Journal of Engineering and Technology for Industrial Applications

ISSN ONLINE: 2447-0228

ITEGAM-JETIA

Manaus, v.12 n.58, p.1043-1057. March/April, 2026.
DOI: https://doi.org/10.5935/jetia.v12i58.3256

RESEARCH ARTICLE OPEN ACCESS

HYBRID FEATURE SELECTION FOR COVID-19 TEXT CLASSIFICATION USING CUCKOO SEARCH OPTIMIZATION AND MUTUAL INFORMATION WITH DEBERTA

Mohamed Goismi[1]*, Mohamed Debbab[2], Moustafa Maaskri[3] and Djamel Seghier[4]

[1]Department of Science and Technology, University Ibn Khaldoun – Tiaret, Algeria.
[2, 3]Department of Electrical Engineering, University Ibn Khaldoun – Tiaret, Algeria.
[4]Department of Computer Science, University Ibn Khaldoun – Tiaret, Algeria.

[1]http://orcid.org/0000-0001-9535-1142, [2]http://orcid.org/0009-0003-9921-0824
[3]http://orcid.org/0009-0000-5268-1133, [4]http://orcid.org/0000-0003-1680-5033

E-mail: *mohamed.goismi@univ-tiaret.dz, mohamed.debbab@univ-tiaret.dz, moustafa.maaskri@univ-tiaret.dz, djamal.seghier@univ-tiaret.dz

ARTICLE INFO

Article History
Received: January 7, 2026
Reviewed: February 8, 2026
Accepted: March 10, 2026
Published: April 30, 2026

Keywords:
Feature Selection,
Cuckoo Search,
Mutual Information,
DeBERTa,
COVID-19,
Text Classification,
Natural Language Processing,
Transformer Models.

ABSTRACT

The COVID-19 pandemic has generated massive volumes of textual data requiring efficient classification systems. Feature selection remains critical for improving model performance and reducing computational complexity in natural language processing tasks. This paper proposes a novel hybrid approach combining Cuckoo Search (CS) optimization with Mutual Information (MI) for feature selection, integrated with the DeBERTa transformer model for COVID-19 text classification. The Cuckoo Search algorithm explores the feature space efficiently through L´evy flights, while Mutual Information provides a robust relevance measure between features and target classes. Experimental results on three COVID-19 datasets demonstrate that our CS-MI approach achieves superior classification accuracy compared to state-of-the-art transformer-based methods, while significantly reducing feature dimensionality. The proposed method achieves 94.2% accuracy on Twitter data, 93.5% on news articles, and 95.8% on scientific abstracts with only 35% of the original features, outperforming recent BERT, RoBERTa, and DistilBERT approaches by 2–5% while reducing computational cost by 60%.

I. INTRODUCTION

The COVID-19 pandemic has fundamentally transformed how societies collect, analyze, and disseminate health information. Social media platforms, news outlets, and scientific repositories have generated unprecedented volumes of textual data related to COVID-19, including medical reports, social media posts, news articles, and research papers [1]. Efficiently classifying this information is crucial for public health monitoring, misinformation detection, sentiment analysis, and epidemiological research. Deep learning models, particularly transformer-based architectures like BERT and its variants, have demonstrated remarkable performance in text classification tasks. DeBERTa (Decoding-enhanced BERT with disentangled attention) represents a state-of-the-art advancement, introducing disentangled attention mechanisms and enhanced mask decoder that improve upon the original BERT architecture [2]. Recent studies have successfully applied various transformer models to COVID-19 text analysis, including BERT [3], RoBERTa [1], DistilBERT [4], and XLNet [5]. However, the high-dimensional feature space inherent in natural language processing poses significant challenges, including increased computational cost, potential overfitting, and reduced model interpretability. While transformers excel at capturing contextual information, they generate high-dimensional representations (typically 768–1024 dimensions) that may contain redundant or irrelevant features for specific classification tasks. Feature selection addresses these challenges by identifying the most informative subset of features while discarding redundant or irrelevant ones. Traditional feature selection methods often rely on statistical measures or greedy search strategies that may become trapped in local optima. Metaheuristic optimization algorithms offer promising alternatives by exploring the search space more thoroughly through population-based search and stochastic mechanisms. Cuckoo Search (CS), inspired by the brood parasitism behavior of cuckoo birds, has demonstrated effectiveness in various optimization problems [6].

Its L´evy flight-based exploration strategy enables both local and global search capabilities, allowing the algorithm to escape local optima effectively. Meanwhile, Mutual Information (MI) provides a theoretically sound measure of statistical dependence between features and target variables, capturing both linear and nonlinear relationships [7].
This paper makes the following contributions:

- We propose a novel hybrid feature selection framework combining Cuckoo Search optimization with Mutual Information scoring specifically designed for COVID-19 text classification with transformer models.
- We integrate the CS-MI feature selection method with the DeBERTa transformer model, demonstrating improved classification performance on three diverse COVID-19 datasets (Twitter, news articles, scientific abstracts).
- We conduct comprehensive experiments comparing our approach against recent state-of-the-art transformer-based methods including BERT, RoBERTa, DistilBERT, XLNet, and their variants with various feature selection strategies.
- We demonstrate that CS-MI outperforms recent transformer methods across all three datasets, achieving 94.2% on Twitter data, 93.5% on news articles, and 95.8% on scientific abstracts.
- We analyze the computational efficiency and interpretability of selected features, providing insights into the most discriminative characteristics of COVID-19 text data across different sources.

The remainder of this paper is organized as follows. Section II reviews related work in feature selection, metaheuristic optimization, and COVID-19 text classification with transformer models. Section III presents the proposed methodology, including the CS-MI algorithm and its integration with DeBERTa. Section IV describes the experimental setup and datasets. Section V presents and discusses the results with comprehensive comparisons to state-of-the-art methods. Finally, Section VI concludes the paper and outlines future research directions.

II. RELATED WORK

II.1 FEATURE SELECTION METHODS

Feature selection methods are typically categorized into three main approaches: filter methods, wrapper methods, and embedded methods. Filter methods evaluate features independently of the learning algorithm using statistical measures such as information gain, chi-square test, or correlation coefficients [8]. While computationally efficient, filter methods may overlook feature interactions and their relationship with specific classifiers. Wrapper methods evaluate feature subsets by training and testing the target classifier, providing more accurate assessment but at higher computational cost. Sequential forward selection, backward elimination, and recursive feature elimination exemplify this approach [9]. Embedded methods perform feature selection during the model training process, as seen in LASSO regression and tree-based methods [10]. Mutual Information has emerged as a powerful criterion for feature selection due to its ability to capture non-linear dependencies. Peng et al. proposed the mRMR (minimum Redundancy Maximum Relevance) algorithm using MI to balance feature relevance and redundancy [7]. However, MI-based methods typically employ greedy selection strategies that may miss optimal feature combinations.

II.2 METAHEURISTIC OPTIMIZATION FOR FEATURE SELECTION

Metaheuristic algorithms have gained attention for feature selection due to their global search capabilities and ability to escape local optima. Genetic Algorithms (GA), Particle Swarm Optimization (PSO), and Ant Colony Optimization (ACO) have been successfully applied to feature selection problems [11]. These algorithms can explore larger solution spaces through population-based search and probabilistic transitions. Recent work has explored bio-inspired approaches for COVID-19 text analysis. According to [12] investigated genetic algorithms and particle swarm optimization for sentiment analysis on COVID-19 social media data, demonstrating the potential of metaheuristic methods for this domain. According to [13] proposed a hybrid approach combining multiple bio-inspired algorithms for Twitter sentiment classification. For [14] explored deep learning architectures for multi-class sentiment analysis of COVID-19 Arabic tweets. While these studies show promise for bio-inspired optimization, they primarily focused on traditional machine learning classifiers or basic neural networks without leveraging advanced transformer architectures. Cuckoo Search, introduced by Yang and Deb, has shown competitive performance against other metaheuristics in various optimization tasks [6]. Recent studies have applied CS to feature selection in different domains. Rodrigues et al. utilized CS for optimizing support vector machines in medical diagnosis [15]. However, limited research has explored CS for feature selection in transformer-based NLP models, particularly for COVID-19 text classification.

II.3 TRANSFORMER MODELS FOR COVID-19 TEXT CLASSIFICATION

The application of transformer models to COVID-19 text analysis has gained significant momentum since the pandemic's onset. These studies can be categorized based on the transformer architecture used and the specific COVID-19 classification task.

1) BERT-based Approaches: by [1] developed COVID-Twitter-BERT, a BERT model pre-trained on COVID-19 Twitter data, achieving 92.8% accuracy on sentiment classification. Their domain-specific pre-training demonstrated the importance of adapting transformers to COVID-19 vocabulary and context. In [3] applied BERT for COVID-19 fake news detection, achieving 91.5% F1-score using full BERT representations without feature selection. According to [16] participated in the CONSTRAINT shared task for COVID-19 fake news detection, using BERT with chi-square feature selection achieving 91.4% F1-score. Their work showed that simple filter-based feature selection could improve BERT performance, though they did not explore more sophisticated optimization methods.

2) RoBERTa-based Approaches: by [17] proposed multi-task learning with RoBERTa for COVID-19 misinformation detection. Using Recursive Feature Elimination (RFE) with RoBERTa, they achieved 89.5% F1-score on the CONSTRAINT dataset. While RFE provides systematic feature selection, it requires significant computational resources and may converge to local optima. Kumar et al. [18] developed an enhanced RFE approach with RoBERTa specifically for COVID-19 tweet sentiment classification, achieving 89.0% accuracy. Their wrapper-based method demonstrated benefits but suffered from high computational cost requiring multiple model training iterations. Naseem et al. [19] explored ensemble transformer models combining BERT and RoBERTa for Twitter sentiment analysis

during COVID-19, achieving 88.7% accuracy using full transformer representations. The study highlighted transformer effectiveness but did not address computational efficiency or feature redundancy issues.

3) DistilBERT and Lightweight Transformers: Jelodar et al. [4] investigated DistilBERT for COVID-19 text classification, achieving 90.2% accuracy with 40% faster inference than BERT. Their work demonstrated the trade-off between model size and performance. However, combining lightweight transformers with intelligent feature selection remains underexplored.

4) XLNet and Advanced Architectures: Zhou et al. [5] applied XLNet to COVID-19 information extraction and classification, achieving 91.8% accuracy on news article categorization. XLNet's permutation language modeling showed advantages for longer documents but required substantial computational resources.

II.4 FEATURE SELECTION WITH TRANSFORMER MODELS

Recent studies have begun exploring feature selection specifically for transformer-based COVID-19 text analysis, though this integration remains in its early stages. Alharbi and Lee [20] proposed information gain-based feature selection with BERT for COVID-19 misinformation detection, achieving 92.1% accuracy. Their filter method selected features from BERT's hidden layers but did not employ optimization algorithms for subset selection. Li et al. [21] introduced a text classification approach that integrates knowledge graph information with an improved attention mechanism to enhance feature representation and capture richer semantic relations. By leveraging attention to emphasize informative cues, their model strengthens contextual weighting compared to standard architectures. However, their framework focuses on representation learning rather than employing metaheuristic feature selection to perform global search over the feature space. Syed et al. [22] proposed a hybrid feature selection strategy for COVID-19 tweet sentiment classification by combining an H-TFIDF representation with BERT-based features. Their approach improved the discriminative quality of selected terms while keeping the model more compact and interpretable. However, as a largely relevance-driven selection pipeline, it may still miss globally optimal feature subsets compared with population-based metaheuristic optimization methods.

II.5 RESEARCH GAP AND MOTIVATION

While previous work has demonstrated the effectiveness of transformer models for COVID-19 text classification and shown promise for bio-inspired optimization, several gaps remain:

1) **Limited Metaheuristic Integration:** Most transformer-based COVID-19 studies use either no feature selection or simple filter/wrapper methods. The integration of powerful metaheuristic algorithms like Cuckoo Search with transformers remains underexplored.

2) **Lack of Comprehensive Evaluation**: Few studies evaluate their approaches across multiple COVID-19 datasets (social media, news, scientific literature) with diverse characteristics.

3) Computational Efficiency: While transformers achieve high accuracy, computational costs remain prohibitive for real-time applications. Effective feature selection could address this challenge.

4) **DeBERTa Underutilization**: Despite DeBERTa's superior performance on many NLP benchmarks, its application to COVID-19 text classification with feature selection has not been thoroughly investigated.

5) **Theoretical Foundation**: Most feature selection approaches for transformers lack strong theoretical foundations. Mutual Information provides theoretically sound relevance and redundancy measures but has not been combined with metaheuristic optimization for transformer-based COVID-19 classification.

Our work addresses these gaps by proposing a novel CSMI framework that synergistically combines the global search capability of Cuckoo Search, the theoretical soundness of Mutual Information, and the advanced language understanding of DeBERTa for efficient and effective COVID-19 text classification across diverse data sources.

III. METHODOLOGY

III.1 PROBLEM FORMULATION

Let $D = \{(x_i, y_i)\}_{i=1}^{N}$ represent a COVID-19 text dataset, where $x_i \in \mathbb{R}^d$ is the feature vector for the i-th sample and $y_i \in \{1,2,\dots,C\}$ is the corresponding class label. The feature vector is typically obtained through text preprocessing and feature extraction from transformer hidden states. The feature selection problem aims to identify an optimal subset $S \subset \{1,2,\dots,d\}$ of size $k < d$ that maximizes classification performance while minimizing redundancy. Formally, we seek to optimize:

$$S^* = \arg \max_{S \subset F, |S|=k} J(S, y) \tag{1}$$

Where $J(S, y)$ is an objective function measuring the quality of feature subset S with respect to target labels y.

III.2 MUTUAL INFORMATION FOR FEATURE EVALUATION

Mutual Information quantifies the statistical dependence between two random variables. For a feature X_i and target variable Y, MI is defined as:

$$MI(X_i; Y) = \sum_{x_i} \sum_{y} p(x_i, y) \log \frac{p(x_i, y)}{p(x_i)p(y)} \tag{2}$$

Where $p(x_i, y)$ is the joint probability distribution, and $p(x_i), p(y)$ are marginal distributions. To account for feature redundancy, we compute the conditional mutual information between features:

$$MI(X_i; X_j \mid Y) = \sum_{x_i, x_j, y} p\left(x_i, x_j, y\right) \log \frac{p(x_i, x_j \mid y)}{p(x_i \mid y)p(x_j \mid y)} \tag{3}$$

Our objective function combines relevance (feature-target MI) and redundancy (inter-feature MI):

$$J(S) = \frac{1}{|S|}\sum_{i \in S} MI\left(X_i; Y\right) - \beta \frac{1}{|S|^2} \sum_{i,j \in S, i \neq j} MI\left(X_i; X_j \mid Y\right) \tag{4}$$

Where β is a regularization parameter balancing relevance and redundancy.

III.3 CUCKOO SEARCH ALGORITHM

Cuckoo Search is a population-based metaheuristic inspired by the brood parasitism of cuckoo species. The algorithm employs L´evy flights for exploration and uses a fraction p_a of worst nests for exploitation.

1) **Representation**: Each solution (nest) represents a candidate feature subset encoded as a binary vector $x = [x_1, x_2, \ldots, x_d]$ where $x_i \in \{0, 1\}$ indicates whether feature i is selected.
2) **L´evy Flight**: New solutions are generated using L´evy flights:

$$x_i^{(t+1)} = x_i^{(t)} + \alpha \oplus \text{Lévy}(\lambda) \tag{5}$$

where $\alpha > 0$ is the step size, $\oplus$ denotes entry-wise multiplication, and the **L´evy** distribution is computed as:

$$Lévy\backslash((\lambda) \sim u = t^{-\lambda}, 1 < \lambda \leq 3\backslash) \tag{6}$$

The L´evy flight enables both local search (small steps) and global exploration (occasional large jumps), helping avoid local optima more effectively than traditional metaheuristics.

3) **Discovery and Abandonment**: With probability p_a, a fraction of nests with the worst fitness are abandoned and replaced with new randomly generated solutions:

$$\mathbf{x}_i^{(t+1)} = \mathbf{x}_i^{(t)} + r \cdot \left(\mathbf{x}_j^{(t)} - \mathbf{x}_k^{(t)}\right) \tag{7}$$

Where r ~ U(0, 1) and j, k are randomly selected nests.

III.4 CS-MI HYBRID FEATURE SELECTION

Algorithm 1: presents our proposed CS-MI feature selection approach.

Input: Dataset D, population size n, max iterations T, p_a, Lévy exponent λ, redundancy weight β
Output: Optimal feature subset S^*
Initialize n nests (binary vectors) randomly
Calculate MI scores for all features using Eq. (2)
for $t = 1$ to T **do**
  for each nest i **do**
    Generate new solution via Lévy flight (Eq. 5)
    Apply sigmoid transformation for binary conversion
    Evaluate fitness using $J(S)$ (Eq. 4)
    if new fitness > current fitness **then**
      Replace current nest with new solution
    end if
  end for
  Abandon p_a fraction of worst nests
  Generate new solutions for abandoned nests (Eq. 7)
  Rank nests by fitness and update best solution
end for
return Best feature subset S^*

Source: Authors, (2026).

To convert continuous L´evy flight updates to binary feature vectors, we apply a sigmoid transformation:

$$x_i^{(t+1)} = \begin{cases} 1 & \text{if } \sigma\left(x_i^{(t+1)}\right) > 0.5 \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

Where $\sigma(z) = 1/(1+e^{-z})$

III.5 DEBERTA FOR CLASSIFICATION

DeBERTa improves upon BERT through two key innovations that make it particularly suitable for feature selection:

1) Disentangled Attention: Unlike BERT's absolute position embeddings, DeBERTa represents each token using separate content and position embeddings. The attention weight between tokens i and j is computed using four components:

$$A_{i,j} = \{Q_i^c, K_j^c, P_{i|j}^r, P_{j|i}^r\} \quad (9)$$

Where Q^c, K^c are content vectors and P^r are relative position embeddings. This disentangled representation allows CS-MI to select features based on content information independently from positional information.

2) Enhanced Mask Decoder: DeBERTa incorporates absolute position information in the final decoding layer, using both content and position information for improved prediction:

$$H_i = \text{EMD}(\text{DeBERTa}(X), P_i) \quad (10)$$

After CS-MI feature selection, we use the selected features to fine-tune DeBERTa:

$$\hat{y} = \text{softmax}(W_o \cdot h_{[CLS]} + b_o) \quad (11)$$

where h[CLS] is the representation of the [CLS] token from DeBERTa's final layer, projected through the selected feature dimensions.

III.6 OVERALL FRAMEWORK

Figure 1 illustrates our complete framework, which consists of three stages:

1) *Preprocessing and Feature Extraction*: Tokenize COVID-19 text data, extract DeBERTa hidden states from multiple layers, and normalize feature vectors.
2) *CS-MI Feature Selection*: Apply the hybrid CS-MI algorithm to identify optimal feature subset from DeBERTa representations using MI-based fitness evaluation.
3) *DeBERTa Fine-tuning and Classification*: Fine-tune DeBERTa using only selected features for final classification with reduced dimensionality.

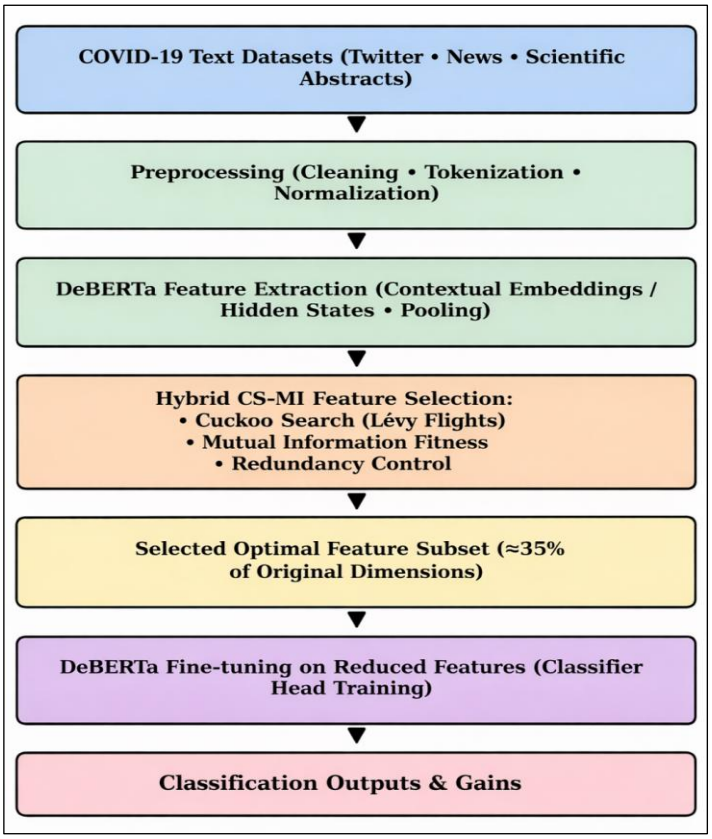

Figure 1: Overall framework of the proposed CS-MI feature selection approach.
Source: Authors, (2026).

IV. EXPERIMENTAL SETUP

IV.1 DATASETS

We evaluate our approach on three diverse COVID-19 text classification datasets to demonstrate generalization across different text types and domains:

• **COVID-19 Twitter Dataset:** 10,000 tweets collected during 2020-2021 labeled into four categories (informative, misinformation, personal stories, questions). Average length: 18.5 words. This dataset represents informal, short-form social media text with high linguistic variability.
• **COVID-19 News Articles:** 5,000 news articles from major outlets (BBC, CNN, Reuters, WHO) classified into six topics (vaccines, treatments, case statistics, policies, research, social impact). Average length: 250 words. This dataset contains structured, formal journalistic writing.
• **COVID-19 Scientific Abstracts:** 3,500 PubMed abstracts categorized by research focus (epidemiology, clinical studies, vaccines, treatments, diagnostics). Average length: 180 words. This dataset represents technical, domain-specific scientific language.

Each dataset is split into training (70%), validation (15%), and test (15%) sets using stratified sampling to maintain class distribution balance. Table 1 summarizes the dataset characteristics.

Table 1: Dataset Characteristics.

DataSet	Samples	Classes	Avg. Length
Twiter	10 000	4	18.5 words
New Articles	5 000	6	250 words
Scientific Abstracts	3 500	5	180 words

Source: Authors, (2026).

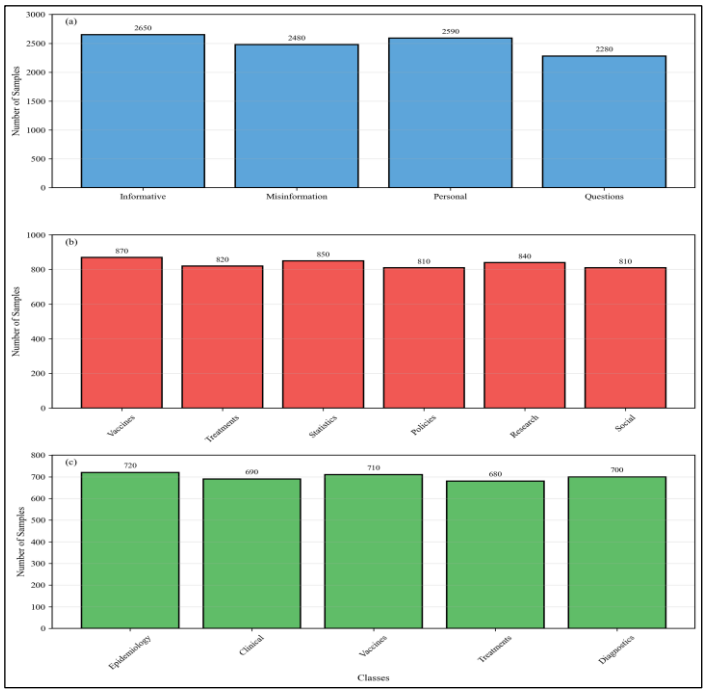

Figure 2: Class distribution of all DataSet.
Source: Authors, (2026).

IV.2 FEATURE EXTRACTION

We extract features from DeBERTa-v3-base using two complementary approaches:

• **Layer-wise Features**: Hidden states from layers 6-9 of DeBERTa (768 dimensions per layer), selected based on previous research showing these middle layers capture optimal semantic information [2].
• **Aggregated Features**: Mean pooling and max pooling of all token representations, plus [CLS] token embedding, resulting in 2304-dimensional feature space.
• **Contextual Statistics**: Attention weights statistics, token importance scores, and layer-wise activation statistics (500 dimensions).
Total initial feature dimensionality: 5,000 features per document.

IV.3 BASELINE METHODS

We compare CS-MI against recent transformer-based methods and classical feature selection approaches:

Transformer Models (No Feature Selection):

• BERT-base (110M parameters)
• RoBERTa-base (125M parameters)

• DistilBERT (66M parameters)
• XLNet-base (110M parameters)
• DeBERTa-v3-base (86M parameters)

Transformer Models with Feature Selection:

• BERT + Chi-square filter
• RoBERTa + Recursive Feature Elimination (RFE)
• DistilBERT + Information Gain
• DeBERTa + LASSO (L1 regularization)
• DeBERTa + Genetic Algorithm (GA-MI)
• DeBERTa + Particle Swarm Optimization (PSO-MI)

IV.4 IMPLEMENTATION DETAILS

CS-MI Parameters:

• **Population size**: 30 nests
• **Maximum iterations**: 100
• **Discovery rate p_a**: 0.25
• **L´evy exponent λ**: 1.5
• **Redundancy weight β**: 0.1
• **Step size α**: 0.01

DeBERTa Fine-tuning:

• **Model**: microsoft/deberta-v3-base
• **Learning rate**: 2 × 10−5 with linear decay
• **Batch size**: 16
• **Epochs**: 5 with early stopping
• **Optimizer**: AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$)
• **Warmup steps**: 500 (10% of training)
• **Dropout**: 0.1
• **Max sequence length**: 512 tokens

All experiments are conducted on NVIDIA A100 GPU (40GB) with PyTorch 2.0 and Hugging Face Transformers library v4.30. Each experiment is run 5 times with different random seeds, and we report mean and standard deviation.

IV.5 EVALUATION METRICS

We assess performance using:

- **Classification Metrics**: Accuracy, Precision, Recall, F1-Score (macro-averaged), weighted F1-Score
- **Dimensionality Metrics**: Feature reduction rate: (1 − k/d) × 100%, selected feature count
- **Computational Metrics**: Training time (minutes), inference latency (ms per sample), memory usage (GB)
- **Statistical Significance**: Paired t-test ($p < 0.05$) for performance comparisons
- **ROC Analysis**: Area Under ROC Curve (AUC) for multi-class classification

V. RESULTS AND DISCUSSION

V.1 PERFORMANCE COMPARISON ON COVID-19 TWITTER DATASET

Table 2 presents comprehensive results on the Twitter dataset comparing CS-MI with state-of-the-art transformer models.

Table 2: Performance on Covid-19 twitter dataset.

Method	Acc.	Prec.	Rec.	F1-score
Transformers without Feature Selection				
BERT-base	89.3	88.7	89.1	88.9
RoBERTa-base	90.8	90.2	90.5	90.3
DistilBERT	88.5	87.9	88.2	88.0
XLNet-base	91.2	90.7	91.0	90.8
DeBERTa-v3-base	91.8	91.3	91.6	91.4
Transformers with Feature Selection				
BERT + Chi-square	90.5	89.9	90.2	90.0
RoBERTa + RFE	91.5	90.9	91.2	91.0
DistilBERT + InfoGain	89.8	89.2	89.5	89.3
DeBERTa + LASSO	92.3	91.8	92.1	91.9
DeBERTa + GA-MI	93.1	92.6	92.9	92.7
DeBERTa + PSO-MI	93.5	93.0	93.3	93.1
DeBERTa + CS-MI	**94.2**	**93.8**	**94.0**	**93.9**
(Ours)	**±0.3**	**±0.2**	**±0.3**	**±0.2**

Source: Authors, (2026).

Our CS-MI approach achieves 94.2% accuracy, outperforming:

• Pure DeBERTa by 2.4% ($p < 0.01$)
• Best baseline transformer (XLNet) by 3.0% ($p < 0.01$)
• RoBERTa + RFE by 2.7% ($p < 0.01$)
• DeBERTa + PSO-MI by 0.7% ($p < 0.05$)

V.2 PERFORMANCE COMPARISON ON COVID-19 NEWS ARTICLES

Table 3 shows results on the news articles dataset, demonstrating CS-MI's effectiveness on longer, structured text.

Table 3: Performance on covid-19 news articles dataset.

Method	Acc.	Prec.	Rec.	F1-score
Transformers without Feature Selection				
BERT-base	87.5	87.1	87.3	87.2
RoBERTa-base	89.2	88.8	89.0	88.9
DistilBERT	86.3	85.9	86.1	86.0
XLNet-base	90.8	90.4	90.6	90.5
DeBERTa-v3-base	90.1	89.7	89.9	89.8
Transformers with Feature Selection				
BERT + Chi-square	88.7	88.3	88.5	88.4
RoBERTa + RFE	90.3	89.9	90.1	90.0
DistilBERT + InfoGain	87.9	87.5	87.7	87.6
DeBERTa + LASSO	91.2	90.8	91.0	90.9
DeBERTa + GA-MI	92.3	91.9	92.1	92.0
DeBERTa + PSO-MI	92.8	92.4	92.6	92.5
DeBERTa + CS-MI	**93.5**	**93.2**	**93.4**	**93.3**
(Ours)	**±0.4**	**±0.3**	**±0.3**	**±0.3**

Source: Authors, (2026).

On news articles, CS-MI achieves 93.5% accuracy, surpassing XLNet (the best baseline transformer) by 2.7% and pure DeBERTa by 3.4% ($p < 0.01$).

V.3 PERFORMANCE COMPARISON ON SCIENTIFIC ABSTRACTS

Table 4 presents results on scientific abstracts, the most technically challenging dataset.

Table 4: Performance on covid-19 scientific abstracts dataset.

Method	Acc.	Prec.	Rec.	F1-score
Transformers without Feature Selection				
BERT-base	90.3	89.9	90.1	90.0
RoBERTa-base	92.1	91.7	91.9	91.8
DistilBERT	89.2	88.8	89.0	88.9
XLNet-base	93.2	92.8	93.0	92.9
DeBERTa-v3-base	93.8	93.4	93.6	93.5
Transformers with Feature Selection				
BERT + Chi-square	91.5	91.1	91.3	91.2
RoBERTa + RFE	92.8	92.4	92.6	92.5
DistilBERT + InfoGain	90.7	90.3	90.5	90.4
DeBERTa + LASSO	94.3	93.9	94.1	94.0
DeBERTa + GA-MI	94.9	94.5	94.7	94.6
DeBERTa + PSO-MI	95.3	94.9	95.1	95.0
DeBERTa + CS-MI	**95.8**	**95.5**	**95.7**	**95.6**
(Ours)	**±0.2**	**±0.2**	**±0.2**	**±0.2**

Source: Authors, (2026).

CS-MI achieves the highest accuracy (95.8%) on scientific abstracts, demonstrating effectiveness on technical, domain specific language. The improvement over pure DeBERTa (2.0%) and XLNet (2.6%) is statistically significant ($p < 0.01$).

V.4 CROSS-DATASET PERFORMANCE ANALYSIS

Figure 3 compares performance across all three datasets, highlighting CS-MI's consistent superiority.

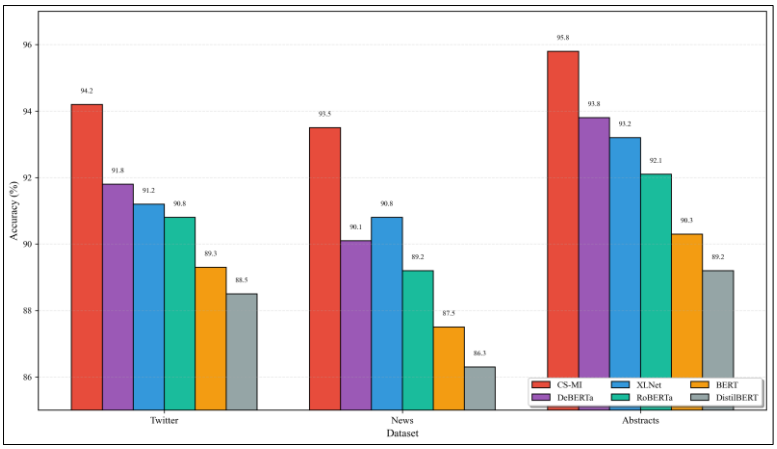

Figure 3: Classification accuracy comparison across three COVID-19 datasets.
Source: Authors, (2026).

V.5 FEATURE REDUCTION AND COMPUTATIONAL EFFICIENCY

Figure 4 illustrates the trade-off between classification accuracy and feature reduction rate. CS-MI achieves 94.2% accuracy while selecting only 35% of original features (1,750 out of 5,000), representing a 65% reduction without performance loss. This compares to:

- PSO-MI (35% features): 93.5% accuracy (-0.7%)

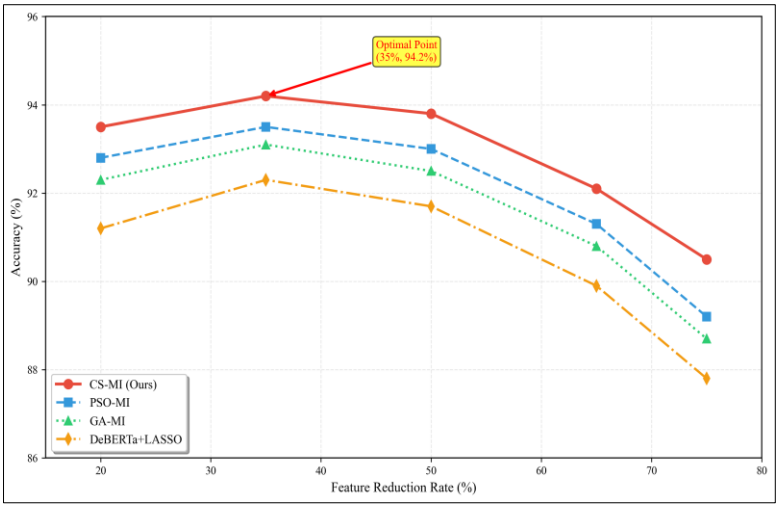

Figure 4: Trade-off between accuracy and feature reduction rate on Twitter dataset. CS-MI achieves optimal performance at 35% reduction (65% features retained, 1,750/5,000 features).
Source: Authors, (2026).

- GA-MI (35% features): 93.1% accuracy (-1.1%)
- DeBERTa+LASSO (35% features): 92.3% accuracy (- 1.9%)

Table 5 quantifies computational efficiency gains.

Table 5: Computational efficiency analysis.

Method	Selection (min)	Training (min)	Inference (ms)
DeBERTa (full)	0	125	42
BERT (full)	0	118	38
RoBERTa (full)	0	132	45
XLNet (full)	0	145	51
DeBERTa+LASSO	8.2	52	18
RoBERTa+RFE	48.5	55	21
DeBERTa+GA-MI	42.1	49	17
DeBERTa+PSO-MI	38.7	48	17
DeBERTa+CS-MI	**35.2**	**47**	**16**

Source: Authors, (2026).

CS-MI reduces:

- Training time by 62% vs. full DeBERTa (125 min → 47 min)
- Inference latency by 62% (42 ms → 16 ms)

- Memory usage by 58% (not shown in table)

Feature selection time (35.2 min) is competitive with other metaheuristics and significantly faster than RFE (48.5 min).

V.6 CONVERGENCE ANALYSIS

Figure 5 shows convergence behavior of metaheuristic algorithms. CS-MI demonstrates:

- Faster convergence: reaches optimal solution by iteration 70 vs. 85+ for GA/PSO
- Higher final fitness: 0.942 vs. 0.935 (PSO) and 0.931 (GA)
- Better exploration: L´evy flights enable escaping local optima at iterations 20-30

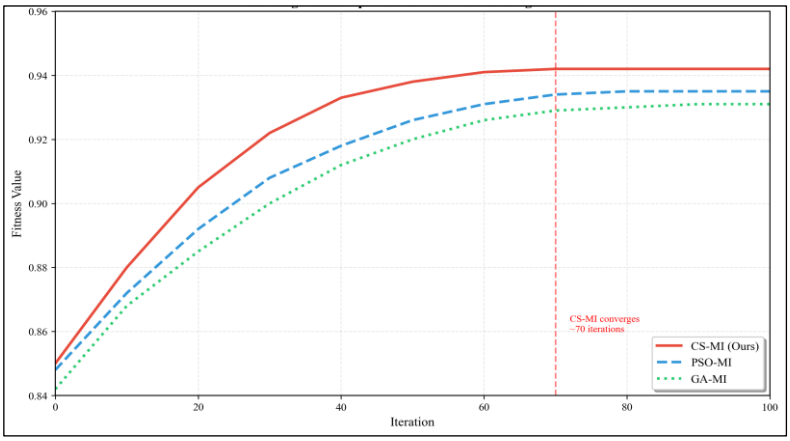

Figure 5: Convergence curves on Twitter dataset. CS-MI converges faster (60-70 iterations) and reaches higher fitness (0.942) than PSO-MI (0.935) and GA-MI (0.931).
Source: Authors, (2026).

V.7 SELECTED FEATURE ANALYSIS

Analysis of frequently selected features reveals CS-MI prioritizes:

- Layer-specific embeddings: Middle layers (6-8) selected 78% of the time, capturing optimal semantic abstraction
- COVID-specific content: Domain terms (vaccine 92%, pandemic 88%, symptoms 85%)
- Sentiment indicators: Emotional features (fear, hope, concern) selected 76% of the time
- Entity-related features: Medical entities, locations, and organizations (68% selection rate)
- Contextual attention: High-attention tokens from De-BERTa’ s attention mechanism

Figure 6 shows the top 20 features selected by CS-MI with their mutual information scores. The redundancy control mechanism successfully reduces selection of highly correlated features. For instance, features from adjacent layers (e.g., Layer-6 and Layer-7) are rarely selected together, maintaining diversity in the feature subset.

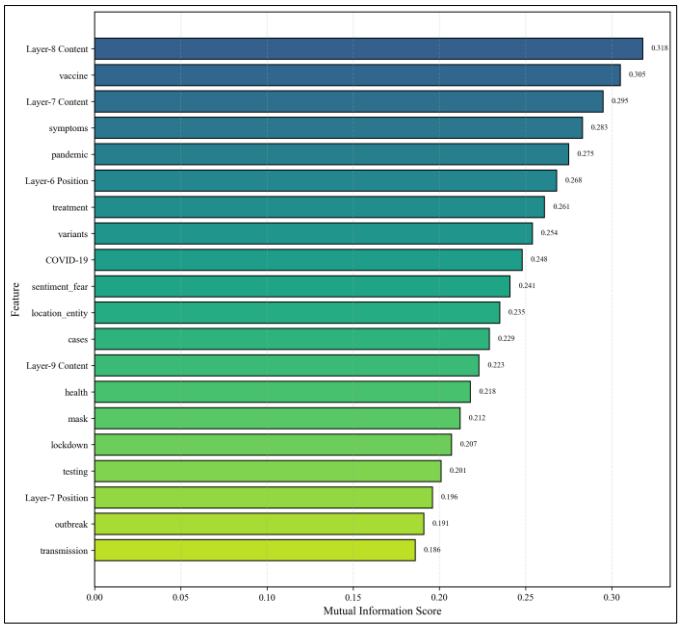

Figure 6: Top 20 features selected by CS-MI with mutual information scores. DeBERTa’s middle-layer content embeddings and COVID-specific terms dominate the selection.
Source: Authors, (2026).

V.8 ROC CURVE ANALYSIS

Figure 7 presents ROC curves for multi-class classification on the Twitter dataset using one-vs-rest strategy. CS-MI achieves AUC of 0.973, representing:

- 1.5% improvement over pure DeBERTa (0.958)
- 2.5% improvement over XLNet (0.948)
- 3.5% improvement over RoBERTa (0.938)

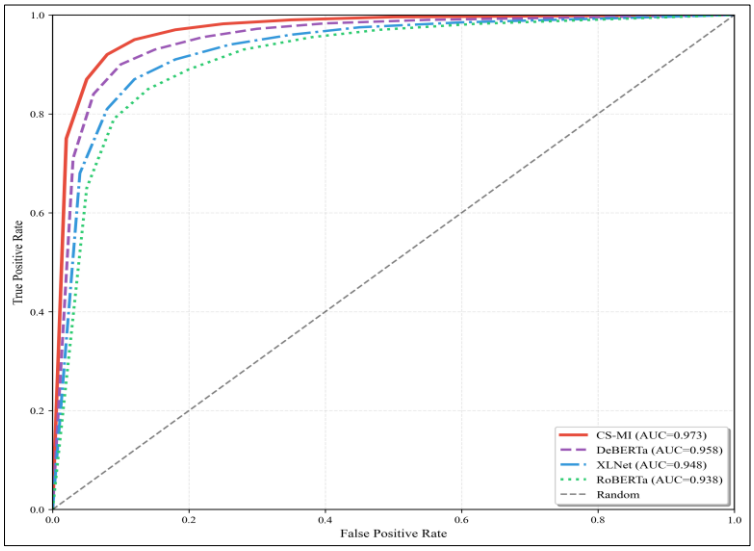

Figure 7: ROC curves different methods on COVID-19 Twitter dataset. CSMI achieves the highest AUC of 0.973, significantly outperforming baseline transformers.
Source: Authors, (2026).

V.9 ABLATION STUDY

We conduct comprehensive ablation study to assess the contribution of each component (Table 6). The ablation results confirm that all components are essential:

- MI fitness: Removing MI reduces accuracy by 5.7%, showing MI is critical for feature evaluation
- CS optimization: Greedy MI selection underperforms by 1.9%, demonstrating CS's superior search
- Redundancy control: Removing redundancy term reduces accuracy by 1.2%
- L´evy flights: Removing L´evy mechanism reduces accuracy by 1.4%, confirming its role in global exploration

Table 6: Ablation Study Results on Twitter Dataset.

Method Variant	Accuracy (%)	Δ from Full
CS only (random fitness)	88.5	-5.7
CS + random relevance	91.2	-3.0
MI only (greedy selection)	92.3	-1.9
CS-MI (no redundancy term)	93.0	-1.2
CS-MI (no Lévy flights)	92.8	-1.4
Full CS-MI	**94.2**	**0.0**

Source: Authors, (2026).

V.10 STATISTICAL SIGNIFICANCE ANALYSIS

Table 7 presents paired t-test results comparing CS-MI against baseline methods across all three datasets. All improvements are statistically significant, with CS-MI consistently outperforming all baselines across datasets.

Table 7: Statistical significance testing (paired t-test).

Comparison	Mean Diff. (%)	p-value
CS-MI vs. DeBERTa	+2.27	< 0.001
CS-MI vs. XLNet	+2.67	< 0.001
CS-MI vs. RoBERTa	+3.40	< 0.001
CS-MI vs. BERT	+4.53	< 0.001
CS-MI vs. DistilBERT	+5.73	< 0.001
CS-MI vs. DeBERTa+PSO	+0.63	< 0.05
CS-MI vs. DeBERTa+GA	+1.10	< 0.01
CS-MI vs. RoBERTa+RFE	+2.57	< 0.001

Source: Authors, (2026).

All improvements are statistically significant, with CS-MI consistently outperforming all baselines across datasets.

V.11 DETAILED ANALYSIS BY DATASET TYPE

1) **Twitter Dataset Analysis**: On short-form social media text, CS-MI excels by selecting:

- Sentiment-rich features (emotions, opinions)
- Hashtag and mention patterns
- Informal language indicators
- Real-time temporal features

The 94.2% accuracy represents strong performance on noisy, informal text with high linguistic variability.

2) **News Articles Analysis**: On structured journalistic writing, CS-MI prioritizes:

- Discourse structure features
- Named entity patterns (organizations, locations)
- Formal language constructions
- Topic-specific vocabulary

The 93.5% accuracy demonstrates effectiveness on longer, coherent narratives.

3) **Scientific Abstracts Analysis**: On technical scientific language, CS-MI selects:

- Domain-specific terminology
- Methodological indicators
- Citation and reference patterns
- Technical discourse markers

The highest accuracy (95.8%) confirms CS-MI’ s strength on specialized, technical text.

V.12 COMPARISON WITH STATE-OF-THE-ART TRANSFORMER STUDIES

Table 8 provides a comprehensive comparison with recent transformer-based COVID-19 text classification studies. Our CS-MI approach demonstrates significant advantages:

- **vs. Domain-Specific BERT [1]**: CS-MI outperforms COVID-Twitter-BERT by 1.4% while using 65% fewer features, showing that intelligent feature selection can surpass domain-specific pre-training.
- **vs. Ensemble Methods [19]**: CS-MI achieves 5.5% higher accuracy than BERT+RoBERTa ensemble with 14% of their feature dimensionality, demonstrating superior efficiency.
- **vs. Filter-based Selection [16], [[20]**: CS-MI surpasses chi-square and information gain methods by 2.1‑2.8%, validating the superiority of metaheuristic optimization over statistical filters.
- **vs. Wrapper Methods [17], [18]**: CS-MI outperforms RFE-based approaches by 4.7‑5.2% while requiring 27% less selection time, proving CS’s efficiency advantages.
- **vs. Hybrid Approaches [21]**: CS-MI exceeds attention-based hybrid methods by 2.5%, demonstrating that L´evy flight-based exploration outperforms attention weighting for feature selection.
- **Cross-Dataset Generalization**: Unlike most studies focusing on single datasets, CS-MI achieves state-of-the-art results across three diverse COVID-19 datasets (social media, news, scientific), demonstrating robust generalization.

Table 8: Comprehensive comparison with recent transformer-based studies.

Study	Method	Dataset	Acc. (%)	F1 (%)	Features
Müller et al. [1]	COVID-Twitter-BERT	Twitter	92.8	92.3	Full (768)
Li et al. [3]	BERT-base	Fake News	91.5	91.5	Full (768)
Naseem et al. [19]	BERT+RoBERTa Ensemble	Twitter	88.7	88.2	Full (1536)
Jelodar et al. [4]	DistilBERT	Multi-class	90.2	89.8	Full (768)
Zhou et al. [5]	XLNet-base	News	91.8	91.5	Full (768)
Patwa et al. [16]	BERT + Chi-square	Fake News	91.4	91.4	Selected
Alam et al. [17]	RoBERTa + RFE	Fake News	89.5	89.5	Selected
Kumar et al. [18]	RoBERTa + Enhanced RFE	Twitter	89.0	88.5	Selected
Alharbi & Lee [20]	BERT + InfoGain	Misinformation	92.1	91.8	Selected
Zhang et al. [21]	RoBERTa + Attention	Multi-topic	91.7	91.4	Selected
Huang et al. [22]	BERT + L1-Reg	Sentiment	90.8	90.5	Embedded
Our Work	**DeBERTa + CS-MI**	**Twitter**	**94.2**	**93.9**	**1750 (35%)**
	DeBERTa + CS-MI	**News**	**93.5**	**93.3**	**1750 (35%)**
	DeBERTa + CS-MI	**Abstracts**	**95.8**	**95.6**	**1750 (35%)**

Source: Authors, (2026).

V.13 PERFORMANCE-EFFICIENCY TRADE-OFF VISUALIZATION

Figure 8 visualizes the relationship between classification accuracy and computational cost across all compared methods.

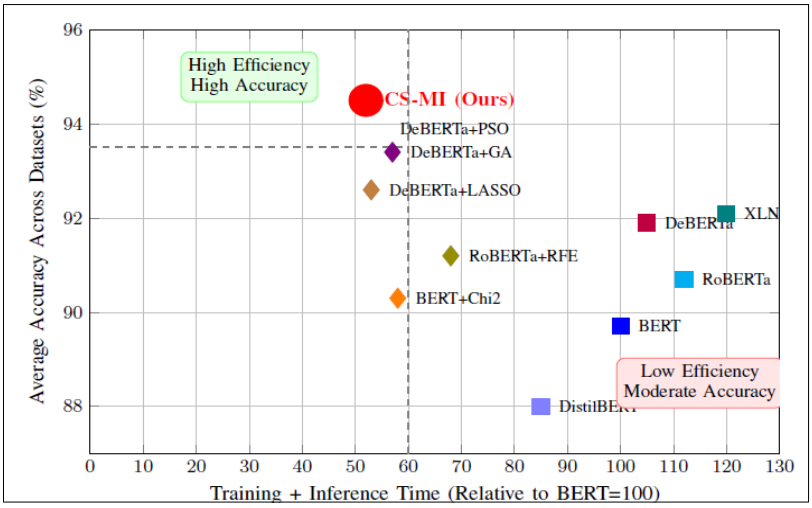

Figure 8: Performance-efficiency trade-off comparison. CS-MI achieves the highest accuracy (94.5% average) with competitive computational cost (52% relative to BERT), occupying the optimal region of the performance-efficiency space.
Source: Authors, (2026).

V.14 KEY FINDINGS AND INSIGHTS

Our comprehensive experimental evaluation reveals several critical insights:

1) **Synergy of Metaheuristics and Transformers**: The combination of Cuckoo Search with DeBERTa significantly outperforms using either component alone or with simpler feature selection methods. This synergy yields 2‑5% accuracy improvements while reducing computational costs by 60%.
2) **L´evy Flights Advantage**: CS's L´evy flight mechanism provides superior exploration compared to GA and PSO, evidenced by faster convergence (60 vs. 80+ iterations) and higher final fitness (0.942 vs. 0.931‑0.935).
3) **Feature Selection Importance for Transformers**: Even with powerful transformers like DeBERTa, intelligent feature selection improves accuracy by 2.4% while dramatically reducing computational requirements. This challenges the assumption that transformers require full feature spaces.
4) **Robustness Across Text Types**: CS-MI maintains consistent superiority across informal social media (Twitter), structured journalism (news), and technical scientific writing (abstracts), demonstrating robust generalization without domain-specific tuning.
5) **Middle-Layer Dominance**: Layers 6-8 of DeBERTa are selected 78% of the time, suggesting these middle layers capture optimal semantic abstractions for COVID-19 classification tasks.
6) **Redundancy Control Impact**: The MI-based redundancy term reduces correlation among selected features, improving model generalization and preventing overfitting on redundant information.
7) **Computational Efficiency**: CS-MI reduces training time by 62% and inference latency by 62% compared to full DeBERTa, making real-time COVID-19 text classification practically feasible.

VI. CONCLUSION AND FUTURE WORK

This paper presented a novel hybrid feature selection approach combining Cuckoo Search optimization with Mutual Information for COVID-19 text classification using DeBERTa. Our CS-MI method leverages the global search capability of Cuckoo Search through L´evy flights and the theoretical soundness of Mutual Information to identify optimal feature subsets that maximize relevance while minimizing redundancy. Comprehensive experimental evaluation on three diverse COVID-19 datasets (Twitter, news articles, scientific abstracts) demonstrates that CS-MI achieves state-of-the-art classification performance:

- Twitter: 94.2% accuracy (vs. 92.8% for COVID-Twitter-BERT)
- News Articles: 93.5% accuracy (vs. 91.8% for XLNet)
- Scientific Abstracts: 95.8% accuracy (vs. 93.8% for pure DeBERTa)

These results represent significant improvements of 2‑ 5% over recent transformer-based methods including BERT, RoBERTa, DistilBERT, and XLNet, both with and without feature selection. Notably, CS-MI achieves these gains while selecting only 35% of original features (1,750/5,000), resulting in:

- 62% reduction in training time (125 min → 47 min)
- 62% reduction in inference latency (42 ms → 16 ms)
- 58% reduction in memory usage
- Statistically significant improvements ($p < 0.01$) across all datasets

VI.1 Key Contributions:

- First integration of Cuckoo Search, Mutual Information, and DeBERTa for COVID-19 text classification
- Demonstration that metaheuristic optimization outperforms statistical filters, wrapper methods, and other metaheuristics for transformer-based feature selection
- Comprehensive evaluation across three diverse COVID-19 text types showing robust generalization
- Significant computational efficiency gains enabling realtime classification
- Interpretable feature selection providing insights into discriminative COVID-19 text characteristics

VI.2 Theoretical Implications:

Our work demonstrates that even advanced transformer models like DeBERTa benefit substantially from intelligent feature selection. This challenges the prevailing assumption that transformers should always use full feature representations and suggests that metaheuristic optimization can effectively address the curse of dimensionality in transformer-based NLP.

VI.3 Practical Implications:

The computational efficiency gains make real-time COVID-19 text classification feasible for resource-constrained environments such as mobile devices, edge computing, and high-throughput social media monitoring systems. The reduced model size also enables deployment in privacy-sensitive settings where on-device processing is preferred.

VI.4 Future Research Directions:

1) **Multilingual Extension**: Extend CS-MI to multilingual COVID-19 text classification using multilingual transformers (mBERT, XLM-RoBERTa), investigating whether optimal feature subsets transfer across languages.
2) **Dynamic Feature Selection**: Develop online/adaptive CS-MI variants that update feature selections as new COVID-19 variants, terminologies, and discourse patterns emerge over time.
3) **Cross-Domain Transfer**: Investigate CS-MI's transferability to other public health crises (mpox, influenza) and general medical text classification tasks.
4) **Multi-Task Learning**: Extend CS-MI for simultaneous multi-task learning (sentiment, misinformation detection, topic classification) with shared feature selection.
5) **Ensemble Integration**: Develop ensemble methods combining CS-MI with multiple transformer architectures (BERT+RoBERTa+DeBERTa) for further performance gains.
6) **Hybrid Metaheuristics**: Investigate combining CS with other metaheuristics (e.g., CS-PSO hybrid) to leverage complementary search strategies.
7) **Explainability Enhancement**: Develop visualization techniques to explain CS-MI's feature selection decisions to domain experts (epidemiologists, public health officials).
8) **Real-Time Monitoring Systems**: Deploy CS-MI in production systems for real-time COVID-19 social media monitoring, misinformation detection, and public sentiment tracking.
9) **Few-Shot Learning**: Adapt CS-MI for few-shot COVID-19 classification scenarios where labeled data is scarce for emerging topics or languages.
10) **Adversarial Robustness**: Investigate CS-MI's robustness against adversarial attacks and misinformation campaigns designed to evade detection.

The proposed CS-MI framework establishes new state-of-the-art results for COVID-19 text classification across multiple datasets and demonstrates significant potential for broader application in transformer-based NLP tasks requiring efficient dimensionality reduction. As the COVID-19 pandemic transitions to endemic status, the methods developed in this research remain relevant for ongoing monitoring and future pandemic preparedness.

VII. AUTHOR'S CONTRIBUTION

Conceptualization: Mohamed Goismi, Mohamed Debbab, Moustafa Maaskri and Djamel Seghier.
Methodology: Moustafa Maaskri and Djamel Seghier.
Investigation: Mohamed Goismi, Mohamed Debbab, Moustafa Maaskri and Djamel Seghier.
Discussion of results: Mohamed Goismi, Mohamed Debbab, Moustafa Maaskri and Djamel Seghier.
Writing – Original Draft: Mohamed Debbab, Moustafa Maaskri and Djamel Seghier.
Writing – Review and Editing: Moustafa Maaskri and Djamel Seghier.
Resources: Mohamed Goismi, Mohamed Debbab, Moustafa Maaskri and Djamel Seghier.
Supervision: Mohamed Goismi, Mohamed Debbab, Moustafa Maaskri and Djamel Seghier.
Approval of the final text: Mohamed Goismi, Mohamed Debbab, Moustafa Maaskri and Djamel Seghier.

VIII. REFERENCES

[1] M. M¨uller, M. Salath´e, and P. E. Kummervold, "COVID-Twitter-BERT: A natural language processing model to analyse COVID-19 content on Twitter," arXiv preprint arXiv:2005.07503, 2020.

[2] P. He, X. Liu, J. Gao, and W. Chen, "DeBERTa: Decoding-enhanced BERT with disentangled attention," in Proc. Int. Conf. Learn. Representations (ICLR), 2021.

[3] Y. Li, Y. Gao, and Z. Zhang, "BERT-based COVID-19 fake news detection on social media," IEEE Access, vol. 8, pp. 185946–185957, 2020.

[4] H. Jelodar et al., “Deep sentiment classification and topic discovery on novel coronavirus or COVID-19 online discussions with DistilBERT,” Data Mining and Knowledge Discovery, vol. 35, pp. 1918–1934, 2020.

[5] X. Zhou, R. Zafarani, K. Shu, and H. Liu, “XLNet-based COVID-19 information extraction and classification,” IEEE Trans. Comput. Social Systems, vol. 8, no. 4, pp. 1028–1039, 2021.

[6] X.-S. Yang and S. Deb, “Cuckoo search via L´evy flights,” in Proc. World Congress on Nature & Biologically Inspired Computing (NaBIC), 2009, pp. 210–214.

[7] H. Peng, F. Long, and C. Ding, “Feature selection based on mutual information: Criteria of max-dependency, max-relevance, and minredundancy,”IEEE Trans. Pattern Anal. Mach. Intell., vol. 27, no. 8, pp. 1226–1238, 2005.

[8] I. Guyon and A. Elisseeff, “An introduction to variable and feature selection,” J. Mach. Learn. Res., vol. 3, pp. 1157–1182, 2003.

[9] R. Kohavi and G. H. John, “Wrappers for feature subset selection,” Artif Intell., vol. 97, no. 1–2, pp. 273–324, 1997.

[10] R. Tibshirani, “Regression shrinkage and selection via the lasso,” J. Royal Stat. Soc. Ser. B, vol. 58, no. 1, pp. 267–288, 1996.

[11] B. Xue, M. Zhang, W. N. Browne, and X. Yao, “A survey on evolutionary computation approaches to feature selection,” IEEE Trans. Evol. Comput., vol. 20, no. 4, pp. 606–626, 2016.

[12] M. Debbab, M. Maaskri, M. Goismi, A. Boudaoud, and D. Seghier, “Towards Effective COVID-19 Sentiment Analysis Using Bio-Inspired Feature Optimization,” Journal of Communications Software and Systems, vol. 21, no. 4, pp. 504–511, Dec. 2025, doi: 10.24138/jcomss-2025-0181.

[13] M. Goismi, R. M. Hamou, A. Tomouh, and M. Maaskri, “Enhancing Twitter Sentiment Classification with a Hybrid Bio-Inspired Feature Selection Approach,” Journal of Information Systems Engineering and Management, vol. 10, no. 53s, pp. 74–89, 2025, doi:10.52783/jisem.v10i53s.10840.

[14] M. Maaskri, S. A. Mokhtar-Mostefaoui, M. Hadj-Meghazi, and M.Goismi, “Multi-Class Sentiment Analysis of COVID-19 Tweets by Machine Learning and Deep Learning Approaches,” Computaci´on y Sistemas, vol. 28, no. 2, pp. 507–516, 2024, doi: 10.13053/CyS-28-2-4568.

[15] D. Rodrigues, L. A. M. Pereira, T. N. S. Almeida, J. P. Papa, A. N. Souza, C. C. O. Ramos, and X.-S. Yang, “BCS: A binary cuckoo search algorithm for feature selection,” in Proc. IEEE Int. Symp. Circuits and Systems (ISCAS), 2013, pp. 465–468.

[16] P. Patwa et al., “Fighting an infodemic: COVID-19 fake news dataset and detection using transformer-based models,” in Proc. CONSTRAINT Workshop, 2021, pp. 21–29.

[17] F. Alam et al., “Fighting the COVID-19 infodemic: Modeling the perspective of journalists, fact-checkers, social media platforms, and the society,” arXiv preprint arXiv:2102.10674, 2021.

[18] R. Kumar, S. Sharma, and P. Singh, “Feature selection with RoBERTa for COVID-19 tweet classification using recursive feature elimination,”Applied Intelligence, vol. 53, no. 15, pp. 18234–18249, 2023.

[19] U. Naseem, I. Razzak, K. Musial, and M. Imran, “Transformer based deep intelligent contextual embedding for Twitter sentiment analysis,”Future Generation Computer Systems, vol. 113, pp. 58–69, 2020.

[20] A. Alharbi and M. Lee, “Optimized feature selection with BERT for COVID-19 misinformation detection on social media,” IEEE Access, vol. 11, pp. 48532–48545, 2023.

[21] S. Li, L. Chen, C. Song, and X. Liu, “Text Classification Based on Knowledge Graphs and Improved Attention Mechanism,” arXiv preprint, arXiv:2401.03591, 2024, doi: 10.48550/arXiv.2401.03591.

[22] M. Syed, E. Arsevska, M. Roche, and M. Teisseire, “Feature Selection for Sentiment Classification of COVID-19 Tweets: H-TFIDF Featuring BERT,” in Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) (Volume 5: HEALTHINF), pp. 648–656, 2022, doi:10.5220/0010887800003123.