Multiset Dual Summarization for Incongruent News Article Detection

Sujit Kumar, Rohan Jaiswal, Mohit Ram Sharma and Sanasam Ranbir Singh
Department of Computer Science and Engineering
Indian Institute of Technology, Guwahati, Assam, India
{sujitkumar, mohit.rs, ranbir}@iitg.ac.in, therohanjaiswal@gmail.com

Abstract

The prevalence of deceptive and incongruent news headlines has highlighted their substantial role in the propagation of fake news, exacerbating the spread of both misinformation and disinformation. Existing studies on incongruity detection primarily concentrate on estimating the similarity between the encoded representation of headlines and the encoded representation or summary representative vector of the news body. In the process of obtaining the encoded representation of the news body, researchers often consider either sequential encoding or hierarchical encoding of the news body or to acquire a summary representative vector of the news body, they explore techniques like summarization or dual summarization methods. Nevertheless, when it comes to detecting partially incongruent news, dual summarization-based methods tend to outperform hierarchical encoding-based methods. On the other hand, for datasets focused on detecting fake news, where the hierarchical structure within a news article plays a crucial role, hierarchical encoding-based methods tend to perform better than summarization-based methods. Recognizing this contradictory performance of hierarchical encoding-based and summarization-based methods across datasets with different characteristics, we introduced a novel approach called "Multiset Dual Summarization" (MDS). MDS combines the strengths of both hierarchical encoding and dual summarization methods to leverage their respective advantages. We conducted experiments on datasets with diverse characteristics, and our findings demonstrate that our proposed model outperforms established state-of-the-art baseline models.

1 Introduction

Misleading and deceptive news headlines have emerged as a powerful force in the propagation of false information, resulting in a double-pronged effect that intensifies the proliferation of misinformation and disinformation (Wang et al., 2021). Firstly, when integrated into current news stories gaining popularity, such headlines entice readers to interact with the material, thereby extending the influence of false information. Secondly, they perpetuate a cycle of fake news by twisting facts and manipulating how readers perceive them, ruining the trustworthiness of reputable news outlets. These dual negative consequences significantly bolster the spread of misinformation and disinformation in the contemporary media (Chesney et al., 2017; Effron and Raj, 2020) [1, 2]. A news article is said to be incongruent if the news headline misrepresents its body through fabrication, manipulation, false connections[3], or incorrect context[4] (Ecker et al., 2014; Chesney et al., 2017; Wei and Wan, 2017). As outlined in research studies (Rieis et al., 2015; Gabielkov et al., 2016; Wei and Wan, 2017), deceptive headlines play a significant role in vitality on social media and also influence readers' opinions (Tannenbaum, 1953). Inconsistent news can adversely impact readers, leading to false beliefs and erroneous opinions [5, 6] (Ecker et al., 2014, 2022; Tsfati et al., 2020). Once this misleading information spreads, it becomes challenging to rectify, as research (Ecker et al., 2020) has found that corrective measures may not have a substantial impact and can, in some instances, even reinforce individuals' misconceptions. Consequently, the detection of deceptive and incongruent news articles (Chesney et al., 2017; Ecker et al., 2014; Horner et al., 2021; Bago et al., 2020; Guess et al., 2020) has become a crucial research problem in combating the dissemination of misinformation in digital media. According to findings in the studies (Chesney et al.,

[1] Misleading headline fake news regarding WHO
[2] Misleading headlines as fake news
[3] When the caption of the image does not align with its image or the headline does not support its content.
[4] Legitimate information is presented in the wrong context
[5] Effects of the misleading headline on health
[6] Impact of misleading headlines related to economy

2017; Kumar et al., 2022), there exist four primary characteristics associated with deceptive incongruent news articles: (i) The claim presented in the headline either has no relation to or directly contradicts the claim articulated in the body of the news article. (ii) News headlines and the body text pertain to the same topic or event, yet the content in the headline and the body are unrelated to each other. (iii) While both the headline and body describe a genuine event, the dates or entities mentioned in them have been manipulated. (iv) There are cases where certain paragraphs in the news body align with the headline while others do not, resulting in what is termed as "partially incongruent.

In the early stages (Pomerleau and Rao, 2017; Hanselowski et al., 2018a; Riedel et al., 2017) of research on detecting incongruent news, researchers employed basic n-gram features to assess the similarity between news headlines and the body of news articles. Further, studies on incongruent news article detection can be grouped into three categories: similarity-based, summarization-based, and dual summarization-based. Similarity-based studies can be further grouped into two categories: non-hierarchical encoding and hierarchical encoding-based methods. Non-hierarchical encoding-based studies, as seen in works by Hanselowski et al. (2018)(Hanselowski et al., 2018b) and Borges (2019)(Borges et al., 2019), aim to obtain sequential encoding of the news body and headline. They then estimate the similarity between the encoded representations of the headline and news body. On the other hand, hierarchical encoding-based methods, as explored in studies by, (Karimi and Tang, 2019a; Conforti et al., 2018; Yoon et al., 2019). define a news article as a hierarchical structure where the body is a collection of paragraphs, and each paragraph is a collection of sentences. They then obtain hierarchical encodings of the news and an encoding of the news headline. Subsequently, these methods estimate the similarity between the encoded representation of the headline and the news body for incongruent news detection. While hierarchical encoding aids in obtaining a better encoded representation of the news body for incongruent news article detection, studies (Mishra et al., 2020; Yoon et al., 2021; Kumar et al., 2022) report that the aforementioned similarity-based methods often struggle to detect incongruent news in articles with larger paragraphs and sentences. To overcome the limitation of similarity-based methods, summarization-based studies (Sepúlveda-Torres et al., 2021b; Mishra et al., 2020; Kim and Ko, 2021a) first summarize 'the news body to generate a synthetic news headline and then estimate the similarity between the generated news headline and the actual headline to detect incongruent news. Though summarization-based studies overcome the limitations of similarity-based studies, summarization-based studies fail to detect partially incongruent news articles (Kumar et al., 2022). Dual summarization-based study (Kumar et al., 2022) split the sentences of news body into two sets, positive and negative sets, based on contextual similarity of sentences and headlines. Next, generate the summary of both positive and negative sets separately and estimate the similarity between the headline and summary of positive and negative for incongruent news detection. However, the Dual summarization-based study (Kumar et al., 2022) splits the news articles into two sets based on the similarity between sentences and headline, which leads to a loss of hierarchical information present in a news article and also, the positive and negative sentences are biased toward headline which leads to loss of news body context information. Our experimental results in Table 1 also suggest that dual summarization-based methods outperform hierarchical encoding-based methods for partial incongruent news detection and hierarchical encoding-based methods outperform summarization-based methods on fake news detection datasets where hierarchical information within a news article is critical. To overcome the limitations of dual summarization-based method, in this study we combine the hierarchical encoding of news body and dual summarization-based methods. We proposed *Multiset Dual Summarization* (**MDS**), which combines both hierarchical encoding and dual summarization to leverage the advantages of each. Hierarchical encodings of the news body help obtain an encoded representation that captures hierarchical information present in the news body and the contextual information within the news body. Similarly, dual summarization aids in separating news body sentences that are congruent and incongruent with respect to the headline. We conduct our experiments over datasets of different natures, and our experimental results over datasets of different natures suggest that the proposed models outperform existing state-of-the-art models in the literature.

2 Related Work

In this section, we will briefly review prior research related to the detection of incongruent news articles. Initial studies on incongruent news article detection mainly consider bag-of-words-based features, such as n-grams, term frequency-inverse document frequency (TF-IDF), and topic modelling features (Sepúlveda-Torres et al., 2021a; Riedel et al., 2017; Hanselowski et al., 2018a). Moreover, research on detecting incongruent news articles can be categorized into three main groups: similarity-based, summarization-based, and dual summarization-based methods. Similarity-based studies can be further grouped into two categories: non-hierarchical encoding and hierarchical encoding-based methods. Non-hierarchical encoding-based studies, as seen in works by Hanselowski et al. (2018)(Hanselowski et al., 2018b) and Borges (2019)(Borges et al., 2019), aim to obtain sequential encoding of the news body and headline. They then estimate the similarity between the encoded representations of the headline and news body. On the other hand, hierarchical encoding-based methods, as explored in studies by, (Karimi and Tang, 2019a; Conforti et al., 2018; Yoon et al., 2019). define a news article as a hierarchical structure where the body is a collection of paragraphs, and each paragraph is a collection of sentences. They then obtain hierarchical encodings of the news and an encoding of the news headline. Subsequently, these methods estimate the similarity between the encoded representation of the headline and the news body for incongruent news detection. While hierarchical encoding aids in obtaining a better encoded representation of the news body for incongruent news article detection, studies (Mishra et al., 2020; Yoon et al., 2021; Kumar et al., 2022) report that the aforementioned similarity-based methods often struggle to detect incongruent news in articles with larger paragraphs and sentences. To overcome the limitation of similarity-based methods, summarization-based studies (Sepúlveda-Torres et al., 2021b; Mishra et al., 2020; Kim and Ko, 2021a) first summarize 'the news body to generate a synthetic news headline and then estimate the similarity between the generated news headline and the actual headline to detect incongruent news. Though summarization-based studies overcome the limitations of similarity-based studies, summarization-based studies fail to detect partially incongruent news articles (Kumar et al., 2022). Dual summarization-based study (Kumar et al., 2022) split the sentences of news body into two sets, positive and negative sets, based on contextual similarity of sentences and headlines. Next, generate the summary of both positive and negative sets separately and estimate the similarity between the headline and summary of positive and negative for incongruent news detection. However, the Dual summarization-based study (Kumar et al., 2022) splits the news articles into two sets based on the similarity between sentences and headline, which leads to a loss of hierarchical information present in a news article and also, the positive and negative sentences are biased toward headline which leads to loss of news body context information. Motivated by the above limitations of dual summarization-based methods, this study proposes We proposed *Multiset Dual Summarization* (**MDS**), which combines both hierarchical encoding and dual summarization to leverage the advantages of each.

3 Proposed Method

As discussed above, this study aims to combine hierarchical encoding of the news body with dual summarizations. Our proposed model, **MDS**, first obtains a hierarchical encoding of the news body, encodings for each sentence in the news body and headline. It then divides the sentences of the news body into four sets based on the similarity between the encoded representation of the headline and the encoded representation of the sentences in the news body, as well as the similarity between the hierarchical encoded representation of the news body and the encoded representation of the sentences in the news body. By following the above objective, our proposed model takes advantage of both a dual summarization approach and a hierarchical encoding-based approach for incongruent news article detection. Figure 2 presents the block diagram of our proposed model, *Multiset Dual Summarization* **MDS**. Given a pair of news body $\mathcal{B}$ and headline, $\mathcal{H}$ our proposed model **MDS** first splits the news article into four sets, namely head positive $\mathcal{H}^+$, head negative $\mathcal{H}^-$, body positive set $\mathcal{B}^+$ and body negative set $\mathcal{B}^-$. Sentences of news body are placed into head positive $\mathcal{H}^+$ and head negative $\mathcal{H}^-$ based on similarity between the encoded representation of the sentence and the encoded representation of the headline. In contrast, sentences of news body are placed into body positive, $\mathcal{B}^+$ and body negative $\mathcal{B}^-$ sets based on the similarity

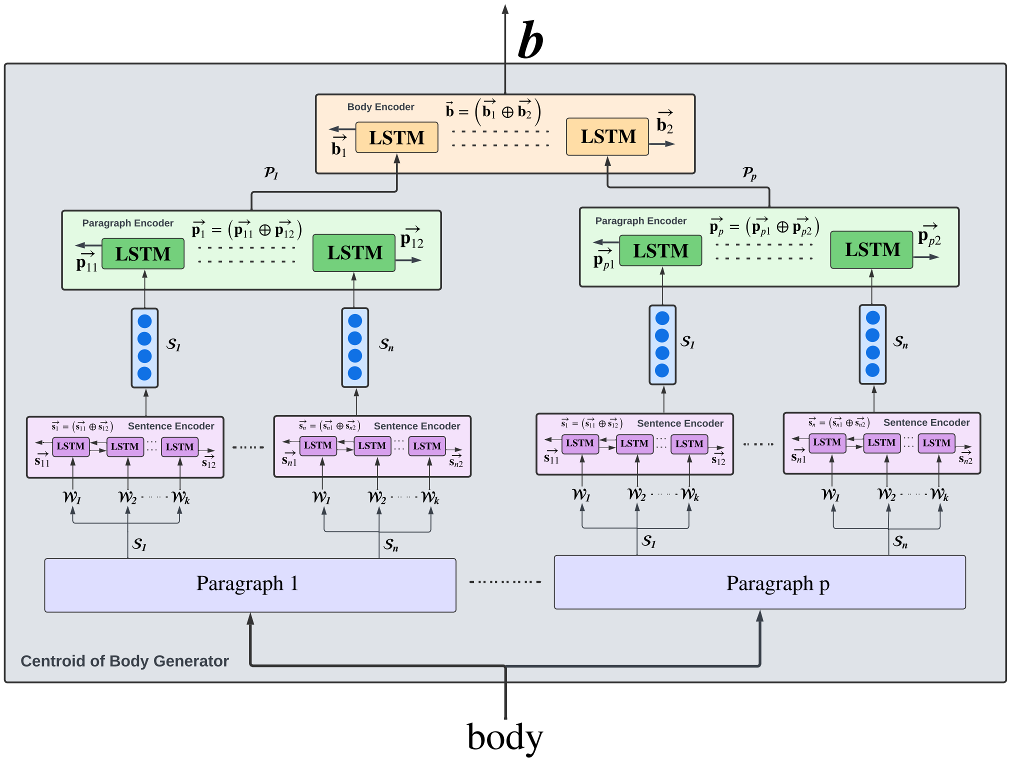

Figure 1: present the hierarchical encoding of the news body. Given a new body **MDS** first splits the news body $\mathcal{B}$ into a set of sentences $\mathcal{S}_i$ and then obtains an encoded representation $\mathbf{s_i}$ of sentences $\mathcal{S}_i$. Next, a paragraph $\mathcal{P}_i$ is defined as a sequence of sentences within a paragraph. accordingly, encoded representation $\mathbf{p_i}$ of paragraphs $\mathcal{P}_i$ is obtained by applying BiLSTM over encoded representation $\mathbf{s_i}$ of sentences $\mathcal{S}_i$ within paragraphs $\mathcal{P}_i$. Subsequently, encoded representation $\mathbf{b}$ of news body $\mathcal{B}$ is obtained by applying BiLSTM over encoded representation $\mathbf{p_i}$ of paragraphs $\mathcal{P}_i$.

between the encoded representation of the sentence and an encoded representation of a full news body. The primary rationale for categorizing body sentences into head positive $\mathcal{H}^+$ and head negative $\mathcal{H}^-$ sets lies in the identification of incongruities within news articles. When a news article is partially incongruent, the sentences that align with the headline are placed in the head positive $\mathcal{H}^+$, while the sentences that deviate from the headline are classified in the negative set head negative $\mathcal{H}^-$. Similarly, in the case of a fully congruent news article, the majority of the sentences in the body should belong to the head positive $\mathcal{H}^+$, with only a few sentences residing in the head negative $\mathcal{H}^-$. However, in the scenario where a news article is entirely incongruent, all the sentences in the body should contrast with the headline and, therefore, belong to the negative set head negative $\mathcal{H}^-$, with the exception of one or a few sentences that align with the headline and are placed in the head positive $\mathcal{H}^+$. Similarly, in the case of congruent news articles, the sentences of the news body must be correlated with each other. Accordingly, in the case of congruent news, most of the sentences will be in body positive set, $\mathcal{B}^+$, and only a few sentences will be body negative set $\mathcal{B}^-$. Whereas in the case of a partially incongruent news article, one or more paragraphs of the news body will not correlate with other paragraphs of the news body. Consequently, sentences which are highly similar to the encoded representation of the news body will be placed in body positive set, $\mathcal{B}^+$, and sentences which do not align with other paragraphs of news body, i.e. least similar to the encoded representation of news body will be placed in body negative set $\mathcal{B}^-$.

4 Similarity between headline and sentence of news body

We apply *Bidirectional Long Short-Term Memory* (BiLSTM) (Hochreiter and Schmidhuber, 1997) over the headline and sentences of news body to obtain encoded representations $\mathbf{h}$ and $\mathbf{s}_i$ of headline $\mathcal{H}$ and sentence $\mathcal{S}_i$, respectively. Subsequently, fol-

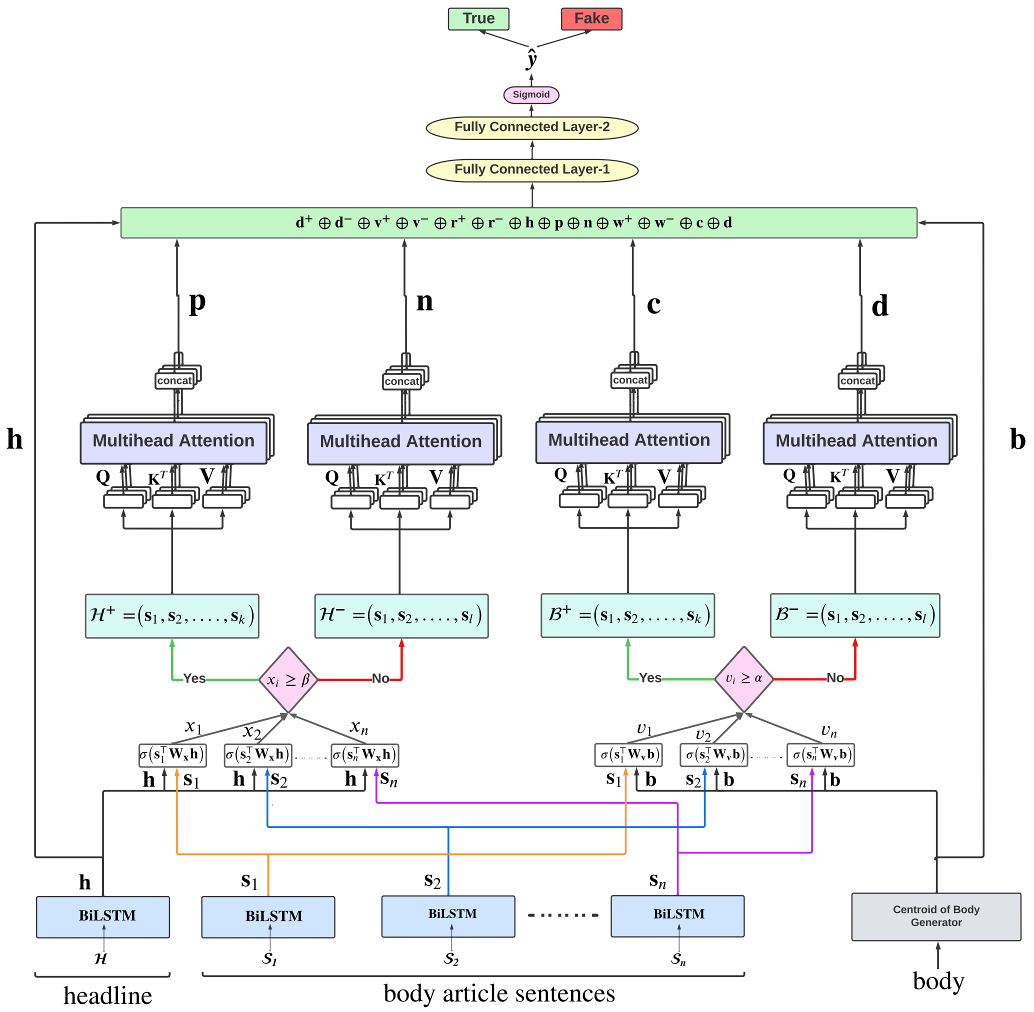

Figure 2: present the working of **MDS** models. First, **MDS** obtains and encoded representations $\mathbf{h}$, $\mathbf{s}_i$ and $\mathbf{b}$ of headline $\mathcal{H}$, sentences $\mathcal{S}_i$ and news body $\mathcal{B}$. Subsequently, estimate similarity $\mathbf{x_i}$ between $\mathbf{h}$ and $\mathbf{s}_i$, and similarity $\mathbf{v_i}$ between $\mathbf{b}$ and $\mathbf{s}_i$. If $x_i \geq \boldsymbol{\beta}$, then the sentence $\mathbf{s}_i$ is added to head positive set; otherwise, it is added to set head negative set $\mathcal{H}^-$. Similarly, If $v_i \geq \boldsymbol{\alpha}$, then the sentence $\mathbf{s}_i$ is added to the body positive set; otherwise, it is added to set body negative set $\mathcal{B}^-$. Next, obtain a summary representative vector $\mathbf{p}$, $\mathbf{n}$, $\mathbf{c}$, and $\mathbf{d}$ and form a feature vector $\mathbf{e}$ and pass it to a fully connected neural network for incongruent news classification.

lowing the steps reported in the studies (Tay et al., 2018; Luong et al., 2015) we estimate similarity score $\mathbf{x}_i$ between $\mathbf{h}$ and $\mathbf{s}_i$ as defined below.

$$\mathbf{x}_i = \sigma\left(\mathbf{s}_i^\top \mathbf{W}_x \mathbf{h}\right) \tag{1}$$

where $\mathbf{W}_x$ is a learnable parameter matrix, $\boldsymbol{\sigma}$ is the sigmoid function and $\top$ is a transpose operation over a vector. If $\mathbf{x}_i \geq \boldsymbol{\beta}$, then the sentence $\mathbf{s}_i$ is added to head positive set $\mathcal{H}^+$, otherwise it is added to set head negative set $\mathcal{H}^-$. Next, to split the sentences of news body into body positive set, $\mathcal{B}^+$ and body negative set, $\mathcal{B}^-$, we first obtain an encoded representation of news body $\mathbf{b}$ applying hierarchical encoding over the news body. Our hierarchical encoding of news body is similar to *Recursive and Sequential Deep Hierarchical Encoding* (RaSHE) model as defined in the study (Kumar et al., 2023) except TreeLSTM is replaced by bidirectional LSTM (BiLSTM) to encode sentences within paragraphs of news body. Figure 1 presents the working of hierarchical encoder model used to encode news body. Once we obtain an encoded representation $\mathbf{b}$ of news body by applying hierarchical encoding over news body $\mathcal{B}$, next we estimate the similarity score $\mathbf{v}_i$ between $\mathbf{b}$ and $\mathbf{s}_i$ as defined below.

$$\mathbf{v}_i = \sigma\left(\mathbf{s}_i^\top \mathbf{W}_v \mathbf{b}\right) \tag{2}$$

where $\mathbf{W}_v$ is a learnable parameter matrix. If $\mathbf{v}_i \geq \boldsymbol{\alpha}$, then the sentence $\mathbf{s}_i$ is added to the body positive set $\mathcal{B}^+$, otherwise it is added to the body negative set $\mathcal{B}^-$.

4.1 Summarization

We derive summary representative vectors from four sets of sentences: $\mathcal{H}^+$, $\mathcal{H}^-$, $\mathcal{B}^+$, and $\mathcal{B}^-$, using a *multi-head attention-based summary* approach for each set individually. The characteristics of the dual summary over head positive $\mathcal{H}^+$, head negative $\mathcal{H}^-$, body positive set $\mathcal{B}^+$ and body negative set $\mathcal{B}^-$ sets are defined as follows. (i) A sentence which is highly similar to other sentences in the head positive set $\mathcal{H}^+$ should be given high attention weight while generating a summary of the head positive set $\mathcal{H}^+$. (ii) A sentence which is least similar to other sentences in the head negative set $\mathcal{H}^-$ should be given high importance while generating a summary of the head negative set $\mathcal{H}^-$. The characteristics of dual summary for body positive set $\mathcal{B}^+$ are similar to characteristics for head positive, $\mathcal{H}^+$ and the characteristics of summary for body negative set $\mathcal{B}^-$ are similar to characteristics for the head negative set $\mathcal{H}^-$. The primary motivation behind creating dual summaries lies in the comparison between these two summary types. If a summary generated by a highly influential sentence (sentences with high similarity with all other sentences in the set) from a head positive set $\mathcal{H}^+$ and a summary generated by a sentence that is not similar or at least similar to other sentences in the $\mathcal{H}^-$ are highly similar with the headline, then the news article is congruent, otherwise incongruent. We apply multi-head attention (Vaswani et al., 2017) over $\mathcal{H}^+$, $\mathcal{H}^-$, $\mathcal{B}^+$, and $\mathcal{B}^-$ to obtain a summary representative vector by capturing different aspects of sentences within the set. Given a sequence of sentences $(\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_k)$, we establish a matrix $\mathbf{H}$ in which each row represents a sentence encoding. Where k is the number of sentences in the respective set. Subsequently, we derive the query matrix $\mathbf{H}^q$, the key matrix $\mathbf{H}^k$, and the value matrix $\mathbf{H}^v$ using the following expression.

$$\mathbf{H}_c^q, \mathbf{H}_c^k, \mathbf{P}_c^v = \mathbf{H} \cdot \mathbf{W}_c^q, \mathbf{H} \cdot \mathbf{W}_c^k, \mathbf{P} \cdot \mathbf{W}_c^v \quad (3)$$

Here, $\mathbf{W}_c^q$, $\mathbf{W}_c^k$, and $\mathbf{W}_c^v$ represent learnable parameter matrices for query, key, and value projections, respectively, for the c^{th} attention head of the multi-head attention. The $\cdot$ denotes the dot product operation between matrices. Subsequently, the attention weights $\mathbf{A}_c$ are defined as follows:

$$\mathbf{S} = \left(\frac{\mathbf{H}_c^q\, (\mathbf{H}_c^k)^\top}{\sqrt{z}}\right) \quad (4)$$

$$\mathbf{A}_{c,i,j} = \left(\frac{\mathbf{exp}(\mathbf{S}_{ij})}{\sum_{k,l}\mathbf{exp}(\mathbf{S}_{k,l})}\right) \quad (5)$$

In this context, $\mathbf{S}$ stands for the similarity matrix, and $\mathbf{A}_c$ denotes the attention weight matrix for the c^{th} attention head. Each entry in the $\mathbf{A}_c[i, j]$ matrix represents the similarity probability between the $\mathbf{i^{th}}$ and $\mathbf{j^{th}}$ sentence in the set $\mathcal{H}^+$. The dimension of $\mathbf{H}_c^q$ is denoted as z. Subsequently, a weighted summation is applied over the sentence encodings $\mathbf{s}_i$ based on their similarity with other sentences within the set.

$$\mathbf{u}_{c,i} = \left(\sum_{j=1, i\neq j}^{k} \mathbf{A}_{c,ij}\mathbf{P}_{c,i}^v\right) \quad (6)$$

Where $\mathbf{u}_{c,i}$ is the sentence representation obtained after weighted summation between i^{th} sentence of $\mathbf{H}_c^v$ and attention weight $\mathbf{A}_{c,ij}$ between i^{th} sentence with all other sentences j in $\mathbf{H}_c^v$ of attention head c. Similarly, by following equation 6, representation of other sentences in a respective set are also obtained to form a sentence representation matrix $\mathbf{U}_c = \{\mathbf{u}_{c,1}, \mathbf{u}_{c,2}, ..., \mathbf{u}_{c,k}\}$ of attention head c. Now we concatenate the sentence representation obtained by different attention heads and pass it to a dense layer to obtain the final sentence representation $\mathbf{U}$.

$$\mathbf{U} = \Big(\mathbf{U}_1 \oplus \mathbf{U}_2 \oplus ..\mathbf{U}_c \oplus . \oplus \mathbf{U}_l\Big)\mathbf{W}_u \quad (7)$$

where, $\mathbf{W}_u$ represents the trainable parameter matrix, and $\mathbf{U}_c$ is the representation derived for the $\mathbf{c}^{th}$ attention head. $\mathbf{U}$ is a sentence representation matrix obtained by concatenating the representations of the $\mathbf{i^{th}}$ sentence derived from l different attention heads. Subsequently, the representations of the sentences $\mathbf{u}_i$ within the sentence representation matrix $\mathbf{U}$ are concatenated, and the resulting matrix is passed through a dense layer to generate a summary $\mathbf{p}$ for the head positive set $\mathcal{H}^+$.

$$\mathbf{p} = \Big(\mathbf{u}_1 \oplus \mathbf{u}_2 \oplus .. \oplus \mathbf{u}_i \oplus . \oplus \mathbf{u}_k\Big)\mathbf{W}_m \quad (8)$$

Here, $\mathbf{u}_i$ represents a row vector in the matrix $\mathbf{U}$, and $\mathbf{W}_m$ is a learnable parameter matrix and where $\oplus$ is a vector concatenation operation. Similarly, we also obtain a summary representative vector $\mathbf{c}$

for sentences in the body positive set $\mathcal{B}^+$ by following Equations 3, 4, 5, 6, 8 and 7. Similarly, to extract a summary representative vector $\mathbf{n}$ for the head negative set $\mathcal{H}^-$, equation 5 is replaced by equation 9. This modification is made to prioritize the sentence with the least similarity score to other sentences within the set $\mathcal{H}^-$ when generating a summary $\mathbf{n}$ for the set $\mathcal{H}^-$. Similarly, we also obtain a summary representative vector $\mathbf{d}$ for body negative set $\mathcal{B}^-$ by following Equations 3, 4, 9, 6, 8 and 7.

$$\mathbf{A_{c,i,j}} = \left(\frac{\mathbf{exp(1 - S_{ij})}}{\sum_{k,l}\mathbf{exp(1 - S_{k,l})}}\right) \tag{9}$$

4.2 Aggregation and Classification

After obtaining the summary representative vectors $\mathbf{p}$, $\mathbf{n}$, $\mathbf{c}$, and $\mathbf{d}$ for the positive and negative sets of both the headline and body, we construct a feature vector $\mathbf{e}$ by assessing the angle and difference between the encoded representation of the headline and the summary representative vectors $\mathbf{p}$, $\mathbf{n}$, $\mathbf{c}$, and $\mathbf{d}$ of the news body. The primary motivation behind estimating these measures is as follows: for a congruent news article, the encoded representation of the headline and the summary representative vectors will indeed show a high degree of similarity. Conversely, in the case of an incongruent news article, the summary derived from the encoded representation of the headline will exhibit the least similarity to the summary of the representative vectors. In the instances of partially incongruent news articles, the encoded representation of the headline will demonstrate the least similarity to the summary of the representative vectors from the negative sets, while displaying a high level of similarity to the summary of the representative vectors from the positive sets.

$$\mathbf{d^+}\ ,\mathbf{d^-} = \mathbf{h} \odot \mathbf{p}\ ,\mathbf{h} - \mathbf{p} \tag{10}$$

$$\mathbf{v^+}\ ,\mathbf{v^-} = \mathbf{h} \odot \mathbf{n}\ ,\mathbf{h} - \mathbf{n} \tag{11}$$

$$\mathbf{r^+}\ ,\mathbf{r^-} = \mathbf{h} \odot \mathbf{c}\ ,\mathbf{h} - \mathbf{c} \tag{12}$$

$$\mathbf{w^+}\ ,\mathbf{w^-} = \mathbf{h} \odot \mathbf{d}\ ,\mathbf{h} - \mathbf{d} \tag{13}$$

$$\begin{aligned}\mathbf{e} = \big(&\mathbf{d^+} \oplus \mathbf{d^-} \oplus \mathbf{v^+} \oplus \mathbf{v^-} \oplus \mathbf{r^+} \oplus \mathbf{r^-}\\ &\oplus \mathbf{h} \oplus \mathbf{p} \oplus \mathbf{n} \oplus \mathbf{w^+} \oplus \mathbf{w^-} \oplus \mathbf{c} \oplus \mathbf{d}\big)\end{aligned} \tag{14}$$

where $\oplus$ is a vector concatenation operation. Following the estimation of the feature vector $\mathbf{e}$, it is then input into a three-layer fully connected neural network for the purpose of incongruent news classification.

5 Experimental setups and discussions

5.1 Dataset Characteristics

We consider five datasets of different nature from both Hindi and English language. We consider (ISOT)[7](Ahmed et al., 2018, 2017), Fake News Challenge (FNC) dataset[8] (Pomerleau and Rao, 2017), and NELA-17 (News Landscape) dataset (Horne et al., 2018; Yoon et al., 2019) for English Language fake and incongruent news article detection. The NELA dataset is compiled following the methodology outlined in (Yoon et al., 2019) over the news article corpus released by (Horne et al., 2018; Yoon et al., 2019). In this dataset, news articles from reputable media sources are classified as congruent (Cong.), while incongruent news articles are generated by inserting a paragraph from a randomly selected news article into a congruent news article. As only one paragraph is inserted into a congruent news article to create incongruent samples, all other paragraphs, except the inserted one, remain congruent with the headline. Therefore, the incongruent samples in the NELA dataset are considered partially incongruent. The FNC dataset comprises four distinct classes: agree, disagree, discuss, and unrelated. Samples from agree, disagree, and discuss classes are amalgamated and termed as the congruent class, while the unrelated class samples are regarded as incongruent class. An essential characteristic of the FNC dataset is that the unrelated class consists of samples created by pairing headlines and bodies from different news articles on unrelated topics, hence referred to as fully incongruent news articles (Pomerleau and Rao, 2017).

We also curate fake news detection datasets for the Hindi language. The two synthetic Hindi fake news datasets are curated using the following two methods on publicly available Hindi news articles from the BBC news Corpus: (i) *Split and Merge* (SM) and (ii) *Named-Entity Replacement*(NE-R). The *split and merge*(SM) method is inspired by the NELA-17 dataset by following the procedure reported in the study (Yoon et al., 2019). In this method, news articles published by the *British Broadcasting Corporation* (BBC) are considered as true news, and a fake news article is generated by inserting random paragraphs from random news articles into a true news article. Thus, this approach creates partially incongruent news articles unless

[7]ISOT Fake News Dataset Repository Source
[8]Fake News Challenge (FNC)

the randomly selected article is similar to the true news article. To avoid this situation, the dataset is curated by inserting three random paragraphs from random news articles into one true news article's body to make it fake. The Named-Entity Replacement(NE-R) method generates fake news articles by replacing associated entities in true news articles with different entities. Replacing entities in a true news article may make the article fake. For example, a claim made in a news article [9] *Corona affects Hindus, Muslims have dua, don't need vaccine: Kolkata's Maulana Barkati* if the name Maulana Barkati is replaced by another entity World Health Organization, then it becomes fake news of serious concern. Motivated by such scenarios, the proposed NE-R method generates fake news articles by impersonating an individual, personality, celebrity or organization. Both BBC SM and BBC NE-R datasets are balanced datasets, as one fake news article is generated for each true news article. Table 3 presents the characteristics of our experimental datasets. We also curate a real fake news dataset by manually collecting the Hindi language real fake news under circulation from different digital platforms such as Facebook, Twitter, Reddit, Koo . For true news articles, news articles from reputed media houses were collected.

5.2 Experimental setups

The details of experimental hyperparameters are presented in appendix section A.

5.3 Baseline

This study considers hierarchical encoding-based studies (Hierarchical Discourse level Structure Learning) `HDSF` (Karimi and Tang, 2019a) (Attentive Hierarchical Dual Encoder) `AHDE` (Yoon et al., 2019)(Graph-based Hierarchical Dual Encoder) (Yoon et al., 2021) `GHDE`, `HoBERT` (Hierarchy over BERT) (Kumar et al., 2023) and `HeLSTM` (hierarchical encoding using LSTM) (Kumar et al., 2023) as hierarchical encoding-based methods from literature to study the response of hierarchical encoding-based methods over datasets of different nature. Similarly, we consider summarization-based methods `GFND` (Graph-based Fake News Detection using a summarization) (Kim and Ko, 2021b,a) and dual summarization base method `MADS` (Multi-head Attention-based Dual Summarization) (Kumar et al., 2022) to study the response of summarization and dual summarization-based methods over datasets of different nature. We also consider baseline models `BERT` and `RoBERT` as defined in studies (Kumar et al., 2022, 2023) as *Bidirectional Encoder Representations from Transformers* (BERT) (Devlin et al., 2019) encoding-based baseline models. Our setting for `BERT` and `RoBERT` baseline models is similar to as defined in studies (Kumar et al., 2022, 2023).

5.4 Results and discussion

Table 1 presents the performance baseline and proposed models over datasets of different natures from both the Hindi and English languages. All the baseline models are grouped into three categories, namely *Encoding*, *Hierarchical*, *Summarization*. While inspecting the performance of the baseline models across the datasets, the following interesting observations can be made. From Table 1 it is evident that among the encoding-based baseline model $RoBERT$ outperforms $BiLSTM$ over *SM*, *NE-R*, *NELA*, *ISOT*, and *FNC* datasets. From Table 1, it is also evident that the performance of hierarchical encoding-based models is superior to the performance of summarization-based baseline model over *SM*, *NE-R* dataset. In contrast, the performance of summarization-based baseline model models is superior to the performance of hierarchical encoding-based models of NELA and FNC datasets. Relating such contradictory performance of hierarchical encoding and summarization-based models to characteristics of datasets, discussed in subsection 5.1 it can be concluded that the nature of datasets heavily influences the performance of models. Motivated by such observations, we propose the *Multiset Dual Summarization* **MDS** model, which combines both hierarchical encoding and dual Summarization to take advantage of both hierarchical encoding and dual summarization-based approaches. From Table 1 it is evident that our proposed model **MDS** outperforms both hierarchical encoding and summarization-based baseline model over *SM*, *NELA*, *ISOT*, and *FNC* datasets. Table 2 presents the performance of models trained on dataset curated using *SM* and *NE-R* methods using BBC news corpus and tested over real fake news datasets. From Table 2, it is evident that our hierarchical encoding-based methods outperform summarization and dual summarization-based studies. However, our proposed model **MDS** outperforms both hierarchical, summarization and dual Summarization methods over real fake news datasets.

[9] Named entity example

Table 1: Comparing the performance of models trained over synthetic dataset and tested over synthetic datasets only. (i) **Acc** : indicates the accuracy, (ii) **F** indicates F-measure score. color indicates the best peromance.

Models		BBC				English					
		SM		NE-R		NELA - 17		ISOT		FNC	
		Acc	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**	**Acc**	**F**
Encoding	**BiLSTM** (Kumar et al., 2022, 2023)	0.840	0.839	0.926	0.926	0.555	0.550	0.990	0.990	0.616	0.504
	RoBERT (Kumar et al., 2022, 2023)	**0.894**	**0.893**	**0.983**	**0.982**	**0.615**	**0.613**	**0.996**	**0.996**	**0.664**	**0.583**
Hierarchical	**AHDE** (Yoon et al., 2019)	0.691	0.671	0.869	0.869	0.606	0.606	0.913	0.913	0.691	0.454
	HDSF (Karimi and Tang, 2019b)	0.889	0.888	0.983	0.983	0.517	0.494	0.720	0.712	0.758	0.666
	HoBERT (Kumar et al., 2023)	0.899	0.898	**0.985**	**0.984**	**0.635**	**0.634**	0.991	0.991	0.686	0.632
	RaSHE (Kumar et al., 2023)	**0.909**	**0.908**	0.985	0.984	0.652	0.652	**0.999**	**0.999**	0.805	0.805
Summarization	**GFNDS** (Kim and Ko, 2021b,a)	0.514	0.502	0.505	0.504	0.533	0.532	0.998	0.998	0.878	0.837
	MADS (**BiLSTM**) (Kumar et al., 2022)	0.898	0.897	0.934	0.934	0.63	0.628	0.984	0.984	0.971	0.963
	MADS (**BiLSTM**) (Kumar et al., 2022)	0.851	0.850	0.505	0.496	0.641	0.640	0.998	0.998	0.969	0.960
Proposed	**MDS** ($\alpha = 0.5, \beta = 0.5$)	0.916	0.915	0.971	0.971	0.656	0.652	0.999	0.999	0.972	0.971
	MDS ($\alpha = 0.25, \beta = 0.25$)	**0.922**	**0.911**	0.975	0.975	**0.657**	**0.656**	**0.999**	**0.999**	**0.973**	**0.974**

Table 2: Comparing the performance of models trained over synthetic dataset and tested over real fake news samples. (i) **Acc** : indicates the accuracy, (ii) **T** and **F** indicates F-measure score for *True* news and *Fake* news class respectively. color indicates the best peromance.

		SM			NE-R		
	Model	**Acc**	**T**	**F**	**Acc**	**T**	**F**
BBC	**BiLSTM** (Kumar et al., 2022, 2023)	0.603	0.703	0.402	0.688	0.730	0.630
	RoBERT (Kumar et al., 2022, 2023)	0.775	0.722	0.811	0.705	0.673	0.732
	AHDE (Yoon et al., 2019)	0.699	0.572	0.768	0.691	0.723	0.650
	HDSF (Karimi and Tang, 2019b)	0.473	0.640	0.022	0.673	0.694	0.649
	HoBERT (Kumar et al., 2023)	0.783	0.734	0.817	0.753	0.691	0.794
	RaSHE (Kumar et al., 2023)	0.640	0.631	0.648	**0.756**	**0.692**	**0.799**
	GFNDS (Kim and Ko, 2021b,a)	0.500	0.667	0.502	0.500	0.666	0.498
	MADS (**BiLSTM**) (Kumar et al., 2022)	0.691	0.590	0.752	0.516	0.643	0.249
	MADS (**BiLSTM**) (Kumar et al., 2022)	0.716	0.643	0.765	0.500	0.666	0.498
	MDS ($\alpha = 0.5, \beta = 0.5$))	0.861	0.861	0.860	0.682	0.680	0.549
	MDS ($\alpha = 0.25, \beta = 0.25$)	**0.882**	**0.881**	**0.874**	0.651	0.649	0.645

From such observations, it can be concluded that combining both hierarchical encoding and dual summarization based improved the performance of the proposed model **MDS** over datasets of different natures. Also, our proposed model **MDS** is effective in incongruent and fake news detection.

6 Conclusions and Future Works

This study proposes *Multiset Dual Summarization* (**MDS**), which combines both hierarchical encoding and dual summarization approaches to take advantage of incongruent news detection. In our proposed model, we initially acquire hierarchical encoding of the news body. Subsequently, we divide the sentences of the news body into four distinct sets: head positive, head negative, body positive, and body negative. The news body sentences are placed into head positive and head negative based on the similarity between the headline and sentences of the news body. Similarly, the news body sentences are placed into body positive and body negative based on the similarity between the sentences and an encoded representation of the full news body. Subsequently, we obtain a summary representative vector for each set and estimate the similarity between the headline and summary representative vectors for incongruent news detection. We conducted experiments on datasets with varying characteristics, and our results indicate that our proposed model surpasses existing state-of-the-art baseline models. It efficiently detects both incongruent and partially incongruent news articles.

References

Hadeer Ahmed, Issa Traore, and Sherif Saad. 2017. 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*, pages 127–138. Springer.

Hadeer Ahmed, Issa Traore, and Sherif Saad. 2018.

Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1(1):e9.

Bence Bago, David G Rand, and Gordon Pennycook. 2020. Fake news, fast and slow: Deliberation reduces belief in false (but not true) news headlines. *Journal of experimental psychology: general*, 149(8):1608.

Luís Borges, Bruno Martins, and Pável Calado. 2019. Combining similarity features and deep representation learning for stance detection in the context of checking fake news. *Journal of Data and Information Quality (JDIQ)*, 11(3):1–26.

Sophie Chesney, Maria Liakata, Massimo Poesio, and Matthew Purver. 2017. Incongruent headlines: Yet another way to mislead your readers. In *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, pages 56–61, Copenhagen, Denmark. Association for Computational Linguistics.

Costanza Conforti, Mohammad Taher Pilehvar, and Nigel Collier. 2018. Towards automatic fake news detection: cross-level stance detection in news articles. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 40–49.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Ullrich KH Ecker, Stephan Lewandowsky, and Matthew Chadwick. 2020. Can corrections spread misinformation to new audiences? testing for the elusive familiarity backfire effect. *Cognitive Research: Principles and Implications*, 5:1–25.

Ullrich KH Ecker, Stephan Lewandowsky, Ee Pin Chang, and Rekha Pillai. 2014. The effects of subtle misinformation in news headlines. *Journal of experimental psychology: applied*, 20(4):323.

Ullrich KH Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K Fazio, Nadia Brashier, Panayiota Kendeou, Emily K Vraga, and Michelle A Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology*, 1(1):13–29.

Daniel A Effron and Medha Raj. 2020. Misinformation and morality: Encountering fake-news headlines makes them seem less unethical to publish and share. *Psychological science*, 31(1):75–87.

Maksym Gabielkov, Arthi Ramachandran, Augustin Chaintreau, and Arnaud Legout. 2016. Social clicks: What and who gets read on twitter? In *Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science*, SIGMETRICS '16, page 179–192, Antibes Juan-les-Pins, France. Association for Computing Machinery.

Edouard Grave, Piotr Bojanowski, Prakhar Gupta, Armand Joulin, and Tomas Mikolov. 2018. Learning word vectors for 157 languages. In *Proceedings of the International Conference on Language Resources and Evaluation (LREC 2018)*.

Andrew M Guess, Michael Lerner, Benjamin Lyons, Jacob M Montgomery, Brendan Nyhan, Jason Reifler, and Neelanjan Sircar. 2020. A digital media literacy intervention increases discernment between mainstream and false news in the united states and india. *Proceedings of the National Academy of Sciences*, 117(27):15536–15545.

Andreas Hanselowski, PVS Avinesh, Benjamin Schiller, and Felix Caspelherr. 2018a. Description of the system developed by team athene in the fnc-1, 2017. 1(1):03–13.

Andreas Hanselowski, Avinesh PVS, Benjamin Schiller, Felix Caspelherr, Debanjan Chaudhuri, Christian M. Meyer, and Iryna Gurevych. 2018b. A retrospective analysis of the fake news challenge stance-detection task. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 1859–1874, Santa Fe, New Mexico, USA. Association for Computational Linguistics.

Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation*, 9(8):1735–1780.

Benjamin Horne, Sara Khedr, and Sibel Adali. 2018. Sampling the news producers: A large news and feature data set for the study of the complex media landscape. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 12.

Christy Galletta Horner, Dennis Galletta, Jennifer Crawford, and Abhijeet Shirsat. 2021. Emotions: The unexplored fuel of fake news on social media. *Journal of Management Information Systems*, 38(4):1039–1066.

Divyanshu Kakwani, Anoop Kunchukuttan, Satish Golla, Gokul N.C., Avik Bhattacharyya, Mitesh M. Khapra, and Pratyush Kumar. 2020. IndicNLPSuite: Monolingual Corpora, Evaluation Benchmarks and Pre-trained Multilingual Language Models for Indian Languages. In *Findings of EMNLP*.

Hamid Karimi and Jiliang Tang. 2019a. Learning hierarchical discourse-level structure for fake news detection. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 3432–3442, Minneapolis, Minnesota. Association for Computational Linguistics.

Hamid Karimi and Jiliang Tang. 2019b. Learning hierarchical discourse-level structure for fake news detection. *arXiv preprint arXiv:1903.07389*.

Gihwan Kim and Youngjoong Ko. 2021a. Effective fake news detection using graph and summarization techniques. *Pattern Recognition Letters*, 151:135–139.

Gihwan Kim and Youngjoong Ko. 2021b. Graph-based fake news detection using a summarization technique. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3276–3280, Online. Association for Computational Linguistics.

Sujit Kumar, Durgesh Kumar, and Sanasam Ranbir Singh. 2023. Gated recursive and sequential deep hierarchical encoding for detecting incongruent news articles. *IEEE Transactions on Computational Social Systems*.

Sujit Kumar, Gaurav Kumar, and Sanasam Ranbir Singh. 2022. Detecting incongruent news articles using multi-head attention dual summarization. In *Proceedings of the 2nd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 12th International Joint Conference on Natural Language Processing*, pages 967–977.

Minh-Thang Luong, Hieu Pham, and Christopher D Manning. 2015. Effective approaches to attention-based neural machine translation. In *Proceedings of the 2015 Conference on Empirical Methods in Natural Language Processing*, pages 1412–1421.

Rahul Mishra, Piyush Yadav, Remi Calizzano, and Markus Leippold. 2020. Musem: Detecting incongruent news headlines using mutual attentive semantic matching. In *2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pages 709–716. IEEE.

Dean Pomerleau and Delip Rao. 2017. The fake news challenge: Exploring how artificial intelligence technologies could be leveraged to combat fake news. *Fake News Challenge*.

Benjamin Riedel, Isabelle Augenstein, Georgios P Spithourakis, and Sebastian Riedel. 2017. A simple but tough-to-beat baseline for the fake news challenge stance detection task. *arXiv preprint arXiv:1707.03264*, 1(1):1–6.

Julio Rieis, Fabrício de Souza, Pedro Vaz de Melo, Raquel Prates, Haewoon Kwak, and Jisun An. 2015. Breaking the news: First impressions matter on online news. In *Proceedings of the international AAAI conference on web and social media*, volume 9, pages 357–366.

Robiert Sepúlveda-Torres, Marta Vicente, Estela Saquete, Elena Lloret, and Manuel Palomar. 2021a. Exploring summarization to enhance headline stance detection. In *International Conference on Applications of Natural Language to Information Systems*, pages 243–254. Springer.

Robiert Sepúlveda-Torres, Marta Vicente, Estela Saquete, Elena Lloret, and Manuel Palomar. 2021b. Headlinestancechecker: Exploiting summarization to detect headline disinformation. *Journal of Web Semantics*, page 100660.

Percy H Tannenbaum. 1953. The effect of headlines on the interpretation of news stories. *Journalism Quarterly*, 30(2):189–197.

Yi Tay, Anh Tuan Luu, and Siu Cheung Hui. 2018. Hermitian co-attention networks for text matching in asymmetrical domains. In *IJCAI*, volume 18, pages 4425–31.

Yariv Tsfati, Hajo G Boomgaarden, Jesper Strömbäck, Rens Vliegenthart, Alyt Damstra, and Elina Lindgren. 2020. Causes and consequences of mainstream media dissemination of fake news: literature review and synthesis. *Annals of the International Communication Association*, 44(2):157–173.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Andrew Wang, Phillip Guo, and Ethan Wong. 2021. Fake news detection: Misleading headlines and satire. *International Journal of Computational and Biological Intelligent Systems*, 3(2).

Wei Wei and Xiaojun Wan. 2017. Learning to identify ambiguous and misleading news headlines. In *Proceedings of the 26th International Joint Conference on Artificial Intelligence*, pages 4172–4178.

Seunghyun Yoon, Kunwoo Park, Minwoo Lee, Taegyun Kim, Meeyoung Cha, and Kyomin Jung. 2021. Learning to detect incongruence in news headline and body text via a graph neural network. *IEEE Access*, 9:36195–36206.

Seunghyun Yoon, Kunwoo Park, Joongbo Shin, Hongjun Lim, Seungpil Won, Meeyoung Cha, and Kyomin Jung. 2019. Detecting incongruity between news headline and body text via a deep hierarchical encoder. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 791–800.

A Details of Experimental Hyperparameters

This study uses FastText (Grave et al., 2018) pre-trained word embeddings to represent words in the Hindi news article, and we used pre-trained IndicBERT IndicBERT (Kakwani et al., 2020) to encode sentences of news body, headline or news body. Table 4 presents the details of hyperparameters used to produce the results presented in this paper. Table 3 presents the characteristics of experimental datasets.

Table 3: Characteristics of Experimental Datasets

	Dataset	True	Fake	Total	#Head	#Body	#Para	#Sen
ISOT	Train	17083	18232	35315	9.438	244.325	3.799	16.955
	Test	1726	1815	5313	9.377	236.379	3.729	16.606
	Dev	2607	2706	3541	9.388	241.136	3.733	16.607
FNC	Train	40321	15161	55482	11.133	361.326	10.782	19.113
	Test	11039	4038	15077	8.503	365.027	10.950	19.331
	Dev	3533	1292	4825	11.174	363.417	10.916	19.203
NELA-17	Train	35710	35710	71420	10.558	551.923	13.494	26.649
	Test	3151	3151	6302	10.529	566.921	13.851	27.526
	Dev	3151	3151	6302	10.547	541.188	13.49	26.256
Split and Merge	Train	6242	3108	3134	7.427	383.64	5.756	21.851
	Test	1734	882	852	7.455	376.84	5.369	21.145
	Dev	694	2706	3541	9.388	241.136	5.713	22.407
NE-R	Train	6241	3128	3113	7.55	222.37	5.081	16.838
	Test	1734	862	872	7.403	273.13	5.134	17.163
	Dev	694	349	345	7.599	719.18	4.952	17.42
Real fake news	-	3984	1992	1992	9.656	291.86	3.021	10.066

Table 4: Details of hyperparameters used to produce results

Hyperparameters	**Values**
Epoch	40
Batch Size	50
Word Embedding Dimension	300
Sentence Embedding Dimension	768
Learning Rate	0.01
Loss Function	Cross Entropy
Number of Layers in MLP	2