Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
February 2024 5(2), 211–221

# FAKE NEWS DETECTION USING MACHINE LEARNING MULTI-MODEL METHOD

Laxminarayan Sahu [1], Dr. Bhavana Narain [2]

[1] Research Scholar, MSIT, MATS University, Raipur, Chhattisgarh, India
[2] Professor, MSIT, MATS University, Raipur, Chhattisgarh, India

**Corresponding Author**
Laxminarayan Sahu,
pro.lns711@gmail.com

**DOI**
**10.29121/shodhkosh.v5.i2.2024.1811**

**Funding:** This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.



## ABSTRACT

A fake news article that originates from an WhatsApp source is known as fake news. Fake news is becoming more and more prevalent on social media and other platforms, and this is a serious worry since it has the potential to have devastating effects on society and the country. This is why there has already been a lot of research done on its detection. This study uses supervised machine learning techniques to develop a product model through research and implementation of false news detection system. To put it briefly, this work will use a Naive Bayes classifier to build a model that can identify fake news by measuring its words and phrases against a set of criteria. that uses techniques like a count vectorizer (using word tallies) or a (Term Frequency Inverse Document Frequency) tfidf matrix to categorize bogus news as real or false It's highly likely that the meaning of two papers with comparable word counts will be entirely different.

**Keywords:** Fake News, WhatsApp, Machine Learning, Random Forest

## 1. INTRODUCTION

The spread of misinformation, particularly through social media, poses a significant challenge in today's digital age. Fake news can manipulate public opinion, erode trust in institutions, and even have real-world consequences. The proliferation of fake news, particularly on social media platforms, necessitates robust detection methods. Traditional approaches relying solely on text analysis have limitations. This paper proposes a novel multi-modal approach for detecting fake news by leveraging the power of both textual and visual information. Fake news detection is critical for combating misinformation in the digital age. While traditional methods analyze textual content for deception cues, they struggle with content that employs manipulative visuals. This paper proposes a multi-modal approach that utilizes machine learning techniques to analyze both text and images for a more comprehensive understanding of the information.

Usually, the intention behind fake news items is to trick or mislead readers. As an illustration, prior instances of creating false news items were utilized to boost revenue by driving website traffic through "Clickbait" content. They were typically crafted and written to become viral by targeting controversial issues. The majority of individuals get their news these days from social media. 62% of American people in the country use social networking platforms to obtain news. Humans have a 70% success rate in identifying fake news, and most non-expert users are unable to determine the veracity of the material they read. Consequently, fake news has grown increasingly prevalent on social media platforms, posing a growing threat to cyber security. It is now crucial to have technologies for automatically identifying false information.

## 2. FAKE NEWS

Fake news, also known as misinformation and disinformation, is a growing problem in today's digital age. It refers to false or misleading information presented as news, often spread deliberately to deceive or influence people. The practice of using newspapers began in the early 19th century, which is when bogus information first appeared. Yellow journalism is the process by which rumors are reported as factual information and a sensation is sparked among the public. In 1898, yellow journalism helped to set the stage for the Spanish-American War (A SurveyIhsan Ali et al., 2022).

When consumers started paying attention to web-based news, Yellow Journalism started to wane. Since fake news frequently features inflated stories with eye-catching headlines and images, it has gained widespread attention. During a 2017 interview, US President Trump and Hillary Clinton came under fire for spreading false information that quickly gained popularity on the internet and social media.

### 2.1. TYPES OF FAKE NEWS

Fake news is made-up information that mimics the content of verified news sources but exposes unreliable information and lacks the media community's editorial standards. It combines with unverified material that is disseminated with the intent to deceive and mislead people. Fake news starts on social media and spreads quickly, both purposefully and inadvertently.

### 2.2. MISINFORMATION

False information that is spread unintentionally by people who believe it to be true. Misinformation, a subset of the larger fake news category, is incorrect or misleading information that is spread unintentionally. People who share misinformation often believe it to be true and lack the intent to deceive. Here's a deeper dive into misinformation (Akyon et al., 2019).

**Figure 1**

**Figure 1** Sample Fake Image (Its Canals Appear Like They'd Be Perfect for Ice Skating in The Winter)
**Source** https://images.search.yahoo.com/search/images

**Figure 2**

**Figure 2** Real Image (This Image's Frozen Water Truly Originates from Russia's Lake Baikal)
**Source** https://images.search.yahoo.com/search/images

Misinformation lacks the malicious intent to deceive present in disinformation. People sharing misinformation might be genuinely mistaken or simply haven't verified the information before passing it on. 1). Reliance on social media and echo chambers for information. 2). Confirmation bias – the tendency to favor information that confirms existing beliefs (A. Kesharwani et al., 2020). 3). Lack of media literacy – difficulty evaluating the credibility of online sources.

## 3. LITERATURE SURVEY

In this work, we present a comprehensive overview of methods for detecting fake news from a novel angle: the inherent features of the fake news dissemination process, such as deliberate production, heteromorphic dissemination, and contentious reception. In addition to helping researchers better understand this topic, this review can be used to identify trends in technical advancements and provide insights into how to create methods for detecting false news that are both efficient and comprehensible (Bo Hu, et. al., 2024). By comparing different approaches, the effectiveness of these frameworks in identifying and stopping the spread of erroneous information may be evaluated. During the review process, a specific method was applied to produce an in-depth and educational evaluation. More research can be done in the future to identify false news, even though a lot of it has already been done. Numerous drawbacks characterize these investigations, including skewed data, adaptability, credibility analysis, and the fact that their solutions are the only ones that determine if the news is true or not. However, a practical method (Malek algabri, et. al., 2024). Deep Neural Network (DNN), passive-aggressive, and Naïve Bayes classifiers are used to classify the link between missing data variables and valuable data characteristics. According to the study's findings, the suggested method's total computation for identifying fake news was 99.8% accurate when evaluating the dataset's numerous claims, including those that are barely true, half true, true, largely true, and untrue. Lastly, the suggested approach's effectiveness is contrasted with the current approaches, demonstrating the proposed method's superior efficiency (Balshetwar SV et al., 2023).

One of the biggest issues of our day is the dissemination of false information on social media. Furthermore, although some research have offered a multimodal strategy that integrates textual and picture data for the purpose of detecting fake news, others have relied on attributes based on the user or the social network. In this study, the sentiments expressed in the comments made by the public were examined in order to determine their opinions regarding news. Given that the majority of remarks regarding fake news express feelings like shock, contempt, and fear, whereas remarks regarding true news (Hamed, et. al., 2022). Given social media's increasing popularity, it is anticipated that more people will be getting their news via these platforms than from traditional news outlets. According to the research, false information spreading on social media has had a significant negative impact on both people and society at large. In this study, the problem of fake news was investigated through a literature review that was split into two stages: detection and characterization, followed by a discussion of our results (Steni et. al., 2022). We have suggested a method for Facebook users to identify fake news that makes use of deep learning and machine learning classifiers in a Chrome environment. Our method examines the aspects of the news content as well as the user profile. We have created a Chrome plugin that makes use of data that our crawler has collected and retrieved. We have also employed the Long Short-Term Memory deep learning method to improve the speed of the Chrome addon (S. R. Sahoo, B.B. Gupta, 2021).

## 3.1. LEARNING ALGORITHMS

Algorithms for learning such as Random Forest, J48, SVM, and Logistic Regression plays an important role in this research work. Given are some discussions of these algorithms.

## 3.2. MULTI-MODEL METHOD FOR TEXT FEATURE

Evidence Based on Syntax: According to studies that have been published, syntax-based features include a variety of linguistic characteristics that follow specific patterns for the categorization of false information. In order to create information that is deceptive, writers intentionally use components like headline words and capitalized keywords. They also take into account the text's length and word density, as shown in Table 3.2. Because of these innate characteristics, digital natives are drawn to and influenced by news, which emphasizes the value of statistical proof in spotting false information. In this chapter, we give a formal characterization of the syntax-based characteristic that is being studied, $Xisy$ in the context of fake news.

**Table 1**

**Table 1 Grammatical Features**

| S. No. | Grammatical-Based feature | Example | Represents |
|---|---|---|---|
| 1 | Noun | The Candidate Eloquently Addressed the Nation, Inspiring Them with His Passionate Speech | Candidate, nation Addressed, Inspiring Them, His passionate, eloquently |
| 2 | Verb | | |
| 3 | Pronoun | | |
| 4 | Adjective | | |
| 5 | Adverb | | |

## 3.3. DEFINITION 3: PROOFS GRAMMATICAL

The grammatical features $Xgr$= {(X$nou$), (X$ver$), (X$adx$), X$xpru$)} for a given news $A$ correspond to the following features references: noun count (X$nou$), verb count (X$ver$), adjective count (X$auj$), adverb count (X$adv$), and pronoun count (X,).

## 3.4. EVIDENCE-BASED ON READABILITY

The degree to which a reader comprehends the written content is measured by readability. Simply put, the subject matter of a book and the complexity of its language and syntax determine how readable it is. Multiple readability indices (word and sentence score) are utilized to measure the abstract's readability and can be used to establish any text's grade. It helps us assess if the intended audience will be able to understand it completely. Table 3.5 offers further details. The features needed to determine the readability measures score are listed in Table 3.1.

**Table 2**

**Table 2 Readability Feature**

| S. No. | Readability Based Feature | Description | Formulation |
|---|---|---|---|
| 1 | Flash Reading Ease | Analyses the written text level of difficulty | 207.835 – (1.015 × ASL) – (84.6 × ASW) |
| 2 | Automated Readability Index | Determined the Level of Comprehension of English text | $4.17\left(\frac{X_{cc}}{X_{wc}}\right)+\left(\frac{X_{wc}}{X_{st}}\right)$ |
| 3 | Gunning Fog Index | Observing the Issue with The Writing | 0.4 (ASL + PHW) |
| 4 | Coleman Liau | Focus on The Text Character | ((0.0588 × L) – (0.296 × S) – 15.8) |
| 5 | Flesch-Kincaid Score | Examine The Written Text Complexity | 3 + Square Root of Polysyllable Count |
| 6 | The SMOG Index | Understand the Formulation of Written Content | (0.39 × ASL) + (11.8 × ASW) – 15.56 |
| 7 | Linsear Write Formula | Developed to Assess the Readability of Technical manual for the United State Air Force | $\left[\frac{\left(100-\frac{100\times N_{wsy<3}}{N_w}\right)+\left(3\times\frac{100\times N_{wsy\leq3}}{N_w}\right)}{\left(100\times\frac{N_{st}}{N_w}\right)}\right]$ |

## 3.5. DEFINITIONS 4: PROOFS OF READABILITY

The readability features $XR$= {X$\boldsymbol{U}e$, X$ari$, X$gl$, X$fki$, X$smg$, X$\boldsymbol{X}wf$ are comprised of (X$\boldsymbol{U}$E) Felsch Reading Ease, (X$aii$) for a particular news article $A$. The Automated

Readability Index, Coleman Liau, Gunning Fog, Flesch-Kincaid Index, and ($Xbmg$) SMOG Index, Linear Write formula ($XLwf$). The readability indexes are listed in Table 3.4.

## 4. METHODOLOGY

The reference frame is computed using the collusion approach by applying a statistical operation on a predetermined number of consecutive frames. The flowchart shown in Fig. 3.3 helps to convey the complete description. The motion residual extraction procedure with the collusion approach is illustrated in Fig. 3.3 through the use of flowcharts. The first step in the extraction procedure is to turn a given video into grayscale frames. Choosing the current frame that the motion residual is taken from is another step. To create a temporal window of frames, an equal number of frames are chosen before and after the chosen frame, which is considered the middle frame. To obtain a reference frame for motion residual extraction, the pixel-by-pixel median operation is applied to this chosen window of frames. To create the motion residual corresponding to the present frame, the reference frame is subtracted from it. Up until all of the video frames are transformed into motion residual frames, the process is repeatedly carried out.

**Figure 3**

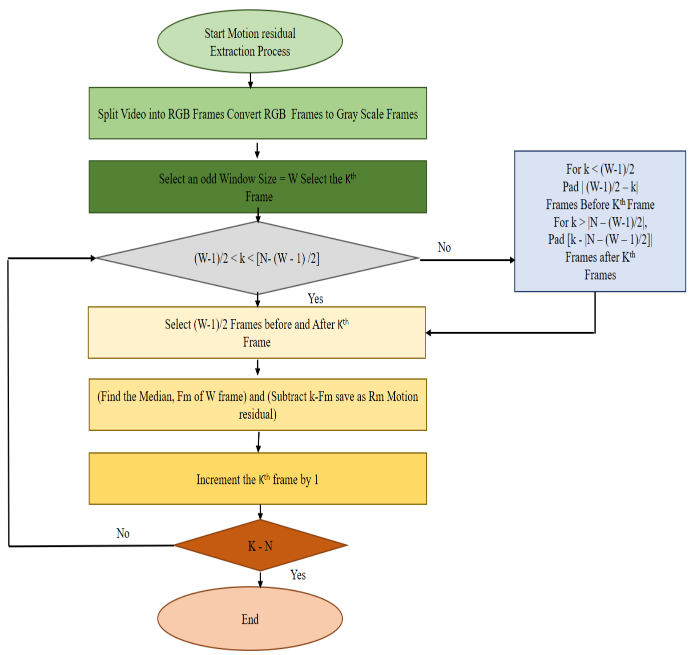



**Figure 3** Shows the Motion Residual Extraction Method Flow Chart

## 5. MULTI-MODEL METHOD FOR VIDEO FEATURE

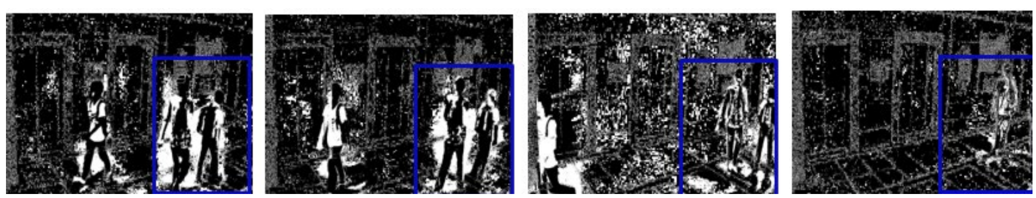
Object-based video falsification is a sort of falsification in which the target object is purposefully added or removed from the video content. Surveillance films typically show object-based forgeries, particularly at ATMs, banks, traffic signals, roadside locations, etc. These videos are essential pieces of evidence in legal proceedings. Consequently, in order to validate the object-based counterfeit films, a scientific approach is needed. To identify forgeries, researchers use crucial video metrics like Group of Pictures (GOP), codec, data rate, and details. The authors of (Bayar B. et al., 2016, Feng C., et al., 2017, Long C. et al., 2017, Haralick R. M., et al., 1987), and other publications define frame-level forgeries as the process of identifying inserted, duplicated, or removed frames from a movie. The majority of these detection techniques use fixed GOP-length films (GOP and iFrame are the two types of codecs. I-Frame essentially holds unique data for every single video frame. GOP - holds the data for the first frame, and every other frame refers to how it differs from that first frame.)

Found frame-level forgeries in (Jia et al., 2018), using an optical flow-based technique to identify copy-move frames in a video. They have developed their approaches for the SULFA (320×240). (Quadir G. et al., 2016) and DERF (176×144) datasets. Chen et al.'s study SULFA (Quadir G. et al., 2016) used video steganalytic features to identify clean, double-compressed, and fabricated frames in a film.

A Multi-Model Method is made to classify frames as fabricated or double compressed by using the motion residuals of a video, in order to accomplish the aforementioned objectives. Comprehensive assessments are carried out for diverse testing situations. The evaluation reveals the consistency of the classification accuracy in contrast to the videos with varying durations.

**Figure 4**

(a) The immaculate video frame numbers 10, 20, 40, and 60 are shown in the upper row, going from left to right.



(b) The motion residual frames in the binary form image that corresponds to are shown in the row above, going left to right.

**Figure 4** Using Frames from the SYSU-OBJFORG Dataset, Displays Both Authentic (00055.Mp4) and Fake (00055 016-117.mp4) video frames (Chen S. et al., 2016).

The recovered motion residuals for the frames from the SYSU-OBJFORG dataset are shown in Fig. 3.1 (Chen S. et al., 2016). Fig. 3.1(a) depicts the film with the file number 00055.mp4, whereas Figure 3.1b shows the extracted motion residual frames as a binary form image. Fig. 3.1(a). Two moving people are indicated by the blue rectangle box in immaculate frames (10, 20, 40, and 60) in Fig. 3.1(a), and by the double compressed frame (10 numbered frame). The blue box-marked individuals are missing from the frames 20, 40, and 60, indicating that they are fake. The area with the green box deleted was the motion residual matching the forged frames which indicated two moving people.

## 6. RESIDUAL MOTION EXTRACTION LEVEL

In order to extract motion residuals for each frame in the video, the given video is first pre-processed. The motion residuals that are retrieved from a video serve to eliminate superfluous background information while also emphasizing any artifacts that may be present in the object-based fabricated frame of the video. Let's say that a video V has N total frames and may be expressed as follows:

V = [F(N), F (1), F (2),……] ------------------(3.1)

In which the Ith position frame is denoted by F(i): I am 1, 2, ..., N.

Choose a temporal window, W(t), with a length of (2p+1), and use F(t) as the middle frame.

W(t) = [F(p), F(t), F (t +1), ...., F (t +1+ p)] ------------------------- (3.2)

Thus, only when (t > p) and t < (N − p) is the equation (3.2) true. Pre-pad the frame sequence with (t − p) zero-valued frames in the situation where t < p. Post-pad frame sequence with (t − (N − p)) zero-valued frames for the case t > (N − p). This implies that seven zero-valued frames must be added at the start if the temporal window is (2 ∗9+1) long and the motion residual of the second frame is to be recovered. Similarly, six zero-valued frames must be added at the conclusion if the motion residual of the third-to-last frame is to be extracted. The motion residual relating to a frame can only be calculated with a reference frame. The reference frame in the suggested method is W(t), or the median frame of the temporal window that was chosen. Pixel per pixel, the median frame is calculated.

Pixel per pixel, the median frame is calculated. Using the formula for pixel (x, y) as follows, let Fm(t)x,y represent the median frame:

Fm (t)x,y = median[F(1)x,y,--------, F(p)x,y, F(t)x,y, F(t +1)x,y, ---------, F(t +1+ p)x,y]----------------(3.3)

The motion residual frame Rm(t) is now generated by subtracting the median frame, Fm(t), from the current frame F(t) as follows:

Rm(t) = F(t)−Fm(t)---------------------(3.4)

The statistical process known as the median is used to determine the center tendency among an ordered set of data. When one performs a median operation on the ordered series of frames that make up the temporal window W(t), the relevant information of W(t) is preserved in the form of Fm(t). Furthermore, a frame called Rm(t), which is an integer-valued frame that can have both positive and negative values is produced when Fm(t) is subtracted from F(t).

Motion residuals are retrieved in this way for every frame in a particular video. The following formula represents the set of motion residuals for a given video V:

Rm = [Rm (1), Rm(2), ......, Rm(N)]------------------------(3.5)

After being saved as a 2D matrix, these extracted motion residuals are then supplied as an input to the Multi-Model Method that is being suggested for frame classification.

## 7. WHATSAPP DATASET

The WhatsApp dataset (WD) includes both visual and textual elements. Accuracy, precision, recall, and F1-Score of the Multi-Model Method technique are analyzed using textual features from the WhatsApp dataset. The performance evaluation of the WD dataset with textual features is displayed in Table 4.5.

**Table 3**

| | Accuracy (in %) | Precision (in %) | Recall (in %) | F1-Score (in %) |
|---|---|---|---|---|
| SVM | 74 | 76 | 78 | 79 |
| Classifier J48 | 67 | 68 | 67 | 63 |
| Logistic Rogation | 68 | 69 | 70 | 75 |
| **Random Forest (Multi-Model Method)** | **90** | **92** | **91** | **95** |
| Cluster Based Stacking Classification | 85 | 87 | 88 | 83 |

**Table 3** Performance Evaluation of WhatsApp Dataset with Text Input Using All Machine Learning Models

Using the Instagram dataset, the Multi-Model Method outperforms SVM, Classifier J48, Logistic Regression, Random Forest, and Cluster-Based Stacking Classification in terms of accuracy, Precision, Recall, and F1-Score.

## 8. WHATSAPP RESULT

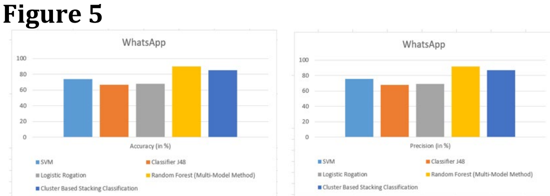
The WhatsApp dataset (WD) includes both visual and textual elements. Accuracy, precision, recall, and F1-Score of the Multi-Model Method technique are analyzed using textual features from the WhatsApp dataset. The performance evaluation of the WD dataset with textual features is displayed in Table 4.5.

**Table 4**

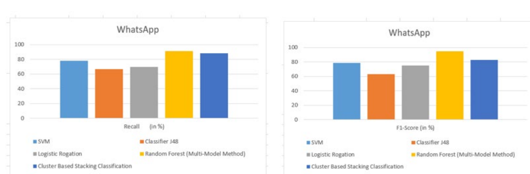
| | Accuracy (in %) | Precision (in %) | Recall (in %) | F1-Score (in %) |
|---|---|---|---|---|
| SVM | 74 | 76 | 78 | 79 |
| Classifier J48 | 67 | 68 | 67 | 63 |
| Logistic Rogation | 68 | 69 | 70 | 75 |
| **Random Forest (Multi-Model Method)** | **90** | **92** | **91** | **95** |
| Cluster Based Stacking Classification | 85 | 87 | 88 | 83 |

**Table 4** Performance Evaluation of WhatsApp Dataset with Text Input Using All Machine Learning Models

**Figure 5**




(a) Accuracy (b) Precision




(c) Recall (d) F1-Score

**Figure 5** Performance Analysis in Terms of (a) Precision, (b) Accuracy, (c) Recall, and (d) F1-Score of WhatsApp Dataset using Multi-Model Method

Using the Instagram dataset, the Multi-Model Method outperforms SVM, Classifier J48, Logistic Regression, Random Forest, and Cluster-Based Stacking Classification in terms of accuracy, Precision, Recall, and F1-Score.

## 9. CONCLUSION

Next, the Multi-Model Method performs well on Instagram, Twitter, and WhatsApp datasets. The verifiable results provided by ImageNet weights demonstrate that both the Multi-Model Method and Random Forest models have distinct qualities when applied to the Instagram, Twitter, and WhatsApp dataset. These networks are beneficial primarily due to their rapid learning, which makes timing easier to manage. The findings also demonstrate that the Multi-Model Method model, which learned from incredibly customized datasets to recognize hidden false information in test samples with more than 80% accuracy. A general deepfake detection system utilizing 26 distinct deep convolutional networks—some of which are subsequently ensembled to enhance predictions—is presented in this thesis as an alternative method. For improving dependability of the system, models are trained on three unique datasets. As far as we can tell, this is a work that covers practically every well-liked classification model in the field of image classification.

Additionally, a method is used to eliminate the term "Black Box," which is typically connected to deep learning models and illustrates how these models truly forecast if an image is real or phony. Additionally, this thesis built a unique architecture that combined segmentation and classification models that are evaluated and trained on both the Instagram, Twitter, and WhatsApp is utilized to raise the model's accuracy. Additionally, the models are tuned for hardware implementation. Through quantizing, gathering, and integrating the Multi-Model Method AI's segmentation and classification model. Reverse card inference. The Instagram, Twitter, and WhatsApp photos and movies can be effectively classified by the model The task of deep false audio detection is also highlighted in this paper. Two techniques have been utilized to discern between real and fake audio: a feature-based approach based on machine learning algorithms as well as a deep learning approach based on images. Since then, both techniques have been contrasted. The results showed that the deep learning algorithms performed better than the machine learning techniques. Outcomes seen above. The Multi-Model Method model had the highest test accuracy, at 93%. Considering that Multi-Model Method is a Multi-Model Method may function if there is a sequential network and sequential data input.

## CONFLICT OF INTERESTS

None.


## ACKNOWLEDGMENTS

None.


## REFERENCES


"Local Tampering Detection in Video Sequences," In 2013 Ieee 15th International Workshop on Multimedia Signal Processing (Mmsp), 2013, Pp. 488–493.

Chen S., Tan S., Li B., and Huang J., "Automatic Detection of Object-Based Forgery in Advanced Video," IEEE Transactions on Circuits and Systems for Video Technology, vol. 26, no. 11, pp. 2138–2151, Nov 2016.

Jia S., Xu Z., Wang H., Feng C., a nd Wang T., "Coarse-to-fine Copy-Move Forgery Detection for Video Forensics," IEEE Access, vol. 6, pp. 25 323–25 335, 2018.

D'Avino D., Cozzolino D., Poggi G., and Verdoliva L., "Autoencoder with Recurrent Neural Networks for Video Forgery Detection," Electronic Imaging, vol. 2017, pp. 92–99, 01 2017.

Johnston P., Elyan E., and Jayne C., "Video Tampering Localisation Using Features Learned from Authentic Content," Neural Computing and Applications, pp. 1–15, 2019.

Zhang J., Su Y., and Zhang M., "Exposing Digital Video Forgery By Ghost Shadow Artifact," in Proceedings of the First ACM Workshop on Multimedia in Forensics, ser. MiFor '09. New York, NY, USA: Association for Computing Machinery, 2009, p. 49–54. [Online]. Available: https://doi.org/10.1145/1631081.1631093

Hsu C.-C., Hung T.-Y., Lin C.-W., and Hsu C.-T., "Video Forgery Detection Using Correlation