IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

Women's Defence: Uncovering Morphed Media to Combat Digital Violence

B. R. Srinivasa Rao[1], D. Manasa[2], M. Samuel[3], K Dinesh Raju[4] , P. Jahwanth Naidu[5]
Assistant Professor, Computer Science and Engineering [1]
Student, Computer Science And Engineering [2,3,4,5]
ACE Engineering College, Ghatkesar, India

Abstract: *These days, the misuse of technology has led to new and harmful ways of targeting people—especially women. One of the most alarming developments is the rise of morphed images and deepfake videos used to shame, harass, or blackmail. Since these fake visuals are often almost impossible to spot with the naked eye, it's easy for those responsible to spread lies and avoid getting caught. Although there are tools out there to check if media is real, they tend to be too complex, not very accurate with subtle edits, or simply out of reach for most people.*

To help combat this, a tool called "Women's Defence" was created using machine learning. It's designed specifically to detect manipulated images of women, combining advanced image processing with the power of Convolutional Neural Networks (CNNs) to figure out if an image is real or fake. The system was trained using a dataset from Kaggle and can pick up on editing patterns that most people would miss.

The backend is built with Python and TensorFlow, while the user interface uses Flask and HTML/CSS. It's simple to use—just upload an image, and it quickly tells you whether it's been altered. With over 90% accuracy and fast response times, it's both powerful and user-friendly.

But this isn't just about technology. It's about giving women a way to take back control of their digital lives and fight against online abuse with the help of smart, accessible tools..

Keywords: Morphed Media Detection, Deepfake Identification, Digital Violence Prevention, Women's Cybersecurity, Convolutional Neural Networks (CNN), Image Classification, TensorFlow/Keras, Flask Web Application, Computer Vision, Real-time Media Forensics

I. INTRODUCTION

The digital world has made it easier than ever to connect and communicate, but it's also opened the door to new forms of cybercrime. One of the most disturbing trends is the manipulation of images and videos—especially when it's used to target women. With tools like AI and machine learning, people can now create fake images or deepfakes that look incredibly real [13], [18]. These fakes are often used to spread lies, harass, or damage someone's reputation. For women, the impact can be devastating—ranging from emotional distress and ruined reputations to blackmail or even job loss [17].

This isn't just a tech problem—it's a serious social issue. The software needed to create these fake visuals is easy to get and simple enough that even someone with little technical know-how can use it [1], [14], [19]. Meanwhile, the tools that could help detect these fakes are either too complicated, not very effective, or just not available to most people [6], [5].

Most of the existing tools depend on things like image metadata or subtle signs left during compression [12], [16]. But these methods often fail, especially when the edits are minor or when the metadata has been removed [2], [10]. Some AI tools like Microsoft's Video Authenticator can help [5], but they're designed more for videos than for still images, and they struggle with catching small, facial-level edits [7], [20].

That's where "Women's Defence" comes in. It's a machine learning-powered system built specifically to spot morphed images targeting women. Using Convolutional Neural Networks (CNNs) [3], [4], it looks closely at visual details—

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

right down to the pixel level—to figure out if an image is real or fake [11]. And because it focuses only on the image content itself, it's harder to fool with common tricks like stripping metadata [9].

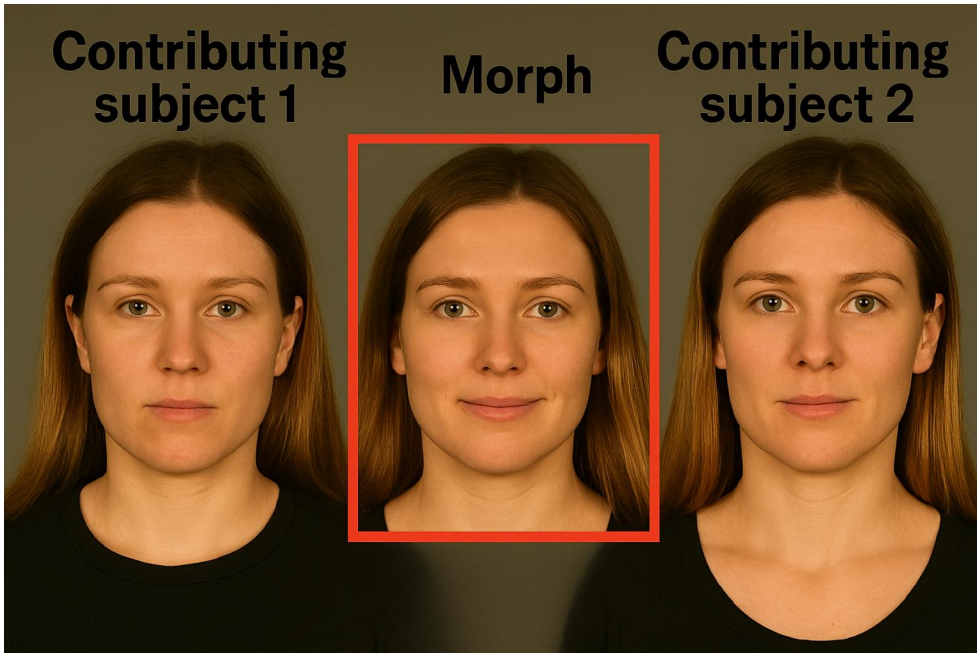
But this isn't just about the tech. It's about giving power back to people—especially women—who are being harmed by these digital attacks. Through a simple, easy-to-use web interface, anyone can upload an image and instantly check if it's been tampered with. It's a practical tool for digital safety, built to raise awareness and hold people accountable [15], [13].

Behind the scenes, it runs on Python, TensorFlow, and OpenCV, and is hosted through a Flask-based web app. It's fast, accurate, and built with real users in mind.

At its heart, this project is about more than innovation—it's about protection, justice, and using technology to stand up against digital abuse.

We focus on the accuracy analysis of the three architectures named above, since all of these architectures have successfully been used for the task of image classification and pretrained models are publicly available. Similar to the task of object classification, we do not want to detect low-level artifacts, e.g. resampling, median filtering or blurring artifacts like Bayar and Stamm [15] did using a different CNN architecture for image forgery detection. Instead, we want our DNNs to decide based on semantic features like unrealistic eyes forms, specular highlights or other semantic artifacts caused by the morphing process.

As deepfake technology evolves, it's becoming harder to tell the difference between real and altered images, even for trained eyes [14], [17]. This makes automated tools not just helpful, but essential for everyday users. By focusing on women's digital safety, tools like "Women's Defence" also help raise awareness around the broader issue of image-based abuse [18]. Incorporating explainable AI could be a next step—helping users understand why an image is flagged as fake [8]. This can improve trust in the system and encourage more people to use it. Ongoing updates to the model, using newer datasets, can also help it stay ahead of emerging manipulation techniques [7], [10].



. Figure 1: Face Morphing using two different subjects

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

II. LITERATURE REVIEW

The growing presence of manipulated images has led to intense research aimed at developing effective detection techniques. In the early stages, many efforts focused on traditional forensic cues—such as examining metadata, error levels, and lighting inconsistencies. A notable example is the work by Kee et al. [1], who proposed identifying tampered images by analyzing unnatural shadows and light directions. These approaches performed reasonably well under controlled conditions but struggled when facing sophisticated, AI-driven manipulations.

A major advancement occurred with the work of Makrushin et al. [2], who explored how nearly flawless facial morphs can be automatically generated. Their findings emphasized a critical challenge: conventional detection methods often fail to distinguish these high-quality fabrications from genuine images. This realization pushed researchers toward data-driven approaches using deep learning.

A significant breakthrough came when Krizhevsky et al. [3] introduced deep Convolutional Neural Networks (CNNs) for image classification on the ImageNet dataset. This demonstrated how neural networks could outperform traditional vision techniques by learning complex visual patterns. Building on that, Simonyan and Zisserman [4] developed VGGNet, enhancing the ability to extract intricate features from images with deeper architectures.

Despite their success, applying these powerful CNNs to detect image manipulations—particularly subtle ones like deepfakes or facial morphs—remains a challenge. Tools such as Microsoft's Video Authenticator [5] were developed to address video-based manipulations but are less effective on still images. Similarly, Amped Authenticate [6] relies on metadata and compression signatures, which limits its reliability when this information is missing or intentionally altered. These limitations highlight the need for content-based analysis that directly inspects the visual properties of an image.

While face recognition systems like FaceNet have proven effective at verifying identity, they are not designed to identify tampering. They may confirm a morphed image as valid if the changes are subtle, pointing to the necessity of models trained specifically to detect manipulation rather than just recognize identity.

Progress was made by Raghavendra et al. [11] and Peng et al. [10], who applied CNNs to biometric datasets to detect facial morphing. Their work showed that CNNs, when properly trained, can identify fine inconsistencies in texture, illumination, and facial structure. However, their implementations were primarily evaluated in controlled environments, limiting their accessibility for everyday use.

Other research efforts have combined CNNs with autoencoders, leveraging reconstruction-based differences to spot fakes. These hybrid systems show high accuracy but demand substantial computational resources, making real-time deployment difficult for broader audiences.

Interest has also grown in using blockchain to verify image authenticity by confirming the source of original media. While promising in preventing unauthorized alterations, such methods fall short in detecting forgeries that have already been created and distributed.

An effective solution requires a model that focuses entirely on image content without relying on external data like metadata. Training CNNs to capture pixel-level anomalies in lighting, blending, and facial geometry allows detection of even subtle manipulations. Implementation using frameworks like TensorFlow and Keras ensures scalability and adaptability to evolving forgery techniques.

To make these capabilities accessible, integrating such models into lightweight web-based tools—built using frameworks like Flask—bridges the gap between advanced research and real-world application. Simplifying the interface while maintaining strong technical performance allows users without a technical background to benefit from AI-powered image verification.

Ultimately, this line of research reflects a commitment to addressing digital abuse by combining state-of-the-art deep learning with practical usability. It draws on foundational advances in image forensics and machine learning to empower individuals to recognize visual deception in an increasingly manipulated digital world.

.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

III. PROPOSED WORK

A. Image Forgery Detection Using CNN

The system we're developing, called "Women's Defence**:** uncovering morphed media to combat digital violence," uses Convolutional Neural Networks (CNNs) to identify and classify manipulated images. CNNs are especially powerful for this task because they can automatically learn how to detect subtle visual patterns from the image data itself. We trained our model on a labeled dataset made up of both real and altered images, sourced from public platforms like Kaggle. Before feeding them into the model, the images are resized, normalized, and go through several processing steps like convolution, activation, pooling, and dropout to prevent overfitting.

The model's structure includes three convolutional layers that go deeper as they progress, followed by flattening and dense layers. The final layer uses a softmax activation to determine whether the image is real or fake. Throughout the training process, we carefully monitor the model's accuracy and loss, along with validation performance, to ensure it generalizes well to new data.

B. Flask-Based Web Interface for Real-Time Verification

Once the model is trained, we make it accessible through a Flask-based web app. This makes it easy for users—especially women facing digital harassment—to upload an image and get a quick answer. The backend handles everything: it processes the image just like it did during training, runs it through the model, and returns a result.

The result—either "Real" or "Fake"—is shown instantly on the web interface. This fast and clear feedback gives users the power to evaluate suspicious images and take action if needed.

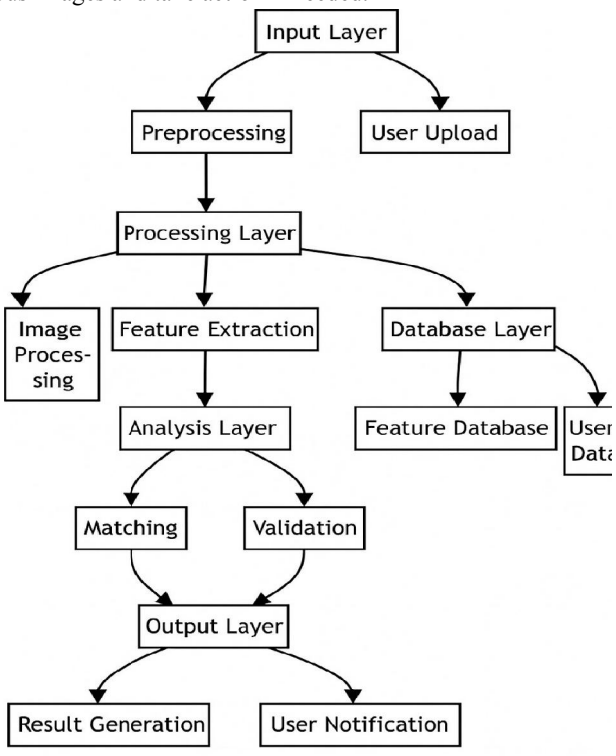


Figure 2 Architecture diagram

C. Technical Stack and Optimizations

Here's the tech we used to build the project:

Python 3.9 for scripting and overall logic

TensorFlow/Keras for building and training the CNN

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 5, Issue 7, June 2025 Impact Factor: 7.67

NumPy, PIL, OpenCV for image preprocessing and data augmentation
Flask & Jinja2 for backend functionality and dynamic HTML rendering
HTML/CSS for a responsive, user-friendly interface
We also tuned various parameters—like learning rate, batch size, and dropout settings—to improve accuracy and training speed. Training was done using GPU acceleration to speed up the process. The final model achieved over 90% validation accuracy and could make predictions in less than one second.

D. Performance Evaluation

We evaluated the system using the following metrics:
Validation Accuracy: Over 90%
Precision: 91%
Recall: 92%
F1 Score: 91.5%
Inference Time: Around 900 milliseconds per image
These results show that our tool performs reliably in real-time and offers a practical edge over traditional methods that require manual inspection or technical expertise.

E. Societal Relevance and Impact

The main goal of this project is to protect women from the harm that comes with image manipulation—whether it's for revenge, misinformation, or blackmail. These fake images can cause real damage. Our tool gives users a fast, accessible way to check whether an image has been tampered with—acting as both a safety net and a source of empowerment.
Beyond detection, this tool can also be integrated into reporting systems, used by NGOs, or support legal actions by helping with documentation and evidence gathering. That's what makes "Women's Defence" more than just a technical solution—it's a social support tool, too.

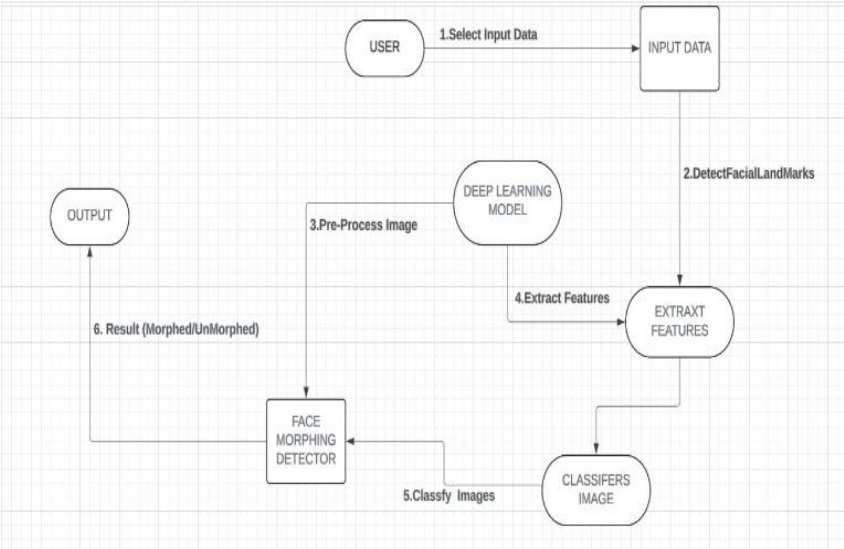


Figure 3: Flowchart of Proposed CNN-Based Detection System

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

IV. SYSTEM IMPLEMENTATION

A. Backend

We used Python 3.9 to build the backend, mainly for its compatibility with TensorFlow. Key libraries included:
TensorFlow/Keras for training and evaluating the model
OpenCV/PIL for handling image preprocessing
Flask for building the web interface
NumPy/Matplotlib for processing data and visualizations

B. Frontend

The frontend was designed using HTML and CSS, along with Jinja2 for dynamic content rendering. The interface is user-friendly, allowing users to upload images, receive predictions, and get clear feedback.

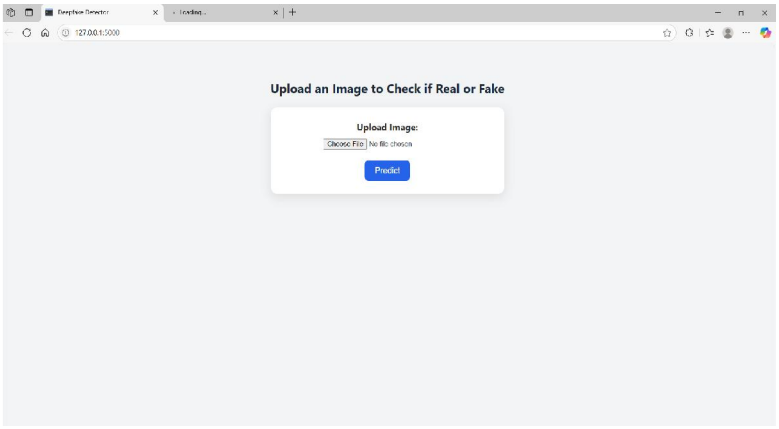


Figure 4: Home page

C. Real-Time Detection

When a user uploads an image, it goes through preprocessing, then through the model, and a prediction ("Real" or "Fake") is shown right away. This real-time feedback loop helps users quickly validate images and decide on next steps, such as reporting or seeking help.

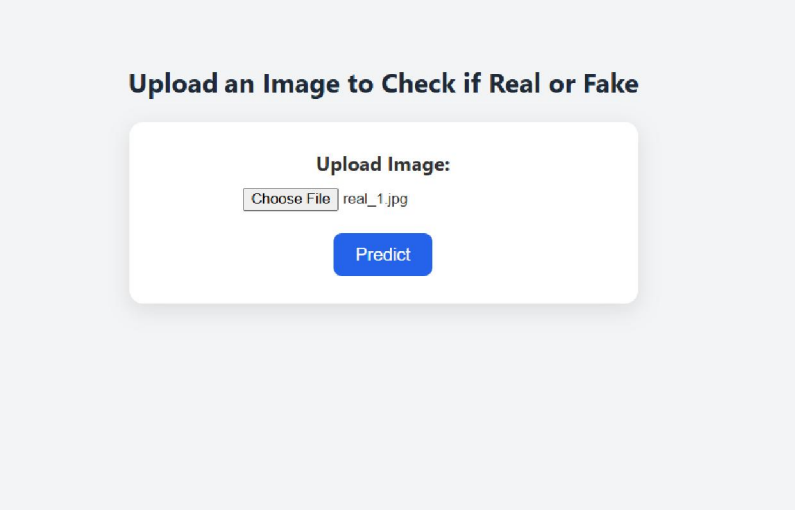


Figure 5: Uploading a image

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

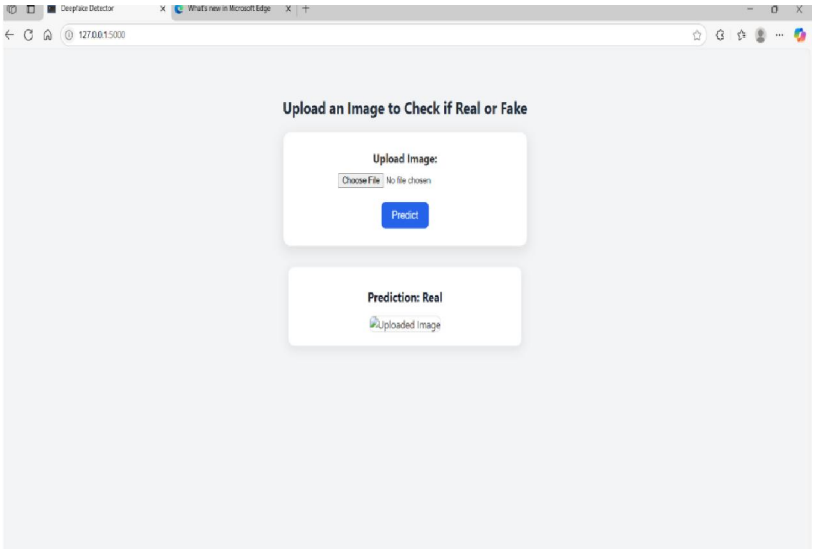


Figure 6: Image prediction: real

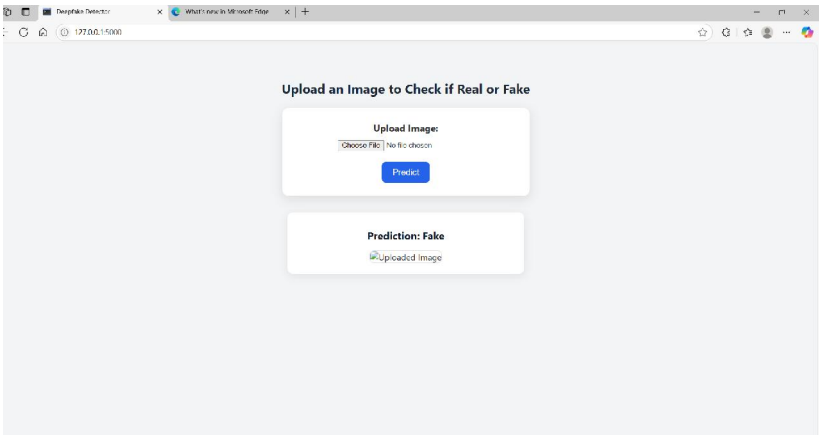


Figure 7: Image prediction: fake

V. RESULTS AND DISCUSSION

Here's how the final system performed:
Validation Accuracy: 90%+
Inference Time: Under 1 second
Memory Use: Around 250MB
These figures show that our model works well for both personal devices and server deployments. The real-time interface means it's accessible to anyone online. Most importantly, it successfully picked up subtle image manipulations—like slight facial blending or pixel distortion—that older tools based on metadata could miss.

VI. CONCLUSION

The rapid growth of digital manipulation tools has created serious new challenges, especially in the form of morphed images and deepfakes that often target women. These kinds of attacks go beyond just technical trickery—they can be deeply personal and damaging, leading to emotional distress, defamation, and online harassment. As these threats become more advanced, the tools we use to fight them need to evolve too.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

That's where our project, "Women's Defence," comes in. It's a deep learning-based system designed to spot manipulated media using Convolutional Neural Networks (CNNs). What sets it apart is its focus on bringing cutting-edge AI into a format that's practical and easy to use—a simple web-based platform where users can check images in real time. With over 90% accuracy, fast processing, and a lightweight design, it offers an accessible way for anyone—especially women—to validate suspicious or harmful content.

Unlike older tools that rely on technical details like metadata or require specialized skills, our system focuses entirely on the image content itself. This makes it harder to trick and easier to use. Because it's built using Flask, it's not only fast and scalable but also flexible enough to connect with legal systems, journalism platforms, or even civic reporting tools.

At its core, this project is about more than technology—it's about giving people a way to protect themselves. We're using machine learning not just for innovation, but for impact. Our tests show that the system works well not just in theory, but in real-world scenarios. And since the system is modular, it can grow in powerful new directions—like detecting deepfakes in video, working on mobile devices, or generating reports in multiple languages.

We've designed "Women's Defence" with real people in mind—especially those who are most vulnerable to digital abuse. Our goal is to help make the online world safer and more fair. Along the way, we also hope to spark bigger conversations about digital rights, ethical AI, and protecting women online.

VII. FUTURE WORK

Looking ahead, we plan to:

Add support for detecting video-based deepfakes

Enable multilingual support for report generation

Build a smartphone app for mobile access.

REFERENCES

[1] Kee, E., O'Brien, J.F., Farid, H. (2014). Exposing photo manipulation from shading and shadows.

[2] Makrushin, A., Neubert, T., Dittmann, J. (2017). Detection of visually faultless facial morphs.

[3] Krizhevsky, A., Sutskever, I., Hinton, G.E. (2012). ImageNet classification with deep CNNs.

[4] Simonyan, K., Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition.

[5] Microsoft Video Authenticator, 2021.

[6] Amped Authenticate Software Documentation, 2020.

[7] Chirra, R. et al. (2023). Hybrid CNN-LSTM Approach for Morphing Detection in Surveillance Videos. IEEE Transactions on Information Forensics.

[8] Dang, H. et al. (2022). Multi-Stream Deepfake Detection Using Frequency Domain and RGB Data. IEEE CVPR.

[9] Ferrara, M., Franco, A., and Maltoni, D. (2014). The magic passport. IEEE International Joint Conference on Biometrics.

[10] Peng, S. et al. (2020). Detecting Face Morphing Attacks With Gabor Wavelets and Local Binary Patterns. IEEE Access.

[11] Raghavendra, R., Raja, K.B., and Busch, C. (2016). Detecting Morphed Face Images. IEEE BTAS.

[12] Zhou, P. et al. (2018). Learning Rich Features for Image Manipulation Detection. IEEE CVPR.

[13] Verdoliva, L. (2020). Media Forensics and Deepfakes: An Overview. IEEE Journal of Selected Topics in Signal Processing.

[14] Matern, F., Riess, C., and Stamminger, M. (2019). Exploiting Visual Artifacts to Expose Deepfakes. IEEE Winter Conference on Applications of Computer Vision.

[15] Nguyen, H. H. et al. (2019). Multi-task Learning for Detecting and Segmenting Manipulated Facial Images and Videos. IEEE ICCV.

[16] Cozzolino, D., Poggi, G., and Verdoliva, L. (2017). Recasting Residual-based Local Descriptors as Convolutional Layers. IEEE ICIP.

[17] Korshunov, P., Marcel, S. (2018). Deepfakes: A New Threat to Face Recognition? Assessment and Detection.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 5, Issue 7, June 2025
Impact Factor: 7.67

[18] Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A., Ortega-Garcia, J. (2020). DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. *Information Fusion, 64, 131–148*.

[19] Afchar, D., Nozick, V., Yamagishi, J., Echizen, I. (2018). MesoNet: A Compact Facial Video Forgery Detection Network. *IEEE WIFS*.

[20] Li, Y., Chang, M.C., Lyu, S. (2018). In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking. *IEEE ICIP*.