Journal of Multimedia Information System VOL. 11, NO. 4, December 2024 (pp. 229-240): ISSN 2383-7632 (Online)
https://doi.org/10.33851/JMIS.2024.11.4.229

Evaluating Deepfake Images: An Empirical Evaluation of Select Methods with Data Engineering

Louis Echefu[1], Qingsong Zhao[1], Subhajit Chakrabarty[1*]

Abstract

In the rapidly evolving digital landscape, visual content—especially images and videos—plays a crucial role in online communication. However, the rise of deepfake technology, which employs deep learning techniques to create realistic manipulated media, raises significant ethical concerns due to its potential for misuse. This study conducts an empirical study of methods/tools for deepfake generation and detection, focusing on three prominent political figures: Vladimir Putin, Joseph Biden, and Narendra Modi. Using authentic images from the internet, we generated fake images using various deepfake tools and constructed a dataset comprising 600 real and 600 deepfake images. One of the key contributions of this paper was to integrate a data engineering approach. Among the models evaluated, the InceptionV3 model achieved the highest detection accuracy of 98.97%. Upon evaluating cross-datasets and combined datasets, we found that focused datasets improved model performance, emphasizing the importance of robust data engineering methodologies in addressing deepfake threats. This research contributes to the broader field of deepfake detection, with potential applications for other similar tasks.

Key Words: Deepfake, Machine Learning, Deepfake Creation, Deepfake Detection.

I. INTRODUCTION

In today's digital world, the internet has become a daily source for information, entertainment, and education. Visual content, particularly images and videos, has become dominant in online communication, significantly influencing how we consume information [1]. Platforms like YouTube, TikTok, Instagram, X (formerly as Twitter) and Facebook et al. have significantly contributed to this trend. In 2023, the average person spent 17 hours per week watching online videos, and 86% of consumers report spending at least a quarter of their social media time watching videos. For instance, Facebook users collectively streamed over 2 billion videos each month, with video content accounting for 50% of their time on the platform [2].

The volume of data generated online is overwhelming. In 2020, Internet users generated 64.2 zettabytes (ZB) of data, an amount that experts predict will more than double to 147 ZB by the end of 2024 [3]. A large portion of this data is composed of images and videos. However, in recent years, the rise of visual content forgery has emerged as a significant issue within the internet community and social media applications, raising serious ethical and social concerns. Online applications and Social Medias, such as Snapchat, Instagram, Facebook, Reddit et al. have used deep learning (DL) techniques to develop tools to facilitate users to create fake images and videos, which is usually called "Deepfake". The easy access of these tools makes the situation worse [4].

While deepfakes can enhance user experiences in various legitimate contexts, such as entertainment, education, industry, and marketing, they also present serious threats when exploited maliciously. Examples of misuse include spreading misinformation, inciting political discord, and even harassment [6-7]. Reports indicate that adult content platforms hosted thousands of deepfake videos, illustrating the breadth of this issue [6].

As a combination of "deep learning" and "fake," deepfakes refer to highly realistic audiovisual content created using deep learning techniques. Initially derived by a Reddit user for face-swapping in videos, the term now encompasses a range of manipulations including facial expression re-enactment, body and background alteration, and audio synthesis. While deepfakes are a product of advancements in AI, machine learning, and deep learning, the term often implies misuse for unethical or illicit purposes [4-6]. Deepfake technology typically utilizes Generative Adversarial Networks (GANs), which involve two neural networks: a

Manuscript received October 31, 2024; Revised December 07, 2024; Accepted December 10, 2024. (ID No. JMIS-24M-11-032)
Corresponding Author (*): Subhajit Chakrabarty, +1-318-795-4829, Subhajit.Chakrabarty@lsus.edu
[1]Department of Computer Science, Louisiana State University at Shreveport, Shreveport, LA, USA, echefuc43@lsus.edu, qingsong.zhao@lsus.edu, Subhajit.chakrabarty@lsus.edu

generative network and a discriminative network. The generative network, using an encoder and decoder, creates fake images or videos, while the discriminative network assesses their authenticity [10-22].

Despite extensive research on deepfake creation, detection, and dataset development, few studies have fully integrated all three aspects in a single paper. Our study aims to bridge this gap by presenting a comprehensive process for generating fake images, developing a dataset, and training detection models to improve accuracy. We focused on three prominent political figures: Vladimir Putin, Joseph Biden, and Narendra Modi. Real images were collected from Google Photos, Instagram, and YouTube to ensure a wide range of facial expressions and characteristics. High-quality fake images were generated using DeepFaceLab and FaceSwap, creating a comprehensive dataset for testing, which included 600 real and 600 deepfake images. We employed data augmentation techniques as a key part of the data engineering process, ensuring better diversity and increase robustness in training samples.

Three pre-trained models—VGG16, MobileNet, and InceptionV3—were used for deepfake detection, with InceptionV3 achieving the highest accuracy of 98.97%. While cross-dataset evaluations revealed limitations in the model's generalizability, training on a combined dataset improved accuracy to 72.76% with batch normalization. However, further modifications such as dropout and unfreezing pre-trained layers led to a drop in performance, emphasizing the importance of preserving critical pre-trained features. Our findings contribute to the broader field of deepfake detection and may be generalized to similar detection task.

One of the key contributions of this paper is that we used a combination of data engineering techniques to generate deepfake datasets that are balanced. So, in this paper we are not using any benchmark dataset, but showcasing our data engineering approach to create our own deepfake datasets, suitable for our purpose. Another key contribution is that we perform empirical evaluation of existing state-of-the-art deep learning architectures on the generated dataset, for our task.

The remainder of the paper is organized as follows. In Section II, we review related works – mentioning some deepfake datasets, generation tools, detection tools and cross-dataset evaluation. In Section III, we describe our dataset and data engineering methods. In Section IV, we described the experiments and results on cross dataset and combined dataset. In the last Section, we discussed our overall findings.

II. RELATED WORKS

This section reviews deepfake datasets, generation models, detection techniques, and cross-dataset generalization issues. Key datasets like FaceForensics++ and DFDC have driven advancements in detection, while tools such as DeepFaceLab and FaceSwap use GANs for high-quality deepfake creation. CNN models like XceptionNet, VGG16, and InceptionV3 show strong detection results on individual datasets but struggle with generalization across unseen datasets. Combining multiple datasets improves model robustness and accuracy.

2.1. Deepfake Datasets

In the research of deepfake, dataset plays critical role in improving the accuracy of detection algorithms. There are several public datasets have been widely adopted and each offers unique characteristics and challenges for model training and evaluation. Here we check out FaceForensics++, the DeepFake Detection Challenge Dataset (DFDC), Celeb-DF, and DF-TIMIT.

FaceForensics++ is a benchmark dataset consisting of manipulated videos created using four different methods, including DeepFakes [12], Face2Face [14], FaceSwap [14] and NeuralTextures [15]. This dataset provides a large corpus of manipulated and original videos, making it one of the most utilized resources for training and testing deepfake detection models [15]. Similarly, the DFDC dataset [16], released by Meta in 2020 in collaboration with various academic institutions, provides a diverse set of real and deepfake videos aimed at supporting research in developing more robust detection algorithm

Celeb-DF dataset includes 590 original videos collected from YouTube with subjects of different ages, ethnic groups and genders, and 5,639 corresponding DeepFake videos [18]. Celeb-DF was introduced to address limitations in earlier datasets such as low visual quality.

This high-quality dataset consists of celebrity videos and has become a key benchmark in evaluating the performance of deepfake detection models [20].

Compared to other datasets, DF-TIMIT is smaller but it offers a quality-focused dataset where GAN-based methods are used for face swapping. This feature makes it a useful tool, especially for assessing deepfake manipulation techniques [21].

2.2. Deepfake Generation Tools

Generative Adversarial Networks (GANs) and its variants are the most popular ones used in the generation of deepfake images and videos since they can produce high-quality, realistic images [22]. The most prominent tools for generating deepfakes are DeepFaceLab, DeepFakeSwap, DeepSwap, and FaceSwap. DeepFaceLab is an open-source framework that allows for the creation of realistic deepfakes by using deep learning models for face swapping. It

has become a popular tool due to its user-friendly interface and the high quality of generated content [23]. Similarly, FaceSwap is another widely used platform for facial manipulation, leveraging GANs to align facial features and produce photorealistic results [16].

DeepSwap can also be used to create high-resolution face swaps with minimal effort. This tool is user friendly and accessible to both casual users and professionals, enabling the creation of highly realistic outputs with ease. On the other hand, DeepFakeSwap allows users to fine-tune aspects such as facial alignment and blending. This makes it particularly appealing to advanced users who require greater control over the details in generating deepfakes [40].

Another model worth mentioning is the First Order Motion Model for Image Animation, which has also gained popularity for its ability to animate still images by generating motion fields from source videos, further enhances the quality of deepfakes [19].

2.3. Deepfake Detection Tools

As a primary defense against the growing threat of deepfakes, the field of deepfake detection has gained significant attention from researchers and experts in recent years. This focus has led to the development of numerous detection techniques aimed at identifying manipulated media. Tools like Microsoft's Video Authenticator is used to analyze manipulation confidence scores, while FakeBuster detects fake video conferencing through deep learning and facial segmentation. Similarly, FakeCatcher innovatively uses photoplethysmography to identify subtle biological cues like pulse variations in manipulated videos. These tools help in detecting deepfakes [5].

Various deep learning architectures have been extensively studied for deepfake detection, demonstrating varying degrees of success depending on the specific dataset and application.

In Rana et al. [5] the systematic review of 112 studies (2018−2020) highlights that deep learning, particularly CNNs, is widely used for deepfake detection. The most commonly used dataset is FaceForensics++. Detection accuracy is the key performance metric, and deep learning models generally outperform non-deep learning approaches. XceptionNet, VGG16, InceptionV3 and MobileNet are some key deep learning architectures that have been deployed in in this domain. They are briefly introduced as follows.

XceptionNet, a convolutional neural network (CNN) architecture, has been one of the most frequently used models for detecting manipulated images and videos. Some studies show that XceptionNet has achieved detection accuracies ranging from 90% to 99% on datasets such as FaceForensics++ and Celeb-DF, which makes it one of the top-performing models in Deepfake detection area [24]. VGG16 is a convolutional neural network (CNN) model with 16 layers, designed for image classification tasks. It can categorize images into 1,000 object classes, including animals, objects like keyboards, and more. VGG16 improves upon AlexNet [25]. by using smaller 3×3 kernel-sized filters and consistently applying convolution and max-pooling layers throughout the network. Despite being an older architecture, VGG16 remains effective in detecting manipulated content, particularly in specialized datasets [26]. InceptionV3, a convolutional neural network (CNN) that helps with image analysis and object detection, has demonstrated exceptional performance in detecting deepfakes, achieving an accuracy of over 98% in some experiments. However, its generalization to unseen datasets is often limited, as its performance tends to degrade when tested across different datasets [26]. EfficientNet, a family of convolutional neural networks (CNNs) developed by Google in 2019, designed for image classification tasks, has become a powerful tool for detecting deepfakes, especially when working with larger datasets [27]. MobileNet is also used in deepfake detection due to its lightweight and efficient design, making it particularly suitable for deployment in resource-constrained environments like mobile devices [9].

Another approach that has been used in this domain, was Vision Transformers (ViTs). This method effectively captured fine-grained patterns in manipulated images and videos, achieving a high detection accuracy [36]. Error-level analysis (ELA) combined with deep learning, has been performed to leverage compression artifacts to identify inconsistencies in manipulated content. This enhanced both the accuracy and computational efficiency of deepfake detection [37]. Additionally, blockchain-based methods have successfully been used in fake news detection [41].

2.4. Cross-Dataset Evaluation

Cross-dataset evaluations refer to the process of assessing the performance of machine learning models, particularly in deepfake detection, across different datasets. Cross-dataset evaluations have shown that even high-performing models like XceptionNet and InceptionV3 experience a significant drop in accuracy when applied to unseen datasets, which indicates the limited generalizability of models across different datasets. Combining multiple datasets for training has proven to improve model generalization. For example, combining data from FaceForensics++ and Celeb-DF has been shown to enhance performance by incorporating a broader range of manipulated and real images, resulting in higher detection accuracy [31].

III. METHODS

The study collected data on Vladimir Putin, Joseph Biden, and Narendra Modi from sources like Google Photos, Instagram, and YouTube, focusing on diverse images. Tools such as DeepFaceLab and FaceSwap generated high-quality deepfakes. The final dataset included 600 real and 600 fake images, providing a comprehensive base for training deepfake detection models.

3.1. Data Collection

The process of gathering data for our study was essential since the performance of deepfake detection algorithms is directly impacted by the quality and diversity of the photos. We concentrated on three well-known political figures: Vladimir Putin, Joseph Biden, and Narendra Modi, to create an extensive dataset of authentic pictures of these individuals. The images were taken from publicly available sites such as YouTube, Instagram, and Google Photos, thus a wide range of each person's expressions and face characteristics were captured.

3.1.1. Google Photos

We collected high-resolution photos of the three presidents using Google Photos. The search was customized to encompass a range of situations, including formal functions, private gatherings, and public appearances. To ensure the diversity of the dataset, keywords covering a wide range of expressions and viewpoints were carefully chosen.

After retrieving the images, we filtered the data to get rid of duplicates and poor-quality photos that might interfere with the training process. Images were also examined to ensure they were clear and accurately represented the subject without significant distortion.

3.1.2. Instagram

Popular social media site Instagram offered a wealth of photos of the presidents in more relaxed environments. To collect real photos, we concentrated on official accounts and verified posts from reliable sources. Using this site to gather pictures of the presidents at different times of day and in different lighting situations was helpful.

3.1.3. Youtube

Frames from the three presidents' speeches, interviews, and public broadcasts were taken from YouTube, a huge collection of video footage. Videos were chosen for extraction based on their clarity and resolution and verified sources.

We took frames at predetermined intervals, concentrating on the presidents' distinct face features and clear visibility as in Fig. 1. The dynamic range of expressions and situations typical of video content was ensured in the dataset by this strategy, which is important for training models that will be applied to both static photos and video streams in the future.

(a) Biden (b) Putin

(c) Modi

Fig. 1. Presidents' distinct faces.

3.2. Deepfake Generation

The quality of the training data has a big impact on how accurate deepfake detection models are. Two well-known programs, DeepFaceLab and FaceSwap, were employed to make sure our deepfake dataset was comprehensive. These tools were picked because they have a track record of creating high-quality deepfakes, which are essential for carefully evaluating and testing the limits of detection algorithms [4].

The first step in creating a deepfake is to identify and align faces in the original photos as shown in Fig. 2. After aligning, the tools transfer the presidents' facial traits onto target images to accomplish face swapping. This technique necessitates exact control over facial features including lighting and expression. The exchanged faces were flawlessly integrated with their backdrops using advanced blending techniques to produce as realistic looking deepfakes as possible [4]. Post-processing techniques like color correction and edge blending further refine this integration, which enhance the visual fidelity of the deepfakes and make them more challenging for detection models to identify.

Fig. 2. Deepfake image of Putin using FaceSwap.

Table 1. Images in the dataset.

	Number of real images	Number of deepfake images
Modi dataset	200	200
Biden dataset	200	200
Putin dataset	200	200
Combined	600	600

The balanced proportion of real and fake images was included in the datasets for each president, which was essential for developing and testing the detection models. The Modi Dataset was made up of 200 deepfake images and 200 genuine images that were gathered from different sources. The 200 genuine photos in the Biden Dataset were gathered from several sources, and 200 deepfake images were generated. The Putin Dataset included 200 deepfake photos in addition to 200 authentic images gathered from other sources (Table 1).

After preparing individual datasets for each president, they were combined into a comprehensive dataset containing 600 real images and 600 fake images. This merged dataset is designed to train and evaluate deepfake detection models on a larger scale.

Our method for detecting deepfakes involved a systematic approach using a combination of pre-trained convolutional neural network (CNN) models, tailored preprocessing, and extensive evaluation across multiple datasets. The following (Fig. 3) outlines the key steps we took to develop and evaluate the accuracy of our deepfake detection models.

3.3. Preprocessing

The preprocessing method involved several key steps Fig. 3. The first step in our method was to preprocess the image data to ensure consistency and optimize the input for the models. We resized all images to 255×255 pixels, a standard size that balances computational efficiency with the need to maintain sufficient detail in the images. This resizing step was critical for ensuring that the pre-trained models could process the images effectively and make accurate predictions.

Each pixel value in the images was normalized to a range of [0, 1] by dividing by 255. This step helps in speeding up the convergence of the training process by ensuring that the input data has a consistent scale, preventing any one feature from dominating the learning process.

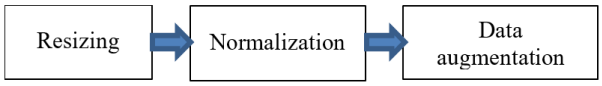


Fig. 3. Data processing pipeline.

To enhance the performance of the models, data augmentation techniques such as random rotations, flips,and shifts were applied to the training images. In practice, augmentations help in simulating different real-world scenarios, making the models more resilient to variations in input image [32].

3.4. Data Engineering

Our data was pre-processed through resizing and data augmentation, as mentioned. Our data engineering approach was to perform (a) single dataset learning with cross dataset evaluation, and (b) combined dataset learning with cross dataset evaluation.

In the single dataset learning, we identified images that were difficult to learn, augmented them, and then identified the best model. In the combined dataset learning, we took the best performing model, added layers that improved accuracy, and retrained the model.

We implemented MixBoost, a mask-based augmentation technique targeting critical image regions like facial features prone to manipulation. This enhanced the model's ability to detect subtle differences between real and fake images [38].

Further, we implemented Smart Augmentation, combining existing samples to create hybrid examples that amplify challenging features in the dataset [39].

3.5. Models Used

To detect deepfakes, we selected three widely recognized pre-trained models: VGG16 [26], MobileNet [24] and InceptionNet [27]. They were chosen for their architectures and track records in broader image classification tasks. Their architectures are shown in Fig.s 4-6. This paper did not seek to report the performance of all known models, exhaustively, but to showcase the integration of data engineering in this domain.

The models were trained on dataset of each president and evaluated for identifying deepfakes within the specific context [32]. We experimented with adding dense layer, batch normalization and dropout to improve the best model.

IV. EXPERIMENTS AND RESULTS

This section presents a systematic approach to evaluating and improving deepfake detection models across various datasets, emphasizing both achievements and opportunities for further enhancement.

4.1. Same Dataset Experiment and Evaluation

We used the same three pre-trained models with each dataset. After each model was refined using the president-spe-

cific datasets, it was evaluated.

4.1.1. Narendra Modi Dataset

Using this dataset, each model was trained, and its performance was evaluated using the validation set. The models were able to distinguish between real and fake images with varying degrees of accuracy. According to the results (Table 2), InceptionV3 and VGG16 were the two models that performed the best on this dataset.

The MobileNetV2 model showed lower accuracy, potentially due to its lightweight architecture, which might not have captured the complex features of the dataset as effectively as the other models.

4.1.2. Vladimir Putin Dataset

The InceptionV3 model significantly outperformed (98.26%) the others on the Putin dataset, suggesting its superior ability to learn from the complex patterns in this dataset. MobileNetV2 also performed well, while VGG16 showed comparatively lower accuracy (Table 3).

4.1.3. Joseph Biden Dataset

The models were trained on the Biden dataset yielded the highest accuracy scores across all the datasets.

Among the models for the Biden dataset, the Inception V3 has the highest accuracy score of 98.97% (Table 4).

4.2. Cross Dataset Experiment and Evaluation

Since the InceptionV3 model performed the best on each individual dataset, we chose to use this model exclusively for our cross-dataset evaluation. The goal was to assess how well the model, trained on one president's dataset, could generalize when tested on the datasets of the other two presidents.

When the InceptionV3 Model trained on the Biden dataset was tested on the datasets of Putin and Modi, it achieved accuracy rates of 48.7% and 50%, respectively, as seen in Fig. 4. This suggests a little decline in performance, underscoring the difficulties in predicting to new data. These findings verify that although the InceptionV3 model performs well on individual datasets, its generalizability across many datasets is constrained.

The differences in accuracy point to the need for more improvement in deepfake detection algorithms' cross-dataset generalization.

4.3. Combined Dataset Experiment and Evaluation

We combined the datasets of the three presidents to produce a dataset containing 600 real photographs and 600

Table 2. Results of modi dataset.

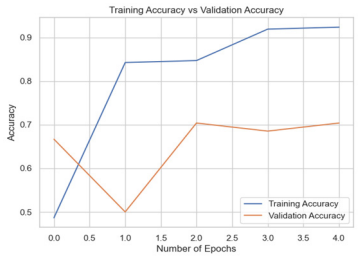
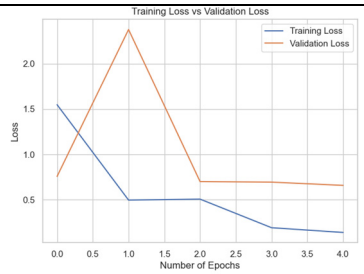
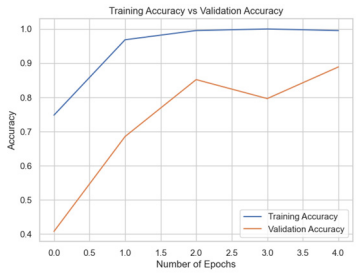
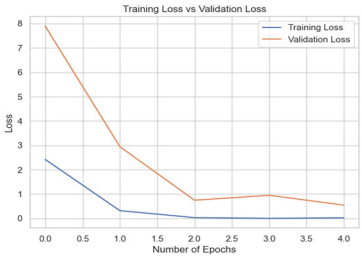
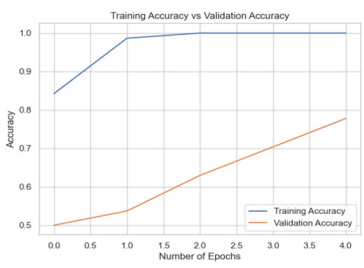
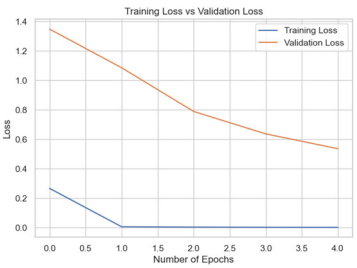
Model	Training accuracy and validation accuracy graph	Training loss and validation loss graph	Result (%)
VGG16			70.37
InceptionNet			77.78
MobileNet			68.52

Table 3. Results of putin dataset.

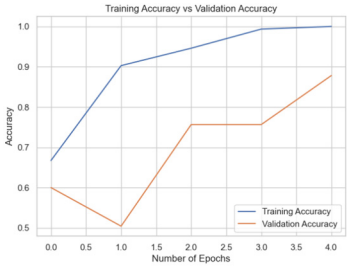
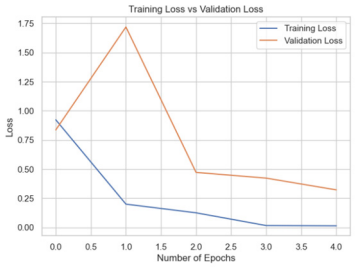
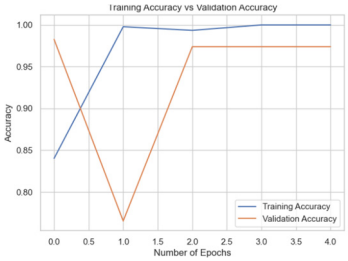
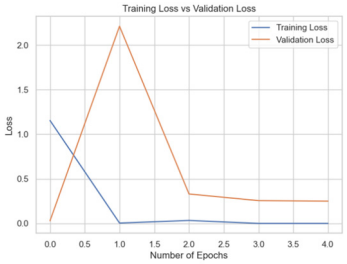
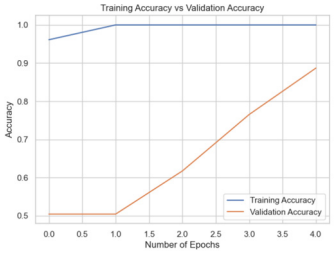
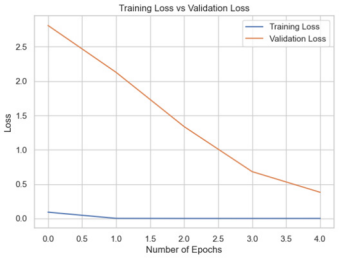
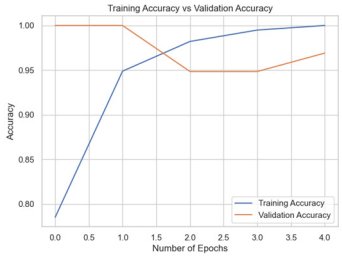
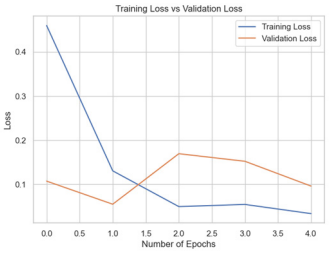
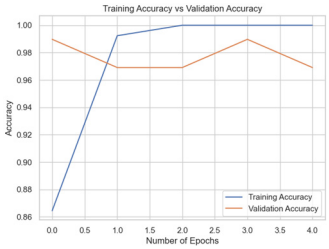
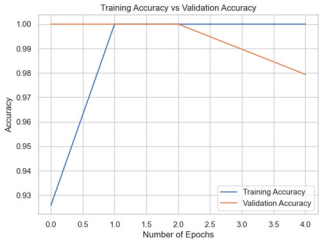
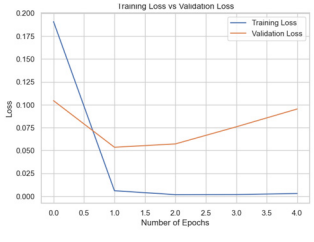
Model	Training accuracy and validation accuracy graph	Training loss and validation loss graph	Result (%)
VGG16			89.57
InceptionNet			98.26
MobileNet			88.70

Table 4. Results of biden dataset.

Model	Training accuracy and validation accuracy graph	Training loss and validation loss graph	Result (%)
VGG16			96.91
InceptionNet			98.97
MobileNet			97.94

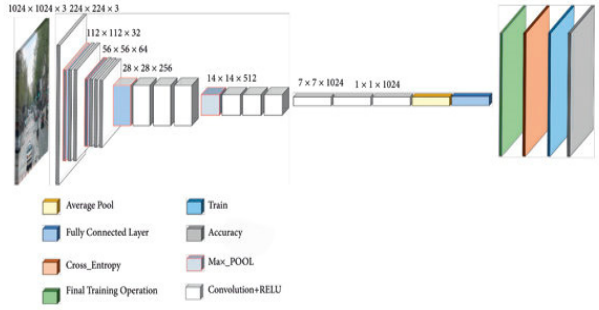


Fig. 4. MobileNet architectural diagram.

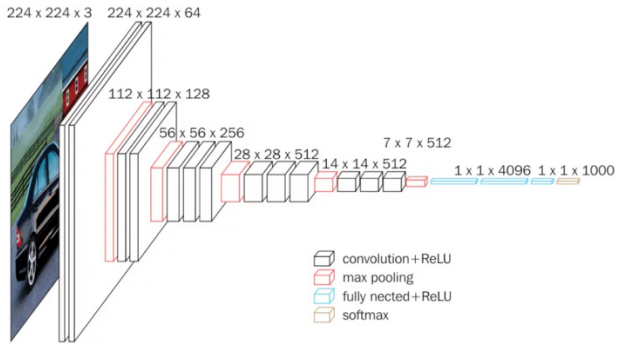


Fig. 5. VGG16 architectural diagram.

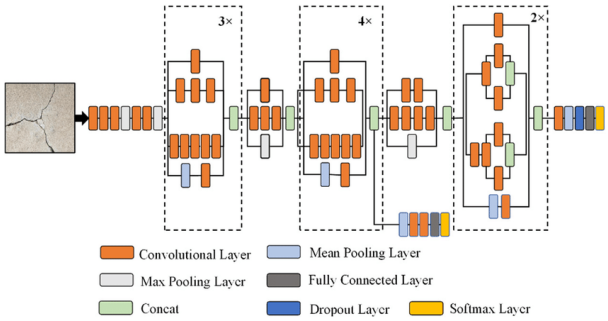


Fig. 6. InceptionNet architectural diagram.

fake images to build a more broadly applicable model. For additional testing, the InceptionV3 model was selected since it had demonstrated the highest accuracy in individual tests (Table 5).

By including Batch Normalization in between the dense layers, we were able to increase the accuracy of the model to 72.76% for the entire dataset. Reducing overfitting by adding dropout layers led to a minor decline in accuracy to 71.27%. We retrained the InceptionV3 model on the combined dataset by unfreezing its previous layers. However, this method resulted in a significant reduction in accuracy to 41%, indicating that unfreezing the layers disrupted significant characteristics included in the pre-trained layers. The curve shows an overfitting issue. This experiments also highlights that training on focused datasets would likely do better.

4.4. Discussion

The InceptionV3 model outperformed the other models consistently, as demonstrated by the experiments conducted on all datasets. It achieved the best accuracy rates on the Biden and Putin datasets and showed strong performance on the Modi dataset. The results of the cross-dataset experiments (Fig. 7) indicate that while the InceptionV3 model performs well on individual datasets, its generalizability across different datasets is constrained.

When tested on the combined dataset, the InceptionV3 model's performance initially improved with modifications such as adding Batch Normalization, but accuracy decreased when additional techniques like dropout were applied. Moreover, unfreezing the pre-trained layers and retraining the model resulted in a significant drop in accuracy. This suggested that using InceptionV3 with dense layer and batch-normalization layer worked the best.

However, the algorithms likely picked up the differences between Biden and Putin well, but not the differences between real and fake. The training curve on combined datasets showed high accuracy, but it could generalize to unseen data well on the combined dataset – indicating the importance for data engineering.

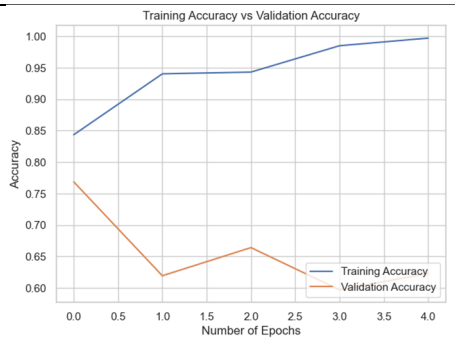
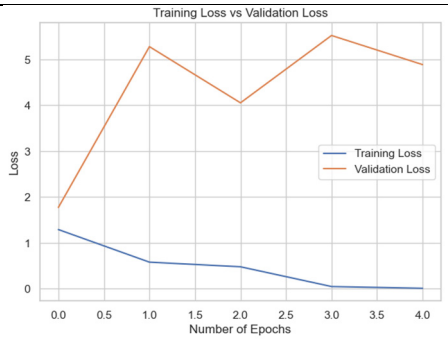
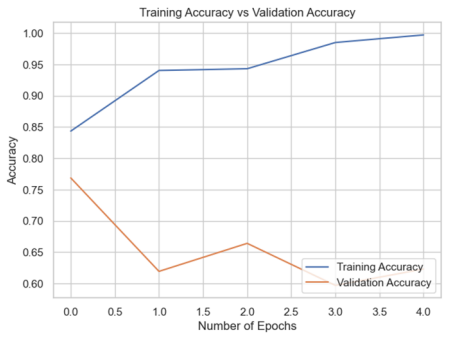
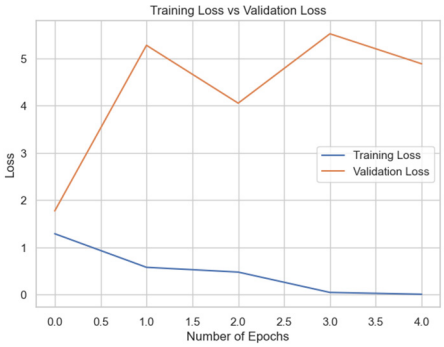
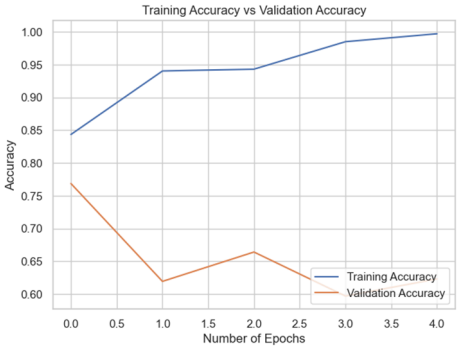
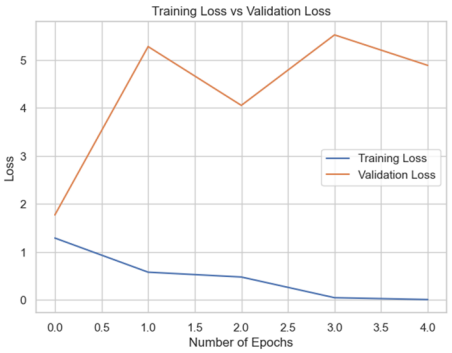
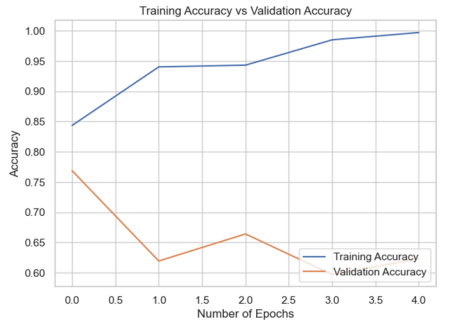
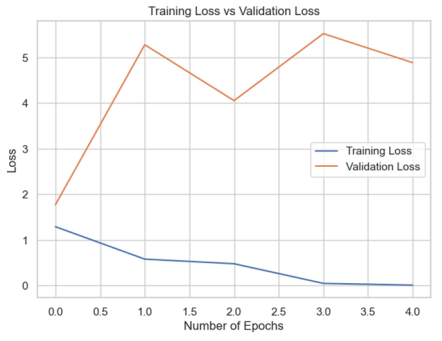
V. DISCUSSION AND CONCLUSION

This study evaluates deepfake generation and detection methods, focusing on political figures Vladimir Putin, Joseph Biden, and Narendra Modi. We created a dataset of 600 real and 600 deepfake images using tools like DeepFaceLab and FaceSwap. Our experiments with pre-trained CNN models revealed that InceptionV3 achieved the highest accuracy at 98.97%. However, while it excelled on individual datasets, its performance declined in cross-dataset evaluations, highlighting challenges in generalization. With combining datasets, our own pre-trained models performed well with accuracy of 72.76%. But, applying dropout layers and unfreezing model layers, decreased performance, indicating the importance of preserving the pre-trained features. To improve the model's generalization capabilities, we can enhance the dataset by collecting more focused and representative data. Diversity in images (different persons) is not contributing to detection of fake. In layman language, we will get higher accuracy if we use real and fake data of one person, rather than using a larger real and fake dataset of multiple persons.

Our findings emphasize the need for data engineering and focused training datasets to enhance deepfake detection methodologies, contributing valuable insights applicable across various domains affected by manipulated media.

Future research will aim to enhance deepfake detection by expanding the dataset to include a broader range of po-

Table 5. Results of combined dataset.

Model	Training accuracy and validation accuracy graph	Training loss and validation loss graph	Result (%)
Inception model with dense layer			61.57
Add batch Normalization between dense layers			72.76
Drop out layer between dense layers			71.27
Unfreeze and retrain the best (batch normalization between dense layers)			41.04

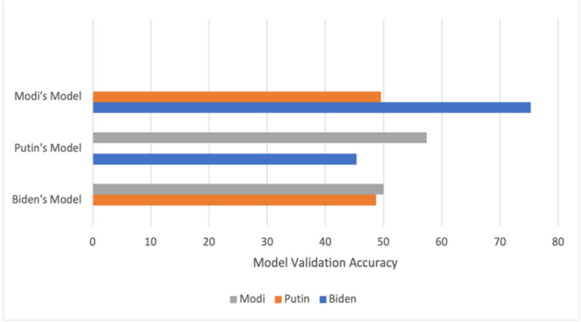


Fig. 7. InceptionV3 model performance on cross-dataset.

litical figures and contexts, which will improve model generalization.

REFERENCES

[1] A. A. Mohammad, "Impact of new digital media on conventional media and visual communication in Jordan," *Journal of Engineering, Technology, and Applied Science (JETAS),* vol. 4, no. 3, pp. 105-1139, Dec. 2022.

[2] SproutSocial, 50+ Social Media Video Marketing Statistics for 2024, Mar. 2024. https://sproutsocial.com/ insights/social-media-video-statistics.

[3] Edge Delta, Breaking Down The Numbers: How Much Data Does The World Create Daily in 2024? Mar. 2024. https://edgedelta.com/company/blog/how-much-datais -created-per-day.

[4] S. S. Henrique, M. Bethany, A. M. Votto, I. H. Scarff, N. Beebe, and P. Najafirad, "Deepfake forensics analysis: An explainable hierarchical ensemble of weakly supervised models," *Forensic Science International: Synergy,* vol. 4, p. 100217, 2022.

[5] M. S. Rana, M. N. Nobi, B. Murali, and A. H. Sung, "Deepfake detection: A systematic literature review," *IEEE Access*, vol. 10, pp. 25494-25513, 2022.

[6] G. Patrini, F. Cavalli, and H. Ajder, The State of DEEPFAKES: Reality under Attack, 2018. https://s3.eu-west-2.amazonaws.com/rep2018/2018-the-state-of-deepfakes.pdf.

[7] J. Kietzmann, L. Lee, I. P. McCarthy, and T. C. Kietzmann, "Deepfakes: Trick or treat?," *Business Horizons*, vol. 63, no. 2, pp. 135-146, 2020.

[8] M. Westerlund, "The emergence of deepfake technology: A review," *Technology Innovation Management Review,* vol. 9, no. 11, pp. 39-52, 2019.

[9] B. U. Mahmud and A. Sharmin, "Deep insights of deepfake technology: A review," *arXiv Prep. arXiv:2105. 00192*, 2021.

[10] G. Oberoi, Exploring DeepFakes, Jan. 2021. https:// goberoi.com/exploring-deepfakes-20c9947c22d9.

[11] J. Hui, How Deep Learning Fakes Videos (Deepfake) and How to Detect It, Jan. 2021. https://medium. com/how-deep-learning-fakes-videos-deepfakes-and-how-to-detect-itc0b50fbf7cb9.

[12] Deepfakes GitHub, Faceswap, Oct. 2018. https://github.com/deepfakes/faceswap.

[13] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2Face: Real-time face capture and reenactment of RGB videos," in *IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2016, pp. 2387-2395.

[14] M. Kowalski, "FaceSwap," Github, 2018. https://github.com/MarekKowalski/FaceSwap.

[15] J. Thies, M. Zollhöfer, and M. Nießner, "Deferred neural rendering: Image synthesis using neural textures," *ACM Transactions on Graphics (TOG)*, vol. 38, no. 4, pp. 1-12, 2019.

[16] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Seoul, Korea, Nov. 2019, pp. 1-11.

[17] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C. Ferrer, "The deepfake detection challenge dataset," *arXiv Prep. arXiv:2006.07397,* 2020.

[18] Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics, 2024. GitHub Repository, https://github.com/yuezunli/celeb-deepfakeforensics.

[19] First Order Motion Model, GitHub Repository. 2024. https://aliaksandrsiarohin.github.io/first-order-model-website/.

[20] Y. Li, P. Sun, H. Qi, S. Lyu, and J. Yang, "Celeb-DF: A large-scale challenging dataset for deepfake forensics," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3207-3216, Jun. 2020.

[21] P. Korshunov and S. Marcel, "Deepfakes: A new threat to face recognition? Assessment and detection," *arXiv Prep. arXiv:1812.08685* 2018.

[22] I. Goodfellow, J. P. Abadie, M. Mirza, B. Xu, D. W. Farley, and S. Ozair, et al., "Generative adversarial nets," in *Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS)*, Montreal, QU, Dec. 2014, pp. 2672-2680.

[23] DeepFaceLab, GitHub Repository, https://github.com/ iperov/DeepFaceLab, 2020.

[24] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, Jul. 2017, pp. 1251-1258.

[25] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Advances in Neural Information Processing Systems*, 2012, pp. 1097-1105.

[26] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prep. arXiv:1409.1556,* 2014.

[27] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, and D. Anguelov, et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, Jun. 2015, pp. 1-9.

[28] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: A compact facial video forgery detection network," *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2018, pp. 1-7.

[29] A. Sankaranarayanan, M. Groh, R. Picard, and A. Lippman, "The presidential deepfakes dataset," in *CEUR Workshop Proceedings*, 2021, vol. 2942, pp. 57-72.

[30] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the International Conference on Machine Learning (ICML)*, 2019.

[31] P. Korshunov and S. Marcel, "Vulnerability assess-

ment and detection of deepfake videos," *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020, pp. 16-19.

[32] F. M. Abu-Naser, "Classification of real and fake human faces using deep learning," *International Journal of Academic Engineering Research*, pp. 1-14, 2022.

[33] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv Prep. arXiv:1409.1556,* 2014.

[34] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, and T. Weyand, et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," *arXiv Prep. arXiv:1704.04861*, 2017.

[35] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, and D. Anguelov, et al., "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 1-9.

[36] Y. J. Heo, W. H. Yeo, and B. G. Kim, "Deepfake detection algorithm based on improved vision transformer," *Applied Intelligence*, vol. 53, no. 7, pp. 7512-7527, 2023.

[37] R. Rafique, R. Gantassi, R. Amin, J. Frnda, A. Mustapha, and A. H. Alshehri, "Deep fake detection and classification using error-level analysis and deep learning," *Scientific Reports*, vol. 13, no. 1, p. 7422, 2023.

[38] Y. Liu and H. Zhu, "MixBoost: Improving the robustness of deep neural networks by boosting data augmentation," *arXiv Prep. arXiv:2212.04059,* 2022.

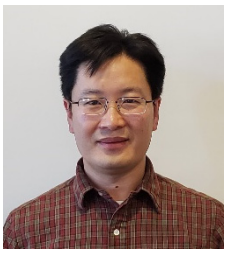
[39] J. Lemley, S. Bazrafkan, and P. Corcoran, "Smart augmentation: Learning an optimal data augmentation strategy," in *IEEE International Conference on Machine Learning for Signal Processing (MLSP)*, 2017.

[40] S. Alanazi and S. Asif, "Exploring Deepfake technology: Creation, consequences and countermeasures," *Human-Intelligent Systems Integration*, pp. 1-12, 2024.

[41] S. K. Kim, J. H. Huh, and B. G. Kim, "Artificial intelligence blockchain based fake news discrimination," *IEEE Access*, vol. 12, pp. 53838-53854, 2024.

AUTHORS

Louis Echefu received his B.S. degree from Bells University, Nigeria, and his M.S. degree in Computer Science from Louisiana State University at Shreveport, USA, in 2024. His research interests include machine learning, cybersecurity, and robotics.



Qingsong Zhao received his Ph.D. degrees in the Institute of Software Chinese Academy of Sciences in 2023. In 2020, he joined the Department of Computer Science at Louisiana State University at Shreveport as an Assistant Professor. His research interests include Cybersecurity, Operating System, and Healthcare IT.

Subhajit Chakrabarty is an Associate Professor with the Dept. of Computer Science at LSU Shreveport Louisiana, USA. With dual Ph.D. in Computer Science and International Business, he has over three decades of experience spanning IT management, academia, and government service. His research focuses on inter-disciplinary applications of Artificial Intelligence / Machine Learning.