

Classificação de Dígitos Manuscritos em Imagens Ruidosas

Explorando o uso e a adaptação conjunta de redes neurais para classificação e remoção de ruído

Gabriel Eduardo Lima
Universidade Federal do Paraná
Curitiba, Paraná, Brasil
gabriel.lima1@ufpr.br

Ricardo de la Rocha Ladeira
Instituto Federal Catarinense Campus
Blumenau
Blumenau, Santa Catarina, Brasil
ricardo.ladeira@ifc.edu.br

Éder Augusto Penharbel
Instituto Federal Catarinense Campus
Blumenau
Blumenau, Santa Catarina, Brasil
eder.penharbel@ifc.edu.br

ABSTRACT

Image classification entails assigning labels to an image based on its visual representation. Among the approaches used for this problem, the use of neural networks is particularly notable. However, these algorithms can be hampered by noisy data. Preprocessing techniques, such as autoencoder neural networks, are employed to mitigate this impact. Therefore, the problem of classifying noisy images is approached using two models: one for classification and the other for noise removal. The research question asks about the feasibility of incorporating information from the classification model into the development and adaptation of the noise removal model. The general objective is to explore the possibility on accuracy improving of an image classifier using an autoencoder for noise removal, adapted by introducing the classification error metric into its learning algorithm. The scope of this research is limited to the challenging problem of classifying handwritten digits in images. The adopted method is an experimental research consisting of four stages: (I) evaluation of the classifier's performance in noise-free images; (II) analysis of classifier performance in noisy images; (III) investigation of the impact on classification performance after image pre-processing; (IV) exploration of the adaptation of the noise removal model and analysis of its impact on classification. It stands out as a significant result that the proposed adaptation can improve the rate of correct classifications for pre-processed noisy images, in addition to showing signs of making the development of noise removal models more efficient.

KEYWORDS

Classificação de Imagens, Remoção de Ruído, Redes Neurais, Aprendizado Profundo

1 INTRODUÇÃO

A classificação de imagens envolve a atribuição de um ou mais rótulos a uma imagem com base na informação que ela representa. Sua importância é destacada pelo potencial de aplicação em problemas práticos e pelo interesse no estudo de métodos diversos [1]. Como exemplo de técnicas aplicadas são destacadas as redes neurais, modelos originalmente inspirados na maneira em que o cérebro biológico resolve tarefas [2, 3].

Apesar do progresso na aplicação dessas técnicas, existem situações que apresentam desafios à classificação de imagens por meio de redes neurais. Como exemplo, e foco dessa pesquisa, cita-se o uso de imagens ruidosas. Nesse cenário, ruídos interferem no padrão dos dados e afetam negativamente a taxa de classificações corretas

obtidas pelos modelos de classificação baseados em redes neurais [3, 4].

Uma abordagem para reduzir o impacto negativo é incorporar ruídos nos dados ao modelar as soluções [3]. No entanto, essa abordagem demanda acesso e modificações na rede neural. Uma alternativa adicional envolve a aplicação de técnicas de pré-processamento de imagens para reduzir o ruído presente nos dados, viabilizando assim o uso de um modelo previamente desenvolvido [4, 5].

Há uma variedade de algoritmos destinados à remoção de ruído [6]. Dentre eles, merece destaque o contínuo avanço de abordagens fundamentadas no Aprendizado Profundo [7], em especial os modelos baseados na arquitetura de rede neural *autoencoder*. Essa arquitetura demonstra relevância pela sua robustez e flexibilidade para resolução de problemas [8, 9].

Portanto, uma possível solução para a classificação de imagens ruidosas pode consistir no desenvolvimento individual de duas redes neurais: uma para classificação e outra para remoção de ruído. Observando que para ambas é possível calcular uma métrica de desempenho durante a fase de treinamento, esse trabalho levanta o questionamento: é viável empregar a medida proveniente do modelo de classificação para ajustar o modelo de remoção de ruído com o objetivo de incrementar o percentual de classificações corretas?

Delimitando o escopo do trabalho, essa pesquisa estuda o questionamento supracitado na classificação de dígitos manuscritos; escolhida pela sua aplicação em soluções de Reconhecimento Óptico de Caracteres e pela existência de pesquisas relacionadas para referência [10, 11]. Além disso, a classificação de dígitos manuscritos é um problema desafiador, mas que não demanda uma tremenda capacidade computacional para elaboração de uma solução [12].

1.1 Objetivos

Considerando o problema de classificação supramencionado, o objetivo geral dessa pesquisa é explorar a viabilidade de aprimorar a taxa de classificações corretas de um modelo de classificação de imagens (treinado com imagens sem ruído) por meio da utilização e do ajuste de uma rede neural *autoencoder* para pré-processamento das imagens.

Quanto aos objetivos específicos, destacam-se:

- Determinar o grau de ruído que mais exerce impacto negativo na taxa de classificações corretas;
- Avaliar o maior aumento na taxa de classificações corretas obtido com o pré-processamento das imagens ruidosas utilizadas pelo modelo de classificação.

Portanto, o presente trabalho não busca elaborar um classificador de imagens para aprimorar o desempenho obtido na literatura

estado da arte. O foco da pesquisa elaborada é analisar a adaptação do algoritmo de treino do modelo de remoção de ruído e seu impacto na classificação correta de imagens. Para isso, o problema de classificação de dígitos manuscritos é utilizado como um contexto inicial para a discussão de resultados.

2 TRABALHOS CORRELATOS

Existem diferentes pesquisas que abordam o problema de classificação de dígitos manuscritos. Em Liu et al. [10] e Baldominos, Saez e Isasi [11] são destacados trabalhos que realizam uma revisão da literatura sobre o tema. Neles é explicitada a existência de diferentes abordagens para resolver o problema de classificação de dígitos, onde as técnicas baseadas em redes neurais realizam classificações com taxas superiores a 98% de sucesso.

No trabalho de Bottou et al. [13], observa-se uma pesquisa comparativa que avalia o desempenho de diversos algoritmos para o mesmo propósito. Além de introduzir e empregar o conjunto de dados *Modified National Institute of Standards and Technology* (MNIST), o estudo enfatiza a utilização de uma rede neural *Multi-layer Perceptron* (MLP) para abordar o problema. Entre os principais resultados, destaca-se mais uma vez a obtenção de elevadas taxas de classificações corretas.

No entanto, as pesquisas mencionadas não abordam a problemática de imagens ruidosas. Nesse contexto, destacam-se os estudos de Zhou, Song e Cheung [14] e Momeny et al. [15]. Ambos tratam da classificação de dígitos do conjunto MNIST, incorporando a adição de ruído gaussiano nos dados. Os resultados e conclusões dessas pesquisas evidenciam o impacto negativo da classificação em dados ruidosos.

No que diz respeito à aplicação de métodos de remoção de ruído usando redes *autoencoders*, especificamente para classificar imagens ruidosas de dígitos, podem ser mencionados os estudos de Roy, Ahmed e Akhand [16, 17]. Os resultados dessas pesquisas demonstram melhorias nas taxas de classificação, inclusive para o conjunto de dados MNIST, ao empregar o pré-processamento dos dados por meio de redes neurais.

Todavia, esses estudos não consideram o uso de informações provenientes da rede de classificação no desenvolvimento da rede de remoção de ruído. Em contrapartida, uma abordagem semelhante a estudada é encontrada no Aprendizado Multitarefa [18], onde são construídas simultaneamente redes neurais para problemas relacionados pelo compartilhamento de informações entre os modelos. Li et al. [19] e Xu et al. [20] ilustram a eficácia desse método na classificação de imagens, porém com técnicas e escopos distintos em relação à presente pesquisa.

Dessa forma, torna-se evidente a presença de estudos que tratam da classificação de dígitos manuscritos, incluindo cenários com dados ruidosos. Contudo, a abordagem explorada pela pesquisa apresentada neste documento é destacada como uma característica distintiva, diferenciando-se de trabalhos em Aprendizado Multitarefa devido às técnicas e problemas examinados.

3 MATERIAIS E MÉTODOS

Conforme a classificação proposta por Wazlawick [21], a pesquisa em questão é categorizada como de natureza primária e objetivos exploratórios. No que diz respeito aos procedimentos, além da

condução de uma revisão bibliográfica, destaca-se a aplicação do método experimental. A pesquisa experimental é baseada na execução e na análise de resultados obtidos com quatro experimentações, sendo elas resumidas da seguinte maneira:

- Experimento I: Implementar e avaliar o desempenho de uma rede neural treinada em imagens sem ruído para resolução do problema de classificação de imagens de dígitos manuscritos;
- Experimento II: Analisar o desempenho do classificador de imagens obtido no Experimento I ao ser utilizado para classificar imagens ruidosas;
- Experimento III: Investigar o desempenho do classificador do Experimento I ao ser aplicado em imagens ruidosas após serem pré-processadas por redes neurais de remoção de ruído;
- Experimento IV: Identificar se alguma modificação nos diferentes modelos de remoção de ruído, adaptados a partir do Experimento III, melhora o desempenho na classificação de imagens. Este experimento explora o objetivo geral dessa pesquisa.

Para conduzir os experimentos, são implementados dois modelos de redes neurais — um classificador de imagens e um removedor de ruído — ambos baseados no modelo MLP. Esse modelo é caracterizado por camadas interligadas de neurônios artificiais, sendo capaz de aproximar funções arbitrárias [2, 3, 22]. A sua escolha é justificada pelos seguintes motivos: (I) Ele é reconhecido como um dos modelos mais fundamentais em Aprendizado Profundo [3, 22]; (II) Sua aplicabilidade em problemas de classificação é comprovada, assim como seu uso no desenvolvimento de redes *autoencoders* [23].

Para o desenvolvimento das redes neurais, são utilizadas imagens do conjunto de dados MNIST, composto por 70 mil imagens (60 mil para treino e 10 mil para validação) de dígitos manuscritos e seus respectivos rótulos de classificação. Cada imagem possui dimensões 28x28, está em escala de cinza e representa um único dígito manuscrito (de 0 a 9) (Fig. 1). A sua escolha é feita pelos dados serem públicos¹ e serem utilizados em pesquisas na área de Inteligência Artificial.

Além da necessidade dos dados, um modelo de ruído é utilizado para gerar imagens ruidosas. Com essa finalidade destaca-se o modelo de Ruído Gaussiano Estacionário Aditivo. A escolha é respaldada por: (I) ser um modelo utilizado em diversas pesquisas [24]; (II) situações reais de corrupção, como má iluminação, alta temperatura e interferência de circuito eletrônico são simuladas pelo modelo [25].

A métrica de avaliação de resultados utilizada é a acurácia da classificação de imagens. Sua escolha é feita por ser adequada a problemas de classificação [3] e por ser considerada mais intuitiva em relação a outras métricas [23]. Além disso, considerando a relativa homogeneidade da distribuição de classes do conjunto MNIST, a acurácia já é capaz de validar as experimentações realizadas.

Por fim, a condução dos experimentos e o desenvolvimento dos códigos são realizados com a linguagem Python (3.11.5); escolhida pela sua popularidade como ferramenta e pela abundância de material para embasamento e consulta. Cabe destacar que não são utilizadas bibliotecas como PyTorch e Tensorflow, pois a adaptação

¹Disponível em: <http://yann.lecun.com/exdb/mnist/>.

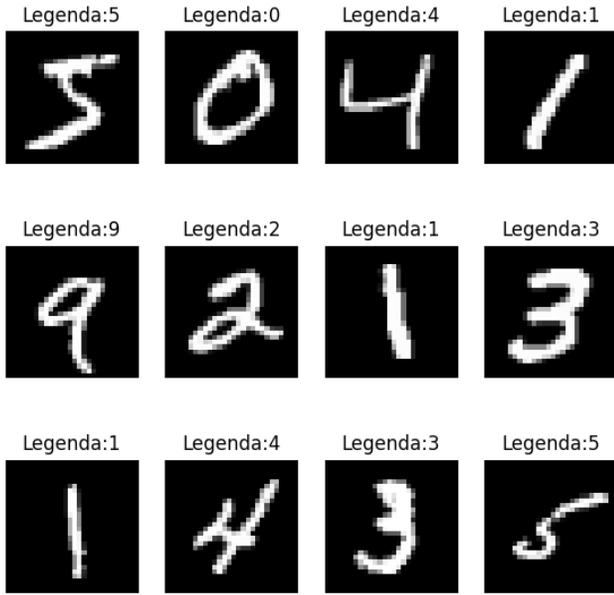


Figura 1: Exemplos de imagens pertencentes ao conjunto de dados MNIST e suas respectivas legendas.

de uma MLP no Experimento IV é facilitada por meio da implementação própria do modelo.

4 PROJETO E RESULTADO DOS EXPERIMENTOS

Dá-se sequência ao trabalho apresentando a pesquisa experimental conduzida. Cada experimento é descrito e seguido pela análise de seus resultados. É crucial destacar que a execução dessa pesquisa segue uma abordagem linear, onde a realização de um experimento depende do experimento anterior.

4.1 Experimento I

O experimento implementa e avalia o desempenho de uma rede neural MLP na resolução do problema de classificação de dígitos manuscritos do conjunto MNIST. A arquitetura da rede neural desenvolvida é baseada nos modelos apresentados por Nielsen [12] e Ekman [23], consistindo em uma rede MLP *full connected* [22] com duas camadas de neurônios (Fig. 2). A entrada da rede é um vetor de 785 sinais, onde o primeiro sinal refere-se ao cálculo do *bias*² e os demais estão associados aos *pixels* da imagem de entrada, que são normalizados para pertencerem ao intervalo de 0 a 1.

A primeira camada de processamento possui 25 neurônios com a função de ativação tangente hiperbólica (*tanh*), enquanto a segunda (e última) camada é composta por 10 neurônios utilizando a função de ativação sigmóide logística (*logsig*). Portanto, as saídas da rede neural são associadas às probabilidades de classificação de cada

²Parâmetro de neurônios artificiais. Representa uma transformação afim no processamento do neurônio artificial.

dígito. Sendo assim, é destacado que os rótulos do conjunto MNIST são codificados com *one-hot-encoding*³.

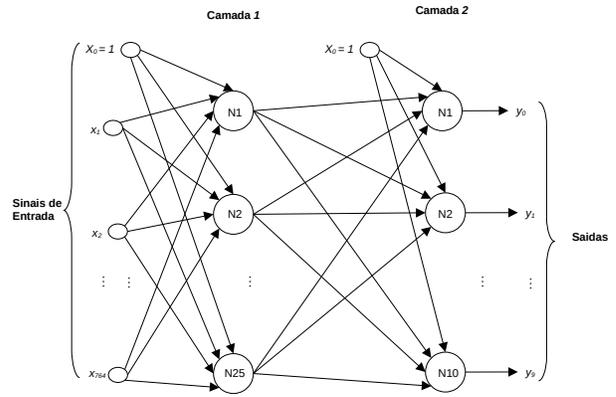


Figura 2: Arquitetura do modelo de classificação desenvolvido.

Sobre o seu processo de aprendizagem, o treinamento é realizado com o algoritmo *backpropagation* estocástico e a função de erro *Mean Squared Error* (MSE) aproximada para cada entrada. Adotando a configuração proposta por Ekman [23], a taxa de aprendizado é estabelecida em 0,01. Por fim, o treino do modelo é efetuado com a passagem de 10 épocas⁴, consideradas suficientes para produzir resultados satisfatórios e reduzir o tempo despendido no processo de treino.

Por fim, a Tabela 1 exibe as acurácias de treino e validação do modelo de classificação para as épocas 1, 5 e 10, visando destacar a evolução do desempenho da rede neural conforme treinamento. Percebe-se que em uma única época, o desempenho em ambos os cenários é satisfatório e supera a taxa de 10% de uma predição aleatória. Ademais, o aumento de épocas está associado ao incremento das classificações corretas observadas no treino e na validação, indicando que não houve *overfitting*⁵.

Época	Acurácia média Treino (%)	Acurácia média Validação (%)
1	85,63	91,57
5	93,56	92,97
10	94,57	93,61

Tabela 1: Acurácia (treino e validação) de classificação em relação às épocas 1, 5 e 10.

³Codifica um valor em um vetor atribuindo o valor 1 para apenas um componente (componente excitado) e o restante com o valor 0 [23]. Valores são distinguidos pela diferença do componente excitado.

⁴Indica quantas vezes a rede neural é treinada no conjunto de dados de treino.

⁵Ocorre quando o modelo apresenta desempenho satisfatório na fase de treino, mas insatisfatório na validação. Identifica que a rede neural está memorizando os dados de treino e não generalizando o problema para o qual foi desenvolvida [23].

4.2 Experimento II

O experimento tem como objetivo examinar a influência do ruído introduzido nas imagens de entrada durante a validação da rede neural para classificação de imagens. Portanto, utilizando o modelo treinado no Experimento I, diferentes versões do conjunto de validação MNIST são criadas. Cada versão compreende a introdução de ruído em diferentes níveis geradas pelo modelo de Ruído Gaussiano Estacionário Aditivo.

A variação nos níveis de ruído é realizada mediante a alteração do desvio padrão (std), um parâmetro utilizado pelo mecanismo gerador de ruído. Para o experimento, são estabelecidas algumas configurações de desvio padrão, pertencentes ao intervalo de 0,1 a 1,0. À medida que essa configuração aumenta, a degradação na imagem é intensificada, dificultando o reconhecimento da informação armazenada (Fig. 3).

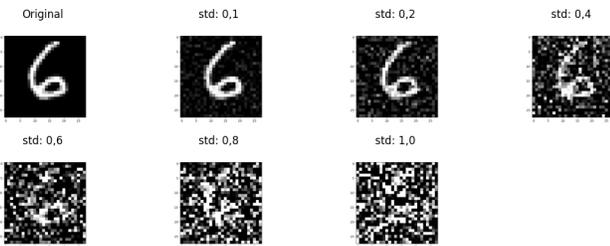


Figura 3: Imagem (dígito 6) do conjunto de validação para diferentes níveis de ruído.

Como observação destaca-se que o vetor de dados representando a imagem ruidosa é obtido pela soma do vetor original com o vetor gerado pelo modelo de ruído. Posteriormente, é aplicada uma restrição nos valores do vetor resultante para pontos flutuantes, mantendo-os no intervalo [0, 1]. Essa prática preserva o padrão de entradas aceito pelo classificador de imagens.

Concluindo, a Tabela 2 fornece a média da acurácia obtida em 10 etapas de validação para cada conjunto de dados ruidosos. O cálculo da média é essencial, uma vez que os valores de ruído são distribuídos aleatoriamente, gerando conjuntos distintos para cada validação. Assim, é importante destacar que o aumento no nível de ruído está diretamente associado à diminuição da acurácia em relação ao valor base obtido no experimento anterior.

Varição do nível de ruído (std)	Acurácia média de classificação (%)
0,1	88,18
0,2	81,66
0,4	61,38
0,6	44,58
0,8	33,72
1,0	27,02

Tabela 2: Acurácia média de classificação em relação a diferentes níveis de ruído.

4.3 Experimento III

Este experimento avalia o desempenho na classificação de imagens ruidosas submetidas à etapa de pré-processamento, efetuada por uma rede neural. Para cada versão dos conjuntos de validação ruidosos obtidos no Experimento II, é treinado um removedor de ruído (de mesma configuração inicial) projetado especificamente para detectar e eliminar aquele nível específico de ruído nas imagens.

A rede neural para essa tarefa consiste de uma MLP estruturada como um *autoencoder* — que é caracterizado pela capacidade de representar uma função identidade através da codificação e da decodificação de informações [8, 22, 23]. Para a remoção de ruído, o modelo é adaptado; em vez de apresentar o dado original como entrada, é fornecida uma versão contendo ruído. Dessa forma, a rede neural é otimizada para minimizar a diferença entre um vetor ruidoso e o mesmo vetor sem ruído, resultando no aprendizado do processo para a remoção de ruído [8].

Os modelos *autoencoders* utilizados nessa pesquisa são baseados em uma única arquitetura adaptada de Ekman [23] (Fig. 4). A rede é composta por duas camadas de neurônios, recebendo 785 sinais de entrada (assim como o modelo de classificação) que são direcionados a 64 neurônios, uma configuração arbitrária. O espaço latente é, portanto, o vetor composto pelos potenciais de ativação desses neurônios. Na última camada, são utilizados 784 neurônios para garantir que a saída tenha a mesma dimensão da imagem de entrada.

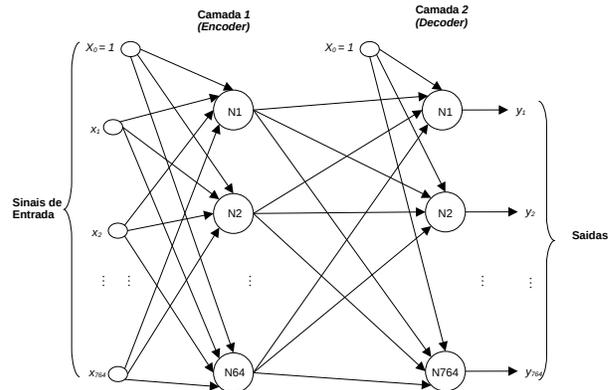


Figura 4: Arquitetura *autoencoder* MLP implementada pelos modelos de remoção de ruído.

As demais configurações, como funções de ativação, taxa de aprendizado, algoritmo de aprendizagem, função de erro e conjunto de dados, seguem as definições do modelo de classificação. A título de exemplificação, a Fig. 5 destaca imagens ruidosas após o processamento pelos removedores de ruído.

Nos resultados, de maneira semelhante ao Experimento II, a avaliação da acurácia de classificação é realizada em relação à média obtida em 10 etapas de validação para cada conjunto de dados ruidosos. Considerando isso, os resultados são apresentados na Tabela 3. Além disso, são destacadas as acurácias considerando as redes neurais após o treino em 1 e 10 épocas; como no Experimento I, a informação é usada para destacar a evolução do desempenho da rede neural conforme treinamento.

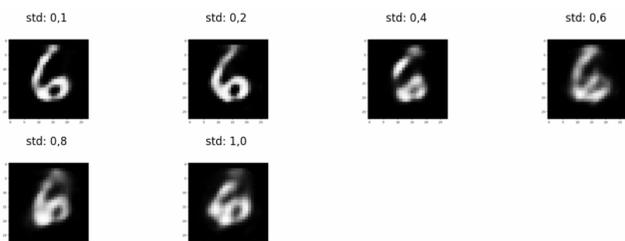


Figura 5: Saída dos removedores de ruído (10 épocas) para entrada ruidosa referente ao dígito 6.

Em termos de resultados, primeiramente destaca-se que, na maioria dos casos, um maior número de épocas de treino resulta em taxas de classificação superiores. Por fim, são apresentados os cenários menos favoráveis e mais favoráveis para a aplicação do pré-processamento na classificação. Para ambos os casos, o pior cenário é identificado para um nível de ruído de 0,1. Quanto aos cenários mais favoráveis, para a passagem de 1 época, destaca-se o nível de ruído de 0,8, que representa um aumento de 19,14% na acurácia de classificação em relação ao Experimento II. Já para 10 épocas, menciona-se o nível de ruído de 1,0, que representa um incremento de 23,84% na acurácia de classificação em comparação ao segundo experimento.

Variação do nível de ruído (std)	Acurácia média de classificação (%) Removedor de ruído 1 época	Acurácia média de classificação (%) Removedor de ruído 10 épocas
0,1	76,92	86,73
0,2	73,71	84,76
0,4	61,69	77,47
0,6	55,89	63,71
0,8	52,85	54,58
1,0	39,70	50,85

Tabela 3: Acurácia média de classificação em relação a imagens ruidosas pré-processadas pelos respectivos modelos de remoção de ruído treinados em 1 e 10 épocas.

4.4 Experimento IV

O experimento final investiga a viabilidade de empregar o processamento do classificador de imagens para auxiliar o processo de aprendizagem dos modelos de remoção de ruído. Essa abordagem é inspirada em trabalhos do campo de pesquisa de Aprendizado Multitarefa, como Liao et al. [26] e Teichmann et al. [27]. No entanto, é importante ressaltar que a pesquisa em questão não aplica o treinamento simultâneo de duas redes neurais.

A adaptação proposta explora o classificador de imagens previamente treinado, concentrando-se em utilizar o resultado de seu processamento para o treinamento de modelos de remoção de ruído. Com isso, é esperado que as imagens pré-processadas conduzam a melhores taxas de classificações corretas. Em termos técnicos, o

experimento utiliza o valor do erro aproximado $E^c[k]$ do classificador de imagens juntamente com o erro aproximado $E^a[k]$ do removedor de ruído no algoritmo *backpropagation* para as redes *autoencoders*.

Matematicamente é possível destacar que para as redes neurais desenvolvidas nessa pesquisa — usando o erro MSE aproximado e a função de ativação *logsig* —, para cada neurônio i da última camada é calculado um termo de erro [23] definido pela Equação 1

$$eNi = (-y_i + \hat{y}_i) \cdot \text{logsig}'(v_i) \quad (1)$$

onde y_i , \hat{y}_i e v_i são respectivamente a saída esperada, a saída obtida e o potencial de ativação do neurônio i . Esse valor é utilizado durante o algoritmo *backpropagation* para treinar o modelo. Sendo assim, a alteração explorada para introduzir a medida de erro da classificação $E^c[k]$ é definida pela Equação 2

$$eNi = (\varphi(-y_i + \hat{y}_i) + \omega E^c[k]) \cdot \text{logsig}'(v_i) \quad (2)$$

onde φ e ω representam duas configurações que devem ser determinadas durante o projeto da rede neural.

Assim, o experimento envolve aplicar a adaptação do algoritmo de treino variando as configurações supramencionadas, reproduzindo algo similar ao Experimento III. No entanto, em vez de treinar os modelos de remoção de ruído em 10 épocas, o processo de aprendizagem é realizado em apenas uma época; restrição imposta para limitar o tempo gasto com treino, uma vez que para cada par de φ e ω , são desenvolvidos 6 *autoencoders* (um para cada nível de ruído estudado nessa pesquisa).

Finalizando a descrição do experimento, o conjunto C de valores φ e ω é dado pelos pares apresentados pela Equação 3. A título de especificação, essas configurações são escolhidas arbitrariamente por meio de experimentação.

$$C = \left\{ \begin{array}{l} (+1, 0; -1, 0), (+0, 8; -0, 2), (+0, 5; -0, 5), \\ (+1, 0; -0, 5), (+1, 0; +1, 0), (+0, 5; 0, 5), \\ (+1, 0; +0, 5), (+0, 8; +0, 2) \end{array} \right\} \quad (3)$$

Como resultados, a Tabela 4 apresenta as médias de 10 validações da taxa de classificações corretas para cada configuração de valores (φ, ω) nos diferentes níveis de ruído explorados nessa pesquisa. Inicialmente, destaca-se que a variação dos valores utilizados resulta em diferentes desempenhos de classificação. Comparando os resultados com as acurácias apresentadas na Tabela 3 para redes *autoencoders* treinadas em uma época, observa-se o seguinte:

- O par $(\varphi = 1, \omega = -0, 5)$ apresentou os maiores incrementos em relação ao modelo padrão. No pior caso, houve um ganho de aproximadamente 5,7% na acurácia para o nível de ruído 0,1. No melhor caso, a acurácia foi incrementada em mais de 14% (nível de ruído 0,4).
- O par $(\varphi = 0, 5, \omega = 0, 5)$ demonstrou os maiores decrementos em relação ao modelo padrão. No pior caso, houve uma redução de aproximadamente 46% na acurácia para o nível de ruído 0,6. No melhor caso, a acurácia foi reduzida em 13% (nível de ruído 0,2).

Por fim, uma breve comparação é feita entre o melhor desempenho identificado na Tabela 4 $(\varphi = 1, \omega = -0, 5)$ e os desempenhos obtidos ao aplicar os removedores de ruído treinados em 10 épocas

Configuração (ϕ, ω)	Acurácia média de classificação (%)					
	0,1	0,2	0,4	0,6	0,8	1,0
($\phi = +1, 0, \omega = -1, 0$)	75,80	73,96	68,29	66,31	52,64	41,46
($\phi = +0, 8, \omega = -0, 2$)	80,38	79,62	65,84	61,36	52,40	42,91
($\phi = +0, 5, \omega = -0, 5$)	70,05	74,12	69,86	68,78	54,77	42,14
($\phi = +1, 0, \omega = -0, 5$)	82,66	80,82	75,75	67,75	60,69	48,26
($\phi = +1, 0, \omega = +1, 0$)	66,17	63,33	26,90	15,07	28,58	15,91
($\phi = +0, 5, \omega = +0, 5$)	57,30	60,60	42,93	9,74	24,91	9,74
($\phi = +1, 0, \omega = +0, 5$)	69,82	70,55	51,29	33,60	29,64	28,24
($\phi = +0, 8, \omega = +0, 2$)	72,49	72,57	54,69	46,16	39,24	34,13

Tabela 4: Acurácia média de classificação de imagens (com diferentes níveis de ruído) pré-processadas pelos respectivos modelos de remoção de ruído treinados com as configurações (ϕ, ω).

(Tabela 3). O modelo de 10 épocas resulta em acurácias superiores às obtidas com o modelo adaptado de 1 época nos níveis de ruído 0,1, 0,2, 0,4 e 1,0. Nos casos restantes, o modelo adaptado resulta em métricas melhores; por exemplo, para o nível de ruído 0,8, observa-se uma diferença positiva de 7% na acurácia.

5 CONCLUSÃO

O presente trabalho conduziu uma investigação sobre o problema de classificação de imagens de dígitos manuscritos, empregando técnicas de Aprendizado Profundo. Além da avaliação do desempenho da rede neural para essa tarefa específica, foram propostos cenários adversos pela introdução de ruído nos dados, visando analisar o comportamento do modelo nessa situação. Por fim, foi explorada a aplicação de pré-processamento dos dados por meio de redes neurais *autoencoder*, com o intuito de destacar melhorias no desempenho do modelo de classificação.

Dentre os resultados obtidos, destacam-se as seguintes conclusões: (I) A introdução de ruído nas imagens de entrada na rede de classificação não robusta a esses ruídos resulta na redução da sua taxa de acertos; (II) A utilização de modelos de pré-processamento demonstra um aumento na taxa de sucesso da rede neural de classificação; (III) A inclusão de informações do modelo de classificação no processo de aprendizado dos modelos de remoção de ruídos indicam melhoria na taxa de classificações corretas.

O último resultado mencionado representa o diferencial dessa pesquisa. Os dados empíricos obtidos indicam a existência de um caminho promissor a ser explorado, que pode aprimorar a eficiência de modelos relacionados a redes neurais. Além da obtenção de taxas de classificação superiores, a adaptação do algoritmo *backpropagation* explorada sugere a redução das épocas necessárias para alcançar resultados finais comparáveis aos obtidos com o treino em mais épocas com o algoritmo original.

Por fim, para pesquisas futuras, destaca-se a necessidade de reproduzir e expandir os experimentos realizados neste trabalho, buscando validar e formalizar os indícios observados. Seria interessante estudar os mesmos cenários, mas utilizando modelos treinados (principalmente os removedores de ruído adaptados) em mais épocas. Além disso, realizar um teste *post-hoc* entre classificadores distintos pode expandir os resultados e a análise realizada. Por fim,

um estudo formal do algoritmo explorado é crucial para compreender suas capacidades de forma analítica e explicar os resultados obtidos empiricamente.

REFERÊNCIAS

- [1] Richard Szeliski. *Computer Vision*. Springer, London, England, 2010.
- [2] Martin T. Hagan, Howard B. Demuth, Mark H. Beale, and Orlando De Jesús. *Neural Network Design (2nd Edition)*. Martin Hagan, 2 edition, 2014.
- [3] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. MIT Press, London, England, 2016.
- [4] Tiago S. Nazaré, Gabriel B. Paranhos da Costa, Welinton A. Contato, and Moacir Ponti. *Deep convolutional neural networks and noisy images*, volume 10657, page 416–424. Springer International Publishing, 2 2018.
- [5] Samuel Dodge and Lina Karam. Understanding how image quality affects deep neural networks. In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, pages 1–6. IEEE, 2016. doi: 10.1109/QoMEX.2016.7498955.
- [6] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. A review of image denoising algorithms, with a new one. *Multiscale modeling and simulation*, 4(2): 490–530, 2005. doi: 10.1137/040616024. URL <http://dx.doi.org/10.1137/040616024>.
- [7] Shuhang Gu and Radu Timofte. A brief review of image denoising algorithms and beyond. In *Inpainting and Denoising Challenges*, page 1–21. Springer International Publishing, 2019.
- [8] Pascal Vincent, Hugo Larochelle, Yoshua Bengio, and Pierre-Antoine Manzagol. Extracting and composing robust features with denoising autoencoders. In *International Conference on Machine Learning*, page 1096–1103, New York, NY, USA, 2008. Association for Computing Machinery. doi: 10.1145/1390156.1390294. URL <https://doi.org/10.1145/1390156.1390294>.
- [9] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, and Pierre-Antoine Manzagol. Stacked denoising autoencoders. *J. Mach. Learn. Res.*, 11: 3371–3408, 12 2010.
- [10] Cheng-Lin Liu, Kazuki Nakashima, Hiroshi Sako, and Hiromichi Fujisawa. Handwritten digit recognition. *Pattern recognition*, 36(10):2271–2285, 2003. doi: 10.1016/S0031-3203(03)00085-2. URL [http://dx.doi.org/10.1016/S0031-3203\(03\)00085-2](http://dx.doi.org/10.1016/S0031-3203(03)00085-2).
- [11] Alejandro Baldominos, Yago Saez, and Pedro Isasi. A survey of handwritten character recognition with mnist and emnist. *Applied sciences*, 9(15):16, 8 2019. doi: 10.3390/app9153169. URL <http://dx.doi.org/10.3390/app9153169>.
- [12] Michael A. Nielsen. *Neural Networks and Deep Learning*. Determination Press, 2015. URL <https://books.google.com.br/books?id=STDBswEACAAJ>.
- [13] L. Bottou, C. Cortes, J. S. Denker, H. Drucker, I. Guyon, L. D. Jackel, Y. LeCun, U. A. Muller, E. Sackinger, P. Simard, and V. Vapnik. Comparison of classifier methods. In *IAPR*, volume 3, pages 77–82. IEEE, 1994.
- [14] Yiren Zhou, Sibong Song, and Ngai-Man Cheung. On classification of distorted images with deep convolutional neural networks. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2017.
- [15] Mohammad Momeny, Ali Mohammad Latif, Mehdi Agha Sarram, Razieh Sheikhpour, and Yu Dong Zhang. A noise robust convolutional neural network for image classification. *Results in engineering*, 10:100225, 2021. doi: 10.1016/j.rineng.2021.100225. URL <http://dx.doi.org/10.1016/j.rineng.2021.100225>.
- [16] Sudipta Singha Roy, Mahtab Ahmed, and Muhammad Aminul Haque Akhand. Classification of massive noisy image using auto-encoders and convolutional neural network. In *International Conference on Information Technology (ICIT)*, pages 971–979. IEEE, 10 2017.
- [17] Sudipta Singha Roy, Mahtab Ahmed, and Muhammad Aminul Haque Akhand. Noisy image classification using hybrid deep learning methods. *Journal of*

- Information and Communication Technology*, 17(2), 4 2018. doi: 10.32890/jict2018.17.2.8253. URL <http://dx.doi.org/10.32890/jict2018.17.2.8253>.
- [18] Rich Caruana. Multitask learning. *Machine learning*, 28(1):41–75, 1997. doi: 10.1023/a:1007379606734. URL <http://dx.doi.org/10.1023/a:1007379606734>.
- [19] Ya Li, Xinmei Tian, Xu Shen, and Dacheng Tao. Classification and representation joint learning via deep networks. In *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence*, California, 2017. International Joint Conferences on Artificial Intelligence Organization.
- [20] Qiang Xu, Yu Zeng, Wenjun Tang, Wei Peng, Tingwei Xia, Zongrun Li, Fei Teng, Weihong Li, and Jinhong Guo. Multi-task joint learning model for segmenting and classifying tongue images using a deep neural network. *Journal of biomedical and health informatics*, 24(9):2481–2489, 2020. doi: 10.1109/JBHI.2020.2986376. URL <http://dx.doi.org/10.1109/JBHI.2020.2986376>.
- [21] Raul Sidnei Wazlawick. *Metodologia de Pesquisa para Ciência da Computação*. GEN LTC, Rio de Janeiro, RJ, 3 edition, 2020.
- [22] Simon O. Haykin. *Neural Networks and Learning Machines*. Pearson, Upper Saddle River, NJ, 3 edition, 2008.
- [23] Magnus Ekman. *Learning deep learning*. Addison Wesley, Boston, MA, 1 edition, 2021.
- [24] David A. Forsyth and Jean Ponce. *Computer Vision*. Pearson, Upper Saddle River, NJ, 2 edition, 2011.
- [25] Philippe Cattin. Image restoration, 2016. URL <https://docplayer.net/85966756-Image-restoration-introduction-to-signal-and-image-processing-prof-dr-philippe-cattin-miac-university-of-basel-april-19th-26th-2016.html>.
- [26] Yiyi Liao, Sarath Kodagoda, Yue Wang, Lei Shi, and Yong Liu. Understand scene categories by objects. In *International Conference on Robotics and Automation (ICRA)*, pages 2215–2221. IEEE, 2016.
- [27] Marvin Teichmann, Michael Weber, Marius Zollner, Roberto Cipolla, and Raquel Urtasun. Multinet. In *2018 IEEE Intelligent Vehicles Symposium*, pages 1013–1020. IEEE, 2018.