AN INTELLIGENT SYSTEM TO IDENTIFY FAKE VIDEOS ON ONLINE SOCIAL NETWORKS USING MACHINE LEARNING

OLUWAFOLAKE ESTHER OJO[1*], OLUWATOBI ADEDAMOLA AYILARA-ADEWALE[1], YUSUF OWOLABI OLATUNDE[2], ZAINAB OYINLOLA OTUN[3]

[1]Department of Information Technology, Osun State University, Osogbo, Nigeria
[2]Department of Cyber Security, Osun State University, Osogbo, Nigeria
[3]Department of Computer Science, Federal University of Agriculture, Abeokuta, Nigeria

**Corresponding author: oluwafolake.ojo@uniosun.edu.ng*

ABSTRACT: Advances in emerging technologies have led to the wide dissemination of fake videos on online social networks. This research employs the hybrid machine learning and deep learning algorithms to recognise video forgery on social media. This research presents a hybrid deep learning and machine learning approach for detecting manipulated videos. The proposed model employs the OpenCV deep neural network (DNN) face detector to locate facial regions in video frames, after which a pretrained ResNet50 convolutional neural network is applied to extract deep features, and a Support Vector Machine (SVM) is used to classify authentic and fake content in a binary fashion. The model was trained and evaluated on a dataset of publicly available 128 short-form videos gathered from TikTok and Facebook platforms. The five-fold stratified cross-validation results generated an average accuracy of 89.1%, precision of 87.6%, recall of 96.1%, F1-score of 91.5%, and an AUC of 0.95 for the SVM model. Furthermore, the comparative analyses showed that SVM outperformed Logistic Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbour classifiers. The findings demonstrate that combining automated face detection with deep feature extraction and classical machine learning significantly improves fake-video detection and contributes to preserving authenticity in digital communication.

KEY WORDS: *Network security, Social Networks, Machine learning, Fake Videos, ResNet50*

1. INTRODUCTION

The ability of a machine to perform cognitive tasks like perception, learning, reasoning, and problem-solving that are completed by humans is referred to as artificial intelligence. AI speech, vision, and thinking teams need to function at least as effectively as human ones (Markauskaite et al., 2024). Currently, artificial intelligence (AI) is widely employed across various industries, offering significant technological benefits to organizations that extensively utilize it. The potential of AI is to increase value of the banking sector by 50% (Christensen, 2021) and the retail sector by $600 billion (Donepud, 2019) when compared to conventional analytic techniques. Furthermore, research revealed that higher percentage rise in potential income in logistics and transportation due to the adoption of AI (Kuberkar & Singhal, 2020; Yaiprasert & Hidayanto, 2024).

AI systems employ machine learning (ML) that can acquire knowledge autonomously and continually improve. The practice of ML includes designing programs that can access data and learn autonomously. To begin the learning process, one must gather observations or data, such as examples, personal experiences, or instruction, to identify patterns and improve decision-making in the future. The main aim is for computers to gain knowledge independently and adjust their actions as necessary, without depending on humans (Sarker, 2021). Deep learning in machine learning focuses on constructing models for intricate datasets that involve creating relationships using multiple-level representational learning approaches. Higher-level concept and functionalities are expressed concerning those at a lesser level in a deep architecture. Most of these models are based on unsupervised learnt representations (Ahmed et al., 2023). The phrase “deep-fake” originates from the concepts “deep learning” and “fake”. Deep neural networks have made it easier and faster to fabricate realistic-looking fake photos and videos. It is a method whereby someone’s video or image is modified by another person’s image by using deep learning (Mitra et al., 2021). Deep neural network-based encoders, which are commonly employed in picture compression, impose a compressed version of the original input by creating a bottleneck in the network. More advanced encoders have made it possible to compress high-quality images, which reduces the computing power required for deep-fake jobs. To produce deep fakes, two auto-encoders are trained. The deep-fake image is created by reconstructing the target image using the decoder of the source image. This yields an image of the target that contains elements of the source image. The characteristics of the source picture are learned by the first auto-encoder, the attributes of the target image are learned by the second encoder, and the settings of the two encoders are shared (Khalil & Maged, 2021). Nowadays, social networking networks are a multipurpose tool utilised for telemarketing, business, education and regrettably, illicit activity. Users usually use social media to communicate with colleagues and peers who share their interests. It also acts as a channel for client communication, and the information it gathers could help identify new trends in business insights. Social media is a major aspect of contemporary living. Social media users post news, videos, and thoughts about a wide range of activities on Facebook, WhatsApp, Instagram, Twitter, and other online networks. Even though many individuals enjoy using social media, many dishonest activities could lead users to believe misleading information, such as rumours or fake news (Islam et al, 2020).

The methods used in creating deepfake videos are either the target’s real video is used as the source, in which case the target is manipulated to say and do things they have never done before, or the target’s face is replaced with a video of someone else. This has prompted the creation of fake videos which are almost real and hard to tell the difference between them and the original ones (Botha & Pieterse, 2020). They have become paramount due to their ability to spread fake news, sway opinions, instigate social and political unrest. Hence, this research presents a robust and intelligent model suitable for detecting fake video clips on different social media platforms. The rise in the proliferation of doctored videos via the internet has augmented the necessity to have effective detection systems that can maintain the authenticity of visual data. Although numerous deep learning approaches have been created to identify deepfakes, most of them do not cope with variability in video quality, lighting, and motion in the real world, and their end-to-end training is

very expensive. These problems may have the propensity to limit the scale of their operations and application to social media data, which is brief, compressed, and exceedingly diverse.

This paper introduces a hybrid detection system to overcome these limitations, which is a fusion of deep feature and classical machine learning classification. The model unites representational power of a trained ResNet50 model and efficiency of a Support Vector Machine (SVM) classifier. This balanced structure gives a boost to the detecting power incurring less computation price which is more adaptive to the dynamism of the online video materials. The emphasis on the short format and user-created videos of TikTok and Facebook enables the research to bring a realistic and varied dataset, which represents the actual complexity of social media settings. Therefore, this approach is capable of achieving a lightweight and yet very efficient model to locate manipulated videos and protect digital authenticity. The purpose of this work is to optimise the chances of detecting fake videos and contribute to preventing the spread of false information, protecting individuals' rights to privacy and maintaining the credibility of visuals disseminated on social media. The remaining part of the paper is structured as follows: in Section 2, the literature review is provided. Section 3 provides methods and approaches, and Section 4 provides findings and analysis. Section 5 is the conclusion and potential future research directions.

2. RELATED WORKS

The growing availability of artificial intelligence products has facilitated the ability of even non-experts to generate and share fake content, such as manipulated videos, content and fabricated news. As social media becomes more essential for accessing important information, the danger of deepfakes sharing misleading information is increasing (Jing & Murugesan, 2021). Authentic face-swapping effects are achieved in media using AI software like FaceApp, AgingBooth, and others. Through swapping, it is possible to change facial structure, hairstyle, sex, age and other factors (Yavuzkiliç et al., 2021). The mass distribution of such distorted media brings up severe issues touching on misinformation, privacy, and online trust. Consequently, detecting falsified video content has become an urgent

research priority given society's growing reliance on social media for communication and information exchange. Even though some studies have examined deepfake detection methods, few studies have examined deepfake detection specifically in the context of social-media-based video datasets, which are characterised by compression artefacts, varied lighting, and other user-induced variations that present special detection problems. Recent research on fake media detection varies across learning paradigms and data types. This section reviews key developments within these categories and situates the contribution of the current study. Although numerous studies have proposed deepfake detection systems, relatively few have targeted social-media-based videos, where compression artifacts, spontaneous lighting, and short duration complicate analysis. The manufacturing of deepfakes frequently fails to maintain temporal consistency between frames, and several works have leveraged this limitation to design detection models (Sabir et al., 2019).

2.1. Machine Learning Approaches

Machine learning (ML) techniques have been widely applicable and used to identify deepfakes by analysing frame-level data and classifying it based on labelled data (Altaei, 2023; Hamza et al, 2022). These techniques typically involve two processes, that is, discrimination feature extraction and classification. Features extracted can be hand-crafted, like edge descriptors, motion vectors, or texture patterns, or high-level features, which are the output of tested deep networks. Support Vector Machines (SVM), Logistic Regression (LR), Decision Trees, Random Forest (RF), Gradient Boosting, and K-Nearest Neighbours (KNN) are some of the classifiers that have been used to differentiate between authentic and cheated content (Alemerien and Al-Mahadin, 2025; Suryani et al, 2025). As an example, (Kharbat et al, 2019) used SVM regression to detect videos containing fake content. The conventional edge feature detectors were used to train the model to find counterfeit videos. Moreover, other scholars had conducted an extensive survey of various deepfake recognition systems based on ML and found that there are still primary issues, such as the imbalanced distribution of classes, overfitting, and low generalisation across datasets (Rana et al. 2022). Moreover, another researcher did a study to investigate the use of ensemble models (RF and Gradient Boosting) to detect fake content; the experimental findings showed that the ensemble models offer higher detection accuracy and recall compared to single-model models (Ali et al., 2024).

2.2. Deep Learning Approaches

The deep learning (DL) methods performed a bit differently from ML; DL architectures learn spatial and temporal features directly from video data without the need for manual feature engineering. These models employ convolutional and recurrent neural networks to detect inconsistencies that arise during manipulation. The process of creating fake videos usually involves altering real video to depict fabricated actions and generating highly realistic results that are difficult to detect (Botha & Pieterse, 2020).

Over the years, several DL studies have attempted to address these challenges; the study conducted by Sabir et al. (2019) used spatio-temporal relations between frames and observed that fake videos tend to have temporal discontinuities. The autoencoder-decoder pipeline employed by Khalil and Maged (2021) recreates the target frames and reveals the differences between real and fake videos. Moreover, Elhassan et al. (2022) developed the DFT-MF model which is a CNN based model that utilises mouth region motion to identify abnormal lip synchronisation. On the same note, Yadav and Salmani (2019) examined Generative Adversarial Networks (GANs) in which the generator creates counterfeit content, and the discriminator identifies it.

Other modern DL works are Lewis et al. (2020), the authors combined spatial, spectral and temporal images with a multimodal CNN architecture, Deressa et al. (2024), the authors combined CNN and vision transformer modules to align pixels sequentially and Aymen and Hussein (2024), the authors used discrete wavelet transforms and CNNs to detect forgery at the frame level. In addition, El-Gayar et al. (2024) suggested a graph neural network based on video-level deepfake detection, whereas Ismail et al. (2021a, 2021b) used YOLO and YOLOv2 as the architectures on the image of precisely face localisation and manipulation detection.

Moreover, Mitra et al. (2021) and Malathi (2023) created CNN-based classifiers with the help of XceptionNet, InceptionV3, and ResNet50 with high accuracy to detect fake frames. Guera and Delp (2018) used CNN and RNN models, whereby convolutional layers obtained features, and recurrent layers modelled temporal variations, which resulted in high accuracy. Amerini et al. (2019) proposed an optical flow based approach that compares changes between two successive frames and discovered that VGG16 was superior to ResNet50 in detecting inconsistency on a frame-by-frame basis. Nguyen et al. (2022) also applied autoencoders to predict encoding and decoding mechanisms during generating and detecting manipulated videos.

Despite the high performance of the DL models, they can be characterised by high requirements in computational resources and voluminous, high-quality datasets. According to the research conducted by Amerini et al. (2019), the accuracy of ResNet50 was only 76% in certain cases, which demonstrates that deep learning does not necessarily provide high accuracy, particularly social-media videos that have compression artefacts.

2.3. Hybrid Approaches

Combining the representational power of deep neural networks and efficiency of machine learning classifiers, hybrid models represent a new approach to machine learning. Such frameworks use pretrained CNNs like ResNet, VGG or Inception to extract features and use the final classification with classical algorithms like SVM, Logistic Regression or Gradient Boosting. Ahmed et al. (2023) proved that the integration of CNN-generated embeddings and SVM enhanced the classification on small datasets. To achieve a high rate of accuracy of more than 86% on benchmark videos, Alzurfi and Altaei (2025) introduced a hybrid system that employed CNN-extracted features and a series of ML classifiers. Short-form content on social media is particularly well-placed in hybrid approaches due to their balanced accuracy in detection and computational efficiency. The present study adheres to this paradigm when applying a ResNet50-SVM model that has been trained on videos on social-media and collected through cross-validation. This composite system manages drawbacks of both pure ML and DL models, providing a high-performing and lightweight architecture to be deployed in practise.

2.4. Datasets Used in Prior Studies

The available literature is based on well-organised benchmark datasets, including FaceForensics++, DeepFake Detection Challenge (DFDC), and Celeb-DF (AbdElfattah et al., 2025), which consists of high-resolution face-centric clips under controlled circumstances. Although these datasets are helpful in benchmarking, they are not realistic in view of the diversity, compression artefact, and lighting anomalies of social-media videos created by users (Khan et al., 2023). According to Dadkhah et al., 2023, the models that are trained on only such datasets are poor when they are tested on real-life content on platforms such as Tik Tok and Facebook. The current study aims to address this shortcoming by coming up with a new dataset of 128 short-form videos, which were gathered on TikTok and Facebook, consisting of 78 real and 50 fake samples. These videos are between 5 to 30 seconds, and this is what is characteristic of social-media. There was extraction of frames, annotation of the frames and classification into real and

fake directories which offered realistic basis on assessing the performance of detection under real online circumstances.

2.5. Summary and Identified Gap

The literature reviewed has shown that there has been a significant advance in using machine learning, deep learning, and hybrid-based architectures in detecting deepfakes. However, key gaps persist. The majority of models rely on high-quality benchmark videos instead of social-media videos. Deep learning systems are also associated with high computational cost and inability to generalise. Machine learning classifiers are efficient but mostly use handcrafted or pre-extracted features and do not reflect temporal dynamics. Compromise Hybrid frameworks are a promising middle ground involving CNN-based feature extraction and classical ML classification. Based on this observation, the current research will develop a hybrid ResNet50-SVM model that will be trained and evaluated on a novel Tik Tok and Facebook dataset to create an effective and stable fake-video detector that applies in real-life social-media scenarios.

3. METHODOLOGY

This paper proposes a hybrid architecture that combines deep learning and machine learning models in detecting manipulated videos. As illustrated in Fig. 1, the workflow consists of five steps namely, dataset collection, preprocessing and face detection, deep feature extraction, machine learning classification, and experimental evaluation. The architecture uses deep neural representations of ResNet50 and the discriminative ability of a Support Vector Machine (SVM) classifier to identify visual irregularities associated with fake content.

3.1. Dataset Collection

The data used in this paper was a manual collection of publicly available short-form videos on Tik Tok and Facebook two widely used platforms where real and fake content is postponed in the same degree. These platforms were selected based on the idea that the dataset should consist of real-world diversity and the dynamism of social media videos. The number of short-form videos collected was 128 representing 78 real and 50 fake clips. The videos were of between 5 and 30 seconds, which is common with social media posts. With the openCV library, every video was broken down into still image frames with a fixed sampling rate. The count of the extracted frames per video varied between 2 and 220 and the median count was around 15, and the total number of annotated frames was around 2,440 because of the difference in the video duration and motion activity. Each frame had all its dimensions shrunk to 224 x 224 pixels and saved as JPEG files. The frames were put on two directories RealFrames (around 1,490 images) and FakeFrames (around 950 images).

a) Annotation and Verification: Two reviewers independently checked each video and assigned labels to them. A video was regarded to have been faked in case it presented two or more signs of manipulation including facial boundary artefacts, lip movements that were asynchronous, irregular blinking, or uneven illumination and texture. The disagreements that arose among reviewers were solved by discussing the issues, and Cohen, Kappa ($k = 0.91$) was obtained, which means a high level of inter-rater reliability.

b) Video-Level Organisation: The difference in the number of frames per video was significant, hence the dataset was processed at video level rather than frame level. Frame embeddings were summarised so that short and long videos have the same contribution in the model training.

c) Ethical Implications: The videos were all publicly available, and any personal identifiable information (such as the usernames, captions or comments) was removed prior to processing. The data was gathered exclusively to be used in academic research.

3.2. Data Preprocessing and Face Detection

Preprocessing has been done so that it will be uniform, improve the image, and concentrate on areas of the face that are the most vulnerable to manipulations. OpenCV was used to deconstruct each video into still frames. Face location in each frame was done by a Deep Neural Network (DNN)-based face detector, which was developed based on the Single Shot Multibox Detector (SSD) architecture, and used a ResNet backbone. The identified faces were cropped, rescaled to 224 x 224 pixels, and brought to the [0,1] interval to make them uniform in their light and size. Histogram equalisation and denoising philtres were used in order to increase contrast and decrease artefacts and make the picture more visual. This preprocessing made sure that the extracted facial regions had enough visual fidelity to be used in extracting features later.

3.3. Deep Feature Extraction Using ResNet50

Deep feature extraction involved a pretrained ResNet50 model which was trained on ImageNet dataset. The last connected classification layer was eliminated and features were obtained out of the penultimate global average pooling layer, which generated a 2048-dimensional feature vector per frame. The architecture of the ResNet50 was chosen due to its depth, residual connexions, and demonstrated capability to represent hierarchical features- low-level pixel irregularities and high-level semantic inconsistency, which is typical of deepfake videos. The weights of the model were not trained (no fine-tuning) and it was possible to use the pretrained network as a feature extractor only. This approach gave consistent and non-selective representations and minimised computational expenses. For each video, all frame-level embeddings were averaged to create one representation vector, which captured the overall spatio-temporal properties of it.

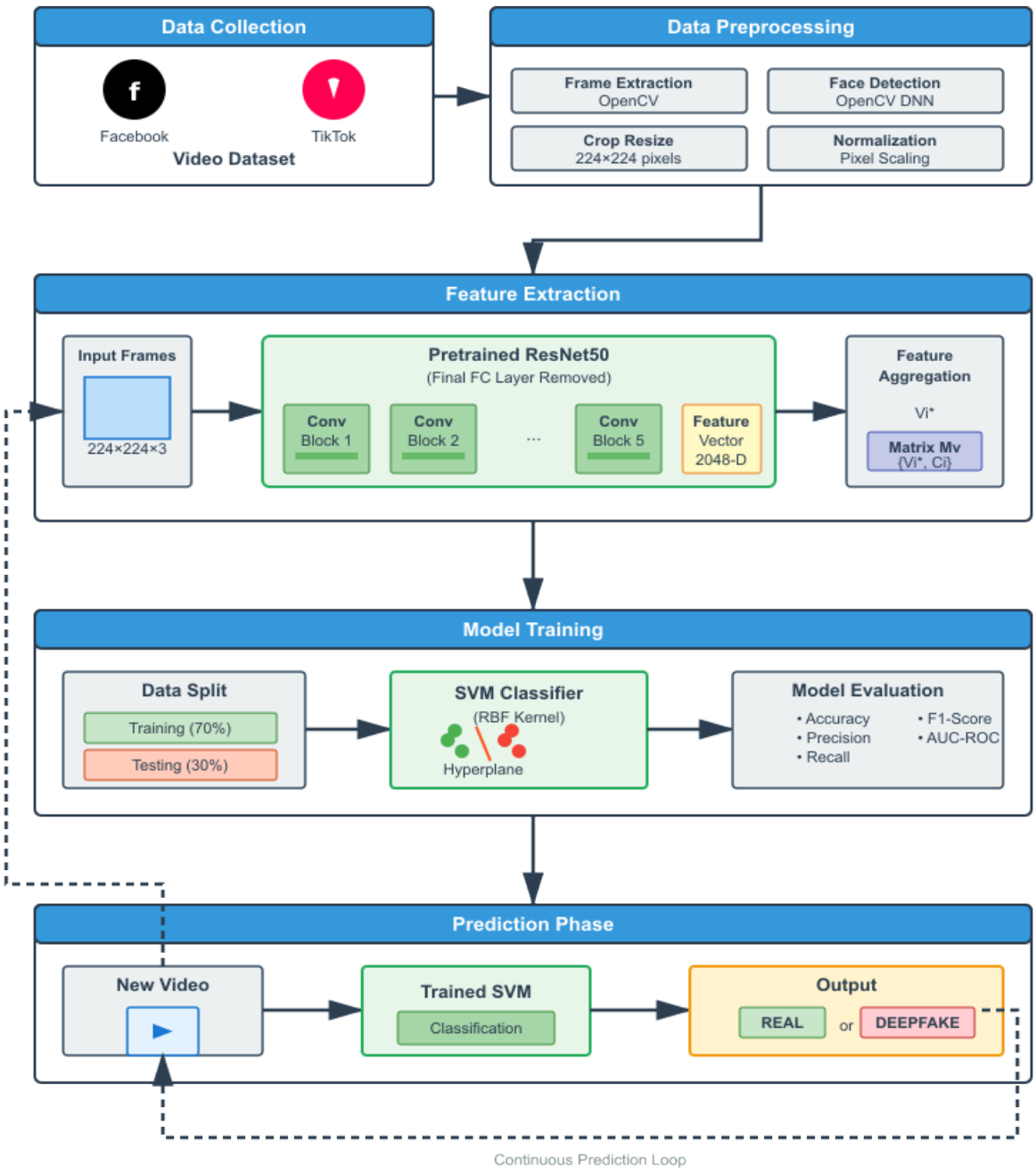

Fig. 1. Fake Video Detection Model

3.4. Machine Learning Classification

Stratified sampling was used to divide the aggregate feature matrix into training (70%) and a testing (30%) sample to ensure there was an equal number of classes in the real and fake videos. Classification was based on a Support Vector Machine (SVM) classifier using a Radial basis function (RBF) kernel. The model was selected as it is useful in processing high-dimensional, nonlinear feature spaces based on deep networks and due to its good generalisation on rather small datasets.

The penalty constant C and kernels coefficient g that have been hyperparameterized were optimised with grid search and five-fold cross-validation. This made sure that the boundaries of classification were as high as possible with a minimum overfitting. When making an inference new videos were processed through the same feature extraction pipeline and the trained SVM generated binary predictions on whether the video was real or fake. The general workflow of the suggested model is described in Algorithm 1. It begins by decoding frames of every video and identifying areas of faces to analyse them. The faces are preprocessed and then subjected to the ResNet50 model to get the deep feature representations and the same is summed up into a single feature vector per video. With the help of these vectors, an SVM classifier is trained to recognise real and fake videos. New videos are subjected to the same procedure during testing and the trained model estimates their authenticity.

Algorithm 1: Fake Video Detection

```
Input: V = {V1, V2, ..., Vn}      // Set of social media videos
Output: C ∈ {0, 1}                // 0 = Fake, 1 = Real

1.  Begin
2.  For each video Vi ∈ V:
3.      Extract frames Fi using OpenCV
4.      For each frame f ∈ Fi:
5.          Detect and crop face using OpenCV DNN face detector
6.          Resize to 224×224 and normalize pixel values
7.      End For
8.  End For
9.  Load pretrained ResNet50
10. For each frame f ∈ Fi:
11.     Extract 2048-dimensional feature vector
12. End For
13. Aggregate frame features to obtain Vi*
14. Construct feature matrix Mv = {Vi*, Ci}
15. Split Mv into Train (70%) and Test (30%) sets
16. Initialize SVM classifier with RBF kernel
17. Train SVM using Train set
18. Evaluate on Test set using metrics
19. For each new video Vnew:
20.     Repeat Steps (3–13)
21.     Predict Ŷ = SVM(Vnew)
22.     Display Ŷ
23. End For
24. End
```

4. IMPLEMENTATION

This part reports on how the proposed fake video detection framework can be implemented in practise. The entire experiment was carried out on a Google Colab environment to take advantage of the acceleration provided by GPUs to efficiently train the model and make inferences. It was developed based on TensorFlow, Keras, OpenCV, scikit-learn, and NumPy to preprocess data, create models, as well as to evaluate their performance. The data, which has been described in the section 3.1, was short-form videos recorded on Tik Tok and Facebook, comprising both real and faked samples. OpenCV was used to break down each video into

frames and a DNN face detector was used to obtain facial regions. The features Before extracted faces were normalised to a 224 x 224 pixel size and before features extraction. The video dataset in terms of images is shown in Fig. 2, whereas Fig. 3 demonstrates the number of classes of real and fake samples employed in the process of the model training.

Representation of features was done by using a pretrained ResNet50 model that was initialised using ImageNet weights. The last classification layer was then eliminated in order to have the network acting as a pure feature extractor hence producing a 2048-dimensional embedding of every detected face. This made the model both able to predict pixel-level deviations and high-level semantic indicators that define deepfake manipulations. The use of fine-tuning was avoided since ResNet50 actually offered good generalizable features to tell between the real and the altered visual contents. These embeddings were extracted and then trained a SVM classifier that was based on RBF kernel. Optimization of hyperparameters, such as the penalty parameter (C) and kernel coefficient (g) was done using grid search and a five-fold cross-validation to ensure good generalisation and avoid overfitting. The data was divided into training and test set in a ratio of 70:30 in order to have equal representation of the two classes.

Model assessment was based on standard performance measures such as accuracy, precision, recall, F1-score and the area under the receiver operating characteristic curve (AUC). To obtain statistical reliability and reduce bias in the performance reported, a five-fold cross-validation strategy was used. In order to measure the robustness of the hybrid architecture, ResNet50-SVM was compared with other supervised models such as Logistic Regression, K-Nearest Neighbour, Random Forest, and Gradient Boosting models using the same extracted features. The hybrid model was significantly more accurate and recalled and had a better AUC, indicating its suitability in automated recognition of manipulated videos on a variety of social media content. Overall, the practicality of hybridising deep convolutional feature extraction and a classical machine learning classifier to enable efficient and accurate fake video content detection was proven by the implementation.

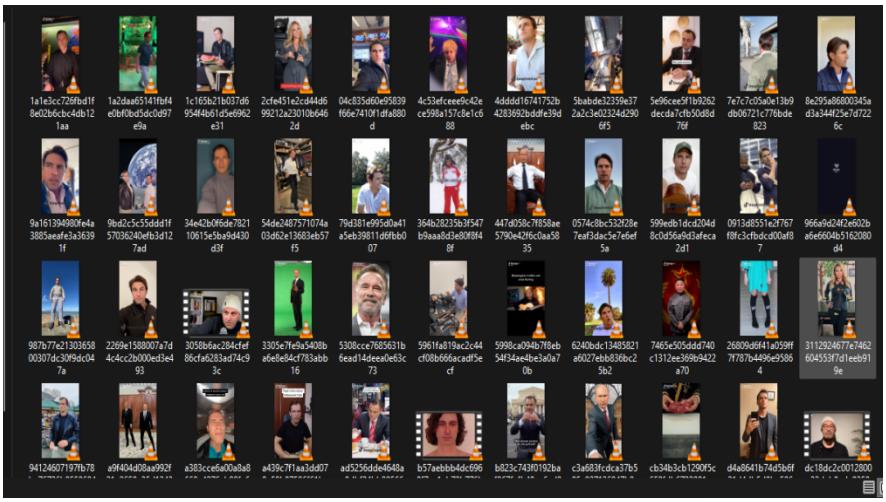

Fig. 2. Sample Images of the Video Dataset

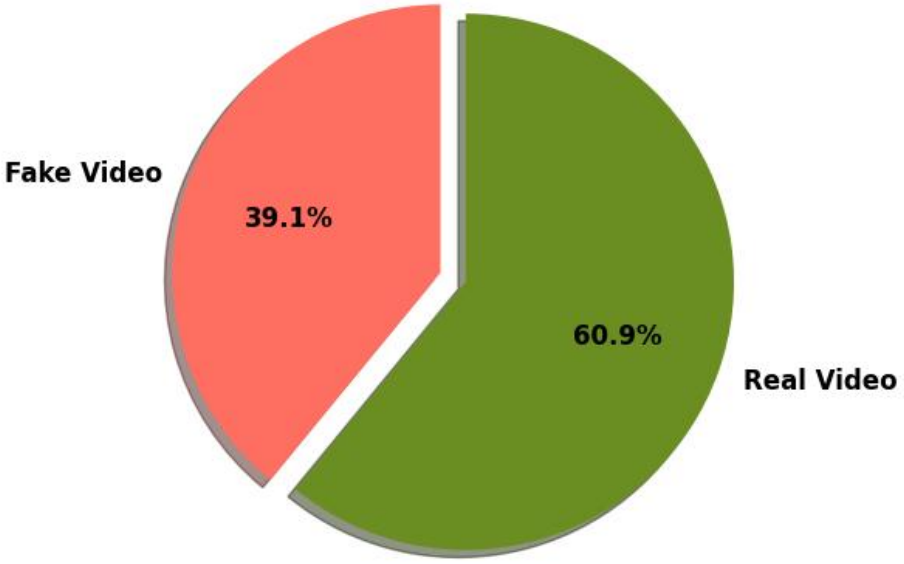

Fig. 3. Class Distribution of Video Dataset

5. RESULT AND DISCUSSION

5.1. Implementation Results

To assess the performance of the hybrid ResNet50-SVM model, several analyses such as the confusion matrix, ROC curve, and visualisation of features through PCA were used to give both quantitative and qualitative information on the classification's capabilities of the hybrid model. The accumulated confusion map (see Fig. 4) demonstrates that the model made correct decisions in 78 fake videos (True Fake) and 96.2 real videos (True Real) and 22 fake videos which were wrongly classified as real (False Real) and 3.8 real videos which were wrongly classified as fake (False Fake). These figures point toward the fact that the classifier has a high accuracy and mostly it identifies the authentic material and false video content, with a minor percentage of misclassifications occurring due to videos containing milder or more realistic manipulations.

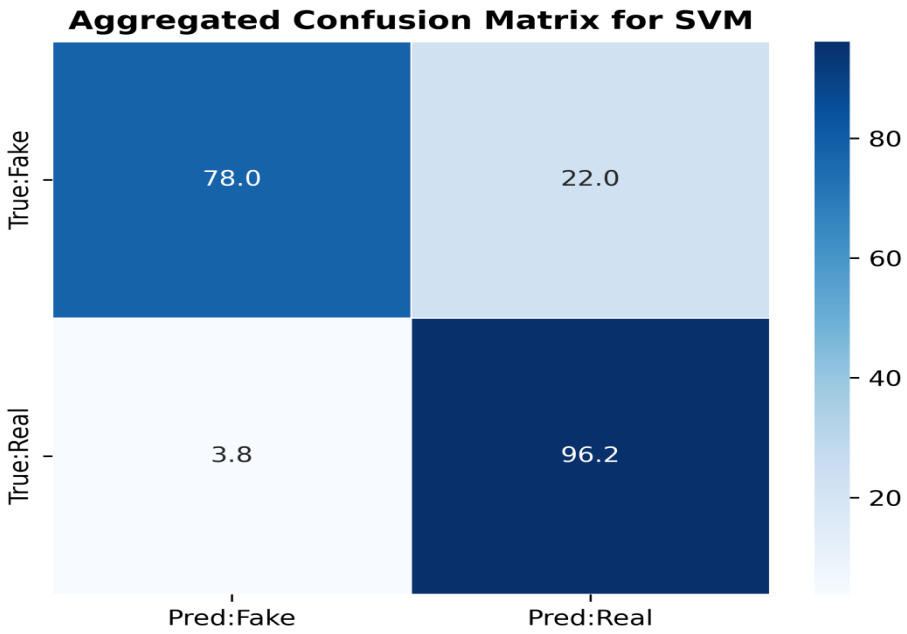

Fig. 4. Confusion Matrix for SVM

The Receiver Operating Characteristic (ROC) curve (see Fig. 5) also indicates the strength of the model with an AUC of 0.996 that indicates almost perfect separation between authentic and fake videos. This suggests that the classifier has high sensitivity and specificity under varying thresholds, which further substantiates the use of the classifier in real world video authentication activities. Further qualitative support is in the form of a 2D Principal Component Analysis (PCA) projection of the extracted video features (Fig. 6). This projection reveals that fake and real video are separated into different clusters with little overlap which means that features that are learned by ResNet50 are highly discriminative. The fake videos are clustered in focused parts of the feature space which show that they share common features that are not similar to the genuine ones and the real videos are more varied but still show certain patterns that can be separated. The small ratio of overlap is in line with the good classification accuracy, which is a confirmation that the first two principal components hold the most pertinent discriminative information.

Going further with this analysis, a 3D PCA projection (see Fig. 7) shows that the fake and real video features can be further separated into three major components. False videos create compact clusters in particular areas, whereas the real video is presented in various clusters, which means more variation in authentic materials. Such a conspicuous difference in 3D space further illustrates the high classification rate of the SVM model, where the learned feature representations are resilient and discriminative in higher dimensional feature space.

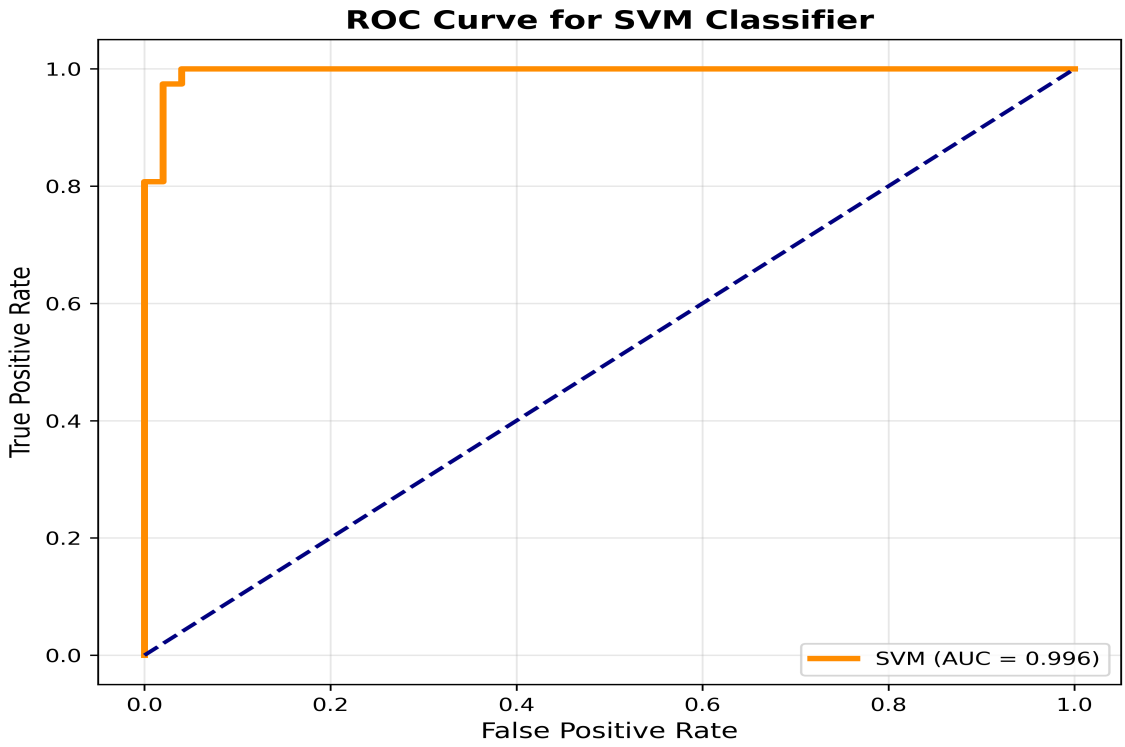

Fig. 5. ROC Curve for the SVM Classifier

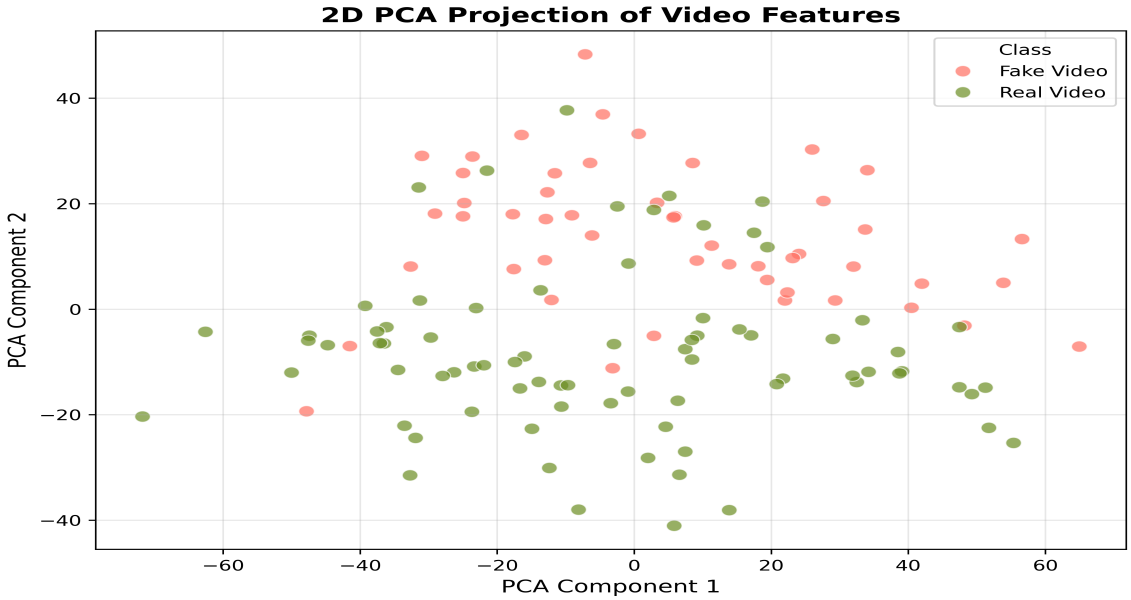

Fig. 6. 2D PCA Projection of Video Features

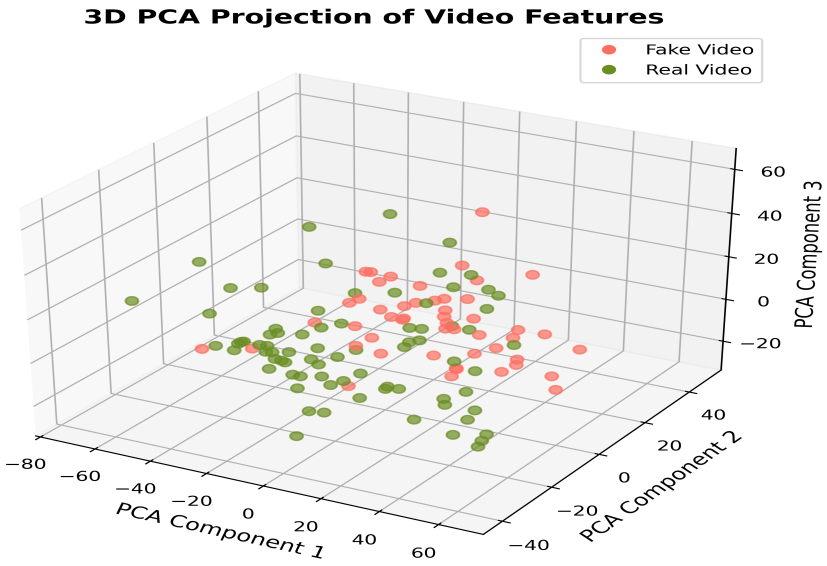

Fig. 7. 3D PCA Projection of Video Features

Further examination of misclassified points by 2D PCA (see Fig. 8) indicates that misclassified points fall in areas where the two classes also overlap, especially along particular sets of combinations of the two major components. This indicates that the combinations of features cause some videos to be more challenging to categorise by nature, which can be used to understand the remaining errors in the confusion matrix. Knowledge of these overlapping regions can be used to direct subsequent feature engineering or model refinement to further minimise misclassification. Altogether, these results indicate that the hybrid ResNet50-SVM model is very efficient in differentiating between real and fake videos, has good generalisation, has well-represented features, and can provide clear-cut answers about the origins of the residual classification errors.

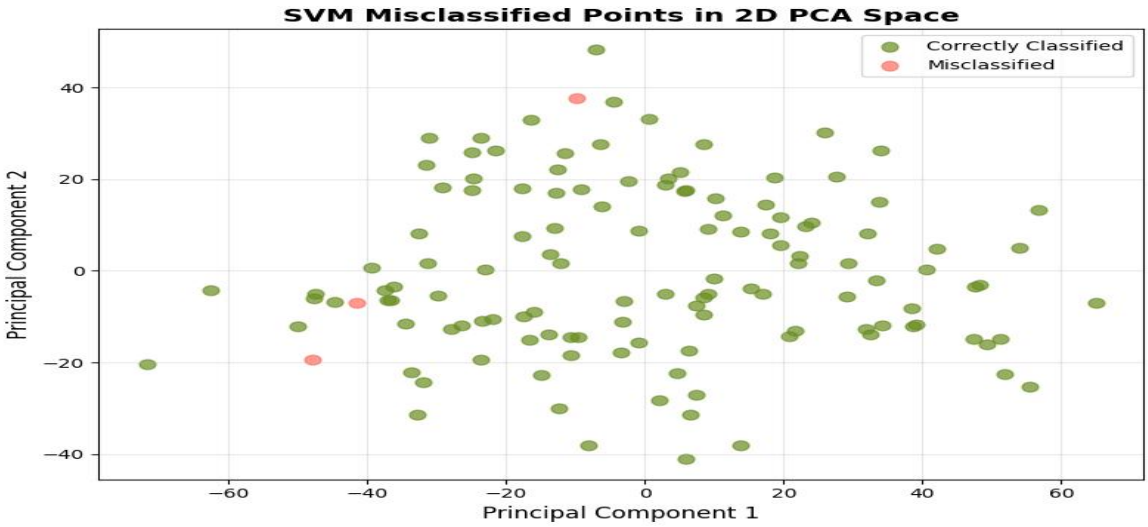

Fig. 8. SVM Misclassified Points

5.2. Comparative Analysis with Baseline Models

In a further attempt to confirm the strength and stability of the hybrid ResNet50-SVM model, other baseline classifiers were tested namely: Random Forest (RF),

Gradient Boosting (GB), K-Nearest Neighbours (KNN), and Logistic Regression (LogReg). Though SVM is the primary classifier of the approach to the methodology in Section 3.4, these other models were also added to the study on comparative grounds, in order to compare SVM with other standard supervised learning methods. Each of the models was trained and evaluated on the same 70:30 stratified portion of the data, and with five-fold cross-validation repeated ten times. Averaging of the performance metrics over the iterations was done to ascertain the statistical reliability and reduce the variability. SVM, Random Forest and Gradient Boosting are compared in Fig. 9. The SVM model has the most balanced performance with a recall of 96.1, F1-score of 91.5 and AUC of 95.2, and are very good at correctly classifying fake videos and separating between classes. Random Forest has better raw accuracy (96.1%), but poorer recall (73.4) and precision (75.3), which means that it has a higher number of misclassified positive samples. Gradient Boosting does average as far as it has gradual yet less pronounced outcomes.

Fig. 10 extends this comparison to include KNN and Logistic Regression. Here, SVM maintains its superior performance across all metrics. KNN, although showing high precision (92.6%), suffers from poor recall (60.3%) and accuracy (72.7%), indicating difficulty capturing all fake videos. Logistic Regression demonstrates consistent performance across metrics and ranks second overall, with accuracy of 83.9%, recall of 88.3%, and AUC of 91.5%. These comparative analyses clearly demonstrate that the hybrid ResNet50–SVM model consistently outperforms classical and ensemble classifiers, providing the best trade-off between precision, recall, and overall discriminative ability for video forgery detection. Fig. 11 shows the interface of the designed web application that will be the front-end platform for real-time fake videos detection. The interface permits the user to add video content and obtain the results of the classification with the best-performing model (SVM). The implementation brings about a cross between model creation and functional application and serves to show the feasibility and functionality of a system in a real-world implementation. The combination of machine learning with an interactive platform allows identifying the full potential of the project regarding data preprocessing, model training, and inference in real-time and interaction with users.

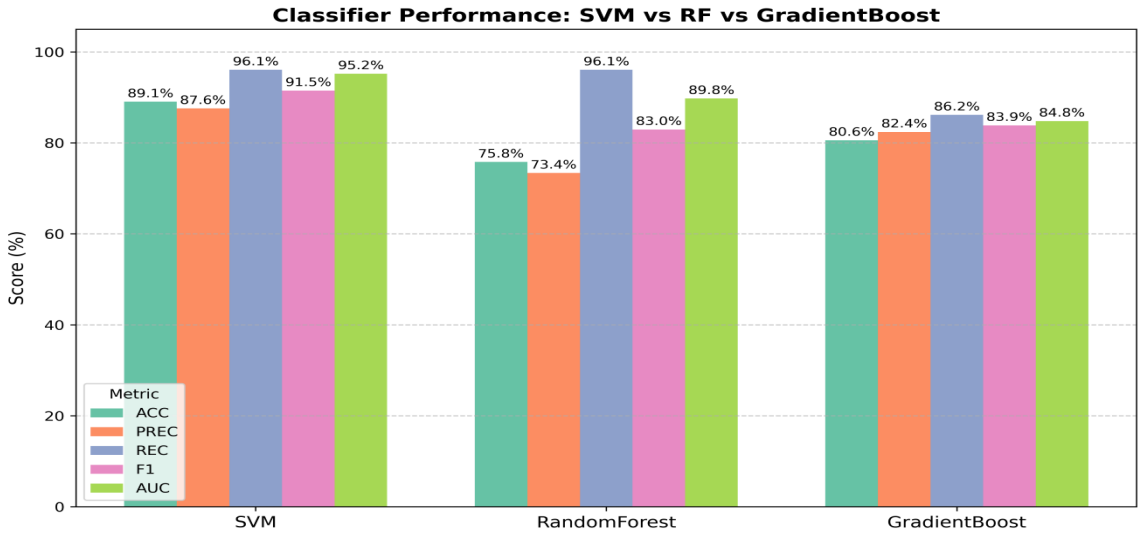

Fig. 9. SVM vs Random Forest vs Gradient Boosting performance

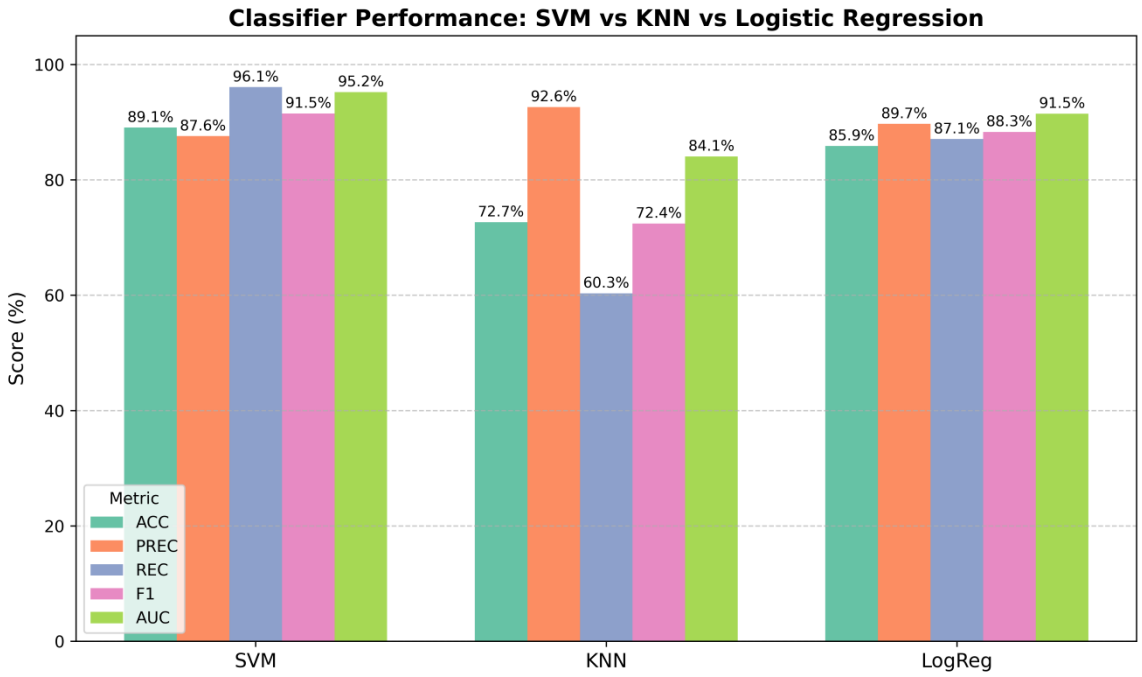

Fig. 10. SVM vs KNN vs Logistic Regression performance

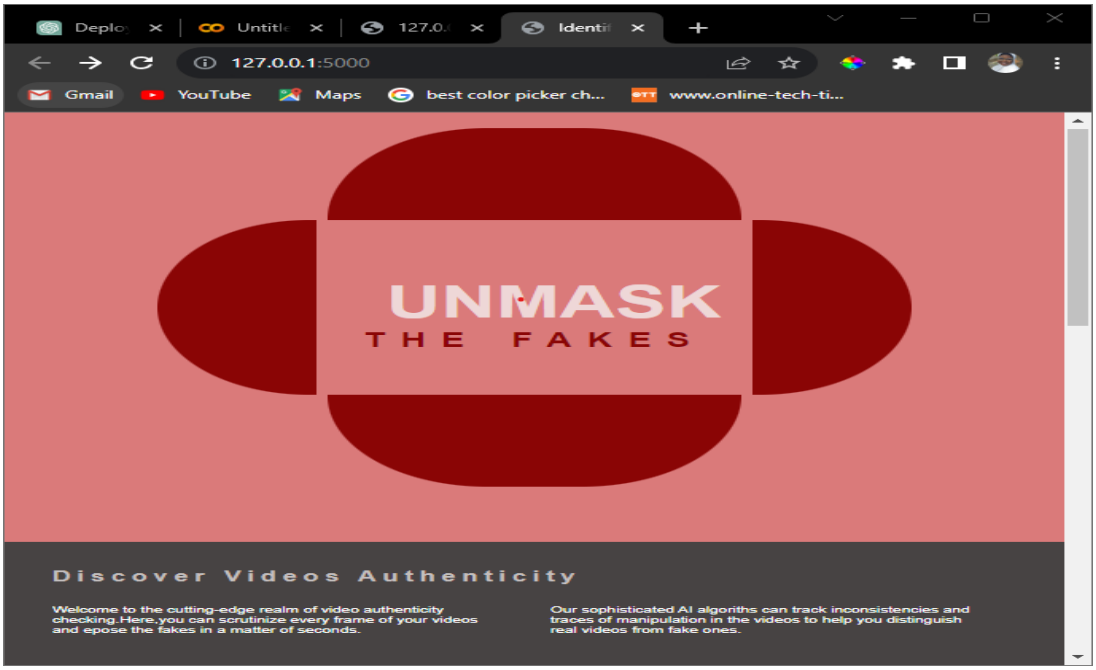

Fig. 11. Web Application User Interface

6. CONCLUSION

This study presents a hybrid framework that integrates deep feature extraction using pretrained ResNet50 with a Support Vector Machine SVM classifier for detecting manipulated videos. The framework was evaluated using a curated dataset from TikTok and Facebook, where complex video content was transformed into meaningful feature representations, enabling accurate distinction between real and fake videos. Quantitative analyses, including the confusion matrix, ROC curve, and PCA projections, demonstrated high accuracy, recall, and AUC, while feature visualizations confirmed clear class separation and highlighted regions contributing to residual misclassifications. Comparative evaluation with baseline classifiers, including Random Forest, Gradient Boosting, K-Nearest Neighbours, and Logistic Regression, showed that the SVM model consistently outperformed alternative approaches, providing the best balance between precision, recall, and overall discriminative capability. The successful deployment of the model within a functional web application demonstrates its practical applicability for real-time video verification. These results underscore the effectiveness, robustness, and scalability of combining deep convolutional features with classical machine learning techniques in multimedia forensics, with potential for further improvements through expanded datasets, real-time deployment, and advanced architectures.

REFERENCES

Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review, 56(11), 13521-13617.

Aymen, F., & Hussein, W. (2024). Application of spatial and Wavelet transforms for improved Deep Fake Detection. In 2024 5th International Conference on Artificial Intelligence, Robotics and Control (AIRC) (pp. 13-17). IEEE.

Amerini, I., Galteri, L., Caldelli, R., & Del Bimbo, A. (2019). Deepfake video detection through optical flow based cnn. In Proceedings of the IEEE/CVF international conference on computer vision workshops .

Altaei, M. S. M. (2023). Detection of deep fake in face images based machine learning. Al-Salam Journal for Engineering and Technology, 2(2), 1-12.

Ali, G., Rashid, J., Hussnain, M. R. U., Tariq, M. U., Ghani, A., & Kwak, D. (2024). Beyond the illusion: ensemble learning for effective voice deepfake detection. IEEE Access, 12, 149940-149959.

Alemerien, K., & Al-Mahadin, M. (2025). Machine learning-based approaches for manipulated image and video forensics in digital criminal investigation. Multimedia Tools and Applications, 84(27), 32619-32641.

AbdElfattah, E., Mahmoud, N., M Mousa, H., & Elsisi, A. (2025). A Comprehensive Overview of Deep Learning for Deepfakes: Generation, Detection, Dataset: A Survey. IJCI. International Journal of Computers and Information.

Botha, J., & Pieterse, H. (2020). Fake news and deepfakes: A dangerous threat for 21st century information security. In ICCWS 2020 15th International Conference on Cyber Warfare and Security. Academic Conferences and publishing limited (p. 57).

Christensen, J. (2021). AI in financial services. In Demystifying AI for the Enterprise. Productivity Press, 149-192.

Gignac, G. E., & Szodorai, E. T. (2024). Defining intelligence: Bridging the gap between human and artificial perspectives. Intelligence, 104, 101832.

Donepudi, P. K. (2019). Automation and machine learning in transforming the financial industry. Asian Business Review, 9(3), 129-138.

Deressa, D. W., Lambert, P., Van Wallendael, G., Atnafu, S., & Mareen, H. (2024). Improved Deepfake Video Detection Using Convolutional Vision Transformer. In 2024 IEEE Gaming, Entertainment, and Media Conference (GEM) (pp. 1-6). IEEE.

Dadkhah, S., Zhang, X., Weismann, A. G., Firouzi, A., & Ghorbani, A. A. (2023). The largest social media ground-truth dataset for real/fake content: Truthseeker. IEEE Transactions on Computational Social Systems, 11(3), 3376-3390.

Elhassan, A., Al-Fawa'reh, M., Jafar, M. T., Ababneh, M., & Jafar, S. T. (2022). DFT-MF: Enhanced deepfake detection using mouth movement and transfer learning. SoftwareX, 19, 101115.

El-Gayar, M. M., Abouhawwash, M., Askar, S. S., & Sweidan, S. (2024). A novel approach for detecting deep fake videos using graph neural network. Journal of Big Data, 11(1), 22.

Güera, D., & Delp, E. J. (2018). 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.

Hamza, A., Javed, A. R. R., Iqbal, F., Kryvinska, N., Almadhor, A. S., Jalil, Z., & Borghol, R. (2022). Deepfake audio detection via MFCC features using machine learning. IEEE Access, 10, 134018-134028.

Islam, M. R., Liu, S., Wang, X., & Xu, G. (2020). Deep learning for misinformation detection on online social networks: a survey and new perspectives. Social Network Analysis and Mining, 10(1), 82.

Ismail, A., Elpeltagy, M., S. Zaki, M., & Eldahshan, K. (2021a). A new deep learning-based methodology for video deepfake detection using XGBoost. Sensors, 21(16), 5413.

Ismail, A., Elpeltagy, M., Zaki, M., & ElDahshan, K. A. (2021b). Deepfake video detection: YOLO-Face convolution recurrent approach. PeerJ Computer Science, 7, e730.

Jing, T. W., & Murugesan, R. K. (2021). Protecting data privacy and prevent fake news and deepfakes in social media via blockchain technology. In Advances in Cyber Security: Second International Conference, ACeS 2020, Penang, Malaysia, December 8-9, 2020, Revised Selected Papers 2 (pp. 674-684). Springer Singapore.

Kharbat, F. F., Elamsy, T., Mahmoud, A., & Abdullah, R. (2019, November). Image feature detectors for deepfake video detection. In 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA) (pp. 1-4). IEEE.

Khalid, H., & Woo, S. S. (2020). Oc-fakedect: Classifying deepfakes using one-class variational autoencoder. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops (pp. 656-657).

Kuberkar, S., & Singhal, T. K. (2020). Factors influencing adoption intention of AI powered chatbot for public transport services within a smart city. International Journal of Emerging Technologies in Learning, 11(3), 948-958.

Khalil, H. A., & Maged, S. A. (2021). Deepfakes creation and detection using deep learning. In 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC) (pp. 1-4). IEEE.

Khan, S. A., Sheikhi, G., Opdahl, A. L., Rabbi, F., Stoppel, S., Trattner, C., & Dang-Nguyen, D. T. (2023). Visual user-generated content verification in journalism: An overview. IEEE Access, 11, 6748-6769.

Lewis, J. K., Toubal, I. E., Chen, H., Sandesera, V., Lomnitz, M., Hampel-Arias, Z., ... & Palaniappan, K. (2020). Deepfake video detection based on spatial, spectral, and temporal inconsistencies using multimodal deep learning. In 2020 IEEE Applied Imagery Pattern Recognition Workshop (AIPR) (pp. 1-9). IEEE.

Markauskaite, L., Marrone, R., Poquet, O., Knight, S., Martinez-Maldonado, R., Howard, S., ... & Siemens, G. (2022). Rethinking the entwinement between artificial intelligence and human learning: What capabilities do learners need for a world with AI?. Computers and Education: Artificial Intelligence, 3(2022), 100056.

Mitra, A., Mohanty, S. P., Corcoran, P., & Kougianos, E. (2021). A machine learning based approach for deepfake detection in social media through key video frame extraction. SN Computer Science, 2(2), 98.

Malathi, S. (2023). Breast Cancer Detection With Resnet50, Inception V3, And Xception Architecture. Journal of Pharmaceutical Negative Results, 14(4).

Nguyen, T. T., Nguyen, Q. V. H., Nguyen, D. T., Nguyen, D. T., Huynh-The, T., Nahavandi, S., ... & Nguyen, C. M. (2022). Deep learning for deepfakes creation and detection: A survey. Computer Vision and Image Understanding, 223, 103525.

Sabir, E., Cheng, J., Jaiswal, A., AbdAlmageed, W., Masi, I., & Natarajan, P. (2019). Recurrent convolutional strategies for face manipulation detection in videos. Interfaces (GUI), 3(1), 80-87.

Sarker, I. H. (2021). Machine learning: Algorithms, real-world applications and research directions. SN computer science, 2(3), 160.

Suryani, V., Yulianto, F. A., Sukarno, P., & Rizal, A. (2024). A Comparison of Deep Learning and Machine Learning Approaches to Video Injection Detection. Mathematical Modelling of Engineering Problems, 11(12).

Rana, M. S., Nobi, M. N., Murali, B., & Sung, A. H. (2022). Deepfake detection: A systematic literature review. IEEE access, 10, 25494-25513.

Yadav, D., & Salmani, S. (2019). Deepfake: A survey on facial forgery technique using generative adversarial network. In 2019 International conference on intelligent computing and control systems (ICCS) (pp. 852-857). IEEE.

Yaiprasert, C., & Hidayanto, A. N. (2024). AI-powered ensemble machine learning to optimize cost strategies in logistics business. International Journal of Information Management Data Insights, 4(1), 100209.

Yavuzkiliç, S., Akhtar, Z., Sengür, A., & Siddique, K. (2021).DeepFake face video detection using hybrid deep residual networks and LSTM architecture. In AI and Deep Learning in Biometric Security (pp. 81-104). CRC Press.