

# PROJETO 1 — CLASSIFICAÇÃO DE DÍGITOS COM REDES NEURAIS ARTIFICIAIS (RNA) NO MNIST

ALUNO: ADRIANO PEDRO COUTO DOS SANTOS

TEMA: FUNÇÃO DE ATIVAÇÃO

# CONTEXTO/PROBLEMÁTICA

- O MNIST (Modified National Institute of Standards and Technology) é um dos conjuntos de dados mais populares para treino e teste de algoritmos de classificação de imagens.
- Ele é composto por **70.000 imagens** de dígitos manuscritos de **0 a 9**, sendo **60.000 imagens** para treinamento e **10.000 imagens** para teste.
- Cada imagem é em **escala de cinza**, com tamanho **28x28 pixels**.  
As imagens são simples e centralizadas, focadas em representar a forma dos dígitos com diferentes estilos de escrita manual.
- O MNIST é amplamente utilizado como um **padrão de referência ("benchmark")** para testar algoritmos de Machine Learning, especialmente Redes Neurais, por ser fácil de manipular mas ainda desafiador o suficiente para validar modelos de classificação.

[HTTPS://DOCS.ULTRALYTICS.COM/PT/DATASETS/CLASSIFY/MNIST/#EXTENDED-MNIST-EMNIST](https://docs.ultralytics.com/pt/datasets/classify/mnist/#extended-mnist-emnist)

# CONTEXTO/PROBLEMÁTICA

- A escolha da função de ativação é um dos fatores críticos para o desempenho de Redes Neurais Artificiais (RNA), principalmente em tarefas de classificação como no dataset MNIST.
- Cada função de ativação (ReLU, Sigmoid e Tanh) possui características distintas que afetam a capacidade de aprendizado, a propagação do gradiente e a robustez do modelo a ruídos nos dados de entrada.
- Além disso, em cenários práticos, os dados reais frequentemente apresentam algum nível de ruído, o que pode impactar significativamente a performance de classificadores.
- Avaliar como diferentes ativações reagem a esse ruído permite entender melhor a estabilidade e a capacidade de generalização dos modelos.
- Portanto, o problema tratado é:
  - **Comparar** o comportamento de diferentes funções de ativação em termos de **acurácia** e **robustez** em Redes MLP.
  - **Analisar** o impacto da introdução de ruído gaussiano nos dados sobre a performance das redes com diferentes ativações.
- Essa análise é relevante para projetar sistemas de classificação mais eficientes e confiáveis em aplicações práticas que envolvam dados ruidosos

# CONTEXTO/PROBLEMÁTICA

Critério	ReLU	Sigmoid	TanH
Fórmula	$f(x) = \max(0, x)$	$\sigma(x) = \frac{1}{1 + e^{-x}}$	$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$
Intervalo de saída	$[0, \infty)$	$(0, 1)$	$(-1, 1)$
Vantagens	Simples e eficiente	Suave e diferenciável	Saída centrada em zero
	Evita saturação positiva	Boa para saída binária	Melhor que sigmoid para redes profundas
	Rápido de treinar		
Desvantagens	Pode causar “neurônios mortos” (gradiente 0 para $x < 0$ )	Gradiente desaparece em extremos	Também sofre com gradiente pequeno
		Saída não centrada	Sensível a ruído
Uso típico	Camadas ocultas em redes profundas ou com ruído	Camada de saída em classificação binária	Camadas ocultas em redes moderadas
Robustez ao ruído	Alta	Média	Baixa

[HTTPS://WWW.AITUDE.COM/COMPARISON-OF-SIGMOID-TANH-AND-RELU-ACTIVATION-FUNCTIONS/](https://www.aitude.com/comparison-of-sigmoid-tanh-and-relu-activation-functions/)

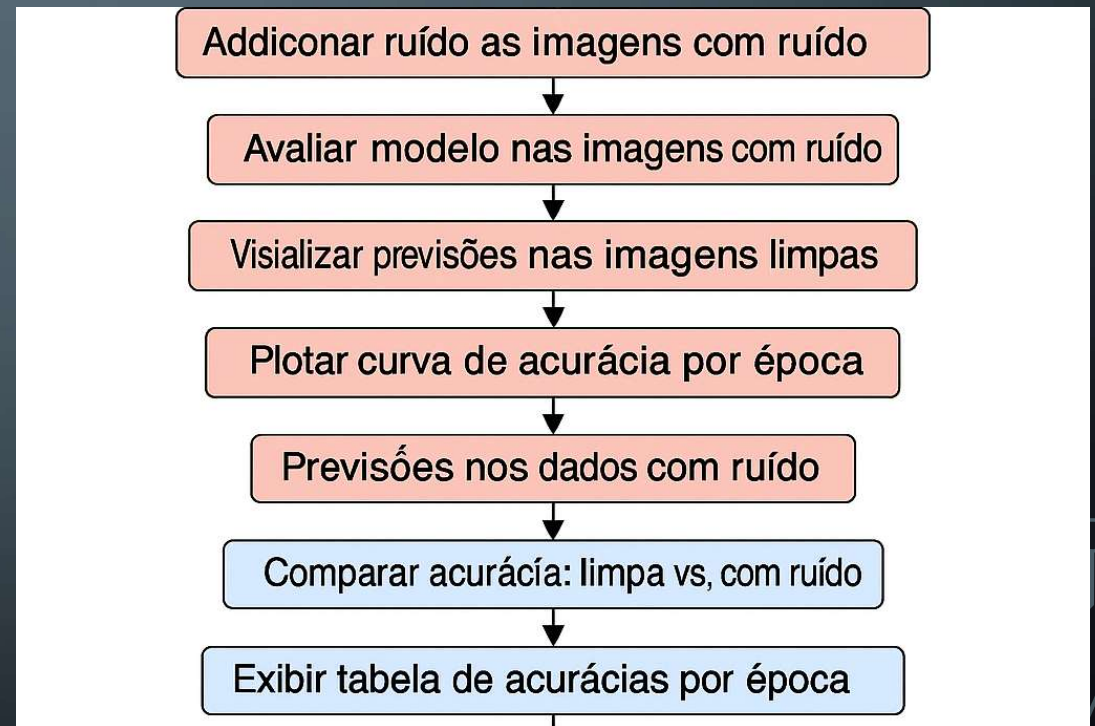
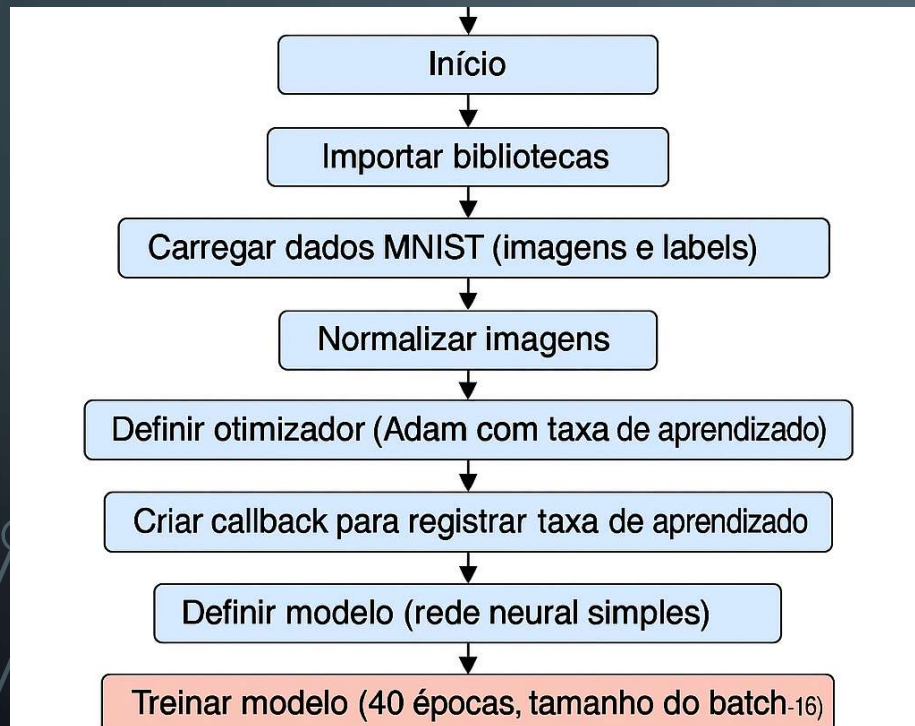
# OBJETIVO GERAL

- Avaliar qual função de ativação (ReLU, Sigmoid ou Tanh) é mais robusta ao ruído em imagens do MNIST e proporciona melhor acurácia de classificação.
- Investigar o impacto da profundidade da rede, comparando redes com uma e múltiplas camadas ocultas para entender se arquiteturas mais profundas conseguem mitigar ou amplificar os efeitos do ruído.
- Identificar padrões de comportamento que possam orientar futuras escolhas de arquiteturas e funções de ativação para problemas de classificação em ambientes ruidosos.

Especificamente, espera-se:

- Determinar se funções não lineares suaves (como Sigmoid e Tanh) apresentam maior degradação de desempenho sob ruído em comparação à ReLU.
- Observar se redes mais profundas melhoram ou pioram a robustez quando expostas a ruído gaussiano.

# METODOLOGIA



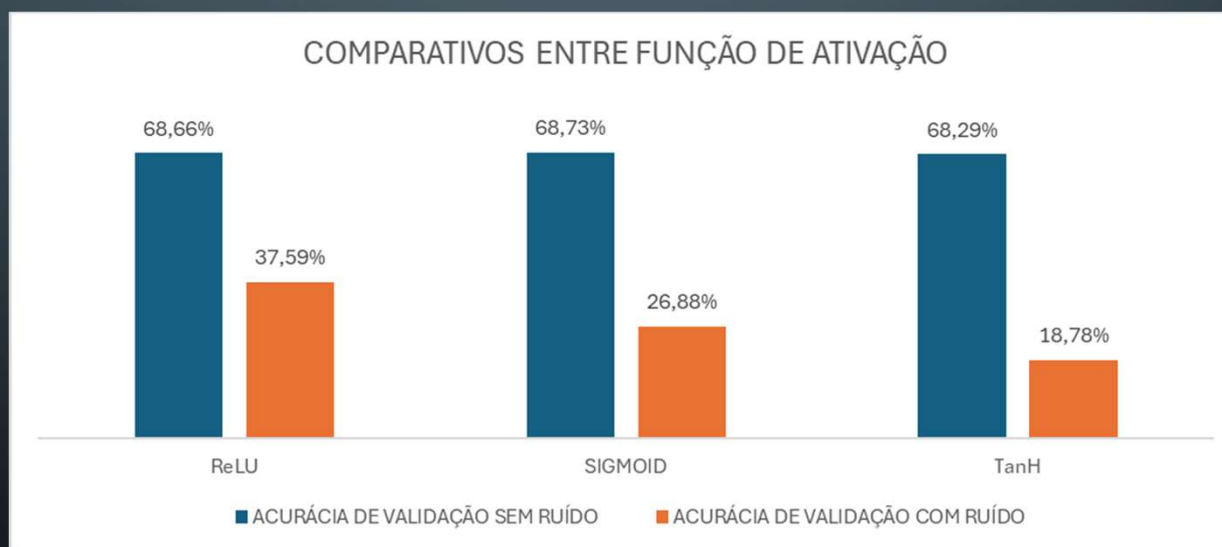
# METODOLOGIA

- Para comparar as funções de ativação Relu, Sigmoid e Tanh, foram realizados 4 experimentos nos quais foram treinados e testados utilizando as seguintes categorias:
  - Quantidade de neurônios na camada oculta
  - Ruído Gaussiano aplicado nos testes.
  - Batch Size utilizado.
  - Learning Rate utilizado.
  - Quantidade de Épocas.
  - Acurácia do treinamento.
  - Acurácia do teste sem ruído.
  - Acurácia do teste com ruído.

# RESULTADOS

- Experimento 1:

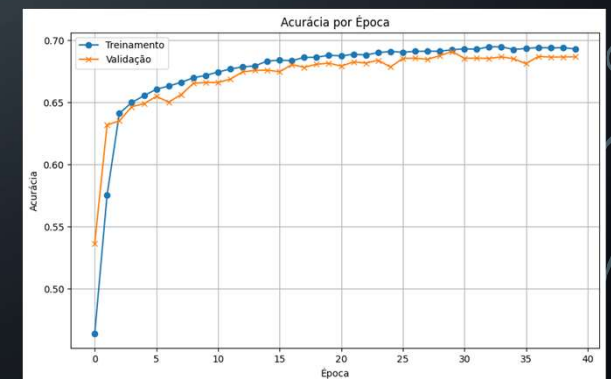
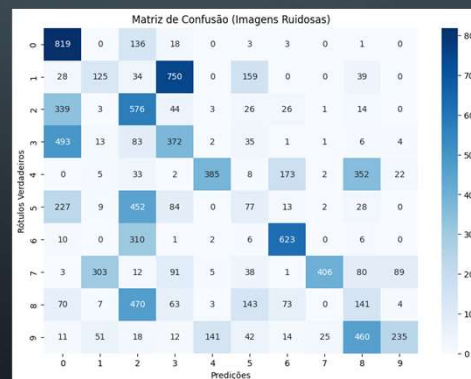
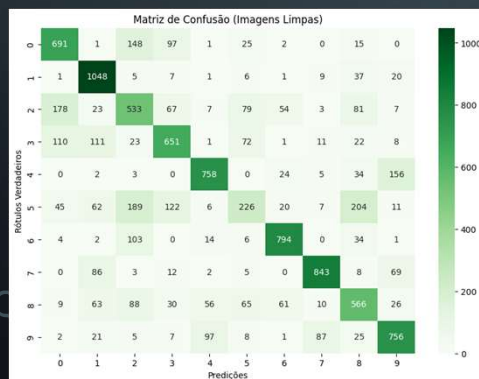
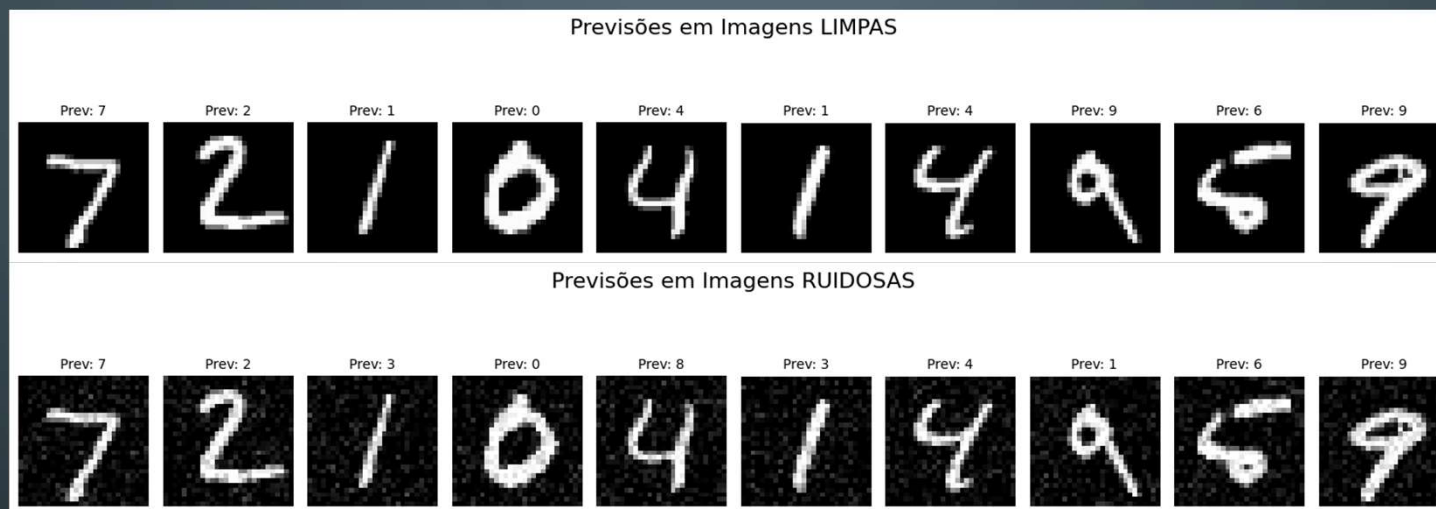
FUNÇÃO DE ATIVAÇÃO	CAMADAS OCULTAS	RÚIDO GAUSSIANO	BATCH SIZE	LEARNING RATE	ÉPOCAS	ACURÁCIA DE TREINAMENTO	ACURÁCIA DE VALIDAÇÃO SEM RÚIDO	ACURÁCIA DE VALIDAÇÃO COM RÚIDO
ReLU	2	0,1	16	0,001	40	69,29%	68,66%	37,59%
SIGMOID	2	0,1	16	0,001	40	68,94%	68,73%	26,88%
TanH	2	0,1	16	0,001	40	68,91%	68,29%	18,78%





# RESULTADOS

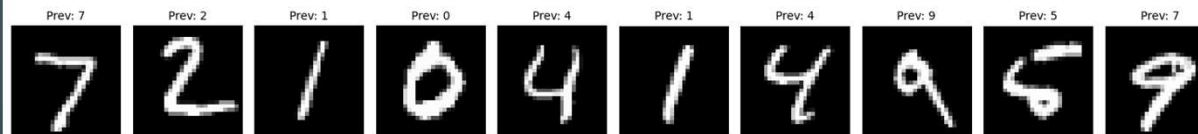
- RELU:



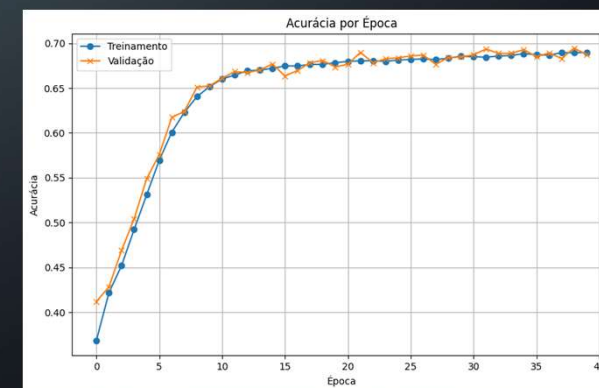
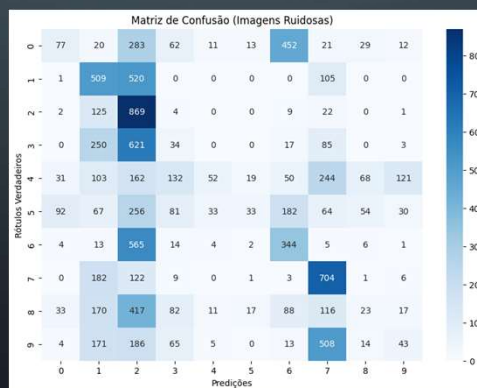
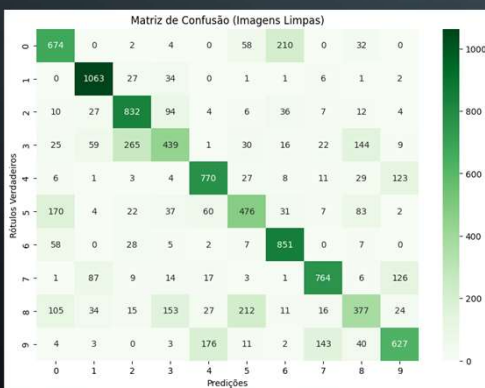
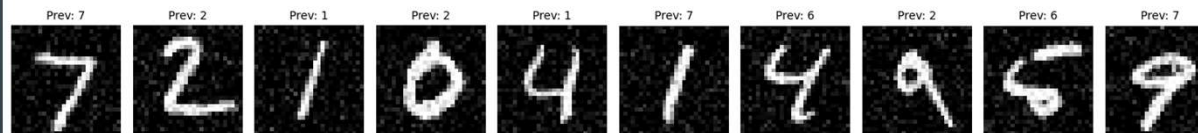
# RESULTADOS

- SIGMOIDE:

Previsões em Imagens LIMPAS

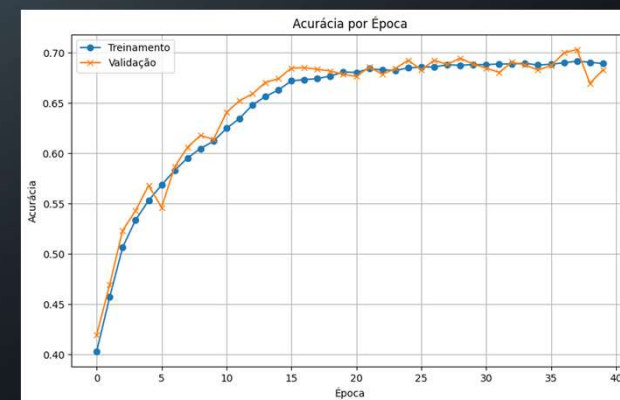
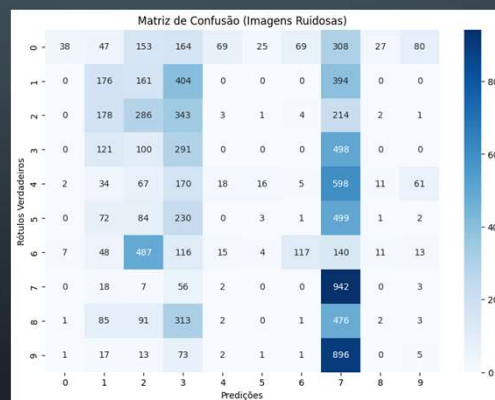
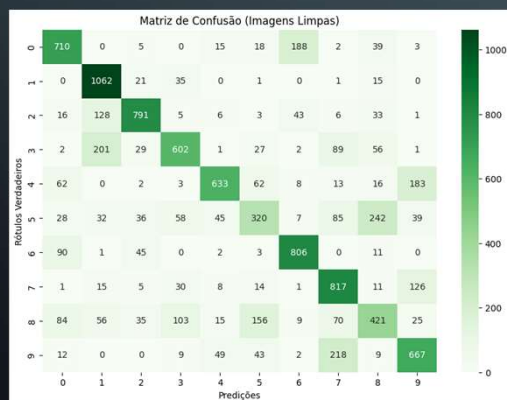
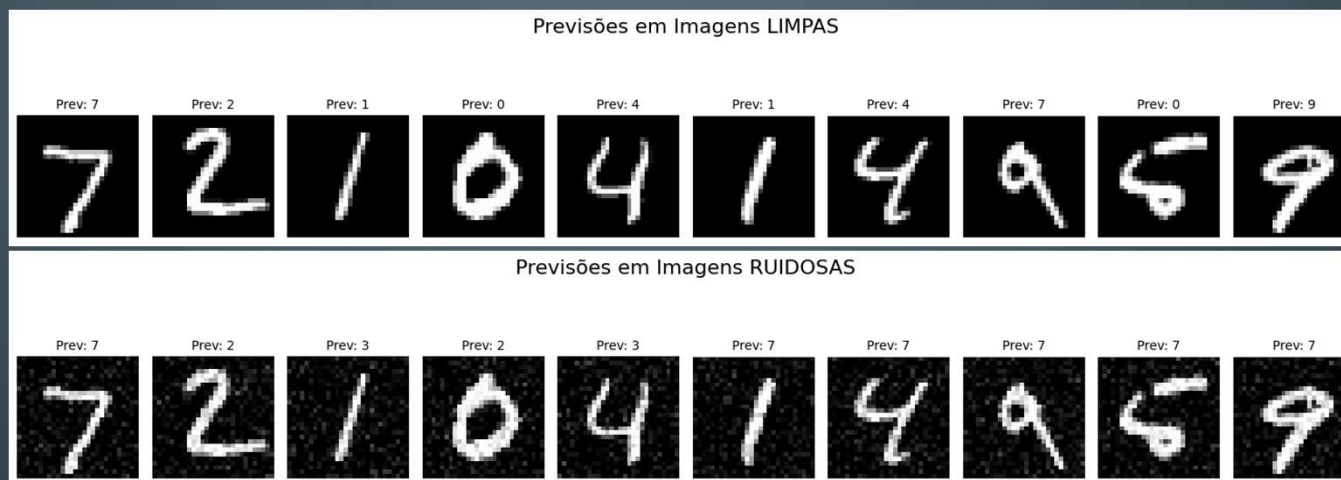


Previsões em Imagens RUIDOSAS



# RESULTADOS

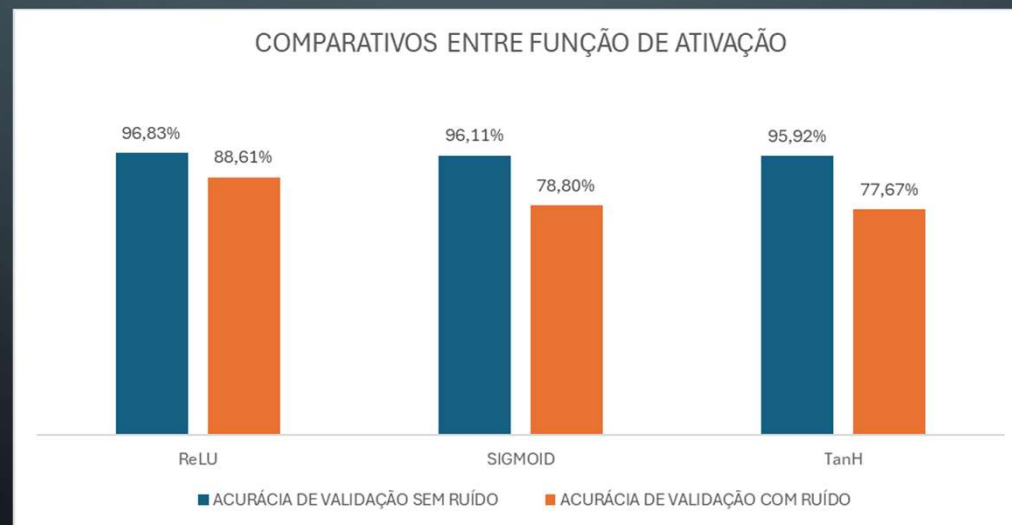
- TANH:



# RESULTADOS

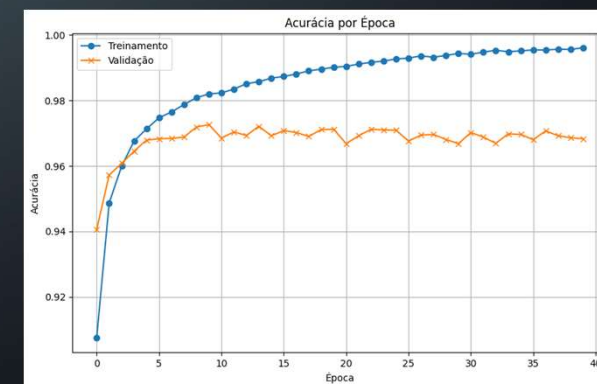
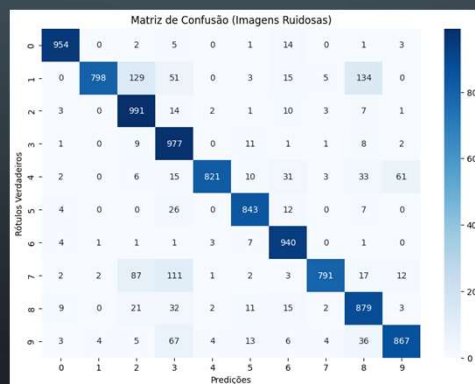
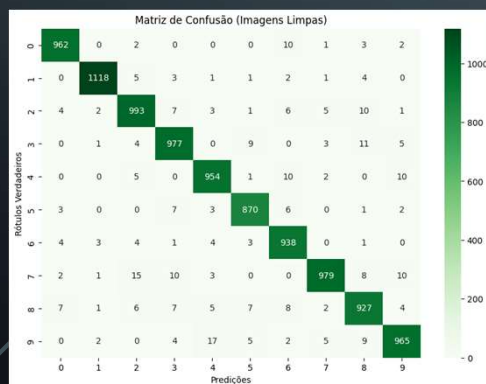
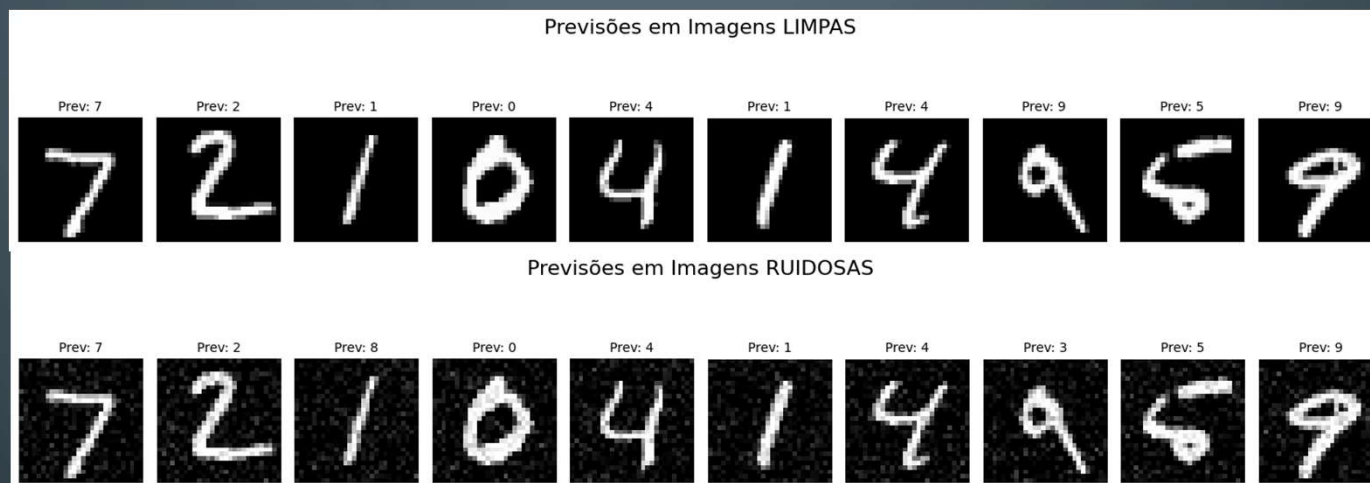
- Experimento 2:

FUNÇÃO DE ATIVAÇÃO	CAMADAS OCULTAS	RÚIDO GAUSSIANO	BATCH SIZE	LEARNING RATE	ÉPOCAS	ACURÁCIA DE TREINAMENTO	ACURÁCIA DE VALIDAÇÃO SEM RUÍDO	ACURÁCIA DE VALIDAÇÃO COM RUÍDO
ReLU	32	0,1	16	0,001	40	99,61%	96,83%	88,61%
SIGMOID	32	0,1	16	0,001	40	99,54%	96,11%	78,80%
TanH	32	0,1	16	0,001	40	99,53%	95,92%	77,67%



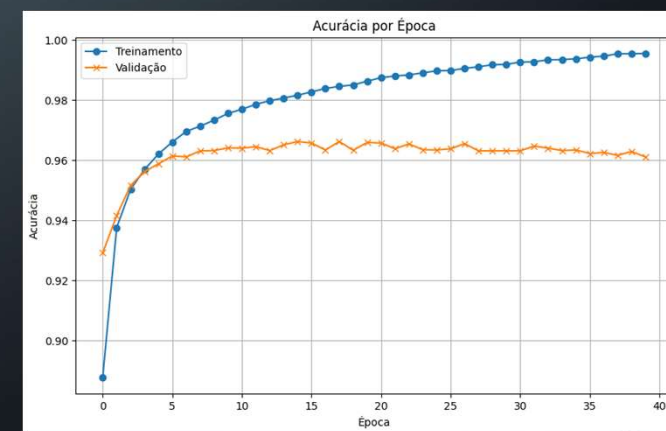
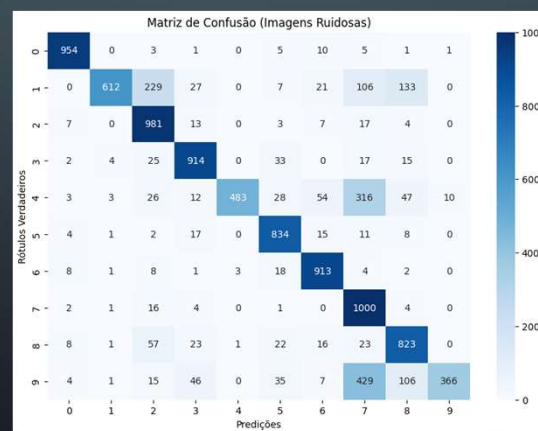
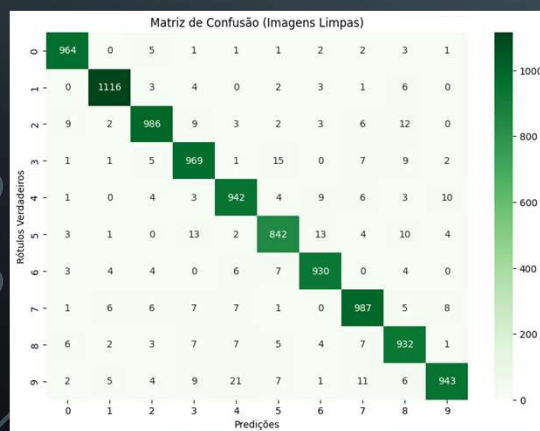
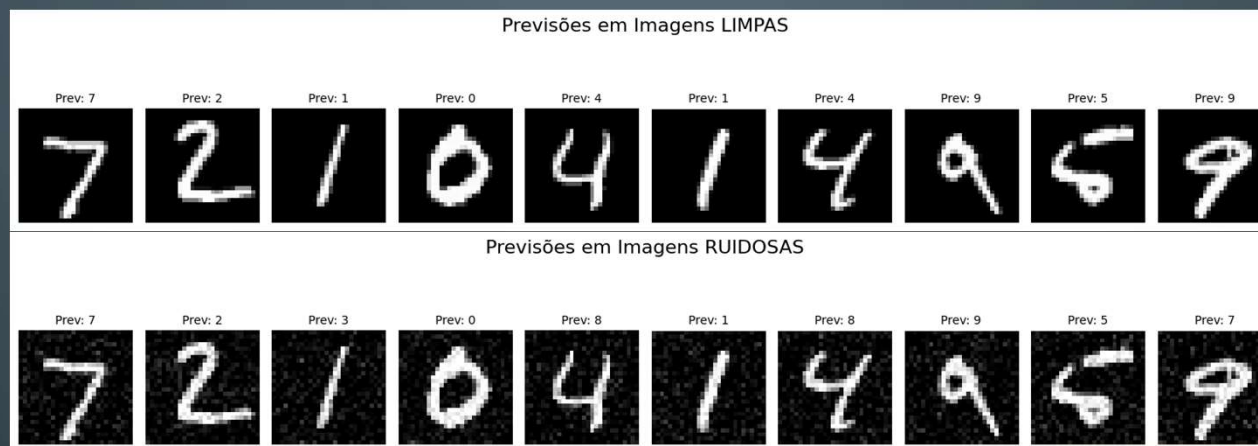
# RESULTADOS

- RELU:



# RESULTADOS

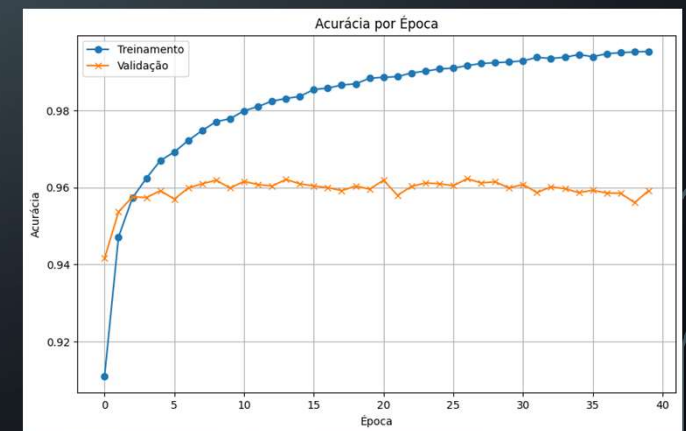
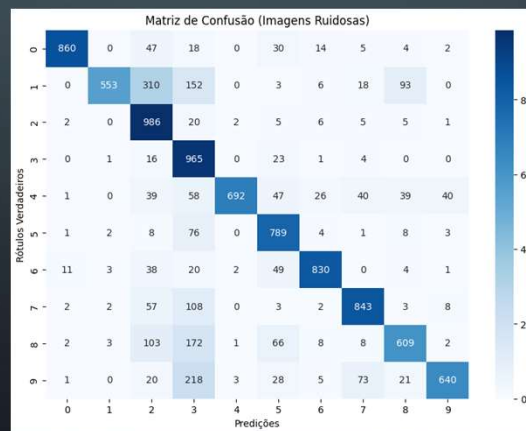
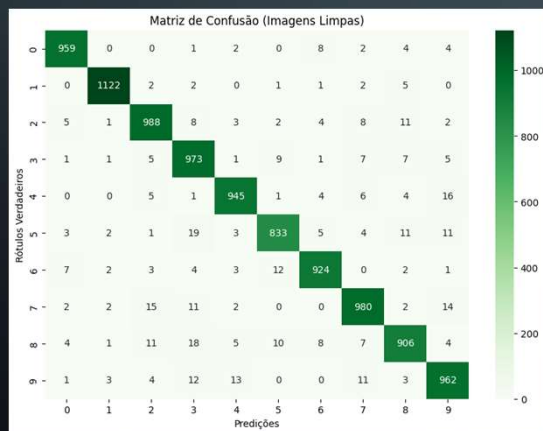
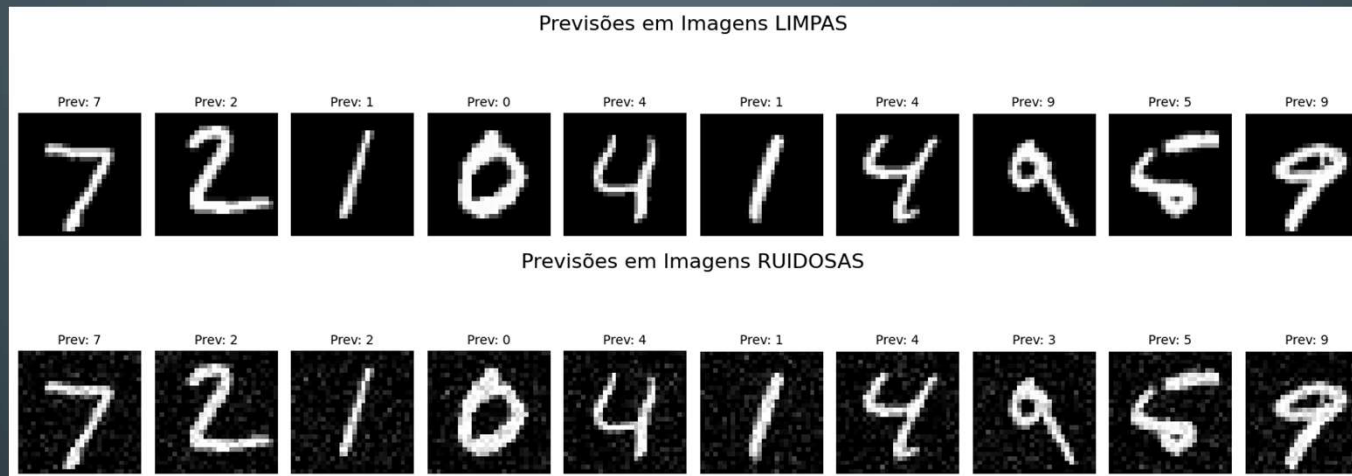
- SIGMOIDE:





# RESULTADOS

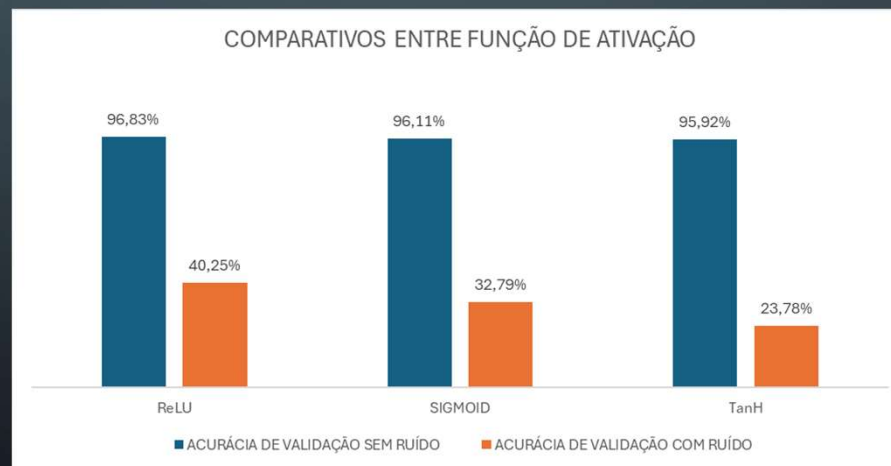
- TANH:



# RESULTADOS

- Experimento 3:

FUNÇÃO DE ATIVAÇÃO	CAMADAS OCULTAS	RÚIDO GAUSSIANO	BATCH SIZE	LEARNIN G RATE	ÉPOCAS	ACURÁCIA DE TREINAMENTO	ACURÁCIA DE VALIDAÇÃO SEM RUÍDO	ACURÁCIA DE VALIDAÇÃO COM RUÍDO
ReLU	32	0,4	16	0,001	40	99,61%	96,83%	40,25%
SIGMOID	32	0,4	16	0,001	40	99,54%	96,11%	32,79%
TanH	32	0,4	16	0,001	40	99,53%	95,92%	23,78%

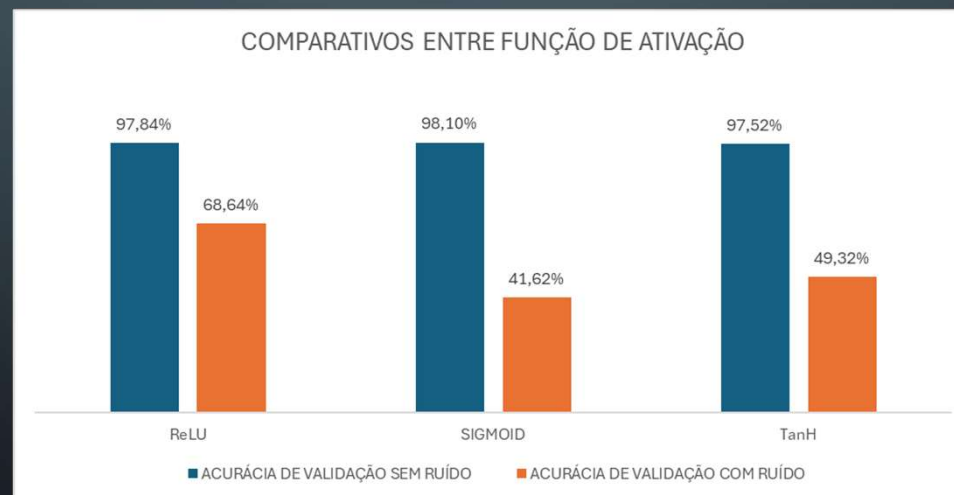




# RESULTADOS

- Experimento 4:

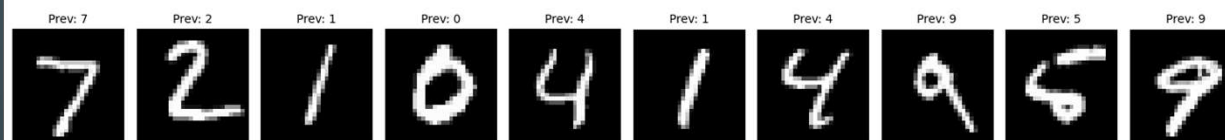
FUNÇÃO DE ATIVAÇÃO	CAMADAS OCULTAS	CAMADAS OCULTAS	RÚIDO GAUSSIANO	BATCH SIZE	LEARNIN G RATE	ÉPOCAS	ACURÁCI A DE TREINAM	ACURÁCIA DE VALIDAÇÃO SEM RUÍDO	ACURÁCIA DE VALIDAÇÃO COM RUÍDO
ReLU	128	64	0,4	16	0,001	40	99,79%	97,84%	68,64%
SIGMOID	128	64	0,4	16	0,001	40	99,90%	98,10%	41,62%
TanH	128	64	0,4	16	0,001	40	99,73%	97,52%	49,32%



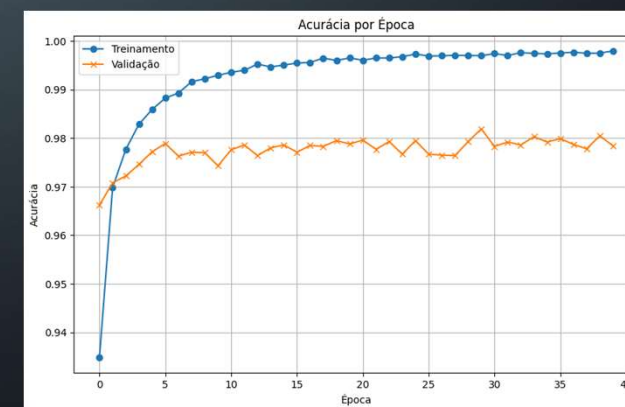
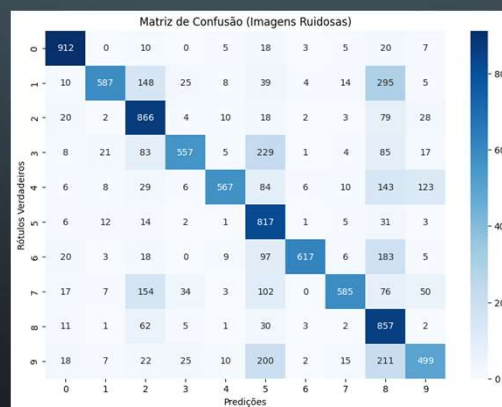
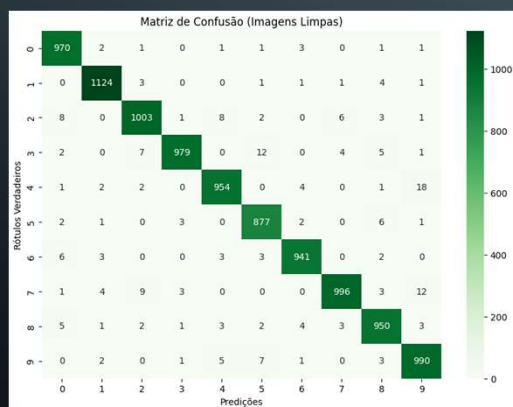
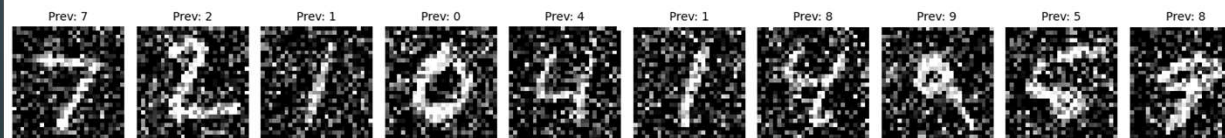
# RESULTADOS

- RELU:

Previsões em Imagens LIMPAS

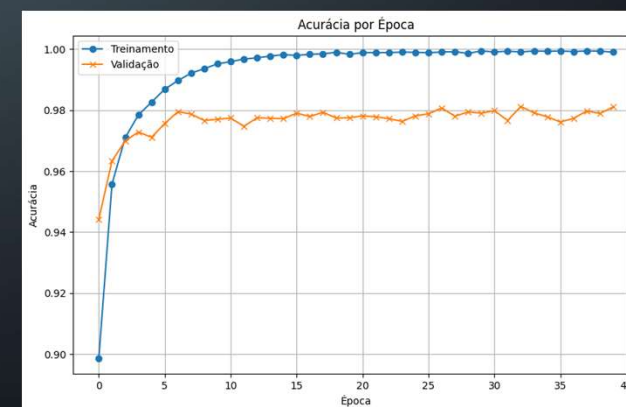
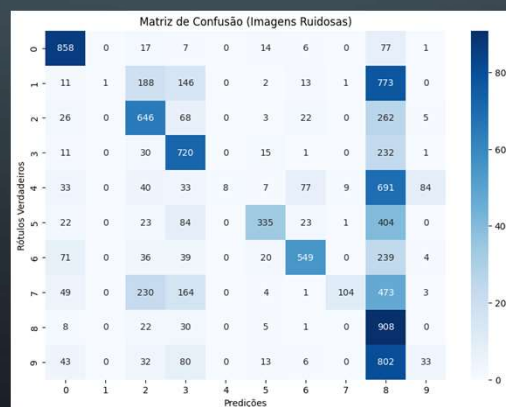
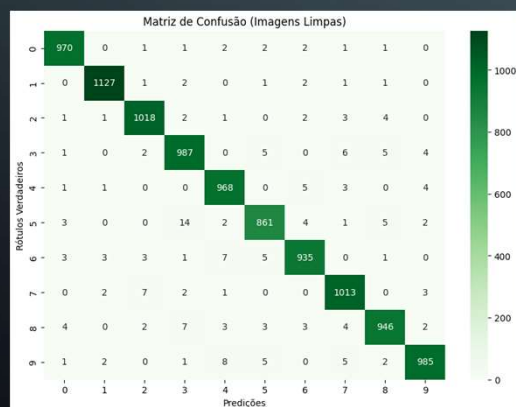
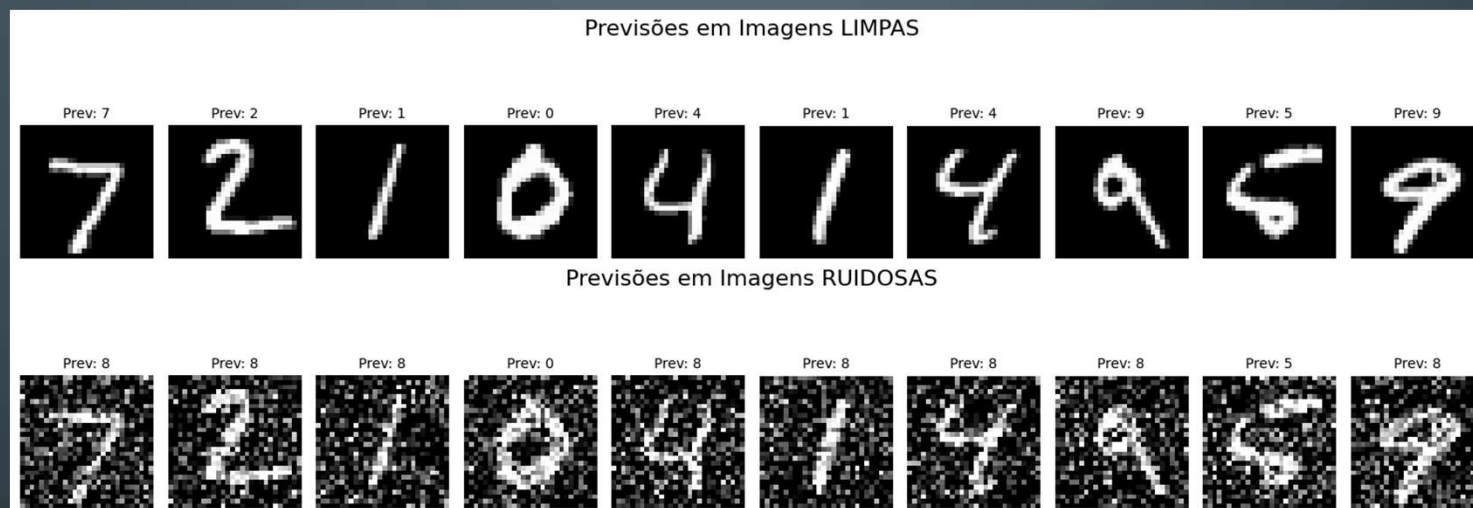


Previsões em Imagens RUIDOSAS



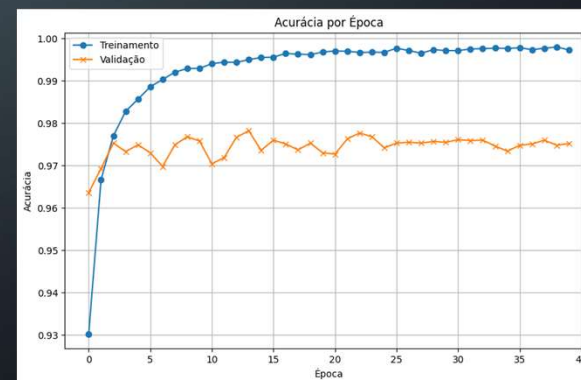
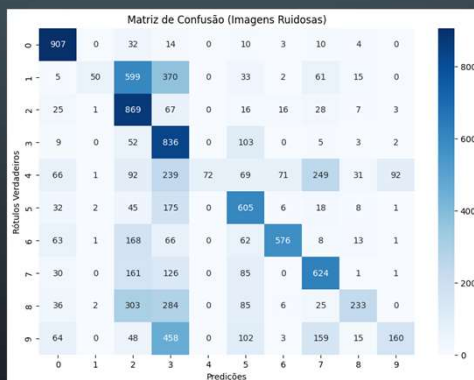
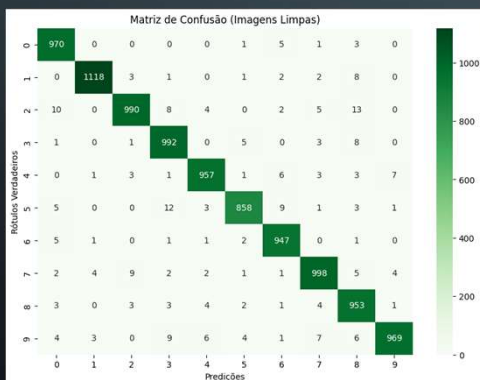
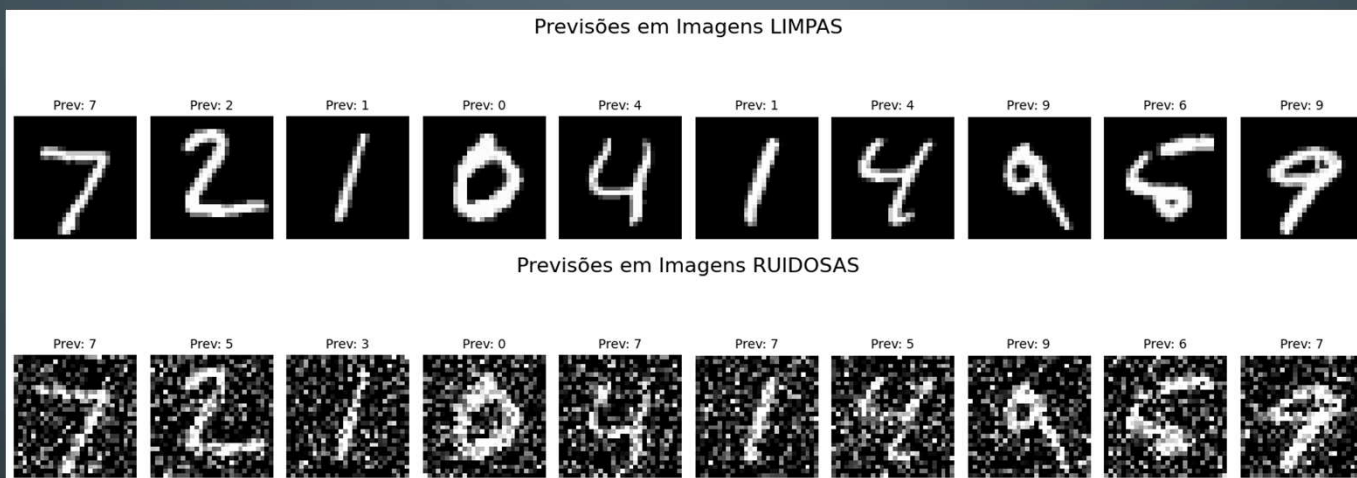
# RESULTADOS

- SIGMOIDE:



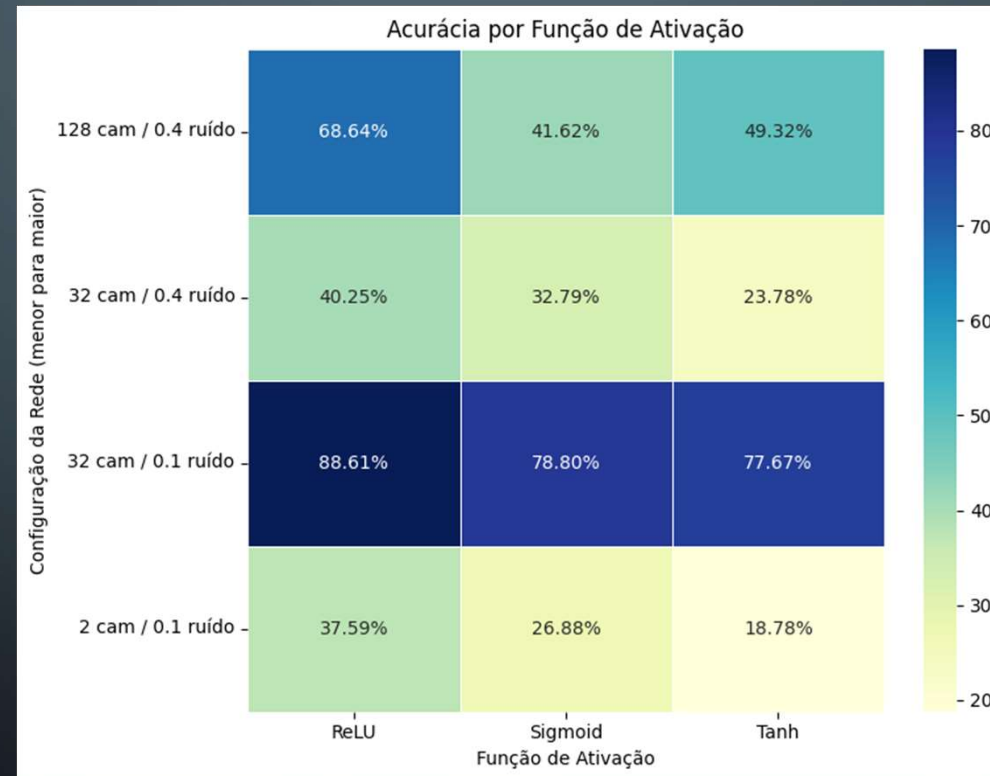
# RESULTADOS

- TANH:



# DISCUSSÃO

- Realizado Comparativo entre as funções de ativação utilizando as seguintes métricas:



# CONCLUSÃO

- Com os resultados obtidos foi possível constatar:
  - Relu com melhor performance: Mantém gradientes significativos para valores positivos de entrada, o que ajuda no aprendizado profundo e mais rápido.
  - Sigmoid com performance intermediária: Valores muito grandes ou pequenos de entrada levam a gradientes quase nulos, logo a um aprendizado mais lento.
  - Tanh com pior performance: Convergência lenta, possível dificuldade de generalização, especialmente em datasets grandes como MNIST.
- Quando inserido uma maior quantidade de camadas ocultas e um erro gaussiano maior:
  - Tanh se tornou melhor que Sigmoid: Em redes mais profundas, a Tanh preserva melhor a informação do gradiente, A Tanh suaviza o ruído porque suas saídas estão mais equilibradas entre -1 e 1, reduzindo o impacto de pequenas perturbações.
- Possíveis melhorias futuras:
  - Aumento de quantidade de camadas ocultas e neurônios.

[HTTPS://WWW.GEEKSFORGEEKS.ORG/TANH-VS-SIGMOID-VS-RELU/](https://www.geeksforgeeks.org/tanh-vs-sigmoid-vs-relu/)