ORIGINAL

Hybrid Feature Extraction and Capsule Neural Network Model for Fake News Detection

Modelo híbrido de extracción de características y red neuronal de cápsulas para la detección de noticias falsas

R. Uma Maheswari[1] ✉, N. Sudha[1] ✉

[1]Bishop Appasamy College of arts and science, Department of Computer Science. Coimbatore, Tamil Nadu, India.

ABSTRACT

The introduction and widespread use of social media has altered how information is generated and disseminated, along with the expansion of the Internet. Through social media, information is now more quickly, cheaply, and easily available. Particularly harmful content includes misinformation propagated by social media users, such as false news. Users find it simple to post comments and false information on social networks. Realising the difference between authentic and false news is the biggest obstacle. The current study's aim of identifying bogus news involved the deployment of a capsule neural network. However, with time, this technique as a whole learns how to report user accuracy. This paper offers a three-step strategy for spotting bogus news on social networks as a solution to this issue. Pre-processing is executed initially to transform unstrsuctrured data into a structured form. The second part of the project brought the HFEM (Combined Feature Extraction Model), which also revealed new relationships between themes, authors, and articles as well as undiscovered features of false news. based on a collection of traits that were explicitly and implicitly collected from text. This study creates a capsule neural network model in the third stage to concurrently understand how creators, subjects, and articles are presented. This work uses four performance metrics in evaluations of the suggested classification algorithm using on existing public data sets.

Keywords: Fake News; Preprocessing; Hybrid Feature Extractor; Capsule Neural Network; Fake News Detection.

RESUMEN

La introducción y el uso generalizado de los medios sociales han alterado el modo en que se genera y difunde la información, junto con la expansión de Internet. A través de los medios sociales, la información es ahora más rápida, barata y fácil de conseguir. Entre los contenidos especialmente nocivos se encuentra la desinformación propagada por los usuarios de las redes sociales, como las noticias falsas. A los usuarios les resulta sencillo publicar comentarios e información falsa en las redes sociales. Comprender la diferencia entre noticias auténticas y falsas es el mayor obstáculo. El objetivo del presente estudio de identificar noticias falsas implicaba el despliegue de una red neuronal de cápsulas. Sin embargo, con el tiempo, esta técnica en su conjunto aprende a informar con precisión al usuario. Este trabajo ofrece una estrategia de tres pasos para detectar noticias falsas en las redes sociales como solución a este problema. En primer lugar, se realiza un preprocesamiento para transformar los datos no estructurados en una forma estructurada. La segunda parte del proyecto aporta el HFEM (Modelo Combinado de Extracción de Rasgos), que también revela nuevas relaciones entre temas, autores y artículos, así como rasgos no descubiertos de las noticias falsas. basándose en una colección de rasgos recogidos explícita e implícitamente del texto. Este estudio crea un modelo de red neuronal en cápsula en la tercera etapa para comprender simultáneamente cómo se presentan los creadores, los temas y los artículos. Este trabajo utiliza cuatro métricas de rendimiento en las

evaluaciones del algoritmo de clasificación sugerido utilizando en conjuntos de datos públicos existentes.

Palabras clave: Noticias Falsas; Preprocesamiento; Extractor Híbrido de Características; Red Neuronal de Cápsulas; Detección de Noticias Falsas.

INTRODUCTION

More and more individuals are choosing to find and consume information on social networks rather of traditional news outlets as we spend more and more time interacting online through social media platforms system.[1] One of the biggest challenges to media, democracy, and the economy today is fake news. This has eroded public confidence in governments, and it hasn't yet been realised how it can affect contentious subjects. Fake news has an influence on the stock market and leads to significant acquisitions, so our economy is not exempt.[2] Interdisciplinary research has focused on human vulnerability to fake news, which can provide helpful signals or make it more difficult to identify fake news. For example, advances in forensic psychology, such as the Undeutsch hypothesis, have emphasized stylistic differences between truth, deception, and misinformation.[3] Similarly, multidisciplinary studies have investigated the motives behind the spread of fake news, contending that the distinction between good users and bad users is becoming increasingly blurred. Regular users can frequently and unknowingly engage in fake news activities, for example because of their social identity or social identity.

Fake news, which is disseminated as facts or actual occurrences, is becoming more and more prevalent in social networks and the media. False news is presented as an evasive and theatrical idea that fails to identify the information's source, context, or URL [4] spreads rumours, ideas, or conspiracies by often using ambiguous formulations like "according to an expert," "research shows," or "a journalist says." Offensive headlines, sometimes capitalised or including capital letters, are a common feature of fake news. The following list provides an imperfect classification of false news:

- Misinformation based on genuine information that has been taken out of context or presented sensationally.
- Partisan news based on interpretations of real events that are manufactured to fit a political or commercial objective. News that only contains invented information.
- Clickbait news [5] draws readers in with surprising, sensational, or convincing headlines but frequently lacks the supporting facts that the headlines' promises call for. count and improve the site's position, luring users to purchase advertisements there).

The nature of fake news makes their detections a complex issue. Detection algorithms use a variety of information from news (such as headlines, content, and publishers) and social networks (such as comments, viral links, and broadcasters). Each type of information can be in the form of text, multimedia, network, etc. and involves many different methods and resources. Accordingly, researchers, practitioners and business leaders must place a high priority on understanding and detecting fake news to prevent it from harming democracy, journalism and the economy.[6,7] Let's go through the theoretical and technological components of false news identification below to help future research in this area:

1. Promote multidisciplinary study into fake news, since this will enable the creation of methodologies for detecting false news that are logical and explicable, as well as qualitative and quantitative research.
2. A comprehensive framework and strategy is needed to thoroughly analyze and identify fake news. Such strategies attract and bring together fake news seekers who are knowledge and technology experts in related fields.
3. To highlight future research paths and goals, outstanding problems and challenges linked to fake news research need to be explained.

The ability of certain human moderators to check and validate facts has traditionally been used to determine if information is true, however owing to the amount and diversity of information, this approach has limitations. The expansion of the Internet has resulted in an abundance of information that has to be analysed nowadays.[8] There is a definite trend towards the use of software programmes that are based on artificial intelligence, particularly the branches of AI known as data mining and machine learning. Humans are not as adept at spotting lies in text as one might believe, which is another argument in favour of switching to automatic fake news detection.

Several previous studies explain fake news detection in many aspects such as knowledge, style, transmission and reliability.[9] From a knowledge point of view, fake news detection is a "comparison" between relational information extracted from news to be verified and the knowledge base representing the truth/fact. true in fact. The goal of style-based fake news detection is to identify and quantify the difference in writing style between fake and true news.[10] Spread-based fake news identification utilizes information provided during

transmission. Finally, trust-based fake news detection evaluates the credibility of headlines (e.g. using click-bait recognition), publishers (i.e. source sites), receiving reviewers (for example, using opinion spam detection) and users to indirectly detect fake news. Each perspective has its own set of tools, datasets, and detection processes in data mining, machine learning, natural language processing, information retrieval, and social research.[11] Different perspectives can be combined into a single framework in analyses by examining time of creation and publication.

This detection approach seeks to forecast deliberate disinformation based on the study of verified accurate and incorrect information. Therefore, the accessibility of top-notch training data is a crucial concern. This research project focuses on enhancing detection accuracy by incorporating feature selection and capsule neural network modelling in order to address this issue. The remaining study is structured to include a number of cutting-edge techniques for identifying false news in Part 2. The suggested technique for detecting bogus news is briefly explained in Section 3. The findings and discussion are illustrated in Section 4. Section 5 covers recommendations and ongoing research.

Literature review

In this section, we will look at many suggested strategies for detecting fake news in different types of data. Previous and related studies on identifying fake news have been reviewed.

Granik et al.[12] developed a simple method to detect fake news using a simple Bayesian classifier. This technique is implemented as a software system and tested using the Facebook news post template. In terms of the test set, this means a classification accuracy of about 74 %, which is quite a good result considering the simplicity of the model. Several methods, which are also covered in the article, can be used to enhance these outcomes. The findings produced demonstrate that the issue of identifying false news may be resolved by using AI methods.

Reis et al.[13] proposed a new feature set for automated false news identification,and evaluated the predicted effectiveness of existing techniques and features. Interesting conclusions on the value and significance of false news detection are shown by the data. Finally, discuss the advantages and disadvantages of practical implementations of fake news detection algorithms.

Ahmad et al.[14] proposed a fake news detection technique based on n-gram analysis and machine learning (ML). Review and compare six separate machine classification procedures and two separate feature extraction methods in this section. Experimental evaluation works best with 92 % accuracy when TF-IDF (Term Frequency Inverted Document Frequency) is used as feature extraction strategy and LSVM (Linear Support Vector Machine) as an analysis tool.

Rădescu et al.[15] demonstrate machine learning (ML) applied to an automated fix for the false news detection issue. Two straightforward but nonetheless remarkably potent strategies are described in detail. While the second section is based on probability theory and the Naive Bayes approach, the first section is based on keyword research in the text. His unique contribution to this study was to establish and evaluate the effectiveness of these two news rating methodologies. The functional tests show that the aforementioned approaches are straightforward but extremely efficient. Therefore, both approaches can differentiate articles into different degrees of validity based just on titles: genuineness, low, medium, or high. The outcomes of the two procedures used have produced extremely promising results. If a big and high-quality training database is available and can be produced by extending the training data files being utilised, the performance of the Naive Bayes approach can be significantly improved.

Rubin et al.[16] developed three types of fake news, each of which collates real articles and examines their advantages and disadvantages as a source of text and sample analysis. forecast. Filtering, organizing, and validating online content is essential in LIS (library and information science) as the barrier between traditional and online information is fading.

Ruchansky et al.[17] have suggested a model that incorporated three features for more precise and automated prediction, Integrate both parties' behaviour, including that of users and publications, as well as the collective behaviour of users who propagate false information. Three features drive the suggestion of the CSI paradigm, which consists of three modules: capture, scoring, and integration. The first module is text-based, responsive, and employs recurrent neural networks to track users' temporal activity patterns while reading a particular article. In order to determine if an item is fraudulent or not, the third module integrates the second module, which learns the characteristics of the source based on user behaviour. Real-world data from experimental research shows that CSI is more accurate than current models and extracts meaningful latent representations of users and items.

Tacchini et al.[18] provided two categorization approaches, one based on logistic regression and the other on a community-developed new version of the Boolean algorithm. Even though the training set comprised fewer than 1 % of the posts, higher than 99 % classification accuracy was still attained on a dataset of 15 500 Facebook posts and 909 236 individuals. These strategies still work when limiting attention to those who appreciate

prank postings and don't prank further demonstrates their effectiveness. These findings imply that charting the patterns of information diffusion might be a helpful tool in automated hoax detection systems.

Pérez-Rosas et al.[19] presented two new datasets that included seven distinct information domains for the purpose of identifying false news. Here, the procedure for gathering, annotating, and validating the data is detailed in detail, and some exploratory analyses are offered for spotting linguistic variations in honest and false material. The second step is to carry out a number of learning experiments to develop a reliable false news detector. Additionally, it compares how bogus news is identified manually and automatically.

Two effective deep learning based models have been published by Saikh et al.[20] to address the issue of identifying false news in online news material from various domains. Use the two freshly published FakeNews AMT and Celebrity datasets to evaluate these strategies for spotting false news. The suggested systems surpass today's most cutting-edge, manually constructed, feature-engineered systems by a sizeable margin of 3,08 % and 9,3 % for the two models, respectively. Perform a cross-domain study here to examine the applicability of our system across all domains and to take advantage of the available data sets for related tasks (i.e., a model trained on FakeNews AMT and tested on Celebrities and vice versa).

Potthast et al.[21] developed a new methodology to evaluate stylistic similarities between text kinds through unmasking, a meta-learning technique initially created for authorship authentication and demonstrated that left and right news had more in common with one another than with the mainstream. We further demonstrate that the orthodox tone of hyperpartisan news (F1=0,78) and the irony of both (F1=0,81) may be used to identify it. As expected, style-based false news identification is ineffective (F1=0,46). The first result, however, is crucial for choosing which false news detectors to use.

Goldani et al.[22] created a capsule neural network with the purpose of identifying bogus news. To accommodate news stories of various durations, use several embed formats below. Short news items employ static word embedding, but medium- and long-length articles use non-static word embedding, which allows for incremental training and changes during the training process. Apply various N-gram values for feature extraction as well. Two recent, well-known data sets in the area, ISOT and LIAR, are used to assess the suggested models. The findings beat the top techniques on the ISOT by 7,8 %, the validation set by 3,1 %, and the LIAR data test suite by 1 %, respectively.

According to the research above, this strategy offers certain benefits and drawbacks. The superiority of deep neural models over traditional ML methods has led to the development of increasingly sophisticated deep neural models in order to produce the most cutting-edge outcomes possible. This paper suggests a number of cutting-edge deep learning features and models for spotting false news, which are explained in more detail in the next section.

Suggested approach

The suggested approach for false news identification is covered in great length in this section. The first step is to pre-process the data collection by removing duplicate keywords or characters such stop words, numerals, etc. The false news dataset underwent feature extraction to condense the feature space. Finally, the classifier based on a capsule neural network is used to identify bogus news in the provided data set. This study focuses on false news and offers a three-step strategy for locating it on social media.

1. Pre-processing is used to change the unstructured data set into a structured data set in the initial phase of the procedure.
2. In the second stage, HFEM reveals the different links between articles, producers, and themes as well as the unidentified traits of false news. based on a collection of traits that were explicitly and implicitly collected from text.
3. This study creates a neural network model in the third stage to concurrently understand how creators, themes, and articles are presented. Use several embed formats here for news stories of various durations. Short news items employ static word embedding, but medium- to long-length articles use non-static word embedding, which allows for incremental training and changes during the training process.

The overall process of the suggested HFEM-CNN (Hybrid Feature Extraction model with Capsule Neural Network) is depicted in figure 1.
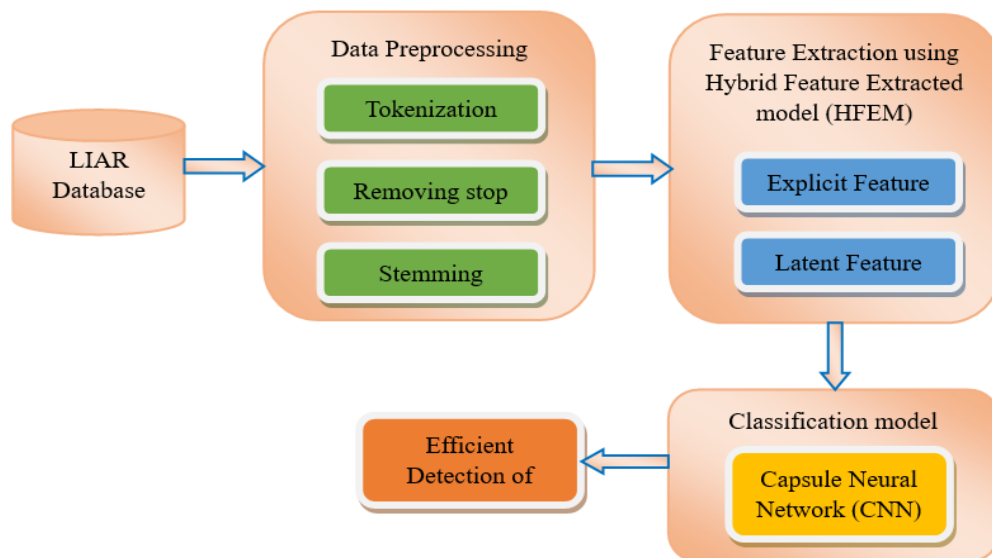
Figure 1. Overall Flow of the suggested HFEM-CNN

1. Pre-processing is used to change the unstructured data set into a structured data set in the initial phase of the procedure.
2. In the second stage, HFEM reveals the different links between articles, producers, and themes as well as the unidentified traits of false news. based on a collection of traits that were explicitly and implicitly collected from text.
3. This study creates a neural network model in the third stage to concurrently understand how creators, themes, and articles are presented. Use several embed formats here for news stories of various durations. Short news items employ static word embedding, but medium- to long-length articles use non-static word embedding, which allows for incremental training and changes during the training process.

The overall process of the suggested HFEM-CNN (Hybrid Feature Extraction model with Capsule Neural Network) is depicted in Figure 1.

Data Collection
The internet datasets utilised in this study are open source and totally free. The information comprises both false and accurate news from many areas. Real-world occurrences are accurately described in published, truthful news stories; yet, false information is presented on websites. On fact-checking websites like politifact.com and snopes.com, you may personally verify whether many of these articles' statements about the policy domain are true.

The "LIAR Fake News Dataset",[22] which includes both factual and fake stories pulled from the Internet, served as the basis for this study. The prominent news website reuters.com is where authentic articles are taken from, whereas politifact.com websites are where fraudulent items are taken from. The entire archive is filled with articles from many sectors, although they are mostly focused on political news.

Data Pre-processing
The first regular step before training is data preparation, and the ML algorithm's evaluation of the data is only useful if the data supports it. In order to provide the greatest results, it is crucial that the data be appropriately prepared and have important characteristics. As demonstrated in the paper, there are several phases involved in the data pretreatment process for ML computer vision systems, including normalising the picture input and shrinking the size.[23] This is done to eliminate certain insignificant details that may be used to identify between photos. In picture labelling, characteristics like darkness or brightness are useless. The text's authenticity can't always be determined by reading some passages, as well.

- Tokenization: The text is divided into tokens during the tokenization process, which also eliminates any punctuation from the text data. Terms with numerical content are filtered out using numerical filters. Text data may be converted to uppercase or lowercase using the uppercase converter. All words in this work have been changed to lowercase. Finally, in this stage, words with fewer than N characters are eliminated using an N-character filter.
- Stop-words removal: Despite being often used to finish sentences and link phrases, stop words

are not necessary. These are language-specific terms that don't communicate anything. Stop words include conjunctions, pronouns, and prepositions. In English, there are 400–500 stop words. Stop words include "a," "an," "about," "by," "but," "this," "done," "on," "above," "one," "after," "until," "again," "when," "where," "what," "all," "am," "and," "everything," "against," and so on.

- Stemming: Different grammatical forms of a word, such as adjectives, adverbs, nouns, verbs, etc., will be produced during this process. has been restored to its initial state. Finding basic forms of words with the same meaning but distinct forms is the goal of derivation. For instance, the word "connection" may be used to create the terms connect, connect, link, connect, and connect. Algorithm 1 provided the methods for data preparation.
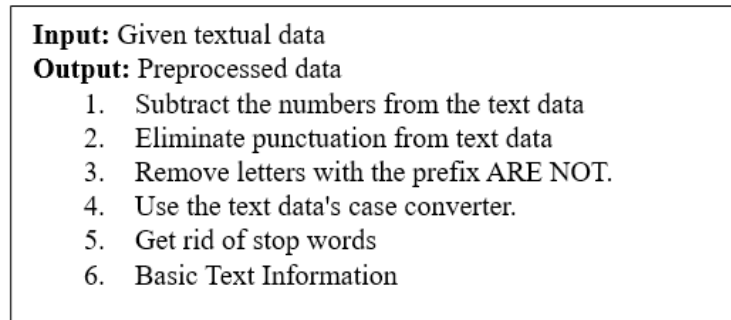
Input: Given textual data
Output: Preprocessed data
  1. Subtract the numbers from the text data
  2. Eliminate punctuation from text data
  3. Remove letters with the prefix ARE NOT.
  4. Use the text data's case converter.
  5. Get rid of stop words
  6. Basic Text Information

Figure 2. Algorithm: Preprocessing process

HFEM

As illustrated in figure 3, we will concentrate on feature learning from textual content information in this section, whereas the following section will employ relationships to build CNN.
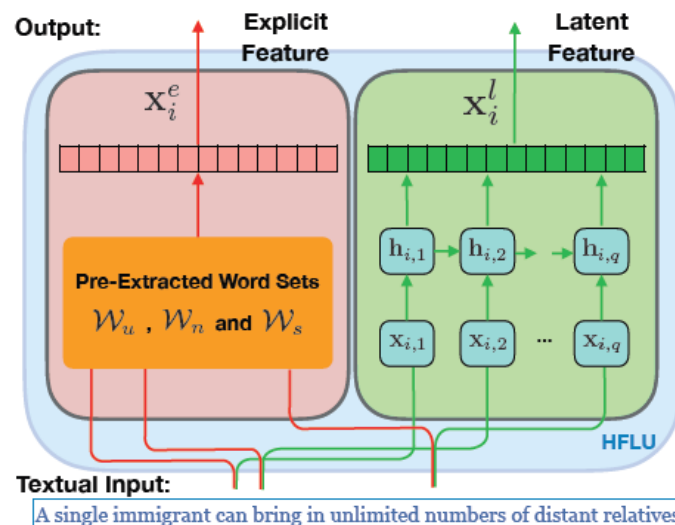

Figure 3. HFEM

1. Explicit Feature Extraction:

The linguistic content of false news can provide significant clues for determining their veracity. Let W stand for the whole vocabulary included in the LIER dataset; from W, one may also extract a particular collection of terms from articles, creator profiles, and literary data. theme version, it should be highlighted sets $W_n \subset W, W_u \subset W$ and $W_s \subset W$ respectively.

The retrieved words and their true/false classifications had the strongest association. Pre-extracted word sets W_n m can represent extracted feature vectors specifically for vector forms as well as article (comments) $X_{n,i}^e \in R^d$, where entries $x_{n,i}^e$ (k) denote counts of word appearance $W_k \in W_n$ in news articles n_i as shown in the left side of figure 3. Similarly extracted word sets W_u (and W_s), represent obtained explicit feature vectors for creators (news ids) u_j as $X_{u,j}^e \in R^d$ (and subjects (semantic meanings) s_l as $X_{n,i}^e \in R^d$).

2. Latent Feature Extraction:

There are subtle signals across articles, creators, and subjects, such as discrepancies, in addition to these

terms that are plainly evident on news story assets, creator biographies, and topic descriptions. It is possible to manage the consistency of information regarding article content and probable profile/description patterns. by using the latent characteristics. Based on this understanding, we suggest in this study that the capsule neural network model be used to extract an additional set of latent properties for articles, creators, and themes.

Formally, for an article $n_i \in N$, represents sequential word contents represented as vectors $(X_{i,1}, X_{i,2}, \ldots, X_{i,q})$ where q stands for articles' max lengths and zeros are padded when words lengths are lesser than q. This is done based on the article's original text content. Each $x_{i,k}$ feature vector corresponds to a word in the article. Based on the latent semantic analysis model, latent semantic characteristics are retrieved.
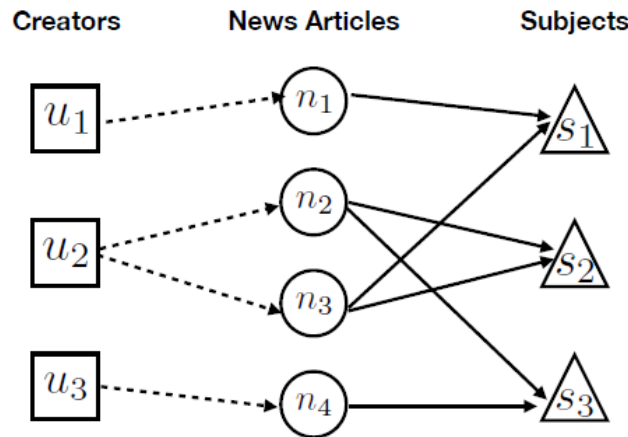

Figure 4. Relationships of Articles, Creators and Subjects

RNN model (with underlying neurons) is used for latent feature extraction, and includes 1 input, 1 hidden, and 1 merged layers as shown in right side of figure 4. Depending on the input vector $(X_{i,1}, X_{i,2}, \ldots, X_{i,q})$ from the input text string, the feature vectors can be represented in the hidden and output layers as follows:

$$\begin{cases} convolution\ layer: X^l_{n.i} = \sigma\left(\sum_{t=1}^{q} W_i h_{i,t}\right), \\ Capsule\ layer: h_{i,t} = GRU(h_{i,t-1}, x_{i,t}; W) \end{cases} \quad (1)$$

Based on components with similar architectures as shown in figure 4, latent feature vectors for news creators $u_j \in U$ (and subjects $s_l \in S$) can be extracted and denoted as vectors $X^l_{u,j}$ (and $X^l_{s,l}$). Formal feature representations of articles, creators, and themes can be described based on [24] and by adding explicit and implicit feature vectors, which will be supplied as input to the CNN model, which will be introduced in the next paragraph.

Capsule neural network model
In actuality, the author and topic of an article have a significant impact on its trustworthiness. The illustration in figure 4 shows the connection between subjects, authors, and publications. There is only one Creator per article, although any creator is allowed to write several articles. Each article may cover a variety of subjects, and several articles may focus on a single subject. We shall introduce the neural network model in the following in order to model the relationship between articles, creators, and themes. This article titled "Dynamic Routing Between Capsules" is where the capsule neural network was initially introduced.

This study showed that the IOST dataset's performance on substantially overlapping digits may be improved using the capsule network. A capsule is a neural network that attempts to make inverse graphics function in computer vision. A large number of capsules make up the capsule network, which attempts to anticipate the initialization parameters and the existence of a certain object at a specific place.[25] Equivalence, which seeks to keep specific information about an object's location and posture over the whole array, is one of the important characteristics of capsule arrays. Recently, The Capsule Network has drawn a lot of attention. By including the following elements into each source and destination node, this design intends to enhance CNNs and RNNs. The destination node may choose how many messages it can receive from various source nodes, and the source node can choose how many messages to forward to the destination nodes.

The textual features are stored in the capsules (groups of neurons) of the capsule network. For text

categorization, the first effort employs a network of capsules. In their study, they assessed the efficacy of the capsule network as a text categorization network for the first time. The n-gram convolution layer, a prominent feature extractor convolution layer, is included in their capsule network architecture. The second layer is a class known as the capsule representation layer that translates scalar-valued qualities to a capsule representation. A layer of convolutional capsules receives the output of these capsules. Each capsule in this layer is only linked to a small area of the bottom layer. The anterior layer's output is flattened and transferred via the transition capsule layer at the very end. Each output capsule is regarded as a distinct class for this class. Capsules can provide a sentence or document as a text classification vector thanks to a few specific characteristics. These characteristics include the expression of semantic meaning over a wide region and the display of incomplete entity properties. These algorithms identify bogus news using numerous parallel networks and statements of varied durations. These parallel networks contain the n-gram convolutional layer, the main follicular layer, the previously utilised convolutional capsule layer, and the transitional capsule layer. Furthermore, the pooled mean was employed in the parallel cyst network's last layer. Depending on the length of the text, the models can learn to represent the text at different n-gram levels in a more extended and understandable manner.

Depending on the length of the release, use either model here. Figure 5 depicts the framework of the suggested model for medium- and long-term information disclosure. In this class, "Glove.6B.300d" is the pre-trained integrator. Four parallel networks with four various filter sizes (2, 3, 4, and 5) as the n-gram convolution layer are used to extract features. In the next layers, each parallel network has a primary shell and a convolutional shell. As the last layer, each parallel network employs a completely bonded shell. The survey average is used to calculate the final result. Because word order is constrained in short phrases, we propose utilising a different structure. Consider simply two parallel networks with filter sizes of three and five.The layers are the same as the model's first layers. Static magnetic integration is employed in this model.
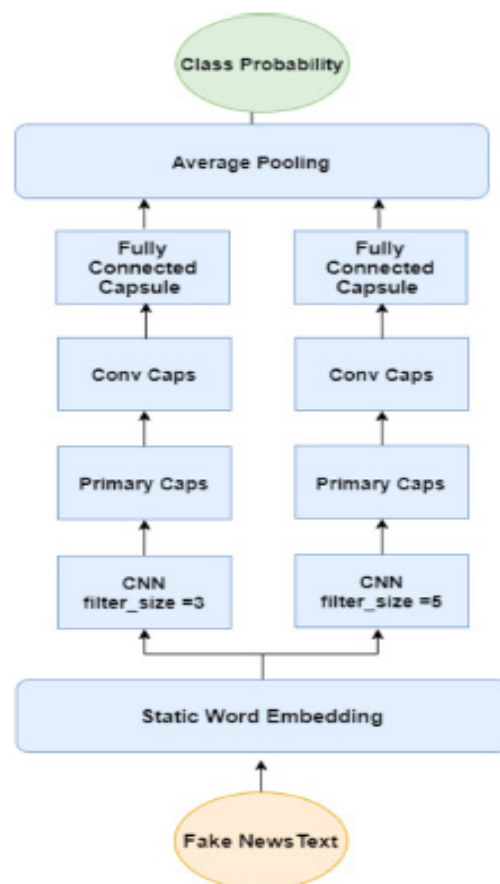
Figure 5. The model of the proposed static capsule network for detecting fake news in short news statements

RESULTS AND DISCUSSION

Many tests were run on the datasets listed below to assess the efficacy of the suggested model. This section explains these experiments and compares the outcomes with those of other fundamental approaches. To identify bogus news, several datasets have been presented. A huge dataset is necessary to train the model when utilising a neural network, and this is one of the primary criteria. In this work, we trained deep models using a

LIAR fake news dataset that contained a substantial quantity of data. Fake news dataset (4.1)

The LIAR fake news dataset
   A brand-new sizable dataset named LIAR, which is made up of 12,8,000 brief statements from the POLITIFACT. COM API that have been manually annotated. POLITIFACT.COM editor assessed claims for their veracity and true news items. The labels in this dataset are distributed based on six terms: fire pants, false, hardly true, half true, mainly true, and true. Labels for various products with flame retardant properties vary from 2063 to 2638.

Evaluation metrics
   The effectiveness of this study is assessed in comparison to the provided dataset and established methodologies like Capture, Score and Integration (CSI), Capsule Neural Network (CNN), and the suggested hybrid feature extraction model with Capsule Neural Network (HFEM-CNN). In the section below, the test metric is specified. I first construct several performance measures using the true positive (TP), false positive (FP), true negative (TN), and false negative (FN) rates. The percentage of linked retrieved versions constitutes the first performance metric, which is accuracy. Recall, which is characterised as the percentage of pertinent versions retrieved, is the second performance metric. Both measurements of accuracy and recall—though sometimes at odds with one another—are crucial for assessing how well a prediction system works. The F-index may be created by adding these two measures and giving them equal weights. Accuracy, which is determined by the ratio of the metrics, is the final performance statistic. among all cases that were predicted, this was the one that came true.
   Precision is defined as the ratio of correctly found positive observations to all of the expected positive observations.

$$Precision = TP/TP+FP \qquad (2)$$

   Sensitivity or Recall is defined the ratio of correctly identified positive observations to the over-all observations.

$$Recall = TP/TP+FN \qquad (3)$$

   F - measure is defined as the weighted average of Precision as well as Recall. As a result, it takes false positives and false negatives.

$$F1\ Score = 2*(Recall * Precision) / (Recall + Precision) \qquad (4)$$

Accuracy is calculated in terms of positives and negatives as follows:

$$Accuracy = (TP+FP)/(TP+TN+FP+FN) \qquad (5)$$

Table 1. Performance comparison results for the proposed fake news detection model

Metrics	CSI	CNN	HFEM-CNN
Accuracy	84,1400	88,7812	90,7623
Precision	85,2500	87,4836	91,5510
Recall	82,3400	87,9223	89,7452
F-measure	83,1900	87,7024	90,6391

   Figure 6 shows how the proposed and current false news detection models compare in terms of accuracy of findings. Based on the results, it can be said that the suggested HFEM-CNN strategy outperforms other classification methods in terms of accuracy.
   Figure 7 depicts a recall comparison of the proposed and existing false news detection methods. In terms of ability recovery capability, the suggested HFEM-CNN technique surpasses previous classification methods, according to the results.
   Figure 8 depicts a comparison of the F measure between the proposed and current fake news detection methods. The suggested HFEM-CNN technique outperforms the present classification methods that differ in F measure, according to the findings.
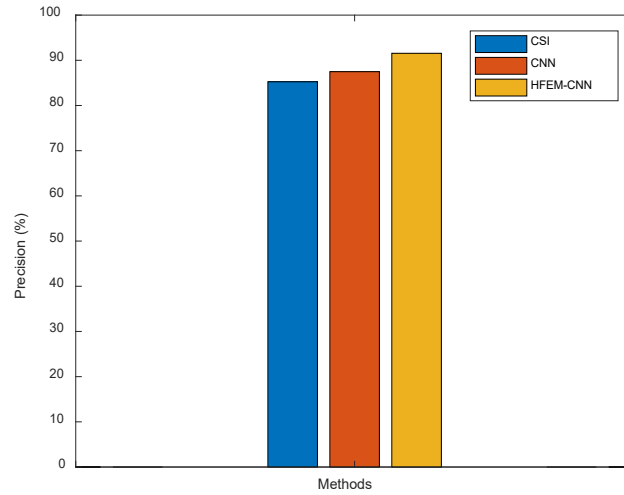
Figure 6. Comparisons of the proposed and existing false news detection models' accuracy findings
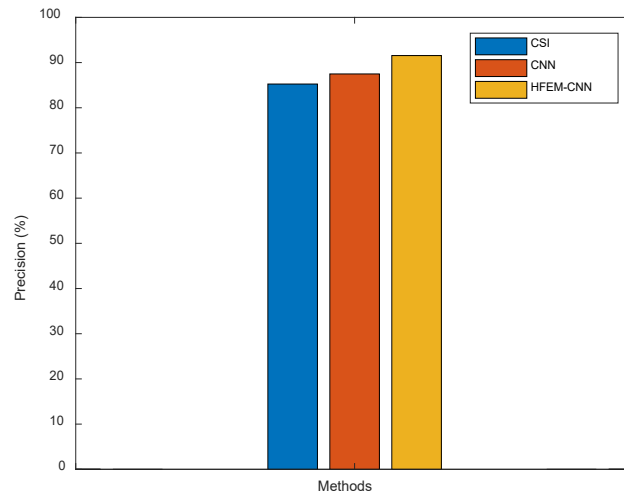

Figure 7. Recall comparative results between suggested and existing models for detecting fake news
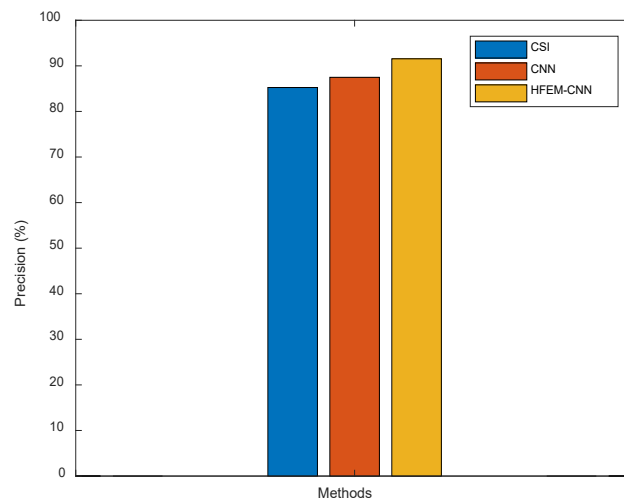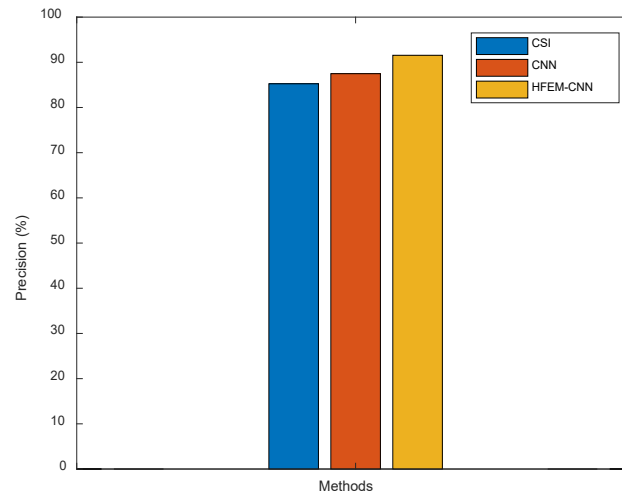

Figure 8. F-measure comparative results between suggested and existing models for detecting fake news

Figure 9. Accuracy comparative results between suggested and existing models for detecting fake news

Figure 9 depicts a comparison of the accuracy of the proposed and current false news detection methods. The suggested HFEM-CNN technique beats other current classification methods in terms of accuracy, according to the results.

CONCLUSIONS

Due to the proliferation of information on social networks, consumers have recently experienced trouble finding accurate and trustworthy information. The text mining techniques and AI algorithms used in this study are merged in this work to create a feature mining model that works in conjunction with the Capsule neural network to identify bogus news on social media. monitor. Separate analyses of supervised AI algorithms and text mining are carried out. Three phases of processing are carried out in this framework. From the textual data of the relevant articles, producers, and themes, a collection of explicit and implicit characteristics may be gleaned. A capsule neural network model was also suggested to include network structure information during model training based on the link between articles, creators, and news subjects. On the LIAR dataset, this combined model was put to the test and measured for accuracy, recall, precision, and F-measure. Future work is already under progress. Finding new methods and combining them with clever optimisation techniques helps enhance placement. overall and using various techniques. It is also possible to incorporate feature extraction techniques to enhance model performance.

REFERENCES

1. 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.

2. 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.

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

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

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

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

7. 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.

8. 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.

9. 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.

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

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

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

13. Goldani, M. H., Momtazi, S., & Safabakhsh, R. (2021). Detecting fake news with capsule neural networks. Applied Soft Computing, 101, 106991.

14. Ozbay, F. A., & Alatas, B. (2020). Fake news detection within online social media using supervised artificial intelligence algorithms. Physica A: Statistical Mechanics and its Applications, 540, 123174.

15. Huancahuire-Vega S, Newball-Noriega EE, Rojas-Humpire R, Saintila J, Rodriguez-Vásquez M, Ruiz-Mamani PG, et al. Changes in Eating Habits and Lifestyles in a Peruvian Population during Social Isolation for the COVID-19 Pandemic. Journal of Nutrition and Metabolism 2021;2021. https://doi.org/10.1155/2021/4119620.

16. Cabrera-Aguilar E, Zevallos-Francia M, Morales-García M, Ramírez-Coronel AA, Morales-García SB, Sairitupa-Sanchez LZ, et al. Resilience and stress as predictors of work engagement: the mediating role of self-efficacy in nurses. Frontiers in Psychiatry 2023;14. https://doi.org/10.3389/fpsyt.2023.1202048.

17. Cayupe JC, Bernedo-Moreira DH, Morales-García WC, Alcaraz FL, Peña KBC, Saintila J, et al. Self-efficacy, organizational commitment, workload as predictors of life satisfaction in elementary school teachers: the mediating role of job satisfaction. Frontiers in Psychology 2023;14. https://doi.org/10.3389/fpsyg.2023.1066321.

18. Plaza-Ccuno JNR, Puri CV, Calizaya-Milla YE, Morales-García WC, Huancahuire-Vega S, Soriano-Moreno AN, et al. Physical Inactivity is Associated with Job Burnout in Health Professionals During the COVID-19 Pandemic. Risk Management and Healthcare Policy 2023;16:725-33. https://doi.org/10.2147/RMHP.S393311.

19. Huaman N, Morales-García WC, Castillo-Blanco R, Saintila J, Huancahuire-Vega S, Morales-García SB, et al. An Explanatory Model of Work-family Conflict and Resilience as Predictors of Job Satisfaction in Nurses: The Mediating Role of Work Engagement and Communication Skills. Journal of Primary Care and Community Health 2023;14. https://doi.org/10.1177/21501319231151380.

20. Dilas D, Flores R, Morales-García WC, Calizaya-Milla YE, Morales-García M, Sairitupa-Sanchez L, et al. Social Support, Quality of Care, and Patient Adherence to Tuberculosis Treatment in Peru: The Mediating Role of Nurse Health Education. Patient Preference and Adherence 2023;17:175-86. https://doi.org/10.2147/PPA.S391930.

21. Morales-García WC, Huancahuire-Vega S, Saintila J, Morales-García M, Fernández-Molocho L, Ruiz Mamani PG. Predictors of Intention to Vaccinate Against COVID-19 in a Peruvian Sample. Journal of Primary Care and Community Health 2022;13. https://doi.org/10.1177/21501319221092254.

22. Chura S, Saintila J, Mamani R, Ruiz Mamani PG, Morales-García WC. Predictors of Depression in Nurses During COVID-19 Health Emergency; the Mediating Role of Resilience: A Cross-Sectional Study. Journal of Primary Care and Community Health 2022;13. https://doi.org/10.1177/21501319221097075.

23. Marquez NM, Saintila J, Castellanos-Vazquez AJ, Dávila-Villavicencio R, Turpo-Chaparro J, Sánchez-Tarrillo JA, et al. Telehealth-based interventions on lifestyle, body mass index, and glucose concentration in university staff during the coronavirus disease 2019 pandemic: A pre-experimental study. Digital Health 2022;8.

https://doi.org/10.1177/20552076221129719.

24. Soto CEH, Lizarme EAV. Administrative management and user satisfaction of tele-consultation in a FEBAN polyclinic in Lima. SCT Proceedings in Interdisciplinary Insights and Innovations 2024;2:217-217. https://doi.org/10.56294/piii2024217.

25. Huirse SAH, Panique JCA. Relationship Marketing and customer loyalty in the company Saga Falabella S.A. Cusco. SCT Proceedings in Interdisciplinary Insights and Innovations 2024;2:206-206. https://doi.org/10.56294/piii2024206.

26. Otero DL, Licourt MT. Clinical and genetic characterization of Duchenne Muscular Dystrophy. SCT Proceedings in Interdisciplinary Insights and Innovations 2024;2:221-221. https://doi.org/10.56294/piii2024221.

27. Gonzalez-Argote J, Castillo-González W. Update on the use of gamified educational resources in the development of cognitive skills. AG Salud 2024;2:41-41. https://doi.org/10.62486/agsalud202441.

28. Quiroz FJR, Gamarra NH. Psychometric evidence of the mobile dependence test in the young population of Lima in the context of the pandemic. AG Salud 2024;2:40-40. https://doi.org/10.62486/agsalud202440.

29. Trovat V, Ochoa M, Hernández-Runque E, Gómez R, Jiménez M, Correia P. Quality of work life in workers with disabilities in manufacturing and service companies. AG Salud 2024;2:43-43. https://doi.org/10.62486/agsalud202443.

30. Olguín-Martínez CM, Rivera RIB, Perez RLR, Guzmán JRV, Romero-Carazas R, Suárez NR, et al. Rescue of the historical-cultural heritage of the Yanesha: interculturality and inclusive education of the oral traditions. AG Multidisciplinar 2023;1:5-5. https://doi.org/10.62486/agmu20235.

31. Cuervo MED. Exclusive breastfeeding. Factors that influence its abandonment. AG Multidisciplinar 2023;1:6-6. https://doi.org/10.62486/agmu20236.

32. Auza-Santivañez JC, Lopez-Quispe AG, Carías A, Huanca BA, Remón AS, Condo-Gutierrez AR, et al. Work of the emergency system in polytraumatized patients transferred to the hospital. AG Multidisciplinar 2023;1:9-9. https://doi.org/10.62486/agmu20239.

33. Saavedra MOR. Revaluation of Property, Plant and Equipment under the criteria of IAS 16: Property, Plant and Equipment. AG Managment 2023;1:11-11. https://doi.org/10.62486/agma202311.

34. Solano AVC, Arboleda LDC, García CCC, Dominguez CDC. Benefits of artificial intelligence in companies. AG Managment 2023;1:17-17. https://doi.org/10.62486/agma202317.

35. Ríos-Quispe CF. Analysis of ABC Cost Systems. AG Managment 2023;1:12-12. https://doi.org/10.62486/agma202312.

36. Ledesma-Céspedes N, Leyva-Samue L, Barrios-Ledesma L. Use of radiographs in endodontic treatments in pregnant women. AG Odontologia 2023;1:3-3. https://doi.org/10.62486/agodonto20233.

37. Figueredo-Rigores A, Blanco-Romero L, Llevat-Romero D. Systemic view of periodontal diseases. AG Odontologia 2023;1:14-14. https://doi.org/10.62486/agodonto202314.

38. Millán YA, Montano-Silva RM, Ruiz-Salazar R. Epidemiology of oral cancer. AG Odontologia 2023;1:17-17. https://doi.org/10.62486/agodonto202317.

39. 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.

40. 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.

41. 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).

42. 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.

43. 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.

44. 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).

45. 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.

46. 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.

47. Zhang, J., Dong, B., & Philip, S. Y. (2020, April). Fakedetector: Effective fake news detection with deep diffusive neural network. In 2020 IEEE 36th International Conference on Data Engineering (ICDE) (pp. 1826-1829). IEEE.

48. 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).

CONFLICT OF INTEREST
The authors declare that there is no conflict of interest.

AUTHORSHIP CONTRIBUTION
Conceptualization: R. Uma Maheswari, N. Sudha.
Research: R. Uma Maheswari, N. Sudha.
Methodology: R. Uma Maheswari, N. Sudha.
Project management: R. Uma Maheswari, N. Sudha.
Original drafting: R. Uma Maheswari, N. Sudha.
Writing-revision and editing: R. Uma Maheswari, N. Sudha.