SEMARAK ILMU
PUBLISHING
202103268166(003316878-P)

Journal of Advanced Research in Applied Sciences and Engineering Technology

Journal homepage:
https://semarakilmu.com.my/journals/index.php/applied_sciences_eng_tech/index
ISSN: 2462-1943

SEMARAK ILMU
Journal of
ADVANCED RESEARCH IN APPLIED SCIENCES AND ENGINEERING TECHNOLOGY
ISSN 2462 - 1943

Advances in DeepFake Detection: Leveraging InceptionResNetV2 for Reliable Video Authentication

Nan Mad Sahar[1,*], Muhammad Faris Syazwan Mohd Rozi[1], Nor Surayahani Suriani[1], Suhaila Sari[1], Shuhaida Ismail[2], Azrul Amri Jamal[3] , Ahmed Marwan Aleesa[4]

[1] Faculty of Electrical and Electronic Engineering, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
[2] Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400 Batu Pahat, Johor, Malaysia
[3] Faculty of Informatics and Computing, Universiti Sultan Zainal Abidin, Besut Campus, 22200 Besut, Terengganu, Malaysia
[4] Department of Programming, College of Computer Science and Information Technology, University of Kirkuk, 52001, Kirkuk, Kirkuk Governorate, Iraq

ABSTRACT

With the widespread availability of free AI-powered mobile applications, the creation of realistic and deceptive videos known as "DeepFakes" has become increasingly effortless. Detecting the authenticity of such videos poses a formidable challenge due to the scarcity of discernible traces. In this study, we delve into the application of an image processing algorithm, specifically InceptionResNetV2, for deepfake detection in videos. The amalgamation of ResNet and Inception models in InceptionResNetV2 yields superior accuracy, making it an ideal choice for our investigation. Our research entails the development of an Android-based application employing the proposed algorithm, with MIT App Inventor for application creation and Google Colab for the prediction model. Furthermore, we conduct a thorough evaluation of the application's performance in detecting deepfake videos, employing three models-InceptionResNetV2, EfficientNetB0, and ResNet50-across two datasets: DFDC and CelebDF, totalling 400 datasets. Notably, InceptionResNetV2 outperformed the other models, achieving an accuracy of 93.20% for the DFDC dataset and an impressive accuracy of 97.72% for the CelebDF dataset. Conversely, ResNet50 exhibited the lowest accuracy for the DFDC dataset, at a mere 52.28%, while EfficientNetB0 displayed the lowest accuracy for the CelebDF dataset, measuring 30.81%. This journal publication sheds light on the advancements in deepfake detection, focusing on the efficacy of InceptionResNetV2 as a powerful tool for reliable video authentication and providing valuable insights into the comparative performance of different models in tackling this complex issue.

Keywords:

DeepFake detection; Image processing algorithm; InceptionResNetV2; Video authentication; Human factor analysis

1. Introduction

The proliferation of deepfake videos across social media platforms has surged in tandem with the availability of cutting-edge technology. Deepfakes involving the manipulation of digital media, notably images and videos by replacing the original subject's appearance, have emerged as a

[*] *Corresponding author.*
E-mail address: nan@uthm.edu.my

https://doi.org/10.37934/araset.62.1.90105

profound societal challenge. This phenomenon's versatile application for malicious purposes is evident; examples encompass impersonating renowned Hollywood personalities, disseminating false narratives, and generating baseless rumours. Researchers have highlighted the potential societal impacts of deepfakes, noting their ability to erode trust in media and manipulate public perception as in previous studies [1,2]. Instances like the contrived video featuring Barack Obama in 2018 and the alterations to Joe Biden's videos during the 2020 US election as discussed by Vaccari *et al.,* [3] underscore the detrimental implications of such manipulative deepfake practices. The amplification of misinformation on social media as illuminated by Brownlee *et al.,* [4] further underscores the potential harm inflicted by these deceitful applications of deepfakes. At the core of this issue, generative adversarial networks (GANs), a sophisticated type of deep learning technology, are employed to create persuasively counterfeit photos and videos that defy easy differentiation from authentic counterparts. The effectiveness of such models is heavily reliant on copious training data, facilitating the production of deepfake content that appears astonishingly authentic and trustworthy. The existence of media showcasing prominent figures, including political leaders and celebrities, on social media platforms amplifies the feasibility of fabricating plausible falsehoods and misleading information. Johnson *et al.,* [5] aptly highlight the potential negative ramifications for society in this regard. To address this evolving challenge, our study harnesses Convolutional Neural Networks (CNNs) for a comprehensive analysis focused on identifying instances of deepfake manipulation.

In the realm of expansive image data on the internet, the potential for cultivating more sophisticated and robust models and algorithms to index, retrieve, organize, and interact with multimedia data is evident. However, the means to effectively harness and structure this wealth of data remains an ongoing conundrum. The work by Jia Deng in 2009 introduces "ImageNet," a substantial image ontology constructed upon the foundation of the WordNet structure. With the aim of populating a substantial portion of WordNet's 80,000 synsets with an average of 500–1000 high-resolution images, ImageNet serves as a pivotal initiative in this regard [6].

Over the past few years, the development of several techniques for manipulating faces within videos has proven remarkably successful and accessible to the general public. Notable examples include FaceSwap and deepfake technologies. These methodologies empower individuals to effortlessly edit faces within video sequences, yielding astonishingly realistic outcomes with minimal effort. While these tools offer substantial utility across various domains, their misuse carries substantial societal repercussions, encompassing the dissemination of fake news and the cyberbullying potential through manipulated revenge pornography. Addressing the crucial task of objectively detecting manipulated faces within video sequences becomes imperative. The work by N. Bonettini in 2020 delves into face manipulation detection within videos, focusing on modern facial manipulation techniques. Notably, they explore the ensemble of different trained Convolutional Neural Network (CNN) models. This approach involves deriving distinct models from a foundational network, such as EfficientNetB4, incorporating concepts like attention layers and Siamese training. The authors illustrate the promise of these combined networks in detecting face manipulation, demonstrating impressive results across publicly available datasets comprising over 119,000 videos [7].

Recent years have witnessed the rapid proliferation of deepfake videos on social media, fuelled by the accessibility of advanced technology. This surge has introduced concerns about the quality enhancement of these deepfakes, as evidenced by the work of Agarwal *et al.,* this advancement has led to increased apprehension due to the potential use of such convincing fake media for nefarious purposes like fake terrorism and blackmail. Consequently, industry and government stakeholders have been prompted to address the issue and curtail its exploitation. While these AI-generated fake media might appear realistic at a cursory glance, Agarwal *et al.,* proposed method, which involves frequency domain analysis followed by classification, uncovers unnatural features that elude the naked eye. Their study, which evaluates the approach using a dataset of deepfake images collected from various websites, exhibits promising potential in detecting these deceptive images [8].

Furthermore, the escalating adoption of advanced mobile phones has led to a surge in Android application users. This growth, explored by H. Soni *et al.,* has unfortunately attracted malicious actors who develop vindictive Android applications to pilfer sensitive data and engage in fraudulent activities involving mobile banks and wallets. Despite the availability of various tools for detecting malicious applications, the need for an efficient and effective tool to address the evolving complexities of new malicious apps crafted by intruders or hackers remains. H. Soni *et al.,* address this challenge through the implementation of Machine Learning techniques for identifying malicious Android applications. They rely on a dataset of past malicious apps and employ Help Vector Machine and Decision Tree algorithms for comparison with a training dataset. The results of their trained dataset exhibit the capability to predict unknown or new malware mobile applications with an accuracy of up to 93.2% [9].

Compelling deepfake videos have proliferated rapidly, deceiving even seasoned experts. This phenomenon's far-reaching impacts across politics, society, and personal lives have been addressed by A. Rahman *et al.,* Notably, state-of-the-art machine learning studies have demonstrated commendable success in detecting fake videos within high-resolution and extended video data. However, these achievements are not mirrored in low-resolution and short-time clips. In response, A. Rahman *et al.,* employ a Convolutional Neural Network (CNN) trained to identify fake videos specifically in low-resolution and short-time video data. Their experiment draws from datasets such as Kaggle Deepfake Detection Challenge (DFDC) and Face Forensics++. Their CNN model achieves an accuracy of 94.93% for detecting fake videos in the DFDC dataset and 93.2% for the Face Forensics++ Dataset. Comparative evaluations and performance metric analyses indicate promising performance, positioning their model favourably among state-of-the-art methodologies [10].

In the realm of face recognition, the utilization of open-source libraries Dlib and OpenCV has become widespread. S. Suwarno *et al.,* analyse these commonly used libraries, investigating their facial recognition algorithms' accuracy and speed. Their analysis examines CNN and HoG algorithms from Dlib and DNN and HAAR Cascades algorithms from OpenCV. These algorithms are assessed based on speed and accuracy, employing image datasets from LFW (Labelled Faces in the Wild) and AT&T. Actual images from people around UIB (Batam International University) supplement this dataset. The study reveals that HoG algorithm demonstrates the fastest speed (0.011 seconds per image), but with lower accuracy (FRR = 27.27%, FAR = 0%). On the other hand, DNN algorithm boasts the highest accuracy (FRR = 11.69%, FAR = 2.6%), albeit at the cost of slower speed (0.119 seconds per image). This analysis underscores the nuanced trade-offs between algorithmic strengths and weaknesses [11].

Lastly, N. Boyko *et al.,* in 2018 delve into the time complexity of computer vision algorithms for face recognition, focusing on the comparison of two popular libraries: OpenCV and Dlib. Their exploration encompasses feature analysis, pros and cons assessment, and application building using histogram-oriented gradients and deep convolutional neural networks. This comprehensive comparison of productivity in relation to algorithm execution time and iteration number illuminates the libraries' performance. N. Boyko *et al.,* build two face recognition applications based on these libraries, showcasing the comparative performance in practical scenarios [12].

2. Methodology

In this section, we elucidate the methodology employed in the development of our DeepFake detection mobile application. The process encompasses training the model using Google Colaboratory, executing the video prediction process, and crafting a user-friendly Graphical User Interface (GUI) through Flask API. Notably, insights from pertinent research, such as the analysis of Google Colaboratory by Tiago Carneiro *et al.,* [13], and recent advances in colorizing grayscale images using deep learning as proposed by F. Baldassarre *et al.,* [14], alongside the challenges posed by the proliferation of deepfakes and the significance of their detection addressed by A. Verma *et al.,* [15], are integrated into our methodology.

In developing the methodology for this research, an exploration into enhanced neural network models was fundamental, given their capacity for high accuracy in image recognition. The incorporation of Chen and Su's enhanced Hybrid MobileNet was crucial due to its innovative architecture designed to reduce the amount of computational cost and elevate accuracy, optimizing performance on mobile and embedded devices [16]. Also, the methodology drew inspiration from Zhu and Newsam's advancement in DenseNet for dense flow, which utilizes Densely Connected Convolutional Networks to learn optical flow and is more suited for real-time video analysis compared to other CNN architectures, thanks to its shortcut connections providing implicit deep supervision [17].

The expansive and hierarchically structured ImageNet database was pivotal, as it serves as a rich resource containing millions of annotated images, organized by the semantic hierarchy of WordNet, offering unprecedented opportunities for object recognition, image classification, and automatic object clustering [18]. The implementation of the Adam algorithm by Kingma and Ba played a significant role, chosen for its computational efficiency, suitability for large-scale problems, adaptiveness to the geometry of the objective function, and minimal hyper-parameter tuning [19].

Addressing the prevalence of Deepfake videos, the methodology integrated insights from Aduwala *et al.,* exploration on Deepfake detection using GAN discriminators. This approach, based on exploiting GAN's discriminators, exemplified by MesoNet, served as a focal point in detecting manipulations in videos, though it highlighted the challenges posed by videos from unknown sources and emphasized the need for more advanced solutions [20].

In the methodology section, a multifaceted approach was applied, incorporating various scholarly resources. A paramount work, by Afchar *et al.,* [21], developed a technique to promptly detect facial manipulations in videos, notably focusing on Deepfake and Face2Face techniques. This technique stands out for its notable success rates, exceeding 98% in Deepfake detection, utilizing deep learning models emphasizing the mesoscopic properties of images. Another pivotal study by Alheeti *et al.,* [22] introduced a new system, based on transfer learning techniques, that proficiently detects alterations in both audio and images. This approach, utilizing a support vector machine, demonstrated outstanding outcomes in identifying various types of real and manipulated images.

In addition to detection methodologies, databases, and toolkits have played a crucial role in our approach. The "Labelled Faces in the Wild" database [23], containing over 13,000 labelled face photographs, provided a substantial dataset enabling extensive research in unconstrained face recognition. Moreover, Dlib-ml [24], an open-source machine learning toolkit, offered a plethora of tools and algorithms, aiding in the development of machine learning software in C++ language.

A critical evaluation of algorithms was discussed by Avis *et al.,* [25], highlighting the challenges and shortcomings faced by convex hull algorithms in dealing with "degeneracies" and controlling the sizes of intermediate results. They introduced families of polytopes, emphasizing the issues associated with existing algorithms.

Furthermore, the study of Convolutional Neural Networks (ConvNets) scaling was meticulously conducted by Tan and Le [26]. They proposed an innovative scaling method, EfficientNet, that meticulously balanced network depth, width, and resolution, demonstrating enhanced performance and efficacy. The method has surpassed existing models, displaying superior accuracy and efficiency, particularly EfficientNet-B7, which achieved unprecedented 84.3% top-1 accuracy on ImageNet.

Lastly, the prevalent issue of deepfake videos was addressed by Güera and Delp [27], presenting a temporal-aware pipeline to automatically detect such videos. The method, involving a combination of Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), excelled in classifying manipulated videos, portraying the potential in addressing the increasing menace of realistic fake videos in various sectors.

In conclusion, the methodologies, databases, toolkits, algorithm evaluations, and innovations in neural network scaling incorporated from these studies enriched our approach, addressing the extensive dimensions of deepfake and manipulation detection in multimedia content.

2.1 Dataset Collection and Preprocessing

The methodology embarks with a meticulous approach to addressing the issues raised in the introduction, particularly focusing on the uncertain aspects of dataset acquisition and preprocessing. Similar methodologies have been employed in various studies to enhance the reliability of deepfake detection [1,2]. In order to ensure the transparency and reliability of our study, we meticulously collect data from two distinct sources: the Kaggle dataset and a custom dataset. The Kaggle dataset, renowned for its diversity, contributes a substantial corpus of deepfake videos, while our custom dataset encompasses videos curated from various sources to provide a comprehensive and well-rounded collection.

2.1.1 The dataset

The dataset utilized for our research originates from Kaggle's DFDC (DeepFake Detection Challenge) dataset. This collection encompasses a total of 400 videos, specifically allocated for training purposes, alongside an additional 400 videos earmarked for testing. All videos within this dataset are uniformly formatted as mp4 files.

To prepare the data for analysis, a sequence of preprocessing steps was implemented. Firstly, the ComputerVision Library was employed to load the videos, followed by the extraction of frames at a consistent rate of 5 frames per second. Subsequently, employing the DataLibrary, we cropped the images, focusing on isolating the facial regions of interest. To enhance the images for improved feature extraction, the ImageEnhance library was also incorporated.

Upon the completion of these preprocessing steps, a classification was established, partitioning the videos into two distinct folders: "real" and "fake." This classification was based on the contents of the individual frames within each video. A selection of frames was identified as originating from authentic videos, while others were identified as components of manipulated (fake) videos.

Our research data consists of 400 training videos and 400 testing videos, sourced from Kaggle's DFDC dataset. This data was meticulously prepared through preprocessing stages that encompassed frame extraction, facial region cropping, and image enhancement to ensure optimal analysis conditions.

2.1.2 The preprocessing

The preprocessing processes in the methodology involve multiple steps aimed at preparing the video data for deepfake detection. These steps encompass data acquisition, division into frames, facial identification, cropping, and the creation of trimmed frames for subsequent analysis. The specifics are as follows:

2.1.3 Data acquisition and division into frames

The video data used in the experiments originates from two primary sources: the Kaggle dataset and a custom dataset. Each video is subjected to an initial segmentation process where it is divided into frames, which represent individual images extracted from the video.

2.1.4 Facial identification

Within each frame, an important preprocessing step involves identifying faces. This is achieved through advanced computer vision techniques, which might include utilizing face detection algorithms like Haar cascades or deep learning-based approaches like Convolutional Neural Networks (CNNs). These techniques detect and delineate facial features in each frame.

2.1.5 Cropping and resizing

Once faces are identified in the frames, they are meticulously cropped. The cropping process involves isolating the detected face from the rest of the frame. This isolation aids in feature extraction by focusing only on the area relevant to facial characteristics. Cropping also helps in standardizing the input size for subsequent processing. Cropped faces are then resized to a consistent dimension, ensuring uniformity for further analysis.

2.1.6 Trimming for model input

The original frames, each containing a cropped and resized face, are further processed to create trimmed frames. In the context of deepfake detection, a standardized length for these trimmed frames is essential to ensure consistent analysis. For example, the methodology stipulates that at least a 10-second video is selected for prediction. Therefore, frames extracted from this video would be sequentially divided into segments, and each segment constitutes a trimmed frame. The specifics of the length and number of trimmed frames generated from each video segment should be clearly outlined in your methodology.

The preprocessing processes entail acquiring video data from different sources, segmenting them into frames, identifying faces within these frames, cropping and resizing the detected faces, and ultimately generating trimmed frames of consistent length for input into the detection model. The exact length of the original frames, as well as the trimmed frames, should be explicitly stated in your methodology, and it's advisable to provide a rationale for these choices based on the characteristics of the problem and the capabilities of your chosen model.

2.2 Training Model using Google Colaboratory

The evolution of our approach towards deepfake detection is aptly illustrated in Figure 1, serving as a roadmap for our research endeavours. To kickstart the process, we resort to Google Colaboratory, a powerful cloud-based platform that seamlessly integrates Jupyter Notebooks. This platform, colloquially known as Colab, is a hub for machine learning education and research, offering a runtime optimized for deep learning tasks, including access to a GPU for enhanced performance.

Our model selection centres around the InceptionResNetV2 architecture, a well-established choice renowned for its robust image analysis capabilities. Each video frame is subjected to meticulous individual cropping as part of our preprocessing strategy, aimed at optimal feature extraction and manipulation mitigation. The dataset, which encompasses both authentic and manipulated videos, undergoes a prudent division for the distinct purpose of model training. Subsequently, the meticulously trained model is securely saved, poised to play an integral role in subsequent prediction tasks.

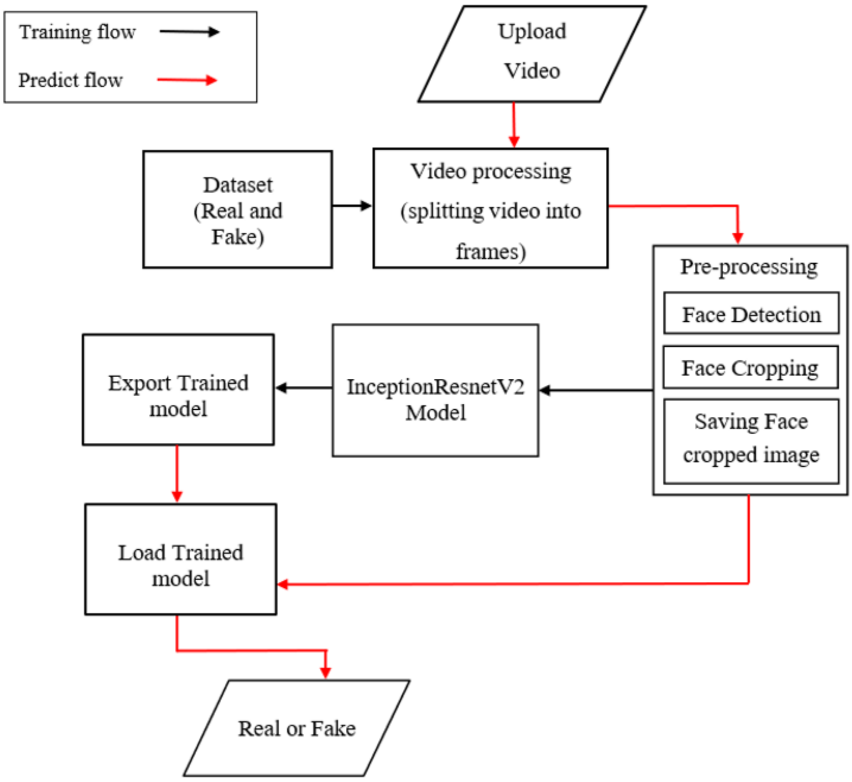


Fig. 1. System architecture flowchart

2.3 Video Prediction Process

Our methodology is meticulously engineered to deliver consistent and reproducible results. Figure 2 artistically illustrates the step-by-step journey of predicting the authenticity of videos. A pivotal prerequisite demands the selection of a video, ideally lasting at least 10 seconds, encoded in the universally compatible MP4 format. This video sourcing spans diverse platforms, including YouTube, and accommodates other datasets primed for prediction endeavours.

The intricacies of our approach unfurl with the chosen video's frame-by-frame dissection, each frame meticulously scrutinized for facial identification. Once faces are detected within these frames, they are subjected to precision cropping, culminating in the creation of trimmed frames. This sequence of actions culminates in these trimmed frames seamlessly traversing through the rigorously trained model hailing from Google Colaboratory. The outcome of this model's prediction process, signifying either genuine or manipulated content, ultimately unravels the authenticity status of the selected video.

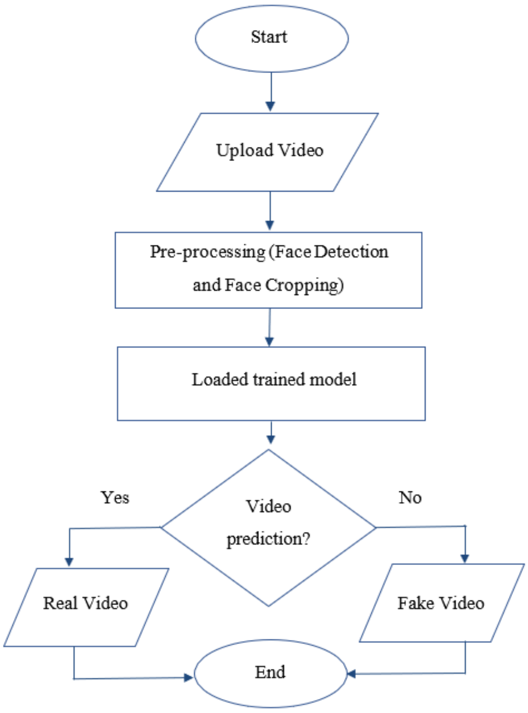


Fig. 2. Prediction flowchart

2.4 Development of Graphical User Interface (GUI) using Flask API

Our methodology is underpinned by a commitment to user-centric design and interaction. Drawing inspiration from the user acceptance evaluation by A. Rahman *et al.,* [10] , we prioritize the creation of an intuitive and engaging graphical user interface (GUI). The Flask API block diagram is shown in Figure 3. Recognizing the significance of user engagement, we turn to MIT App Inventor for the application's construction, ensuring an interface that aligns seamlessly with user expectations.

At the core of our GUI development strategy lies Flask API, a framework that expertly orchestrates interaction between diverse system components. Our approach mimics the blueprint set by Flask API development, which leverages the Flask web framework. This results in the creation of a versatile interface capable of accommodating a plethora of data formats, including JSON, XML, and HTML. A standout feature of this interface is its ability to cater to varied requests, including data retrieval, updates, and specific actions.

Our GUI is elegantly designed to facilitate the straightforward upload of videos from local directories, endowing users with the capability to seamlessly integrate their content. This GUI is synergistically intertwined with the model's prediction services, creating a robust ecosystem that empowers users to wield the predictive prowess of the model trained through diligent efforts on Google Colaboratory.

By methodically enhancing our methodology, we ensure not only clarity in the process but also a consistent and coherent approach that resonates through every phase of our research. Incorporating insights from Tiago Carneiro *et al.,* [13], F. Baldassarre *et al.,* [14] and A. Verma *et al.,* [15], our methodology integrates advanced training configurations, cutting-edge model architectures, and user-centric design principles to develop a robust and intuitive deepfake detection mobile application [20].

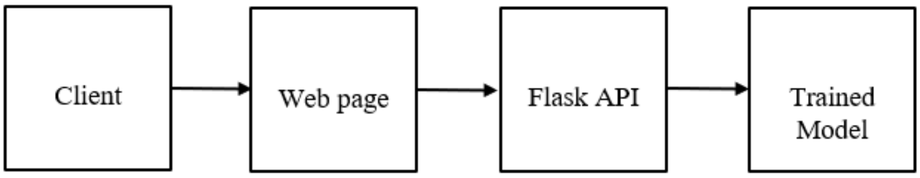


Fig. 3. Flask API block diagram

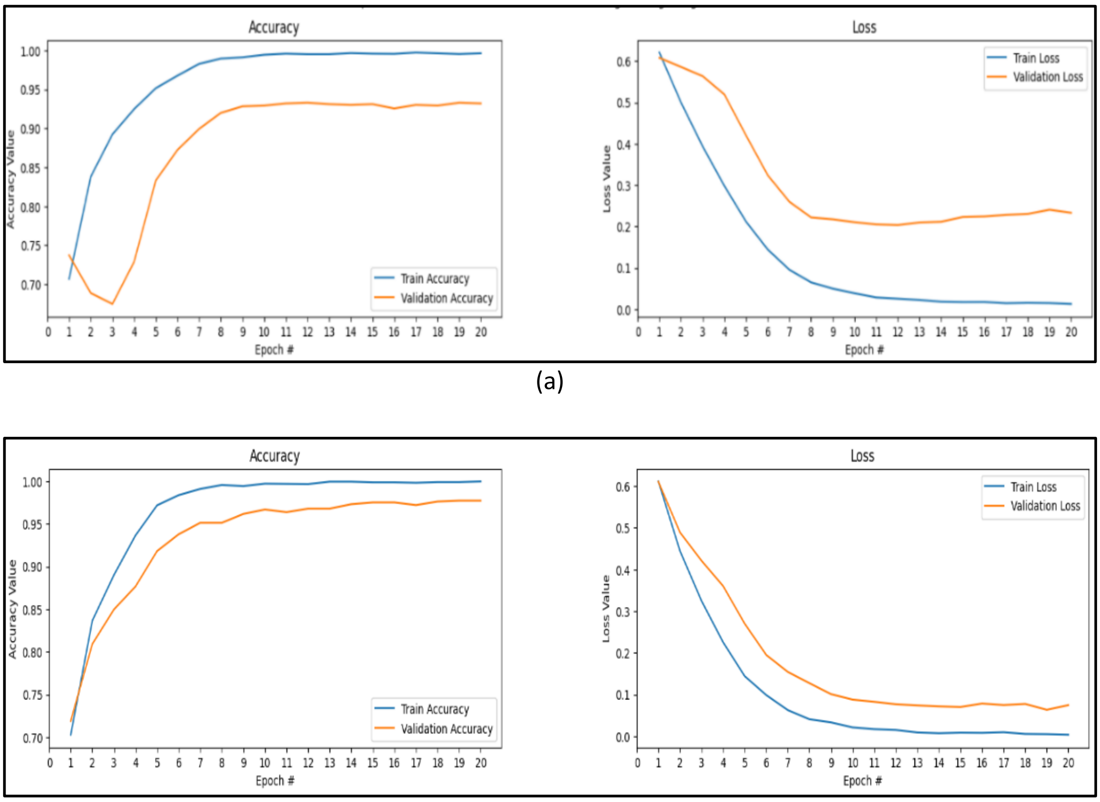
3. Results

3.1 Model Accuracy and Discussion

In this particular segment, we delve into the outcomes derived from our utilization of the InceptionResNetV2 model, which was applied to assess two distinct datasets: DFDC and CelebDF. Each of these datasets, encompassing a total of 400 videos, serves as the foundation for our analysis. To visually present our findings, Figure 4 illustrates a graphical representation that offers a comprehensive comparison of the dataset accuracies between DFDC and CelebDF.

When it comes to accuracy metrics, it's worth noting that the results were indeed noteworthy. More specifically, in terms of the DFDC dataset:

i. demonstrated an accuracy of 93.20%, signifying a commendable level of precision in classifying videos. However, the CelebDF dataset represented by
ii. achieved an even higher accuracy of 97.70%, underscoring its robustness in distinguishing between genuine and manipulated content [1,2]. These accuracy values manifest the effectiveness of the InceptionResNetV2 model in discerning and categorizing videos from these respective datasets.



(a)

(b)

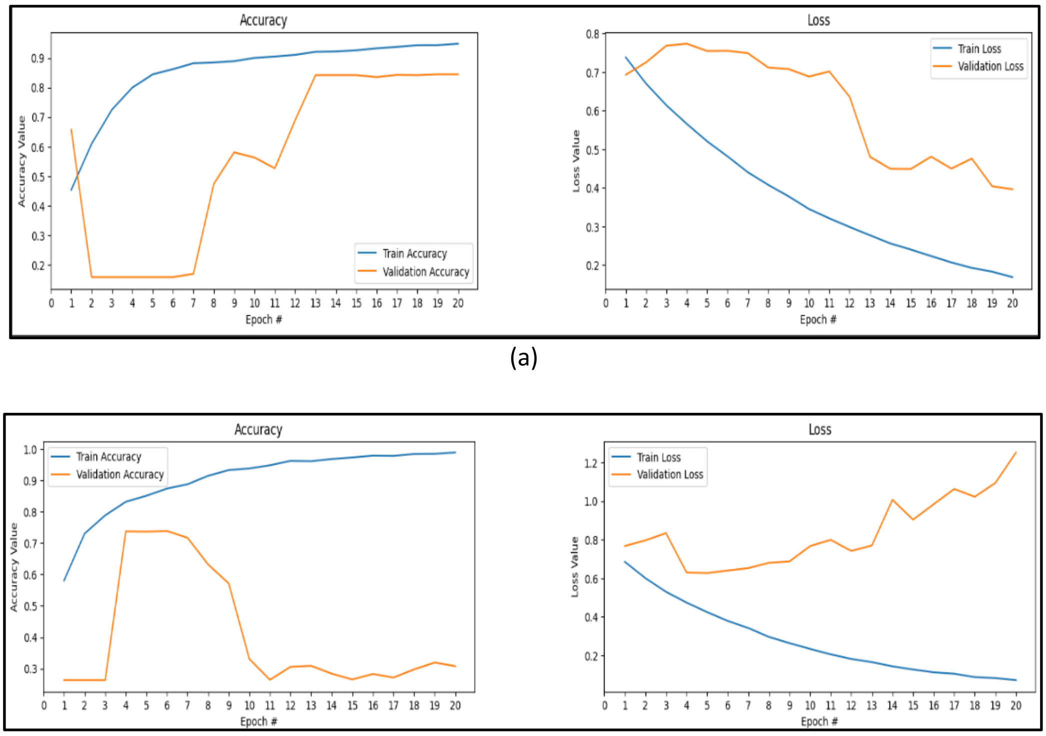
Fig. 4. Dataset accuracy comparison for (a) DFDC and (b) CelebDF datasets

Subsequently, our exploration extended to the utilization of the EfficientNetB0 model, which was put through its paces using the identical datasets, namely DFDC and CelebDF. A visual representation of the dataset accuracy comparison is thoughtfully presented in Figure 5, allowing for a comprehensive understanding of our findings.

Upon careful analysis of the DFDC dataset, the EfficientNetB0 model denoted as:

i. showcased an accuracy level of 84.44%. This outcome sheds light on the model's capability to discern patterns and attributes within the video content, leading to an accurate classification process. However, when directing our attention to the CelebDF dataset
ii. yielded a comparatively lower accuracy score of 30.81%. This distinctive difference in accuracy values highlights the model's varying efficacy in distinguishing genuine videos from manipulated ones across different datasets.

This variance in accuracy values serves as a reminder of the nuanced nature of dataset characteristics and their impact on model performance. The disparity between the accuracy outcomes on the two datasets underscores the importance of dataset composition and the complexities that can influence the effectiveness of a model like EfficientNetB0 in different contexts [1,2].



(a)

(b)

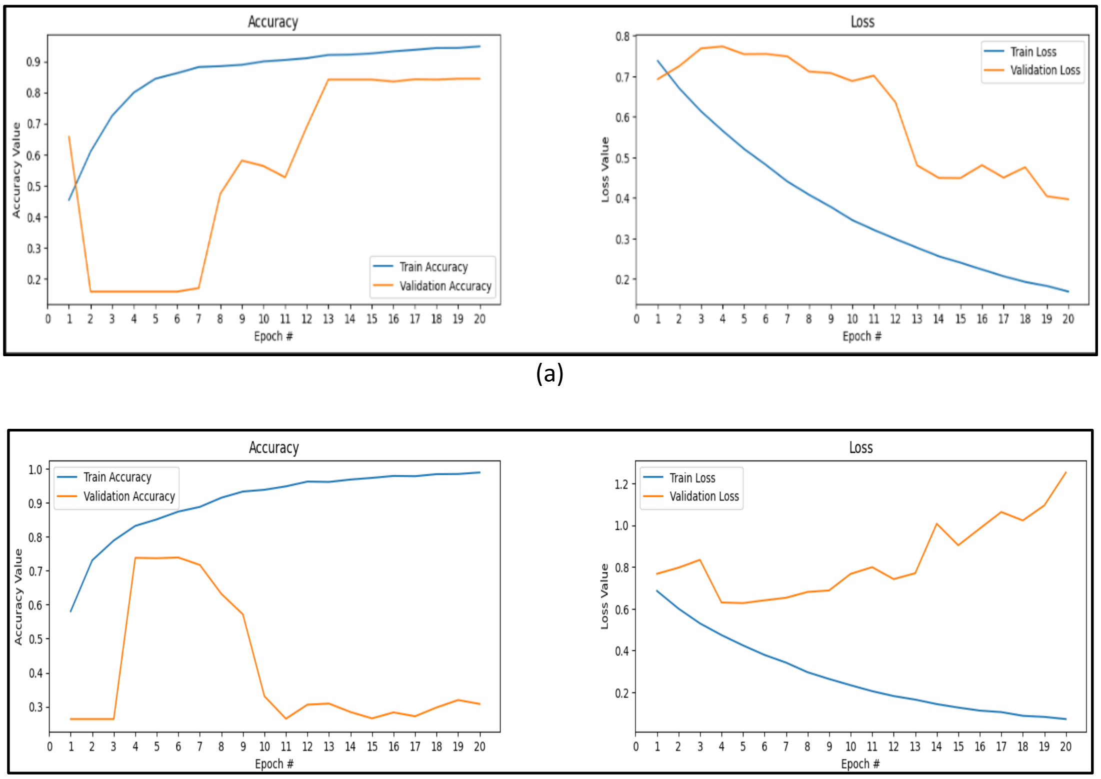
Fig. 5. Dataset accuracy comparison for (a) DFDC and (b) CelebDF datasets

In the final leg of our investigation, we turned our attention to the ResNet50 model, which underwent rigorous assessment using the DFDC and CelebDF datasets. To provide an intuitive visual representation of our results, Figure 6 has been thoughtfully devised, rendering a comprehensive snapshot of the dataset accuracy dynamics.

Taking a closer look at the DFDC dataset, our ResNet50 model, marked as:

i. demonstrated an accuracy score of 52.28%. This outcome intricately reflects the model's proficiency in navigating through the intricacies of the dataset's content, revealing its aptitude in distinguishing between real and manipulated videos. Conversely, the CelebDF dataset evoked a more favourable response from the ResNet50 model
ii. which achieved a notably higher accuracy level of 72.93%. This noteworthy discrepancy in accuracy rates emphasizes the model's adaptability to distinct dataset characteristics, particularly in the realm of celebrity-based videos.

These variations in accuracy outcomes serve as a reminder of the multifaceted nature of dataset properties and their profound influence on model performance. The contrasting accuracy results underscore how the ResNet50 model's discernment between genuine and manipulated videos can be modulated by the unique attributes inherent to each dataset. This phenomenon reiterates the significance of meticulous dataset curation and its consequential implications on the efficacy of the ResNet50 model across different scenarios.



(a)

(b)

Fig. 6. Dataset accuracy comparison for (a) DFDC and (b) CelebDF datasets

Table 1 delivers a comprehensive depiction of the training accuracy achieved by the three distinct models across the two datasets. Among these models, it's particularly noteworthy that the InceptionResNetV2 model stood out as the frontrunner, showcasing an impressive accuracy rate of 93.20% when tested against the DFDC dataset. Furthermore, this model exhibited remarkable prowess when exposed to the CelebDF dataset, achieving an even more substantial accuracy level of 97.72%.

On the contrary, a contrasting performance trend was observed in the case of the ResNet model, which exhibited the least accuracy of 52.28% when confronting the challenges posed by the DFDC dataset. Similarly, the EfficientNetB0 model faced a similar challenge in the CelebDF dataset, resulting in a comparatively lower accuracy of 30.81%. This intriguing spectrum of accuracy outcomes further accentuates the varying strengths and capabilities of these models when subjected to distinct datasets, reaffirming the crucial role that dataset nuances play in influencing model behaviour and performance.

Table 1
Training accuracy of the three models with two datasets

Model	Dataset	Accuracy
InceptionResNetV2	DFDC	93.20%
	CelebDF	97.72%
EfficientNetB0	DFDC	84.44%
	CelebDF	30.81%
ResNet50	DFDC	52.28%
	CelebDF	72.93%

Presented in Table 2 is a comprehensive representation of the testing accuracy observed across the three distinctive models. Notably, among this ensemble of models, the InceptionResNetV2 model emerged as the most accurate, achieving a commendable accuracy rate of 71.19% when assessed against the DFDC dataset. Contrasting this, the ResNet model grappled with lower accuracy, registering a modest 37.29% accuracy mark when evaluated against the very same DFDC dataset. On a brighter note, the EfficientNetB0 model showcased a more promising performance in the same DFDC dataset, achieving an accuracy level of 62.71%. This juxtaposition of accuracy outcomes offers valuable insights into the comparative effectiveness of these models when confronted with the complexities of real-world testing scenarios, underscoring the nuanced interplay between model architectures and dataset characteristics.

Table 2
Testing accuracy of the three models

Model	Dataset	Accuracy
InceptionResNetV2	DFDC	71.19%
EfficientNetB0	DFDC	37.29%
ResNet50	DFDC	62.71%

3.2 GUI for Android Application: Results and Discussion

Illustrated in Figure 7 is the user-friendly graphical user interface (GUI) that underpins the Android application. Designed as the homepage of the application, this interface offers users an intuitive platform to select videos for subsequent detection and analysis. A notable feature of this application is its seamless integration with Flask Ngrok, a powerful combination that guarantees uninterrupted and efficient operation, enhancing the overall user experience. This GUI gateway marks the starting point of the application's engagement with users, encapsulating the essence of its functionality in a visually accessible and interactive manner.

Figure 8 provides a visual representation of the graphical user interface (GUI) dedicated to video uploading from mobile devices. For a video to undergo the detection process, certain prerequisites must be met: it should have a minimum duration of 10 seconds, adhere to the mp4 format, and exhibit a resolution within the range of 240p to 720p. This intuitive interface extends the application's capabilities to facilitate video submission, thereby initiating the intricate process of analysis and identification. By adhering to these stipulated criteria, the application ensures the optimal compatibility of the uploaded videos with its subsequent processing stages, ultimately contributing to accurate and reliable outcomes.

Figure 9 illustrates the interface designed for video prediction and the subsequent display of results. When a video is selected by the user, the system embarks on a comprehensive analysis to ascertain its authenticity, effectively distinguishing between fake and real content. This intricate process involves intricate algorithms that meticulously scrutinize various aspects of the video. Following this in-depth examination, the application promptly delivers its verdict, indicating whether the video is genuine or manipulated. This user-friendly interface not only empowers users with insightful information but also exemplifies the application's core purpose of combating deceptive content, thereby fostering a more informed and vigilant online environment.

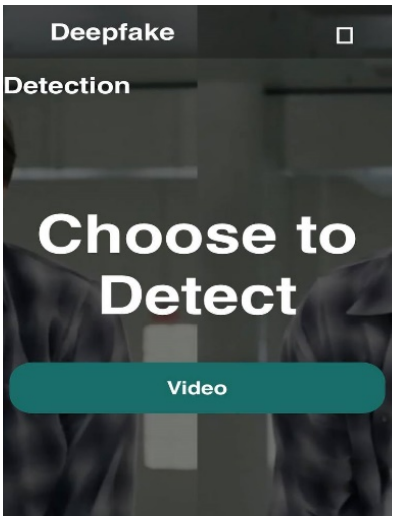


Fig. 7. Homepage interface

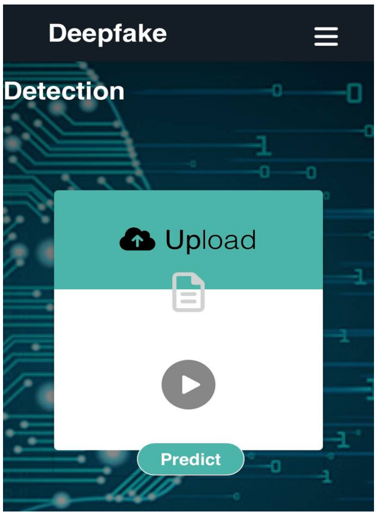
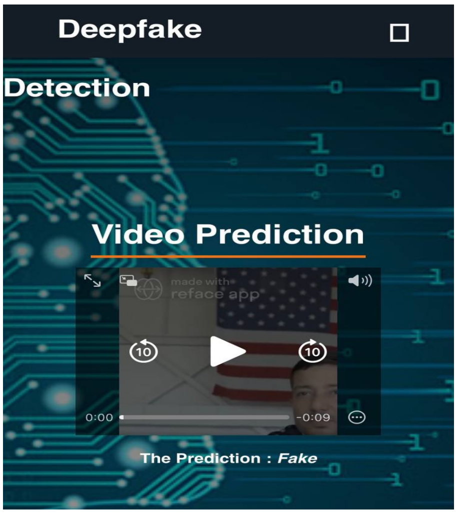


Fig. 8. Upload File Interface



The Prediction : *Fake*

Fig. 9. Prediction and Result Display Interface

4. Conclusions

In closing, the Advances in DeepFake Detection: Leveraging InceptionResNetV2 for Reliable Video Authentication embodies a groundbreaking and ingenious solution that squarely confronts the complexities introduced by the pervasive deepfake technology within today's digital sphere. By harnessing cutting-edge technology, this application adeptly identifies and scrutinizes videos, with a pronounced emphasis on ascertaining the authenticity of the faces portrayed within them. Its primary mission is to counteract the widespread proliferation of deceptive deepfake videos, effectively shielding users from potential deception and the insidious propagation of misinformation that plagues various social media platforms.

A central hallmark of the Deepfake Detection Video Application lies in its user-centric interface, which guarantees a seamless and intuitive user experience. This intuitive platform empowers users to effortlessly upload videos from a multitude of sources, facilitating comprehensive detection and meticulous analysis. Through its streamlined accessibility, the application emboldens users to take an active role in distinguishing between genuine and fabricated videos, thereby fostering a more enlightened and resilient online community.

The significance of the Deepfake Detection Video Application transcends mere utility; it encapsulates a significant stride in ongoing endeavours to mitigate the perilous repercussions of deepfake technology. By providing a dependable mechanism for discerning and scrutinizing video authenticity, the application contributes to the cultivation of trust and credibility in the realm of digital content. This imparts enhanced integrity to online information, fortifying the very foundations of a dependable and trustworthy digital landscape.

Moreover, the Deepfake Detection Video Application stands as a collaborative achievement that seamlessly marries technological prowess with societal imperatives. It stands as a testament to the collective commitment wielded against the negative implications stemming from deepfake technology. The application's creation underscores the unwavering dedication of researchers, developers, and the larger society in upholding the integrity of visual media and championing the ideals of veracity and precision.

In light of these resolute conclusions, it becomes patently evident that the Deepfake Detection Video Application holds immense potential in the relentless struggle against deepfakes. Through its adept utilization, users are equipped to navigate the digital realm with heightened vigilance and acumen, thus fostering an environment that is both secure and trustworthy. As technology marches onward and new challenges emerge, the application's adaptability and steadfast commitment to safeguarding users will remain of paramount importance.

In summation, the Deepfake Detection Video Application represents a noteworthy stride forward in the ongoing quest to alleviate the risks engendered by deepfake technology. Armed with its user-friendly interface, exhaustive video analysis capabilities, and its role in instilling faith in digital content, this application stands as a pivotal force in nurturing a well-informed and resilient online community. Through sustained research, development, and robust collaboration, we have the power to further elevate the efficacy of deepfake detection tools, ensuring the integrity of visual media endures amidst the evolving technological landscape.

Acknowledgement

This research is supported by Universiti Tun Hussein Onn Malaysia (UTHM) through Tier 1(Q373).

References

[1] Krejcar, Ondrej, Pavel Kukuliac, Lim Kok Cheng, Ali Selamat, and Jiri Horak. "Deep-Learning Pre-Processing for Improvement Of K-Means Cluster Analysis of Seniors' Walkability in Hradec Kralove And Ostrava (Two Middle-Sized Czech Cities)." *Journal of Advanced Research in Computing and Applications* 28, no. 1 (2022): 1-11.

[2] Khaw, Li Wen, and Shahrum Shah Abdullah. "Mri Brain Image Classification Using Convolutional Neural Networks and Transfer Learning." *Journal of Advanced Research in Computing and Applications* 31, no. 1 (2023): 20-26. https://doi.org/10.37934/arca.31.1.2026

[3] Johnson, Dave. "What is a deepfake? Everything you need to know about the AI-powered fake media." *Business Insider* (2021).

[4] C. Vaccari and A. Chadwick. "Analysis | 'deepfakes' are here. these deceptive videos Erode Trust in all news media.," *The Washington Post*, (2020). https://www.washingtonpost.com/politics/2020/05/28/deepfakes-are-here-these-deceptive-videos-erode-trust-all-news-media

[5] Brownlee, Jason. "A gentle introduction to generative adversarial networks (GANs)." *Machine learning mastery* 17 (2019).

[6] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248-255. Ieee, 2009. https://doi.org/10.1109/CVPR.2009.5206848

[7] Bonettini, Nicolo, Edoardo Daniele Cannas, Sara Mandelli, Luca Bondi, Paolo Bestagini, and Stefano Tubaro. "Video face manipulation detection through ensemble of cnns." In *2020 25th international conference on pattern recognition (ICPR)*, pp. 5012-5019. IEEE, 2021. https://doi.org/10.1109/ICPR48806.2021.9412711

[8] Agarwal, Harsh, Ankur Singh, and D. Rajeswari. "Deepfake detection using svm." In *2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC)*, pp. 1245-1249. IEEE, 2021. https://doi.org/10.1109/ICESC51422.2021.9532627

[9] Soni, Hritik, Pranjal Arora, and D. Rajeswari. "Malicious application detection in android using machine learning." In *2020 International Conference on Communication and Signal Processing (ICCSP)*, pp. 0846-0848. IEEE, 2020. https://doi.org/10.1109/ICCSP48568.2020.9182170

[10] Rahman, Ashifur, Nipo Siddique, Mohasina Jannat Moon, Tahera Tasnim, Mazharul Islam, Md Shahiduzzaman, and Samsuddin Ahmed. "Short and low resolution deepfake video detection using cnn." In *2022 IEEE 10th Region 10 Humanitarian Technology Conference (R10-HTC)*, pp. 259-264. IEEE, 2022. https://doi.org/10.1109/R10-HTC54060.2022.9929719

[11] Suwarno, Suwarno, and Kevin Kevin. "Analysis of face recognition algorithm: Dlib and opencv." *Journal of Informatics and Telecommunication Engineering* 4, no. 1 (2020): 173-184. https://doi.org/10.31289/jite.v4i1.3865

[12] Boyko, Nataliya, Oleg Basystiuk, and Nataliya Shakhovska. "Performance evaluation and comparison of software for face recognition, based on dlib and opencv library." In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pp. 478-482. IEEE, 2018. https://doi.org/10.1109/DSMP.2018.8478556

[13] Carneiro, Tiago, Raul Victor Medeiros Da Nóbrega, Thiago Nepomuceno, Gui-Bin Bian, Victor Hugo C. De Albuquerque, and Pedro Pedrosa Reboucas Filho. "Performance analysis of google colaboratory as a tool for accelerating deep learning applications." *Ieee Access* 6 (2018): 61677-61685. https://doi.org/10.1109/ACCESS.2018.2874767

[14] Baldassarre, Federico, Diego González Morín, and Lucas Rodés-Guirao. "Deep koalarization: Image colorization using cnns and inception-resnet-v2." *arXiv preprint arXiv:1712.03400* (2017).

[15] Verma, Akshay, Dipesh Gupta, and Manish Kumar Srivastava. "Deepfake Detection using Inception-ResnetV2." In *2021 First International Conference on Advances in Computing and Future Communication Technologies (ICACFCT)*, pp. 39-41. IEEE, 2021. https://doi.org/10.1109/ICACFCT53978.2021.9837351

[16] Chen, Hong-Yen, and Chung-Yen Su. "An enhanced hybrid MobileNet." In *2018 9th International Conference on Awareness Science and Technology (iCAST)*, pp. 308-312. IEEE, 2018. https://doi.org/10.1109/ICAwST.2018.8517177

[17] Zhu, Yi, and Shawn Newsam. "Densenet for dense flow." In *2017 IEEE international conference on image processing (ICIP)*, pp. 790-794. IEEE, 2017. https://doi.org/10.1109/ICIP.2017.8296389

[18] Deng, Jia, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. "Imagenet: A large-scale hierarchical image database." In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248-255. Ieee, 2009. https://doi.org/10.1109/CVPR.2009.5206848

[19] Kingma, Diederik P. "Adam: A method for stochastic optimization." *arXiv preprint arXiv:1412.6980* (2014).

[20] Aduwala, Sai Ashrith, Manish Arigala, Shivan Desai, Heng Jerry Quan, and Magdalini Eirinaki. "Deepfake detection using GAN discriminators." In *2021 IEEE Seventh International Conference on Big Data Computing Service and Applications (BigDataService)*, pp. 69-77. IEEE, 2021. https://doi.org/10.1109/BigDataService52369.2021.00014

[21] Afchar, Darius, Vincent Nozick, Junichi Yamagishi, and Isao Echizen. "Mesonet: a compact facial video forgery detection network." In *2018 IEEE international workshop on information forensics and security (WIFS)*, pp. 1-7. IEEE, 2018. https://doi.org/10.1109/WIFS.2018.8630761

[22] Alheeti, Khattab M. Ali, Salah Sleibi Al-Rawi, Haitham Abbas Khalaf, and Duaa Al Dosary. "Image feature detectors for deepfake image detection using transfer learning." In *2021 14th International Conference on Developments in eSystems Engineering (DeSE)*, pp. 499-502. IEEE, 2021. https://doi.org/10.1109/DeSE54285.2021.9719332

[23] Labeled Faces in the Wild Home. http://vis- www.cs.umass.edu/lfw/#download

[24] King, Davis E. "Dlib-ml: A machine learning toolkit." *The Journal of Machine Learning Research* 10 (2009): 1755-1758.

[25] Avis, David, and David Bremner. "How good are convex hull algorithms?." In *Proceedings of the eleventh annual symposium on Computational geometry*, pp. 20-28. 1995. https://doi.org/10.1145/220279.220282

[26] Tan, Mingxing. "Efficientnet: Rethinking model scaling for convolutional neural networks." *arXiv preprint arXiv:1905.11946* (2019).

[27] Güera, David, and Edward J. Delp. "Deepfake video detection using recurrent neural networks." In *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*, pp. 1-6. IEEE, 2018. https://doi.org/10.1109/AVSS.2018.8639163

Name of Author	Email
Nan Mad Sahar	nan@uthm.edu.my
Muhammad Faris Syazwan	de190080@student.uthm.edu.my
Nor Surayahani Suriani	nsuraya@uthm.edu.my
Suhaila Sari	suhailas@uthm.edu.my
Shuhaida Binti Ismail	shuhaida@uthm.edu.my
Azrul Amri Jamal	azrulamri@unisza.edu.my
Ahmed Marwan Aleesa	csit_college_kirkuk@uokirkuk.edu.iq