

---

# Avaliação do Modelo BTHOWeN em Datasets Multiclasse

---

Breno Tostes      Eduardo Naslausky      Felipe Barrocas  
Giordano Souza      Maria Bianca Irace      Miguel Sousa  
Rafael Paladini

COPPE, Universidade Federal do Rio de Janeiro

## Abstract

Este trabalho avalia o modelo BTHOWeN (Bleached Thermometer-encoded Hashed-input Optimized Weightless Neural Network) em diversos datasets multiclasse. Comparamos a performance do BTHOWeN com o modelo BloomWiSARD, analisando o impacto dos hiperparâmetros na acurácia final. Utilizamos uma abordagem gulosa para otimização, variando sequencialmente os parâmetros de tamanho de endereço, número de discriminadores, funções hash, filtros Bloom e fator de bleaching. Os resultados demonstram as vantagens do BTHOWeN em termos de acurácia e eficiência para diversos conjuntos de dados de classificação.

## 1 Introdução

BTHOWeN vem de Bleached Thermometer-encoded Hashed-input Optimized Weightless Neural Network, e como o nome indica, é uma arquitetura de rede neural sem peso que se diferencia do uso do modelo WiSARD (Wilkes, Stonham and Aleksander Recognition Device) [1] por:

- Incorporar counting Bloom filters para reduzir o tamanho do modelo e permitir bleaching [4].
- Usar função hash que não requer operações aritméticas [5].
- Fazer encoding com termômetro não-linear de distribuição normal para melhorar a acurácia do modelo [5, 4].

## 2 Proposta

### 2.1 Metodologia

Neste trabalho, iremos replicar o BTHOWeN utilizando o BloomWisard como cerne, com os diferentes datasets multiclasse inclusos no artigo original. A proporção entre a massa de treinamento e a massa de teste foi mantida idêntica às respectivas proporções originais, porém os hiperparâmetros foram configurados para verificar o impacto de cada um deles na acurácia final.

Os hiperparâmetros configuráveis são:

- Tamanho do endereço (ou tamanho da tupla).
- Número de discriminadores.
- Número de funções hash.

- Número de filtros de bloom.
- Fator do bleaching.

Os experimentos são feitos alterando um hiperparâmetro de cada vez, de maneira gulosa. Ao atingir um valor máximo variando apenas um hiperparâmetro, este tem seu valor mantido pelo resto do experimento, e o próximo hiperparâmetro passa a ser o variável. Ao realizar este procedimento com todos os hiperparâmetros, identificamos o melhor resultado.

Tomamos o maior valor de acurácia de todos os experimentos e sua configuração de hiperparâmetros como melhor valor obtido e o comparamos com a acurácia obtida no artigo original.

## 2.2 Hiperparâmetros

### 2.2.1 Tamanho do Endereço (Tamanho da Tupla)

O tamanho do endereço, também conhecido como tamanho da tupla, é um parâmetro da WiSARD que determina a quantidade de bits de entrada que são agrupados para formar um endereço para cada RAM no discriminador [1]. No contexto do BTHOWeN, o tamanho do endereço define quantos bits são agrupados para endereçar cada filtro de Bloom, afeta o número total de filtros necessários (número total de entradas dividido pelo tamanho da tupla) e determina o espaço de endereçamento para cada filtro ( $2^{\text{tamanho da tupla}}$  possíveis endereços em WiSARD tradicional) [3].

Tuplas menores resultam em mais filtros e melhor generalização, enquanto tuplas maiores reduzem o número de filtros, mas podem afetar a capacidade de generalização [4]. Nos experimentos, este parâmetro varia conforme o dataset, desde valores pequenos (2 para Iris) até valores maiores (28 para MNIST).

### 2.2.2 Número de Discriminadores (OWeN)

O número de discriminadores no BTHOWeN está relacionado à estrutura de entrada do modelo, em que discriminador é responsável por processar uma parte específica do vetor de entrada, segmentando os dados de entrada em subconjuntos de bits [1, 4].

O número de discriminadores para cada classe é determinado pela divisão da dimensão total da entrada pelo número de bits por filtro [4]. Por exemplo, em um modelo para MNIST com entrada de 784 pixels (28x28) e 28 bits por filtro, teríamos 28 discriminadores por classe.

O valor do parâmetro OWeN representa o número de bits de entrada que cada discriminador recebe [4, 5]. Quanto maior este valor, mais granular é a segmentação dos dados, o que pode melhorar a capacidade de distinção, mas aumenta a complexidade computacional.

### 2.2.3 Número de Funções Hash (FH)

As funções hash são utilizadas nos filtros de Bloom para mapear os dados de entrada em posições de memória [4]. O parâmetro FH determina quantas diferentes funções hash são utilizadas em cada filtro de Bloom.

BTHOWeN implementa a família de funções hash H3, conforme descrita por Carter e Wegman, que não requer operações aritméticas complexas, sendo ideal para implementações em hardware [5].

O número de funções hash tem uma relação complexa com a taxa de falsos positivos [4]:

- Valores típicos variam de 1 a 4, dependendo do dataset
- Um maior número de funções hash pode aumentar o poder discriminativo do modelo, mas também eleva o custo computacional
- Os melhores resultados obtidos nos experimentos utilizaram de 1 a 4 funções hash, dependendo da complexidade da tarefa

### 2.2.4 Número de Filtros de Bloom (FE - Filter Entries)

O parâmetro FE define o número de entradas em cada filtro de Bloom, ou seja, o tamanho do vetor utilizado para armazenar os padrões aprendidos [4]. Este tamanho é uma potência de 2 (por exemplo, 128, 256, 512, 1024, 2048, etc.).

Aumentar o tamanho do filtro reduz a probabilidade de colisões e, consequentemente, a taxa de falsos positivos, mas também aumenta a memória necessária [4, 5]. A escolha deste parâmetro afeta diretamente o equilíbrio entre precisão e eficiência de memória do modelo.

Em hardware, este parâmetro se traduz na quantidade de memória alocada para cada filtro, influenciando o consumo de energia [5].

### 2.2.5 Fator de Bleaching (b)

O fator de bleaching é um hiperparâmetro introduzido pela arquitetura BTHOWeN, que não existia em outras redes neurais sem peso anteriores [4]. Em implementações tradicionais de filtros de Bloom, cada posição de memória armazena apenas um bit (0 ou 1), indicando se um padrão foi visto ou não [4, 2].

Na BTHOWeN, cada posição armazena um contador, permitindo registrar quantas vezes um determinado padrão foi encontrado durante o treinamento [4]. O fator de bleaching define o limiar mínimo de ocorrências para que um padrão seja considerado válido durante a inferência.

Funcionamento [4, 5]:

- Durante o treinamento, os contadores são incrementados cada vez que um padrão é encontrado
- Na inferência, um padrão é considerado presente apenas se seu contador for  $\geq b$
- Aumentar o valor de b pode melhorar a precisão ao reduzir falsos positivos
- Valores comuns de b variam de 1 a 10, sendo o valor ótimo determinado durante o treinamento

A técnica de bleaching permite ao modelo distinguir padrões frequentes e relevantes em detrimento de ocorrências aleatórias e ruidosas, melhorando a capacidade de generalização do modelo [4].

## 2.3 Datasets

### 2.3.1 MNIST

O dataset **MNIST** (Modified National Institute of Standards and Technology) é uma coleção de dígitos manuscritos. Ele inclui **60k imagens de treinamento** e **10k imagens de teste**. Todas as imagens estão em escala de cinza e possuem tamanho de **28×28 pixels**.

### 2.3.2 Ecoli

O dataset **Ecoli** é usado para prever onde proteínas celulares se localizam com base em suas sequências de aminoácidos. Ele contém **336 proteínas**, cada uma descrita por **sete atributos numéricos** derivados da sequência. As proteínas são classificadas em **oito possíveis locais celulares**.

### 2.3.3 Iris

O dataset **Iris** contém **150 observações** de flores de íris, cada uma descrita por:

- Comprimento da sépala.
- Largura da sépala.
- Comprimento da pétala.
- Largura da pétala.

A classificação é feita em uma de **três espécies**: **Iris Setosa**, **Versicolor** ou **Virginica**.

#### 2.3.4 Glass

O dataset **Glass** contém **214 instâncias** de fragmentos de vidro, cada uma descrita por **10 atributos**. Com ele, conseguimos prever o tipo de vidro com base em sua composição química e índice de refração.

#### 2.3.5 Letter

O dataset **Letter** contém letras manuscritas. As imagens dos caracteres foram baseadas em **20 fontes diferentes**, e cada letra dentro dessas fontes foi distorcida aleatoriamente para produzir **20k entradas**, onde **16k** foram usadas para treinamento e **4k** para teste.

#### 2.3.6 Wine

O dataset **Wine** reúne **178 amostras** de vinho de três cultivares de uvas da região de Piemonte, Itália. Cada amostra é descrita por **13 atributos químicos contínuos** – como teor de álcool, magnésio e intensidade de cor. O objetivo é identificar a cultivar correta entre as **três classes** disponíveis.

#### 2.3.7 Segment

O dataset **Image Segmentation** (Segment) contém **2.310 segmentos** de imagens externas, distribuídos igualmente entre **sete classes** de região: tijolo, céu, folhagem, cimento, janela, caminho e grama. Cada segmento é representado por **19 atributos numéricos** que descrevem características de cor e textura. A tarefa é prever a classe da região na qual o segmento se enquadra.

#### 2.3.8 Shuttle

O dataset **Statlog Shuttle** possui cerca de **58.000 registros** de telemetria do sistema de controle do ônibus espacial. Cada registro é descrito por **nove atributos numéricos** derivados de sensores a bordo e está rotulado em um de **sete possíveis estados operacionais**. O conjunto é fortemente desbalanceado, com predominância da classe 1.

#### 2.3.9 Vehicle

O dataset **Vehicle Silhouettes** contém **846 silhuetas** de veículos, cada uma descrita por **18 medidas geométricas** extraídas da imagem, como área, compacidade e momentos. Os veículos devem ser classificados em uma de **quatro categorias**: ônibus, Opel, Saab ou van.

#### 2.3.10 Vowel

O dataset **Vowel Recognition** inclui **990 amostras** de fala gravadas por 15 locutores. Cada amostra é caracterizada por **10 atributos acústicos contínuos** (coeficientes derivados do espectro) e deve ser classificada em uma de **11 vogais** do inglês, como /a/, /e/ ou /i/.

## 3 Resultados

### 3.1 Resultados por dataset

Nas tabelas a seguir, apresentamos os resultados obtidos com diferentes configurações do BTHOWeN para cada dataset. As configurações são organizadas da seguinte forma:

- **Bloom WiSARD**: Resultados obtidos pela implementação original do Bloom WiSARD, apresentados como referência para comparação.
- **Base**: Configuração inicial do BTHOWeN utilizada como ponto de partida para os experimentos, com parâmetros utilizados nos experimentos originais publicados pelo autor.

- **Var. X:** Variações da configuração base, onde modificamos sistematicamente um hiperparâmetro por vez, seguindo a abordagem gulosa descrita na metodologia. Os números indicam a sequência das variações testadas.

Para cada configuração, apresentamos os valores dos hiperparâmetros (b, OWeN, FE, FH), a acurácia obtida, o percentual de empates ocorridos durante a classificação e o valor de bleaching que produziu os melhores resultados. A coluna “Bleaching” indica o valor ótimo encontrado durante o treinamento, que permite ao modelo obter a melhor acurácia possível.

### 3.1.1 Tabelas

Tabela 1: Parâmetros e métricas do dataset Iris

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	960	2.0	3
Base	3	2	128	1	980	12.0	2
Var. 1	4	2	128	1	920	16.0	1
Var. 2	3	2	256	1	980	14.0	2
Var. 3	3	2	128	2	980	8.0	2
Var. 4	4	2	128	2	860	4.0	9
Var. 5	3	2	256	2	980	0.0	2
Var. 6	4	2	256	2	900	12.0	3

Tabela 2: Parâmetros e métricas do dataset Ecoli

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	799	N/A	N/A
Base	10	128	2	10	786	8.9	7
Var. 1	4	128	2	11	821	10.7	1
Var. 2	3	256	2	10	813	19.6	1
Var. 3	3	128	3	10	786	15.2	7
Var. 4	4	128	3	11	839	17.9	1
Var. 5	3	256	3	10	848	10.7	1
Var. 6	4	256	4	10	830	13.4	1

Tabela 3: Parâmetros e métricas do dataset Glass

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	726	N/A	N/A
Base	3	128	3	9	577	39.4	1
Var. 1	4	128	3	10	563	38.0	1
Var. 2	3	256	3	9	493	33.8	1
Var. 3	3	128	4	9	549	19.7	4
Var. 4	4	128	4	10	592	40.8	1
Var. 5	3	256	4	9	676	29.6	1
Var. 6	4	256	4	10	676	28.2	1

### 3.2 Gráficos

Nas figuras a seguir, apresentamos gráficos de dispersão que ilustram a relação entre taxa de erro (1 - acurácia) e taxa de empates para diferentes configurações do BTHOWeN em cada dataset. Os pontos são categorizados em:

- **Melhor ponto:** Configuração que obteve o melhor desempenho (azul)
- **Melhor que as bases:** Configurações que superaram a linha base (verde)
- **Demais variações:** Outras configurações testadas (vermelho)

As linhas de referência indicam:

Tabela 4: Parâmetros e métricas do dataset Letter

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	848	N/A	N/A
Base	3	128	3	9	734	18.6	6
Var. 1	4	128	3	10	736	17.4	5
Var. 2	3	256	3	9	789	15.2	3
Var. 3	3	128	4	9	707	19.2	5
Var. 4	4	128	4	10	719	18.7	6
Var. 5	3	256	4	9	775	15.6	4
Var. 6	4	256	5	12	811	11.3	4
Var. 7	11	256	5	18	840	7.6	3
Var. 8	15	256	5	35	884	3.9	3

Tabela 5: Parâmetros e métricas do dataset Wine

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	926	N/A	N/A
Base	9	13	128	3	983	N/A	1
Var. 1.1	10	13	256	4	1.000	1.7	1
Var. 1.2	10	13	256	4	949	3.39	1
Var. 2	11	9	256	4	983	1.69	1
Var. 3	11	13	256	2	966	3.38	1
Var. 4	9	9	256	4	966	1.69	1
Var. 5	10	17	256	4	966	11.8	1
Var. 6	11	15	256	2	966	N/A	N/A
Var. 7	7	9	256	4	932	5.08	1

Tabela 6: Parâmetros e métricas do dataset Segment

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	N/A	N/A	N/A
Base	9	27	1024	2	925	N/A	1
Var. 1	10	16	256	4	938	9.35	1
Var. 2	8	20	512	3	924	9.24	1
Var. 3.1	10	18	1024	3	942	8.57	2
Var. 3.2	10	18	1024	3	944	6.49	2
Var. 4	10	20	512	2	944	8.96	1
Var. 5	16	16	256	3	941	10.8	8
Var. 6	10	14	512	4	939	8.10	2
Var. 7	9	20	1024	2	937	23.7	1
Var. 8	15	15	256	4	936	9.8	1
Var. 9	9	32	2048	4	936	34.9	1

Tabela 7: Parâmetros e métricas do dataset Shuttle

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	868	N/A	N/A
Base	9	27	1024	2	0	0	0
Var. 1.1	11	29	1024	2	999	0.11	1
Var. 1.2	11	29	1024	2	998	0.17	1
Var. 2	11	25	1024	3	999	0.10	1
Var. 3	8	27	1024	1	998	0.21	4
Var. 4	9	23	512	3	998	0.21	8
Var. 5	8	23	2048	1	998	0.70	1
Var. 6	7	27	1024	2	989	2.55	5
Var. 8	11	27	1024	2	976	4.99	276

Tabela 8: Parâmetros e métricas do dataset Vehicle

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	926	N/A	N/A
Base	16	16	256	3	N/A	N/A	N/A
Var. 1	14	14	512	4	755	32.1	1
Var. 11	15	12	256	2	755	20.2	1
Var. 19	18	16	512	2	748	30.3	1
Var. 9	18	18	512	3	737	25.5	1
Var. 18	16	14	512	2	734	20.2	1
Var. 10	15	16	512	3	726	32.8	1
Var. 15	14	12	512	3	726	27.5	1

Tabela 9: Parâmetros e métricas do dataset Vowel

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
Bloom WiSARD	N/A	N/A	N/A	N/A	876	N/A	N/A
Base	15	15	256	4	0	0	0
Var. 16	15	13	512	5	924	24.4	1
Var. 8	16	11	256	3	918	21.8	1
Var. 9	16	11	256	5	918	0	0
Var. 19	17	11	512	5	918	23.0	1
Var. 18	14	11	128	3	912	28.8	1
Var. 17	16	13	512	5	909	0	0
Var. 5	16	17	256	5	906	0	0

Tabela 10: Parâmetros e métricas do dataset MNIST

Configuração	b	OWeN	FE	FH	Acurácia	Empates (%)	Bleaching
ULEEN	6	49	8192	4	952	N/A	N/A
Base	2	28	1024	2	929	1.61	8
Var. 1	16	16	256	3	915	N/A	N/A
Var. 2	15	15	256	4	913	N/A	N/A
Var. 3	9	27	1024	2	933	N/A	N/A
Var. 4	4	16	512	2	918	2.1	16
Var. 5	8	20	512	3	921	N/A	N/A
Var. 6	4	24	256	2	916	N/A	N/A
Var. 7	8	32	2048	4	943	0.36	6

- **Referência Bloom WiSARD:** Desempenho da implementação de referência do Bloom WiSARD (linha pontilhada)
- **BTHOWeN Base Estudo:** Desempenho da configuração base do BTHOWeN utilizada como ponto de partida (linha tracejada)

### 3.3 Resultados agregados

Na tabela a seguir, apresentamos os parâmetros ótimos identificados para cada dataset, junto com as métricas de desempenho obtidas:

Também apresentamos uma comparação dos resultados entre as diferentes implementações:

## 4 Conclusão

...

Tabela 11: Parâmetros ótimos para cada dataset

Dataset	b	OWeN	FE	FH	Melhor Bleaching	Execução	Empates (%)
Iris	3	2	256	2	2	1	0
Ecoli	3	256	3	10	1	1	10.7
Glass	4	256	4	10	1	1	28.2
Letter	15	256	5	35	3	1	3.9
Wine	10	13	256	4	1	—	1.7
Segment	10	18	1024	3	2	—	6.49
Shuttle	11	25	1024	3	1	—	0.1
Vehicle	15	12	256	2	1	—	20.2
Vowel	15	13	512	5	1	—	24.4
MNIST	6	49	8192	4	—	—	—

Tabela 12: Comparação de acurácia e empates entre implementações

Dataset	Acurácia Ref	Acurácia Base	Acurácia Melhor	Empates Ref	Empates Base	Empates Melhor
Iris	960	<b>980</b>	<b>980</b>	2	12	0
Ecoli	799	786	<b>848</b>	—	8.9	0
Glass	<b>726</b>	577	676	—	39.4	0
Letter	848	734	<b>884</b>	—	18.6	0
Wine	926	983	<b>1000</b>	—	—	0
Segment	—	925	<b>944</b>	—	—	0
Shuttle	868	0	<b>999</b>	—	0	0
Vehicle	<b>926</b>	—	755	—	—	0
Vowel	876	0	<b>924</b>	—	0	0

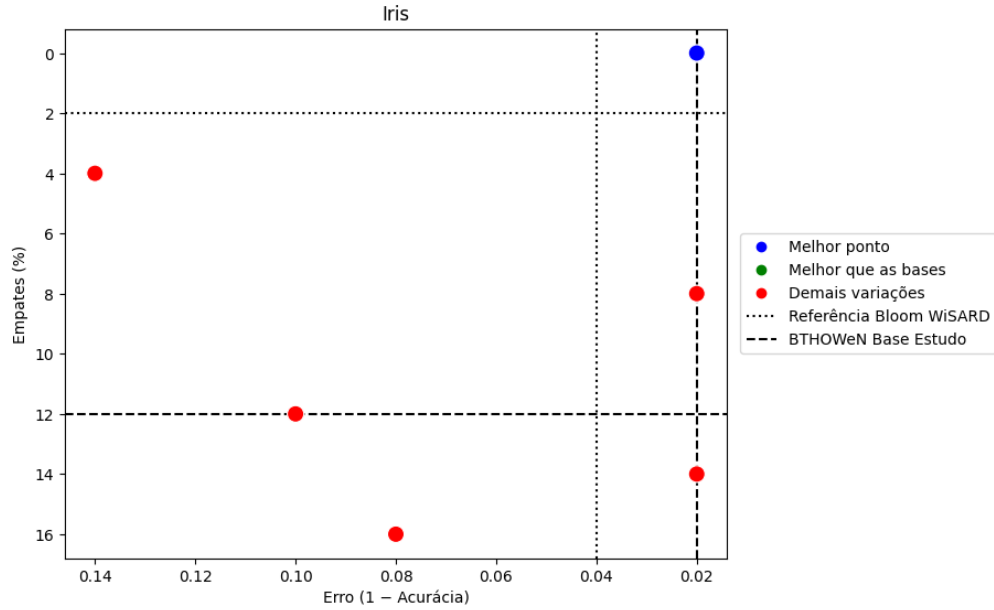


Figura 1: Relação entre erro e empates para o dataset Iris



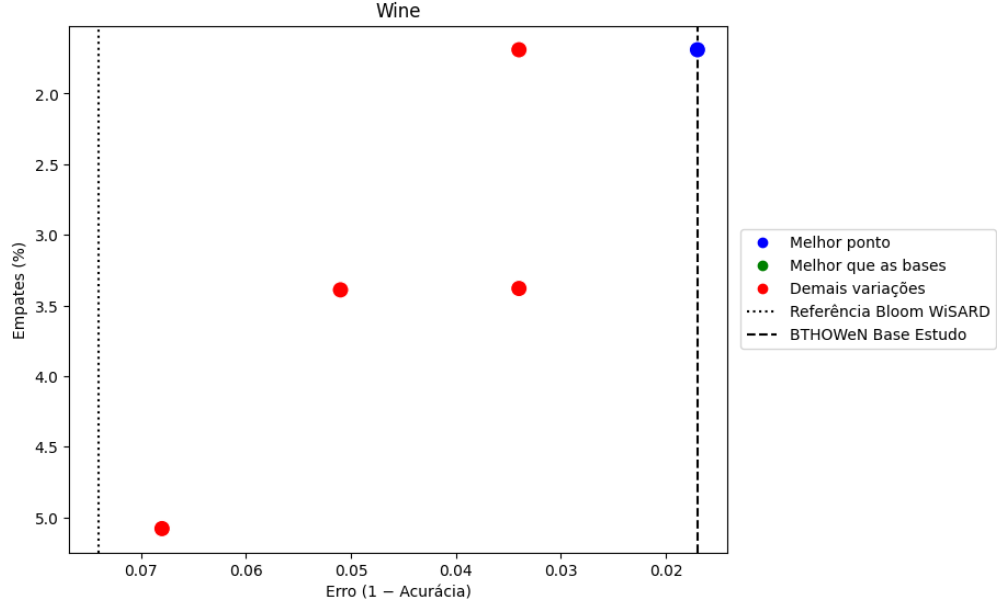


Figura 2: Relação entre erro e empates para o dataset Wine

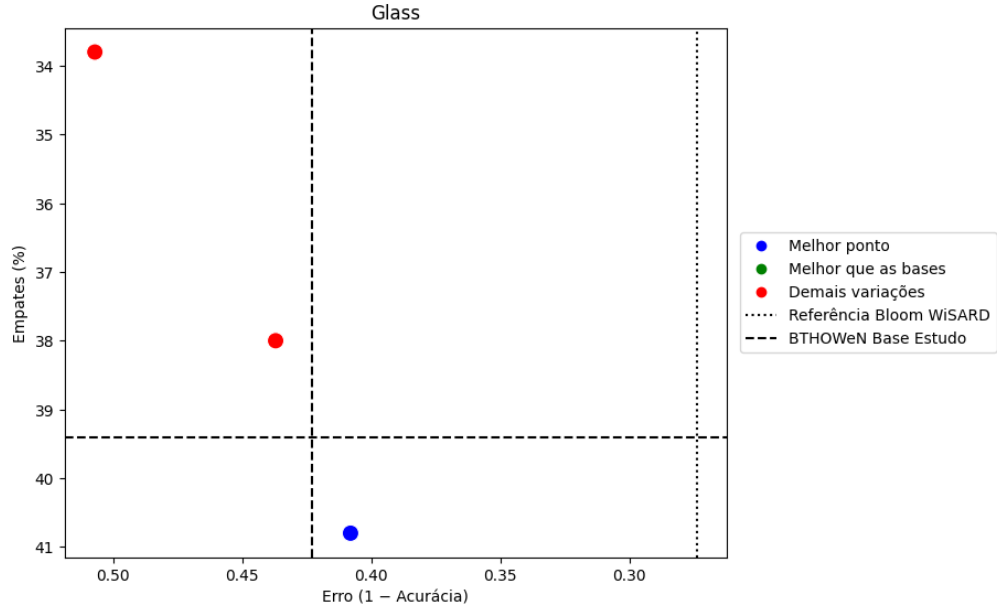


Figura 3: Relação entre erro e empates para o dataset Glass

## Referências

- [1] Aluizio S. Lima Filho, Gabriel P. Guarisa, Leopoldo A. D. Lusquino Filho, Luiz F. R. Oliveira, Felipe M. G. França, and Priscila M. V. Lima. wisardpkg - a library for wisard-based models. *arXiv preprint arXiv:2005.00887*, 2020.
- [2] Leopoldo A. D. Lusquino Filho, Luiz F. R. Oliveira, Aluizio Lima Filho, Gabriel P. Guarisa, Lucca M. Felix, Priscila M. V. Lima, and Felipe M. G. França. Extending the weightless wisard classifier for regression. *Neurocomputing*, 419:125–138, 2020.

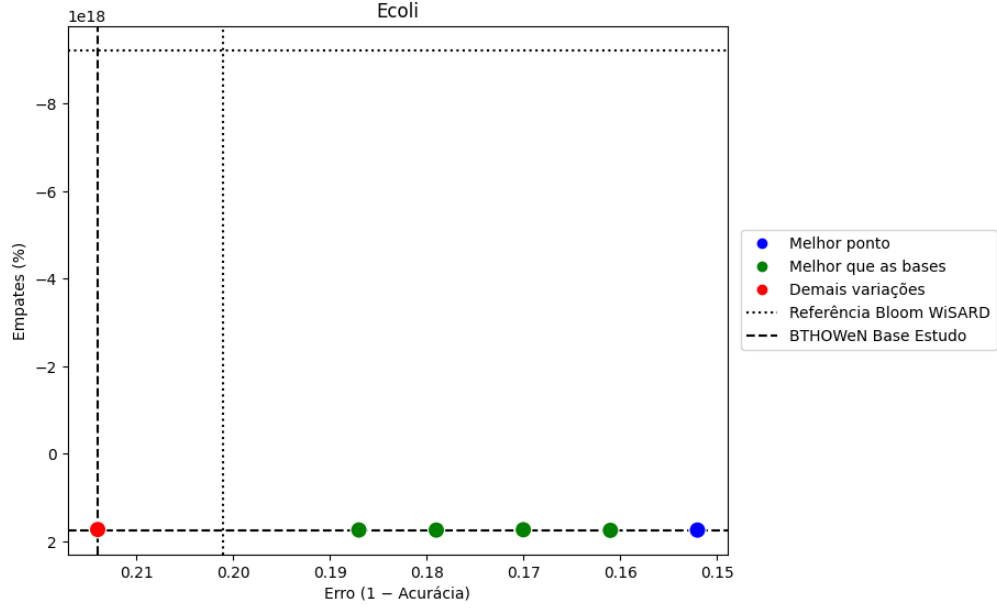


Figura 4: Relação entre erro e empates para o dataset Ecoli

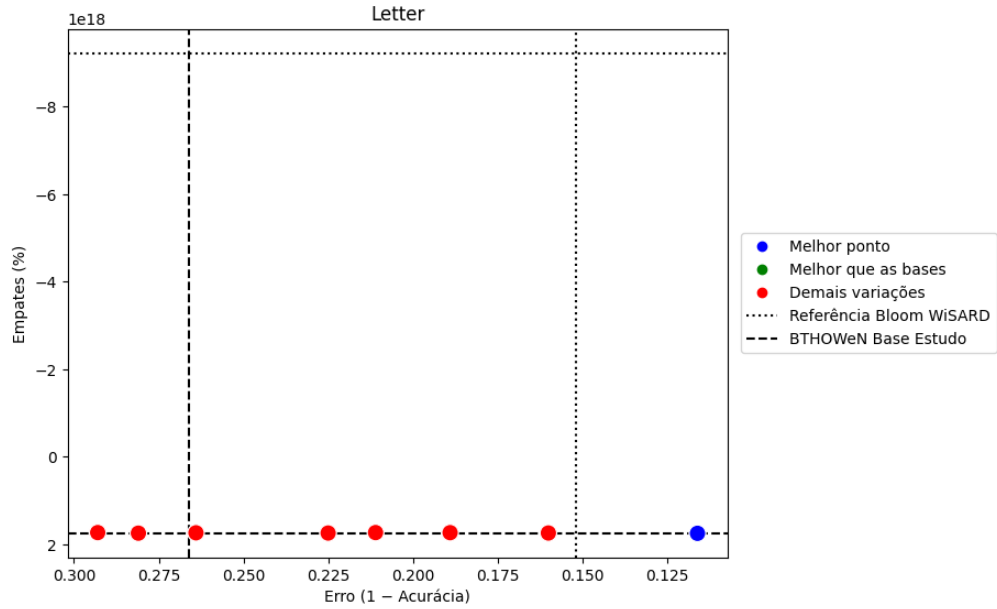


Figura 5: Relação entre erro e empates para o dataset Letter

- [3] Leandro Santiago. Redes sem peso wisard - soluções para memória eficiente, 2023. Apresentação sobre modelos WiSARD e otimizações de memória.
- [4] Leandro Santiago, Leticia Verona, Fabio Rangel, Fabrício Firmino, Daniel S. Menasché, Wouter Caarls, Mauricio Breternitz Jr, Sandip Kundu, Priscila M. V. Lima, and Felipe M. G. França. Weightless neural networks as memory segmented bloom filters. *Neurocomputing*, 416:354–366, 2020.
- [5] Zachary Susskind, Aman Arora, Igor D. S. Miranda, Luis A. Q. Villon, Rafael F. Katopodis, Leandro S. Araújo, Diego L. C. Dutra, Priscila M. V. Lima, Felipe M. G. França, Mauricio Breternitz Jr, and Lizy K. John. Weightless neural networks for efficient edge

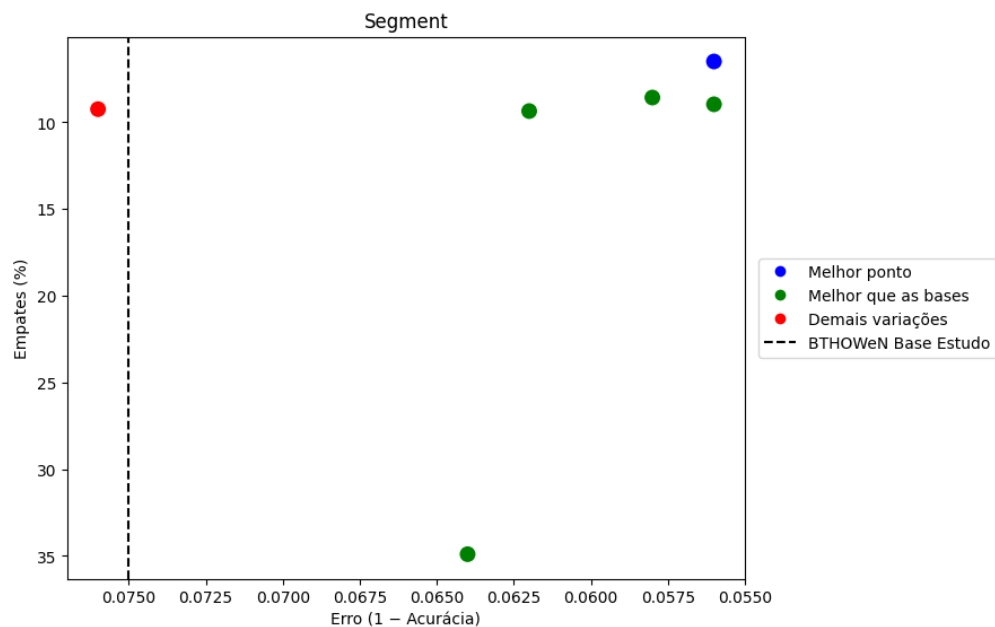


Figura 6: Relação entre erro e empates para o dataset Segment

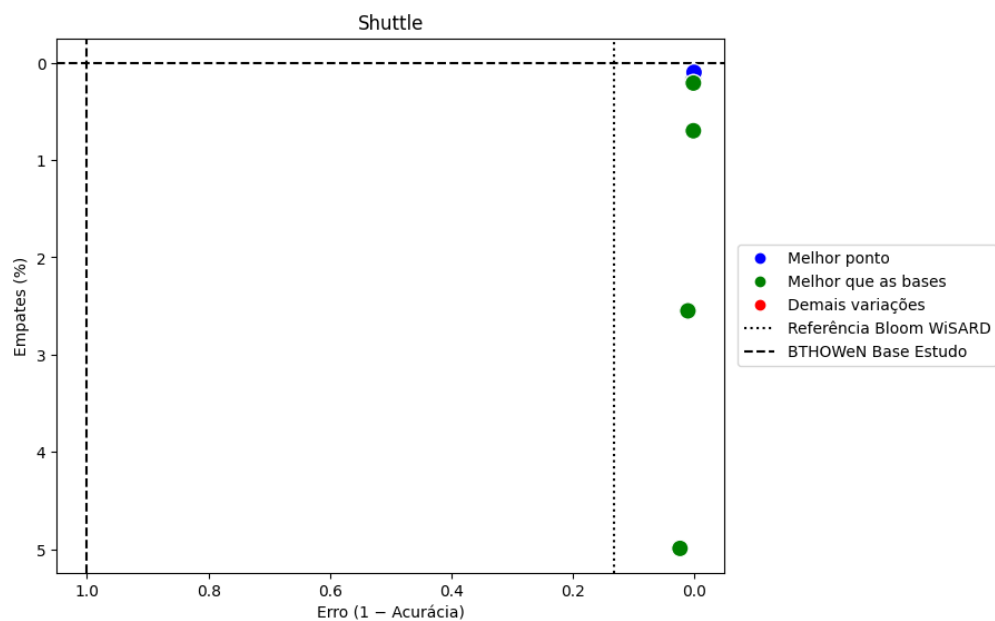


Figura 7: Relação entre erro e empates para o dataset Shuttle

inference. In *PACT '22: International Conference on Parallel Architectures and Compilation Techniques*, New York, NY, USA, 2022. ACM.

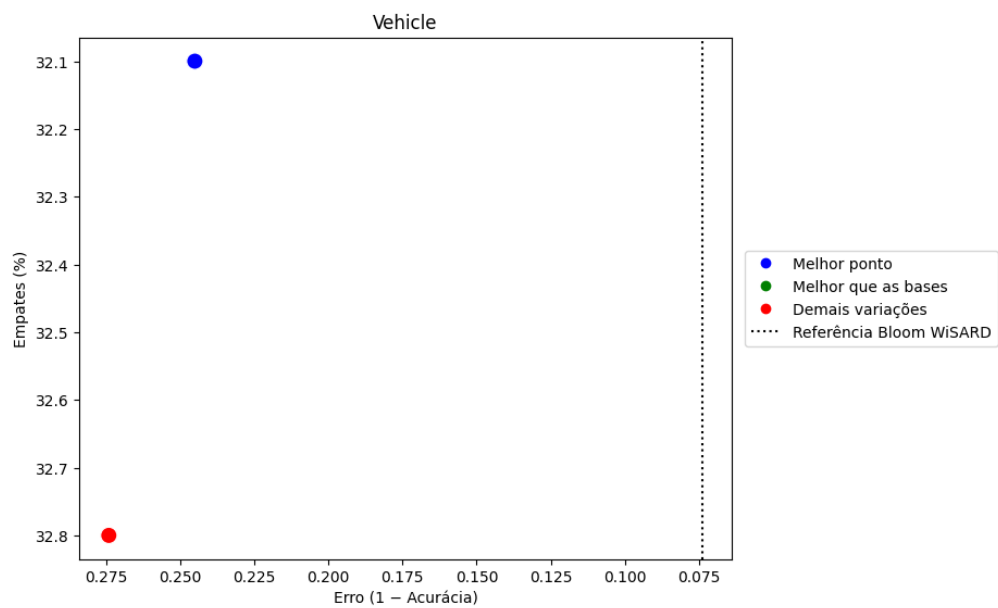


Figura 8: Relação entre erro e empates para o dataset Vehicle

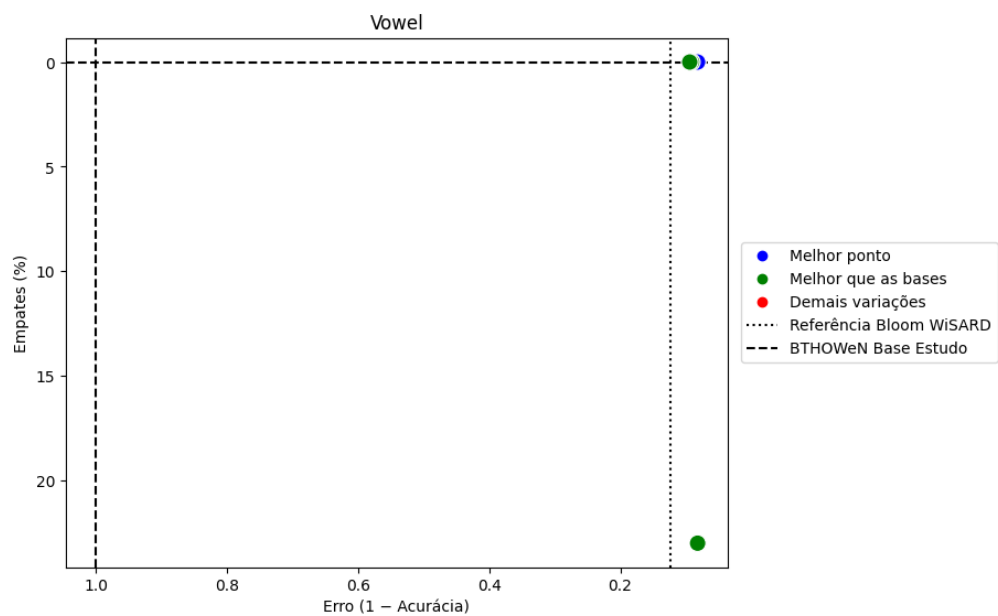


Figura 9: Relação entre erro e empates para o dataset Vowel

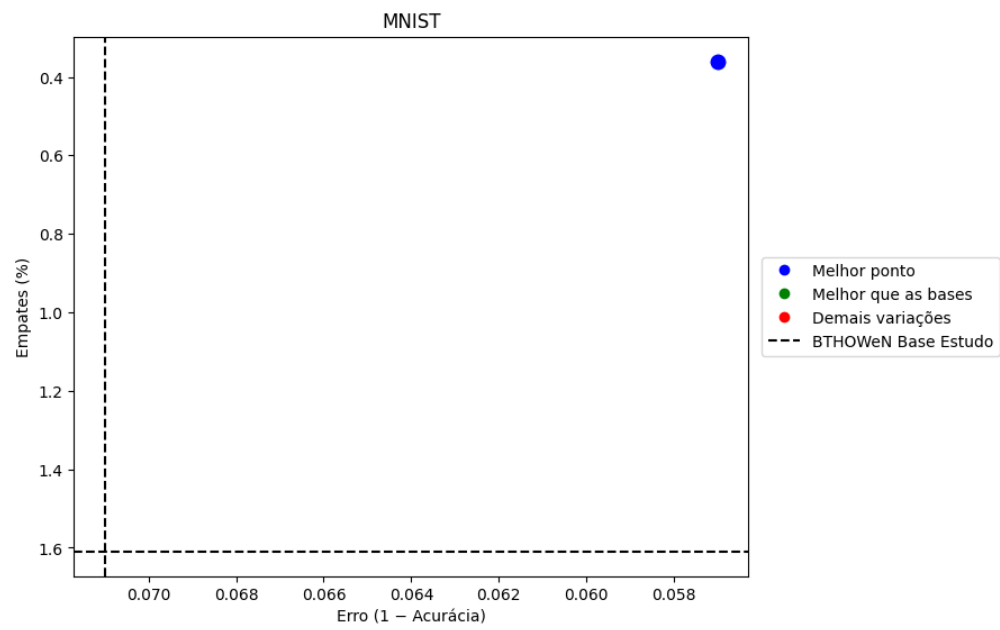


Figura 10: Relação entre erro e empates para o dataset MNIST