

## **A systemic literature overview of Fake News Challenge (FNC-1) dataset and its use in fake news detection schemes**

Zainab A. Jawad \*

*Department of Computer Science*

*Faculty of Computer Science and Mathematics*

*University of Kufa*

*Najaf*

*Iraq*

Ahmed J. Obaid <sup>†</sup>

*Department of Computer Science*

*Faculty of Computer Science and Mathematics*

*University of Kufa*

*Najaf*

*Iraq*

*and*

*National University of Science and Technology*

*Thi-Qar*

*Iraq*

---

### **Abstract**

Communication among people has been significantly enhanced by the recent advancements in social media. Social media platforms have enabled users to share information, connect with others, and stay informed about the latest events. However, social media can be a two-edged sword as it can be misused to spread unreliable information, such as fake news, which may be intended to deceive people, as witnessed in the 2016 U.S. presidential elections. Consequently, a considerable number of researchers are working towards identifying patterns and traits that fake news may exhibit to develop automated detection systems, given the gravity of this issue. However, fake news datasets have been created and made available to help devise such systems. In this study, we have described

---

\* E-mail: [Zainaba.alhasani@student.uokufa.edu.iq](mailto:Zainaba.alhasani@student.uokufa.edu.iq) (Corresponding Author)

<sup>†</sup> E-mail: [ahmedj.aljanaby@uokufa.edu.iq](mailto:ahmedj.aljanaby@uokufa.edu.iq)

the FNC-1 dataset and given an overview of the competitive attempts to build a fake news detection system using the FNC-1 dataset. Besides, it highlights the strengths and weaknesses of each of these works.

---

**Subject Classification:** *Primary 93A30, Secondary 49K15.*

**Keywords:** *Fake news, FNC-1 dataset, Disinformation detection, Natural language processing, Artificial intelligence.*

## 1. Introduction

Nowadays, the advancements in digital media have transformed the way of information exchange, where this news can be created and published online faster than traditional news sources such as newspapers and television [1]. Unfortunately, this innovation can be utilized to distribute false information. Fake news consists of fake tales disguised as actual news to gain readership, internet click revenue, and online sharing. This news can be regarded as one of the greatest threats to democracy, journalism, and freedom of expression. Researchers have found that humans are illogical and unable to differentiate between truth and falsehood when overloaded with misleading information [2]. The impact of fake news can be witnessed clearly in the “Brexit” referendum and the 2016 U.S. presidential elections [3]. As a result, public trust in media content has been undermined. The conventional approach for this problem is to build a fake news detection system. For this purpose, several computational techniques are available. These techniques can classify the article into fake and real depending on their textual content [4]. Another type of technique implies using fact-checking websites which means it maintains a list of websites that are classified as fake [5]. However, these solutions are impractical as it needs human efforts to identify fake websites or fake articles. Additionally, because fact-checking websites are solely focused on specific areas, they are not equipped to detect false news pieces across different sectors, such as entertainment, sports, and technology.

Consequently, recent works attempted to extract linguistic features from textual articles and use these features with different machine learning algorithms. The accuracy of which a system can identify fake news is one measure of that system’s efficacy. For this purpose, researchers have collected a set of datasets that can be used for evaluating fake news detection systems, verifying that the headline and body text match, identifying mismatched sentences in the body text, and determining the spread of fake news via social media [6].

This paper provides an overview of the several competing attempts to construct a fake news detection system using the FNC-1 dataset. It also provides a description of the FNC-1 dataset. Besides, it highlights the strengths and weaknesses of each of these works.

This paper is classified as follows: Section 2 describes the FNC-1 dataset, and section 3 presents the previous attempts that must be created to detect fake news using the FNC-1 dataset and discusses these works in a different characteristic. Finally, section 4 presented the conclusion.

2. Description of the FNC-1 Dataset

FNC-1 is focused on exploiting artificial intelligence technologies, especially machine learning and natural language processing, to tackle fake news [7]. This dataset contains 75,385 instances of different types of news, which are divided into training, testing, and competitive data (see Figure 1). These instances were labeled with agree, disagree, discuss, or unrelated.

Figure 1 demonstrates the main parts of the FNC-1 dataset, and the number of instances in each part.

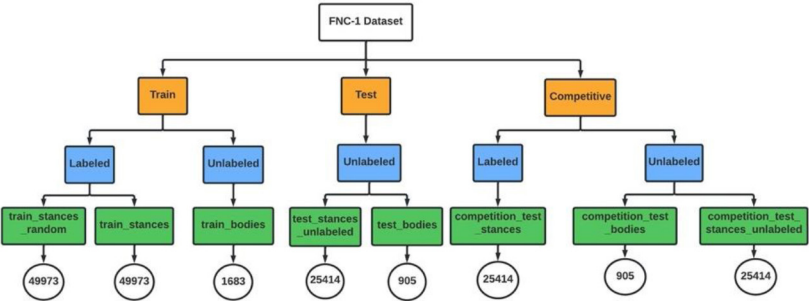


Figure 1  
The Classification of the FNC-1 Dataset Parts

This dataset contains two independent features and one feature which can be used for labeling purposes.

Input

- Both a headline and a body text - either from the same or two different news articles.

**Table 1**  
**Comparison Between Parts of FNC-1 Dataset**

| FNC-1 dataset parts                    | Size     | No. of word | Preposi-tions | Pro-nouns | Definite and in-definite articles | Auxilia-ry verbs |
|----------------------------------------|----------|-------------|---------------|-----------|-----------------------------------|------------------|
| Train-body                             | 3.58 MB  | 620911      | 94322         | 125740    | 69045                             | 48993            |
| Train-stances                          | 4.06 MB  | 555977      | 75131         | 48051     | 35369                             | 38394            |
| Train-stances.<br>random               | 4.01 MB  | 555977      | 75131         | 48051     | 35369                             | 38394            |
| Test-bodies                            | 1.95 MB  | 337064      | 51277         | 70653     | 29704                             | 25100            |
| Test-stances-<br>unlabeled             | 1.85 MB  | 282917      | 32533         | 38929     | 7678                              | 14804            |
| Competition-test-<br>bodies            | 1.95 MB  | 337064      | 51277         | 70653     | 29704                             | 25100            |
| Competition-test-<br>stances           | 2.08 MB  | 282917      | 32533         | 38929     | 7678                              | 14804            |
| Competition-test-<br>stances-unlabeled | 1.85 MB  | 282917      | 32533         | 38929     | 7678                              | 14804            |
| Total                                  | 21.33 MB | 3255744     | 444737        | 924672    | 222225                            | 220393           |

### Output

- There are four categories of stance in the body text according to the claim made in the headline:
  - a. Agrees: The body text and the headline have been agreed.
  - b. Disagrees: The body text and the headline have disagreed.
  - c. Discusses: the headlines discuss the body text.
  - d. Unrelated: The body text and the headline explain a different subject.

Table 1 summarized the size, number of words, prepositions, pronouns, definite and indefinite articles, and auxiliary verbs of each part of the FNC-1 dataset.

Table 2 represented the FNC-1 dataset's data, where the rows  $I = \{I_1, I_2, \dots, I_n\}$  represent the instances of the dataset, while the columns represent the weight of each feature  $W = \{W(t)_1, W(t)_2, \dots, W(t)_n\}$  represent the features extracted from the dataset after excluding the prepositions, pronouns, definite and indefinite articles, and auxiliary verbs that are summarized in Table 1.

**Table 2**  
**Representation of FNC-1 Dataset’s Data**

| Data  | $W(t)_1$    | $W(t)_2$    | $W(t)_3$    | $W(t)_4$    | ..... | $W(t)_n$    |
|-------|-------------|-------------|-------------|-------------|-------|-------------|
| $I_1$ | $W(t)_1I_1$ | $W(t)_2I_1$ | $W(t)_3I_1$ | $W(t)_4I_1$ | ..... | $W(t)_nI_1$ |
| $I_2$ | $W(t)_1I_2$ | $W(t)_2I_2$ | $W(t)_3I_2$ | $W(t)_4I_2$ | ..... | $W(t)_nI_2$ |
| $I_3$ | $W(t)_1I_3$ | $W(t)_2I_3$ | $W(t)_3I_3$ | $W(t)_4I_3$ | ..... | $W(t)_nI_3$ |
| $I_4$ | $W(t)_1I_4$ | $W(t)_2I_4$ | $W(t)_3I_4$ | $W(t)_4I_4$ | ..... | $W(t)_nI_4$ |
| ..... | .....       | .....       | .....       | .....       | ..... | .....       |
| $I_m$ | $W(t)_1I_m$ | $W(t)_2I_m$ | $W(t)_3I_m$ | $W(t)_4I_m$ | ..... | $W(t)_nI_m$ |

Furthermore, we have used the term frequency identification method to find the frequency of common words in train bodies data.

Figure 2, represents the instances in Table 3 with curves with different colors, where the figure shows how these words are frequent in fake news.

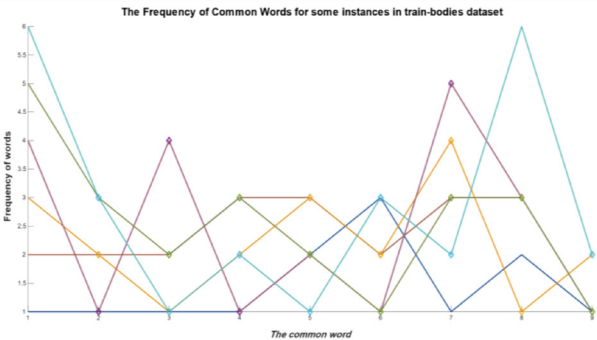


**Figure 2**  
**Representation of Data After Choosing Some Instances from Train\_bodies Dataset.**

**3. An Overview of the FNC-1 Dataset**

In this section, we presents an review of the previous attempts to build a fake news detection system with the FNC-1 dataset. Specially we discuss the approaches that are based on using natural language processing methods with deep learning techniques and machine learning algorithms.

An interesting attempt was made by Riedel et. al. 2017 [8], where they suggested using the UCL Machine Reading’s (UCLMR) system. The

proposed system implies using deep learning techniques with lexical and similarity features. Their system adds the benefits of simplicity as opposed to most sophisticated neural network algorithms. Although the implementation was basic, the authors achieved an FNC-1 score of 81.72%. The researchers believed that their work would serve as a platform for competing attempts on the FNC-1 dataset, although the precision that has been obtained still requires further improvements.

Another team, Gaurav Bhatt and his co-workers 2017 [9], presented a model based on statistical, neural, and feature engineering heuristics methods with deep learning techniques. The proposed model was simple. Besides, it achieved up to 83.08% accuracy, where the accuracy was high in all cases except for the disagree instances.

In this article [10], Benjamin Schiller et. al. 2018 applied a novel approach using memory networks with Convolutional Neural Networks (CNN) and Long short-term memory (LSTM) techniques for stance detection. The model was further extended using the similarity-based matrix and a stance filtering component. The proposed expansion not only contributes to a considerable improvement in performance gains but also brings the benefit of increased competitiveness. The results demonstrated that the approach is capable of remembering past information, and the score that it achieved was 81.23 percent. Nevertheless, the model that has been proposed cannot deal with a collection of instances; rather, it only deals with a single instant at a time.

On the other hand, Andreas Hanselowski et. al. 2018 [11] proposed a new F1-based metric and suggested a novel feature-rich stacked LSTM approach. The proposed system significantly improves the system performance in detecting the minority classes. As a result, they were able to achieve a score of 80.2%. However, the system doesn't perform well under challenging cases, requiring more advanced machine learning methods with further semantic understanding.

Borges et. al. 2018 [12] introduced a novel approach combining Deep Neural Network (DNN) approaches with string external similarity features. The proposed model slightly enhances the results compared to previous works, where they obtained a score of 82.23%. However, researchers stated that advanced techniques for sentence modeling are required. Besides, the recurrent neural network (RNN) method must be used for text encoding instead of feed-forward approaches.

In [13] Benjamin Schiller et. al. 2020, introduced a system that combines Transfer Learning (TL) and Multi Dataset Learning (MDL), enabling the addition and evaluation of low-resource experiments and adversarial attack sets. The researcher accomplished 88.82%, but the proposed system

needs to be combined with sampling methods and methods that cope with biased data to reduce training data and solve the overfitting problem in the MDL approach.

Ronald Denaux et. al. 2020 in [14] introduced a system based on the core concept of Credibility Reviews (CRs) for misinformation detection to build networks of distributed bots. The acred+ system achieved 71.6% accuracy, but the team did not solve all the problems needed to achieve high accuracy. However, they need to improve their system by testing more datasets and fine-tuning.

In this paper [15], Thomas Felber 2021 attempted to apply several linguistic features and machine learning methods to the COVID-19 dataset for stance detection. Results proved the efficiency of the proposed approach. The Support Vector Machines (SVMs) classifier achieves the best among the other classifiers, giving an F1 score of 95.19%. However, the performance is still not enough and needs more improvement.

On the other hand, Alexander Michael 2021 in [16] Presented the canonical machine learning approach, sentiment analysis, and stance detection techniques for an automated fake news detection system. According to the FNC-1 competition test set, the algorithm achieves 90.8% accuracy. However, it showed some dropped in performance compared to the validation dataset when tested on a specific task dataset, and this is due to the difference between data sources.

**Table 3**

**Summarized the competitive attempts on the FNC-1 dataset during the past five years, where it highlighted these works' different aspects.**

| Ar-<br>ticles | Year | Dataset | Technique                                              | Strength                                                                                                                   | Weakness                                                      | Result           |
|---------------|------|---------|--------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------|------------------|
| [8]           | 2017 | FNC-1   | ULCMR                                                  | The purpose approaches were simple                                                                                         | Result still need further improvement                         | F1 score (81.72) |
| [9]           | 2017 | FNC-1   | Statistical, neural and feature engineering heuristics | Got 7% accuracy higher in unrelated categories, but Skip-thought achieved accuracy higher in agree and discuss categories. | Accuracy was high in all labels except for the disagree ones. | F1 score (83.08) |

*Contd...*

|      |      |                                                                                 |                                                                             |                                                                                                      |                                                                                                                               |                  |
|------|------|---------------------------------------------------------------------------------|-----------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------|------------------|
| [10] | 2018 | FNC-1                                                                           | CNN and LSTM are used to improve the performance of a novel memory network. | Extension offers sizable performance gains and models are designed to remember previous information. | Not suited for the complex task                                                                                               | WA. (81.23)      |
| [11] | 2018 | FNC-1                                                                           | A new F1-based metric                                                       | Rich stacked LSTM approach                                                                           | Doesn't perform well in a difficult case                                                                                      | F1m (80.2)       |
| [12] | 2018 | FNC-1                                                                           | Combining DNN approaches with string similarity features                    | Results given the external features essential to a good performance.                                 | Don't present advanced techniques for modeling sentences and matching representation.                                         | WA. (82.23)      |
| [13] | 2020 | ARC, FNC-1, iac1, IBMCS, Snopes, SCD, PERSPECTRUM, semeval2016t6, semeval2019t7 | Combines TL and MDL                                                         | Enabling the addition and evaluation of low-resource experiments and adversarial attack sets.        | The prepose system needs to combine with based data to teach a task and reduce training data to a sample.                     | F1m (88.82)      |
| [14] | 2020 | Clef'18 CheckThat! , FakeNewsNet, coinform250                                   | Based on the core concept of Credibility Reviews (CRs).                     | Build networks of distributed bots.                                                                  | Cannot solve all the issues that need to be solved and need to improve their system by testing more datasets and fine-tuning. | Acc. (71.6)      |
| [15] | 2021 | Covid-19 fake news                                                              | Applying several linguistic features and machine learning methods.          | Got the best performance system on linear SVM.                                                       | The performance isn't high enough and needs more improvement.                                                                 | F1 score (95.19) |

*Contd...*

|      |      |                          |                                                                                      |                                                                                                                                               |                                                                                                                                                      |             |
|------|------|--------------------------|--------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|
| [16] | 2021 | FNC-1, ARC, PERSPEC-TRUM | The canonical machine learning techniques, sentiment analysis, and stance detection. | Could be impactful for how analysts use machine learning to identify disinformation and thus provide reasonable motivation for further study. | Proposed system showed some dropped in performance compared to the validation dataset but was not unreasonable given the difference in data sources. | Acc. (90.8) |
|------|------|--------------------------|--------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|-------------|

In summary, we noted that most of the studies could not detect fake news efficiently in complex cases requiring attempting more advanced machine learning approaches.

#### 4. Conclusion

Now, fake news is regarded as one of the greatest threats to democracy, journalism, and freedom of expression, especially on social media platforms, where news spreads fast at a low cost. Consequently, the global problem of fake news has grown. In order to create fake news detection systems, researchers used a variety of NLP, ML, and DL methodologies. The competitive studies using the FNC-1 dataset were the main emphasis of this research. Also, it emphasized both the virtues and defects of these works. The previous works have been studied, and it has been determined that the suggested methodologies fall short of creating an effective system for finding false news. Future work will require the development of new technology to accurately identify bogus news.

#### References

- [1] Shu, K., et. al., Fake news detection on social media: A data mining perspective. *ACM SIGKDD Explorations Newsletter*, 2017. 19(1): p. 22-36.
- [2] Rubin, V.L., On deception and deception detection: Content analysis of computer-mediated stated beliefs. *Proceedings of the American society for information science and technology*, 2010. 47(1): p. 1-10.
- [3] Pogue, D., How to Stamp Out Fake News. *Scientific American*, 2017. 316(2): p. 24-24.

- [4] Conroy, N.K., V.L. Rubin, and Y. Chen, Automatic deception detection: Methods for finding fake news. *Proceedings of the association for information science and technology*, 2015. 52(1): p. 1-4.
- [5] Asr, F. and M. Taboada, Misinfotext: a collection of news articles, with false and true labels. 2019.
- [6] Zhou, X. and R. Zafarani, A survey of fake news: Fundamental theories, detection methods, and opportunities. *ACM Computing Surveys (CSUR)*, 2020. 53(5): p. 1-40.
- [7] Fake News Challenge Stage 1 (FNC-I): Stance Detection. Available from: <http://www.fakenewschallenge.org>.
- [8] Riedel, B., et. al., A simple but tough-to-beat baseline for the Fake News Challenge stance detection task. arXiv preprint arXiv:1707.03264, 2017.
- [9] Bhatt, G., et. al., On the benefit of combining neural, statistical and external features for fake news identification. arXiv preprint arXiv:1712.03935, 2017.
- [10] Mohtarami, M., et. al., Automatic stance detection using end-to-end memory networks. arXiv preprint arXiv:1804.07581, 2018.
- [11] Hanselowski, A., et. al., A retrospective analysis of the fake news challenge stance detection task. arXiv preprint arXiv:1806.05180, 2018.
- [12] Borges, L., B. Martins, and P. Calado, Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality (JDIQ)*, 2019. 11(3): p. 1-26.
- [13] Schiller, B., J. Daxenberger, and I. Gurevych, Stance detection benchmark: How robust is your stance detection? *KI-Künstliche Intelligenz*, 2021. 35(3): p. 329-341.
- [14] Denaux, R. and J.M. Gomez-Perez. Linked credibility reviews for explainable misinformation detection in International Semantic Web Conference. 2020. Springer.
- [15] Felber, T., Constraint 2021: Machine learning models for COVID-19 fake news detection shared task. arXiv preprint arXiv:2101.03717, 2021.
- [16] Daniel, A.M., Human-in-the-Loop Disinformation Detection: Stance, Sentiment, or Something Else? arXiv preprint arXiv:2111.05139, 2021.
- [17] Shubham Sharma & Ahmed J. Obaid (2020) Mathematical modelling, analysis and design of fuzzy logic controller for the control of ventilation systems using MATLAB fuzzy logic toolbox, *Journal of Interdisciplinary Mathematics*, 23:4, 843-849, DOI: 10.1080/09720502.2020.1727611.

*Received March, 2022*

*Revised April, 2022*