International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

An Ensemble Classification and Hybrid Feature Selection Approach for Fake News Stance Detection

A. Vaideghy[1], C. Thiyagarajan[2]
[1]Assistant Professor
Department of Computer Science
PSG College of Arts & Science
Coimbatore ,Tamilnadu,India.
Email: vaideghy@gmail.com
[2]Associate Professor,
Department of Computer Science,
PSG College of Arts & Science,
Coimbatore, Tamilnadu,India.
Email: pcthiyagu@gmail.com

Abstract- The developments in Internet and notions of social media have revolutionised representations and disseminations of news. News spreads quickly while costing less in social media. Amidst these quick distributions, dangerous or seductive information like user generated false news also spread equally. on social media. Distinguishing true incidents from false news strips create key challenges. Prior to sending the feature vectors to the classifier, it was suggested in this study effort to use dimensionality reduction approaches to do so. These methods would not significantly affect the result, though. Furthermore, utilising dimensionality reduction techniques significantly reduces the time needed to complete a forecast. This paper presents a hybrid feature selection method to overcome the above mentioned issues. The classifications of fake news are based on ensembles which identify connections between stories and headlines of news items. Initially, data is pre-processed to transform unstructured data into structures for ease of processing. In the second step, unidentified qualities of false news from diverse connections amongst news articles are extracted utilising PCA (Principal Component Analysis). For the feature reduction procedure, the third step uses FPSO (Fuzzy Particle Swarm Optimization) to select features. To efficiently understand how news items are represented and spot bogus news, this study creates ELMs (Ensemble Learning Models). This study obtained a dataset from Kaggle to create the reasoning. In this study, four assessment metrics have been used to evaluate performances of classifying models.

Keywords: Fake news, preprocessing, Principal Component Analysis (PCA), Fuzzy Particle Swarm Optimization (FPSO), Ensemble Learning Model (ELM), fake news detection.

I. INTRODUCTION

As more people converse online using social media platforms, these platforms host news items in place of traditional news media [1]. Fake news is one of today's most serious problems to the media, democracy, and economy. It has eroded public trust in governments, and it has to be seen how it impacts contested topics. Fake news also affects economies, producing volumes of transactions and influencing stock markets [2]. Interdisciplinary studies have examined human susceptibility to fake news which can provide useful hints for detecting fake news items. For example, advancements in forensic psychology like the Undeutsch theory have highlighted the differences in presentation between truth and false information [3]. Similar to this, interdisciplinary research has examined why people spread false information, taking into account that it can be difficult to distinguish between legitimate users and malicious ones. Even legitimate users might unintentionally engage themselves in distributing fake news items.

Fake news or connecting to obscure websites are presented as actual events or happenings. These are quickly invading social and mainstream media, but their unknown origins and contexts of material make them elusive and bombastic [4]. The use of ambiguous phrases like "according to experts," "studies demonstrate," or "a journalist declared" frequently assist in spreading rumours, presumptions, or conspiracies. Headlines of false news items are typically in uppercase or include capital letters. A partial list of false news items are listed below:

- Partisan news items are produced by manipulating real political or business events
- News items are generated with imaginary information
- Sensationalizing news contents pulled out from factual data
- Biased news based on reverse interpretations of genuine events to match political/corporate agendas

- Creation of information with wrong facts
- Misleading news contents based on true facts for creating sensations

Due to the features of false news, detecting it is a difficult process with many facets. The detection algorithms make use of a variety of news- and social-related information kinds, including feedback, propagation pathways, and spreaders [6]. Information content types can take texts or multimedia contents or networked or correlate various tactics and resources that can be accessed. As a result, researchers, practitioners, and business leaders must prioritise understanding and recognising fake news if they are to prevent it from having a negative impact on democracy, media, and economies [7]. A few further enhancements in identifying false news are detailed below:

(1) Interdisciplinary studies on false news can be encouraged since it may aid in qualitative and quantitative investigations, as well as creation of well-supported and easily comprehensible fake news detection tools.

(2) To properly comprehend and identify bogus news, an all-encompassing framework and approach are required. Such tactics draw and bring together scholars working on the issue of false news who are experts in relevant information and technology.

(3) To identify future research objectives and directions, it is important to clarify open questions and obstacles for fake news studies.

Traditional method for determining genuineness of news item depend on human expert knowledge who can verify and validate facts, but become complex when volumes and variations of news items increase. The growth of the Internet [8] , has resulted in software applications using AIs (Artificial Intelligences) and MLTs (Machine Learning Techniques) for detecting false news contents.

Studies describe identifications of false news from four perspectives namely credibility, styles, diffusions, and knowledge [9]. Knowledge based approaches on fake news detections entail "comparisons" between relational information received from news items that need to be authenticated and knowledge-bases reflecting facts/ground truths. PCAs when used minimize noises and irrelevant characteristics from feature vectors resulting in improved effectiveness of classifications for false news items. Such techniques for reducing data dimensionalities help in enhancing accuracies up to 98.24% and much higher than prior studies. This research work also takes into account interactions of textual features for improving classifier’s performances.

These detection models analyse true and fake news that are deceptive and purposeful where high-quality training data is a significant concern. This research work proposes a hybrid feature selection and ensemble based classifier model for identifying false news. Following this introductory section, Section 2 goes through some existing strategies for detecting false news. Section 3 presents an overview of the suggested technique for identifying false news. Section 4 displays experimental findings and discussions. This work concludes in Section 5.

II. LITERATURE REVIEW

The discussion of the numerous methods that have been put out to identify false news in different forms of data, a survey of current and relevant efforts in domains related to this work.

Granik et al. [12] used NBs (Naive Bayes) in their simple approach for detecting false news items. The study implemented their technique as a software solution which was tested on Facebook news posts. Their results scored 74% accuracy on test sets. Several strategies, which are also discussed in the article, can be utilised to improve these results. Artificial intelligence tools, according to the findings, can be utilised to tackle the challenge of detecting bogus news. Reis et al. [13] proposed a new set of criteria and assessed how well existing methodologies and features perform for automatic false news detection. The findings on the value and significance of traits for spotting bogus news are fascinating. Finally, talk about the potential and problems of using false news detection methods in practise. A false news detection algorithm utilising n-gram analysis and MLTs were proposed by Ahmed et al. [14]. Investigate and contrast six distinct machine classification strategies in addition to two different features extraction techniques. With an accuracy of 92%, the experimental evaluation's best results are obtained utilising the feature extraction method TF-IDFs (Term Frequency-Inverse Document Frequencies) and LSVMs (Linear Support Vector Machines). Rădescu et al. [15] described MLTs as they are used in the automatic detection of bogus news. Two straightforward yet incredibly effective strategies are discussed in detail, along with their implementations. First, a keyword search of the text is used, and second, the NBs approach of probability theory is used. The goal of this study, which makes up its original contribution, is to put these two ways of categorising the news into practise and assess how well they function. According to the functional testing trials, the aforementioned procedures are straightforward yet quite effective. Therefore, both approaches are capable of categorising news stories into 4 categories of authenticity based merely on their titles: news that is not false (genuine), news items that are moderately or safely untrue. Their method’s results demonstrated the effectiveness of NBs in

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

executions and significantly enhanced on high quality training samples.

To develop a corpus for predictive modelling and text analytics, Rubin et al. [16] developed three different forms of fake news, each compared against actual, serious reporting. As the distinctions between conventional news and online information become hazier making filtering, vetting, and validating online material remains crucial in library and information science. Ruchansky et al. [17] proposed models for enhanced accuracies and automatic predictions by including activities of users, articles and group behaviours on spread false information. Their modular paradigm (Capture, Score, and Integrate) reacted to the afore described parameters. RNNs (Recurrent Neural Networks) were used in the first module and captured user activity's temporal patterns from articles based on responses. Based on user behaviour, the second module detects source characteristics, which are then integrated into the third module for false news identifications. Their experimental results showed that their three modules recovered useful latent representations of users and articles while outperforming prior models in terms of accuracies. Tacchini et al. [18] proposed two categorization schemes where one was based on LRs (Logistic Regressions) while the other was based on novel usage of Boolean crowd sourcing technique. On datasets consisting of 15,500 Facebook postings and 909,236 participants, the research was able to obtain classification accuracies of above 99%, even on small training sets or comprising of small fractions of postings. Furthermore, they demonstrate the robustness of these strategies by functioning even when attention is focused just on people who like both hoax and genuine messages. The study's findings showed that information transmission patterns might be useful techniques for automated false news detections. Pérez-Rosas et al. [19] produced two fresh datasets with various news categories in order to detect bogus news where data gathering, annotations, and validations were detailed, as well providing variety of exploratory studies on identifying linguistic differences in false and actual news contents. Their future scope was to conduct series of learning experiments for developing reliable false news detections and additionally compare automations and human methods in bogus news identifications.

Saikh et al. [20] introduced two efficient deep learning-based models for tackling the issue of false news identification in online news items from various domains. Use the two previously provided datasets Fake News AMT and Celebrity for this evaluation of the methodologies for spotting false news. 3.08% and 9.3%, respectively, by significant margins, the suggested systems outperformed existing methods and manually created feature engineering-based systems. The study evaluated applicability of their systems across domains and used datasets available for pertinent tasks (e.g., models trained on Fake News AMT and tested on Celebrity, and vice versa). They used unmasking, a meta-learning methodology for verifying authorship verifications, Potthast et al. [21] offered a new method of style similarity assessments between texts and found that styles of left- and right-wing journalism had greater similarity. Furthermore, they showed that satires can be identified with high accuracies from both mainstream and hyper partisan news (F1=0.78 and F1=0.81, respectively). It is not surprising that style-based false news recognition fails (F1=0.46). The prior discoveries, however, are critical for pre-screening for fake news detectors. Goldani et al. [22] employed dimensionality reduction techniques and a dataset with four different types of attitudes before delivering the feature vectors to the classifier —agree, disagree, discuss, and unrelated—from Fake News Challenges website. The study loaded non-linear features to add more contextual characteristics for identifying bogus news using PCAs and chi-square test. The study aimed at ascertaining relation between news reports and headlines. The study showed results in terms of accuracies (4%) and F1-scores (20%). The study also found that PCAs outperformed Chi-square with 97.8% accuracy.

According to above listed studies, each strategy has certain benefits and drawbacks. Advanced deep neural models were used due to their advantages over traditional MLTs.

III. Proposed Methodology

In this study, a hybrid ensemble-based classifier model for the identification of false news is introduced. This model's suggested four-step process for spotting bogus news on social media. The steps followed in this study are detailed below:

- The first phase is executions of pre-processes on data for transforming unstructured data into structured data.
- PCAs are introduced in the second phase for extracting features of false news and their connections with news contents.
- The third step introduces FPSO based feature selection approach for reducing features.
- This work used Fake News Challenges website data containing different views for reasoning namely agree, disagree, debate, and irrelevant.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

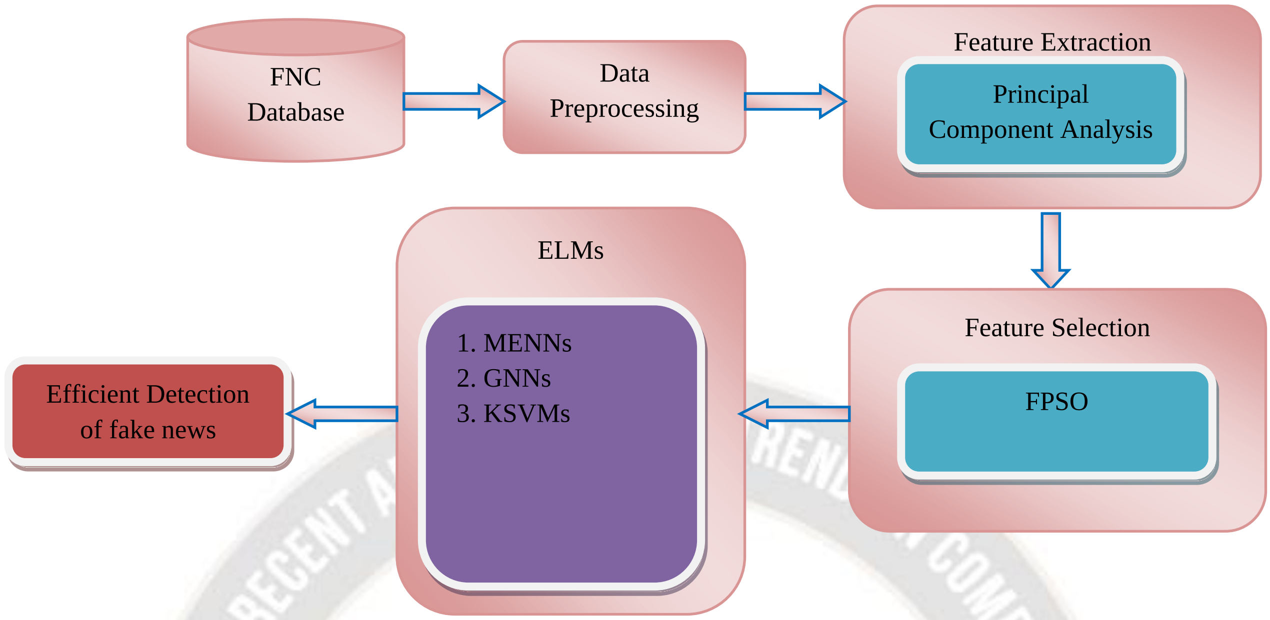

Figure 1. The process of Fake news detection model

3.1. DATASET

The dataset used in this study, a benchmark dataset from Fake News Challenges had 75, 385 tagged occurrences and 2, 587 article bodies corresponding to three hundred headlines [23]. It had between fie to twenty articles for allegations. According to Table1, of these headlines, 7:4% are in agreement, 2% are in disagreement, 17:7% are being debated, and 72:8% are unconnected. Manual labelling is used for the assertions pertaining to the body of the articles. These are the details of the labels:

Agree – implies relationships between headlines and article's body.

Disagree - implies there are no relationships between headlines and article's body.

Discuss – implies partial matches between headlines and article's body, neutralizing them.

Unrelated- Implies no connections between headlines and body of the article.

The data set was split into 25/ 413 instances for testing and 49/ 972 instances for training. This training and testing data distributions were created using FNC-1 challenge's guidelines. In training, article bodies were 1; 683, whereas headlines were 1; 648. The test data encompassed 880 headlines and 904 articles.

TABLE 1. Dataset statistics.

Dataset	Headlines	Tokens	Instances	Agree	Disagree	Discuss	Unrelated
FNC-1	2,587	372	75,385	7.4%	2.0%	17.7%	72.8%

3.2. PRE-PROCESSING

Pre-processes are data mining steps that transform raw unstructured and incomplete data into machine readable formats. On the database provided by the Fake News Challenge, a variety of pre-processing methods were applied to the text. NLPs such as stop word removal, stemming, tokenization, and text character conversion to lowercase letters were used to solve these tasks using Keras library algorithms. Stop words are extremely prevalent words found in the text that are unnecessary to this work and have very little significance in terms of characteristics, such as "of," "the," "and," and "an." By eliminating stop words, processing times were reduced freeing up space that would have been occupied by unnecessary words. Words with similar meanings, such as games, may appear many times throughout the text. If this is the case, it is highly efficient to reduce the terms to their most basic versions. This operation, known as stemming, is done out using the NLTK's open-source version of the Porter stemmer method.

The headlines counts were reduced to 372 after the executions of pre-processes. Headlines were divided into word vectors using tokenizers. Word embeds (word2vec) mapped words or texts to vector sets after pre-processes. A dictionary of the 5,000 unigram terms was generated in the study from article headlines and body texts. The maximum headline length is the length that all of the headlines are set to. Zero padding is used for headlines that are less than the maximum length. Next, the ensemble learning model receives the features (ELM).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

3.3. DIMENSIONALITY REDUCTION METHODS

This work used two strategies for reducing dimensionalities of texts namely selections and extractions of features. During feature selections, only important and pertinent features were maintained. Feature extractions involved changing initial vector spaces to produce new vector spaces which were compressed and with unique properties. Minimizing features taken for processing benefits faster executions. Performances of text classifications are significantly impacted by feature reductions. Selecting proper features are crucial for reducing dimensions of data. This work's dimensionality reductions based on PCAs and FPSO in conjunction with deep learning models boost scalability of text classifiers.

- ***PCAs***

PCAs are standard strategies that reduce dimensionalities of features using linear transformations. The features of original datasets are preserved even if resulting datasets are smaller [24]. New datasets are generated with reduced feature counts when compared to original dataset feature counts where fundamental components are computed using covariance matrices which list important features based on criteria [25]. Equation (1) may be used to determine transformations of original matrices into a "t"-dimensional sub-spaces with "a" dimensions and "b" observations.

$$Y = (E^Z X) \qquad (1)$$

Where, *Ea_t represents* projection matrix containing *the largest t eigen values' associated t eigen vectors*, and *Xa_b* stands for mean centered data matrices.

3.4. FEATURE SELECTIONS USING FPSO

Fuzzy based features of behavioural patterns are characterised by fuzzy matrices, where n represents data item counts and c stands for cluster counts, and matrices have n rows and c columns. The ith object's level of connections or membership functions with jth clusters are indicated by elements in ith row and jth column in the expression ij [26]. This work's suggested approach uses global search capabilities of PSO (particle swarm optimization) to overcome drawbacks of fuzzy based feature processes. PSO can be used to handle a variety of function optimizations as well as solve issues converted as function optimization problems.

Fish schools and avian flocking behaviours served as inspiration for PSOs, population-based stochastic optimization approaches. On iterations or generations, they are based. Particle population is where the algorithmic flow of PSOs starts and locations serve as potential solutions to investigated problems. Velocities are initialised at random in the search space. Iteratively searching for the best spot is accomplished by altering locations and speeds of particles. Additionally, iterations use fitness functions to determine locations of particle fitness values. Individual best positions and overall best positions are used to update particle velocities. The best positions in swarms since first time steps are s best, and best locations of particles ever visited are known as p best while g best is the best global position. The positions and speeds of particles can be updated using:

$$V(t+1) = w.V(t) + c_1 r_1\big(pbest(t) - X(t)\big) + c_2 r_2\big(gbest(t) - X(t)\big); \quad k = 1,2,\ldots..P \qquad (2)$$

$$X(t+1) = X(t) + V(t+1) \qquad (3)$$

Where V stands for particle velocities, X are particle positions. P implies the counts of particles in the swarm, w represents inertia weight while c1, c2 are positive acceleration coefficient constants that help search engines find optimal values for p and g, respectively, and r1, r2 are random values in the range [0, 1]. The suggested FPSO technique redefines particle locations and velocities to express fuzzy relationships between variables.

The location of the particles in the FPSO algorithm X, which depicts the relations between sets of data items, are hazy, $o = \{o_1, o_2, \ldots \ldots, o_n\}$, with cluster centers, $Z = \{z_1, z_2, \ldots \ldots, z_c\}$. X which can be expressed as:

$$X = \left[\mu_{1l} \;\ldots\; \mu_{1c} \;\vdots \ddots \vdots\; \mu_{nl} \;\ldots\; \mu_{nc}\right] \qquad (4)$$

Where, μ_{ij} represents membership functions of i^{th} objects with j^{th} clusters with constraints. This results in positional matrices of particles are the same as fuzzy matrices μ. Velocities of particles are described using matrices of elements that fall between [-1, 1] in c columns and n rows. Based on matrix operations, the equations for location, velocity updates of particles can be:

$$V(t+1) = w \otimes V(t) \oplus (c_1 r_1) \otimes \big(pbest(t) \ominus X(t)\big) \oplus (c_2 r_2) \otimes \big(gbest(t) \ominus X(t)\big) \qquad (5)$$

$$X(t+1) = X(t) \oplus V(t+1) \qquad (6)$$

After changing the position matrix, the limits in (5) and (6) may no longer be followed. As a result, the position matrix must be normalised. First, we set the negative elements of the matrix to zero. To modify the matrix without violating the constraints, all zero elements in a row of the matrix must be revaluated using a sequence of random values between [0, 1].

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

$$Xnormal = \left[\frac{\mu_{11}}{\sum_{j=1}^{c} \mu_{1j}} \cdots \frac{\mu_{1c}}{\sum_{j=1}^{c} \mu_{1j}} \vdots \ddots \vdots \frac{\mu_{n1}}{\sum_{j=1}^{c} \mu_{nj}} \cdots \frac{\mu_{nc}}{\sum_{j=1}^{c} \mu_{nj}}\right] \quad (7)$$

Like other evolutionary algorithms, the FPSO method requires a fitness function to evaluate the generalised solutions. Eq. (8) is utilised in this study to evaluate the answers.

$$f(X) = \frac{K}{Jm} \quad (8)$$

The goal function is Jm, and K is a constant. The clustering effect is better and the individual fitness is higher the smaller Jm is (X).

Algorithm 2 Fuzzy PSO for feature selection

Inputs: Extracted features
Results: enhanced features

1. Set the settings, including the maximum iteration count, population size P, 1c, 2c, and w.
2. Assemble P particles into a swarm (X, pbest, gbest, and V are all nC matrices).
3. Set X, V, pbest for each particle, and gbest for the swarm as initial values.
4. Determine each particle's cluster centres.
5. Utilizing Eq., determine each particle's fitness value (8).
6. Determine each particle's pbest.
7. Determine the swarm's gbest.
8. Using Eq., modify the velocity matrix for every particle (5).
9. Using Eq., modify the position matrices for each particle (6).
10. Go to step 4 if the terminating condition is not satisfied.

The suggested method's termination condition is either reaching observing no improvement over time or doing the maximum amount of iterations.

We can extract the set of all normal characteristics for each group based on the normal feature set of each cardholder:

$$G_j = \cup_{id \in j} G^{id}, \forall j \in V \quad (9)$$

It's challenging to assign these common feature sets to particular patterns in the actual world (words). The explanation for this is that when patterns (or articles) are drawn using human domain expertise, their definitions may be ambiguous. To handle this supervised learning challenge, which can automatically arrange highly abstract knowledge, it is more practical to employ a classification algorithm.

3.5. CLASSIFICATION USING ELMs

Ensemble methodologies combine many models where they address different aspects of original issues, in order to create composite global models that provide estimates and decisions that are more accurate and trustworthy than those produced using single models. MENNs (Modified Elman Neural Networks), GNNs (Granular Neural Networks), and KSVMs (Kernel Support Vector Machines) which are employed in this ensemble model are detailed below:

3.5.1. ENNs

Fig. 2 depicts fundamental composition of ENNs. Input, hidden, context, and output layers make up the Elman network's four fundamental layers. Pairs of adjacent levels are connected by movable weights [28] where it is widely acknowledged to be specific classes of local feedbacks and extra memory neurons in feed-forward neural networks. ENNs are sensitive to history of inputs due to the context nodes' internal connections, which extremely useful for modelling dynamic systems. The following notations are used in this section:

$w1_{ij}$: weight connecting input layer's ith node hidden layer's jth node.

$w2_{ij}$: weight connecting hidden layer's ith node output layer's jth node.

$w3_{ij}$: weight connecting ith context node to jth hidden layer node.

m, n, r: Counts of inputs, outputs, and hidden layers.

$u_i(k), y_i(k)$: Inputs and outputs of ENNs, where i=1,2,.....m, and j=1,2,.......n.

$x_i(k)$: Outputs of hidden nodes i, where i=l, 2,......r.

$c_i(k)$: Outputs of context nodes i or outputs of hidden nodes i of last time. z^{-1}**:** A unit delay.

Context units are additional units that are added for each buried layer unit. All hidden units are inextricably linked to context units which implies context and hidden unit have weights with recurring connections between them. Hidden units, however, are only related to their corresponding context units (seeFig. 2).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

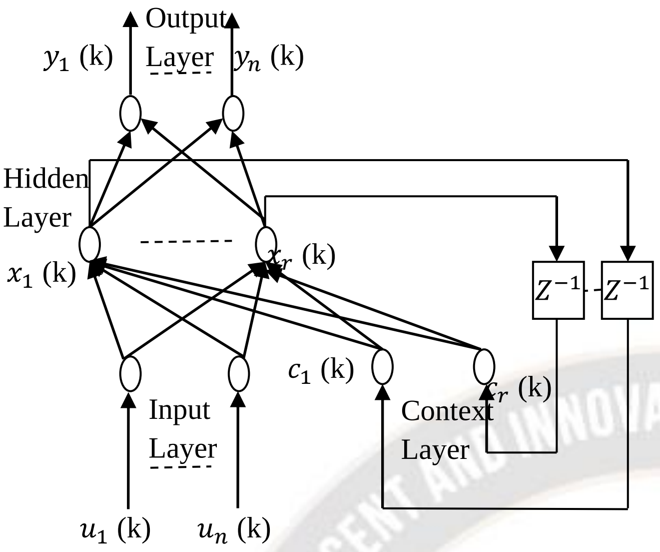

Fig. 2. Elman neural network model structure

BPs (Back Propagations) are used to learn forward weights, whereas recurrent connection weights are fixed. The context units serve as input units during forward phases. Similar to feed forward networks, values of output units and hidden units are computed. Through recurrent connections, current values are copied into their related context units once hidden units have computed their outputs (through unit delays). These variables need to be initialized with certain values before they may be utilised in following time steps. The use of output targets and the modification of forward weights, during backward training phases via BPs. The inputs of network are: $(k) \in R^m$ $y(k) \in R^n$, $x(k) \in R^r$, then each layer's outputs may be provided by

$$x_j(k) = f(\sum_{i=1}^{m} w2_{i,j} u_i(k) + \sum_{i=1}^{r} w1_{i,j} c_i(k)) \quad (10)$$

$$c_i(k) = x_i(k-1) \quad (11)$$

$$y_j(k) = g(\sum_{i=1}^{r} w3_{i,j} x_i(k) \quad (12)$$

where, respectively, f(.) and g(.) are the hidden and output layers' linear or nonlinear output functions. The state need not be used as input or training signals since the dynamic properties of ENNs are completely governed by internal connections. ENNs have advantages over static feed-forward networks. Because of instabilities of learning processes caused by absence of complete gradient information, ENNs can only identify single order linear dynamic systems when utilising traditional learning approaches of BPs, which only have first-order gradients. To solve these issue, two techniques, one employing dynamic BP algorithm and modified ENNs have been developed.

- **Modified ENNs**

Two modified network models are suggested to boost the dynamic properties and convergence rate of the original Elman network. One change possibly, Elman neural network shown in Fig. 3 (ME).

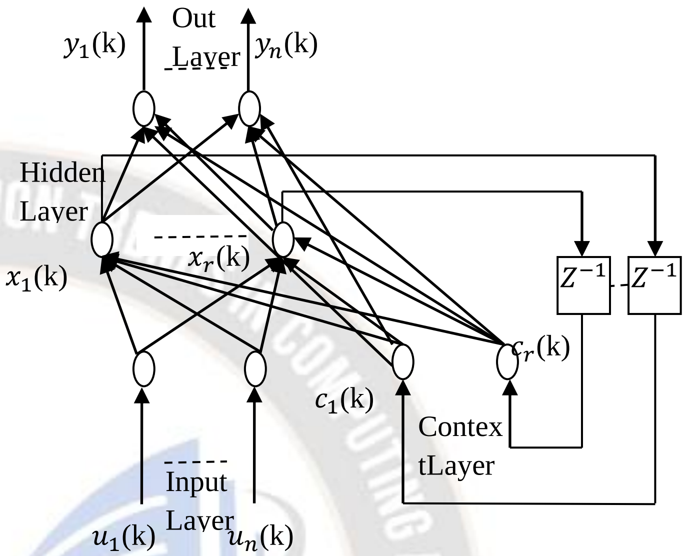

Figure 3. Modified architecture of ENNs

Compare Figs. 2 and 3 to determine the differences between the two networks. In context units of modified Elman networks, auto feedback connections with fixed gains are present. The outputs of context layers at time k are equal to outputs of the hidden layers at time k-1 as a result of additions and multiplications of outputs from context layers at previous times. The modified Elman networks revert to standard ENNs when fixed gains equal to 0.

The updated Elman network's formulation of nonlinear state space is similar to the earlier expression and is as follows:

$$x_j(k) = f(\sum_{i=1}^{m} w2_{i,j} u_i(k) + \sum_{i=1}^{r} w1_{i,j} c_i(k)) \quad (13)$$

$$c_i(k) = x_i(k-1) + \alpha \times c_i(k-1) \quad (14)$$

$$y_j(k) = g(\sum_{i=1}^{r} w3_{i,j} x_i(k)) \quad (15)$$

- **Gaussian weight updating function based elman neural network**

The alteration proposed for MENNs is the inclusion of extra configurable weights between the context and output nodes. The weight w4ij in Figure 2 is what connects context nodes I to nodes j in the output layers. Naturally, they utilize outputs of context layers as inputs to output layers. The proposed alterations of the network model are additions of

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

moveable weights between context and output nodes called Gaussian weight updating function (w4ij). their non-linear state space expressions look like this.

$$x_j(k) = f(\sum_{i=1}^{m} \quad w2_{i,j}u_i(k) + \sum_{i=1}^{r} \quad w1_{i,j}c_i(k)) \qquad (16)$$

$$c_i(k) = x_i(k-1) \qquad (17)$$

$$y_j(k) = g(\sum_{i=1}^{r} \quad (w3_{i,j}x_i(k) + w4_{i,j}\, c_i(k)) \qquad (18)$$

The above analysis indicates modified Elman networks possess features of proportion and integral. As weights were altered, their proportionate gains and integral coefficients varied. Controlled increments of updated networks are adjusted with changes to the input units, unlike ordinary PID algorithms, however, it also reinforces or restrains outputs at the final moment using a + wl weighting function. When a + wl > 1 implies minimal increases in control outputs; when a + wl < 1, the last measure lowers control outputs; and when a + wl = 1, are standard variable-parameters. The network maintains its features of integration and proportionality if a = 0, w4 f 0, or if a f 0 and w4 = 0. And if the modified Elman networks revert to original fundamental Elman network at -0 and w4=0, the result is:

$$y(k) = w1 \times y(k-1) + w2 \times w3 \times u(k) \qquad (19)$$

The typical integral equation looks like this. As a result, its dynamic reaction will become slower, which will slow down convergence. According to the aforementioned theoretical investigations, compared to the original Elman networks, the updated Elman networks exhibit improved dynamic characteristics.

3.5.2. GNNs

GNNs are artificial neural networks that can process data that was previously numerical or granular [29]. GNNs focus on incremental learning via internet data sources. GNN learning use similar two-step procedures as shown in Figure 4. Initially, intervals or more generally, fuzzy sets are built as information granules using original numeric representations. Instead of using their initial inputs, neural networks learn, adapt, and improve in contrast to information granules, which are significantly less prevalent, neural networks don't need to be exposed to all data since samples are eliminated when they don't contain any new information.

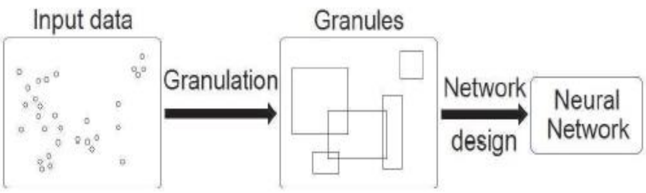

Figure 4. Granular neural network design in two stages

To handle data streams, GNNs rely on rapid incremental one-pass data learning algorithms. They may begin training knowing nothing about the statistical properties of the data or classes. To generate decision boundaries between classes, the technique granulates the feature space with fuzzy hyper boxes. The following are the primary characteristics of GNNs:

- Interacts with labelled and unlabelled instances at the same time
- Detect variations and are capable of dealing with uncertainty in data
- Shows a capacity for nonlinear separation
- Uses both bottom-up (constructive) and top-down (destructive) strategies for lifelong learning
- Ignores what is no longer important

a) Structure and Processing in GNNs

The GNN learns from a stream of data $x[h], h = 1, 2, ...$ There may or may not be a class designation next to the training examples $C[h]$. Each information granule γ_i of a limited number of granules $\gamma = \{\gamma 1, \dots, \gamma c\}$ in the feature space $X \subseteq Rn$ has a class attached to it Ck from the limited set of classes$C = \{C1, \dots, Cm\}$ in the output space $Y \subseteq N$. Utilizing T-S neurons and granules taken from the input stream, the GNN connects the feature and output spaces [30]. According to Fig. 5, the neural network has 5 layers. In essence, the input layer is a fan of feature vectors $x[h] = (x1, \dots, xj, \dots, xn)[h], h = 1, \dots$ into the network; the collection of information grains that make up the granular layer γi $\forall i$ inside the bounds of the feature space. Granules are permitted to partly overlap, and null neurons are included in the aggregation layer $TSni\, \forall i$. To produce values, they aggregate membership values $oi\, \forall i$ comparing values reflecting class compatibilities between examples and granules in decision layers o_i, and classes $\underline{C}_k$ associated with granules γ_i The output layers, which have greatest compatibility score, contain class label indications. except for input layers evolved as $x[h], h = 1, \dots$, are input.

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

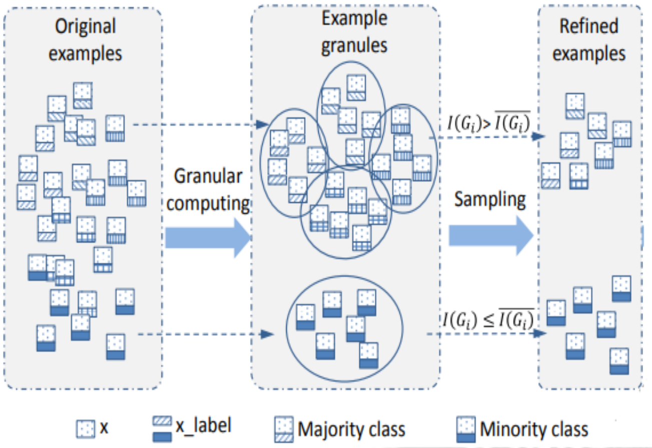

Fig. 5. Structure of evolving GNNs in classifications

Depending on the application, many techniques to structural and parametric modification of GNNs may be utilised in classifications. For example, if the number of courses is known ahead of time, it may be controlled automatically. When memory and processing time can be constrained, granules counts in model structures can also be constrained.

3.5.3. KSVMs

MLTs built on statistical learning theories are SVM models. A collection of support vectors that depict data patterns are used to categorise the data. Finding a discriminant function is a common challenge in two-class classification f(x), such that $y_i = f(x_i)$ given N data samples $(x_1, y_1) \ldots (x_i, yi) \ldots (xN, yN)$. The following is an example of a potential linear discriminant function f(x) = sgn (w · x − b) where w · x − b = 0 in the data space may be seen as a separating hyperplane. Finding a hyperplane with the biggest separation margin between the two classes is the key to selecting a discriminant function [31]. The final formulation of the linear discriminant is f(x) = sgn (Pl i=1 αiyi (xi · x − b), where l denotes how many training recordings there are, yi ∈ {−1, +1} a label that is connected to the training data, 0 ≤ αi ≤ C (constant C > 0), and the support vectors are xi. When the separating surface is not linear, to have two classes data points can be moved to another higher dimensional space for making them linearly separable. The following is the SVM's non-linear discriminant function:

$$f(x) = sgn(\sum_{i=1}^{l} \quad \alpha_i y_i K(x_i, x) + b), \quad (20)$$

where K (x_i, x) stands for kernel procedure that changes data points which may use linear or polynomial or sigmoid function or radial basis functions. These kernel functions don't take into account the variations in data characteristics. As can be seen from the generic SVM kernel function format K (xi, x), all training or test dataset characteristics are given the same consideration. Equal treatment of each feature could not be practical and could reduce SVM accuracy. Adding weights to a kernel function is one method to take the relevance of various attributes into account [32]. Each feature's relevance is determined by the weights. K (wxi, wx), where w is a vector of weights for the features in the data set, represents the new kernel function. A nonlinear discriminant function with feature weights is shown in the formula below,

$$f(x) = sgn(\sum_{i=1}^{l} \quad \alpha_i y_i K(wx_i, wx) + b), \quad (21)$$

This kernel is not required for certain kernel functions. Depending on the application, to apply feature weights to, the best kernel function may be selected. These weights are calculated and generated from training data using rough set theory. According to the basic criteria of weight computation, a characteristic has no weight if it does not occur in any reducts; it has more value if it appears in more reducts and it has a higher importance if there are less features in a reduct. If a reduct just possesses one feature, that trait is of utmost significance.

The identification of false news may be done effectively using an ensemble learning model-based classification approach.

IV. Results And Discussion

A number of tests were run using the FNC-1 datasets listed below to gauge the efficacy of the proposed model. This section describes these trials and compares the findings to other industry standards. For the purpose of detecting false news, several datasets have been presented. Having a big dataset is one of the key prerequisites for employing neural networks while training the model. Use a dataset from Kaggle that contains a huge number of texts for training deep models in this research project.

CNNs (Convolutional Neural Networks) with LSTM (Long Short Term Memory), CSI (Capture, Score, and Integrate), and the specified ELMs were used to evaluate how well this research performed on the provided dataset. The next section specifies the assessment measures utilised in the experiments. The TP (True Positive), FP (False Positive), TN (True Negative), and FN (False Negative) rates were used to calculate different performance indicators. Precisions are fractions of relevant events amongst retrieved instances while Recalls are defined as percentages of relevant occurrences. Despite their frequently contradictory characters, accuracies and recall metrics are both critical for judging efficiencies of prediction approaches. Hence, these two metrics may be added together and given equal weights to create single metrics

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

called F-Measures. Accuracies which are measured as percentages of occurrences that were successfully predicted compared to all instances that were forecasted. These three metrics were used in this study for evaluating the proposed technique

The ratio of accurately discovered positive observations to all anticipated positive observations is what is meant by precision.

Precision = TP/(TP+FP) (22)

The ratio of accurately detected positive observations to all observations is what is known as the sensitivity or recall.

Recall = TP/(TP+FN) (23)

The definition of F-measures is balanced averages of Precisions and Recalls that account for both false positives and false negatives.

F1 Score = 2*(Recall * Precision) / (Recall + Precision) (24)

Accuracies are computed the following list of positives and negatives:

Accuracy = (TP+FP)/(TP+TN+FP+FN) (25)

Table 1. Performance comparison results of proposed and existing methods for the Given FNC-1 dataset

Performance Metrics	CSI	CNN-LSTM	ELM
Accuracy	85.1500	97	98
Precision	82.019	96.90	99
Recall	84.234	96.70	97.90
F-Measure	83.00	96.7996.79	98.44

The performance comparison results of suggested and existing approaches for the provided FNC-1 dataset are shown in table 1. The comparison table demonstrates that the suggested ELM model outperforms the current false news detection methods in terms of detection accuracy.

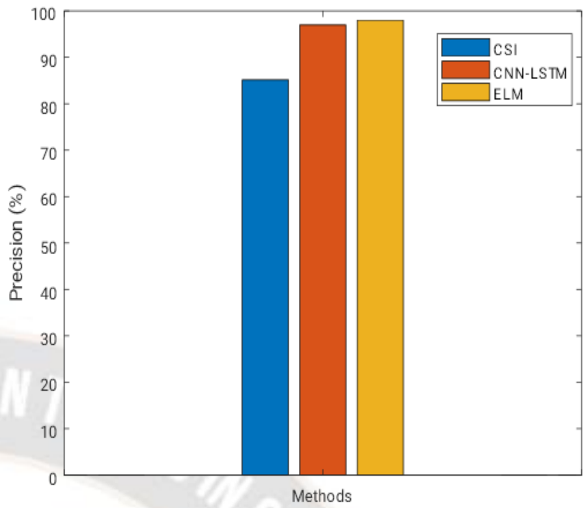

Fig.5. Results of a precise comparison of the proposed and current fake news detection models

The proposed and existing false news detection accuracy comparison models is shown in Figure 5. Using PCAs are more efficient for dimensionality reductions as they considerably increase accuracies which can be inferred from reviews of findings. By delivering results with precisions of 99%, the proposed model beats all competing models. According to the findings, the suggested ELM approach outperforms the already used classification algorithms in terms of precision.

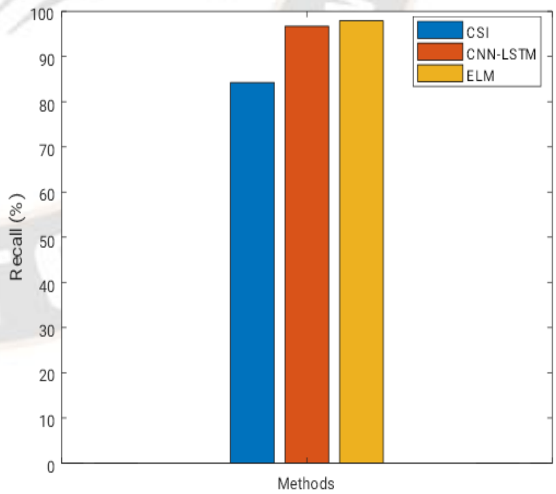

Fig.6. Results of the proposed and current fake news detection models were compared

In Fig. 6, the proposed and existing fake news detection algorithms' recall comparisons are shown. The statistical significance assures that using our suggested approach, any

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

news can be categorised as authentic or phoney with ease. Producing sequences takes a while, and it might over fit if the amount of data is fairly huge. Current CNN architectures, however, lack memory units and cannot handle data sequences. However, important information can be extracted as inputs to ELMs using PCAs and FPSO.

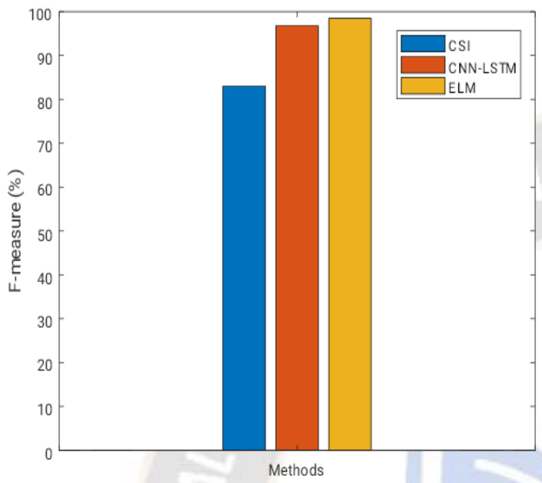

Fig.7. Results of the proposed and current fake news detection models' F-measure comparison

In Fig. 7, the proposed and existing fake news detection methods are compared using the F-measure. Precision, recall, and F1-score all saw significant increases. Furthermore, utilising dimensionality reduction techniques significantly reduces the time needed to complete a forecast. According to the findings, the suggested ELM methodology outperforms existing classification methods by a wide margin.

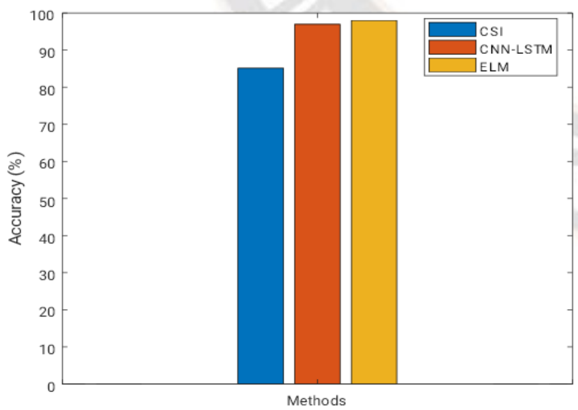

Fig.8. The proposed and existing fake news detection models' accuracy comparison results

In Fig. 8, the proposed and existing fake news detection methods' accuracy comparisons are shown. The results show an accuracy with a low 78% while employing features without data cleaning or preparations. It shows that the original dataset has a large amount of inconsistent, redundant, and noisy data. The accuracy increases considerably to 98% when preprocessing is completed and unnecessary data is removed. According to the findings, the suggested ELM approach provides higher accuracy outcomes than the already used classification techniques.

V. Conclusion

The profusion of information on social media has made it increasingly difficult for customers to get reliable information. In this article, we propose a hybrid feature selection with ensemble learning model to detect fake news on social media by combining text mining approaches with supervised intelligence algorithms. In contrast to earlier study, this research developed an algorithm for spotting false news based on the title of the item and content rather than only examining certain words or phrases. This model continues the feature extraction and selection process after a number of pre-processing steps have been completed. The categorization method based on ELMs is then implemented. PCAs enhance classifier's effectiveness in identifying fake news by reducing redundant, noisy, and unimportant data from feature vectors. The results from this method are good as they achieve 98.24% accuracy, which is substantially better than those from earlier studies. It should be highlighted that dimensionality reduction techniques minimize the amount of features while maintaining good classifier performances. Diverse textual elements will also be the focus of this research, along with how they combine to improve performance.

References

[1] Manzoor, S. I., & Singla, J. (2019, April). Fake news detection using machine learning approaches: A systematic review. In *2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI)* (pp. 230-234). IEEE.

[2] Gilda, S. (2017, December). Notice of Violation of IEEE Publication Principles: Evaluating machine learning algorithms for fake news detection. In *2017 IEEE 15th student conference on research and development (SCOReD)* (pp. 110-115). IEEE.

[3] Khan, J. Y., Khondaker, M., Islam, T., Iqbal, A., & Afroz, S. (2019). A benchmark study on machine learning methods for fake news detection. *arXiv preprint arXiv:1905.04749*.

[4] Zhou, X., Zafarani, R., Shu, K., & Liu, H. (2019, January). Fake news: Fundamental theories, detection strategies and challenges. In *Proceedings of the twelfth ACM international conference on web search and data mining* (pp. 836-837).

International Journal on Recent and Innovation Trends in Computing and Communication
ISSN: 2321-8169 Volume: 11 Issue: 4s
DOI: https://doi.org/10.17762/ijritcc.v11i4s.6304
***Article Received**: 20 December 2022 **Revised**: 18 January 2023 **Accepted**: 02 February 2023*

[5] Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017). Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter*, *19*(1), 22-36.

[6] Zhou, X., & Zafarani, R. (2018). Fake news: A survey of research, detection methods, and opportunities. *arXiv preprint arXiv:1812.00315*, *2*.

[7] Oshikawa, R., Qian, J., & Wang, W. Y. (2018). A survey on natural language processing for fake news detection. *arXiv preprint arXiv:1811.00770*.

[8] Parikh, S. B., & Atrey, P. K. (2018, April). Media-rich fake news detection: A survey. In *2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR)* (pp. 436-441). IEEE.

[9] Bondielli, A., & Marcelloni, F. (2019). A survey on fake news and rumour detection techniques. *Information Sciences*, *497*, 38-55.

[10] Elhadad, M. K., Li, K. F., & Gebali, F. (2019, August). Fake news detection on social media: a systematic survey. In *2019 IEEE Pacific Rim Conference on Communications, Computers and Signal Processing (PACRIM)* (pp. 1-8). IEEE.

[11] Lillie, A. E., & Middelboe, E. R. (2019). Fake news detection using stance classification: A survey. *arXiv preprint arXiv:1907.00181*.

[12] Granik, M., & Mesyura, V. (2017, May). Fake news detection using naive Bayes classifier. In *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)* (pp. 900-903). IEEE.

[13] Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, *34*(2), 76-81.

[14] Ahmed, H., Traore, I., & Saad, S. (2017, October). Detection of online fake news using n-gram analysis and machine learning techniques. In *International conference on intelligent, secure, and dependable systems in distributed and cloud environments* (pp. 127-138). Springer, Cham.

[15] Rădescu, R. (2020, July). AUTOMATIC ADAPTIVE SYSTEM FOR FILTERING FAKE NEWS BY METHODS BASED ON ARTIFICIAL INTELLIGENCE USING MACHINE LEARNING. In *Proceedings of EDULEARN20 Conference* (Vol. 6, p. 7th).

[16] Rubin, V. L., Chen, Y., & Conroy, N. K. (2015). Deception detection for news: three types of fakes. *Proceedings of the Association for Information Science and Technology*, *52*(1), 1-4.

[17] Ruchansky, N., Seo, S., & Liu, Y. (2017, November). Csi: A hybrid deep model for fake news detection. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management* (pp. 797-806).

[18] Tacchini, E., Ballarin, G., Della Vedova, M. L., Moret, S., & de Alfaro, L. (2017). Some like it hoax: Automated fake news detection in social networks. *arXiv preprint arXiv:1704.07506*.

[19] Pérez-Rosas, V., Kleinberg, B., Lefevre, A., & Mihalcea, R. (2017). Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.

[20] Saikh, T., De, A., Ekbal, A., & Bhattacharyya, P. (2020). A deep learning approach for automatic detection of fake news. *arXiv preprint arXiv:2005.04938*.

[21] Potthast, M., Kiesel, J., Reinartz, K., Bevendorff, J., & Stein, B. (2017). A stylometric inquiry into hyperpartisan and fake news. *arXiv preprint arXiv:1702.05638*.

[22] Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A., Choi, G. S., & On, B. W. (2020). Fake news stance detection using deep learning architecture (CNN-LSTM). *IEEE Access*, *8*, 156695-156706.

[23] D. Pomerleau and Rao. (2017). *Fake News Challenge Dataset*. Accessed: Oct. 29, 2019. [Online]. Available: http://www.fakenewschallenge.org/.

[24] S. Deegalla and H. Bostrom, ``Reducing high-dimensional data by principal component analysis vs. Random projection for nearest neighbor classification,'' in *Proc. 5th Int. Conf. Mach. Learn. Appl. (ICMLA)*, Dec. 2006, pp. 245_250.

[25] D. Hou, J. Zhang, Z. Yang, S. Liu, P. Huang, and G. Zhang, ``Distribution water quality anomaly detection from UV optical sensor monitoring data by integrating principal component analysis with chisquare distribution,'' *Opt. Express*, vol. 23, no. 13, p. 17487, Jun. 2015, doi: 10.1364/OE.23.017487.

[26] Bai, Q. (2010). Analysis of particle swarm optimization algorithm. *Computer and information science*, *3*(1), 180.

[27] Lingras, P., & Jensen, R. (2007, July). Survey of rough and fuzzy hybridization. In *2007 IEEE International Fuzzy Systems Conference* (pp. 1-6). IEEE.

[28] Ren, G., Cao, Y., Wen, S., Huang, T., & Zeng, Z. (2018). A modified Elman neural network with a new learning rate scheme. *Neurocomputing*, *286*, 11-18.

[29] Song, M., & Pedrycz, W. (2013). Granular neural networks: concepts and development schemes. *IEEE transactions on neural networks and learning systems*, *24*(4), 542-553.

[30] Pedrycz, W., & Vukovich, G. (2001). Granular neural networks. *Neurocomputing*, *36*(1-4), 205-224.

[31] Kobayashi, K., & Komaki, F. (2006). Information criteria for support vector machines. *IEEE transactions on neural networks*, *17*(3), 571-577.

[32] Chung, K. M., Kao, W. C., Sun, C. L., Wang, L. L., & Lin, C. J. (2003). Radius margin bounds for support vector machines with the RBF kernel. *Neural computation*, *15*(11), 2643-2681.