Hyperdimensional Computing for Fake News Detection: A High-Efficiency Approach for Large-Scale Data

Naoki Yokoyama[1], **Leandro Santiago de Araújo**[1]

[1]Instituto de Computação – Universidade Federal Fluminense (UFF)
Niterói – RJ – Brazil

naokiyokoyama@id.uff.br, leandro@ic.uff.br

Abstract. *The massive spread of misinformation, intensified during the COVID-19 pandemic, has heightened the need for fake news detection systems that operate at scale. However, the state of the art is dominated by Transformer and Large Language Model architectures with high computational cost, energy consumption, and limited edge applicability. This paper investigates Hyperdimensional Computing (HDC) as an alternative paradigm, framing the problem as a trade-off between accuracy and energy efficiency. We evaluate five HDC variants against classical Machine Learning baselines and Transformers (BERT, RoBERTa) across three benchmarks (ISOT, COVID-19, and a binarized FEVER), totaling 160,051 instances. Transformers achieve the highest F1-score in all three datasets, reaching 99.80% on ISOT, whereas the best HDC configuration achieves 95.77%. Despite this accuracy gap, HDC substantially reduces training time, with reductions of approximately 5.5× to 13× relative to BERT. Under the TDP-based energy estimation adopted in this study, these time reductions translate directly into equivalent estimated energy reductions. The results suggest that HDC is not a direct replacement for Transformer models when maximum predictive performance is the primary objective, but rather a promising efficiency-oriented alternative for large-scale and resource-constrained misinformation detection.*

1. Introduction

The ubiquity of social media has reshaped global communication, creating a paradox in which democratized access to information coexists with the massive dissemination of misleading content. Misinformation, associated with the phenomenon of *fake news*, has shifted from an informational nuisance to a threat to democratic integrity and public health [Olan et al. 2024]. This challenge became especially evident during the COVID-19 "infodemic", characterized by the excessive circulation of false or manipulated information across digital platforms [Shahi et al. 2021]. In response, the research community has developed a wide range of automated detectors based on Natural Language Processing (NLP) and Machine Learning (ML) algorithms. Early approaches relied primarily on statistical text representations, such as n-grams and term-frequency features, whereas more recent methods leverage deep neural architectures and Transformer-based models capable of learning contextualized semantic representations through attention mechanisms [Vaswani et al. 2017, Devlin et al. 2019]. These advances have substantially improved detection performance, establishing Transformer architectures as the current state of the art in many misinformation detection tasks.

Despite the substantial gains in detection accuracy achieved in recent years, the literature reveals a persistent trade-off between predictive performance and computational efficiency. Current state-of-the-art approaches are largely dominated by deep learning architectures, particularly Transformer-based models and Large Language Models (LLMs), which require substantial computational resources, memory capacity, and energy consumption [Strubell et al. 2019]. Although LLMs provide powerful semantic reasoning capabilities that can enhance misinformation detection, their practical deployment remains challenging due to limitations in scalability, inference latency, operational cost, and energy efficiency, especially in real-time and resource-constrained environments [Hu et al. 2024]. Consequently, the growing computational demands of modern misinformation detection systems raise concerns regarding their long-term scalability and sustainability, motivating the search for more efficient learning paradigms.

Hyperdimensional Computing (HDC) is an emerging paradigm that represents and processes information through similarity-based operations on high-dimensional hypervectors, combining noise-resistant redundancy with distributed representations to enable robust and energy-efficient learning [Kanerva 2009]. Recent studies have demonstrated the potential of HDC across various machine learning applications, showing competitive performance when compared to conventional approaches while requiring lower computational resources [Ge and Parhi 2020]. These characteristics make HDC particularly attractive for text classification tasks performed in dynamic and resource-constrained environments. Despite the growing interest in HDC, its application to fake news detection remains relatively underexplored compared to traditional machine learning and deep learning approaches. Therefore, significant opportunities exist to investigate, evaluate, and develop novel HDC-based models capable of improving misinformation detection performance while maintaining computational efficiency.

To address this gap, this work investigates HDC classification models as an energy-efficient alternative to conventional deep learning approaches for large-scale fake news detection. Rather than pursuing marginal gains in predictive performance, the proposed approach focuses on achieving a favorable balance between detection effectiveness and computational efficiency, exploring the potential of HDC to remain competitive in realistic classification scenarios while substantially reducing computational and energy costs compared to Transformer-based models. In this context, the study aims to assess the suitability of HDC for scalable misinformation detection and to identify the conditions under which its efficiency advantages outweigh potential accuracy trade-offs. The main contributions of this work are threefold:

1. Integration of NLP embeddings into the HDC pipeline. We compare Bag-of-Words, TF-IDF, Word2Vec, and N-Gram representations within a common HDC pipeline and identify dataset-dependent representation choices that substantially affect performance;
2. A comparative study of five HDC-based classification models. We evaluate Vanilla HDC, AdaptHD, OnlineHD, NeuralHD, and DistHD under a consistent experimental protocol, providing an empirical comparison of their behavior across heterogeneous misinformation datasets;
3. Comprehensive evaluation of the proposed models against traditional machine learning and Transformer-based methods across three heterogeneous fake news detection benchmark datasets. We show that Transformer models consistently

achieve higher predictive performance, while HDC can substantially reduce training time and corresponding TDP-based energy estimates, particularly on the balanced ISOT benchmark.

The remainder of this paper is organized as follows. The related work of fake news detection is disccussed in Section 2. Basic concepts related to this work are presented in Section 3, providing a foundational understanding of Hyperdimensional Computing. Section 4 introduces the details of the methodology adopted in this work. Section 5 describes the experiments and results, evaluating the performance of HDC models against already established state-of-the-art models. Finally, Section 6 concludes the paper, discussing the potential future work.

2. Related Work

Research on automated fake news detection has evolved from traditional machine learning techniques to increasingly sophisticated deep learning architectures. Early approaches relied on handcrafted textual representations, including Bag-of-Words (BoW), TF-IDF representations, and N-gram statistics, to transform news articles into vector-space representations suitable for classification. Among the most widely adopted algorithms are Support Vector Machines (SVM), Logistic Regression (LR), and Random Forests (RF). Studies using benchmark datasets such as ISOT demonstrated that these methods can achieve strong classification performance, often exceeding 90% accuracy when discriminating between legitimate and deceptive news content [Ahmed et al. 2017]. Despite their effectiveness, these approaches are limited by their inability to capture contextual meaning and long-range semantic dependencies.

The emergence of deep learning significantly improved representation learning for misinformation detection. Convolutional Neural Networks (CNNs) demonstrated the ability to automatically extract local linguistic patterns associated with deceptive content, while recurrent architectures based on Long Short-Term Memory (LSTM) networks enabled modeling of sequential dependencies within textual data. Notably, the CSI framework proposed by Ruchansky et al. [Ruchansky et al. 2017] showed that combining textual information with temporal and social-context signals can substantially improve detection performance. However, recurrent models often suffer from high training costs and limited parallelization capabilities.

A major breakthrough occurred with the introduction of the Transformer architecture [Vaswani et al. 2017]. Through self-attention mechanisms, Transformers effectively capture global semantic relationships while enabling efficient parallel processing. This paradigm led to the development of Bidirectional Encoder Representations from Transformers (BERT) [Devlin et al. 2019], which rapidly became a standard baseline for misinformation detection and fact-checking tasks. Subsequent variants, including RoBERTa [Liu et al. 2019], consistently achieved state-of-the-art performance across multiple fake news benchmarks. More recently, Large Language Models (LLMs) have demonstrated competitive results in few-shot and zero-shot fake news detection scenarios [Hu et al. 2024]. The COVID-19 pandemic further accelerated research in misinformation detection, leading to the creation of specialized datasets and benchmarks. Patwa et al. [Patwa et al. 2021] introduced one of the most widely used COVID-19 fake news datasets, while Transformer-based approaches reported strong performance in identifying

misleading health-related information [Alghamdi et al. 2023]. In parallel, the FEVER benchmark [Thorne et al. 2018] scaled claim verification to large volumes of Wikipedia-derived factual statements with manually annotated labels, becoming a standard resource for large-scale veracity assessment. These studies reinforced the effectiveness of contextual language representations for misinformation detection under rapidly evolving information environments.

Despite their remarkable predictive performance, modern Transformer-based models present significant challenges related to computational complexity, memory consumption, inference latency, and energy requirements. Recent surveys consistently identify computational efficiency as one of the major open challenges in the field [Strubell et al. 2019, Alghamdi et al. 2024].

Hyperdimensional Computing offers a promising alternative to conventional deep learning approaches due to efficient learning, low memory requirements, robustness to noise, and fast inference. Nevertheless, the application of HDC to fake news detection remains largely unexplored. In particular, there is a lack of systematic investigations comparing HDC-based classifiers with state-of-the-art Transformer models across heterogeneous fake news datasets, including both general-domain and COVID-19 misinformation benchmarks. Addressing this gap is the primary objective of the present work.

3. Hyperdimensional Computing

Hyperdimensional computing, also known as Vector Symbolic Architecture, encodes data and objects into high-dimensional vectors, referred to as hypervectors. These hypervectors typically contain thousands of dimensions (e.g., 10,000 or more), where the concentration-of-measure phenomenon makes random vectors nearly orthogonal, providing robustness and noise tolerance [Kanerva 2009]. HDC-based applications generally consist of three main phases: encoding, training, and inference. Among these, encoding is the fundamental step, as it transforms raw input data into a high-dimensional representation suitable for learning and reasoning tasks. Given an input feature vector $\vec{F} = \{f_1, f_2, f_3, \ldots, f_n\}$, the encoding process maps it into a D-dimensional hypervector $\vec{H} \in \{0, 1\}^D$ which preserves both the feature values and their positional information within the original vector $\vec{F}$ [Li et al. 2024].

Depending on the application, hypervectors may be represented using binary, bipolar, integer, real-valued, or complex-valued elements. HDC encoding relies on three fundamental operations that enable the representation of complex data structures within a single holistic hypervector:

- **Binding** ($\otimes$): associates two pieces of information, such as a variable and its value, producing a hypervector that is nearly orthogonal to its operands. For binary hypervectors, binding is commonly implemented using the element-wise XOR operation. This choice is principled rather than incidental: XOR is self-inverse, $(\vec{A} \otimes \vec{B}) \otimes \vec{B} = \vec{A}$, so bound information can be recovered exactly; it preserves Hamming distance, $d(\vec{A} \otimes \vec{C},\ \vec{B} \otimes \vec{C}) = d(\vec{A}, \vec{B})$; and its result is nearly orthogonal to both operands, preventing collisions between bound pairs and their constituents in the hyperspace;
- **Bundling** ($\oplus$): aggregates multiple hypervectors into a single representative vector, functioning as an associative memory and enabling the construction of class

prototypes. For binary hypervectors, bundling is typically performed through element-wise addition followed by a thresholding operation (majority voting);

- **Permutation (Π, cyclic rotation)**: encodes order by rearranging hypervector components. The result remains approximately orthogonal to the original, making permutation useful for representing sequences and constructing N-gram encodings.

By combining these operations, HDC encodes an entire document into a single fixed-dimensional hypervector, enabling efficient classification through similarity-based comparisons using metrics such as Hamming distance or cosine similarity.

In classification tasks, HDC relies on an Associative Memory (AM) to store class representations [Ge and Parhi 2020]. During training, samples are encoded using the basis hypervectors of the item memory and those of the same class are aggregated into a class prototype stored in the AM. During inference, a query hypervector is compared against all prototypes, and the most similar one — by cosine similarity or Hamming distance — determines the prediction, involving only lightweight vector operations.

4. Methodology

The binary classification approach employed in this study distinguishes between two classes of news content: Real and Fake. The methodology encompasses data preprocessing, model training, and performance evaluation stages to effectively identify and detect fake news. Finally, the evaluation metrics are defined to compare HDC and traditional ML models in terms of classification accuracy and relevant performance metrics. These techniques enhance the representational capabilities of the models and provide a more comprehensive comparison of their effectiveness in fake news detection tasks.

4.1. Dataset Collection

Three datasets were selected to isolate distinct stress variables, each emphasizing a different challenge in text classification: content learning under balanced conditions (ISOT), minority-class detection under class imbalance (COVID-19), and scalability in large-scale, context-dependent scenarios (FEVER).

ISOT Fake News Dataset [Ahmed et al. 2017] consists of complete news articles collected from verified news agencies, such as Reuters, and from sources identified as unreliable by fact-checking organizations. The version used in this study contains 44,267 instances, with a nearly balanced class distribution (22,851 fake and 21,416 real news articles). Its long texts, rich vocabulary, and balanced classes make it an ideal benchmark for evaluating intrinsic learning capability under controlled conditions.

COVID-19 Fake News Dataset [Patwa et al. 2021] contains social media posts and short news snippets related to the COVID-19 pandemic, vaccines, and treatments. The version employed in this study comprises 5,975 instances and exhibits a significant class imbalance (4,532 real and 1,443 fake instances, approximately a 3:1 ratio). We note that this distribution differs from the original, approximately balanced 10,700-post release described by [Patwa et al. 2021]: the corpus used here is a consolidated collection in the lineage of that benchmark, and its pronounced imbalance is precisely the property exploited in this work as a minority-class stress test. Its short texts, medical terminology, and skewed distribution provide a challenging setting for assessing minority-class detection.

FEVER Dataset (binarized) [Thorne et al. 2018] contains factual claims derived from Wikipedia and manually labeled as SUPPORTS, REFUTES, or NOT ENOUGH INFO. From the original training partition (145,449 claims), we (i) discarded NOT ENOUGH INFO, which has no equivalent in the real/fake dichotomy; (ii) mapped SUPPORTS $\rightarrow$ real and REFUTES $\rightarrow$ fake; and (iii) removed exact duplicates via SHA-256 hashing of the normalized text, yielding 109,809 claims (80,034 real, 29,775 fake, $\approx 2.7{:}1$). The stratified 70/30 split was performed *after* deduplication, guaranteeing zero intersection between partitions. Models operate on the claim text only (*claim-only*), without the evidence-retrieval stage of the original task, so results are not comparable to full FEVER-Score systems and should be read as a deliberately harder, text-only stress test of scalability.

4.2. Preprocessing and encoding

A uniform preprocessing pipeline was applied to all datasets and models: removal of digital artifacts (URLs, HTML tags, mentions, hashtags), lowercasing, stopword filtering, and Porter stemming — which benefits HDC twofold, through semantic density and a smaller Item Memory. Four textual representations were evaluated: Bag-of-Words (BoW), TF-IDF (both limited to 10,000 terms), 300-dimensional Word2Vec embeddings, and a native trigram-permutation N-Gram encoding applied directly to the token stream. For HDC, an *Item Memory* maps each unique token to a static atomic hypervector $\vec{H} \in \{0,1\}^D$. Two dimensionalities were evaluated, $D \in \{1{,}000;\ 3{,}000\}$: the usual literature reference $D = 10{,}000$ [Kanerva 2009] was not feasible for the full experimental grid on the available hardware, since memory and per-epoch cost grow linearly with D; near-orthogonality is nonetheless preserved (expected Hamming distance $\approx D/2$), and the main experiments in Section 5 use D=3,000, selected as the largest dimensionality feasible under the available hardware constraints. For token-based representations (BoW, TF-IDF), local syntactic structure is captured via trigram permutation encoding,

$$\vec{H}_{\text{tri}} = \Pi^2(\vec{H}_{w_1}) \otimes \Pi(\vec{H}_{w_2}) \otimes \vec{H}_{w_3}, \tag{1}$$

and the document hypervector is obtained by bundling all trigrams, $\vec{H}_{\text{doc}} = \text{threshold}\left(\sum_i \vec{H}_{\text{tri},i}\right)$, with TF-IDF weights, when present, scaling each trigram's contribution in the aggregation. The continuous Word2Vec representation, which has no token-level correspondence, is projected by *feature encoding*: each component j of the document vector $\vec{v} \in \mathbb{R}^{300}$ is associated with a random position hypervector $\vec{P}_j$ and a level hypervector $\vec{L}(v_j)$, obtained by quantizing the component's value over a set of correlated level vectors, yielding

$$\vec{H}_{\text{doc}} = \text{threshold}\Big(\sum_{j=1}^{300} \vec{L}(v_j) \otimes \vec{P}_j\Big), \tag{2}$$

as implemented in the `torchhd` library [Heddes et al. 2023]. Both encoding paths preserve the near-orthogonality properties required by the paradigm.

4.3. Classification Models

To contextualize the performance of HDC, this study evaluates three families of comparative models in addition to five HDC variants. The first consists of classical baselines:

Logistic Regression, SVM, and Random Forest, widely used in foundational studies on fake news detection, enabling direct comparability with prior work [Ahmed et al. 2017]. The second group comprises Transformer-based models: BERT, extensively used in fake news detection [Alghamdi et al. 2023] and thus a standard benchmark, and RoBERTa, which extends it through optimized pretraining. Both were selected for their feasibility on a single NVIDIA Tesla T4 GPU, ensuring fair comparison under controlled hardware conditions; larger models such as GPT-3.5 and GPT-4 were excluded due to proprietary APIs and multi-GPU requirements. The third group consists of HDC models, including Vanilla HDC and four advanced variants: AdaptHD, OnlineHD, NeuralHD, and DistHD. These models explore different aspects of hyperdimensional learning, including adaptability, robustness, scalability, and dynamic representation learning.

AdaptHD [Imani et al. 2019] introduces iterative error correction inspired by the perceptron algorithm: class prototypes are adjusted whenever a misclassification occurs, pushing incorrect centroids away from ambiguous samples and pulling correct ones closer.

OnlineHD [Hernández-Cano et al. 2021] adopts an online learning strategy in which update magnitudes depend on sample novelty: frequent patterns induce minimal updates, while rare or divergent instances trigger stronger adjustments, mitigating majority-class dominance without external resampling.

NeuralHD [Zou et al. 2021] addresses representation-capacity limits by regenerating poorly informative dimensions during training: those with low discriminative power are replaced with new random vectors, improving separability in semantically subtle datasets. NeuralHD employs a non-linear sinusoidal encoding capable of capturing more complex and dynamic relationships in the data.

DistHD [Wang et al. 2023] performs learner-aware *dynamic encoding*: during training, it identifies the hypervector dimensions that mislead classification and regenerates them, while a top-2 decision rule considers the two most similar class prototypes, increasing robustness in regions of class overlap.

A consistent experimental protocol is adopted across all models, including a 70/30 train–test split with stratification and a fixed random seed to ensure reproducibility. Classical machine learning models use default scikit-learn configurations, while BERT and RoBERTa are fine-tuned for two epochs with a batch size of 16 — the largest budget compatible with the session limits of the experimental environment; since the budget was identical across paradigms, the comparison remains internally valid. For the adaptive HDC variants, the number of retraining epochs was likewise fixed at two, with the remaining hyperparameters kept at the defaults of their reference implementations.

4.4. Evaluation Metrics

Model performance was evaluated using the confusion matrix, which summarizes classification outcomes in terms of True Positives (TP), True Negatives (TN), False Positives (FP, Type I errors), and False Negatives (FN, Type II errors). From these quantities, accuracy and F1-score metrics were computed to provide a comprehensive assessment of classification effectiveness. The F1-Score is computed for the positive class (*fake*); on the imbalanced COVID-19 dataset it is the primary metric, since a trivial majority-class classifier reaches $\approx 75.85\%$ accuracy with null fake-class F1. Confusion matrix analysis

is particularly important under class imbalance: examining the distribution of FP and FN errors reveals whether a classifier misclassifies real news as fake or fails to detect fake news, providing insight beyond overall accuracy.

Energy consumption was estimated using the nominal Thermal Design Power (TDP) of the graphics processing unit (GPU), according to

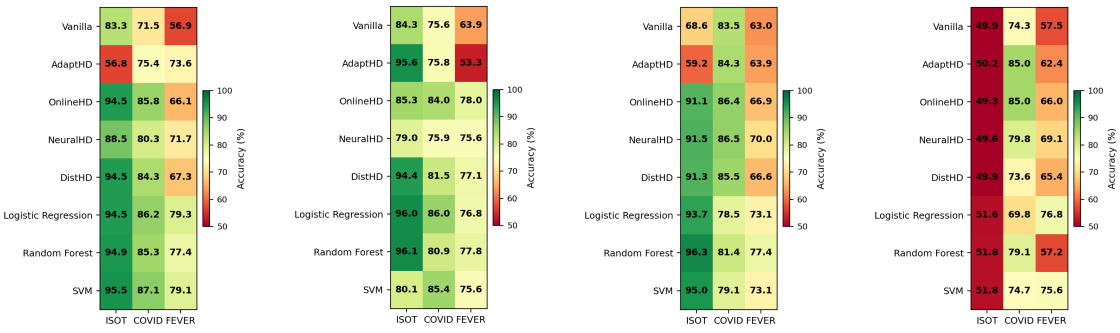
$$E\ [\mathrm{Wh}] = \frac{\mathrm{TDP} \times t}{3600}, \tag{3}$$

where t denotes the execution time in seconds. The experiments were conducted on an NVIDIA Tesla T4 GPU with 16 GB of VRAM, running in a Google Colab Pro environment equipped with 12 GB of RAM, an Intel Xeon processor, and 2 virtual CPUs. A TDP value of 70 W was assumed for the Tesla T4. Consequently, the reported energy values should be interpreted as theoretical upper-bound estimates, since they assume continuous operation at maximum power. Because the TDP is a constant, energy ratios between any two models are, by construction, identical to their training-time ratios; all figures reported in Section 5 satisfy this identity. More precise measurements based on real-time power monitoring tools, such as NVIDIA Management Library (NVML) or CodeCarbon, are left for future work.

5. Experiments and Results

5.1. Experimental Setup

To ensure that efficiency metrics are valid and directly comparable, all experiments were conducted in the standardized environment described in Section 4.4. HDC models were executed primarily on the GPU to accelerate high-dimensional vector operations, while Transformer-based models used the same hardware for training and inference, ensuring a fair comparison in terms of wall-clock time and estimated energy. All experiments followed a standardized experimental setup: a hold-out split with 70% of the data allocated to training and 30% to testing, maintaining a stratified distribution of classes across all datasets. To ensure reproducibility, random_state = 42 was used in all scenarios. The BERT and RoBERTa models were trained for 2 epochs with a batch size of 16, a learning rate of 2e-5, and a max_length of 512. Code, preprocessing scripts, and the binarized FEVER split are available at github repository [1].



(a) Word2Vec (b) Bag-of-Words (c) TF-IDF (d) N-Gram

Figure 1. Accuracy heatmaps for the four text representations ($D = 3{,}000$).

[1] Github repository: `https://github.com/naokityokoyama/fake_news_hdc`

5.2. Performance Results by Text Representation Method

Figure 1 compares the accuracy of the HDC and classical ML models under the four textual representations of Section 4.2, all at $D = 3{,}000$. Among the HDC variants, AdaptHD with BoW led on ISOT (95.6%), NeuralHD led on COVID-19 with TF-IDF (86.5%), and OnlineHD led on FEVER with BoW (78.0%); classical baselines remained strong, with Random Forest reaching 96.3% on ISOT.

The N-Gram panel requires explicit qualification. On ISOT, *every* model — classical baselines included — lands within one percentage point of the majority-class baseline (51.62%), i.e. the classifiers collapse to the majority class, whereas Ahmed et al. [Ahmed et al. 2017] report $\approx 92\%$ with n-gram features. We therefore regard these results as likely affected by an implementation-related issue and refrain from drawing conclusions from this setting. It is reported for transparency and because it is informative in a second respect: under the deduplicated FEVER partition, the best N-Gram result drops to 65.40% (DistHD), far below values from a non-deduplicated corpus. This discrepancy suggests that previously reported high scores may have been influenced by overlap between partitions rather than by the intrinsic effectiveness of the encoding.

For comparability, Figure 1 reports accuracy across the evaluated representations. Since accuracy can be misleading under class imbalance, the positive-class F1-score is used as the primary effectiveness metric in Table 2. The best-performing HDC configuration for each dataset was selected for the final comparison with the Transformer baselines, highlighting that no single text representation is universally optimal across datasets.

Table 1 shows the effect of hypervector dimensionality on the selected HDC configurations. Increasing D from 1,000 to 3,000 substantially improves ISOT performance, with F1 increasing from 88.11% to 95.77%, while the gain on COVID-19 is modest (72.43% to 73.16%). For FEVER, F1 remains essentially unchanged, decreasing slightly from 50.52% to 50.28%. These results indicate that higher hypervector dimensionality does not consistently improve performance and that its impact is dataset-dependent.

Table 1. Effect of hypervector dimensionality on the best-performing HDC configuration for each dataset. Results are reported for $D = 1{,}000$ and $D = 3{,}000$ in terms of accuracy and F1-score. The HDC configuration is fixed within each dataset to isolate the effect of hypervector dimensionality.

Dataset	Model	Representation	Dimension	Acc. %	F1 %
ISOT	AdaptHD	Bag-of-Word	1000	87.68	88.11
	AdaptHD	Bag-of-Word	3000	**95.56**	**95.77**
COVID-19	NeuralHD	TF-IDF	1000	85.95	72.43
	NeuralHD	TF-IDF	3000	**86.50**	**73.16**
FEVER	OnlineHD	Bag-of-Word	1000	77.35	**50.52**
	OnlineHD	Bag-of-Word	3000	**78.00**	50.28

5.3. Performance Results of the Transformers versus HDC models

Table 2 consolidates the final comparison across the three datasets. The results reveal two consistent patterns.

First, Transformer architectures retain the best F1-Score in every scenario: BERT obtained 99.80% on ISOT, 83.24% on COVID-19, and 70.09% on FEVER, with

Table 2. Final comparison across paradigms in terms of accuracy, F1-Score, training time, and estimated energy: Transformers versus the best HDC configuration per dataset ($D = 3{,}000$). Energy ratios equal time ratios by construction.

Dataset	Model	Acc. %	F1 %	Time (s)	E (Wh)
ISOT	BERT	**99.79**	**99.80**	4357.80	84.73
	RoBERTa	99.73	99.74	4450.00	86.53
	AdaptHD (BoW)	95.56	95.77	**796.35**	**15.48**
COVID-19	BERT	**91.47**	**83.24**	123.60	2.40
	RoBERTa	90.96	81.55	126.50	2.46
	NeuralHD (TF-IDF)	86.50	73.16	**9.37**	**0.18**
FEVER	BERT	**85.99**	**70.09**	2532.40	49.24
	RoBERTa	85.41	69.35	2585.00	50.26
	OnlineHD (BoW)	78.00	50.28	**447.04**	**8.69**

RoBERTa consistently within one percentage point (99.74%, 81.55%, and 69.35%), confirming the effectiveness of contextualized representations, particularly under class imbalance.

Second, HDC remains competitive in the balanced scenario at a fraction of the cost. On ISOT, AdaptHD with BoW reached F1 = 95.77%, approximately 4 percentage points below BERT, while reducing training time from 4,357.8 s to 796.35 s and estimated energy from 84.73 Wh to 15.48 Wh — an $\approx 5.5\times$ reduction. On COVID-19 the gap narrows to ≈ 10 points of F1 ($\approx 13\times$ reduction) and on FEVER it widens to ≈ 29 points ($\approx 5.7\times$). The low absolute F1 of *all* models on FEVER, BERT included, is consistent with the difficulty of claim-only verification, in which veracity depends on external evidence deliberately withheld from the input.

Analysis of the confusion matrices clarifies the behaviour under imbalance: on COVID-19 the *fake*-class prototype of NeuralHD, built from only 1,443 training vectors, attains a recall of 77.46% at a precision of 69.33% (146 false positives against 96 false negatives) — the model detects most fake posts but over-flags legitimate ones. This error profile supports the use of HDC as a low-cost first-line filter whose positive flags are confirmed by a Transformer on ambiguous cases, rather than as a standalone replacement.

5.4. Limitations

Several limitations qualify these results. The N-Gram encoding pipeline is faulty on ISOT, as discussed in Section 5.2, and its results are reported for transparency only. The HDC models rely solely on textual representations and do not incorporate external evidence, limiting reasoning-heavy fact-checking. Energy was estimated from TDP rather than measured directly, and only *training* cost was quantified: inference latency and energy — the quantities that matter for edge deployment — remain to be characterized. All experiments ran on a single platform (Tesla T4), so edge, FPGA, or neuromorphic benefits were not validated; the dimensionality study was restricted to $D \in \{1{,}000;\ 3{,}000\}$ by memory constraints; and a single random seed precludes significance testing. Finally, the high ISOT performance must be read with caution: real-class articles come predominantly from a single agency (Reuters), so classifiers may capture source-style markers rather than veracity — a known confound that our preprocessing does not remove.

6. Conclusion

This study investigated Hyperdimensional Computing as an efficiency-oriented alternative to Transformer-based models for fake news detection. Experiments across three heterogeneous benchmark datasets, comprising 160,051 instances in total, showed that HDC can provide competitive classification performance while substantially reducing training cost. The results also demonstrated that HDC performance depends strongly on both dataset characteristics and the underlying text representation. Transformer-based models consistently achieved the highest F1-scores across the evaluated datasets, while HDC achieved its most competitive performance on the balanced ISOT dataset, reaching an F1-score of 95.77% compared with 99.80% for BERT. On the more challenging COVID-19 and binarized FEVER settings, the performance gap was larger, highlighting the limitations of text-only HDC models in datasets characterized by class imbalance or context-dependent factual verification. The most consistent finding was the reduction in computational cost: HDC required between $\approx 5.5\times$ and $\approx 13\times$ less training time and estimated energy than BERT, consistent with the lower computational complexity of the underlying HDC operations and suggesting suitability for Green AI and resource-constrained deployments.

Future work will explore more expressive and efficient HDC architectures, including hybrid HDC–Transformer frameworks, Item Memory initialization using pre-trained language embeddings, and multilingual encoders for misinformation detection. Further directions include comparison with compressed Transformers (DistilBERT, TinyBERT), multi-seed stability analysis with confidence intervals, higher dimensionalities ($D = 10{,}000$), measurement of inference latency and energy, deployment on FPGA and neuromorphic hardware, and direct energy measurement via NVML or CodeCarbon. These directions may help further improve predictive performance while preserving the computational efficiency that makes HDC a promising paradigm for sustainable large-scale fake news detection.

References

Ahmed, H., Traore, I., and Saad, S. (2017). Detection of online fake news using n-gram analysis and machine learning techniques. In *Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, pages 127–138, Cham. Springer International Publishing.

Alghamdi, J., Lin, Y., and Luo, S. (2023). Towards covid-19 fake news detection using transformer-based models. *Knowledge-Based Systems*, 274:110642.

Alghamdi, J., Luo, S., and Lin, Y. (2024). A comprehensive survey on machine learning approaches for fake news detection. *Multimedia Tools and Applications*, 83(17):51009–51067.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). Bert: Pre-training of deep bidirectional transformers for language understanding. In *Proc. NAACL-HLT*, pages 4171–4186. Association for Computational Linguistics.

Ge, L. and Parhi, K. K. (2020). Classification using hyperdimensional computing: A review. *IEEE Circuits and Systems Magazine*, 20(2):30–47.

Heddes, M., Nunes, I., Vergés, P., Kleyko, D., Abraham, D., Givargis, T., Nicolau, A., and Veidenbaum, A. (2023). Torchhd: An open source python library to support research on hyperdimensional computing and vector symbolic architectures. *Journal of Machine Learning Research*, 24(255):1–10.

Hernández-Cano, A., Matsumoto, N., Ping, E., and Imani, M. (2021). Onlinehd: Robust, efficient, and single-pass online learning using hyperdimensional system. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 56–61.

Hu, B., Sheng, Q., Cao, J., Shi, Y., Li, Y., Wang, D., and Qi, P. (2024). Bad actor, good advisor: Exploring the role of large language models in fake news detection. In *Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence*, AAAI'24. AAAI Press.

Imani, M., Morris, J., Bosch, S., Shu, H., De Micheli, G., and Rosing, T. (2019). Adapthd: Adaptive efficient training for brain-inspired hyperdimensional computing. In *2019 IEEE Biomedical Circuits and Systems Conference (BioCAS)*, pages 1–4.

Kanerva, P. (2009). Hyperdimensional computing: An introduction to computing in distributed representation with high-dimensional random vectors. *Cognitive Computation*, 1:139–159.

Li, T., Zhong, H., Wu, J., Kämpfe, T., Ni, K., Narayanan, V., Yang, H., and Li, X. (2024). Cafehd: A charge-domain fefet-based compute-in-memory hyperdimensional encoder with hypervector merging. In *2024 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pages 1–6.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L., and Stoyanov, V. (2019). Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.

Olan, F., Jayawickrama, U., Arakpogun, E. O., Suklan, J., and Liu, S. (2024). Fake news on social media: The impact on society. *Information Systems Frontiers*, 26(2):443–458.

Patwa, P., Sharma, S., Pykl, S., Guptha, V., Kumari, G., Akhtar, M. S., Ekbal, A., Das, A., and Chakraborty, T. (2021). Fighting an infodemic: Covid-19 fake news dataset. In *Combating Online Hostile Posts in Regional Languages during Emergency Situation*, pages 21–29, Cham. Springer International Publishing.

Ruchansky, N., Seo, S., and Liu, Y. (2017). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, CIKM '17, pages 797–806, New York, NY, USA. Association for Computing Machinery.

Shahi, G. K., Dirkson, A., and Majchrzak, T. A. (2021). An exploratory study of covid-19 misinformation on twitter. *Online Social Networks and Media*, 22:100104.

Strubell, E., Ganesh, A., and McCallum, A. (2019). Energy and policy considerations for deep learning in nlp. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3645–3650, Florence, Italy. Association for Computational Linguistics.

Thorne, J., Vlachos, A., Christodoulopoulos, C., and Mittal, A. (2018). FEVER: A large-scale dataset for fact extraction and VERification. In *Proc. NAACL-HLT*, pages 809–819. Association for Computational Linguistics.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., and Polosukhin, I. (2017). Attention is all you need. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pages 6000–6010, Red Hook, NY, USA. Curran Associates Inc.

Wang, J., Huang, S., and Imani, M. (2023). Disthd: A learner-aware dynamic encoding method for hyperdimensional classification. In *2023 60th ACM/IEEE Design Automation Conference (DAC)*, pages 1–6.

Zou, Z., Kim, Y., Imani, F., Alimohamadi, H., Cammarota, R., and Imani, M. (2021). Scalable edge-based hyperdimensional learning system with brain-like neural adaptation. In *SC21: Int. Conf. for High Performance Computing, Networking, Storage and Analysis*, pages 1–15.