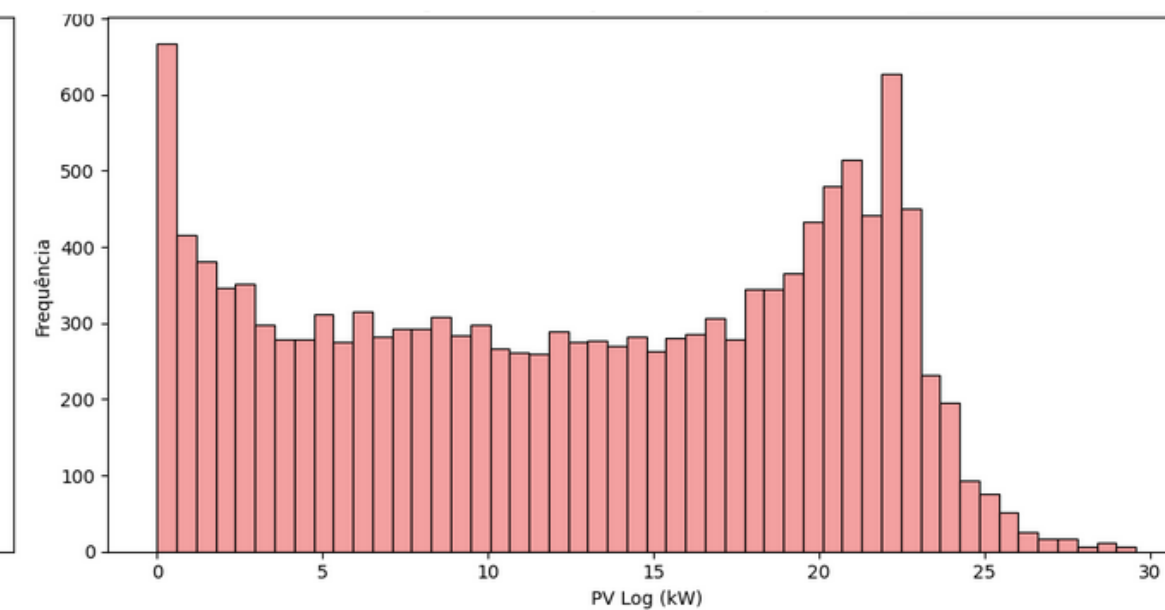
2018-10-01

15:12:30

11.32
kW



Desenvolvimento



Teste

SKIPP'D Benchmark

Amostra e histogramas da geração FV nos conjuntos de desenvolvimento e validação.

Metodologia Experimental

Geral

- Objetivo: O projeto compara o desempenho de duas arquiteturas de CNN: o modelo de referência SUNSET e um modelo que usa como backbone a arquitetura da ResNet50V2.
- Otimizador e Perda: Ambos os modelos foram treinados utilizando o otimizador Adam, com o Erro Quadrático Médio (MSE) como a função de perda a ser minimizada.
- Métricas: RMSE e MAE.
- Validação: A metodologia principal foi a Group K-Fold, agrupando as amostras por dia.
 - Objetivo: Prevenir o vazamento de dados entre os conjuntos de treino e validação, garantindo uma estimativa de desempenho mais robusta.

Metodologia Experimental

SUNSET

- Otimização de Hiperparâmetros: Foi realizada um grid search para encontrar a melhor combinação de taxa de aprendizado e tamanho de lote batch size. Foi utilizada validação k-fold com apenas 2 folds.
- Treinamento Final: Após a seleção dos melhores hiperparâmetros, o modelo foi submetido a um treinamento final com validação cruzada de **10 folds**, alinhando-se à metodologia dos trabalhos de referência.
- Resultados no Teste: A predição final no conjunto de teste consiste na **média das predições dos 10 submodelos** gerados, um para cada fold.

Metodologia Experimental

ResNet50V2

Configuração da Instanciação:

- **Formato de Entrada:** O modelo foi configurado para aceitar imagens de 64x64x3
- **Camadas de Topo:** As camadas de classificação originais foram removidas (`include_top=False`).
- **Pooling de Saída:** Foi adicionada uma camada de **Pooling Médio Global** (`pooling='avg'`) ao final da base convolucional para vetorizar os mapas de features.

Camadas Adicionais e Pré-processamento:

- **Pré-processamento:** O modelo incorpora a função `preprocess_input`, que é específica da arquitetura ResNetV2, para normalizar corretamente as imagens de entrada.
- **Regularização:** Uma camada de **Dropout** foi inserida antes da camada de saída para mitigar o overfitting.
- **Camada de Saída:** Uma camada Densa (`Dense`) final foi utilizada para produzir a predição de regressão.

Metodologia Experimental

ResNet50V2

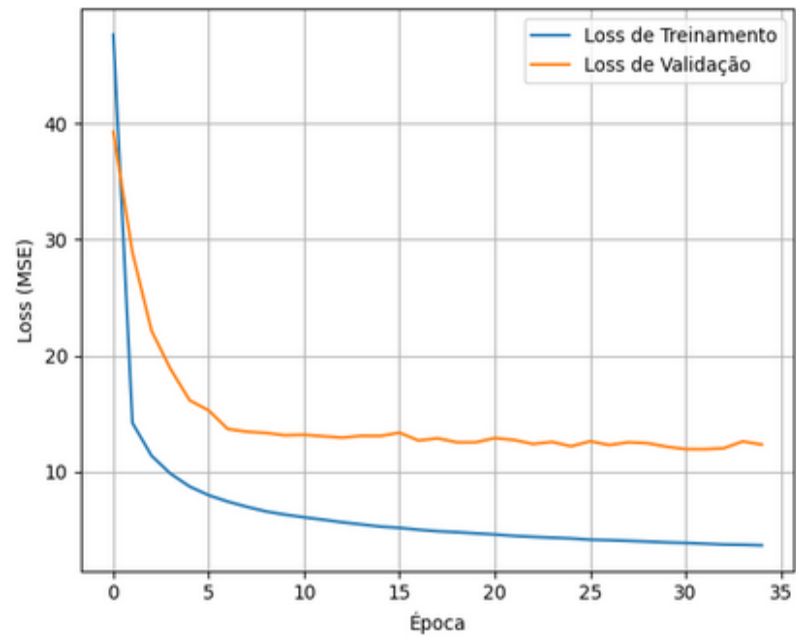
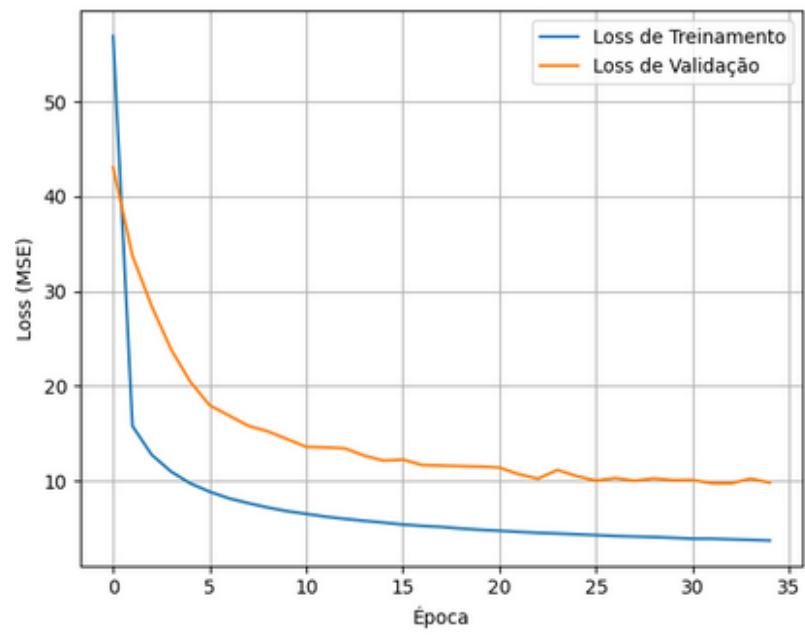
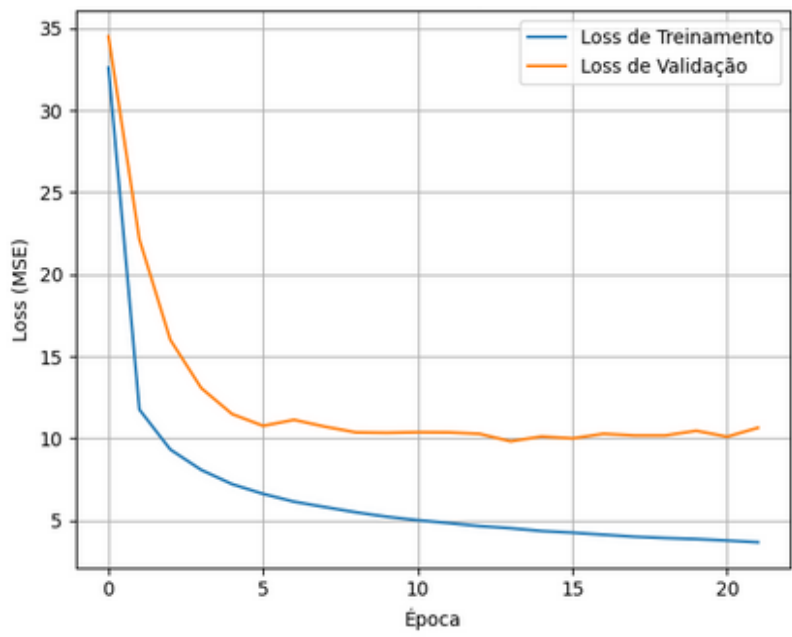
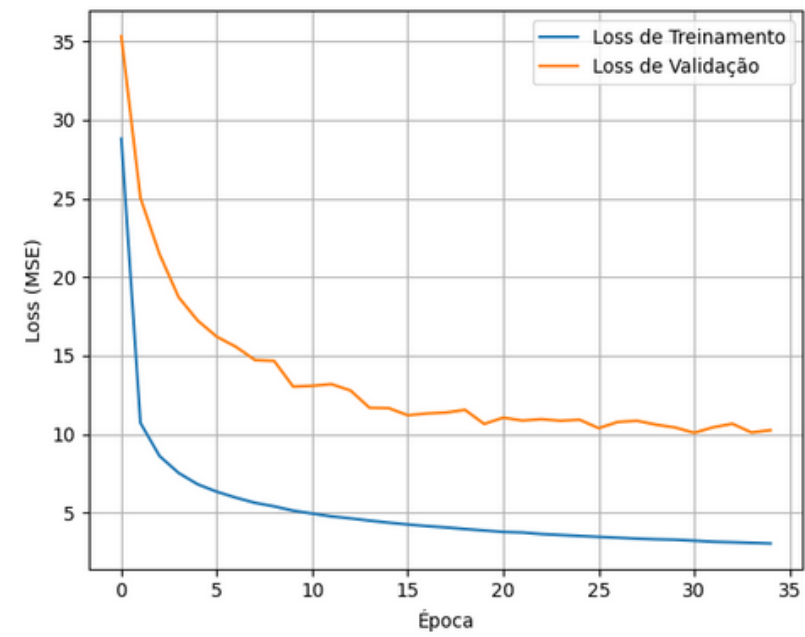
- **Estratégia de Treinamento:** Adotou-se o *fine-tuning* direto, onde os pesos da base convolucional foram descongelados, com exceção das camadas de Batch Normalization.
- **Validação Adaptada:** Devido ao maior custo computacional (tempo por época >2x maior que o SUNSET), a validação cruzada para este modelo foi realizada com **5 folds**.
- **Justificativa:** A escolha por 5 folds representa um *trade-off* para viabilizar o experimento, mantendo a robustez ao treinar cada submodelo com 80% dos dados.

Resultados

SUNSET – Otimização de Hiperparâmetros

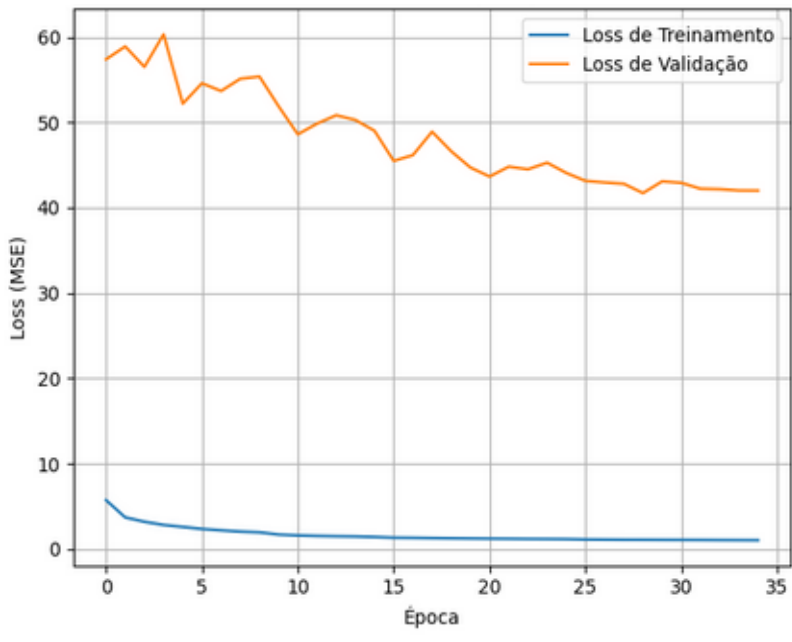
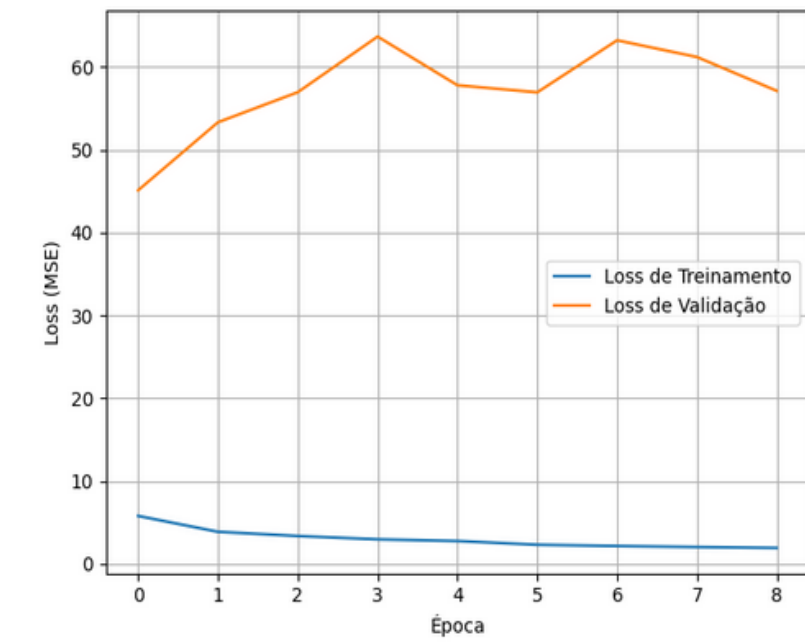
Foi realizado um grid search para encontrar a melhor combinação de taxa de aprendizado e tamanho de lote batch size. Foi utilizada validação k-fold com apenas 2 folds.

Taxa de Aprendizado	Batch Size	Perda Média (MSE)
1e-06	128	9.9525
1e-06	256	10.8420
1e-05	256	14.1517
1e-05	128	16.2486
0.001	32	25.9229
0.001	64	30.3609
0.0001	32	34.1772
0.0001	64	43.4011

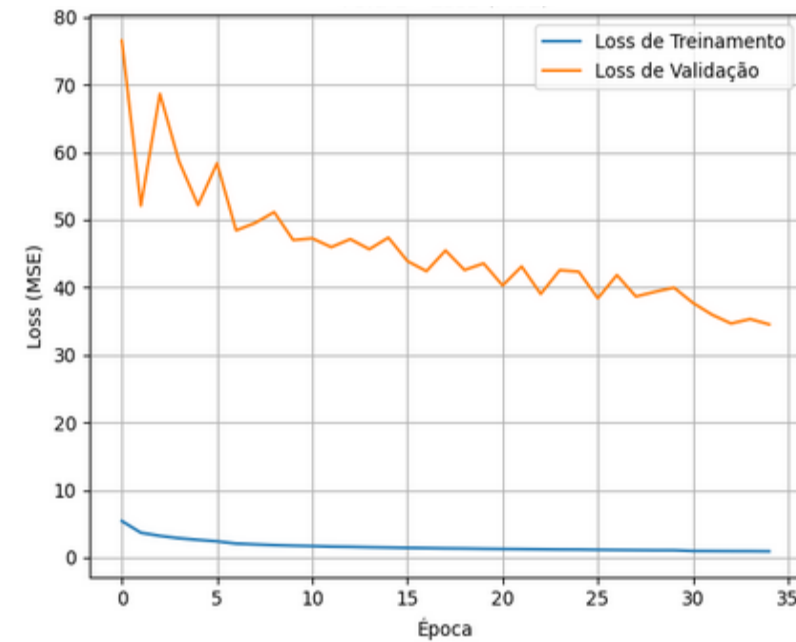
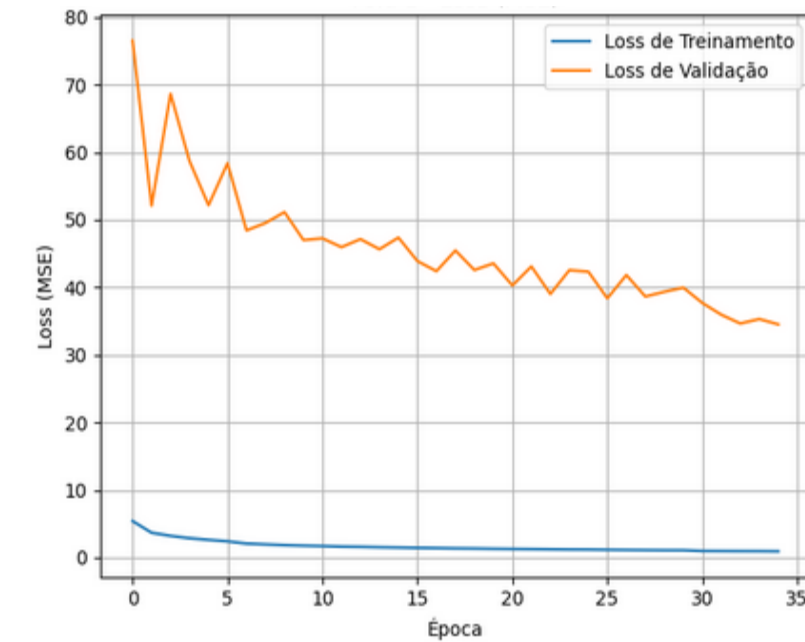


1e-6 - 128

1e-6 - 256



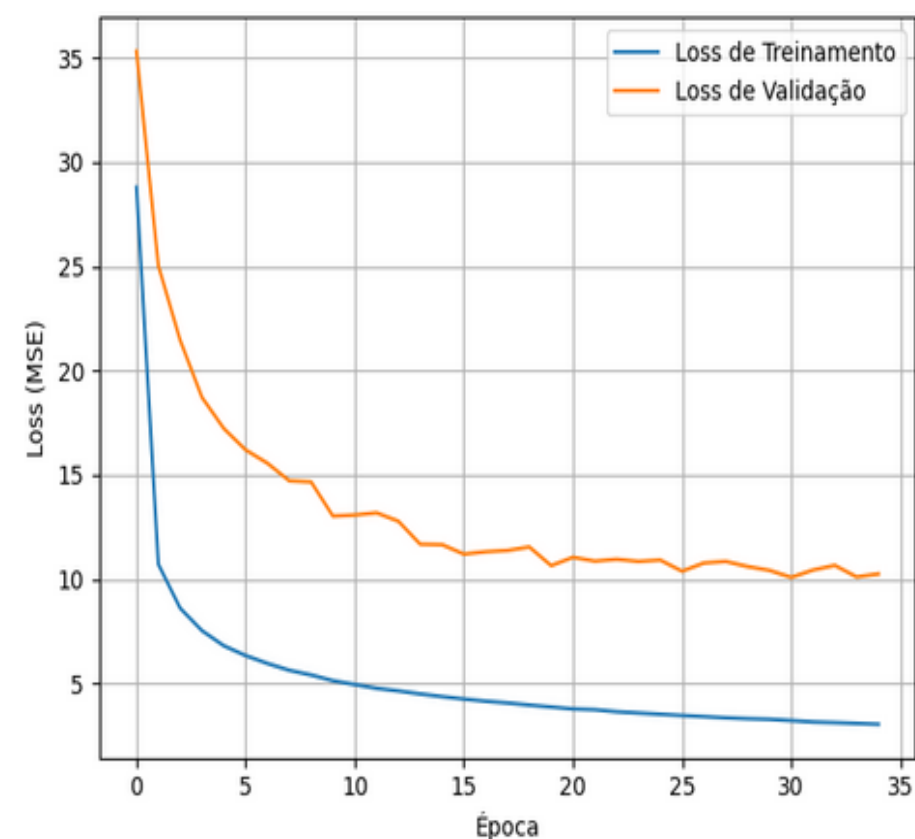
1e-4 - 64



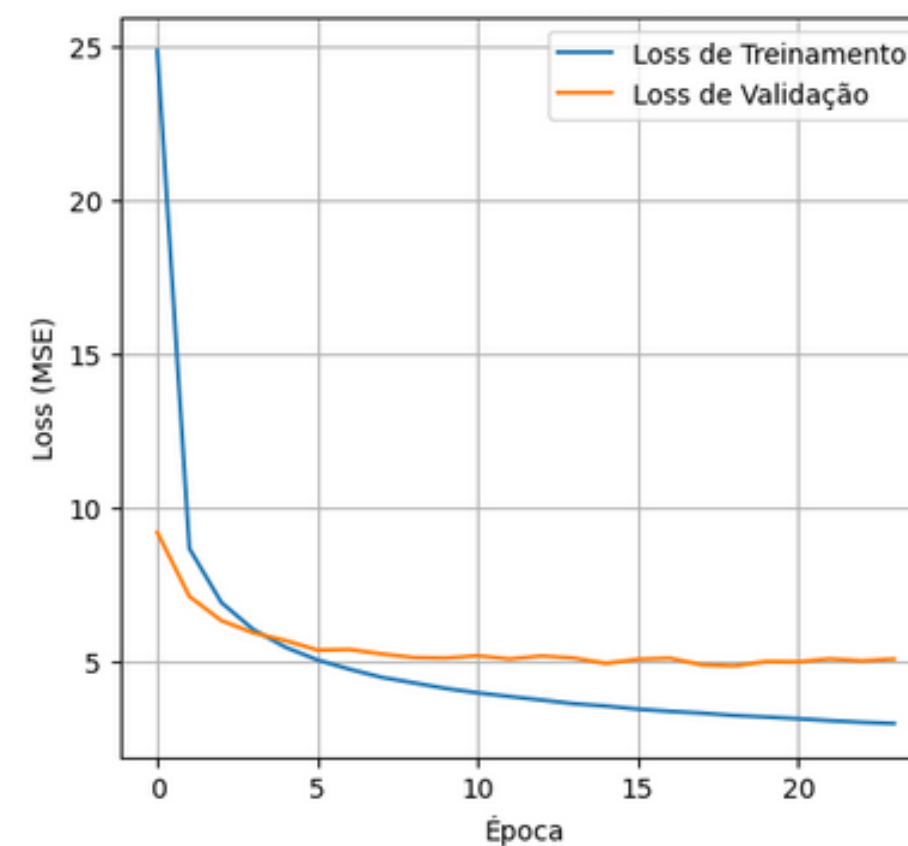
1e-4 - 32

Curvas de Treinamento

Overfitting controlado: otimização da função para criação do dataset.



Antes: Embaralhamento parcial. Buffer size limitado em 5000 amostras.



Agora: Embaralhamento completo do conjunto de treino antes de separá-lo em batches para o treinamento.

Resultados

Controle do overfitting

Resultados

SUNSET – Treinamento Final

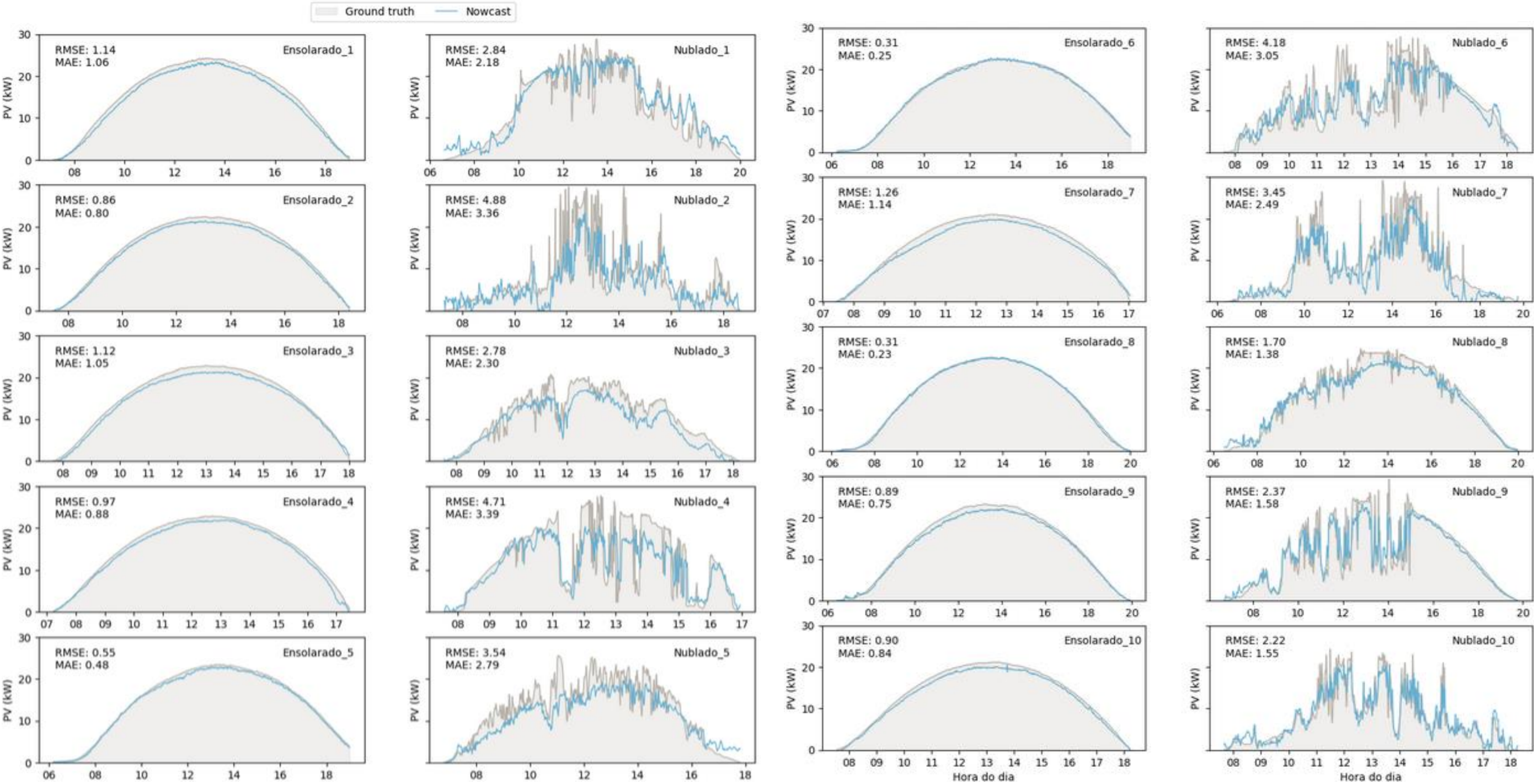
- Limitado em 50 épocas.
- Earlystopping com paciência de 5 épocas.
- ReduceLROnPlateau com paciência de 3 épocas.

Fold	RMSE Treino	RMSE Validação	Melhor Época
1	1.9345	2.3971	12
2	2.0387	1.7304	11
3	2.1941	1.7375	8
4	1.8073	2.2078	19
5	2.0419	1.7398	10
6	2.0169	1.7462	12
7	1.9797	2.0771	12
8	2.0948	2.1477	8
9	2.1608	2.0737	8
10	2.2682	1.7404	6
Média	2.0537	1.9598	–

Desempenho do modelo nos 20 dias de teste

Condição do Céu	RMSE	MAE
Dias Ensolarados	0.865	0.718
Dias Nublados	3.356	2.361
Geral	2.453	1.541

- Bom desempenho para dias ensolarados.
- RMSE aproximadamente 4 vezes maior para dias nublados.
- Captura a tendência geral da curva.
- Tende a suavizar as flutuações.

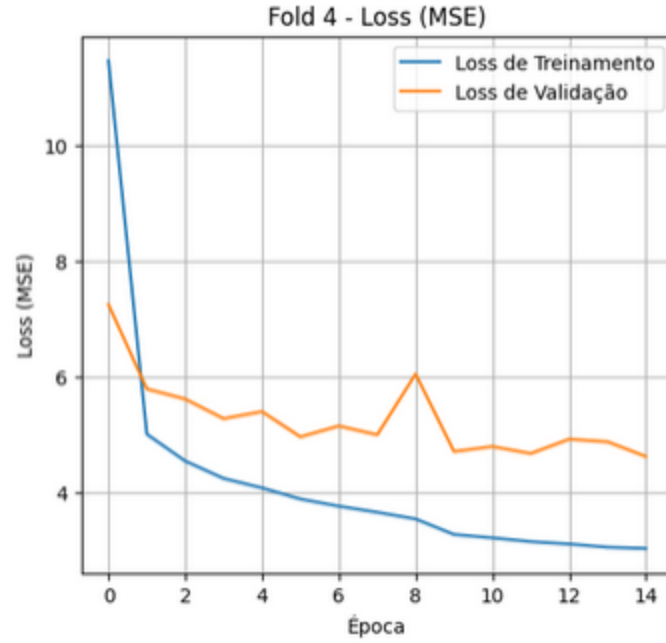
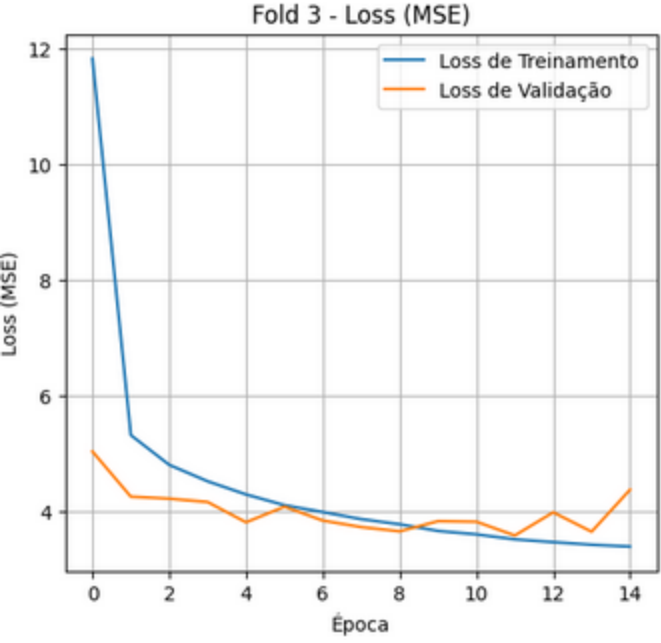
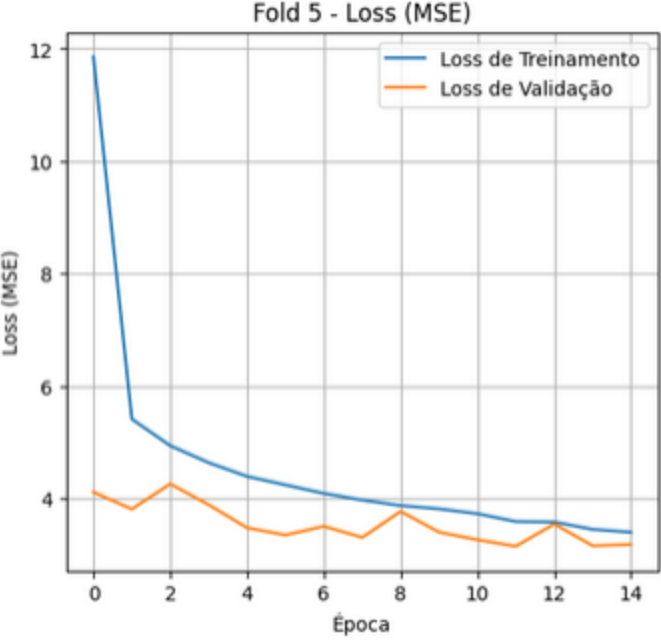
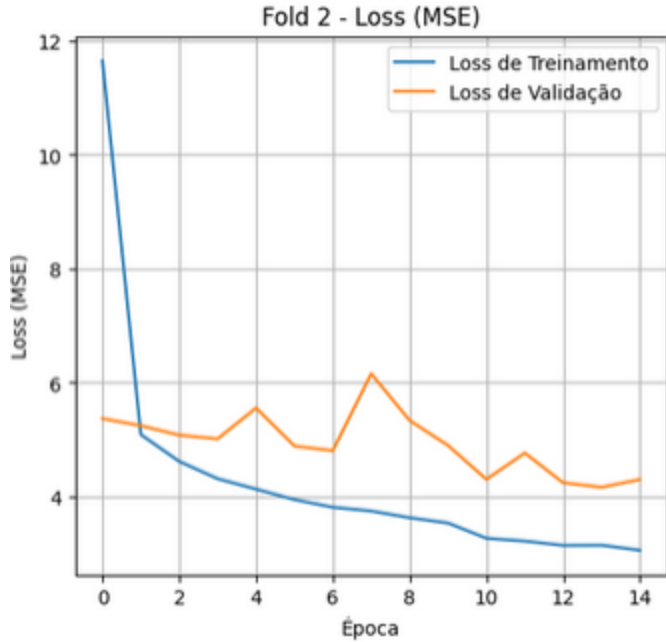
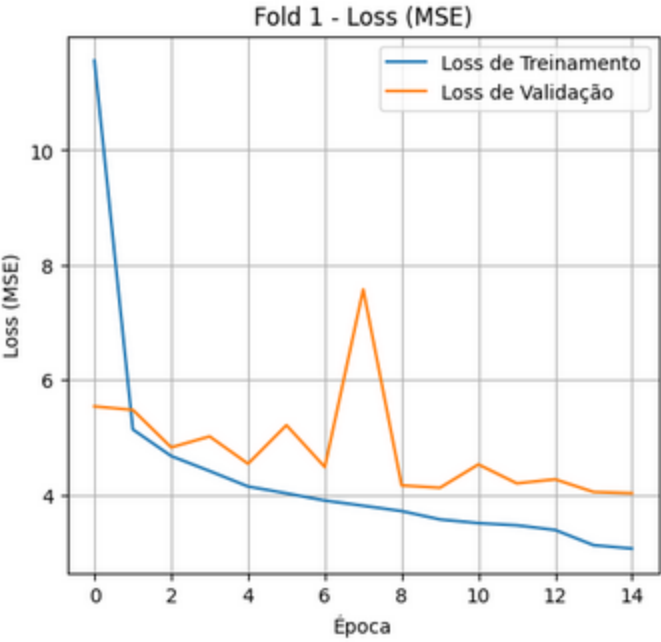


Resultados

ResNet50V2 – Primeira Tentativa (Taxa de Aprendizado: 1e-06):

- Limitado a 15 épocas.
- Earlystopping com paciência de 5 épocas.
- ReduceLROnPlateau com paciência de 3 épocas.

EarlyStopping não foi acionado, e as melhores épocas foram próximas do final, indicando que o modelo poderia se beneficiar de mais treinamento. Além disso, apresentou certa oscilação da perda de validação.



Fold	RMSE Treino	RMSE Validação	Melhor Época
1	1.7500	2.0064	15
2	1.7745	2.0414	14
3	1.8762	1.8940	12
4	1.7377	2.1477	15
5	1.8940	1.7740	12
Média Final	1.8065	1.9727	—

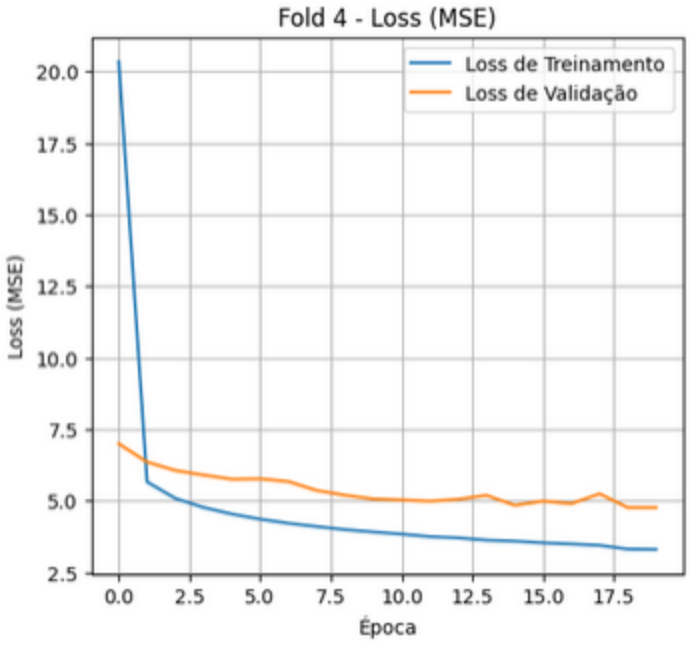
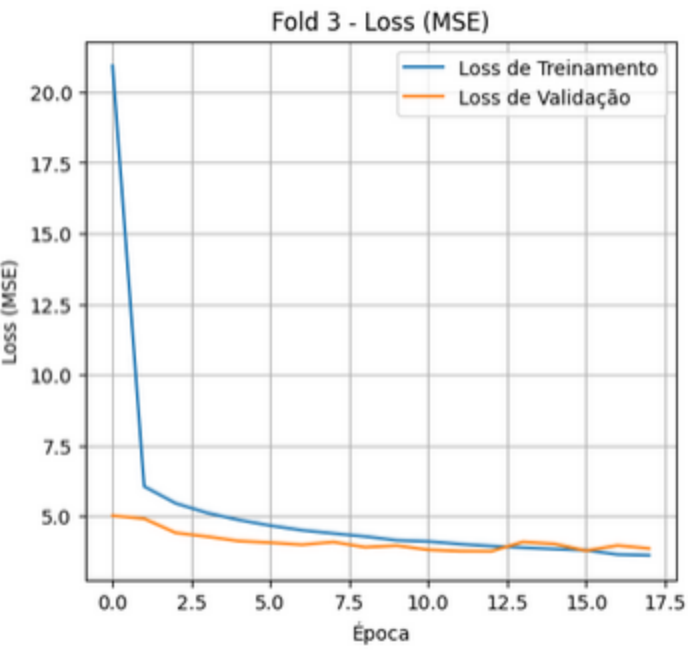
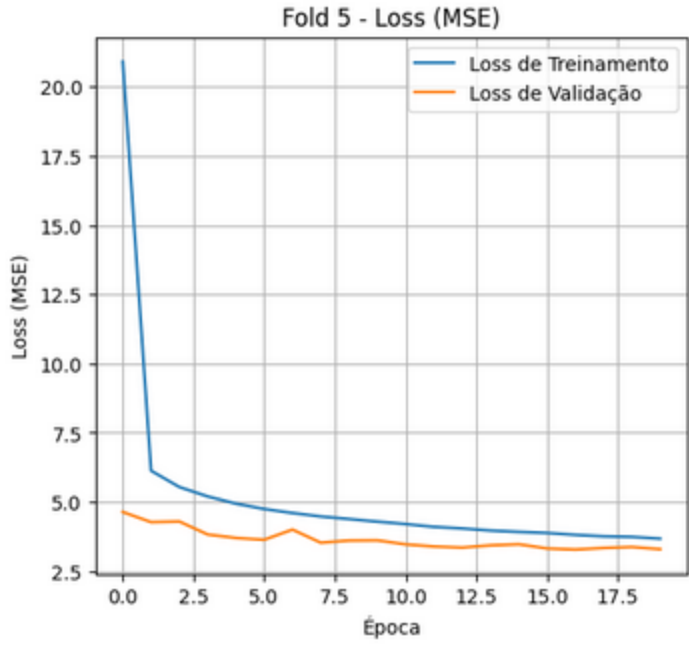
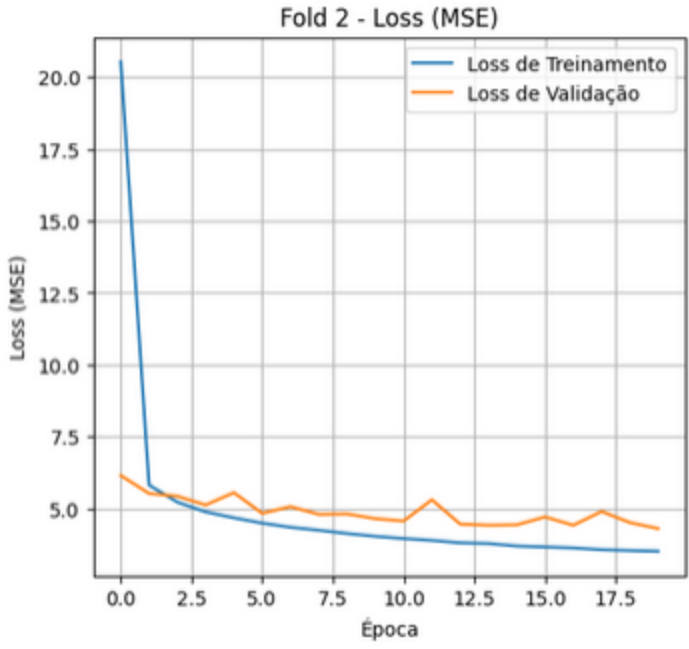
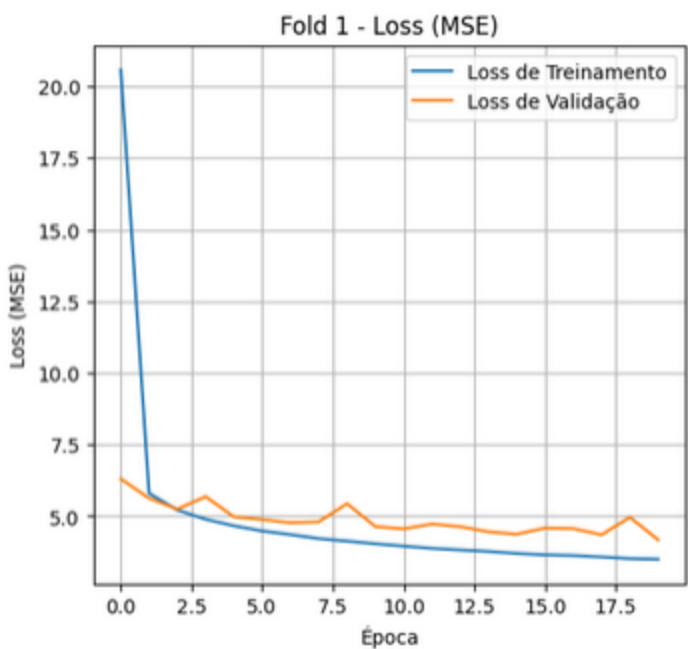
Curvas de Treinamento

Resultados

ResNet50V2 – Segunda Tentativa (Taxa de Aprendizado: $3e-07$):

- Limitado a 20 épocas.
- Earlystopping com paciência de 5 épocas.
- ReduceLROnPlateau com paciência de 3 épocas.

EarlyStopping foi acionado em apenas 1 fold, e as melhores épocas também foram próximas do final. A validação pareceu mais estável.



Fold	RMSE Treino	RMSE Validação	Melhor Época
1,00	1,8763	2,0442	20
2,00	1,8661	2,0737	20
3,00	1,9798	1,9383	13
4,00	1,8091	2,1808	20
5,00	1,9528	1,8101	17
Média Final	1,8968	2,0094	—

Curvas de Treinamento

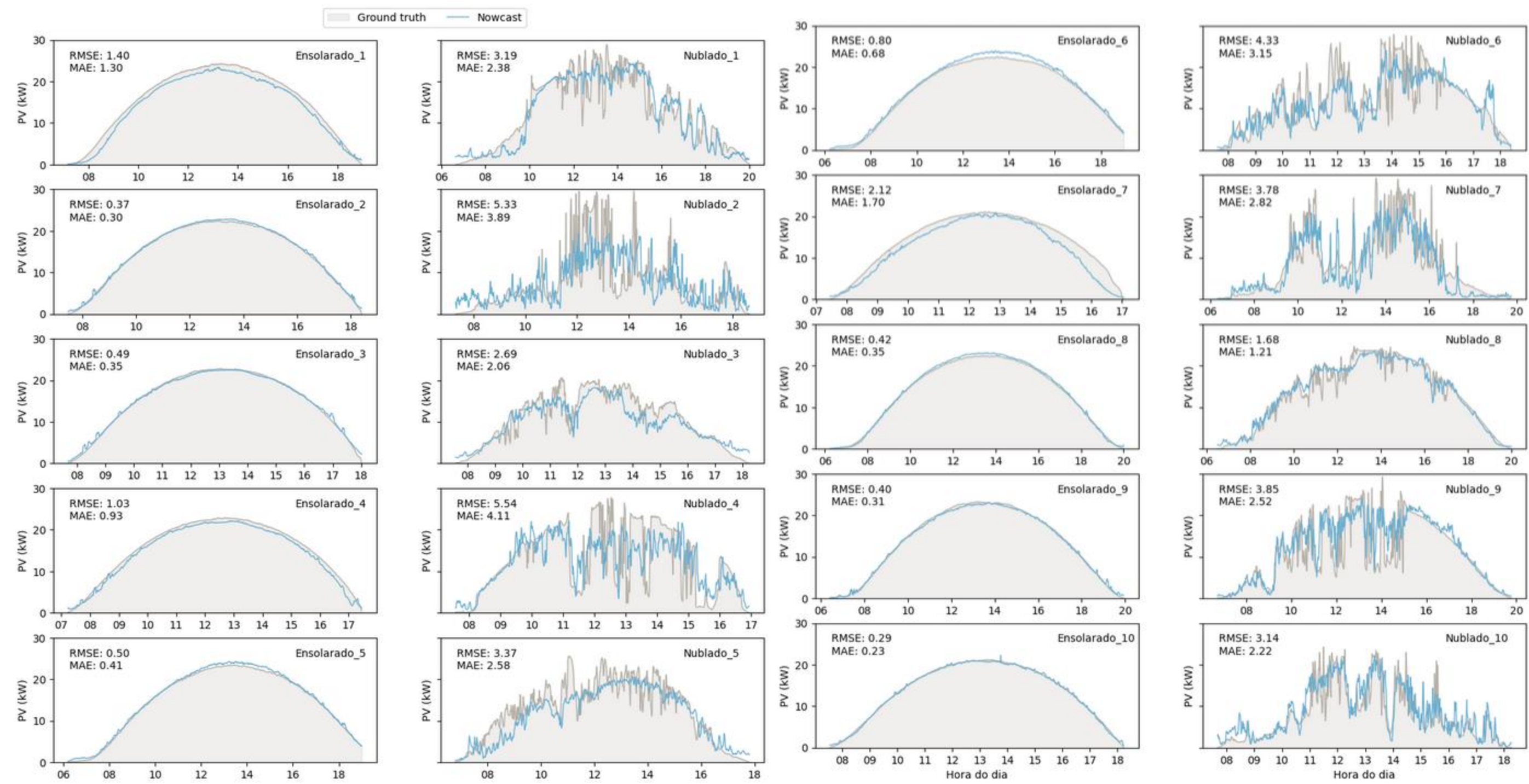
Resultados

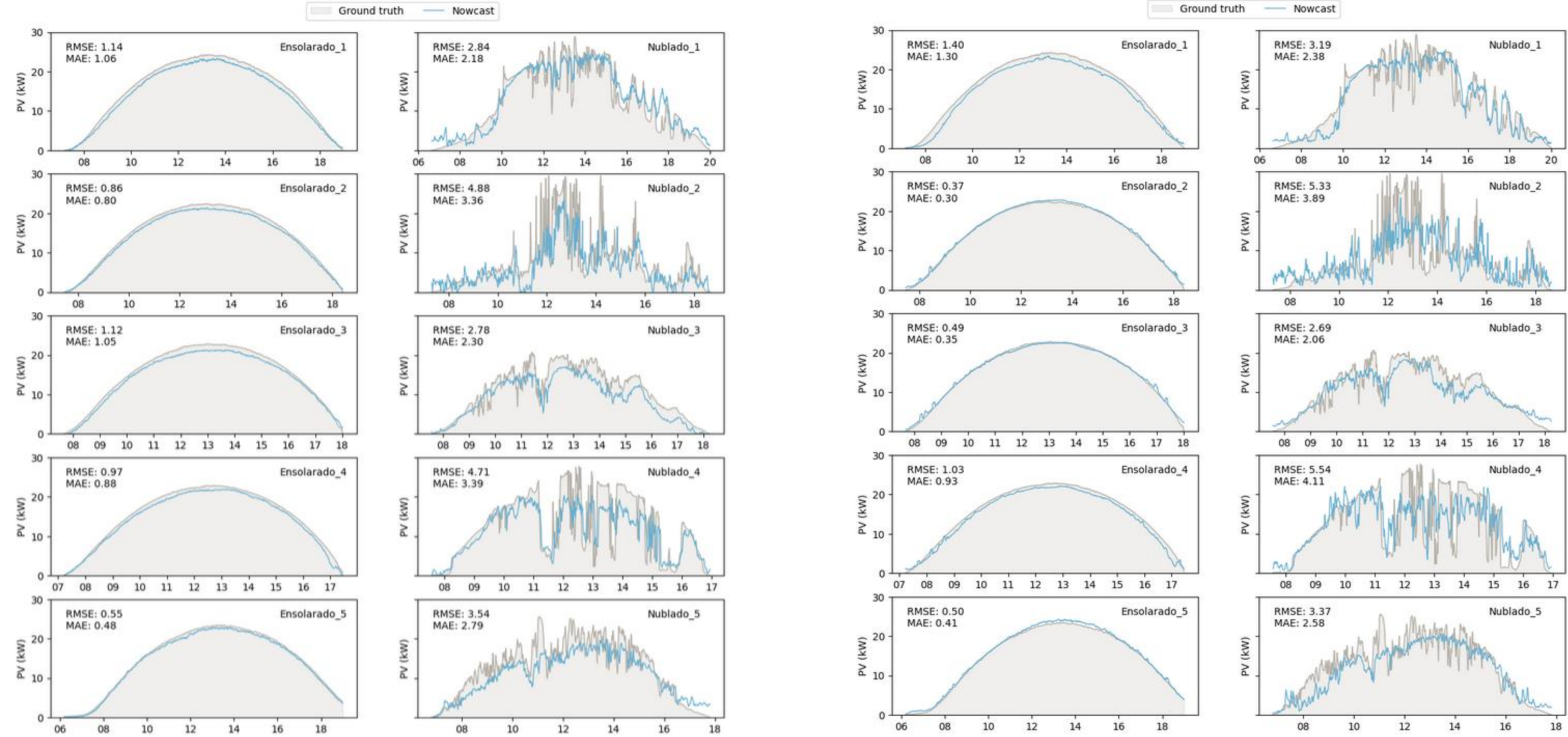
ResNet50V2 – Teste

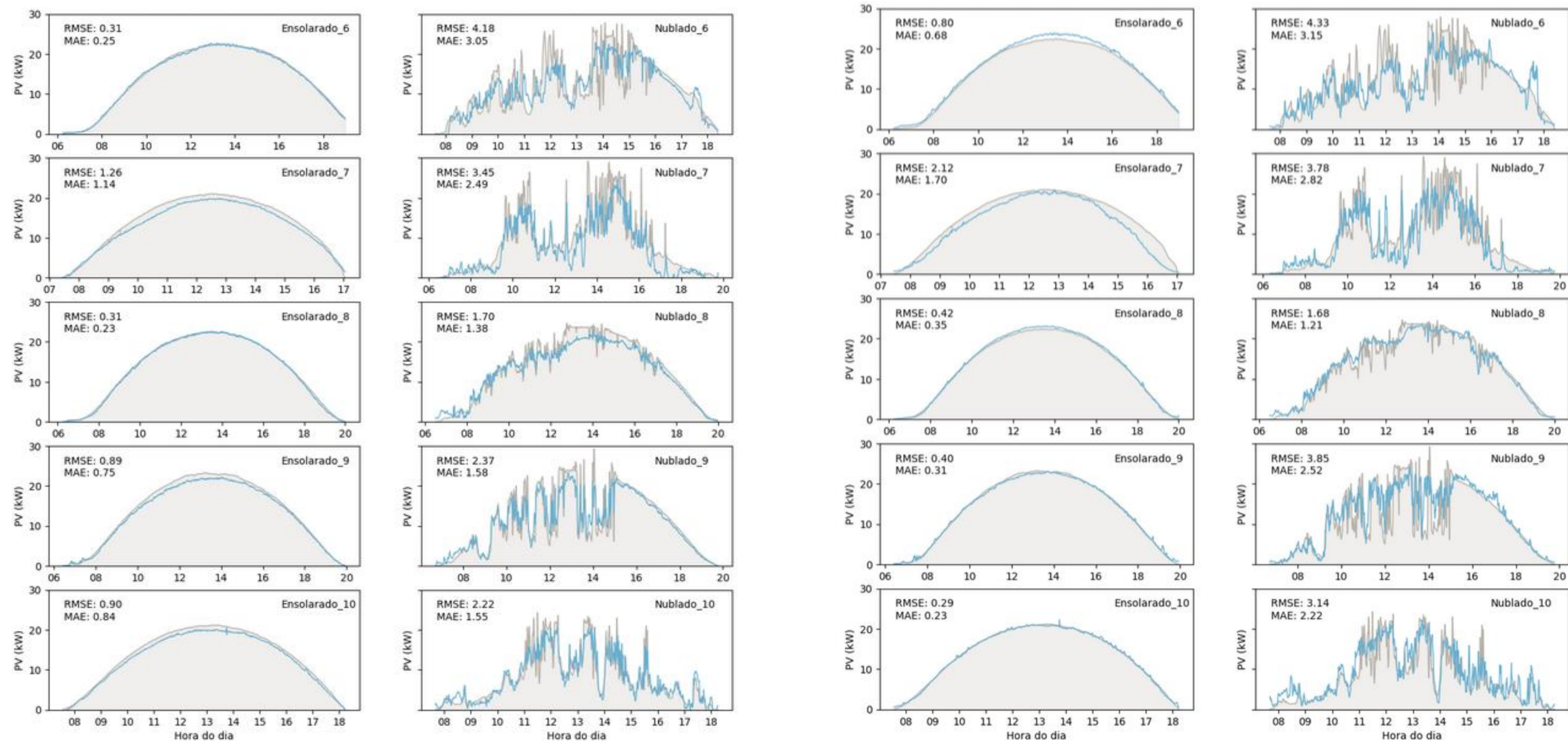
Taxa de Aprendizizado	RMSE de Teste
1e-06	2.7182
3e-07	2.7570

Condição do Céu	RMSE	MAE
Dias Ensolarados	0.917	0.631
Dias Nublados	3.786	2.642
Geral	2.757	1.639

- Bom desempenho para dias ensolarados.
- RMSE aproximadamente 4 vezes maior para dias nublados.
- Captura a tendência gral da curva.
- Tende a suavizar as flutuações.







Resultados Teste

Conclusão

Viabilidade Comprovada:

- O projeto demonstrou com sucesso a viabilidade do uso de Redes Neurais Convolucionais (CNNs) para o "nowcasting" de energia solar a partir de imagens do céu.

Ambos modelos apresentaram bons desempenhos similares e mostraram dificuldade com dias nublados.

ResNet50V2:

- Exibiu resultados promissores mesmo sem uma otimização completa devido ao alto custo computacional.
- Atingiu um desempenho próximo ao da baseline (RMSE 2.757), seguindo um padrão similar: bom em dias ensolarados (RMSE 0.917) e com dificuldades em dias nublados (RMSE 3.786).

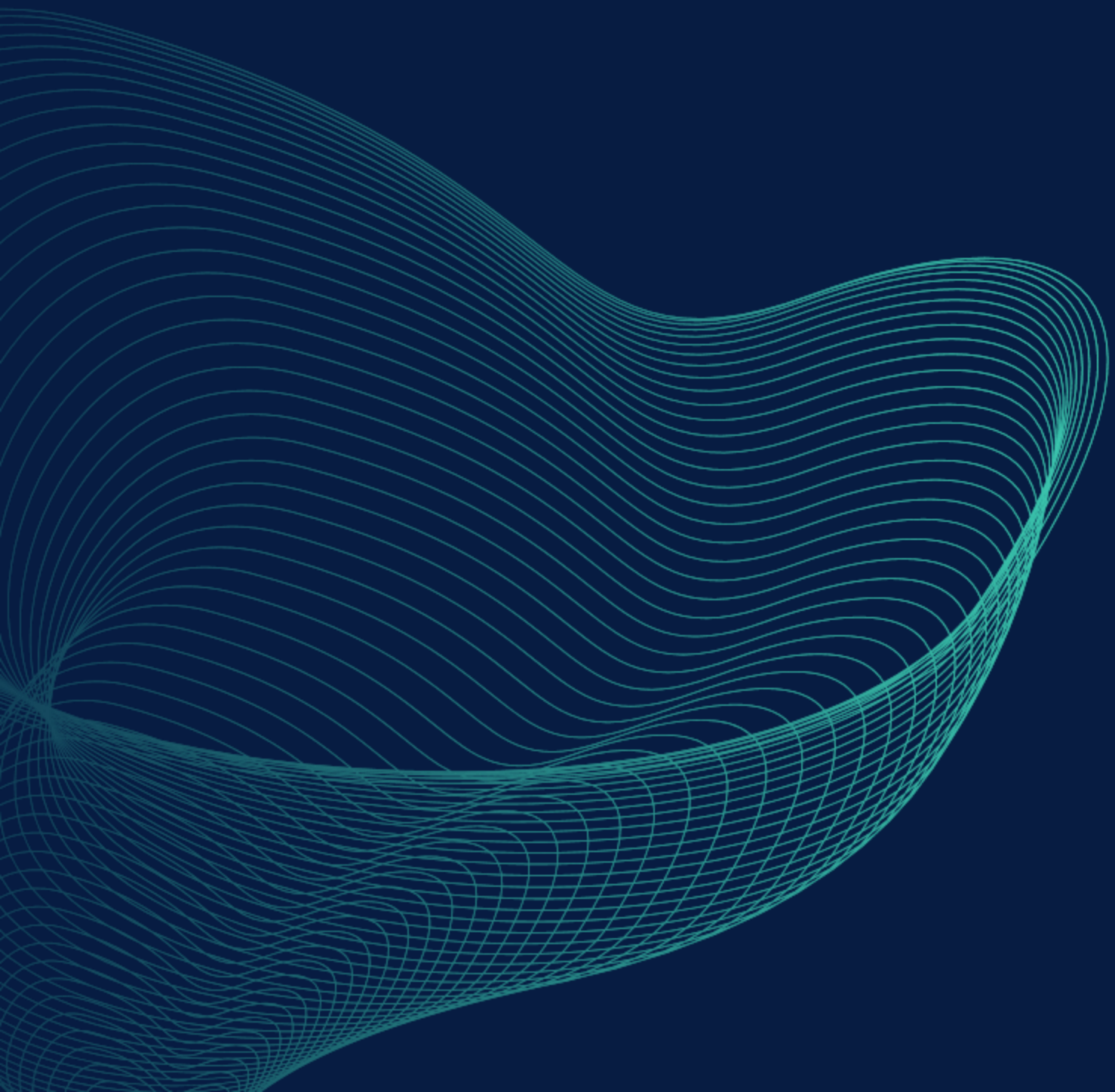
Aprendizados Metodológicos:

- Uma pipeline de dados eficiente foi crucial para superar limites de memória e mitigar o *overfitting*.

Trabalhos Futuros:

- Explorar Data Augmentation para superar as condições climáticas desbalanceadas do dataset.

Obrigado!



Gustavo Friol Bento