



Applied Linguistics Driven Deceptive Content Recognition using Single Valued Trapezoidal Neutrosophic Number with Natural Language Processing

Abdulkhaleq Q. A. Hassan^{*1}

¹Department of English, College of Science and Arts at Mahayil, King Khalid University, Saudi Arabia
Emails: abqaseem@kku.edu.sa

Abstract

Single valued neutrosophic number is a special case of single valued neutrosophic set and are of importance for neutrosophic multi-attribute decision making problem. A single valued neutrosophic number seems to define an ill-known quantity as a generalization of intuitionistic number. Applied linguistics in the context of Natural Language Processing (NLP) comprises the practical applications of linguistic approaches for addressing real time language processing issues. Social media become indispensable components in many people's lives and have been growing rapidly. In the meantime, social networking media have become a widespread source of identity deception. Several social media identity deception cases have appeared presently. The research was performed to detect and prevent deception. Identifying deceptive content in natural language is significant to combat misrepresentation. Leveraging forward-thinking NLP methods, our model contextual cues analyze linguistic patterns, and semantic inconsistencies to flag possibly deceptive contents. By assimilating complex procedures for parameter optimization, feature extraction, and classification, the NLP focused on precisely recognizing deceptive content through different digital platforms, which contributes to the preservation of data integrity and the promotion of digital literacy. This study presents a Single Valued Trapezoidal Neutrosophic Number with Natural Language Processing for Deceptive Content Recognition (STVNNLP-DCR) technique on Social Media. The presented technique includes four important elements: preprocessing, GloVe word embedding, STVN classification, and Chicken Swarm Optimization (CSO) for parameter tuning. The preprocessing stage includes tokenization and text normalization, preparing text information for succeeding analysis. Then, GloVe word embedding represents the word in a continuous vector space, which captures contextual relationships and semantic similarities. The STVN classifier deploys the embedding to discern deceptive patterns within the text, leveraging its capability to effectively manage high-dimensional and sparse datasets. Moreover, the CSO technique enhances the hyperparameter of the STVN classifier, improving its generalization capabilities and performance. Empirical analysis implemented on varied datasets validates the efficacy of the presented technique in precisely recognizing deceptive content. Comparative studies with advanced approaches demonstrate high efficiency. The presented technique shows robustness against different forms of deceptive content, such as clickbait, misinformation, and propaganda

Keywords: Deceptive Content Recognition; Natural Language Processing; Applied Linguistics; Social Media; Word Embedding; Neutrosophic Set

1. Introduction

Generally, a fuzzy set cannot manage situations of inexact data and inconsistencies, however, the Neutrosophic Set (NS) have been utilized to overcome these kinds of problems in real time. The neutrosophic set is an extension of the fuzzy set and the intuitionistic fuzzy set that could handle indeterminate, imperfect, and inconsistent data in the related problem. Interval-valued intuitionistic fuzzy set by expanding the membership and non-membership functions to the interval number. This set could only manage incomplete data, not the indeterminate and inconsistent data. For these reasons, a new theory named neutrosophic set is introduced by adding an independent indeterminacy-membership based on the intuitionistic fuzzy set from the philosophical perspective, which is a

generalization of the concept of probability sets, classical sets, fuzzy sets, rough sets, paraconsistent sets, intuitionistic fuzzy sets, paradoxist sets, tautological sets and dialetheist sets. In concept of falsity-membership, neutrosophic sets, truth-membership and indeterminacy-membership are independently represented.

Social media platforms' huge development and popularity have transformed data access and propagation [1]. These platforms are extremely appealing and very simple to access. So, they attract many consumers and permit rapid broadcasting of content globally. Social media consumers share numerous events and huge breaking news over these platforms [2]. They also present greater visibility of misinformation. Few reports indicate that 65% of data is false news on these platforms. Moreover, misinformation is distributed quicker, wider, and deeper on these platforms [3]. Misinformation has become a serious problem that harmfully affects individuals and society. Indeed, it is very easy to spread misleading data that can influence objective entities. It is attained for dissimilar improvements such as economic and political gains by an organization or individual [4].

Text classification is the procedure of categorizing and arranging tags or texts dependent on the content of the data. Intent classification is one of the vital tasks in natural language processing (NLP), and it has an extensive range of uses that contain spam classification, subject labeling, and sentiment analysis (SA) [5]. NLP permits text analyzers to mechanically discover content, after a pre-defined set of labels or identifications are allocated dependent upon their subjects from medicinal research publications, documents, and other resources from all over the world are recognized. While the classifier chooses which class of textual content is categorized, it is vital to evaluate the grade to which every input in the training dataset is parallel. To mechanically discover and expose patterns in electric texts, NLP, machine learning (ML) models, and data mining (DM) are used. The vital purpose of the technology is to allow consumers to remove data from textual tools and contract with actions using text mining. Information extraction (IE) technologies have been found to remove exact data from textual materials [6]. This is the primary technique, which specifies that the expressions data extraction and text mining may be utilized interchangeably. The major developments in global Internet usage and mobile phones have reformed social connections [7]. Due to their simplicity of access and rapid distribution of news, Twitter has become a more famous methods for persons to distribute and get news. But, the consistency of the news shared on social media platforms has turned into a big anxiety [8]. Due to the increase of social media platforms, real and fake news has been frequently accessible in similar methods, making it incredible to distinguish between the dual. In the lack of difficult verification, well-meant persons may innocently support the spread of fake news [9]. For political, economic, or other reasons, fake news is material that was formed to mislead, deceive, or lure readers. The usage of social media has skyrocketed in present days, due to the benefits of linking people, sharing material, and keeping up-to-date with worldwide happenings [10]. The risk of false data and the spread of fake news have enlarged, which probably produces main social concerns.

This study presents a Single Valued Trapezoidal Neutrosophic Number with Natural Language Processing for Deceptive Content Recognition (STVNNLP-DCR) model on Social Media. The preprocessing stage includes tokenization and text normalization, preparing text information for succeeding analysis. Then, GloVe word embedding represents the word in a continuous vector space, which captures contextual relationships and semantic similarities. The STVN classifier deploys the embedding to discern deceptive patterns within the text, leveraging its capability to effectively manage high-dimensional and sparse datasets. Moreover, the CSO technique enhances the hyperparameter of the STVN classifier, improving its generalization capabilities and performance. Empirical analysis implemented on varied datasets validates the efficacy of the presented technique in precisely recognizing deceptive content.

2. Literature Review

Pospelova et al. [11] deliver a new sine cosine algorithm with DL-based deceptive content detection on social media (SCADL-DCDSM) model. Mainly, the SCADL-DCDSM method pre-processing the input data to alter an input data into a valued layout. Furthermore, the SCADL-DCDSM system tracks the BERT technique for the word embedding procedure. Furthermore, the SCADL-DCDSM system includes an ensemble of 3 approaches for the classification of sentiment like ELM, LSTM, and attention-based RNN. Lastly, SCA is implemented for superior hyperparameter selection of DL methods. The SCADL-DCDSM model mixes the explainable AI (XAI) method LIME was used. Albraikan et al. [12] proposed the Bio-inspired AI with NLP Deceptive Content Detection (BAINLP-DCD) system. In [13], a new fake news recognition system based on the Natural Language Inference (NLI) technique has been projected. The developed model exploits a human-like method, which is dependent upon concluding accuracy utilizing a set of consistent news. In this model, the associated and parallel news distributed in reliable news bases was employed as supplementary knowledge.

Schmidt et al. [14] concentrate on discovering the accuracy of noticing misinformation, which contains the identification of video data into dual types namely genuine misinformation and information. More specially, this study work produces extra metadata embedded within online videos. Employing NLP, the technique removes the

terms of medical subject headings (MeSH) from video records and categorizes videos utilizing 4 ML methods. In [15], an original AI-aided fake news recognition with a deep NLP method has been projected.

Jadhav and Shukla [16] employed a sophisticated technique that combines advanced DL models, analysis of verbal complexity, and innovative NLP models. The paper discovers the assortment and full study of 4 cutting-edge DL structures like LSTM-Attention Mechanism, GPT, and BERT. At the same time, linguistic complexity metrics provide a complete analysis of the text, and the NLP systems aid in discovering semantic forms within the manuscript. In [17], a psycholinguistic features-based and emotion-enhanced technique is presented. It involves identifying rumors using lexicons and numerous verbal features-based learning techniques, mainly by removing the psychological link of words with their emotions. Psycholinguistic and emotional features are removed from both comments and posts to enhance the model and make rumor recognition more effectual.

3. System Model

In this paper, we have presented an STVNNLP-DCR technique on social media. The presented technique includes four important elements: preprocessing, GloVe word embedding, STVN classification, and CSO for parameter tuning.

A. Data Cleaning

Initially, the preprocessing stage includes tokenization and text normalization, preparing text information for succeeding analysis. Generally, the input information is pre-processed in various forms to increase the data quality. The results from the data-mining procedure are based on the various pre-processing. This lacks raw data and adapts the unreliable into typical computational information.

To achieve this, the NLP method includes stemming, tokenization, stop word elimination; character alteration to lowercase letters, and other procedures in the Keras reference library are exploited. Words including 'there,' 'the,' 'of,' etc., are known as stop words and are popular words in daily discussions. Intrinsicly, a stop word might take place in the text repeatedly; however, it probably has a restricted influence on the entire context of phrases.

B. Word Embedding's Generation

Next, GloVe word embedding represents the word in a continuous vector space, which captures contextual relationships and semantic similarities. The objective of GloVe is to analyze the global representation of the whole corpus and integrate the terms' significance into these investigations [18]. While determining the values of a real-valued vector connected to a single word, *co*-occurrence, and word frequency are the most important metrics that are considered. The basis for the calculation is the frequency usage of specific words and the frequency of the word directly nearby to all the words. Then, a *X* *co*-occurrence matrix is built by the corpus and all the phrases. The Pnm specifies the possibility that the *m* word appeared in the identical scenery as the word *b* sometimes.

$$G_{ab} = X_{ab}/X_b \quad (1)$$

Using *b*, *a*, and 3rd words from the *k* context, we compute the possibility of *co*-occurrence of *T*(*b*, *a*, and *k*) as follows:

$$T_{(b,a,k)} = \frac{G_{ba}}{G_{ak}} \quad (2)$$

$$\text{In } M = \sum_{ba=1}^p f(X_{ba})(W_b^T k + B_n + B_m - \log X_{ba})^2 \quad (3)$$

The training objective is to accomplish the great probable decline in the error generated by the least squares algorithm, *f* allocates weights. GloVe assigns a real-valued vector to all the words after finishing the training stage.

C. Development of STVN-based Classification Model

The STVN classifier deploys the embedding to discern deceptive patterns within the text, leveraging its capability to effectively manage high-dimensional and sparse datasets. Some essential results and definitions are specified to distinguish the major outcomes [19].

Definition 1

The fuzzy number *P* is proposed by a couple of bounded functions $P^L(\alpha)$, $P^R(\alpha)$, $\alpha \in [0,1]$ where P^R is monotone decreasing, right continuous and P^L is monotone increasing, left continuous with $P^L(\alpha) \leq P^R(\alpha)$.

A amount of trapezoidal fuzzy is shown as $P = (m_0, n_0, \gamma, \delta)$ where $[m_0, n_0]$ refers to the interval defuzzifier and $\gamma (> 0), \delta (> 0)$ are left and right fuzziness correspondingly, and $(m_0 - \gamma, n_0 + \delta)$ indicates the support of P . The membership function was described by the following expression:

$$P(x) = \begin{cases} \frac{1}{\gamma}(x - m_0 + \gamma), & m_0 - \gamma \leq x \leq m_0, \\ 1, & x \in [m_0, n_0], \\ \frac{1}{\delta}(n_0 - x + \delta), & n_0 \leq x \leq n_0 + \delta, \\ 0, & \text{elsewhere.} \end{cases}$$

In parametric form $P^L(\alpha) = m_0 - \gamma + \gamma\alpha, P^R(\alpha) = n_0 + \delta - \delta\alpha$.

Definition 2

The NS P through the universe U is described using the triplet such as falsity-membership value η_P , truth-membership value μ_P and indeterminacy-membership value ν_P where $\mu_P, \nu_P, \eta_P: U \rightarrow [0, 1]^+$. P is demonstrated by: $P = \{ \langle x, \mu_P(x), \nu_P(x), \eta_P(x) \rangle : x \in U \}$ with $-0 \leq \sup \mu_P(x) + \sup \nu_P(x) + \sup \eta_P(x) \leq 3^+$. Here $1^+ = 1 + \varepsilon$, where 1 and ε refers to the normal and unusual parts. Likewise $\bar{0} = 0 - \varepsilon$, where 0 and ε denote the normal and unusual parts.

However, it is problematic to apply NS with values from the real standard or non-standard subsets of $[0 \text{ and } 1]$ in real-time. To overwhelm this, NS with values from the subset of $[0,1]$ is taken into account.

Definition 3

The single-valued neutrosophic (SVN) set M over the universe U is NS where the component of all the triplets are real standard components of $[0,1]$. Therefore, SVN-set M is implemented by: $M = \{ \langle x, \mu_M(x), \nu_M(x), \eta_M(x) \rangle : x \in U \text{ and } \mu_M(x), \nu_M(x), \eta_M(x) \in [0,1] \}$ so that $0 \leq \sup \mu_M(x) + \sup \nu_M(x) + \sup \eta_M(x) \leq 3$.

Definition 4

Consider $p_i, q_i, s_i, t_i \in R$ with $p_i \leq q_i \leq s_i \leq t_i (i = 1,2,3)$ and $w_{\tilde{m}}, u_{\tilde{m}}, y_{\tilde{m}} \in [0,1]$. Then an SVN-number $\tilde{m} = \langle ([p_1, q_1, s_1, t_1]; w_{\tilde{m}}), ([p_2, q_2, s_2, t_2]; u_{\tilde{m}}), ([p_3, q_3, s_3, t_3]; y_{\tilde{m}}) \rangle$ refers to an SVN set on R whose truth, indeterminacy, and falsity values are correspondingly represented by the mapping $u_{\tilde{m}}: R \rightarrow [0, w_{\tilde{m}}], \nu_{\tilde{m}}: R \rightarrow [u_{\tilde{m}} - 1, 1], \eta_{\tilde{m}}: R \rightarrow [y_{\tilde{m}}, 1]$ as follows:

$$\begin{cases} g_{\mu}^l(x), & p_1 \leq x \leq q_1, & g_{\nu}^l(x), & p_2 \leq x \leq q_2, & g_{\eta}^l(x), & p_3 \leq x \leq q_3, \\ w_{\tilde{m}}, & q_1 \leq x \leq s_1, & u_{\tilde{m}}, & q_2 \leq x \leq s_2, & y_{\tilde{m}}, & q_3 \leq x \leq s_3, \\ g_{\mu}^r(x), & s_1 \leq x \leq t_1, & g_{\nu}^r(x), & s_2 \leq x \leq t_2, & g_{\eta}^r(x), & s_3 \leq x \leq t_3, \\ 0, & \text{otherwise} & 1, & \text{otherwise} & 1, & \text{otherwise} \end{cases}$$

The function $g_{\mu}^l: [p_1, q_1] \rightarrow [0, w_{\tilde{m}}], g_{\nu}^r: [s_2, t_2] \rightarrow [u_{\tilde{m}}, 1], g_{\eta}^r: [s_3, t_3] \rightarrow [y_{\tilde{m}}, 1]$ are non-decreasing and continuous which fulfill the: $g_{\mu}^l(p_1) = 0, g_{\mu}^l(q_1) = w_{\tilde{m}}, g_{\nu}^r(s_2) = u_{\tilde{m}}, g_{\nu}^r(t_2) = 1, g_{\eta}^r(s_3) = y_{\tilde{m}}, g_{\eta}^r(t_3) = 1$. The functions $g_{\mu}^r: [s_1, t_1] \rightarrow [0, w_{\tilde{m}}], g_{\nu}^l: [p_2, q_2] \rightarrow [u_{\tilde{m}}, 1], g_{\eta}^l: [p_3, q_3] \rightarrow [y_{\tilde{m}}, 1]$ are non-increasing and continuous which fulfill the: $g_{\mu}^r(s_1) = w_{\tilde{m}}, g_{\mu}^r(t_1) = 0, g_{\nu}^l(p_2) = 1, g_{\nu}^l(q_2) = u_{\tilde{m}}, g_{\eta}^l(p_3) = 1, g_{\eta}^l(q_3) = y_{\tilde{m}}$.

Definition 5

A neutrosophic series of $\tilde{m} = \langle ([p_1, q_1, \delta_1, \xi_1]; w_{\tilde{m}}), ([p_2, q_2, \delta_2, \xi_2]; u_{\tilde{m}}), ([p_3, q_3, \delta_3, \xi_3]; y_{\tilde{m}}) \rangle$ and described on R is known as SVN-number. $\delta_i (> 0)$ and $\xi_i (> 0)$ are the left and right spreads and $[p_i, q_i]$ denotes the interval for point of truth, indeterminacy, and, falsity-memberships for $i = 1,2,3$ correspondingly in \tilde{m} and $w_{\tilde{m}}, u_{\tilde{m}}, y_{\tilde{m}} \in [0,1]$.

$$T_{\tilde{m}}(x) = \begin{cases} \frac{1}{\delta_1} w_{\tilde{m}}(x - p_1 + \delta_1), & p_1 - \delta_1 \leq x \leq p_1, \\ w_{\tilde{m}}, & x \in [p_1, q_1], \\ \frac{1}{\xi_1} w_{\tilde{m}}(q_1 - x + \xi_1), & q_1 \leq x \leq q_1 + \xi_1, \\ 0, & \text{elsewhere.} \end{cases}$$

$$I_{\tilde{m}}(x) = \begin{cases} \frac{1}{\delta_2}(p_2 - x + u_{\tilde{m}}(x - m_2 + \delta_2)), & p_2 - \delta_2 \leq x \leq p_2, \\ u_{\tilde{m}}, & x \in [p_2, q_2], \\ \frac{1}{\xi_2}(x - q_2 + u_{\tilde{m}}(q_2 - x + \xi_2)), & q_2 \leq x \leq q_2 + \xi_2, \\ 1, & \text{elsewhere.} \end{cases}$$

$$F_{\tilde{m}}(x) = \begin{cases} \frac{1}{\delta_3}(p_3 - x + y_{\tilde{m}}(x - p_3 + \delta_3)), & p_3 - \delta_3 \leq x \leq p_3, \\ y_{\tilde{m}}, & x \in [p_3, q_3], \\ \frac{1}{\xi_3}(x - q_3 + y_{\tilde{m}}(q_3 - x + \xi_3)), & q_3 \leq x \leq q_3 + \xi_3, \\ 1, & \text{elsewhere.} \end{cases}$$

Now \tilde{m} has three different pairs $(T_{\tilde{m}}^l, T_{\tilde{m}}^u)$, $(I_{\tilde{m}}^l, I_{\tilde{m}}^u)$, $(F_{\tilde{m}}^l, F_{\tilde{m}}^u)$ of continuous and bounded functions such that

- (i) $T_{\tilde{m}}^u, I_{\tilde{m}}^l, F_{\tilde{m}}^l$ are monotone non-increasing and $T_{\tilde{m}}^l, I_{\tilde{m}}^u, F_{\tilde{m}}^u$ are monotone non-decreasing.
- (ii) $T_{\tilde{m}}^l(r) \leq T_{\tilde{m}}^u(r), I_{\tilde{m}}^l(r) \geq I_{\tilde{m}}^u(r), F_{\tilde{m}}^l(r) \geq F_{\tilde{m}}^u(r), r \in [0,1]$.

The SVN-number \tilde{m} is transmuted into an SVTN number after three intervals in \tilde{m} are all equivalent. Therefore $\tilde{q} = \langle ([m_0, n_0, \delta_1, \xi_1]; w_{\tilde{q}}), ([m_0, n_0, \delta_2, \xi_2]; u_{\tilde{q}}), ([m_0, n_0, \delta_3, \xi_3]; y_{\tilde{q}}) \rangle$ refers to SVTN-number.

Definition 6

The scoring values of SVTN-number can be assessed from geometrical views and its properties are analyzed. Next, the linear ranking function is by the scoring value.

The maximal height of the neutrosophic component of the SVTN-number is taken as 1 to have a linear ranking function. Therefore, the three elements of SVTN-number $\tilde{p} = \langle [p, q, \delta_1, \xi_1], [p, q, \delta_2, \xi_2], [p, q, \delta_3, \xi_3] \rangle$ are set by.

$$\mu_{\tilde{p}}(y) = \begin{cases} \frac{1}{\delta_1}(y - p + \delta_1), & p - \delta_1 \leq y \leq p, \\ 1, & x \in [p, q], \\ \frac{1}{\xi_1}(q - y + \xi_1), & q \leq y \leq q + \xi_1, \\ 0, & \text{elsewhere.} \end{cases}$$

$$v_{\tilde{p}}(y) = \begin{cases} \frac{1}{\delta_2}(p - y), & p - \delta_2 \leq y \leq p, \\ 0, & y \in [p, q], \\ \frac{1}{\xi_2}(y - q), & q \leq y \leq q + \xi_2, \\ 1, & \text{elsewhere.} \end{cases}$$

$$\eta_{\tilde{p}}(y) = \begin{cases} \frac{1}{\delta_3}(p - y), & p - \delta_3 \leq y \leq p, \\ 0, & x \in [p, q], \\ \frac{1}{\xi_3}(y - q), & q \leq y \leq q + \xi_3, \\ 1, & \text{elsewhere.} \end{cases}$$

Assume SVTN-numbers $\tilde{a} = \langle [a, b, \omega_1, \lambda_1], [a, b, \omega_2, \lambda_2], [a, b, \omega_3, \lambda_3] \rangle$ and $\tilde{c} = \langle [c, d, \xi_1, \kappa_1], [c, d, \xi_2, \kappa_2], [c, d, \xi_3, \kappa_3] \rangle$ Then,

(i) Addition:

$$\tilde{a} + \tilde{c} = \langle [a + c, b + d, \omega_1 + \xi_1, \lambda_1 + \kappa_1], [a + c, b + d, \omega_2 + \xi_2, \lambda_2 + \kappa_2], [a + c, b + d, \omega_3 + \xi_3, \lambda_3 + \kappa_3] \rangle$$

(ii) Scalar multiplication: For real number x ,

$$x\tilde{a} = \langle [xa, xb, x\omega_1, x\lambda_1], [xa, xb, x\omega_2, x\lambda_2], [xa, xb, x\omega_3, x\lambda_3] \rangle \text{ if } x > 0.$$

$x\tilde{a} = \langle [xb, xa, -x\lambda_1, -x\omega_1], [xb, xa, -x\lambda_1, -x\omega_2], [xb, xa, -x\lambda_1, -x\omega_3] \rangle$ if $x < 0$.

(iii) If $a = b = \omega_i = \lambda_i = 0$ for i in \tilde{a} , then it is known as a zero SVTN number and is represented as $\tilde{0} = \langle [0,0,0,0], [0,0,0,0], [0,0,0,0] \rangle$.

Product of two SVTN-numbers

Consider $\tilde{a} = \langle [a, b, \omega_1, \lambda_1], [a, b, \omega_2, \lambda_2], [a, b, \omega_3, \lambda_3] \rangle$ and $\tilde{c} = \langle [c, d, \kappa_1, \zeta_1], [c, d, \kappa_2, \zeta_2], [c, d, \kappa_3, \zeta_3] \rangle$ as SVTN-numbers. The product $\tilde{a} \cdot \tilde{c}$ is represented by:

$$\begin{aligned} \tilde{a} \cdot \tilde{c} = \langle & [ac, bd, a\kappa_1 + c\omega_1 - \omega_1\kappa_1, b\zeta_1 + d\lambda_1 + \lambda_1\zeta_1], \\ & [ac, bd, a\kappa_2 + c\omega_2 - \omega_2\kappa_2, b\zeta_2 + d\lambda_2 + \lambda_2\zeta_2], \\ & [ac, bd, a\kappa_3 + c\omega_3 - \omega_3\kappa_3, b\zeta_3 + d\lambda_3 + \lambda_3\zeta_3] \rangle \end{aligned} \quad (4)$$

If $b + \lambda_i > 0$ and $d + \zeta_i > 0, \forall i = 1,2,3$.

$$\begin{aligned} \tilde{a} \cdot \tilde{c} = \langle & [ad, bc, -a\zeta_1 + d\omega_1 + \omega_1\zeta_1, -b\kappa_1 + c\lambda_1 - \lambda_1\kappa_1], \\ & [ad, bc, -a\zeta_2 + d\omega_2 + \omega_2\zeta_2, -b\kappa_2 + c\lambda_2 - \lambda_2\kappa_2], \\ & [ad, bc, -a\zeta_3 + d\omega_3 + \omega_3\zeta_3, -b\kappa_3 + c\lambda_3 - \lambda_3\kappa_3] \rangle \end{aligned} \quad (5)$$

If $b + \lambda_i < 0$, but $d + \zeta_i > 0, \forall i = 1,2,3$.

$$\begin{aligned} \tilde{a} \cdot \tilde{c} = \langle & [bd, ac, -b\zeta_1 - d\lambda_1 - \lambda_1\zeta_1, -a\kappa_1 - c\omega_1 + \omega_1\kappa_1], \\ & [bd, ac, -b\zeta_2 - d\lambda_2 - \lambda_2\zeta_2, -a\kappa_2 - c\omega_2 + \omega_2\kappa_2], \\ & [bd, ac, -b\zeta_3 - d\lambda_3 - \lambda_3\zeta_3, -a\kappa_3 - c\omega_3 + \omega_3\kappa_3] \rangle \end{aligned} \quad (6)$$

If $b + \lambda_i < 0$ and $d + \zeta_i < 0, \forall i = 1,2,3$.

Assume $b + \lambda_3 < 0, d + \zeta_3 > 0$ and $b + \lambda_k > 0, d + \zeta_k > 0$ or $b + \lambda_k < 0, d + \zeta_k < 0$ for $k = 1,2$ then the product is described as the component in neutrosophic triplet are independent. The product of truth and indeterminacy components in $\tilde{a} \cdot \tilde{c}$ follows either the first or third rules while the product of the falsity component in $\tilde{a} \cdot \tilde{c}$ follows the second rule.

D. Hyperparameter Tuning

At last, the CSO model enhances the hyperparameter of STVN classifier, improving its generalization capabilities and performance. Meng et al. developed the CSO Algorithm [20]. Generally, in a group of chickens, there are numerous types of hens, roosters, and chicks, and every chicken has its equivalent identity as it hunts and acquires from its parents. The CSO algorithm is generally intended to detect the classified order, behaviors of learning, and foraging of chicken is essential for model proposal and position upgrade plan. Fig. 1 depicts the steps involved in CSO.

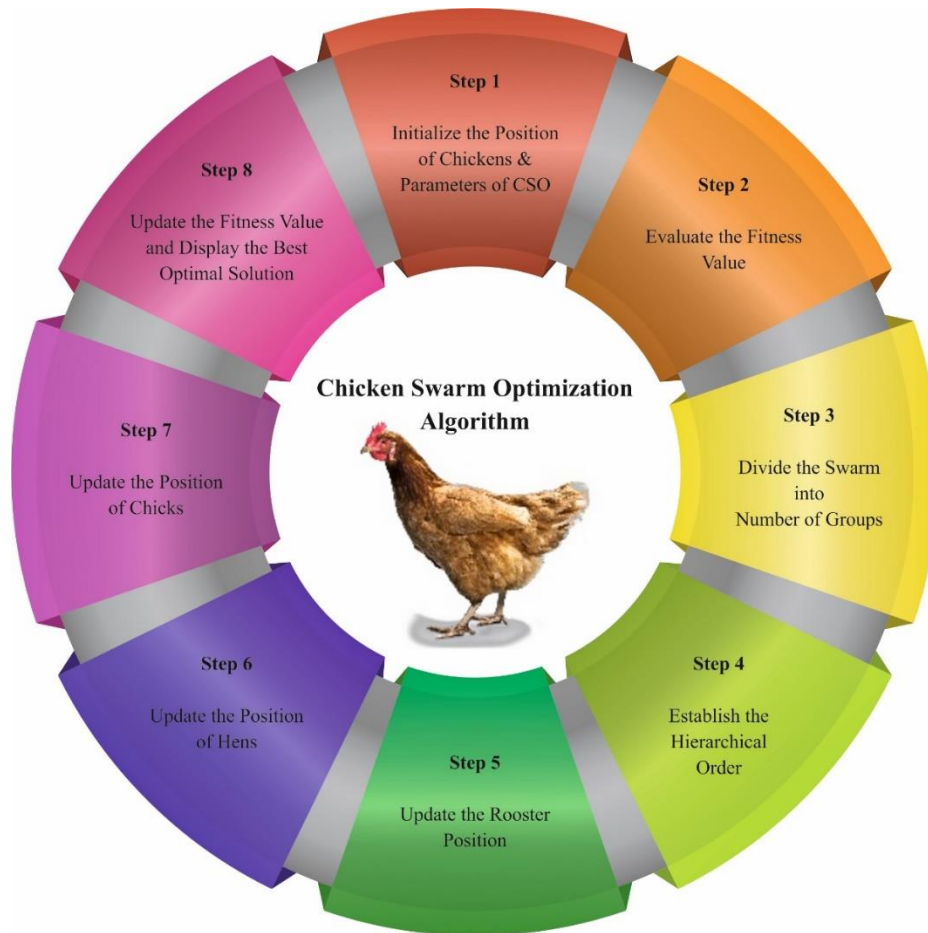


Figure 2: Steps involved in CSO

Mathematical model

The complete chicken group contains N number of individuals, the number of hens, roosters, chicks, and mothers are signified as N_R, N_H, N_M, N_C . Ten $x_{i,j}^t$ represents the location of the i th chicken in j th dimension of t th iteration, and M signifies the highest amount of iterations ($i \in (1,2, \dots, N), j \in (1,2, \dots, D), t \in (1,2, \dots, M)$)

The rooster with the finest fitness in the sub-group can choose the foraging way by itself. The upgrade formulation of this part is exposed in Eqs. (7) and (8):

$$x_{i,j}^{t+1} = x_{i,j}^t \times [1 + N(0, \sigma^2)], \tag{7}$$

$$\sigma^2 = \begin{cases} 1, & f_i \leq f_k \\ \exp\left(\frac{f_k - f_i}{|f_i| - \varepsilon}\right), & f_i > f_k, \\ k \in [1, N_R], k \neq 1 \end{cases} \tag{8}$$

Whereas, $N(0, \sigma^2)$ denotes the usual distribution; σ^2 represents the variance; 0 refers to the mean; k specifies the index of alternative rooster nominated at random, the fitness of rooster k and rooster i are represented by f_k and f_i . ε refers to the least amount in the system. The location upgrade formulation of hen is displayed in Eqs. (9) to (11).

$$x_{i,j}^{t+1} = x_{i,j}^t + S_1 R_1 (x_{r_1,j}^t - x_{i,j}^t) + S_2 R_2 (x_{r_2,j}^t - x_{i,j}^t), \tag{9}$$

$$S_1 = \exp\left(\frac{f_i - f_{r_1}}{abs(f_i) + \varepsilon}\right), \tag{10}$$

$$S_2 = \exp(f_{r_2} - f_i), \tag{11}$$

Whereas, the dual random numbers R_1 and R_2 satisfy the state of $R_1, R_2 \in [0,1]$; r_1 denotes the rooster in the hen's sub-group; ε refers to the lowest number in the system. r_2 is a nominated chicken at random that can be a hen or rooster. The chick location upgrade formulation is in Eq. (12).

$$x_{i,j}^{t+1} = x_{i,j}^t + FL(x_{m,j}^t - x_{i,j}^t). \quad (12)$$

Here, FL signifies its alteration parameter that randomly generated amount amongst $[0,2]$ and $[0,2]$. m denotes the chicken mother. The pseudocode for the CSO algorithm is Algorithm1.

Algorithm1: CSO algorithm
Initialize Complete chicken swarm size: N , highest number of iterations: M , Role assignment factor: G , Dimensions: Dim , Number of roosters: N_R , Number of hens: N_H , Number of chicks: N_C , Number of iterations: t ;
Initialize x , Save best_fitness, best_location;
While $t \leq M$ do
If $t \% G = 1$ then
Role assignment;
Sub-group division;
End
For $i = 1:N$ do
If $1 \leq x_i \leq N_R$ then
Upgrade x_j position utilizing Eqs. (7) ~ (8) ;
End
If $1 + N_R \leq x_i \leq N_R + N_H$ then
Upgrade x_j position utilizing Eqs. (9) ~ (11);
End
If $1 + N_R + N_H \leq x_i \leq N$ then
Upgrade x_i position utilizing Eq. (12);
End
Upgrade fitness based on x ;
Save best_fitness, best_location;
End
End
End

The CSO model originates a fitness function (FF) to get enhanced classifier performance. It defines a positive number to signify the enhanced candidate solution performance. In this study, the minimization of the classifier rate of error was measured as FF, as set in Eq. (13).

$$\begin{aligned} fitness(x_i) &= ClassifierErrorRate(x_i) \\ &= \frac{No. of misclassified samples}{Total no. of samples} * 100 \end{aligned} \quad (13)$$

4. Experimental Validation

The experimentation study for the STVNNLP-DCR system was shown utilizing dual datasets such as Buzz Feed and PolitiFact as definite in Table 1.

Table 1: Details of two datasets

BuzzFeed Dataset	
Classes	No. of Instances
Real News	182
Fake News	91
Total Instances	273
PolitiFact Dataset	
Classes	No. of Instances
Real News	240
Fake News	120
Total Instances	360

The Detection outcome of the STVNNLP-DCR technique on the BuzzFeed dataset is illustrated in Table 2 and Fig. 2. The result depicts that the STVNNLP-DCR system has the recognition of two classes accurately. With 70%TRAS, the STVNNLP-DCR model attains an average $accu_y$ of 93.72%, $prec_n$ of 95.74%, $reca_l$ of 90.32%, and F_{score} of 92.42%, respectively. Moreover, with 30%TESS, the STVNNLP-DCR technique gets an average $accu_y$ of 85.37%, $prec_n$ of 88.64%, $reca_l$ of 80.09%, and F_{score} of 82.33%, correspondingly.

Table 2: Detection outcome of STVNNLP-DCR model on BuzzFeed dataset

Classes	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}
TRAS (70%)				
Real News	93.72	91.49	100.00	95.56
Fake News	93.72	100.00	80.65	89.29
Average	93.72	95.74	90.32	92.42
TESS (30%)				
Real News	85.37	82.54	98.11	89.66
Fake News	85.37	94.74	62.07	75.00
Average	85.37	88.64	80.09	82.33

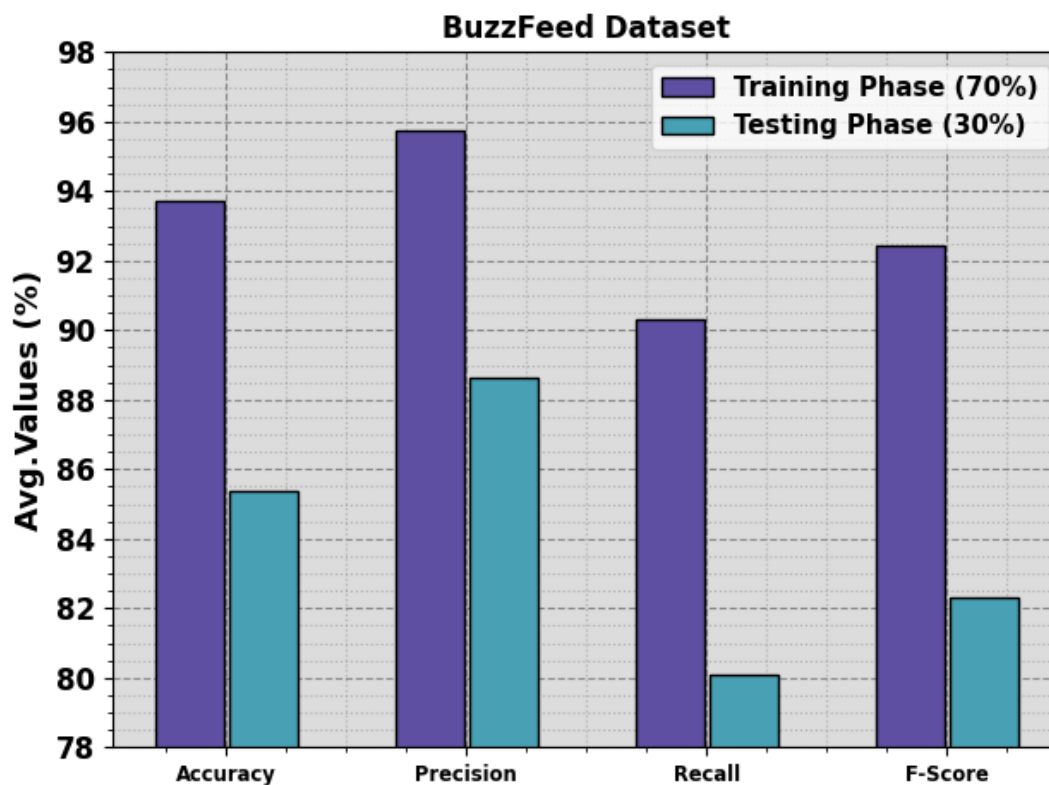


Figure 2: Average outcome of STVNNLP-DCR technique on BuzzFeed dataset

In Table 3 and Fig. 3, the comparative outcome of the STVNNLP-DCR approach with other approaches on the BuzzFeed dataset [12]. The result inferred that the STVNNLP-DCR system has better performances compared with other models. Based on $accu_y$, the STVNNLP-DCR technique has maximum $accu_y$ of 93.72% while the PBIC on Twitter, CIMTDetect, TFLI-FND, NBFND-PDA, DF-IFND DNN, EchoFakeD, and BAINLP-DCD models have lesser $accu_y$ of 90.17%, 90.82%, 91.15%, 90.86%, 91.16%, 91.43%, and 92.26%, respectively. Moreover, based on F_{score} , the STVNNLP-DCR system has the highest F_{score} of 92.42% whereas the PBIC on Twitter, CIMTDetect, TFLI-FND, NBFND-PDA, DF-IFND DNN, EchoFakeD, and BAINLP-DCD approaches have smaller F_{score} of 75.68%, 81.36%, 83.56%, 85.17%, 88.43%, 91.45%, and 92.42%, respectively.

Table 3: Comparative analysis of STVNNLP-DCR technique with other techniques on BuzzFeed dataset

BuzzFeed Dataset				
Models	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}
PBIC on Twitter	90.17	73.58	78.36	75.68
CIMTDetect	90.82	72.96	82.35	81.36
TFLI-FND	91.15	85.28	83.07	83.56
NBFND-PDA	90.86	84.97	85.28	84.26
DF-IFND DNN	91.16	83.40	87.01	85.17
EchoFakeD	91.43	90.54	87.44	88.43
BAINLP-DCD	92.26	94.53	89.25	91.45
STVNNLP-DCR	93.72	95.74	90.32	92.42

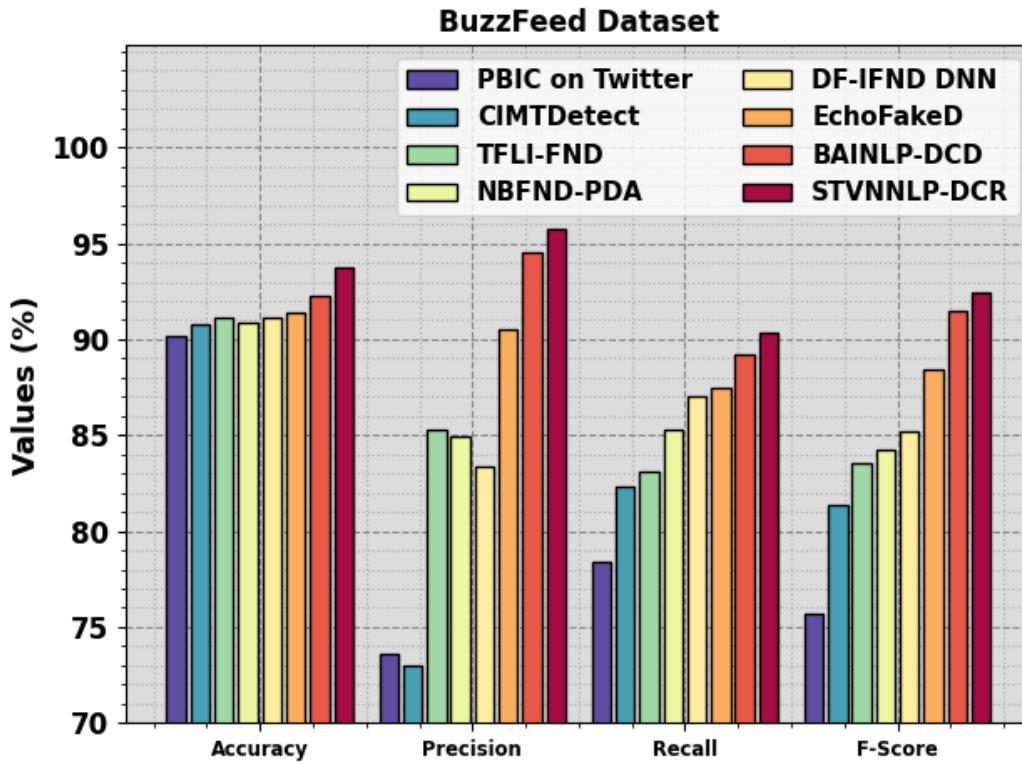


Figure 3: Comparative analysis of STVNNLP-DCR technique on BuzzFeed dataset

The Detection outcome of the STVNNLP-DCR model on the PolitiFact dataset is demonstrated in Table 4 and Fig. 4. The result portrays that the STVNNLP-DCR system has recognition of two classes precisely. With 70%TRAS, the STVNNLP-DCR technique gets an average $accu_y$ of 89.68%, $prec_n$ of 90.08%, $reca_l$ of 87.39%, and F_{score} of 88.46%, correspondingly. Furthermore, with 30%TESS, the STVNNLP-DCR system achieves average $accu_y$ of 93.52%, $prec_n$ of 95.93%, $reca_l$ of 87.93%, and F_{score} of 91.02%, correspondingly.

Table 4: Detection outcome of STVNNLP-DCR technique on PolitiFact dataset

Classes	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}
TRAS (70%)				
Real News	89.68	89.02	95.65	92.22
Fake News	89.68	91.14	79.12	84.71
Average	89.68	90.08	87.39	88.46
TESS (30%)				
Real News	93.52	91.86	100.00	95.76
Fake News	93.52	100.00	75.86	86.27
Average	93.52	95.93	87.93	91.02

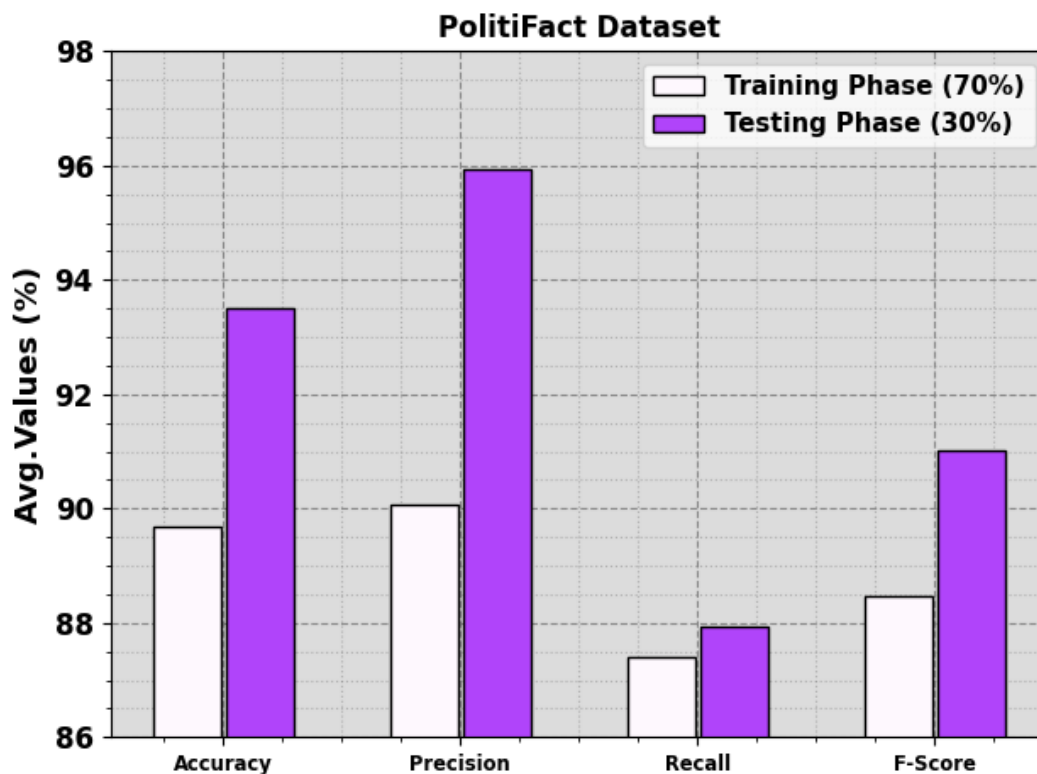


Figure 4: Average outcome of STVNNLP-DCR technique on the PolitiFact dataset

In Table 5 and Fig. 5, the comparative outcome of the STVNNLP-DCR model with other methods on the PolitiFact dataset. The result inferred that the STVNNLP-DCR system has superior performances compared with other methods. Based on $accu_y$, the STVNNLP-DCR technique has the highest $accu_y$ of 93.52% while the PBIC on Twitter, CITDetect, CIMTDetect, TFLI-FND, DF-IFND DNN, EchoFakeD, and BAINLP-DCD approaches has lower $accu_y$ of 91.60%, 88.91%, 90.56%, 91.33%, 86.44%, 90.48%, and 92.62%, correspondingly. Furthermore, based on F_{score} , the STVNNLP-DCR technique has the greatest F_{score} of 91.02% while the PBIC on Twitter, CITDetect, CIMTDetect, TFLI-FND, -PDA, DF-IFND DNN, EchoFakeD, and BAINLP-DCD models have smaller F_{score} of 78.37%, 79.18%, 81.86%, 84.38%, 84.10%, 88.45%, and 90.665%, correspondingly.

Table 5: Comparative analysis of STVNNLP-DCR technique with other approaches on PolitiFact dataset

PolitiFact Dataset				
Models	$Accu_y$	$Prec_n$	$Reca_l$	F_{Score}
PBIC on Twitter	91.60	77.77	79.17	78.37
CITDetect	88.91	67.95	77.55	79.18
CIMTDetect	90.56	80.38	84.28	81.86
TFLI-FND	91.33	87.25	82.18	84.38
DF-IFND DNN	86.44	82.15	84.65	84.10
EchoFakeD	90.48	86.43	80.53	88.45
BAINLP-DCD	92.62	94.47	82.62	90.65
STVNNLP-DCR	93.52	95.93	87.93	91.02

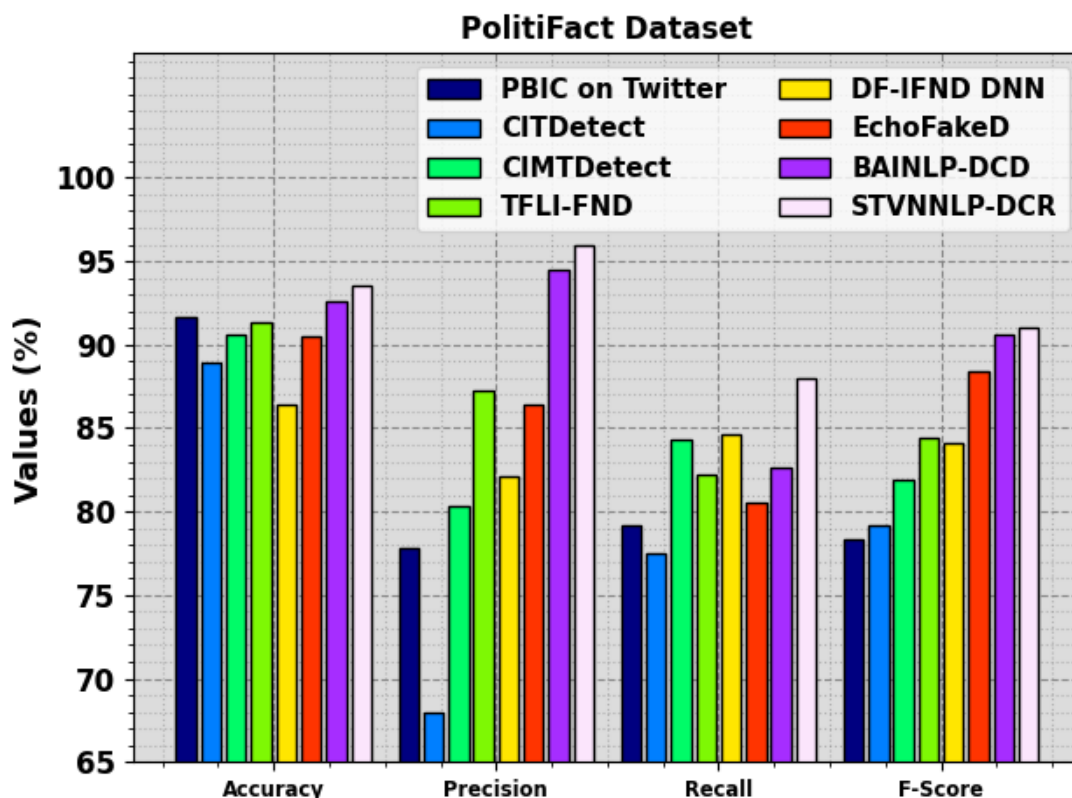


Figure 5: Comparative analysis of STVNNLP-DCR technique on PolitiFact dataset

5. Conclusion

In this study, we have presented an STVNNLP-DCR technique on social media. The presented technique includes four important elements: preprocessing, GloVe word embedding, STVN classification, and CSO for parameter tuning. The preprocessing stage includes tokenization and text normalization, preparing text information for succeeding analysis. Then, GloVe word embedding represents the word in a continuous vector space, which captures contextual relationships and semantic similarities. The STVN classifier deploys the embedding to discern deceptive patterns within the text, leveraging its capability to effectively manage high-dimensional and sparse datasets. Moreover, the CSO technique enhances the hyperparameter of the STVN classifier, improving its generalization capabilities and performance. Empirical analysis implemented on varied datasets validates the efficacy of the presented technique in precisely recognizing deceptive content. Comparative studies with advanced approaches demonstrate high efficiency.

Funding: “The author extends his appreciation to the Deanship of Scientific Research at King Khalid University for funding this work through large group Research Project under grant number (RGP2/ 185 /44)”

Conflicts of Interest: “The authors declare no conflict of interest.”

References

- [1] Hiramath, C.K.; Deshpande, G.C. Fake news detection using deep learning techniques. In Proceedings of the 2019 1st International Conference on Advances in Information Technology (ICAIT), Chikmagalur, India, 25–27 July 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 411–415.
- [2] Aslam, N.; Ullah Khan, I.; Alotaibi, F.S.; Aldaej, L.A.; Aldubaikil, A.K. Fake detect: A deep learning ensemble model for fake news detection. *Complexity* 2021, 2021, 1–8.
- [3] Amine, B.M.; Drif, A.; Giordano, S. Merging deep learning model for fake news detection. In Proceedings of the 2019 International Conference on Advanced Electrical Engineering (ICAEE), Algiers, Algeria, 19–21 November 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 1–4.
- [4] Kaliyar, R.K.; Goswami, A.; Narang, P. EchoFakeD: Improving fake news detection in social media with an efficient deep neural network. *Neural Comput. Appl.* 2021, 33, 8597–8613.

- [5] Dhanalakshmi, G., Sandhiya, S. and Smarandache, F., 2024. Selection of the best process for desalination under a Treesoft set environment using the multi-criteria decision-making method. *International Journal of Neutrosophic Science*, 23(3), pp.140-40.
- [6] Almuhr, E., Miqdad, H., Al-labadi, M. and Idrisi, M.I., 2024. μ -L-Closed Subsets of Noetherian Generalized Topological Spaces. *International Journal of Neutrosophic Science*, 23(3), pp.148-48.
- [7] Tashtemirovich, A.O., Balba, M.E., Ibrohimjon, F. and Batirova, N., Investigating the Impact of Artificial Intelligence on Digital Marketing Tactics Strategies Using Neutrosophic Set.
- [8] Sivakumar, C., Al-Qadri, M.O., Alsarairoh, A.A., Al-Husban, A., Meenakshi, P.M., Rajesh, N. and Palanikumar, M., 2024. q-rung square root interval-valued neutrosophic sets with respect to aggregated operators using multiple attribute decision making. *International Journal of Neutrosophic Science*, 23(3), pp.154-54.
- [9] Gharib, M., Fakhry, A.E., Ali, A.M., Abdelhafeez, A. and Elbehriy, H., 2024. Single Valued Neutrosophic Sets for Analysis Opinions of Customer in Waste Management. *International Journal of Neutrosophic Science*, 23(3), pp.184-84.
- [10] Ali Alqazzaz. "Integrated Neutrosophic methodology and Machine Learning Models for Cybersecurity Risk Assessment: An exploratory study." *International Journal of Neutrosophic Science*, Vol. 23, No. 3, 2024 ,PP. 195-207
- [11] Pospelova, N., Tarasova, A., Subbotina, N., Koroleva, N., Raimova, N. and Lydia, E.L., 2024. Explainable Artificial Intelligence and Natural Language Processing for Unraveling Deceptive Contents. *Fusion: Practice and Applications*, 14(2), pp.146-46.
- [12] Albraikan, A.A., Maray, M., Alotaibi, F.A., Alnfai, M.M., Kumar, A. and Sayed, A., 2023. Bio-Inspired Artificial Intelligence with Natural Language Processing Based on Deceptive Content Detection in Social Networking. *Biomimetics*, 8(6), p.449.
- [13] Sadeghi, F., Bidgoly, A.J. and Amirkhani, H., 2022. Fake news detection on social media using a natural language inference approach. *Multimedia Tools and Applications*, 81(23), pp.33801-33821.
- [14] Schmidt, S., Thoms, B., Eryilmaz, E. and Isaacs, J., 2024. Classifying Vaccine Misinformation in Online Social Media Videos using Natural Language Processing and Machine Learning.
- [15] Devarajan, G.G., Nagarajan, S.M., Amanullah, S.I., Mary, S.S.A. and Bashir, A.K., 2023. AI-assisted deep NLP-based approach for prediction of fake news from social media users. *IEEE Transactions on Computational Social Systems*.
- [16] Jadhav, P. and Shukla, R.K., 2024. Deep Learning Analysis for Revealing Fake News using Linguistic Complexity and Semantic Signatures. *International Journal of Intelligent Systems and Applications in Engineering*, 12(12s), pp.458-465.
- [17] Haque, A. and Abulaish, M., 2023, November. An Emotion-Enriched and Psycholinguistics Features-Based Approach for Rumor Detection on Online Social Media. In *Proceedings of the 11th International Workshop on Natural Language Processing for Social Media* (pp. 28-37).
- [18] Hicham, N., Nasser, H. and Karim, S., 2024. Enhancing Arabic E-Commerce Review Sentiment Analysis Using a hybrid Deep Learning Model and FastText word embedding. *EAI Endorsed Transactions on Internet of Things*, 10.
- [19] Bera, T. and Mahapatra, N.K., 2020. An approach to solve the linear programming problem using single valued trapezoidal neutrosophic number. *International Journal of Neutrosophic Science*, 3(2), pp.54-66.
- [20] Huang, H., Zheng, B., Wei, X., Zhou, Y. and Zhang, Y., 2024. NSCSO: a novel multi-objective non-dominated sorting chicken swarm optimization algorithm. *Scientific Reports*, 14(1), p.4310.