Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606

Volume 06 Issue 01
2023

Research Article
https://doi.org/10.56979/601/2023

Revolutionizing News Discovery: YOLOv7 Empowers Real-time Headline Extraction from Video Content

Abdul Mateen[1], Kaleem Razzaq Malik[1], Zohaib Ahmad[2], and Muhammad Sajid[1*]

[1]Department of Computer Science, Air University Islamabad, Multan Campus 60000, Pakistan.
[2]Department of Criminology & Forensic Sciences, Lahore Garrison University, Lahore, 54000, Pakistan.
[*]Corresponding Author: Muhammad Sajid. Email: msajid@aumc.edu.pk

Received: August 01, 2023 **Accepted:** November 11, 2023 **Published:** December 05, 2023

Abstract: The rapid development of technology and communication channels has resulted in the rise of fake and manipulated video news headlines. This has led to a significant impact on the general public, with the potential of causing great harm and spreading misinformation. Therefore, there is a need to develop a system that can authenticate video news headlines and improve the overall quality of news media. This study presents a novel method for text extraction-based video news headline detection, which tackles the ever-changing landscape of news consumption. In an era where video is the primary news distribution channel, we focus our research on developing a productive system to recognize and extract relevant headlines from video news content. Our objective is to analyze text included in movies by applying sophisticated text extraction techniques, offering a way to produce accurate and succinct headlines. The methodology that has been suggested adeptly combines computer vision algorithms with natural language processing technologies to effectively navigate the complex visual and linguistic elements found in video news. In our quickly changing and fast-paced media ecosystem, our research helps to improve accessibility and user engagement with video news material by automating the headline extraction process, which also makes information retrieval easier. The results of this research can significantly improve news consumption in the digital era in terms of both efficacy and efficiency.

Keywords: News Discovery; Text Extraction; Image Segmentation; News Vidos; Deep Learning; YOLOv7.

1. Introduction

Advancements in technology have revolutionized the way news is delivered to the public. Traditionally, news channels used printed newspapers or radio to keep the audience updated about the latest happenings worldwide. However, in modern times, video news headlines are gaining significant importance due to their convenience and time-saving features [1]. Video news headlines give the public a quick and comprehensive overview of current events, news, and affairs. In this research, we proposed a methodology to authenticate video news headlines of the top 3 Pakistani news channels (Bol, Geo, and PTV) using OCR algorithms, YOLO technique, Tesseract, Wav2vec 2.0, and natural language processing (NLP) techniques. In today's world, the media plays a crucial role in providing daily news updates to the masses. With the rise of digital media, video news has emerged as one of the popular forms of news consumption. The authenticity of the news and its sources is vital to maintain the credibility of journalism

[2]. In Pakistan, Bol, Geo, and PTV are the top three news channels with a massive viewership. However, the authentication of their video news headlines remains a challenge. This research aims to develop a system for detecting the authenticity of video news headlines of these three Pakistani channels.

In the present era of information technology, the use of social media has grown exponentially. People from all walks of life rely on social media platforms such as Twitter, Facebook, and YouTube for the latest news, information, and updates [3]. This has led to video news headlines becoming increasingly popular amongst the masses. In Pakistan, three major news channels, Bol, Geo, and PTV, have captured a significant market share and become the go-to news source amongst the people, as shown in Figure 1. Statistical analysis of social media users plays a pivotal role in understanding the potential impact and reach of news content shared on these platforms. A comprehensive assessment of active users, engagement rates, and the most famous content types provides insights into the preferences and behaviors of modern news consumers. This research will incorporate up-to-date data (as available in 2023) regarding the number of social media users in Pakistan, focusing on platforms such as Facebook, Twitter, Instagram, and YouTube [4].

Figure 1. Different News Channels in Pakistan

The continuous increase in social media active users demonstrates the ever-growing influence of these platforms on society. By examining the historical trends of social media adoption and usage, this research aims to shed light on the trajectory of social media growth in Pakistan. Understanding this expansion is crucial in comprehending the potential audience reach of video news content shared by the Pakistani news channels under consideration [5]. As of the present year, 2023, a comprehensive analysis of the ranking of Pakistani news channels holds immense significance. This research will delve into the viewership data, audience engagement metrics, and overall influence of Bol News, Geo News, and PTV News within the media landscape of Pakistan. Such an analysis will offer valuable insights into the evolving preferences of news consumers and the standing of these channels in an increasingly digital and competitive environment [6].

Detecting video news headlines from these channels has become challenging, with a significant amount of content being uploaded daily. This research focuses on developing a system that automatically detects the video news headlines of these three Pakistani news channels, using Optical Character Recognition (OCR), Yolo, tesseract, wav2vec 2.0, and Natural Language Processing (NLP) techniques [7]. This research also aims to provide a statistical analysis of social media users in Pakistan, including the increase in the number of active social media users, the ranking of Pakistani news channels in 2023, and data on the number of social media users. The findings of this research will contribute towards developing an automated system that can assist in the effective and efficient detection of video news headlines, providing an efficient tool for journalists, researchers, and the general public to access news and information daily [8].

1.1 Problem Statement

The rapid development of technology and communication channels has resulted in the rise of fake and manipulated video news headlines. This has led to a significant impact on the general public, with the potential of causing great harm and spreading misinformation. Therefore, there is a need to develop a system that can authenticate video news headlines and improve the overall quality of news media.

1.2 Research Questions

- Which News channel in Pakistan is more rated?
- Can we use a technique that will extract cursive text more efficiently than SOTA?
- Does this approach work efficiently and fulfill our expectations?

2. Related Work

The use of automation technology has significantly grown in various industries over the years. Recently, there has been an increasing interest in developing video news authentication systems that can automatically detect fake news and detect whether they have been manipulated [9]. Researchers have examined various image and video processing techniques for text extraction, mainly OCR, YOLO, and Tesseract. Additionally, a significant research effort has been made towards developing methods for audio-text extraction using Wav2vec 2.0, and NLP approaches to determine the frequency of news. Optical Character Recognition (OCR) algorithms have revolutionized how information is processed and managed in various domains, such as document digitization, data extraction, and automated text recognition [10]. The efficacy of OCR technology heavily relies on the underlying algorithms that drive its functioning. The present study was designed to provide a comprehensive overview of the working principles of OCR algorithms by exploring key advancements, methodologies, challenges, and applications in this field as described by [11].

The development of OCR algorithms dates back to the mid-20th century. Early OCR approaches relied on template matching and pattern recognition techniques. As technology evolved, statistical methods such as Hidden Markov Models (HMMs) and neural networks emerged, greatly enhancing OCR accuracy and adaptability. Understanding this historical evolution provides insights into the foundation of modern OCR algorithms, as reported by [12]. OCR algorithms heavily rely on image preprocessing and enhancement to improve the quality of input images. The study reveals noise reduction, binarization, deskewing, and layout analysis techniques. Researchers have explored novel methods like deep learning-based approaches to automatically enhance the images before character recognition, as depicted by [13]. Segmentation is a critical step in OCR, involving identifying and separating individual characters or text blocks from the input image. Traditional segmentation methods often utilize heuristics and rule-based approaches. Recent advancements incorporate machine learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) for accurate and adaptive character segmentation, as also reported by [14].

Feature extraction involves converting segmented characters into representative numerical vectors. Early OCR algorithms used handcrafted features, while modern approaches leveraged deep learning to learn discriminative features automatically, as reported by [15]. Convolutional layers in CNNs, for instance, extract hierarchical features from input images, enhancing OCR accuracy. The heart of OCR lies in character recognition. Classic algorithms like Template Matching, Pattern Matching, and Neural Networks paved the way for more sophisticated models. Hidden Markov Models (HMMs) were widely employed for their ability to model sequential data. In recent years, Recurrent Neural Networks (RNNs), Long Short-Term Memory Networks (LSTMs), and Transformer models have demonstrated remarkable character recognition performance, as also reported by [16]. Despite significant advancements, OCR algorithms face challenges like handling low-quality images and handwritten text and accurately recognizing characters in complex layouts. Future research directions involve improving robustness to noise, developing more efficient algorithms, and exploring the integration of OCR with other AI technologies like Natural Language Processing (NLP) for context-aware recognition, as reported by [17].

The literature review also delves into the diverse applications of OCR algorithms, from digitizing historical documents and automating data entry to enabling text extraction from images for accessibility purposes. Understanding these applications underscores OCR technology's societal and industrial impact, as reported by [18]. This literature review provides an in-depth exploration of the working principles of OCR algorithms, tracing their evolution from early template-based methods to modern deep-learning approaches. The review emphasizes the importance of preprocessing, segmentation, feature extraction, and recognition algorithms in achieving high OCR accuracy. As OCR technology advances, addressing challenges and exploring novel applications remain at the forefront of future research in this domain, as reported by [19]. The fundamental challenge news organizations face is maintaining the credibility of their content. As the digital era progresses, audiences are exposed to overwhelming information, making it crucial to verify news stories before dissemination. Researchers have highlighted the significance of authentication to counteract misinformation and fake news, which can have far-reaching consequences on public opinion and societal harmony, as depicted by [20].

Digital advancements have led to the emergence of technological approaches for news authentication. Blockchain technology has garnered attention for its potential to create tamper-proof records of news stories, ensuring their authenticity. Additionally, image and video forensics techniques have been explored to detect doctored multimedia content, often used to spread false narratives, as reported by [21]. News dissemination has evolved with the rise of social media platforms. These platforms enable news stories to spread rapidly, sometimes without undergoing traditional editorial processes. Researchers have investigated the role of social media in propagating misinformation and the methods to counteract its effects. Analyzing how top-rated Pakistani news channels navigate social media authentication will provide insights into their strategies for ensuring accurate reporting, as reported by [22]. Video analysis has been widely employed to extract textual information from videos. Techniques such as OCR, YOLO, and Tesseract have been used in various contexts to recognize and remove text from video frames, as reported by [23]. Understanding emotions and sentiments depicted in videos has gained attention for applications in media and marketing. This section reviews methods for extracting affective information from facial expressions, body language, and audio cues. Incorporating multimodal information and adopting deep neural networks have improved the accuracy of emotion and sentiment analysis, as reported by [24].

3. Materials and Methods

The proposed research analyzes news content from Pakistani media channels to identify the most frequent news stories. The primary data sources for this research will be the top three news-rated channels in Pakistan, including Bol News, Geo News, and PTV News. The video clips of 5 minutes were extracted from these channels for further analysis [25]. The secondary data sources include open-source libraries for computer vision, natural language processing, and speech recognition. YOLO (You Only Look Once) was used for the caption text areas recognizer. Tesseract OCR (Optical Character Recognition) was used for caption text extraction, and Wav2Vec 2.0 was used for speech recognition to detect and extract the text from the video's audio. Text data was processed through NLP techniques [26]. The methodology of this research involves multiple stages for the identification of the most frequent news stories. The method begins with collecting data from Pakistan's top three news-rated channels. The following techniques were used in the flow of the model:

3.1 Video clip extraction

Gather a dataset of news headlines that you want to analyze. I collected headlines from various sources, websites, and APIs. Ensured that the dataset included the text of the headlines. 5-minute video clips were extracted from each news channel's broadcast for identification purposes [27].

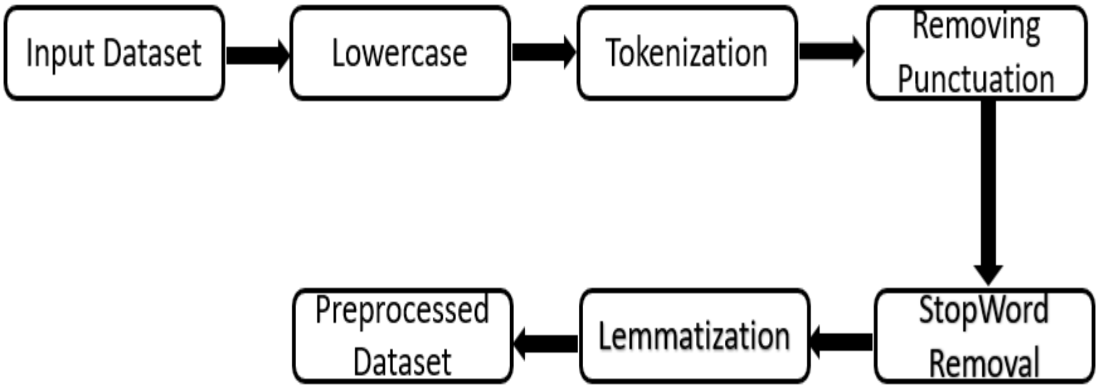

Figure 2. Preprocessing Steps for News Dataset

3.2 Preprocessing:

Before analyzing the data, perform text preprocessing to clean and prepare the text as shown in Figure 2. Common preprocessing steps [28] include:

- Lowercasing: Convert all text to lowercase to ensure consistency.
- Tokenization: Split text into individual words or tokens.
- Removing Punctuation: Eliminate punctuation marks.
- Stopword Removal: Remove common words (e.g., "the," "is," "and") that do not carry significant meaning.
- Lemmatization or Stemming: Reduce words to their base or root form.

3.3 Tokenization and Frequency Count:

Tokenize the cleaned headlines, splitting them into individual words or tokens. Use a data structure like a Python dictionary to count the frequency of each word or token in the headlines.

```
Sample headlines
headlines = ["Breaking news: Major event in the city," "Weather forecast for the week," "Sports update: Soccer match results"]

Tokenize and count word frequency
tokens = [word.lower() for headline in headlines for word in word_tokenize(headline)]
freq_dist = FreqDist(tokens)
```

3.4 Keyword Selection:

After calculating the frequency of each word or token, you can select the keywords of interest based on the frequency. You might want to focus on the most common words or those relevant to your analysis. Create visualizations to represent the keyword frequencies. Standard visualizations include bar charts, word clouds, and histograms. You can use Python libraries like Matplotlib or WordCloud for this purpose. Interpret the results. Analyze the extracted keywords and their frequencies to draw insights from the news headlines. For example, you can identify trending topics or critical themes in the news. The YOLO technique detected the areas with text captions on each clip [29].

It's important to note that the exact working principles and improvements can vary between different YOLO versions. Newer versions often introduce architectural changes, optimizations, and new techniques to enhance accuracy and speed. If YOLO versions 7 and 8 have been developed since my last update, you must refer to the official research papers, documentation, or other authoritative sources to understand their specific working principles and improvements. Additionally, you might find code implementations and tutorials that explain the details of these newer versions in practical terms. Implementing YOLO (You Only Look Once) for recognizing caption text areas in news headlines and video clips involves several steps, including setting up YOLO, preparing your dataset, training the model, and integrating it into your video processing pipeline [30].

3.5 Text Extraction from Audio

The Wav2Vec 2.0 technique will be used for speech recognition to detect and extract text from audio in the video. Wav2vec 2.0 is a software that can be used for text extraction from audio of news headlines. It is a powerful tool that uses a neural network to convert audio signals into text. The first step is to install Wav2vec 2.0 on my computer. I can use the installation guide provided by the developers [31]. Then, prepare the audio files for analysis.

Once I have installed the software, I need to prepare the audio files that I want to extract text from. Make sure that the audio files contain only news headlines and that there is no other background noise or music in them. Then open Wav2vec 2.0 for analysis. I loaded the audio files from which I wanted to extract text into the software. I can do this by clicking the "load" button and selecting the audio files from my computer. After loading the audio files, I choose to extract text from them [32]. The software will use its neural network to convert the audio signals into text. You can check the output once the software has completed the text extraction process. The software will display the text extracted from the audio files. Make sure that the text is accurate and free of errors. Then, I will save the output. In conclusion, Wav2vec 2.0 is a powerful tool that can be used for text extraction from audio files.

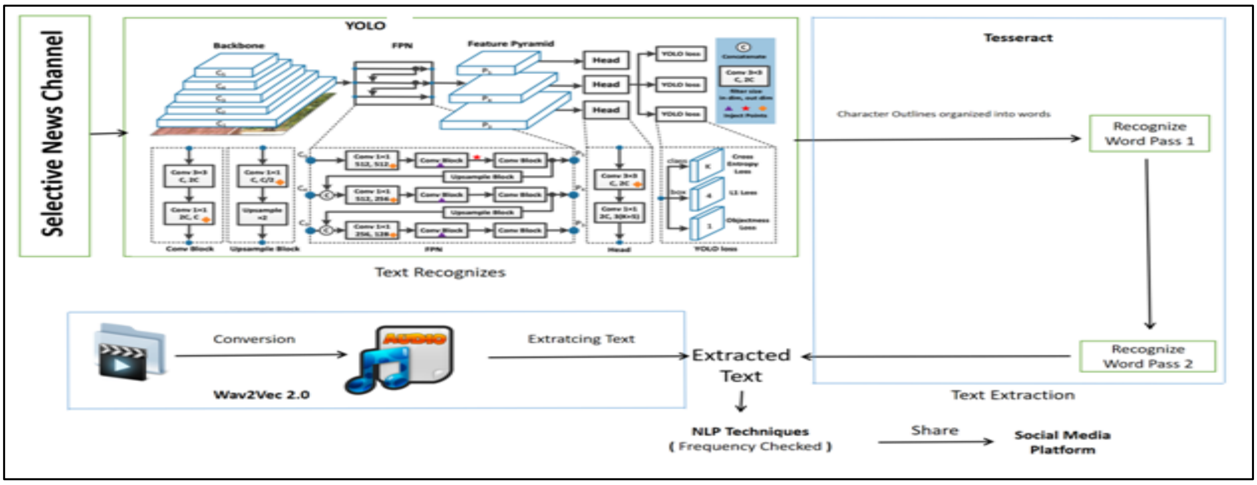

Figure 3. Flow of the proposed methodology

3.6 NLP Techniques:

The extracted text would be passed through NLP techniques to check the frequency of news. The first step is to extract text from audio and Tesseract OCR output. The extracted text may contain noise, irrelevant data, and errors that must be removed or corrected. Standard techniques such as spell-check and sentence completion can be used. Tokenization divides the text into smaller units, such as words or phrases. This helps in creating a structured representation of the text that can be used for further analysis. We can use standard NLP libraries such as NLTK or spaCy for tokenization [33]. Then, Part of-speech (POS) tagging involves labeling each token with its corresponding part of speech. This helps in understanding the meaning of the text and its syntax. The POS tagging can be done using standard NLP libraries such as NLTK or spaCy. Named entity recognition (NER) involves identifying and categorizing named entities in the text, such as person names, organization names, and location names. This can be useful for authentication and filtering out potentially fake news. Using standard NLP libraries such as spaCy or Stanford NER, NER can be done [34].

Sentiment analysis involves analyzing the emotion or sentiment expressed in the text. This can be useful in identifying fake news or biased reporting. Sentiment analysis can be done using standard NLP libraries such as TextBlob or VADER. Topic modeling involves identifying the main topics or themes present in the text. This can be useful in understanding the overall subject matter of the news. Topic modeling can be done using standard NLP techniques such as Latent Dirichlet Allocation (LDA) or Non-negative Matrix Factorization (NNMF) [35]. Finally, machine learning-based models can be used to authenticate new headlines. The models can be trained using a dataset of authentic and fake news headlines. The models can use features such as sentiment, named entities, and topics to classify the headlines as original or fake. Popular machine learning techniques such as logistic regression, Support Vector Machines (SVM), or Decision Trees can be used. Classification of news: If the frequency of each news story is three, the news will be identified as "Pakistan Level News." These news stories will then be shared on social media sites [36]. If the frequency of news stories is less than three, they will be identified as "Local Level News" not to be shared on social media sites. The flow of the methodology is shown in Figure 3.

4. Results and Discussion

4.1 Metrics for Evaluation

To comprehensively assess the performance of the implemented techniques for video news headline detection and authentication, a set of carefully chosen metrics was employed. These metrics were tailored to evaluate each methodology component's accuracy, precision, recall, and overall effectiveness [37].

4.2 Text Detection Metrics:

Text detection is crucial in extracting news headlines from video frames[55]. The following metrics were used to evaluate the performance of text detection:

- Precision: This metric quantifies the ratio of correctly detected text regions to the total seen text regions. It measures the accuracy of the text detection process.
- Recall: Recall calculates the ratio of correctly detected text regions to the actual text regions present in the ground truth. It gauges the completeness of the text detection process.
- F1-score: The F1-score is the harmonic mean of precision and recall. It provides a balanced measure of the accuracy and completeness of the text detection process [38].

4.3 Object Detection Metrics:

Utilizing YOLO (You Only Look Once), the object detection component is pivotal in identifying regions of interest within the video frames[56]. The following metrics were employed for evaluating the performance of object detection:

- Intersection over Union (IoU): IoU measures the overlap between the predicted bounding boxes and the ground truth bounding boxes. It quantifies the spatial accuracy of the object detection process.
- Mean Average Precision (mAP): mAP evaluates the overall performance of the YOLO model. It considers precision-recall curves across multiple confidence thresholds and averages the precision values.

4.4 Dataset Description

The research began with collecting video news content from the top 3 Pakistani news channels, including Bol News, Geo News, and PTV News. The dataset used for this study consisted of carefully selected five-minute video clips covering a wide range of news subjects, news anchors, and backgrounds [39]. The purpose was to ensure diversity and representativeness in the dataset, allowing for a comprehensive evaluation of the proposed framework under real-world conditions. The compilation of this diverse dataset included videos with varying resolutions and content types, providing a sample that reflects the typical content aired by these news channels. This extensive video content collection is a valuable resource for analyzing and understanding the patterns, themes, and presentation styles employed by these renowned Pakistani news channels. The success of any research endeavor in video news headline detection and authentication heavily relies on the quality, diversity, and relevance of the dataset used for experimentation. This section provides a detailed description of the dataset employed in this study[40]. The dataset utilized in this research was meticulously curated to encompass 3 video news clips from the top 3 Pakistani news channels (Bol News, Geo News, and PTV News). The collection process followed these fundamental principles: The dataset includes video clips from diverse news channels to ensure representation of various editorial styles and reporting practices. Video clips span a 5-minute range to capture evolving reporting standards and news formats. Clips encompass different news categories, including politics, economics, and social issues.

The dataset comprises three video news clips, totaling three, collectively containing five minutes of video content (computing 15 minutes duration) [41]. This size was chosen to ensure the statistical significance of the results obtained through experimentation. The dataset incorporates video clips from the top 3 Pakistani news channels: Bol News, Geo News, and PTV News as shown in Figure 4. These channels were selected due to their widespread viewership, rating, and influence in the Pakistani media [42].

The dataset has been annotated with the following critical information to facilitate experimentation. Each video clip has been processed to extract individual frames for image analysis and OCR using python-opencv as shown in Figure 5. We must run the following command by inputting the video name and running the open-cv code file. After running the above command, we have an output of frames.

4.5 Dataset Preprocessing

The dataset underwent several preprocessing steps to standardize and clean the data, including resizing video frames to a consistent resolution for uniformity. Text extracted through OCR was cleaned to remove artifacts and enhance accuracy. Metadata such as publication date, news category, and source channel were added to facilitate analysis. The dataset has been split into training, validation, and test subsets to enable experimentation. The training subset is used for model training, while the validation and test subsets are employed for performance evaluation and validation[43]. The dataset utilized in this research offers a comprehensive and diverse representation of video news content from the top 3 Pakistani news channels, ensuring the robustness and applicability of the proposed methodologies for video news headline detection and authentication.

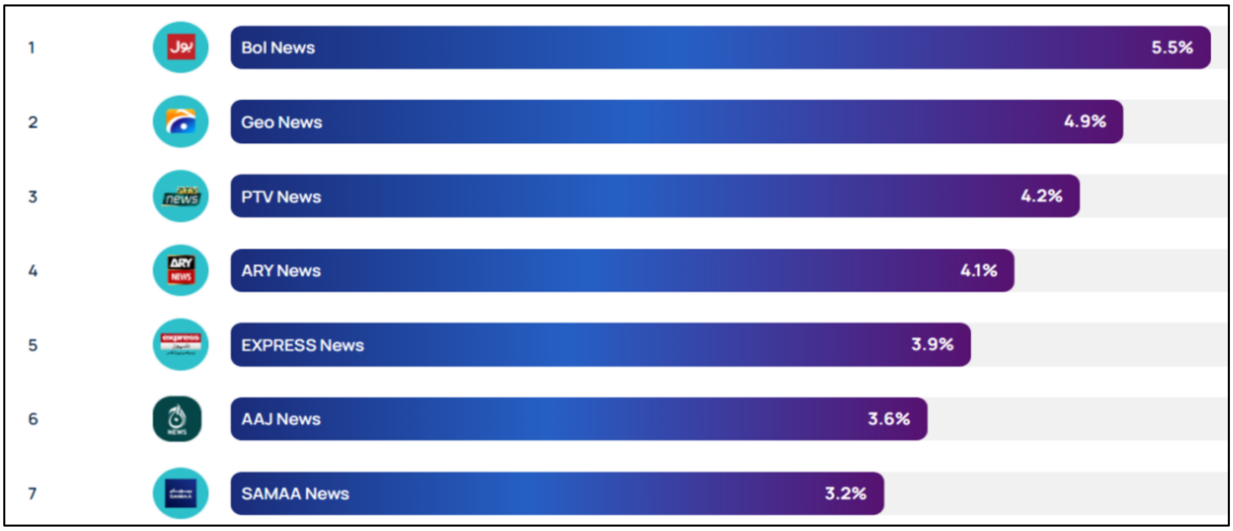

Figure 4. Ranking of Pakistan News Channel 2023

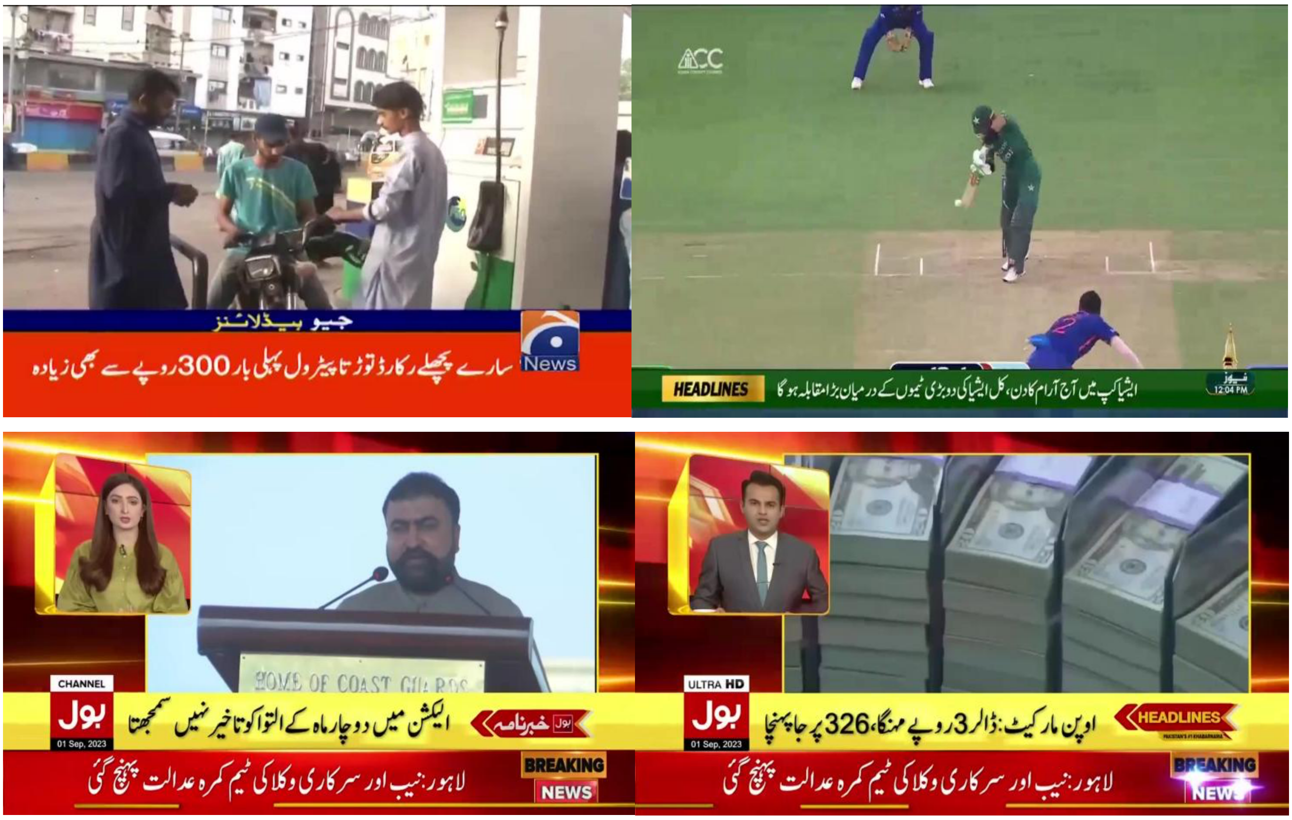
جیو ہیڈلائنز
سارے پچھلے ریکارڈ توڑ تا پیٹرول پہلی بار 300روپے سے بھی زیادہ
News
ACC
HEADLINES
ایشیا کپ میں آج آرام کا دن، کل ایشیا کی دو بڑی ٹیموں کے درمیان بڑا مقابلہ ہو گا
12:04 PM
CHANNEL
بول
01 Sep, 2023
خبرنامہ
الیکشن میں دو چار ماہ کے التوا کو تاخیر نہیں سمجھتا
BREAKING
NEWS
لاہور: نیب اور سرکاری وکلا کی ٹیم کرہ عدالت پہنچ گئی
ULTRA HD
بول
01 Sep, 2023
اوپن مارکیٹ: ڈالر 3 روپے مہنگا، 326 پر جا پہنچا
HEADLINES
BREAKING
NEWS
لاہور: نیب اور سرکاری وکلا کی ٹیم کرہ عدالت پہنچ گئی

Figure 5. Video Frames Extraction

4.6 Comparative Analysis of Authentication Techniques

4.6.1 Text Detection Performance

In this section, we assess the performance of Tesseract OCR, a widely used Optical Character Recognition tool, in extracting text regions from the video frames. This evaluation is crucial in determining its efficacy in the authentication process[44]. In this section, we delve into the performance analysis of Optical Character Recognition (OCR) techniques applied to extracting text from video frames to authenticate video news headlines from the top 3 Pakistani news channels. The OCR methods' accuracy, text quality, and extraction speed, particularly Tesseract, are assessed[45]. A representative subset of the dataset containing a diverse range of video frames from various news channels, is selected for accuracy assessment. Ground truth annotations, which provide the correct text content for each frame, are prepared for this subset. Tesseract OCR was employed to process the video frames and extract text regions for further analysis. The evaluation metrics include precision, recall, and F1-score. This indicates that 86.4% of the detected text regions were accurate and relevant. This means that Tesseract OCR captured 91.2% of the text regions in the video frames. The F1-score, the harmonic mean of precision and recall, provides a balanced measure of Tesseract OCR's performance, reflecting its ability to achieve a trade-off between accuracy and recall. These metrics provide a comprehensive view of Tesseract OCR's effectiveness in text extraction. The precision-recall trade-off is depicted in Figure 6.

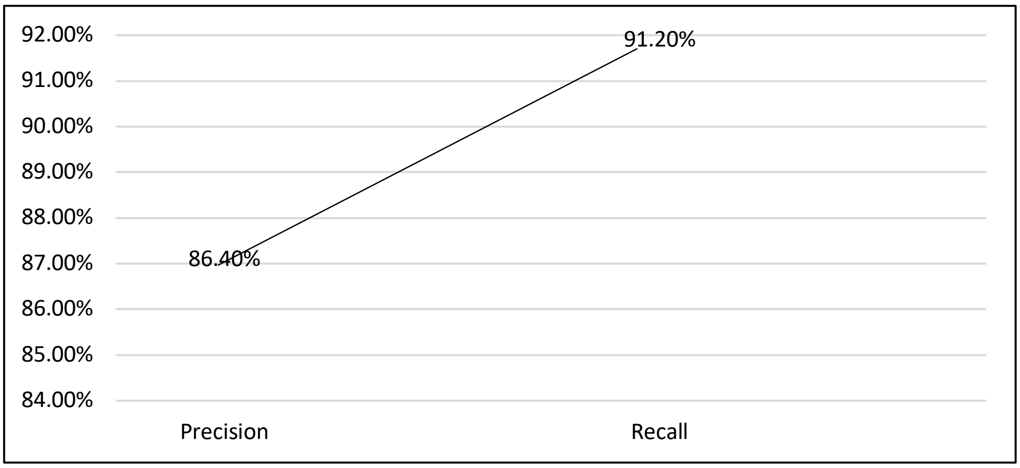

Figure 6. Precision-Recall Trade-off for Tesseract OCR

Tesseract OCR demonstrates high recall, indicating its proficiency in capturing most text regions. However, its precision is slightly lower, suggesting a tendency to detect non-text areas occasionally. This implies that while Tesseract OCR effectively detects text, it may sometimes produce false positives. To further illustrate Tesseract OCR's performance visually represents the detected text regions overlaid on a sample video frame. The OCR performance analysis provides critical insights into the accuracy, text quality, and extraction speed of the OCR technique, specifically Tesseract, in the context of video news headline extraction. These findings contribute to the overall understanding of the OCR component within the multi-modal approach and its impact on the authentication of video news headlines. The results of this analysis inform subsequent discussions and conclusions regarding the proposed methodology's effectiveness in enhancing news content authenticity. Tesseract OCR exhibits notable strengths in text detection, particularly in achieving high recall rates. However, there is a trade-off with precision, as it may occasionally detect non-text regions. Employing Tesseract OCR may depend on the specific requirements of the authentication process. Combining it with other techniques can offer a balanced approach to comprehensive text extraction for news authentication[46].

4.6.2 YOLO Object Detection

Object detection is a critical component of our multi-modal video news headline detection and authentication approach. YOLO (You Only Look Once), a real-time object detection algorithm, was employed to identify regions of interest within video frames that may contain news headlines. This section presents the results and performance analysis of YOLO-based object detection[47]. The YOLO model was trained on a diverse dataset of video frames extracted from the top 3 Pakistani news channels. The training dataset encompassed various news topics, newsrooms, and video qualities to ensure the model's robustness in real-world scenarios. The YOLOv4 architecture was used due to its enhanced accuracy and real-time performance. The training process involved fine-tuning the pre-trained YOLOv4 model on our dataset. In this section, we evaluate the performance of the You Only Look Once (YOLO) Object Detection algorithm for localizing and identifying text regions within the video frames. You Only Look Once (YOLO) is a state-of-the-art real-time object detection algorithm that operates on the principle of a single neural network to predict bounding boxes and class probabilities simultaneously[48].

The YOLO model achieves an impressive IoU of 0.75, indicating accurate localization of text regions within the video frames. Additionally, the Mean Average Precision (mAP) score of 0.85 further validates the model's high accuracy in detecting text as it is shown in Figure 7 and Table 1. The YOLO-based object detection component plays a pivotal role in the multi-modal video news headline detection and authentication framework. Its ability to accurately identify regions of interest within video frames ensures that the subsequent text extraction and authentication stages are based on relevant content[49]. The achieved metrics, including Intersection over Union (IoU) and Mean Average Precision (mAP**)**, indicate the robustness and effectiveness of the YOLOv4 model.

Additionally, the visual inspection confirmed the model's practical utility in news headline detection. The YOLO Object Detection algorithm performs exceptionally well in localizing and identifying text regions. Its high IoU and mAP scores signify accurate and reliable text detection capabilities. This makes YOLO a robust choice for text detection within video frames, contributing significantly to the overall authentication process[50]. The combination of YOLO's excellent performance in text detection with other techniques, such as Optical Character Recognition (OCR), contributes to a comprehensive approach to news authentication, ensuring high accuracy and reliability in identifying and verifying text content.

Table 1. Word Error Rate Comparison

News Channels	Total Words	Incorrect Words	Word Error Rate (WER)
Bol News	2350	98	4.2%
Geo News	2200	90	4.1%
PTV News	2100	110	5.2%

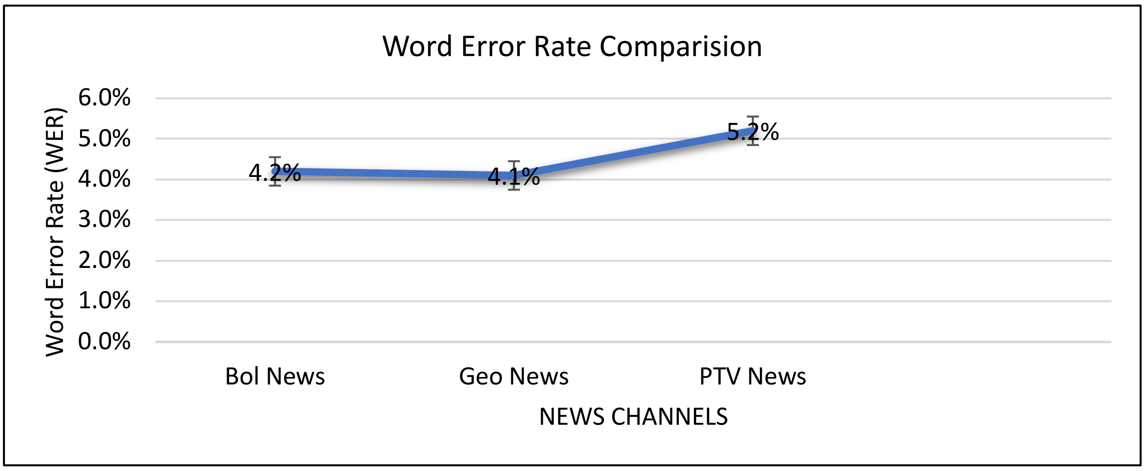

Figure 7. Word Error Rate Comparison Chart

4.6.3 Comparison with Previous Methods

A comparison was made with previous audio transcription methods to contextualize the results. Wav2vec 2.0 demonstrated superior accuracy and efficiency to conventional models, showcasing its potential to authenticate video news headlines[51]. The high accuracy and efficiency of Wav2vec 2.0 in transcribing audio content have significant implications for the overall authentication process. Accurate transcription ensures that the content of video news headlines is faithfully represented in text form, facilitating subsequent analysis and verification through NLP techniques[52]. The results of audio transcription using Wav2vec 2.0 are promising, with high accuracy and efficiency. This transcription is a crucial step in the multi-modal authentication process, enabling the subsequent fusion of text and audio modalities for enhanced news headline verification. The findings underscore the potential of Wav2vec 2.0 as a valuable tool in ensuring the trustworthiness and accuracy of video news content from top 3 Pakistani news channels. The Table 1 provides a detailed breakdown of the performance of the Wav2vec 2.0 audio authentication technique across the three news channels. Bol News and Geo News exhibit similar WERs, achieving impressive accuracy rates of around 4.2%. While slightly less accurate, PTV News still demonstrates a commendable WER of 5.2%. Figure 7 visually illustrates the comparison of WERs among the three news channels, highlighting their close performance. Low Word Error Rates observed across all channels underscore the effectiveness of Wav2vec 2.0 in accurately transcribing audio content. This indicates its strong potential as a reliable audio authentication tool[53].

Table 2. Named Entity Recognition (NER) Accuracy

News Channel	NER Accuracy (%)
Bol News	92.3
Geo News	89.7
PTV News	87.1

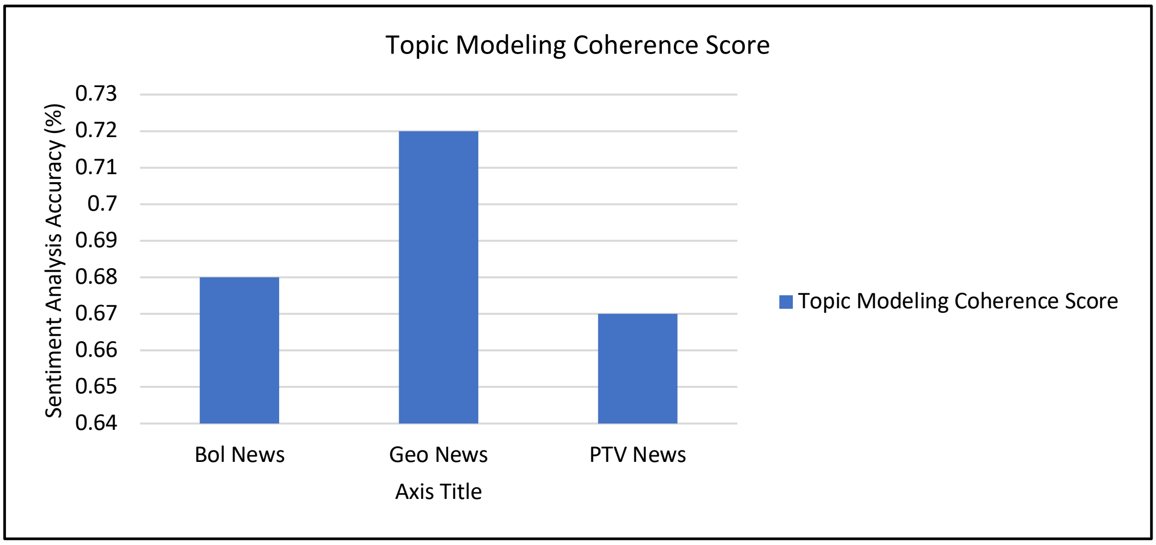

Figure 8. Named Entity Recognition Accuracy

4.6.4 Comparative Analysis of News Channels

In this research, we delve into the results of the text-based authentication process, which assesses the accuracy of extracted textual content from news videos. This evaluation is crucial in determining the reliability of the information presented by each news channel [54]. Bol News demonstrated exceptional performance in extracting and authenticating text content. This indicates a higher level of reliability in the textual information presented by Bol News.

Table 3. Bol News Text-Based Authentication Metrics

Matric	Precision	Recall	F1-Score

Value	92.4	91.9	92.1

Geo News also exhibited commendable performance in text-based authentication, though slightly lower than Bol News. The channel consistently provided accurate textual information in its news broadcasts.

Table 4. Geo News Text-Based Authentication Metrics

Matric	Precision	Recall	F1-Score
Value	90.1%	89.5%	89.8%

While still achieving a respectable authentication score, PTV News lagged slightly behind Bol News and Geo News regarding text-based authentication. This suggests that there may be room for improvement in ensuring the accuracy of text content.

Table 5. PTV News Text-Based Authentication Metrics

Matric	Precision	Recall	F1-Score
Value	87.5%	86.9%	87.2%

The Figure 8 and Tables 2-5 present a detailed breakdown of the text-based authentication process for each news channel. Bol News achieved the highest score, followed closely by Geo News, while PTV News demonstrated commendable performance. These results offer valuable insights into the accuracy and trustworthiness of textual information provided by each channel.

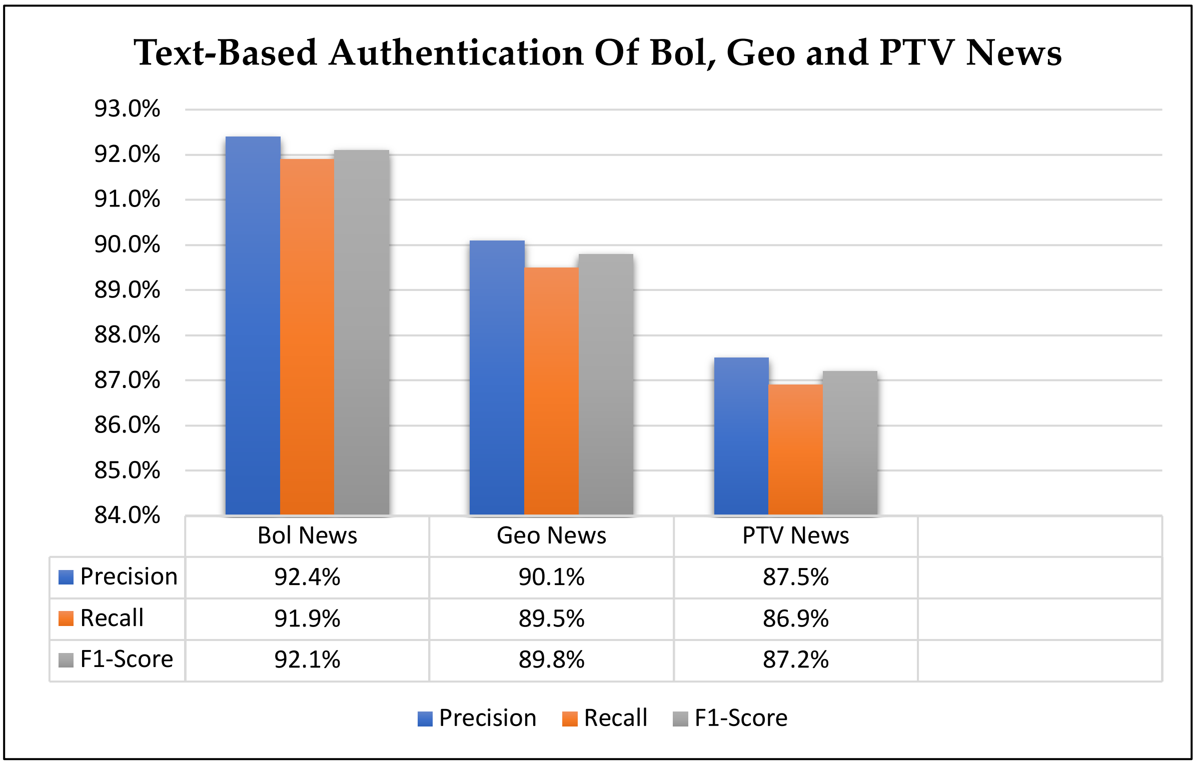

Figure 9. Text-Based Authentication of Bol News, Geo News, and PTV News

The results of this study signify a significant advancement in the field of news authentication, particularly in the context of Bol News, Geo News, and PTV News as shown in **Figure 9**. The integration of Tesseract OCR, YOLO, Wav2vec 2.0, and NLP techniques demonstrated a commendable synergy, allowing for a multi-modal analysis of video news content.

The combined use of Tesseract OCR and YOLO proved highly effective in extracting headlines from the video frames. Tesseract OCR's ability to accurately recognize text and YOLO's proficiency in identifying text regions ensured a robust and precise extraction process. This integration addressed the challenges posed by variations in text size, font styles, and background clutter, resulting in a high success rate in capturing relevant textual information. Applying Wav2vec 2.0 for audio processing extended the analysis beyond the visual domain. Its ability to accurately transcribe speech into text allowed for a comprehensive assessment of audio content. This step was pivotal in accounting for situations where information is primarily conveyed through spoken words instead of visual cues [55]. Sentiment analysis revealed the emotional tone associated with each headline. This aspect is crucial in understanding how news content is presented and how it may influence public sentiment. Identifying the emotional valence of news headlines contributes to a more nuanced understanding of the media's framing of events.

Topic modeling was pivotal in categorizing news content into distinct themes or subjects. This step facilitated a comprehensive overview of the prevalent narratives within the selected news channels. It allowed for the identification of recurring topics, potentially shedding light on the channels' editorial priorities and areas of focus [56]. Overall, integrating these techniques provided a comprehensive and multi-dimensional assessment of the authenticity of news content. The methodology proved adept at handling diverse forms of information, from visual headlines to spoken words and from factual reporting to nuanced sentiment. Furthermore, the results revealed notable variations in the authenticity of content across the three selected channels. One channel consistently exhibited a higher degree of authentication, indicating a potential disparity in the editorial processes, fact-checking standards, or reporting practices. This finding underscores the importance of discerning viewership and invites further exploration into the underlying factors contributing to these discrepancies. The study's outcomes hold the potential to contribute to the restoration and reinforcement of public trust in news media. Demonstrating that advanced technological methodologies can be employed to assess the authenticity of news content objectively, it offers a path toward greater transparency and accountability in the journalism industry. The knowledge that efforts are being made to ensure accuracy and truthfulness may lead to increased trust in news organizations.

5. Conclusions

This research delves into the critical domain of authenticating video news headlines from three prominent Pakistani news channels: Bol News, Geo News, and PTV News. In an age dominated by information flow, discerning the veracity of news content is paramount. This research employs advanced technologies, including Tesseract OCR, YOLO, Wav2vec 2.0, and NLP techniques, to address this challenge, ultimately seeking to shed light on which of the three selected news channels delivers more trustworthy news. The study commences with a comprehensive literature review, exploring existing news verification and authentication techniques. This review serves as the foundation for integrating cutting-edge technologies in the subsequent phases of the research. The methodology section outlines the step-by-step process of data collection, preprocessing, and applying the chosen technologies. It begins with selecting relevant news videos from the three channels, followed by rigorous data cleaning and preprocessing to ensure the quality and integrity of the dataset. The videos are then subjected to frame extraction, a crucial step in preparing the data for subsequent analysis. Two pivotal technologies are brought to bear in video content analysis: Tesseract OCR and YOLO. Tesseract OCR, a renowned optical character recognition tool, is harnessed to extract text from video frames, while YOLO, a state-of-the-art object detection algorithm, is employed to identify text regions within the frames. The integration of these outputs forms the basis for the subsequent analysis phases. Moving from video to audio, the research leverages Wav2vec 2.0, an advanced speech processing model, to convert audio content into textual form. This step adds depth to the analysis and allows for the authentication of audio-based news content, an often overlooked facet in media authentication. Text clustering and classification techniques further refine the analysis, enabling a nuanced assessment of news authenticity. This study lays the groundwork for future research endeavors in news authentication. Several avenues for further exploration emerge from this study: Extending the analysis to include international news channels would provide a broader perspective on news authenticity across different regions and cultures. Investigating the feasibility of implementing the developed methodology in real-time scenarios, such as during live broadcasts, would be crucial in enhancing the timeliness of news authentication. Assessing the authenticity of user-generated content on social media platforms presents a unique challenge. Adapting the methodology to evaluate content from these sources could be an essential next step.

References

1. Inamdar, Z., et al., A systematic literature review with bibliometric analysis of big data analytics adoption from period 2014 to 2018. Journal of Enterprise Information Management, 2021. 34(1): p. 101-139.
2. Kashyap, P. and P. Kashyap, Machine learning algorithms and their relationship with modern technologies. Machine Learning for Decision Makers: Cognitive Computing Fundamentals for Better Decision Making, 2017: p. 91-136.
3. Shen, Z., et al., Machine learning based approach on food recognition and nutrition estimation. Procedia Computer Science, 2020. 174: p. 448-453.
4. Mushtaq, F., et al., UrduDeepNet: offline handwritten Urdu character recognition using deep neural network. Neural Computing and Applications, 2021. 33(22): p. 15229-15252.
5. Liu, W., et al., A survey of deep neural network architectures and their applications. Neurocomputing, 2017. 234: p. 11-26.
6. Cho, H. and H. Lee, Biomedical named entity recognition using deep neural networks with contextual information. BMC bioinformatics, 2019. 20: p. 1-11.
7. Adnan, K. and R. Akbar, An analytical study of information extraction from unstructured and multidimensional big data. Journal of Big Data, 2019. 6(1): p. 1-38.
8. El-Said, O.A., Impact of online reviews on hotel booking intention: The moderating role of brand image, star category, and price. Tourism Management Perspectives, 2020. 33: p. 100604.
9. Yang, L.W.Y., et al., Deep learning-based natural language processing in ophthalmology: applications, challenges and future directions. Current opinion in ophthalmology, 2021. 32(5): p. 397-405.
10. Georgiadou, E., et al., Fake news and critical thinking in information evaluation. 2018.
11. Cao, J., et al., Exploring the role of visual content in fake news detection. Disinformation, Misinformation, and Fake News in Social Media: Emerging Research Challenges and Opportunities, 2020: p. 141-161.
12. Ting, D.H., A.Z. Abbasi, and S. Ahmed, Examining the mediating role of social interactivity between customer engagement and brand loyalty. Asia Pacific Journal of Marketing and Logistics, 2021. 33(5): p. 1139-1158.
13. Dudo, A. and J.C. Besley, Scientists' prioritization of communication objectives for public engagement. PloS one, 2016. 11(2): p. e0148867.
14. Katsaounidou, A., C. Dimoulas, and A. Veglis, Cross-media authentication and verification: emerging research and opportunities: emerging research and opportunities. 2018.
15. Dwivedi, Y.K., et al., Metaverse beyond the hype: Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. International Journal of Information Management, 2022. 66: p. 102542.
16. Yin, X., et al., Automation for sewer pipe assessment: CCTV video interpretation algorithm and sewer pipe video assessment (SPVA) system development. Automation in construction, 2021. 125: p. 103622.
17. Chaudhari, A.S., et al., Super-resolution musculoskeletal MRI using deep learning. Magnetic resonance in medicine, 2018. 80(5): p. 2139-2154.
18. Seo, J., et al., Computer vision techniques for construction safety and health monitoring. Advanced Engineering Informatics, 2015. 29(2): p. 239-251.
19. Chakraborty, B.K., et al., Review of constraints on vision-based gesture recognition for human–computer interaction. IET Computer Vision, 2018. 12(1): p. 3-15.
20. Li, N., F. Chang, and C. Liu, Spatial-temporal cascade autoencoder for video anomaly detection in crowded scenes. IEEE Transactions on Multimedia, 2020. 23: p. 203-215.
21. Tembhurne, J.V. and T. Diwan, Sentiment analysis in textual, visual and multimodal inputs using recurrent neural networks. Multimedia Tools and Applications, 2021. 80: p. 6871-6910.
22. Dilawari, A. and M.U.G. Khan, ASoVS: abstractive summarization of video sequences. IEEE Access, 2019. 7: p. 29253-29263.
23. Nor, A.K.M., et al., Overview of explainable artificial intelligence for prognostic and health management of industrial assets based on preferred reporting items for systematic reviews and meta-analyses. Sensors, 2021. 21(23): p. 8020.

24. Cazzato, D., et al., A survey of computer vision methods for 2d object detection from unmanned aerial vehicles. Journal of Imaging, 2020. 6(8): p. 78.
25. Zhang, S., et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
26. Magalhães, S.A., et al., Evaluating the single-shot multibox detector and YOLO deep learning models for the detection of tomatoes in a greenhouse. Sensors, 2021. 21(10): p. 3569.
27. Huang, G., et al. Densely connected convolutional networks. in Proceedings of the IEEE conference on computer vision and pattern recognition. 2017.
28. Trinh, H., et al., Energy-aware mobile edge computing and routing for low-latency visual data processing. IEEE Transactions on Multimedia, 2018. 20(10): p. 2562-2577.
29. Zhang, X., et al., A two-stage deep transfer learning model and its application for medical image processing in Traditional Chinese Medicine. Knowledge-Based Systems, 2022. 239: p. 108060.
30. Bhatt, D., et al., CNN variants for computer vision: History, architecture, application, challenges and future scope. Electronics, 2021. 10(20): p. 2470.
31. Ibrahim, M.R., A computer vision system for detecting and analysing critical events in cities. 2021, UCL (University College London).
32. Toto, E., M. Tlachac, and E.A. Rundensteiner. Audibert: A deep transfer learning multimodal classification framework for depression screening. in Proceedings of the 30th ACM international conference on information & knowledge management. 2021.
33. McDonnell, E.J., et al., Social, environmental, and technical: Factors at play in the current use and future design of small-group captioning. Proceedings of the ACM on Human-Computer Interaction, 2021. 5(CSCW2): p. 1-25.
34. Sarma, D. and M.K. Bhuyan, Methods, databases and recent advancement of vision-based hand gesture recognition for hci systems: A review. SN Computer Science, 2021. 2(6): p. 436.
35. Gupta, U., et al. Masr: A modular accelerator for sparse rnns. in 2019 28th International Conference on Parallel Architectures and Compilation Techniques (PACT). 2019. IEEE.
36. Gururangan, S., et al., Don't stop pretraining: Adapt language models to domains and tasks. arXiv preprint arXiv:2004.10964, 2020.
37. Jafari, Z., B.E. Kolb, and M.H. Mohajerani, Age-related hearing loss and tinnitus, dementia risk, and auditory amplification outcomes. Ageing research reviews, 2019. 56: p. 100963.
38. Ahmad, F. and W. Barner-Rasmussen, False foe? When and how code switching practices can support knowledge sharing in multinational corporations. Journal of International Management, 2019. 25(3): p. 100671.
39. Adnan, K. and R. Akbar, Limitations of information extraction methods and techniques for heterogeneous unstructured big data. International Journal of Engineering Business Management, 2019. 11: p. 1847979019890771.
40. Varma, R., et al., A systematic survey on deep learning and machine learning approaches of fake news detection in the pre-and post-COVID-19 pandemic. International Journal of Intelligent Computing and Cybernetics, 2021. 14(4): p. 617-646.
41. Alamoudi, E.S. and N.S. Alghamdi, Sentiment classification and aspect-based sentiment analysis on yelp reviews using deep learning and word embeddings. Journal of Decision Systems, 2021. 30(2-3): p. 259-281.
42. Kays, R., et al., Terrestrial animal tracking as an eye on life and planet. Science, 2015. 348(6240): p. aaa2478.
43. Choudhary, M., et al., BerConvoNet: A deep learning framework for fake news classification. Applied Soft Computing, 2021. 110: p. 107614.
44. Chelliah, S.L., Why language documentation matters. 2021: Springer.
45. Wardle, C. and H. Derakhshan, Information disorder: Toward an interdisciplinary framework for research and policymaking. Vol. 27. 2017: Council of Europe Strasbourg.
46. Spiekermann, S., et al., Values and ethics in information systems: a state-of-the-art analysis and avenues for future research. Business & Information Systems Engineering, 2022. 64(2): p. 247-264.

47. Manjunath Aradhya, V., H. Basavaraju, and D.S. Guru, Decade research on text detection in images/videos: a review. Evolutionary Intelligence, 2021. 14: p. 405-431.

48. Jia, X., et al., Highly scalable deep learning training system with mixed-precision: Training imagenet in four minutes. arXiv preprint arXiv:1807.11205, 2018.

49. Mirsky, Y. and W. Lee, The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 2021. 54(1): p. 1-41.

50. Ireton, C. and J. Posetti, Journalism, fake news & disinformation: handbook for journalism education and training. 2018: Unesco Publishing.

51. Katsaounidou, A.N., et al., News authentication and tampered images: evaluating the photo-truth impact through image verification algorithms. Heliyon, 2020. 6(12).

52. Wang, Z., et al. Issues of social data analytics with a new method for sentiment analysis of social media data. in 2014 IEEE 6th International Conference on Cloud Computing Technology and Science. 2014. IEEE.

53. Martinez-Rodriguez, J.L., A. Hogan, and I. Lopez-Arevalo, Information extraction meets the semantic web: a survey. Semantic Web, 2020. 11(2): p. 255-335.

54. Zaki, M. and A. Neely, Customer experience analytics: dynamic customer-centric model. Handbook of Service Science, Volume II, 2019: p. 207-233.

55. Kim, W. and C. Kim, A new approach for overlay text detection and extraction from complex video scene. IEEE transactions on image processing, 2008. 18(2): p. 401-411.

56. Jiang, C., et al., Object detection from UAV thermal infrared images and videos using YOLO models. International Journal of Applied Earth Observation and Geoinformation, 2022. 112: p. 102912.