Original Article
ISSN (Online): 2582-7472

ShodhKosh: Journal of Visual and Performing Arts
July-December 2023 4(2), 873–881

FAKE NEWS DETECTION SYSTEM ON INSTAGRAM USING MACHINE LEARNING MULTI-MODEL METHOD

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

[1] Research Scholar, MSIT, MATS University, Raipur (C.G.)
[2] Professor, MSIT, MATS University, Raipur (C.G.)

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

DOI
10.29121/shodhkosh.v4.i2.2023.1870

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

ABSTRACT

The consumption of news on social media is growing in popularity. Social media appeals to users because it is low-cost, user-friendly, and rapidly disseminates information. False information does, however, also circulate, in part because to social media. It's getting harder and harder to ignore fake news because of the harm it causes to society. However, depending only on news content usually leads to poor detection effectiveness because fake news is designed to look legitimate. As such, a detailed comprehension of the relationship between social media user profiles and fake news is necessary. This study examines the use of machine learning algorithms to detect fake news. It covers significant subjects like user profiles, dataset analysis, and feature integration. The study integrates attributes to provide large feature sets. When dealing with high-dimensional datasets, Principal Component Analysis (PCA) is a helpful technique for dimensionality reduction. The study uses datasets from "Instagram," which include a variety of data processing techniques, to extensively analyze several machine learning models. The evaluation of the Random Forest classification model is further improved via curve analysis. The outcomes show the best feature and model pairings, with our model outperforming the competition.

Keywords: Fake News, Machine learning, Instagram, Multimodal Method

1. INTRODUCTION

Identifying fake news is essential to fighting false information in the digital era. Conventional techniques look for deception cues in textual content, but they have trouble with manipulative visual content. In order to provide a more thorough understanding of the material, this study suggests a multi-modal strategy that makes use of machine learning algorithms to assess both text and images.

Various commercial and scientific initiatives have been made to identify and flag false information. Facebook is one of the venues where a lot of fake news stories are shared, such the ones that were published before the US election in 2016. In response to the threat of fake news spreading, Facebook began labeling stories that fact-checkers may flag as fraudulent. Facebook intends to designate certain postings as "False Information" in the run-up to the US election; however, political Facebook advertisements will not be covered by this new policy. Another initiative to use fact-checking as the initial step in identifying false news was launched in 2017 **(Y. Amit et al., 2023).**

The goal of the Fake News Challenge (FNC-1) is to create novel advancements in intelligent systems for stance recognition, which can be used to forecast one of four stance labels—discuss, agree, disagree, and unrelated—when a document is compared to its title. A gradient-boosting decision tree model and a weighted average Convolutional Neural Network (CNN) model were employed by the top-ranked system. After a more thorough analysis of the top three models, it was determined that even the best-performing features could not properly resolve the challenging situations in which even humans misinterpreted the agree, disagree, and discuss labels. Another investigation on false news identification focuses on linguistic features such as Ngrams, punctuation, grammar, readability metrics, and Psycholinguistic factors (using the Linguistic Inquiry and Word Count (LIWC) tool) **(E. Juandreas et al., 2021).** These characteristics demonstrate how the variation in writing styles can be used to determine the authenticity of the text.

Fake News

In the current digital era, fake news—also referred to as misinformation and disinformation—is becoming a bigger issue. It refers to inaccurate or misleading material that is disseminated as news and is frequently done so on purpose in an effort to trick or sway people. False information originally appeared in newspapers in the early 19th century, which is also when the practice of employing them started. Yellow journalism is the practice of disseminating rumors as true news in order to create a stir among the public. The Spanish-American War began in 1898 as a result of the influence of yellow journalism (**SurveyIhsan Ali et al., 2022).** Yellow journalism started to fade when people began to pay attention to web-based news. Fake news is well-known since it usually consists of exaggerated articles with attention-grabbing titles and graphics. US President Trump and Hillary Clinton faced criticism in 2017 during an interview for disseminating misleading material that immediately went viral on the internet and social media.

1.2.1. Identifying Misinformation

The same techniques used for spotting fake news apply to identifying misinformation:

Is it a credible news organization or a website known for spreading misinformation? Does the information have citations or links to trustworthy sources? Use reputable fact-checking websites. Equipping people with the skills to critically evaluate information online is crucial **(S. Kumar et al., 2024).** These organizations help debunk misinformation and provide reliable information. Before sharing information online, take a moment to verify its source and accuracy.

Fig. 4. Combination of Real and Fake Images for Delhi Chief Minister Arvind Kejriwal and Air Hostess. Source link: https://hi.quora.com

Here's a breakdown of the key aspects of fake news:

1.2.2. Disinformation

False information deliberately spread to deceive people and mislead them. This can be for political gain, financial benefit, or simply to cause chaos. Disinformation, the malicious cousin of misinformation, is false information deliberately spread to deceive people and influence them. Unlike misinformation, which is spread unintentionally, disinformation has a hidden agenda, often aiming to manipulate public opinion, promote a specific viewpoint, or cause social unrest. Let's delve deeper into the world of disinformation **(Villela et al., 2023).**

2. LITERATURE SURVEY

In this study, we provide fake news, an extensive library that includes a variety of models for detecting false news, primarily divided into two categories: social context-based and content-based. Fake news is intended to provide an integrated framework for these algorithms, which includes a series of procedures such as data processing, model training, and evaluation in addition to optional features like logging and visualization **(**Y. Zhu et. al., 2024). The research highlights the significance of combating misinformation and its numerous issues by utilizing machine learning and natural language processing to identify fake news in social media. The study demonstrates that thorough data pretreatment and feature engineering can significantly increase detection accuracy by include information about speakers and their political affiliations **(**S. K. Tummala**, et. al., 2024).** Nowadays, social network environments are where the majority of disinformation is disseminated, yet very few research have been conducted in these settings. The platform with the most often utilized datasets was found to be kaggle; weibo, fnc-1, covid-19 fake news, and twitter followed. Future study should focus on news related to politics, as this was the main topic that drove research development in 2017; moreover, hybrid approaches for spotting fake news should be investigated **(**H. F. Villela, et al., 2023**)** .

The outcome of this work is the formal benchmarking and meta-analysis of fake news detection methods that can be further utilized by the research community, but more importantly by the practitioners and decision-makers that counter fake news on a daily basis, e.g., in press agencies, homeland security agencies, fact-checkers, and so on. This work is the natural extension of the authors' previous systematic analysis of fake news detection methods and authors' own fake news detection methods based on machine learning (ml)/artificial intelligence (ai) techniques **(Rafal Kozik, et. al., 2023).** In this work, the text is converted into a numerical representation using feature extraction approaches like bag of words and tf-idf vectorization. The efficacy of machine learning classifiers, such as random forest, gradient boosting, naïve bayes, decision tree, support vector machine, and logistic regression, in identifying false news is assessed in this research. Notably, good accuracy ratings were obtained when tf-idf vectorization was applied to gradient boosting and decision tree classifiers. Similarly, decision tree, gradient boosting, logistic regression, and bag of words all obtained exceptional accuracy results when combined **(C. Kanwar, Y. Mohan, 2023).** researchers' model has been trained offline thus far. all that's needed to meet the real-time requirement is frequent training and updating of the model. conclusion no. 8we present a novel deep neural architecture in this paper for the detection of bogus news. we recognize and tackle two distinct issues surrounding false news: (1) early detection of bogus news and (2) label scarcity. the news module, the social situations module, and the detection module are the three main components of the system **(Shaina Raza, 2022).** Following an analysis of the five models' results, a combination of machine learning and natural language processing approaches is selected. All of the machine learning methods were combined with a range of natural language processing techniques to construct these five models. A trained machine learning model is connected to an interface that has been constructed. To predict user-inputted news, a passive aggressive classifier with tf-idf vectorizer is trained. This platform allows users to submit news, and it will decide whether or not the story is accurate **(Prof. N.R Hatwar, et al., 2022)**

3. METHODOLOGY

Unlike the traditional definition, which focuses on news articles that are intentionally and verifiably false and could mislead readers, fake news in the context of microblogs refers to user-posted news posts that are typically less than 140 characters. Formally, we state this definition as follows: In this work, we apply the following definition, derived from MediaEval's Verifying Multimedia Use task:

• Real images are consistent with the time and location of the event and are original and not doctored.
Images that are considered to be fake are either altered or mis-contextualized, which is characterized as a misalignment of the image's location and/or time in relation to the location and time of the event that is being reported in the media. 1. Semantic and contextual features are used in the current technique. To express words and sentences that best capture the semantics, BERT is used. A series of words that keep going up the stack are fed into it. After applying self-attention, each layer transfers the information to the subsequent encoder via a feed-forward network. Sentences are turned into tokens in BERT. **(Kim H. et. al., 2018)** Dimension reduction is implemented.

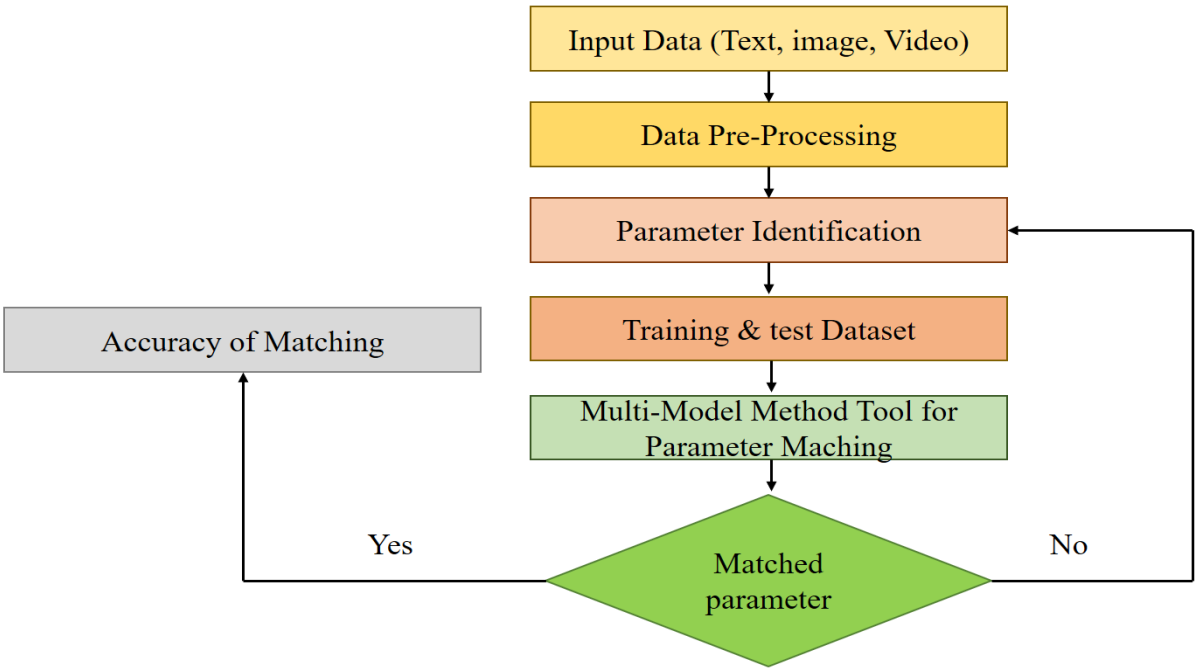


Fig. 4.1: Methodology for Multi-Model Method

Multi-Model Method for Text Feature

The extraction of linguistic feature evidence for the purpose of detecting false news will be the main focus of this section. The diversity of material has made it more vital than ever to identify key characteristics in order to provide crucial details that can be used to assess the reliability of online false news. The main source of support and inspiration for choosing the feature set is the body of existing research and its significant contribution. By using linguistic analysis, the text structure can be extracted most successfully. Through the creative synthesis of words, phrases, and sentences, it effectively conveys the text's intended meaning.

The basic definition in terms of parameters is as follows: A selection of n news stories with fictitious or authentic content is provided. In the case when $\mathcal{A}$=Ain, the issue with fake news identification is determining if the news items in $\mathcal{A}$ are phony or not. This is a binary classification issue, according to the targeted dataset, where label set $\mathcal{Y}$=[1,0], where 1 denotes genuine news and 0 denotes fraudulent news $Ai\epsilon\hbar$. Fake news has been classified according to four different feature categories: Features depending on syntax Grammar evidence features $Xigr$, Sentiment features $Xisen$, and readability features XiR. These properties are the main training parameters for the sequential neural network model.

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 3.1: Syntax-Based Features in Multi-Model Method

S. No.	Syntax-based Features	Description
1	Char Count	Total character count, including spaces
2	Word Count	The total number of words in a sentence
3	Title Word Count	number of words in the title of a phrase
4	Stop Word Count	Total number of stop words in an expression

5	Upper Case Word Count	number of words in a sentence that are capital The proportion of a chosen keyword's occurrences to the text's overall word count and word density

1.1.1. Definition 1: Evidence based on syntax

Syntax-based features $X^{yz} = (X_{cc}, X_{wc}, X_{wd}, X_{twc}, X_{up}, X_{stp})$ for a particular news *A* include character count (X_{cc}), word count (X_{wc}), word density (X_{wd}), title word count (X_{twc}), title uppercase (X_{up}), stop word count (X_{stp}).

1.1.2. Sentiment-Based Evidences:

Using presumptions, the sentiment is expressed when producing a news story and is based on the standards used to determine whether or not the news is fake. The primary sentiment-related traits that influence the assessment and strength of emotions are listed in Table 3.3.

Table 3.2: Semantic-based features of the multi-model method

S. No.	Semantic-Based Features	Description
1	Polarity	It refers to both positive and negative Statements. It falls within the range of -1 and 1.
2	Subjectivity	Reflects the expression of personal feelings, beliefs, of opinions, falling between 0 and 1.

1.2. Working

The spatial domain properties of a face are used by facial forgery detection approaches in literature such, etc., to determine whether a face is immaculate or fabricated. Furthermore, the researchers created intricate networks like XceptionNet **(Agrawal et. al., 2019)** and GAELNet **(Bakej et. al., 2020)** to thoroughly utilize spatial domain artifacts of the facially altered films. Better outcomes, however, come from applying frequency domain analysis to the forgery detection problem. To identify counterfeit photos, authors created a spectrum-based classifier in **(Zhang H. et. al., 2019)**.

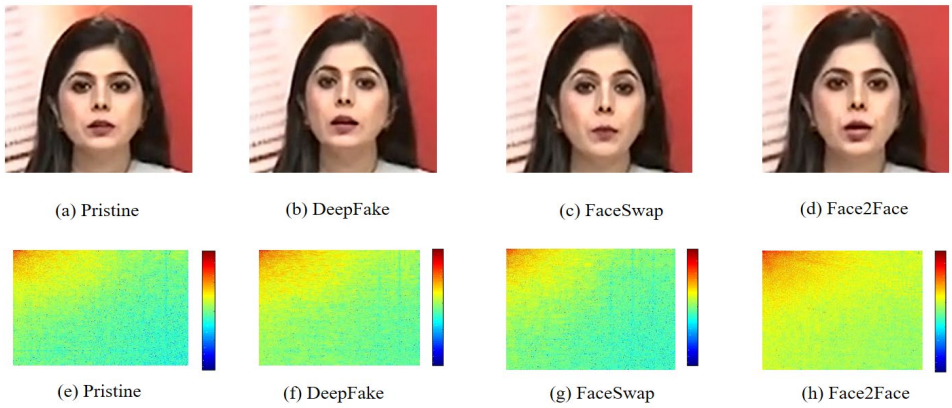


Fig. 4.3: The represents the extracted face from 067 mp4 and its corresponding 2D-GDCT Coefficients in Color map from (a)-(d) represent face of pristine, Deep Fake, Face Swap and Face2Face respectively of Face Forensics ++Dataset (e)-(H) represent 2D-GDCT Coefficients in Color map from corresponding to faces in (a)-(d)

In order to compare frequency domain and spatial domain classifiers, the authors ran experiments. and found that classifiers based on spectrums perform better than those based on pixels. Similar findings are reported in **(Rahmoni M., et. al., 2019)**, where audio recognition is achieved by utilizing frequency domain characteristics, which are shown to be more discriminative than time domain features. Using auto encoders or GANs, facial-altered videos such as Deep Fakes are produced. The Mapping the low-dimension latent space to the high-dimension space is the intuition underpinning autoencoders and GANs. When low-dimensional movies are up-sampled to high-dimensional videos, artificial or

synthetic artifacts are produced. It is easy to see these abnormalities when viewing the movie in frequency domain representations.

1.2.1. Module for Frequency Conversion

GDCT is used to convert the retrieved faces from video frames into the frequency domain. Think about f represents the extracted face.

$$f = [f(x,y), 0 \le x \le X-1, 0 \le y \le \gamma - 1]$$

Then GDCT of the face f_{gdct}

$$f_{gdct}(m,n) = a(m)a(n)\sum_{x=0}^{X-1}\sum_{y=0}^{Y-1} f(x,y)\frac{\cos(\pi m(2x+1))}{2X}\frac{\cos(\pi n(2y+1))}{2Y}$$

Where:

$$a(m) = \begin{cases} \sqrt{\frac{1}{X}} & m = 0 \\ \sqrt{\frac{2}{X}} & 1 \le m \le X-1 \end{cases}$$

$$a(n) = \begin{cases} \sqrt{\frac{1}{Y}} & n = 0 \\ \sqrt{\frac{2}{Y}} & 1 \le n \le \gamma - 1 \end{cases}$$

The converted face frequency coefficients are kept in the form of a size (X ×Y) two-dimensional matrix. Since the value in the upper right corner represents the average of the face, it is essential for frequency domain analysis.

1.3. Dataset

For various circumstances, the suggested Multi-Model Method approach is thoroughly assessed using the Instagram, Facebook, and WhatsApp **(Roseller A., et. al., 2019)** and Celeb-DF(v2) **(Li. Y. et. al., 2020)** datasets. One large and well-known facial augmentation dataset is Instagram, Facebook, and WhatsApp **(Roseller A., et. al., 2019)**. Videos of various resolutions are included in the dataset: 640 x 480 pixels for Video Graphic Array (VGA), 1280 x 720 pixels for High Definition (HD), and 1920 x 1080 pixels for Full High Definition (FHD). Both identity-based and expression-based faked videos are included in this dataset. Face2Face is expression-based, whereas DeepFakes and FaceSwap are identity-based falsified videos. Every facial forgery consists of 1000 videos, 3000 of which are videos with altered faces and the remaining 1000 videos are pure. Three compression grades are applied to these videos: uncompressed (c0), light compressed (c23), and heavy compressed (c40). The suggested method is trained and tested for every compression quality. Two groups—a training group with 850 videos and a testing group with 150 videos—are created from the 1000-video dataset. The 850 videos in the training group are further split into two categories: 700 movies for training and 150 videos for network validation. The frequency domain face samples needed for training and testing are extracted using the face extraction technique covered in section 4.3.2 above.

Table 4.1: The Number of Face Samples on Social Media for Training and Testing of Multi-Model Method for Binary Detection of Facial Forgeries

S. No.	Dataset	DeepFake		FaceSwap		Face2Face		Pristine	
		Training	Testing	Training	Testing	Training	Testing	Training	Testing

1	Instagram	6543	890	4340	619	5479	810	5730	801
2	Facebook	6786	871	4578	679	5671	855	5941	856
3	WhatsApp	639	897	4834	643	5930	827	5140	829

The amount of face samples used for Multi-Model Method training and testing for binary detection of DeepFakes, FaceSwap, and Face2Face forgeries of the Instagram, Facebook, and WhatsApp dataset is listed in table 5.1 **(Roseller A., et. al., 2019).**

Table 4.3: Performance Evaluation of Instagram Dataset with Text Input Using All Machine Learning Model

	Accuracy (in %)	Precision (in %)	Recall (in %)	F1-Score (in %)
SVM	83	80	85	83
Classifier J48	85	83	87	86
Logistic Rogation	79	76	80	81
Random Forest (Multi-Model Method)	**92**	**91**	**94**	**93**
Cluster-Based Stacking Classification	87	85	88	83

For the purpose of multi-class classification of the three facial forgeries, the suggested Multi-Model Method is also assessed. To train and evaluate Multi-Model Method for multi-class classification, four classes—DeepFake, Face2Face, FaceSwap, and Pristine—are created in a dataset. The proposed Multi-Model Method robustness is assessed using the Celeb-DF(v2) **(Li. Y. et. al., 2020)** dataset as well. According to the authors of **(Li. Y. et. al., 2020)** tests, the average Area Among the datasets that are accessible in the literature, the Celeb-DF(v2) dataset has the lowest area under the receiver of curve (AUC). This complicates the task of detecting Celeb-DF(v2). The videos in the Celeb-DF(v2) **(Li. Y. et. al., 2020)** collection feature 59 celebrities representing various geographies, moral codes, and age brackets. The dataset is balanced across films from the male and female communities, so it is not gender biased.

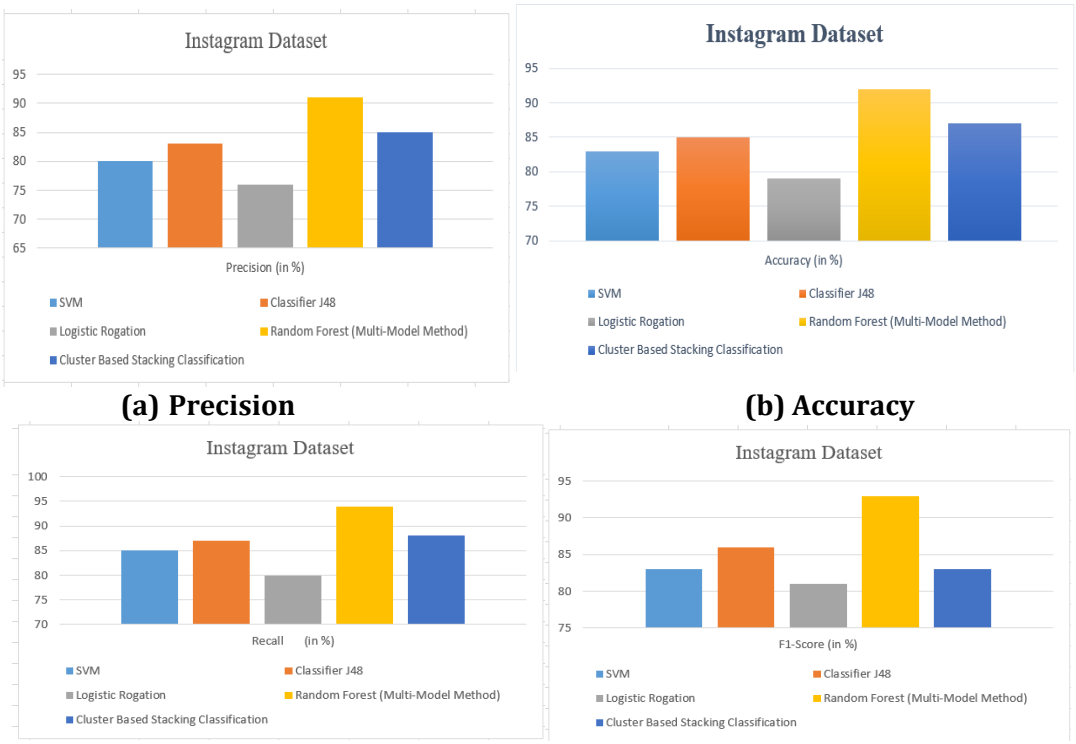
Using the Instagram dataset, the Multi-Model Method outperforms SVM, Classifier J48, Logistic Regression, Random Forest, and Cluster-Based Stacking Classification in terms of accuracy, Precision, Recall, and F1 Score.

1.4. Instagram Result

The Instagram dataset (ID) includes both visual and textual elements.

Table 5.6: Performance Evaluation of Instagram Dataset with Text Input Using All Machine Learning Model

	Accuracy (in %)	Precision (in %)	Recall (in %)	F1-Score (in %)
SVM	83	80	85	83
Classifier J48	85	83	87	86
Logistic Rogation	79	76	80	81
Random Forest (Multi-Model Method)	**92**	**91**	**94**	**93**
Cluster-Based Stacking Classification	87	85	88	83



(a) Precision **(b) Accuracy**

(c) Recall **(d) F1-Score**

Fig 5.2: Performance Analysis in Terms of (a) Precision, (b) Accuracy, (c) Recall, and (d) F1-Score of Instagram Dataset using Multi-Model Method

Accuracy, precision, recall, and F1-Score of the Multi-Model Method technique are analyzed using textual features from the Instagram dataset. The performance evaluation of the ID dataset with textual features is displayed in Table 4.3. 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.

4. CONCLUSION

The saying "Seeing is Believing" was previously prevalent. It's no longer true because Fake News content has increased significantly. People's faith in media material has begun to erode as a result of Deep News. Multi-Model Method films have the potential to disseminate hate speech, create political imbalances, slander individuals, and a host of other negative effects. continues. As a result of its spread, effectively identifying it becomes crucial. Due to the sharp rise one of the worries with the next Multi-Model Method detection technology and quicker calculation rates period. To test the issue of dataset shift, we create a unique test dataset in this thesis. The Multi-Model Method created with Instagram, Twitter, and WhatsApp posted to make up the custom dataset. An autoencoder architecture was used in Instagram, Twitter, WhatsApp to produce Multi-Model Method. We've displayed a few outcomes. From the creation procedure to discuss the superiority of the modifications that can be made with open computational resources and source tools. Fake Detection is valid for all tested distributions. Yields trustworthy outcomes. Likewise, the subpar cross-set outcomes of models that were trained separately on the three the problem of dataset shift brought on by various test settings is illustrated by several distributions. This conveys the idea that a job must be designed to accurately represent the real-world conditions of differently distributed test data. More diversity in pre-processing and generating techniques is needed for something as delicate as facial modification detection. For compiling datasets.

CONFLICT OF INTERESTS

None

ACKNOWLEDGMENTS

None

REFERENCE

Wiegand T., Sullivan G., Bjontegaard G., and Luthra A., *"Overview of the h.264/avc video coding standard,"* IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, no. 7, pp. 560–576, 2003.

Sullivan G. J., Ohm J.-R., Han W.-J., and Wiegand T., *"Overview of the high efficiency video coding (hevc) standard,"* IEEE Transactions on Circuits and Systems for Video Technology, vol. 22, no. 12, pp. 1649–1668, 2012.

Bross B., Wang Y.-K., Ye Y., Liu S., Chen J., Sullivan G. J., and Ohm J.-R.,*"Overview of the versatile video coding (vvc) standard and its applications,"* IEEE Transactions on Circuits and Systems for Video Technology, vol. 31, no. 10, pp. 3736–3764, 2021.

Kim H., Garrido P., Tewari A., Xu W., Thies J., Niessner M., Pe´rez P., Richardt C., Zollho¨fer M., and Theobalt C., *"Deep video portraits,"* vol. 37, no. 4, jul 2018. [Online]. Available: https://doi.org/10.1145/3197517.3201283

Yoo D., Kim N., Park S., Paek A. S., and Kweon I.-S., *"Pixel-level domain transfer,"* in ECCV, 2016.

Yang L.-C., Chou S.-Y., and Yang Y., *"Midinet: A convolutional generative adversarial network for symbolic-domain music generation,"* in ISMIR, 2017.

Schlegl T., Seebo¨ck P., Waldstein S. M., Schmidt-Erfurth U., and Langs G., *"Unsupervised anomaly detection with generative adversarial networks to guide marker discovery,"* in IPMI, 2017.

Wu J., Zhang C., Xue T., Freeman W. T., and Tenenbaum J. B., *"Learning a probabilis-tic latent space of object shapes via 3d generative-adversarial modeling,"* in Proceedings of the 30th International Conference on Neural Information Processing Systems, ser NIPS'16. Red Hook, NY, USA: Curran Associates Inc., 2016, p. 82–90.

Kim H., Garrido P., Tewari A., Xu W., Thies J., Niessner M., Pe´rez P., Richardt C., Zollho¨fer M., and Theobalt C., *"Deep video portraits,"* ACM Trans. Graph., vol. 37, no. 4, Jul. 2018. [Online]. Available: https://doi.org/10.1145/3197517.3201283

Suwajanakorn S., Seitz S. M., and Kemelmacher-Shlizerman I., *"Synthesizing obama: Learning lip sync from audio,"* ACM Trans. Graph., vol. 36, no. 4, Jul. 2017. [Online]. Available: https://doi.org/10.1145/3072959.3073640

"Deepfakes github," https://github.com/deepfakes/faceswap, accessed: 2020-02-01.

Day C., *"The future of misinformation,"* Computing in Science amp; Engineering, vol. 21, no. 01, pp. 108–108, jan 2019.

Newson A., Almansa A., Fradet M., Gousseau Y., and Pe´rez P., *"Video inpainting of complex scenes,"* SIAM Journal on Imaging Sciences, vol. 7, no. 4, pp. 1993–2019, 2014. [Online]. Available: https://doi.org/10.1137/140954933

Ebdelli M., Le Meur O., and Guillemot C., *"Video inpainting with short-term windows: Application to object removal and error concealment,"* IEEE Transactions on Image Processing, vol. 24, no. 10, pp. 3034–3047, 2015.

Xu R., Li X., Zhou B., and Loy C. C., *"Deep flow-guided video inpainting,"* in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019, pp. 3718–3727.

Qadir G., Yahaya S., and Ho A., *"Surrey university library for forensic analysis (sulfa) of video content,"* 01 2012, pp. 1–6.

"Grip dataset," http://www.grip.unina.it/, accessed: 2022-03-10.

Bestagini P., Milani S., Tagliasacchi M., and Tubaro S., *"Codec and gop identification in double compressed videos,"* IEEE Transactions on Image Processing, vol. 25, no. 5, pp. 2298–2310, May 2016.