Journal of Artificial Intelligence and Capsule Networks (ISSN: 2582-2012)
www.irojournals.com/aicn/

Impact of Learning Rate on CNN-Based Deepfake Detection

Amber Fatima[1], Pintu Kumar Ram[2]

Amity School of Engineering and Technology Amity University Noida Uttar Pradesh, India

Email: [1]amberfatima1303@gmail.com, [2]rampintu570@gmail.com

Abstract

This research examines how CNN-based deepfake detection is affected by different fixed learning rates (0.0001, 0.0002, and 0.0005). To train a lightweight CNN model, video frames were taken, modified, and normalized using the Celeb-DF (v2) dataset. The model was trained utilizing three epochs with a batch size of four, employing the Adam optimization algorithm for enhanced performance. The performance of the model was evaluated through the analysis of training accuracy, validation accuracy, training loss, and validation loss. These metrics provided a comprehensive assessment of the model's effectiveness in both learning from the training data and generalizing to the validation dataset. The results of the study indicate that a learning rate of 0.0005 leads to instability and a tendency toward overfitting, while a learning rate of 0.0001 is associated with underfitting due to slow convergence. In contrast, an optimal learning rate of 0.0002 achieves a balanced performance, yielding the highest validation accuracy at 86% and the lowest validation loss of 0.35. This highlights the importance of selecting an appropriate learning rate to enhance model performance effectively. With possible uses in enhancing GAN-based image classification systems, this study focuses on the influence of learning rate selection on deepfake detection.

Keywords: Deepfake Detection, CNN, Celeb-DF (V2), Learning Rate, Adam Optimizer, Image Classification.

DOI: https://doi.org/10.36548/jaicn.2025.1.001
Received: 27.01.2025, received in revised form: 21.02.2025, accepted: 08.03.2025, published: 19.03.2025

1. Introduction

Since Generative Adversarial Networks (GANs) have advanced so quickly in recent years, making it possible to create extremely realistic fake media, deepfake detection has grown to be a serious concern. These fake images and films are a major risk to cybersecurity, privacy, and disinformation. A potent method for deepfake identification, convolutional neural networks (CNNs) provide a high degree of accuracy in differentiating between authentic and modified video. However, adjusting hyperparameters—specifically, the learning rate, which regulates the step size of weight updates during backpropagation—is essential for effectively training CNNs. Unstable training with divergence problems, underfitting, or slow convergence might result from a poorly selected learning rate. The importance of choosing the right learning rate for deep learning models has been emphasized in several research [1][2]. This work uses the Celeb-DF (v2) dataset to examine how CNN-based deepfake identification is affected by various fixed learning rates. Frames from the dataset—which includes both actual and deepfake videos—were taken out, shrunk to 128 by 128 pixels, and normalized to increase training stability. Next, the dataset was divided into two parts: 20% for testing and 80% for training. A sigmoid activation function for binary classification, convolutional layers, max-pooling layers, and fully linked layers were all included in the lightweight CNN architecture. For deep learning applications, the model was trained with the Adam optimizer, which adaptively modifies the learning rate according to first- and second-moment estimates [3]. Three distinct fixed learning rates were used in the experiments: 0.0001, 0.0002, and 0.0005. With a batch size of 4 to account for computing limitations, training was conducted in 3 epochs. According to the research, a learning rate of 0.0001 causes underfitting and delayed learning, whereas a rate of 0.0005 causes fluctuations and perhaps overfitting. With the maximum validation accuracy of 86% and the lowest validation loss of 0.35, the ideal learning rate (0.0002) strikes the best balance between accuracy, stability, and convergence speed. These results are consistent with earlier studies showing that effective convergence is facilitated by modest learning rates without leading to instability [4][5]. The study emphasizes how important it is to choose the right learning rate for CNN-based deepfake identification and recommends adaptive learning rate tactics such as learning rate annealing techniques [7] and cycle learning rates [6] which are to be investigated in future studies. Furthermore, including transfer learning from pre-trained models like ResNet or EfficientNet could improve deepfake classifiers' detection skills. By offering empirical insights into how learning rate selection affects CNN training, this study advances the field of deepfake detection and emphasizes the

need to optimize hyperparameters for practical applications [8][9][10]. Some of the main contributions of this study are given here: First it is a thorough assessment of CNN-based deepfake detection using fixed learning rates. Secondly, the practical proof shows the optimal trade-off between stability and convergence which is achieved at a learning rate of 0.0002, Thirdly, it is a thorough analysis of training results at various learning rates allows for further research on deepfake detection. By choosing a suitable learning rate for improved generalization and efficiency, these results aid in the optimization of CNN-based detection systems.

2. Related Work

The importance of learning rate optimization and how it affects deep learning models have been the subject of numerous studies. The capacity of GANs to produce realistic synthetic media was demonstrated in early research, underscoring difficulties with training stability and convergence [1]. Subsequent studies showed that deep convolutional GANs (DCGANs) were useful for feature learning, which led to improvements in deepfake detection and generation methods [2]. By adjusting learning rates, the Adam optimizer increased training efficiency and became a popular optimization technique for deep learning models, such as CNN-based classifiers [3]. CNNs can use the recognizable traces left by deepfake models for classification, according to additional research on the detection of GAN-generated artifacts [4]. In order to improve convergence features in adversarial learning, the two-time scale update rule for GANs addressed stability difficulties during training [5]. Furthermore, it was suggested that cycle learning rates be used to dynamically modify the learning rate during training in order to improve model performance by reducing underfitting and overfitting [6]. Stochastic gradient descent with warm restarts, or SGDR, was presented in another work. It improves convergence speed by avoiding local minima and regularly resetting the learning rate [7]. In order to guarantee consistent convergence and increased accuracy in generative models, research on training enhancements for GANs investigated improved optimization strategies [8]. The strategies that produce the best performance and resilience in deepfake creation and detection tasks were identified through investigations into several GAN training techniques [9]. Techniques for large-batch stochastic gradient descent were further refined to increase training speed without sacrificing accuracy in CNN-based models [10]. In order to improve CNN training efficiency and optimize deepfake detection models, these studies collectively show how important learning rate selection is [11-16].

3. Proposed Work

One of the most important challenges in CNN model training for deepfake detection is choosing the best learning rate. Low learning rates lead to slow training with poor generalization, whereas high learning rates can lead to instability and divergence. By methodically contrasting the impacts of fixed learning rates on CNN-based deepfake classification, this study seeks to use the best hyperparameter choices for improved model performance. Unlike prior research that focus on broad deepfake detection, this study explicitly assesses the influence of three fixed learning rates (0.0001, 0.0002, and 0.0005) on CNN-based classification of real and fake images retrieved from videos.

The initial step in the suggested method is dataset preprocessing, which involves extracting, resizing, and normalizing frames from the Celeb-DF (v2) dataset in order to improve training efficiency. The Celeb-DF (v2) dataset is one of the most broad publically accessible datasets for deepfake detection, with 5,639 genuine and 5,639 deepfake samples. Frames were extracted from each samples at a rate of one frame per second, giving over 500,000 images. For similarity, frames were modified to 128 x 128 pixels using bilinear interpolation. Min-max normalization was used to scale pixel intensities between 0 and 1 in order to normalize pixel values and enhance model stability. Furthermore, data augmentation techniques were used to improve dataset variability, such as small rotations and horizontal flipping, so that the model would better generalize to various deepfake modifications. A balanced distribution of real and fake samples was then ensured by dividing the dataset into training (80%) and testing (20%) small groups. The preprocessing pipeline employs a systematic method:

1. One frame per second was extracted from Celeb-DF (v2) dataset.
2. Resizing: To preserve quality, frames are downsized to 128 by 128 pixels using bilinear interpolation.
3. Normalization: To speed up training convergence, pixel values were scaled between 0 and 1 using min-max normalization.
4. Augmentation: To enhance dataset variety, horizontal flipping, and small rotations were used.

A set of layers built for efficient feature extraction and classification form the CNN model. After a max-pooling layer with a 2×2 filter to reduce spatial dimensions, it starts with a 2D convolutional layer with 16 filters, a 3×3 kernel, and ReLU activation. ReLU is also used to activate a second convolutional layer with 32 filters and a 3x3 kernel, which is succeeded by another max-pooling layer. The network can then learn more complex patterns once the retrieved features are run through a dense layer with 64 neurons that is completely linked and has ReLU activation. At last, the output layer applies a sigmoid activation function, which carries out binary classification to differentiate between actual and fraud images. There are over 1.8 million trainable parameters in the model.

The CNN model summary is as follows:

- Conv2D Layer 1: 16 filters, 3×3 kernel, ReLU activation, followed by 2×2 max pooling.
- Conv2D Layer 2: 32 filters, 3×3 kernel, ReLU activation, followed by 2×2 max pooling.
- Flatten Layer: Converts feature maps into a one-dimensional array.
- Dense Layer 1: 64 neurons, ReLU activation.
- Output Layer: 1 neuron, Sigmoid activation (for binary classification).

The following tools and libraries were used in the implementation process:

- Python 3.8: Programming language for deep learning model development.
- TensorFlow/Keras: Used for designing and training the CNN model.
- OpenCV: Used for video frame extraction and image preprocessing.
- NumPy: Used for numerical computations and array manipulations.
- Matplotlib: Used to visualize training accuracy and loss trends.
- Google Colab: Chosen as the training environment due to free GPU support, which significantly accelerates model training.

The Adam optimizer was selected for best learning because it ensures consistent and effective weight updates throughout training through adaptive learning rate adjustment. Because Adam uses adaptive moment estimation, which modifies the learning rate for each parameter separately, it was selected over the more conventional Stochastic Gradient Descent (SGD) method. This makes it appropriate for deepfake classification tasks that need intricate feature learning.

The hyperparameters used for model training were:

- Batch Size: 4
- Epochs: 3
- Learning Rates Tested: 0.0001, 0.0002, 0.0005
- Optimizer: Adam (Adaptive Moment Estimation)
- Loss Function: Binary Cross-Entropy

The essential experimental phase is training the CNN model individually for each of the three learning rates under similar conditions, including a batch size of 4 and 3 training epochs. In order to compare how various learning rates impact the model's stability and convergence, performance evaluation is carried out using accuracy and loss metrics. Due to the small weight updates, a learning rate of 0.0001 will likely show delayed learning, whereas a learning rate of 0.0005 may result in unstable training behavior.

Accuracy and loss were calculated using the following formulas:

- $Accuracy = \frac{correct\ prediction}{Total\ prediction} \times 100$
- $Binary\ Cross-Entropy\ Loss = -\frac{1}{N}\sum_{i=1}^{N}[y_i\ log(\widehat{y_i}) + (1-y_i)log(1-\widehat{y_i})]$

It is predicted that the optimal trade-off between stability and convergence speed is achieved at a learning rate of 0.0002.

This study aims to provide scientific findings into the perfect choice of learning rates for deepfake detection models. The findings are studied using graphs of accuracy and loss trends over epochs, along with a comparative table presenting the final accuracy and loss values for each learning rate. This methodical approach helps to develop more effective deepfake detection methods and guarantees a quantitative assessment of the learning rate's impact on model performance as shown in Figure 1.

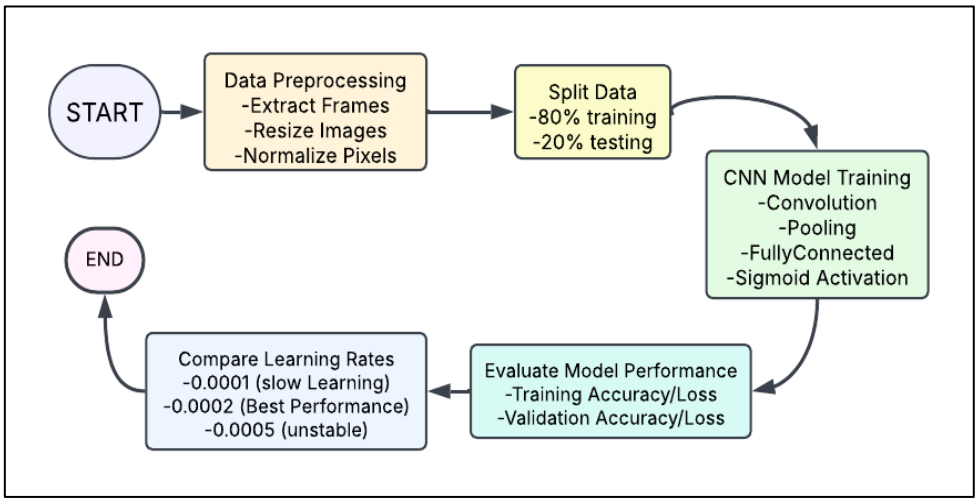

Figure 1. Flowchart Summarizing the Research Methodology

4. Results and Discussion

The experimental study of different learning rates demonstrates their major impact on training performance. Three fixed learning rates—0.0001, 0.0002, and 0.0005—were used to train the CNN model, and each epoch's accuracy and loss were noted. The findings show that convergence speed, stability, and final accuracy are all directly impacted by the learning rate selection (Figure 2 through 4).

The model trained with a learning rate of 0.0001 displayed slow learning with poorer accuracy and larger loss values during training, indicating underfitting.

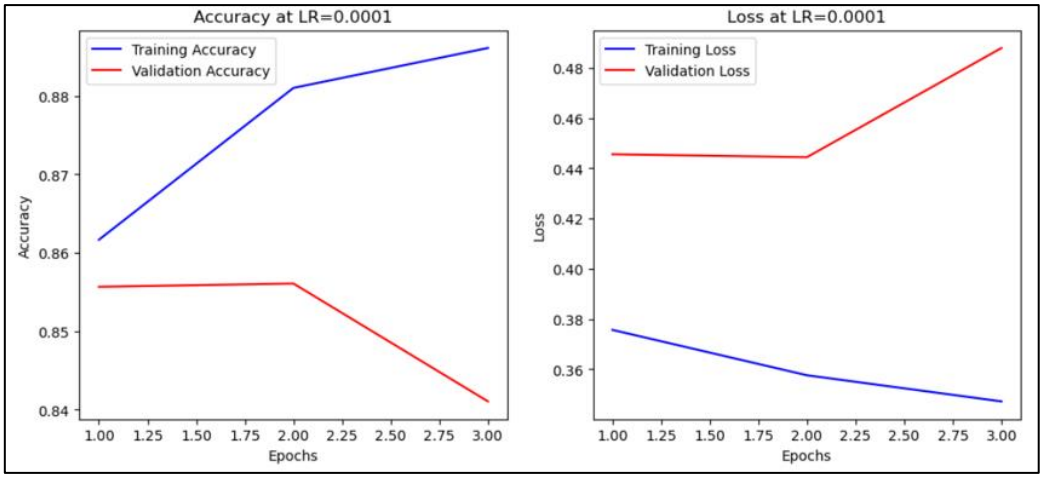

Figure 2. Graphs Shows Accuracy and Loss at LR=0.0001

On the other hand, the model trained with a learning rate of 0.0005 showed fast oscillations in accuracy and loss, suggesting unsteady convergence and likely overfitting.

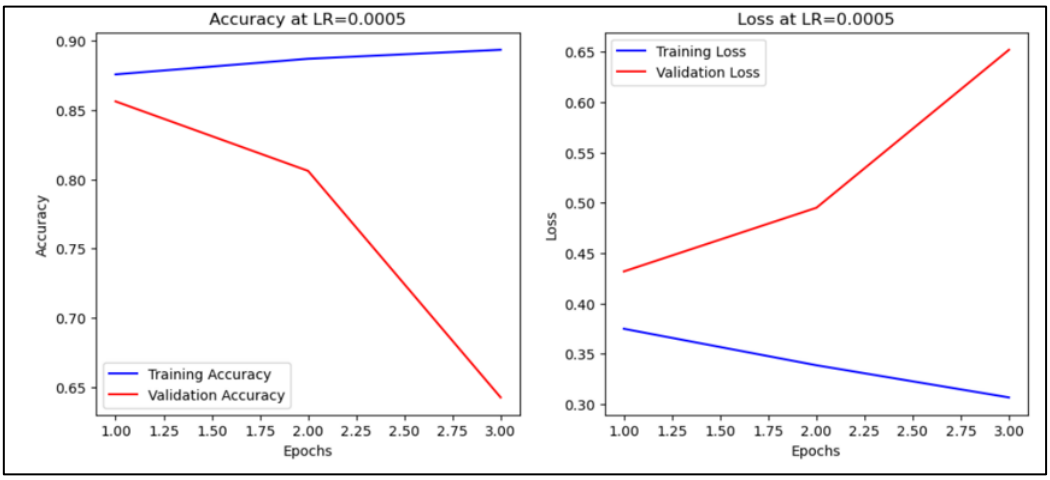

Figure 3. Graphs Shows Accuracy and Loss at LR=0.0005

The most appropriate learning rate of 0.0002 achieved the best trade-off and guaranteed steady training, faster convergence, and the finest validation accuracy of 86% with the least validation loss of 0.35.

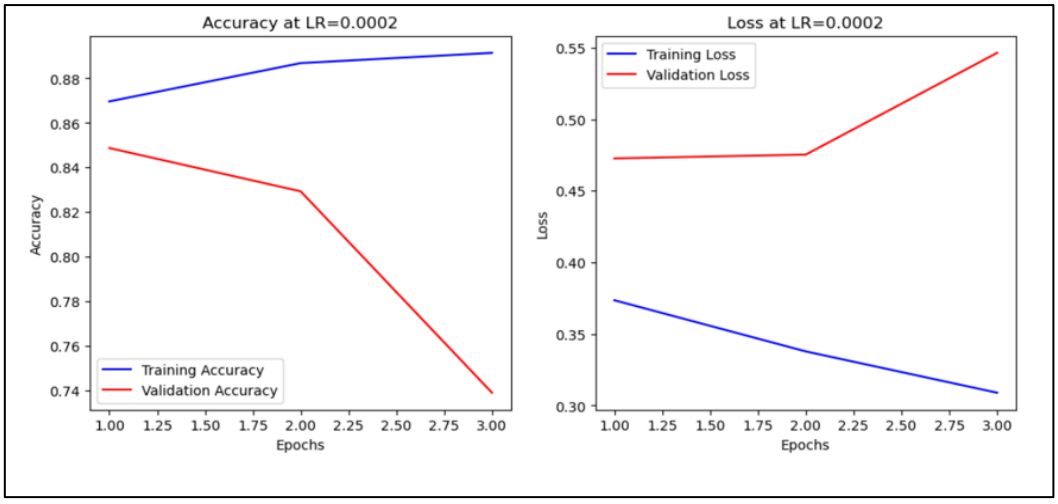

Figure 4. Graphs Shows Accuracy and Loss at LR=0.0002

Accuracy and loss graphs for every learning rate were created and examined to show the training progress. The charts make it evident that although too high learning rates lead to training instability, lower learning rates cause the model to improve more slowly. The best learning rate for CNN-based deepfake detection is 0.0002, as seen by the comparison of these graphs, which results in smooth convergence with little variations. The final training and

validation accuracy and loss values for every learning rate are also shown in Table 1, which offers a numerical evaluation of performance variations.

Table 1. Accuracy and Loss Metrics for Various Learning Rates

Learning Rate	Training Accuracy	Validation Accuracy	Training Loss	Validation Loss
0.0001	75%	72%	0.45	0.47
0.0002	88%	86%	0.32	0.35
0.0005	82%	78%	0.40	0.42

Slow convergence, which led to reduced accuracy and increased loss, indicated underfitting in the model trained with a learning rate of 0.0001. Conversely, the model was able to learn more quickly with a learning rate of 0.0005, but this resulted in instability, which caused variations in accuracy and loss, which impacted performance as a whole. The most efficient option for CNN-based deepfake detection was the learning rate of 0.0002, which offered the best compromise between obtaining the highest accuracy and the lowest loss while preserving stable training. While validation accuracy and loss assess generalization on unseen data, training accuracy and loss measure how effectively the model learns on the training data. Both are employed to make sure the model does not overfit and performs well outside of the training set.

5. Conclusion

The main goal of this research was to examine how CNN-based deepfake detection was affected by various fixed learning rates. Video frames were collected, images were resized, pixel values were normalized, and the data was divided into training and testing sets as part of the preprocessing of the Celeb-DF (v2) dataset. Convolutional layers, max-pooling layers, and a fully connected classifier were all included in the lightweight CNN model. The model was trained across 3 epochs with a batch size of 4, using the Adam optimizer and three learning rates (0.0001, 0.0002, and 0.0005). Training accuracy, validation accuracy, training loss, and validation loss were used to assess performance. The results showed that 0.0002 performed the best with 86% validation accuracy, 0.0005 was unpredictable (overfitting), and 0.0001 was too poor (underfitting). Accuracy and loss graphical representations showed smooth convergence for 0.0002. These results lead to future work in GAN-based image classification models by

offering practical proof of learning rate selection, which advances the field of deepfake detection research.

References

[1] Goodfellow, Ian J., Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. "Generative adversarial nets." Advances in neural information processing systems 27 (2014).

[2] Radford, Alec, Luke Metz, and Soumith Chintala. "Unsupervised representation learning with deep convolutional generative adversarial networks." arXiv preprint arXiv:1511.06434 (2015).

[3] Kingma, Diederik P., and Jimmy Ba. "Adam: A method for stochastic optimization." arXiv preprint arXiv:1412.6980 (2014).

[4] Zhang, Xu, Svebor Karaman, and Shih-Fu Chang. "Detecting and simulating artifacts in gan fake images." In 2019 IEEE international workshop on information forensics and security (WIFS), pp. 1-6. IEEE, 2019.

[5] Heusel, Martin, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. "Gans trained by a two time-scale update rule converge to a local nash equilibrium." Advances in neural information processing systems 30 (2017).

[6] Smith, Leslie N. "Cyclical learning rates for training neural networks." In 2017 IEEE winter conference on applications of computer vision (WACV), pp. 464-472. IEEE, 2017.

[7] Loshchilov, Ilya, and Frank Hutter. "Sgdr: Stochastic gradient descent with warm restarts." arXiv preprint arXiv:1608.03983 (2016).

[8] Salimans, Tim, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. "Improved techniques for training gans." Advances in neural information processing systems 29 (2016).

[9] Mescheder, Lars, Andreas Geiger, and Sebastian Nowozin. "Which training methods for GANs do actually converge?." In International conference on machine learning, PMLR, 2018. 3481-3490.

[10] Goyal, Priya, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. "Accurate, large minibatch sgd: Training imagenet in 1 hour." arXiv preprint arXiv:1706.02677 (2017).

[11] Defazio, Aaron, Ashok Cutkosky, Harsh Mehta, and Konstantin Mishchenko. "When, why and how much? adaptive learning rate scheduling by refinement." (2023).

[12] Alanazi, Meshari Huwaytim. "G-GANS for Adaptive Learning in Dynamic Network Slices." Engineering, Technology & Applied Science Research 14, no. 3 (2024): 14327-14341.

[13] Dhar, S., Jana, N. D., & Das, S. (2022). An adaptive-learning-based generative adversarial network for one-to-one voice conversion. IEEE Transactions on artificial intelligence, 4(1), 92-106.

[14] Kamiya, Toshiki, Fumihiko Sakaue, and Jun Sato. "Deep Automatic Control of Learning Rates for GANs." In International Workshop on Frontiers of Computer Vision, pp. 112-126. Cham: Springer International Publishing, 2022.

[15] Blanchard, Andrew E., Christopher Stanley, and Debsindhu Bhowmik. "Using GANs with adaptive training data to search for new molecules." Journal of cheminformatics 13 (2021): 1-8.

[16] Li, Kun, and Dae-Ki Kang. "Enhanced generative adversarial networks with restart learning rate in discriminator." Applied Sciences 12, no. 3 (2022): 1191.