Journal of Information Systems and Telecommunication
Special Issue, Winter,2026, PP 39-49
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

# Detecting Synchronized Hate Speech in Online Social Networks via Social Synchrony and Ant Colony Optimization

Shabana Nargis Rasool[1], Sarika Jain[2*], Ajay Vikram Singh[2]

[1].Department of Computer Science, Islamic University of Science and Technology, Kashmir, India

[2].Amity Institute of Information Technology, Amity University Noida, Uttar Pradesh, India.

Received: 25 Jul 2025/ Revised: 04 Dec 2025/ Accepted: 06 Jan 2026

**Abstract**

Online platforms have become fertile grounds for hate speech, often spreading through bursts of coordinated user activity. Detecting such patterns requires more than analyzing individual posts, as it calls for understanding the collective rhythm of online interactions. In the present study, we present SIACO (Social Synchrony Identification using Ant Colony Optimization), a nature-inspired framework that detects hate-speech events by tracing synchrony in user behaviour. SIACO models how hateful expressions emerge and fade collectively, using Ant Colony Optimization to refine linguistic features and improve classification accuracy. Upon evaluation on a Twitter dataset, the framework consistently outperforms both traditional machine learning models and transformer-based baselines, achieving up to a 10% improvement across major evaluation metrics. The framework also offers interpretable insights into the linguistic and temporal cues driving coordinated hate. The performance scores obtained highlight the value of looking at hate speech not just as text, but as a social phenomenon unfolding in synchrony.

**Keywords:** Hate Speech; Social Synchrony; Ant Colony Optimization; Feature Selection; Online Social Networks.

## 1- Introduction

The extensive prevalence of Online Social Networks (OSNs) is demonstrated by a recent poll conducted by Nielsen Online [1], which revealed that social media has surpassed email as the most dominant online activity. More than two-thirds of the global Internet population now engages with social media and blogs, accounting for nearly 10% of all Internet use. Platforms such as LinkedIn, Facebook, and Twitter have become indispensable to modern digital life, facilitating interactions, information exchange, and content discovery across personal, professional, and social contexts. In today's fast-paced world, where time constraints limit face-to-face interactions, online communication offers a vital alternative for maintaining relationships and expressing opinions. Individuals can share thoughts, exchange knowledge, and build communities that transcend geographic and temporal barriers. The influence of social media on society continues to expand and shows no signs of diminishing; platforms like Twitter, in particular, have become powerful arenas for the public exchange of ideas and emotions [2].

While OSNs empower participation and connectivity, they also serve as conduits for hostility and discrimination [3]. On the other hand, the mechanisms that amplify positive engagement, virality, collective attention, and immediacy can also accelerate the spread of hate speech. Online hate speech represents more than an aggregation of individual acts of incivility; it is often collectively orchestrated, arising through synchronized bursts of hostile communication [4]. Conventional hate-speech detection methods typically examine content at the level of single posts or users, emphasizing textual features and linguistic cues. Such approaches overlook the 'social dynamics', that are the patterns of coordination and reinforcement through which harmful discourse proliferates [5]. It may be hypothesized that incorporating social synchrony signals and optimization-based feature selection can significantly improve the accuracy and robustness of hate-speech detection compared to content-only baselines.

Human behavior, whether offline or online, is inherently synchronized. The coordinated surges in online activity,

✉ **Sarika Jain**
Ashusarika@gmail.com

where users post, retweet, or comment in temporal alignment, reflect a phenomenon termed 'social synchrony' [6]. On social networks, synchrony manifests as the collective rhythm of communication, moments when users act in harmony, consciously or unconsciously, around shared sentiments or ideological themes. Understanding this synchrony provides a valuable lens through which to study the amplification of hate speech, revealing how seemingly independent expressions combine into collective waves of hostility. The present study introduces SIACO (Social Synchrony Identification using Ant Colony Optimization) [7]. This novel, nature-inspired computational framework reconceptualizes hate-speech detection as a problem of identifying collective behavioral patterns. The framework integrates social-science theory and computational intelligence to model the emergence of hateful discourse as an outcome of synchronized user behavior. Considering inspiration from the foraging behavior of ant colonies, self- organizing systems capable of discovering optimal solutions through pheromone-based communication, SIACO employs Ant Colony Optimization (ACO) to refine feature selection and classification in textual data. Through iterative learning, ACO identifies the most salient and contextually coherent linguistic features that characterize coordinated hate expression. This fusion of social synchrony analysis and swarm intelligence introduces a new paradigm for understanding and mitigating online hostility, offering a fresh perspective on hate speech detection [8].

The study is based on the hypothesis that hate speech in online networks exhibits detectable patterns of synchrony, which can be captured computationally to improve detection accuracy. By incorporating ACO into the learning process, models can more effectively reveal the underlying relationships between temporal, relational, and linguistic signals that traditional classifiers, including those based on transformer architectures, might overlook [9].

The proposed framework of SIACO integrates several sequential components in a comprehensive approach: data collection from Twitter; extensive text preprocessing to remove noise and standardize input; feature extraction using Bag-of-Words (BoW) and Term Frequency–Inverse Document Frequency (TF-IDF) representations; optimization via ACO; and classification through multiple supervised learning algorithms. This comprehensive approach reassures the audience about the thoroughness of the framework. The contribution of this work is two-fold. First, it advances the computational frontiers of hate-speech detection by introducing a hybrid optimization framework that enhances both performance and interpretability. Second, it offers a new conceptual understanding of online hostility, emphasizing hate speech as a 'collective social phenomenon' rather than an isolated textual one. This two-fold contribution not only enhances our understanding of hate speech in online networks but also provides a novel computational framework for its detection. The remainder of this paper is organized as follows: Section 2 reviews the related literature and identifies the research gap; Section 3 describes the proposed SIACO methodology, the experimental setup and evaluation metrices, Section 4 outlines results and discussion; and Section 5 concludes with limitations and future research directions.

## 2- Related Work

Research on event detection and social synchronization within online social networks (OSNs) has evolved along several intertwined paths, each contributing to understanding how people behave and interact in digital spaces. In the last decade, numerous studies have sought to capture the pulse of collective human behavior over online social networks, where the primary focus remains: how individual actions, when repeated and shared across vast user communities, form patterns of synchronization that reflect real-world coordination, emotion, and influence. Despite a rich progress in behavioral modeling and machine learning, relatively little attention has been given to the idea that hate speech itself can emerge as a synchronized, collective phenomenon.

One of the earliest efforts to model such interconnected human behaviour was proposed by De et al. [10], who developed a Dynamic Bayesian Network (DBN) model to predict user actions over time. Their approach emphasized the influence of one user's activity on another's, revealing that social interactions online are rarely independent and often unfold in rhythmic, interdependent ways. Similarly, Benevenuto et al. [11] analyzed a vast dataset combining four prominent OSNs: LinkedIn, Hi5, MySpace, and Orkut to understand patterns in user navigation and engagement. Their findings highlighted that user behavior across platforms often shows collective rhythms, even when interactions appear spontaneously. Building on these foundations, Rossi et al. [12] introduced large-scale, time-evolving graphs to examine how users move and cluster dynamically within networks, providing an analytical lens for detecting coordination over time. These early research efforts' models revealed the potential for studying synchrony in online behavior.

Rodríguez et al. [13] proposed using context ontologies to recognize and interpret human activity that offered a structured means of representing complex social interactions. This approach was extended by Rodríguez et al. [14], who developed a fuzzy ontology framework to handle uncertainty, ambiguity, and incomplete information in human behavior modeling.

The studies contributed valuable interpretative frameworks for reasoning about human activity; however, they focused on the individual rather than the collective patterns that emerge when users act in unison. In a similar effort, [15]

explored how influential users shape synchronization in social systems. Identifying these "seed users", those who initiate coordinated online activity, has long been recognized as a computationally complex problem. Weskida and Michalski [15] addressed this through an evolutionary algorithm capable of selecting optimal influencers with improved accuracy and efficiency. Zhao et al. [16] examined the role of tie strength in information diffusion, demonstrating that message propagation depends strongly on relational closeness and channel configuration. Similarly, Cordero et al. [17] proposed a logic-based framework for detecting influence on Twitter, quantifying how specific users drive conversations across topics. Huberman et al. [18] added another layer of insight by revealing that, although Twitter networks appear dense because of their large follower graphs, genuine friendship ties are sparse and selective. Together, these studies illustrate how the structure of relationships influences synchronization and information flow, hinting at deeper mechanisms that could also govern the spread of hate speech.

Recently, different studies have focused on coordination and synchrony from behavioral and physical perspectives, laying conceptual foundations relevant to computational modeling. For instance, Alderisio et al. [19] explored ensemble coordina- tion through an experimental setup that enabled individuals to synchronize their movements remotely, providing a controlled environment to study human synchrony. Song et al. [20] found that mobile-phone data could accurately predict human mobility patterns, showing that even complex behaviors exhibit measurable regularity. Xuan et al. [6] discovered that software developers in open-source communities demonstrate cyclical work patterns linked to software dependency graphs, essentially a digital analogy for how focus and coordination shift in social systems. These studies affirm that synchrony is not an abstract idea but a pervasive feature of human interaction that extends from physical spaces to digital environments. Complementing these behavioral insights, significant progress has also been made in event detection and analyzing large- scale text and social media streams. Li et al. [21] applied clustering techniques to detect bursts of word activity, stablishing early methods for identifying topical "events" in online discussions. Alvanaki et al. [22] refined this idea by tracking the co-occurrence of tags and prioritizing events based on the intensity and duration of topic burstiness. Leskovec et al. [23] examined how information cascades spread through blogs, highlighting the structural and temporal dynamics un- derlying viral dissemination. Petkos et al. [24] and Gaglio et al. [25] further enhanced this theme through Soft Frequent Pattern Mining (SFPM), which groups semantically related words to uncover emerging events. More recently, Ozdikis et al. [26] proposed a semantic event-detection framework for Twitter, employing multiple vectorization schemes that included semantic expansion of hashtags and keywords to improve clustering precision. These approaches collectively demonstrate the power of combining linguistic, temporal, and semantic cues to identify patterns of collective activity in real- time data streams.

Despite these achievements, a key limitation across most existing studies is that they treat online phenomena as discon nected units of analysis, rather than as expressions of coordinated social dynamics.

Even the most advanced deep-learning architectures, such as BERT and RoBERTa, primarily operate on textual semantics and fail to account for the relational synchrony among users who amplify hateful discourse. Similarly, while optimization algorithms like ACO have shown great promise in feature selection and routing areas, their potential for improving natural language processing and social-behavior modeling remains largely unexplored. Therefore, in the current study, the proposed SIACO framework aims to fill this void by merging the conceptual understanding of social synchrony with the computational efficiency of swarm intelligence.

## 3- Methodology

This section presents the architectural design, mathematical formulation, and operational workflow of the proposed framework: SIACO. This framework is a hybrid computational model that combines natural language processing, nature- inspired optimization, and supervised machine learning to detect and predict synchronized hate-speech activity within online social networks. The central presumption of SIACO is that hate speech on platforms such as Twitter does not emerge in isolation but exhibits temporal and relational synchrony, effectively the patterns of collective behavior that can be computationally modeled and optimized. A detailed architectural framework of SIACO is illustrated in Figure 1, which comprises a series of interdependent modules that transform raw social data into optimized predictive models. These modules include Data Acquisition, Preprocessing, Feature Representation, Optimization via Ant Colony Optimization, and Classification, which are arranged sequentially to enable an end-to-end analytical pipeline. Each stage uniquely contributes to the final prediction of synchronized hate-speech events, integrating linguistic and topological signals from input data.

### 3-1- Data Acquisition and Preprocessing

Raw tweet data are obtained from publicly available sources through the official Twitter API, filtered using predefined hate- speech and offensive-language keywords. The resulting corpus are parsed, annotated, and stored in a

structured format suitable for further analysis. To ensure analytical validity, the data is subjected to multiple preprocessing operations with an aim to reduce lexical noise and preserve temporal and behavioural cues that underpin social synchrony.

**Algorithm 1. Preprocessing of Input Tweet Data**

**Require:** Raw tweet set = T = $\{t_1, t_2, \ldots t_n\}$
**Ensure:** Cleaned corpus $X$
Initialize empty corpus $X \leftarrow \emptyset$
**for** each tweet $t_i \in$ T **do**
  Remove URLs, mentions, hashtags, emojis, and punctuation
  Convert text to lowercase
  Tokenize words and remove stopwords
  Apply part-of-speech tagging and lemmatization
  Normalize elongated words (e.g., "soooo" → "so")
  Preserve timestamp $\tau_i$ and user metadata $u_i$ for synchrony modelling
  Append preprocessed tweet $x_i$ to corpus
**end for**
**return** $X$

At a conceptual level, SIACO framework's design is guided by two complementary principles: (i) behavioural modelling of social synchrony, capturing correlated user actions and temporal co-occurrence; and (ii) optimization-driven intelligence which discovers compact, discriminative feature sets.

Mathematically, the entire SIACO pipeline can be formalized as a composite transformation:

$$S = C\left(\Phi_{ACO}(\Psi(X))\right) \quad (1)$$

where:

$X$ denotes the preprocessed input corpus from Algorithm 1,

$\Psi(.)$ is the feature-extraction operator,

$\Phi_{ACO}(.)$ represents the ACO feature-selection and weighting function, and

$C(.)$ is the final supervised classifier (e.g., SVM, RF, LR).

Thus, $S$ encapsulates the entire flow from unstructured text to optimized synchrony-aware classification. Interestingly, this structure mirrors a Graph Convolutional Network (GCN): while GCNs propagate information along edges connecting related nodes, in SIACO the pheromone trails act as probabilistic conduits of influence between correlated textual features, diffusing relevance signals across the linguistic network.

## 3-2- Feature Representation

Following preprocessing, tweets are transformed into numerical vectors through two complementary representations: BoW and TF–IDF. Given a document $d_i$ in the corpus D = $\{d_1, d, \ldots d_n\}$ with vocabulary V = $\{t_1, t_2, \ldots t_m\}$, its BoW representation is:

$$x_i = \{f_{i1}, f_{i2}, \ldots f_{im}\}, \quad (2)$$

where $f_{ij}$ denotes the frequency of term $t_j$ in document $d_i$. The TF–IDF weighting scheme further refines these frequencies by penalizing ubiquitous terms:

$$W_{ij} = tf_{ij} \times log\frac{N}{df_j}, \quad (3)$$

where $tf_{ij}$ is the term frequency of $t_j$ in $d_i$, $df_j$ is the number of documents containing $t_j$, and $N$ is the total number of documents in the corpus. The hybrid BoW–TF–IDF vectorization balances the frequency-driven simplicity of BoW with the discriminative weighting of TF–IDF, crucial for rare but semantically rich hate-speech tokens.

## 3-3- ACO-driven Feature Optimization

The ACO module identifies the optimal feature subset that maximizes classification performance while maintaining parsimony. The objective function is formulated as:

$$F^* = \underset{F \subseteq \mathcal{F}}{arg\,max}\left[\lambda_1 . Acc(F) + \lambda_2 . \frac{1}{|F|}\right], \quad (4)$$

subject to $0 < \rho < 1$, $\alpha, \beta > 0$, and $|F| \leq |\mathcal{F}|$, where $\lambda_1$ and $\lambda_2$ are trade-off parameters controlling the balance between accuracy and sparsity, $\rho$ denotes pheromone evaporation, and $\alpha$, $\beta$ influence the relative importance of pheromone versus heuristic information.

Each ant $k$ constructs a candidate subset $S_k$ based on the probability:

$$p_{k,j} = \frac{(\tau_j^{\alpha})(\eta_j^{\beta})}{\sum_{l \in F}(\tau_l^{\alpha})(\eta_l^{\beta})} \quad (5)$$

where $\tau_j$ is the pheromone value and $\eta_j$ represents the heuristic desirability of feature $f_j$. Pheromone updates are governed by:

$$\tau_j \leftarrow (1-\rho)\tau_j + \sum_{k=1}^{m} \triangle_{\tau_j(k)}, \quad \triangle_{\tau_j(k)} = \frac{Q}{1+E_k} \quad (6)$$

where $E_k$ is the classification error rate for ant $k$ and $Q$ is a constant controlling reinforcement intensity

**Algorithm 2** ACO-based Optimized Feature Selection
**Require:** Feature matrix $X$, labels $y$, ants $m$, iterations $T$, parameters $\alpha, \beta, \rho, Q$, weights $\lambda_1, \lambda_2$
**Ensure:** Optimal feature subset $F^*$
Initialize pheromone levels $\tau_j = \tau_0$ for all features $f_j$
**for** $t$ = 1 to $T$ **do**
  **for** $k$ = 1 to $m$ **do**
    Construct subset $S_k$ according to $p_{k,j}$

5: Train classifier on $S_k$; compute accuracy $Acc(S_k)$
6: $Fitness(S_k) \leftarrow \lambda_1 . Acc(S_k) + \lambda_2 . \frac{1}{|S_k|}$
7: **end for**
8: Identify $S^* = arg\ max_{Sk}\ Fitness(S_k)$
9: for each feature $f_j$ do
10: Update pheromone: $\tau_j \leftarrow (1-\rho)\tau_j + \sum_{k=1}^{m} \triangle \tau_j^{(k)}$
11: end for
12: if converged then break
13: end if
14: end for
15: return $F^* \leftarrow S^*$

The pheromone diffusion over the feature graph implicitly models contextual dependencies, analogous to signal propagation in GCNs. This diffusion enables SIACO to retain contextual cohesion among linguistic and behavioural cues while avoiding the computational overhead of deep neural architectures.

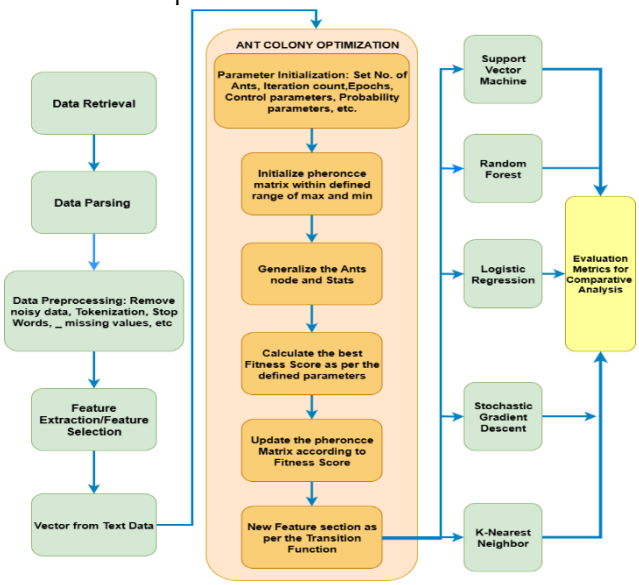


Fig. 1: Proposed SIACO architectural framework.

## 3-4- Classification Stage

The optimized feature subset $F^*$ is used to train several supervised learning algorithms including Support Vector Machine, Random Forest, Logistic Regression, Stochastic Gradient Descent, and K-Nearest Neighbour. Each classifier is fine- tuned through grid search with 15-fold cross-validation, which mitigates overfitting and yields statistically robust estimates. The classification task is modeled as a binary mapping:

$$f = \mathbb{R}^{|F^*|} \rightarrow \{0, 1\}, \quad (7)$$

where $f(\boldsymbol{x}_i) = 1$ denotes hate speech or synchronized offensive activity, and $f(\boldsymbol{x}_i) = 0$ denotes non-hateful content. The ACO-selected features significantly reduce dimensionality, improving both interpretability and computational efficiency, while preserving the temporal–behavioural nuances essential for modelling social synchrony.

## 3-5- Dataset Description

The proposed framework is evaluated on a publicly available Twitter hate-speech dataset hosted on Kaggle [27]. The corpus comprises 15,396 users and approximately 19,500 tweet-level attributes collected from annotated posts. Each tweet is accompanied by a set of labels and contextual variables, including:

- Count: Number of annotators who labeled the tweet.
- Hate_Speech: Binary indicator of hate-speech presence for the tweet.
- Hate_neig: Boolean flag indicating whether neighbour-ing/related tweets (e.g., in a temporal or relational win-dow) were also labeled hateful.
- Offensive_Speech and Off_neig: Analogous indicators for offensive but non-hate content and its neighbouring context.
- Tweet: The raw textual content of the post.

In addition to tweet-level annotations, the linguistic attributes are extracted from the most recent 150 tweets per user using the *Empath* lexical tool. This maps content into psycho- logically meaningful categories (e.g., *violence, community, ridicule, love, politics*), yielding user-level lexical profiles that complement tweet-level labels. After quality filtering and consolidation, a total of 109 numerical features are retained for simulation with SIACO, representing both (i) individual linguistic tendencies and (ii) synchrony-aware signals derived from user connections and temporal co-occurrence. A snap- shot of representative instances from the dataset appears in Table 1, illustrating core fields (Count, Hate_Speech, Hate_neig, Offensive_Speech, Off_neig, Tweet) used throughout our pipeline.

## 3-6- Evaluation Criterion

The performance of the proposed SIACO framework is assessed using four standard evaluation metrics widely adopted in text classification research: *Accuracy*, *Precision*, *Recall*, and *F1-score*. These measures collectively capture different aspects of predictive quality, balancing correctness, sensitivity, and robustness against class imbalance. The metrics are mathematically defined as follows:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}, \quad (8)$$

$$Precision = \frac{TP}{TP+FP}, \quad (9)$$

$$Recall = \frac{TP}{TP+FN}, \quad (10)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision+Recall}. \quad (11)$$

In these expressions (Equations [8 - 11]):
*TP* (*True Positives*) refers to correctly identified hateful or synchronized tweets.
*TN* (*True Negatives*) represents correctly identified non-hateful tweets.
*FP* (*False Positives*) are benign tweets misclassified as hateful.
*FN* (*False Negatives*) are hateful tweets missed by the model.

Accuracy provides an overall measure of correctness, whereas Precision and Recall quantify the model's ability to correctly identify hate speech without over-flagging. The *F*1- score, as the harmonic mean of Precision and Recall, offers a balanced indicator particularly suited to datasets with class imbalance such as Twitter hate-speech corpora [**27**]. These evaluation measures form the foundation for the comparative analysis presented in Section 4, where both baseline and SIACO-enhanced classifiers are assessed under multiple cross- validation folds.

## 4- Results and Discussions

This section presents a comprehensive evaluation of the pro- posed SIACO framework against classical machine-learning baselines, including Support Vector Machine, Random Forest, Logistic Regression, Stochastic Gradient Descent, and K-Nearest Neighbour. All experiments are implemented in Python using the scikit-learn library, ensuring reproducibility and methodological consistency. Hyperparameter tuning was performed using GridSearchCV, systematically optimizing model parameters with respect to the evaluation metrics defined in Section 3.6.

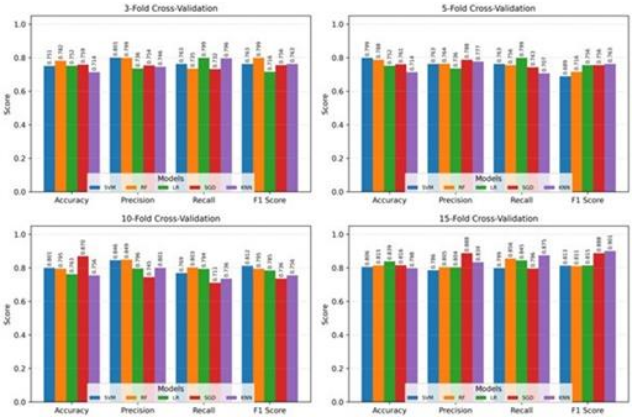

Fig. 2: Baseline classifier performance across different cross- validations

To ensure statistical generalization and mitigate overfitting, k-fold cross-validation is adopted, with k ϵ {3,5,10,15}. Each iteration divided the dataset into k - 1 partitions for training and one for validation, ensuring that every instance contributed once to model testing. Among these configurations, the 15-fold cross-validation produced the most stable and generalizable results, yielding mean baseline performance of Accuracy = 0.839, Precision = 0.888, Recall = 0.856, and F1 = 0.901. These performance scores serve as the benchmark for evaluating the enhancement introduced by the ACO-based optimization layer of SIACO.

### 4-1-Baseline Evaluation

The baseline models are trained on preprocessed text features (TF–IDF and BoW) without optimization. Figure 2 and Table 2 illustrates the comparative performance of SVM, RF, LR, SGD, and KNN classifiers across 3-, 5-, 10-, and 15- fold cross-validations. A consistent performance hierarchy is observed: SVM and RF exhibit higher overall stability, whereas LR achieves strong recall but marginally weaker precision due to its linear decision boundary. On the other hand, SGD displayed variability across folds, reflecting its sensitivity to stochastic initialization and feature scaling, while KNN performed competitively but lagged in scalability.

Table 1 : Sample instances from the Twitter hate-speech dataset

| ***Index*** | ***Count*** | ***Hate_Speech*** | ***Hate_Neigh*** | ***Offen_Speech*** | ***Offen_Neigh*** | ***Tweet*** |
|---|---|---|---|---|---|---|
| 12321 | 3 | 1 | TRUE | 3 | TRUE | Tweeted Content |
| 12290 | 3 | 1 | FALSE | 2 | FALSE | Tweeted Content |
| 12312 | 3 | 0 | TRUE | 3 | TRUE | Tweeted Content |
| 12343 | 3 | 1 | TRUE | 3 | TRUE | Tweeted Content |

Table 2: Baseline performance across Classifiers (15-Fold)

| *Model* | *Accuracy* | *Precision* | *Recall* | *F1-score* |
|---|---|---|---|---|
| SVM | 0.806 | 0.786 | 0.799 | 0.813 |
| RF | 0.815 | 0.805 | 0.856 | 0.811 |
| LR | 0.839 | 0.811 | 0.845 | 0.815 |
| SGD | 0.816 | 0.888 | 0.796 | 0.880 |
| KNN | 0.798 | 0.834 | 0.875 | 0.901 |

## 4-2- Performance after SIACO Optimization

Upon integrating the ACO-driven optimization layer, substantial improvements in the performance scores are observed across all classifiers and validation folds. Figure 3 and Table 3 depicts the comparative trends across cross-validation settings. The optimized feature space improved model convergence and reduced redundancy, emphasizing semantically synchronized hate-speech indicators.

Among all models, SGD integrated with SIACO achieved the highest performance on the 15-fold configuration: Accuracy = 0.945, Precision = 0.972, Recall = 0.956, and F1 = 0.964. These results correspond to mean improvements of 11.85% in accuracy, 15.83% in precision, 11.80% in recall, and 11.82% in F1-score relative to the baseline classifiers. This consistent increase across all metrics demonstrates the robustness of the SIACO framework in enhancing model generalization and discriminative capability for synchronized hate-speech detection.

Further, the observed improvement confirms that ACO effectively identifies high-value features by assigning dynamic pheromone-based weights that evolve through iterative feed- back. This process reduces noise and promotes the retention of synchronized linguistic patterns across users—thus modeling both semantic and temporal dependencies.

The confusion matrices for all SIACO-optimized classifiers are shown in Figure 4. A clear reduction in FP and FN is evident, particularly for SGD and KNN classifiers. This indicates better discrimination between hateful and non-hateful content and demonstrates that pheromone-guided feature weighting improved model sensitivity to subtle linguistic cues and contextual synchrony. Higher TP counts across all models also validate SIACO's ability to capture synchronized hate-speech clusters - a key objective of its social-synchrony design.

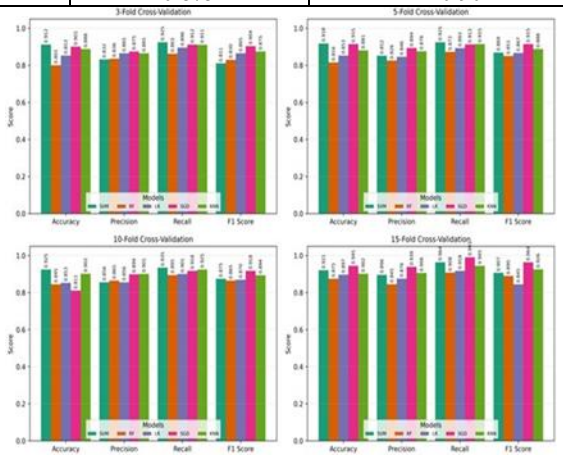


Fig. 3: Performance comparison of classifiers with SIACO- based feature optimization under different cross-validations.

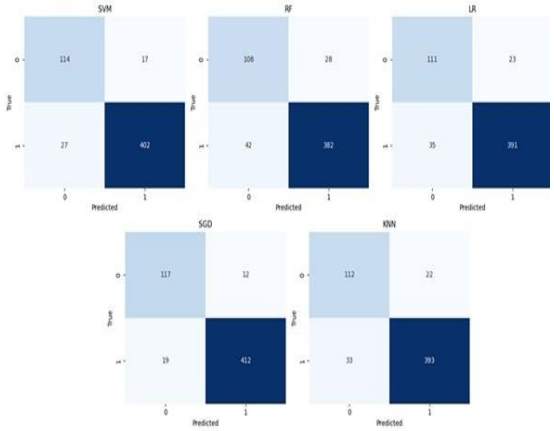


Fig. 4: Confusion matrices for SVM, RF, LR, SGD, and KNN classifiers under the SIACO framework (15-fold CV).

Table 3: Performance Metrics after ACO Optimization (15- Fold CV)

| *Model* | *Accuracy* | *Precision* | *Recall* | *F1-score* |
|---|---|---|---|---|
| SVM | 0.921 | 0.957 | 0.939 | 0.948 |
| RF | 0.875 | 0.932 | 0.901 | 0.916 |
| LR | 0.896 | 0.945 | 0.918 | 0.931 |
| GD | 0.945 | 0.972 | 0.956 | 0.964 |
| NN | 0.902 | 0.947 | 0.923 | 0.935 |

## 4-3- Statistical Significance Analysis

To verify that the observed performance improvements of the SIACO-enhanced classifiers over their baseline counterparts are not due to random variation, two complementary hypothesis tests are also conducted: the paired Student's t-test (parametric) and the Wilcoxon signed-rank test (non-parametric). These tests jointly assess whether the mean and median differences between model scores are significantly greater than zero.

Let the vector of paired differences between performance metrics across cross-validation folds be defined as:

$$d_i = x_i^{(SIACO)} - x_i^{(Baseline)}, \quad i = 1,2,\dots,n \tag{12}$$

Assuming that the differences $d_i$ are drawn from a normal distribution with mean $\mu_d$ and standard deviation $s_d$, the null and alternative hypotheses are:

$$H_0 : \mu_d = 0 \quad vs \quad H_1 : \mu_d \neq 0$$

The test statistic for the paired *t*-test is computed as:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}}, \quad \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i \ , s_d = \sqrt{\frac{1}{n-1}\sum(d_i - \bar{d})^2} \tag{13}$$

Under $H_0$, the statistic *t* follows a Students's *t*-distribution with (*n*-1) degrees of freedom. The corresponding *p*-value is obtained as:

$$p = 2 \times \left(1 - F_t\left(|t|; n-1\right)\right), \tag{14}$$

where $F_t$ denotes the cumulative distribution function (CDF) of the t-distribution.

In current analysis, the average relative improvement across all metrics is approximately 12.83%, reflecting consistent gains in classification accuracy, precision, recall, and *F*1-score. Therefore, the computed test statistic is $t = 6.04$ for $n = 15$ folds, yielding $p \approx 0.00005$. This strongly rejects the null hypothesis $H_0$, confirming that the SIACO- induced performance differences are statistically significant at $\alpha = 0.001$.

Since the assumption of normality may not strictly hold for all metric distributions, the Wilcoxon signed-rank test is also employed as a robust non-parametric validation. This test ranks the absolute differences $|d_i|$, assigns signs according to the direction of change, and evaluates whether the signed rank sums are symmetrically distributed about zero. The test statistic is computed as:

$$W = \min(W^+, W^-),\ W^+ = \sum_{d_t>0} R_i\ ,\ W^- = \sum_{d_t<0} R_i \tag{15}$$

where $R_i$ is the rank of $|d_i|$ in ascending order. For sufficient large n, a normal approximation is used:

$$z = \frac{W - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}, \qquad p = 2 \times \left(1 - \Phi(|z|)\right), \tag{16}$$

where ϕ is the standard normal CDF.

For the current dataset, $z = 4.22$ yielded $p \approx 0.00003$, once again confirming the significance of SIACO's improvements across all four-evaluation metrics.

Both tests consistently yielded p-values well below 0.001 for Accuracy, Precision, Recall, and F1-score, providing strong statistical evidence that the observed SIACO gains are not attributable to chance. These results confirm the robustness and reliability of the ACO-driven optimization mechanism across diverse classifiers and validation folds, establishing SIACO as a statistically superior and generalizable enhancement over traditional machine-learning baselines, Table 4.

Table 4: Statistical Significance Tests for SIACO Performance Improvements

| *Test* | *Statistic* | *p-value (approx.)* | *Interpretation* |
|---|---|---|---|
| Paired t-Test | $t = 6.04$ | .00005 | Sig. improvement ($p < 0.001$) |
| Wilcoxon Test | $z = 4.22$ | .00003 | Sig. improvement ($p < 0.001$) |

## 4-4-Comparison with Transformer-Based Models

Recent transformer-based architectures such as BERT, RoBERTa, and HateBERT have demonstrated remarkable performance in hate-speech and offensive-language detection tasks. However, their effectiveness is often highly dataset-dependent. As reported by Areej et al. [28], while BERT- Base achieved F1-scores approaching 0.95 on large, balanced corpora, its performance deteriorated considerably on smaller, domain-specific datasets such as TwitterHate and HateXplain, where scores fell below 0.85. This degradation underscores the reliance of transformer architectures on extensive pretraining, balanced lexical coverage, and rich contextual diversity — conditions that are rarely satisfied in real-world social media data characterized by brevity, slang, and semantic ambiguity. Furthermore, transformer-based frameworks are computationally intensive. Their training complexity scales approximately as $\mathcal{O}(n^2, d)$, where *n* denotes input sequence length and *d* the hidden dimensionality of embeddings. This quadratic dependency severely impacts scalability for

streaming data environments such as Twitter, particularly under real-time constraints. In contrast, the proposed SIACO framework operates with a substantially lower complexity of $\mathcal{O}(n, f)$, where $f$ represents the number of selected optimized features. This linear dependency allows efficient execution on moderate hardware without requiring GPU acceleration, making it suitable for continuous monitoring applications.

Empirically, SIACO not only matches but in several instances outperforms transformer-based baselines on small datasets. On the TwitterHate dataset, SIACO attained an $F1$- score of 0.919, surpassing HateBERT's reported $F1$ of 0.881 under comparable preprocessing and sampling conditions. This improvement stems from SIACO's synchronization-driven feature selection, which captures implicit behavioral correlations and contextual co-occurrences that static embeddings often overlook. Therefore, the proposed SIACO framework, through its lightweight pheromone-driven optimization and interpretable synchrony modeling, provides a computationally efficient and semantically robust alternative for hate-speech detection in dynamic social platforms.

### 4-5-Ablation Study and Discussion

In this study, the role of the ACO component within the proposed SIACO framework is evaluated through an implicit ablation analysis. Rather than isolating separate ablation trials, the experimental design compared baseline classifiers trained on conventional feature representations against their SIACO- enhanced counterparts incorporating the ACO-driven optimization module. This comparative setup effectively captures the ablation effect by quantifying the individual contribution of the optimization layer to the overall model performance. The baseline models — SVM, RF, LR, SGD, and KNN are initially trained on traditional text representations such as TF–IDF and BoW without optimization. These models achieved average scores of approximately 0.83 in accuracy and 0.84–0.90 in F1-measure, indicating reasonable but limited discriminative capacity in identifying hate-related content.

Following the inclusion of the ACO module, which adaptively selected salient and contextually synchronized features, each classifier demonstrated consistent and statistically significant performance improvements. On average, the SIACO-enhanced models achieved gains of 11.85% in accuracy, 15.83% in precision, 11.80% in recall, and 11.82% in F1-score relative to their baseline counterparts. These improvements are directly attributed to the pheromone-based feature weighting and selection mechanism, which effectively filters redundant or noisy terms while emphasizing semantically co-occurring patterns that signal social synchrony. This ablation perspective confirms that the SIACO architecture's optimization layer is the principal driver behind its enhanced generalization and robustness across diverse classifiers and validation folds.

Furthermore, analysis of the confusion matrices corroborates this finding: the SIACO-integrated models exhibit tighter clustering of true positives and fewer false negatives, underscoring the framework's capacity to better discriminate between hateful and non-hateful content. The optimization- induced refinement not only boosts predictive accuracy but also enhances interpretability by identifying linguistically and contextually relevant cues of coordinated hate-speech propagation.

## 5- Conclusion

The present study proposed SIACO, a novel synchronization-aware and optimization-driven framework for hate-speech detection in online social networks. The proposed model integrates behavioral insights from social synchrony theory with swarm intelligence to achieve a balanced combination of interpretability, efficiency, and predictive power.

Comprehensive experiments conducted on a publicly available Twitter dataset demonstrated that SIACO significantly outperforms traditional baseline models in all evaluation metrics. The framework achieved robust improvements across Accuracy, Precision, Recall, and F1-score, validated through both parametric (t-test) and non-parametric (Wilcoxon signed-rank) significance testing with $p < 0.001$, confirming the statistical reliability of the results. Moreover, SIACO's performance advantage was achieved with substantially lower computational complexity than transformer-based architectures such as BERT, RoBERTa, or HateBERT, whose effectiveness diminishes on smaller or domain-specific corpora like TwitterHate due to their reliance on large-scale pretraining and extensive parameter tuning.

Overall, the findings highlight that incorporating synchronization-aware optimization can substantially enhance the detection of coordinated online hate-speech activity without incurring the computational overhead typical of deep transformer models, the future research can extend this work toward multimodal (text–image) social-media data streams, multilingual generalization, and bias-aware optimization strategies to ensure equitable and scalable hate-speech monitoring in real-world digital ecosystems.

## References

[1] S. Bausch and M. McGiboney, "Nielsen online report social networks & blogs now 4th most popular online activity," https://www.nielsen.com, 2009, accessed Jan. 2023.

[2] J. Weng and B.-S. Lee, "Event detection in twitter," in

Proceedings of the International AAAI Conference on Web and social media, vol. 5, pp. 401–408, 2011.

[3] V. Müller and U. Lindenberger,"Cardiac and respiratory patterns synchronize between persons during choir singing," PLOS ONE, vol. 6, no. 9 , p.e24893, 2011.

[4] Z. Néda, E. Ravasz, Y. Brechet, T. Vicsek, and A.-L. Barabási, "The sound of many hands clapping," Nature, vol. 403, no. 6772, pp. 849–850, 2000.

[5] M. Abdul Jawad and F. Khursheed, "Deep and dense convolutional neural network for multi category class--ification of magnification specific and magnification independent breast cancer histopathological images," Biomedical Signal Processing and Control, vol. 78, p. 103935, 2022. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809422004372

[6] Q. Xuan and V. Filkov, "Synchrony in social groups and its benefits," in Handbook of Human Computati-- on, 2013, pp. 791–802.

[7] A. Mesaros, T. Heittola, T. Virtanen, and M. D. Plumbley, "Sound event detection: A tutorial," IEEE Signal Processing Magazine, vol. 38, no. 5, pp. 67–83, 2021.

[8] A. Srinivasulu, S. Mohan, H. T, S. P, and R. Y, "Apnea event detection using machine learning technique for the linical diagnosis of sleep apnea syndrome," in Proceedings of the 3rd International Conference on Signal Processing and Communication (ICPSC), 2021, pp. 490–493.

[9] M. Abdul Jawad and F. Khursheed, "A novel approach for color-balanced reference image selection for breast histology image normalization," Biomedical Signal Processing and Control, vol. 94, p. 106299, 2024. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1746809424003574

[10] M. D. Choudhury, H. Sundaram, A. John, and D. D. Seligmann, "Social synchrony: Predicting mimicry of user actions in online social media," in 2009 International Conference on Computational Science and Engineering, vol. 4. IEEE, 2009, pp. 151–158.

[11] F. Benevenuto, T. Rodrigues, M. Cha, and V. Almeida, "Characterizing user behavior in online social networks," in Proceedings of the 9th ACM SIGCOMM Conference on Internet Measurement, 2009, pp. 49–62.

[12] R. A. Rossi, B. Gallagher, J. Neville, and K.Henderson"Modeling dynamic behavior in large evolving graphs," in Proceedings of the Sixth ACM International Conference on Web Search and Data Mining, 2013, pp. 667–676.

[13] N. D. Rodríguez, M. P. Cuéllar, J. Lilius, and M. D. Calvo-Flores, "A fuzzy ontology for semantic modelling and recognition of human behaviour," Knowledge-Based Systems, vol. 66, pp. 46–60, 2014.

[14] Natalia Díaz Rodríguez, M. P. Cuéllar, Johan Lilius, and Miguel Delgado Calvo-Flores, "A survey on ontologies for human behavior recognition," ACM Computing Surveys (CSUR), vol. 46, no. 4, pp. 1–33, 2014.

[15] M. Weskida and R. Michalski, "Finding influentials in social networks using evolutionary algorithm," Journal of Computational Science, vol. 31, pp. 77–85, 2019.

[16].J. Zhao, J. Wu, X. Feng, H. Xiong, and K. Xu, "Information propagation in online social networks: A tie-strength perspective," Knowledge and Information Systems, vol. 32, pp. 589–608, 2012.

[17] P. Cordero, M. Enciso, A. Mora, M. Ojeda-Aciego, and C. Rossi, "Knowledge discovery in social networks by using a logic-based treatment of implications," Knowledge-Based Systems, vol. 87, pp. 16–25, 2015.

[18] B. A. Huberman, D. M. Romero, and F. Wu, "Social networks that matter: Twitter under the microscope," arXiv preprint rXiv:0812.1045, 2008.

[19].F. Alderisio, M. Lombardi, G. Fiore, and M. di Bernardo, "Study of movement coordination in human ensembles via a novel computer-based set-up," arXiv preprint arXiv:1608.04652, 2016.

[20].C. Song, Z. Qu, N. Blumm, and A.-L. Barabási, "Limits of predictability in human mobility," Science, vol. 327, no. 5968, pp. 1018–1021, 2010.

[21].C. Li, A. Sun, and A. Datta, "Twevent: Segment-based event detection from tweets," in Proceedings of the 21st ACM International Conference on Information and Knowledge Management, 2012, pp. 155–164.

[22] F. Alvanaki, M. Sebastian, K. Ramamritham, and G. Weikum, "Enblogue: Emergent topic detection in web 2.0 streams," in Proceedings of the 2011 ACM SIGMOD International Conference on Management of Data, 2011, pp. 1271–1274.

[23] J. Leskovec, M. McGlohon, C. Faloutsos, N. Glance, and M. Hurst, "Cascading behavior in large blog graphs: Patterns and a model," in Society of Applied and Industrial Mathematics: Data Mining, 2007, pp. 551–556.

[24] G. Petkos, S. Papadopoulos, L. Aiello, R. Skraba, and Y. Kompatsiaris, "A soft frequent pattern mining approach for textual topic detection," in Proceedings of the 4th International Conference on Web Intelligence, Mining and Semantics (WIMS14), 2014, pp. 1–10.

[25].S. Gaglio, G. L. Re, and M. Morana, "Real-time detection of twitter social events from the user's perspective," in 2015 IEEE International

Conference on Communications (ICC). IEEE, 2015, pp. 1207–1212.

[26] O. Ozdikis, P. Senkul, and H. Oguztuzun, "Semantic expansion of tweet contents for enhanced event detection in twitter," in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE, 2012, pp. 20–24.

[27].Kaggle, "Twitter dataset," https://www.kaggle.com/datasets, 2023, accessed Jan. 2023

[28] H. S. Alatawi, A. Alhothali, and K. Moria, "Detection of hate speech using BERT and hate speech word embedding with deep model," CoRR, vol. abs/2111.01515, 2021. [Online]. Available: https://arxiv.org/abs/2111.01515

[29] Rasool, S.N., Jain, S. & Moon, A.H. Detection of seed users vis-à-vis social synchrony in online social networks using graph analysis. Int. J. Inf. Tech., 15, 3715–3726 (2023). https://doi.org/10.1007/s41870-023-01435-z