DEEP FAKE IMAGE CLASSIFICATION

Introduction

Our project tackles the urgent task of accurately spotting and organizing deepfake images, with the goal of easing worries about their potential misuse and the spread of false information. As deepfake tech gets more sophisticated, traditional detection methods struggle to tell real from fake. So, we're diving into different machine learning and deep learning models to amp up cybersecurity against deepfake trickery. Our aim? To boost accuracy in spotting fakes and set up strong systems to keep our online world secure.

Methods

- **Utilization of Established Architectures**:

- We employed traditional neural network architectures such as AlexNet, VGG-16, and Xception Net, known for their effectiveness in image classification tasks.

- Additionally, we implemented Convolutional Neural Networks (CNNs) from scratch, allowing us to tailor the architecture to the specific requirements of our deepfake detection task.

- **Innovative Modifications and Explorations**:

- In addition to using established models, we introduced modifications to the AlexNet architecture by replacing its fully connected layers with recurrent neural network (RNN) layers to capture temporal dependencies in video-based deepfake detection.

- We also explored the novel approach of Vision Transformers, both from scratch and using pre-trained models, to evaluate their efficacy in image processing tasks.

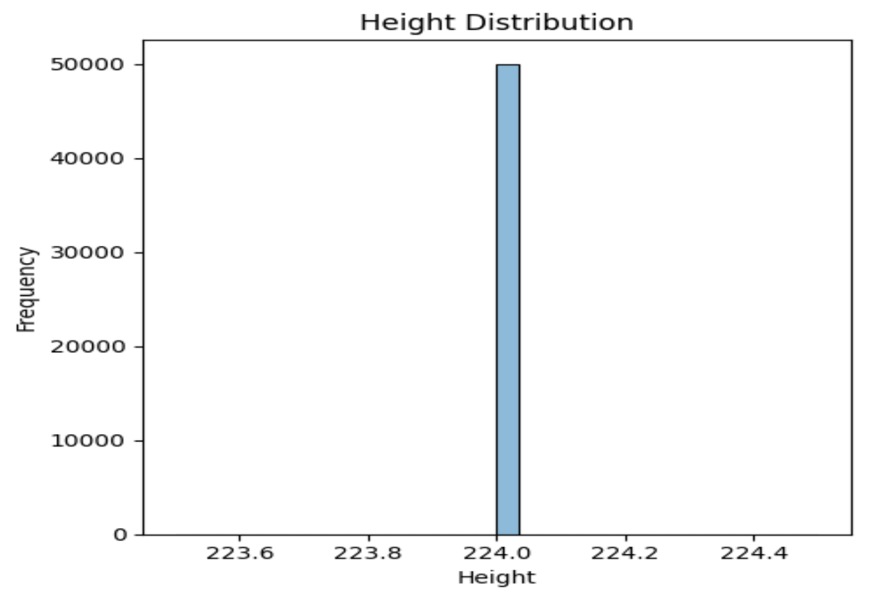
- Furthermore, we implemeted an ensemble model comprising AlexNet, ResNet-18, DenseNet-121, and Vision Transformer architectures to leverage the strengths of each model for improved classification performance.

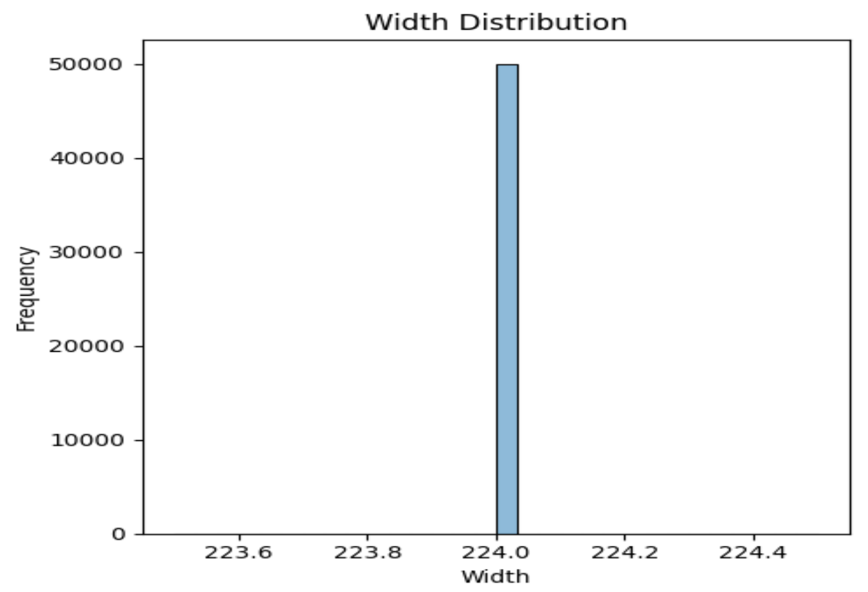
Data Analysis

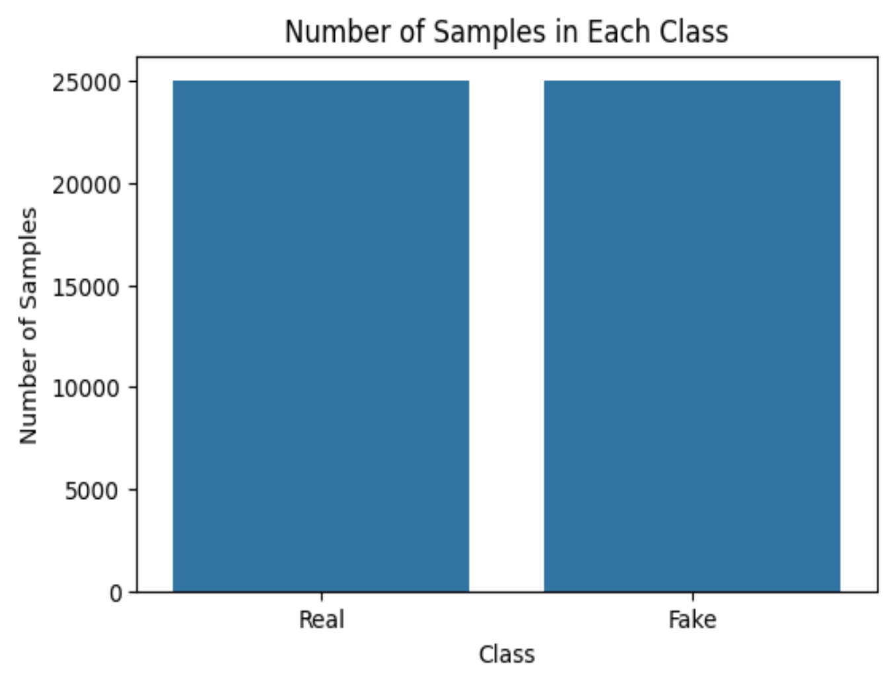
- The dataset used for our project consists of manipulated images and real images. Manipulated images are those generated through various means, such as deepfake techniques or image editing software, while real images depict genuine human faces.
- We obtained this dataset from the source https://zenodo.org/record/5528418#.YpdlS2hBzDd. The dataset contains a collection of JPG images, each depicting a human face, with a resolution of 256 pixels by 256 pixels.

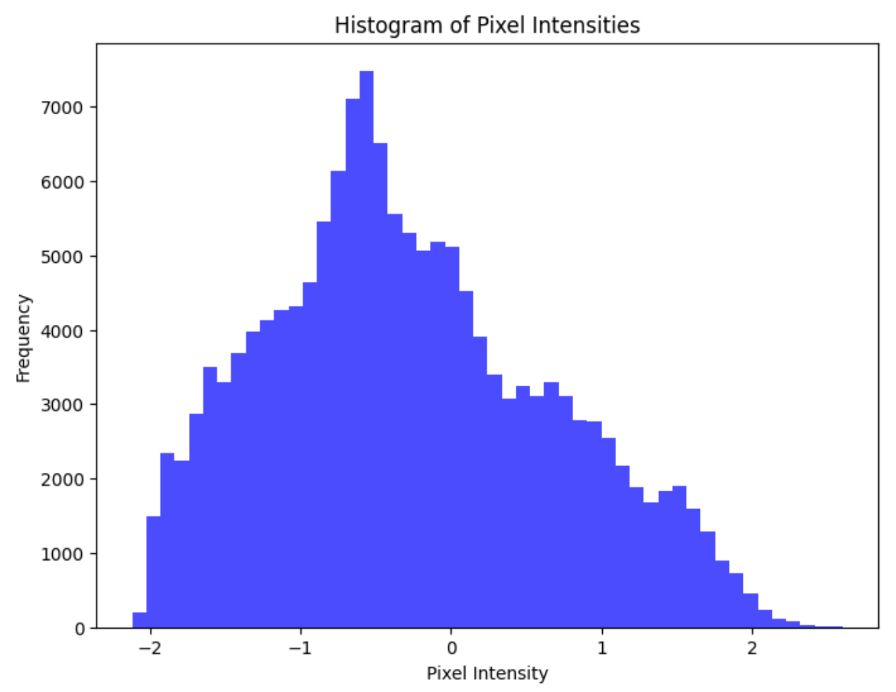
Preprocessing:

The dependent variable was encoded into binary labels (0 and 1), and images were transformed into tensors, normalized, and resized to 224 x 224 pixels. The dataset was balanced by sampling 20,000 images from each category, creating a subset split into 80% training, 10% validation, and 10% test sets, facilitated by dataloaders for model training and evaluation.
Below are the dataset distribution and various graphs.







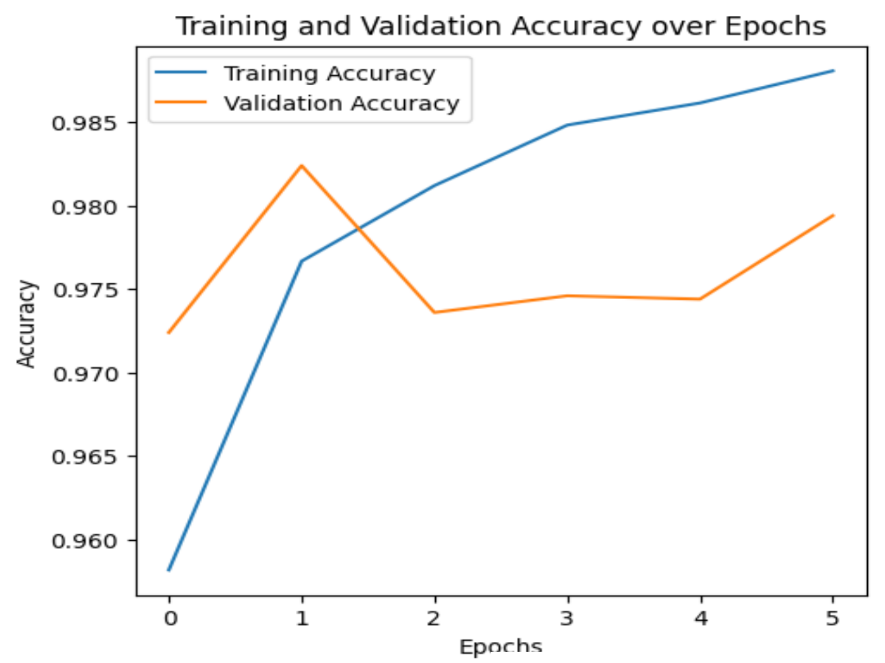


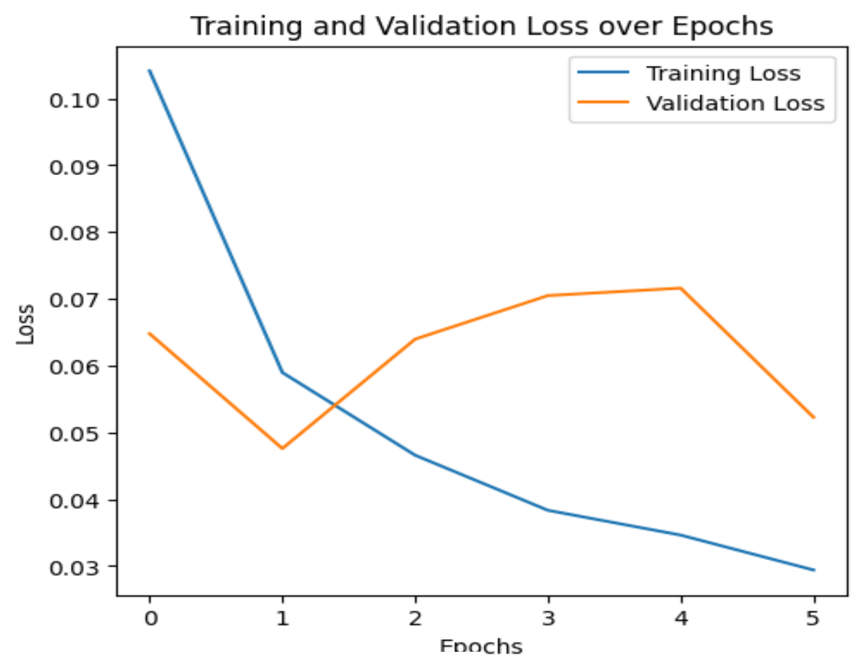
Results

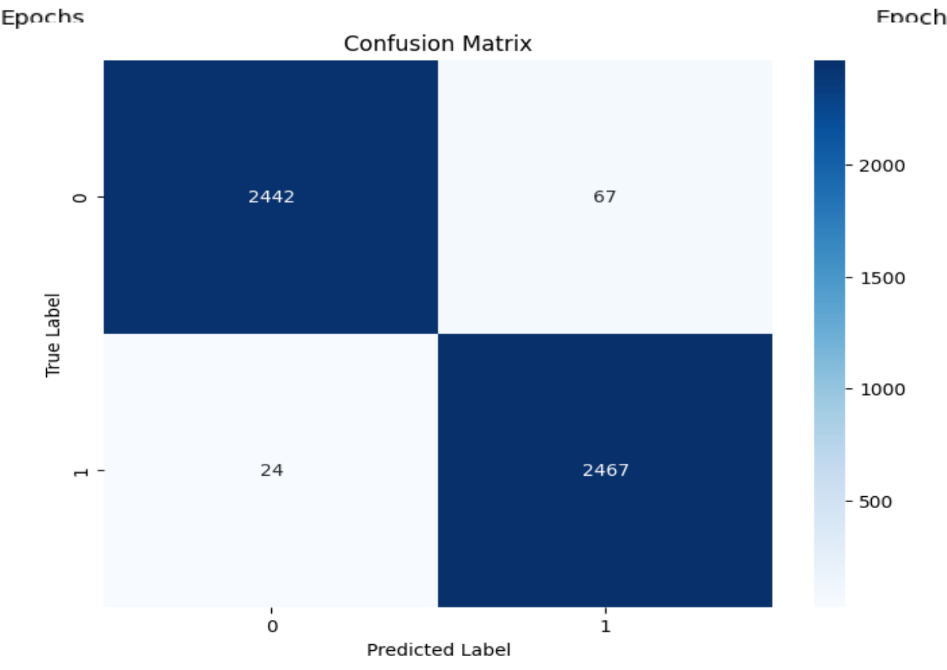
Below, you'll find the performance metrics of several image classification models, highlighting their respective accuracy scores. Further insights, including detailed accuracy breakdowns, confusion matrices, and loss graphs, are provided in the subsequent sections.

Model name	Base model	L2 regularization	Early stopping
Alex Net	95.78%	95.78%	94.58%
VGG-16	77.38%	81.42%	81.45%
Xception Net	89.56%	97.24%	96.96%
AlexNet + RNN	94.74%	94.96%	94.72%
Vit (from scratch)	52%	54%	54%
Vit (Pre trained)	84.68%	88.0%	88.28%
Ensemble	97.78%	98.1%	98.2%

These models represent a diverse range of architectures for image classification tasks. While Vision Transformer (from scratch) achieves a moderate accuracy of 52%, the pre-trained Vision Transformer exhibits notable performance at 84.68%. The Ensemble Method, combining AlexNet, ResNet, DenseNet, and Vision Transformer, achieves a remarkable accuracy of 97.78%, surpassing state-of-the-art accuracy at 98.2%. Detailed analysis including accuracy, confusion matrix, and loss graphs are shown below for this model.







Conclusion

Conclusively, this project explored a variety of image classification models, ranging from traditional architectures like AlexNet and VGG-16 to modern transformer-based approaches. Through comprehensive evaluation, including accuracy assessments and ensemble techniques, we demonstrated the efficacy of different strategies in achieving high classification performance. Moving forward, these findings pave the way for further exploration and refinement in image classification methodologies.

References

1.https://www.kaggle.com/datasets/manjilkarki/deepfake-and-real-images/data
2.https://pytorch.org/vision/main/models/vision_transformer.html
3.https://pytorch.org/vision/stable/models.html
4.https://medium.com/@brianpulfer/vision-transformers-from-scratch-pytorch-a-step-by-step-guide-96c3313c2e0c
5.https://www.geeksforgeeks.org/ensemble-methods-in-python/amp/