

# Federated Optimization for Digital Forensics: Image Forgery Detection Using FedYogi with CNN and ResNet50

Narayani<sup>1</sup>, Charu Gupta<sup>2</sup>, and R. K. Singh<sup>3</sup>

<sup>1</sup> Department of Information Technology,  
Indira Gandhi Delhi Technical University for Women, Delhi, India  
[narayani24082001@gmail.com](mailto:narayani24082001@gmail.com)

<sup>2</sup> School of Open Learning, Department of DDCE,  
University of Delhi, Delhi, India  
[charu.gupta@sol.du.ac.in](mailto:charu.gupta@sol.du.ac.in)

<sup>3</sup> Department of Information Technology,  
Indira Gandhi Delhi Technical University for Women, Delhi, India  
[rksingh@igdtuw.ac.in](mailto:rksingh@igdtuw.ac.in)

**Abstract.** Detection of modified and manipulated images is a challenge in digital forensics as the images are detected using centralized detection models. Forensic datasets require sharing across organizations for aggregating the trained models. However, the aggregation of sensitive data for image forensics increases privacy and security risks. Federated Learning (FL) provides a promising solution by aggregating the trained model parameters in a collaborative manner without sharing the raw data. In this paper, we utilized the collaborative aggregation of FedYogi algorithm to improve the stability and efficiency of deep learning-based image forgery detection under non-IID data distributions. The algorithm FedYogi is an enhanced optimizer that incorporates adaptive learning rates to accommodate variations in client data. Experiments were conducted using CNN and ResNet-50 architectures on the CASIA 1.0 image forgery dataset. The results showed that model updates of ResNet-50 were aggregated using FedYogi provided improved baseline methods in terms of accuracy, precision, recall, and F1-score. The proposed approach effectively extracts image-level features and yields more reliable performance in distributed environments. The experimental results indicate that FedYogi not only preserves data privacy by keeping data local and sharing only model updates, but also enhances overall model performance. The experimental results on image forensic data set showed that federated learning algorithm -FedYogi- is a viable and privacy-preserving framework for detecting forged and manipulated images in digital forensic investigations.

**Keywords:** Federated Learning, Image Forgery Detection, FedYogi, Digital Forensics, ResNet50, CNN, Privacy Preservation

## 1 Introduction

The rapid growth of digital content and proliferation of advanced editing tools facilitates forging the visual evidence making digital forensics domain very challenging. Even a single manipulated image is sufficient to twist the facts,

breach evidence integrity, and mislead investigation or legal processes. Therefore, resilient methods for image forgery detection with privacy considerations have become need of modern forensic investigations.

Traditional AI-based image forensics depends on centralized learning that requires training the sensitive forensic images collected on a common centralized server. Such approach, though effective in model accuracy, raises substantial concerns with respect to privacy, confidentiality, and regulatory compliance. Moreover, aggregation of sensitive evidence images at a central server poses risks such as unauthorized access to the data, leakage of data, and full exposure of raw data in case server gets compromised. Such limitations make server based learning unsuitable for forensic environments where protection of data is of utmost importance.

Federated Learning provides an alternative approach by training the model only at the client’s side, where data resides. Rather than transferring raw forensic images, only model updates are communicated to a central server, which then aggregates them into a global model. FedAvg is the classical aggregation technique used in most FL workflows; however, its performance becomes unstable when client data are highly non-Identical and Independent Distributed (non-IID). In digital forensics, Non-IID data is a common scenario where image characteristics, tampering patterns, and manipulation artifacts vary significantly across sources. A single large or inconsistent update can cause oscillations, reduced convergence speed, or degraded accuracy. To address these challenges, Reddi et al. (2021) introduced adaptive optimization methods such as FedYogi algorithm, which integrates momentum-based stabilization and variance correction to handle heterogeneous client updates more effectively. FedYogi algorithm mitigates the abrupt gradient shifts caused by non-IID distributions and provides smoother, more reliable convergence in federated settings. In this work, we have used the FedYogi optimizer algorithm to aggregate the results of two widely used image classification algorithms namely CNN (Convolutional Neural Network) and ResNet50 to obtain privacy-preserving image forgery detection framework under distributed environments. Combining FedYogi with CNN and ResNet50 augments FedYogi’s stability in the case of Non-IID data to support deep feature extractors that rely on consistent gradient behaviour. CNNs and ResNet50 are well suited at capturing fine-grained forensic cues like edges, textures, and manipulation artifacts, and FedYogi’s adaptive variance control ensures that subtle features are learned reliably across clients with heterogeneous data. Working in tandem, they provide a comprehensive , robust and privacy-preserving approach that offers better accuracy and generalization in real-world image forgery detection.

This work underlines the potential of adaptive federated learning for strengthening digital forensic investigations with strict privacy guarantees and, as such, contributes to the development of secure, scalable, future-ready forensic AI systems.

## 2 Literature Review

Image forgery detection has evolved enormously across both the centralized and distributed AI paradigms. Early CNN-based approaches showed very good performance by learning spatial and texture-level manipulation traces. However, their dependence on centralized data aggregation limits the applicability of such techniques in forensic environments where sharing sensitive evidence is not allowed. Initial image forensic analyses established the potential risks of even subtle manipulations. Farid (2019) presented an inconsistency check at the pixel level, whereas Stamm et al (2019) and Piva et al (2013) described classical methods for detecting copy-move, splicing, and resampling operations. Since the introduction of deep learning, manipulation-specific architectures like the constrained CNN proposed by Bayar and Stamm (2018) and the ManTra-Net from Wu et al (2019) have made extraction of tampering artifacts fully automatic. On the feature extraction side, this process was advanced one step further with residual networks such as ResNet50, allowing for deeper hierarchies without the problem of vanishing gradients to occur.

Meanwhile, the requirement for analysis with privacy has driven growing interest in Federated Learning, which enables decentralized model training without requiring the sharing of raw images. McMahan et al. (2017) proposed FedAvg, which showed effective learning with communication efficiency from decentralized data. However, forensic datasets are inherently Non-IID due to device, manipulation style, and acquisition condition variations, which is the root of client drift and unstable convergence reported in widespread surveys by Kairouz et al. (2021) and Yang et al. (2019). To address such issues, Reddi et al. (2021) introduced adaptive federated optimizers, including FedAdam and FedYogi, which improved the stability of convergence using variance controlled moment estimation. Qin et al. (2020) further pointed out that Non-IID image distributions significantly degrade the performance of the federated model. Therefore, there is an urgent need for optimizers capable of handling conflicting directions of gradients. Very recently, Gautam et al. (2025) proposed FFDL, a feature-fusion-based federated deep learning framework for forged face detection, which showed that FL can successfully support complex forensic models while preserving privacy. Still, such attempts remain few, and explorations of adaptive FL optimizer integrations within deep forensic architectures remain scant.

Despite these advances, the incorporation of adaptive FL optimizers with deep forensic architectures is still bound due to the fact that most federated studies are based either on medical or mobile contexts.

Therefore, combining FL with FedYogi, an adaptive optimizer with deep learning models like CNN and ResNet50 provides a promising solution for developing robust, stable, and privacy-preserving image forgery detection framework for forensic analysis.

## 3 Background Study

In this section, the core ideas guiding the proposed approach for distributed image forensic analysis are briefly discussed. Federated Learning(FL), allows dif-

ferent clients to train a shared model without ever handing over their raw forensic images—an essential requirement when dealing with sensitive or confidential material. The optimizer algorithm used in Federated Learning FedYogi proposed by Reddi et al. (2021) is implemented for non-IID datasets through more balanced and steady updates. The images are trained at client machines using CNN and ResNet50 architectures. The proposed approach focuses on residual connections that improve feature learning for reliable image forgery detection. Federated Learning (FL) enables federated model training without sharing raw data kept on local devices (Mc Mahan et al., 2017) (Yang et al. 2019). Each client  $i$  locally trains a model  $w_i^t$  using its private dataset, it transmits the update to global model/ aggregator. The aggregator is updated iteratively using the equation:

$$w^{t+1} = w^t - \eta \cdot \frac{1}{K} \sum_{i=1}^K \nabla L_i(w_i^t)$$

where  $\eta$  is the learning rate, and  $L_i$  represents the local loss function for client  $i$ . This federated averaging strategy reduce the risks of data leakage and allow the aggregation of update at aggregator from Non IID dataset, which makes it better option for digital forensics. The FedYogi optimizer extends adaptive gradient techniques for federated environments with non-IID data. It tunes the learning rate based on adaptively of clients. FedYogi maintains first and second moment estimates as:

$$\begin{aligned} m_t &= \beta_1 m_{t-1} + (1 - \beta_1) g_t \\ v_t &= v_{t-1} - (1 - \beta_2) \text{sign}(v_{t-1} - g_t^2) \\ w_{t+1} &= w_t - \eta \frac{m_t}{\sqrt{v_t + \epsilon}} \end{aligned}$$

Here:

- $g_t$ : aggregated gradient from clients involved
- $\beta_1, \beta_2$ : momentum coefficients controlling smoothing of updates
- $\eta$ : learning rate of global model
- $\epsilon$ : numerical stability constant

FedYogi’s adaptive variance correction provides smoother convergence and improved generalization in federated settings, demonstrating high robustness on non-IID forensic data. FedYogi (Reddi et al, 2021) is more reliable and finds a solid solution faster, particularly when client data is non-IID by utilizing variance correction (balancing out the updates from all the various clients) and momentum (making slow, consistent progress). FedYogi optimiser algorithm is used in the experiments conducted to develop image forgery detector because of adaptive variance correction. CNN is a basic image based classification model. It trains the model by extracting basic features like lines and curves then shapes using convolution layers but it has only a few layers. CNN model layers have been expanded for deep image analysis using residual connection to reduce vanishing gradients in the deep neural network that introduced Resnet50 by He et al. (2016) it improved performance. This method introduced residual mapping and transfer learning features improved image forgery detection.

## 4 Methodology

In the experiments conducted, performance of Convolutional Neural Networks (CNN) and ResNet50 in both centralized and federated learning (FL) frameworks is evaluated by tuning the hyperparameters. We performed an experiment using the CASIA 1.0 dataset and applied CNN and Resnet50 on the client side. The training parameters are then aggregated on a global server using the fedyogi algorithm. We did a comparative analysis on image forgery detection (Authentic and Modified/Tampered) using Centralized and Federated learning on CNN and Resnet50 Model. We conducted these experiments on the **CASIA 1.0** dataset Dong et al. (2013), and we evaluated our model using Accuracy, Precision, Recall, and F1-Score.

### 4.1 Forensic Dataset: CASIA 1.0

We used CASIA 1.0 dataset serves as a fundamental baseline for training image forgery detection models. CASIA Image Tampering Detection Evaluation Dataset contains 800 authentic images and 921 tampered images. We trained our model more effectively in our experiment by using the FL algorithm Fedyogi, which operates steadily on heterogeneous data with non-IID client distribution. This split ensures clear separation of dataset among clients and their training, testing and evaluation dataset to support efficiency and accuracy.

### 4.2 Hyperparameter Tuning and Model Stability

This hyperparameter selection is critical for making the model stable and preventing overreaction (Smith et al 2018) . The hyperparameters that we adjusted in the experiments are:  $g_t, \beta_1, \beta_2, \eta, \epsilon$ , learning rate, batch size, epochs, client, and round. Number of clients is total no of clients that will be part of training global model. A round is a complete cycle of communication and computation between a central server and participating client devices to update the global model. Epochs per client is total no time client train on its own images in a single round. Batch size is number of images that will train at a time. If learning rate of model is Too fast, it gets confused; if it's too slow, it takes forever so the learning rate of model should be optimal in our experiment it is 0.001. The FedYogi optimizer also had its own set of parameters how it updates the main model. These are the more technical settings, like  $\beta_1, \beta_2, \eta$  (its own learning rate), and  $\tau$  (a smoothing term). The hyperparemters considered for initial set up of the experiments is given in Table 1. The hyperparemters considered for initial set up of the experiments is given in Table 1. The configuration for taring the model at client machine using CNN is config 3 and using the ResNet-50 is config4 as described in Table 2.

## 5 Results and Discussion

The results of the experiments conducted are discussed in this section. We compared two main setups: Federated Learning (using the FedYogi optimizer) and Centralized Learning (the traditional, all-data-in-one-place method) We tested both of these using a simple CNN and the more powerful ResNet50 model. All tests were run on the CASIA 1.0 forgery image dataset , and we evaluated their performance using accuracy, precision, recall, and F1-score.

**Table 1.** Hyperparameter Settings for CNN and ResNet50 Models with FedYogi Optimizer

Parameter	Config 1 (CNN)	Config 2 (CNN)	Config 3 (CNN)	Config 4 (ResNet50)
Clients	4	10	10	10
Clients per round	2	2	2	2
Rounds	15	15	15	15
Epochs per client	1	1	5	3
Batch size	32	64	64	32
Learning rate	0.01	0.01	0.001	0.001
<b>FedYogi Optimizer Parameters</b>				
$\beta_1$			0.9	
$\beta_2$			0.99	
$\eta$ (Learning Rate)			0.01	
$\tau$ (Smoothing Term)			$1 \times 10^{-3}$	

**Table 2.** Centralized Training Configuration for CNN (Config 3) and ResNet50 (Config 4)

Parameter	Config 3 (CNN)	Config 4 (ResNet50)
Training Type	Centralized	Centralized
Dataset	CASIA 1.0	CASIA 1.0
Total Images	1721	1721
Train/Test/Eval Split	1290/344/87	1290/344/87
11.1 Epochs	15	15
Batch Size	64	32
Learning Rate	0.001	0.001
Optimizer	SGD (momentum = 0.9)	SGD (momentum = 0.9)
Input Image Size	128×128	224×224
Classes	2 (Authentic/Tampered)	2 (Authentic/Tampered)

### 5.1 Centralized Learning Results

At first the images are trained at server using the CNN model and ResNet 50 by keeping all the data into one big cluster on a single server. We did this to get a "baseline" score—basically, the "high score" to beat—before we tried our new, privacy-focused federated learning method. The epoch wise evaluation metrics of CNN model are shown in Table 3 and ResNet50 model are shown in Table 4. The results show that the CNN model trains the model with an accuracy of 57.27% while ResNet50 (Config 4) trained model with accuracy of (58.43%) . Training the model with ResNet-50 (residual connections) achieves higher stability as it learns from complex features of the tampered images (He et al. 2016).The centralized training model for detection of forged images for image forensics are limited by minimum data privacy.

### 5.2 Federated Learning using FedYogi Optimizer

The performance of the FedYogi optimizer applied to CNN and ResNet50 architectures under various parameters is assessed in this section. FedYogi includes

**Table 3.** Epoch-wise Centralized Training Results for Simple CNN (Config 3)

Epoch	Accuracy	Precision	Recall	F1-Score
1	0.5436	0.7705	0.5063	0.3635
3	0.5727	0.5763	0.5762	0.5727
5	0.5640	0.5608	0.5606	0.5606
6	0.5698	0.5670	0.5671	0.5619
7	0.5727	0.5775	0.5771	0.5726
10	0.5058	0.5427	0.5282	0.4712
12	0.5610	0.5587	0.5587	0.5587
15	0.5291	0.5075	0.5043	0.4561

**Table 4.** Performance of Centralized ResNet50 Model

Epoch	Accuracy	Precision	Recall	F1 Score
1	0.5349	0.5452	0.5317	0.4956
2	0.5465	0.5574	0.5437	0.5163
3	0.5640	0.5699	0.5620	0.5503
4	0.5727	0.5746	0.5715	0.5677
5	0.5610	0.5612	0.5603	0.5591
6	<b>0.5843</b>	<b>0.5867</b>	<b>0.5831</b>	<b>0.5794</b>
7	0.5610	0.5659	0.5592	0.5487
8	0.5523	0.5553	0.5506	0.5419
9	0.5494	0.5505	0.5482	0.5436
10	0.5552	0.5560	0.5542	0.5511
11	0.5407	0.5422	0.5391	0.5316
12	0.5203	0.5200	0.5199	0.5194
13	0.5378	0.5376	0.5373	0.5367
14	0.5349	0.5349	0.5339	0.5313
15	0.5174	0.5174	0.5174	0.5174

variance correction to stabilize updates on non-IID client data and adaptively modifies local learning rates to extend Adam optimization to federated settings. To evaluate convergence, stability, and generalization behavior, experiments were conducted involving different number of clients, learning rates, and epochs. In table 5, due to the low data diversity and small number of participating clients, performance stayed almost same. FedYogi’s adaptive moment updates were less beneficial when there were fewer clients since the global model received less gradient variance. FedYogi’s small-scale setup hindered its capacity to take use of adaptive optimization, which led to slower convergence and negligible accuracy gains because FedYogi depends on aggregated gradient variance to modify learning rates.

In table 6, The results showed typical FedYogi behavior, with an initial unstable phase due to momentum building followed by eventual stabilization, even though accuracy fluctuated. Although convergence is still slower because local

**Table 5.** Performance Metrics Across Rounds (CNN + FedYogi, Config 1)

Round	Accuracy	Precision	F1-Score
1–15	0.5233	0.2616	0.3435

**Table 6.** Performance Metrics Across Rounds (CNN + FedYogi, Config 2)**Configuration 2 (CNN, 10 Clients, LR = 0.01)**

Round	Accuracy	Precision	F1-Score
1	0.5523	0.2762	0.3558
2	0.4826	0.5254	0.4445
3–15	0.4477	0.2238	0.3092

data distributions fluctuate, the optimizer’s adaptive moment estimation helps reduce oscillations across rounds.

**Table 7.** Performance Metrics Across Rounds (CNN + FedYogi, Config 3)

Round	Accuracy	Precision	F1-Score
1	0.4913	0.2456	0.3294
5	0.5029	0.5086	0.4781
6	0.5349	0.5348	0.5271
10–15	0.5087	0.2544	0.3372

**Configuration 3 (CNN, 10 Clients, LR = 0.001, 5 Epochs)** In table 7, the results showed that FedYogi demonstrated its adaptive optimization capacity more successfully when it had more clients and a moderate learning rate. Variance estimates led to erratic accuracy in the first several rounds, however adaptive momentum correction stabilized client updates after Round 5. Around Round 6, accuracy and F1-score reached their peak; performance then plateaued, showing convergence. Reduced gradient diversity and non-IID client data imbalance cause minor oscillations after this point.

**Configuration 4 (ResNet50, 10 Clients, LR = 0.001, 3 Epochs)** In table 8, the comparison of CNN-based topologies, and the ResNet50 + FedYogi optimiser are presented. Better feature reuse and gradient propagation were made possible by the deeper residual architecture, which made it possible for the model to successfully learn intricate tampering patterns. FedYogi achieved a peak accuracy of 55.52% at Round 12 and maintained constant F1-scores after using variance correction to further reduce gradient noise from non-IID consumers.

**Table 8.** Round-wise Performance Metrics (ResNet50 + FedYogi, Config 4)

Round	Accuracy	Precision	Recall	F1-Score
1	0.5116	0.5061	0.5060	0.5056
5	0.5087	0.4979	0.4981	0.4938
10	0.5233	0.5099	0.5086	0.4988
12	<b>0.5552</b>	<b>0.5463</b>	<b>0.5382</b>	<b>0.5248</b>
15	0.5465	0.5347	0.5249	0.4989

**Table 9.** Comparison between centralized and federated (FedYogi with CNN and ResNet50) configurations

Model	Acc.	Prec.	Rec.	F1	Stability	Privacy	Complexity
CNN (Config 3)	0.5727	0.5775	0.5771	0.5726	M	×	Low
ResNet50 (Config 4)	<b>0.5843</b>	<b>0.5867</b>	<b>0.5831</b>	<b>0.5794</b>	H	×	High
CNN (FedYogi-C1)	0.5233	0.2616	–	0.3435	L	✓	Low
CNN (FedYogi-C2)	0.5523	0.2762	–	0.3558	M	✓	Low
CNN (FedYogi-C3)	0.5349	0.5348	0.5271	0.5271	M-Stable	✓	Medium
ResNet50 (FedYogi-C4)	<b>0.5552</b>	<b>0.5463</b>	<b>0.5382</b>	<b>0.5248</b>	H	✓	High

### 5.3 Comparative Summary

The experimental results obtained by training the model by CNN and Resnet50 and aggregating it with FedYogi optimizer are shown in Table 9. FedYogi maintained competitive performance in spite of distributed and non-IID data situations, while centralized models achieved slightly greater accuracy due to total data availability. Our experiment show that the algorithm fedYogi increased the consistency in convergence and less oscillations when used for collaborative learning at server side and ResNet50 for training the model at client machines.

Federated learning maintained the privacy of data by using only the model updates received from clients without sharing the raw data, while in centralized learning data is collected in raw form at a central server. ResNet50 with FedYogi achieved the best balance between accuracy (55.52%), F1-score (0.5248), stability, and privacy, making it suitable for forensic picture forgery detection that preserves privacy.

## 6 Conclusion and Future Work

We performed experiments on centralized CNN and Resnet50 and the results show that Resnet50 performed better than CNN for detecting the forged images. Once the dataset is trained at local machine, the model parameters are aggregated using Federated optimizer FedYogi algorithm. The experimental results showed that the detecting forged images using ResNet 50 and FedYogi federated algorithm performed better with 55.52% accuracy. The federated algorithm FedYogi helped in maintaining the data privacy by aggregating the model pa-

rameters without sharing the raw data from the clients. The Federated algorithm FedYogi is suitable with non-IID data, allowing us to work on real-life scenarios where image data is non identical and variedly distributed at different client machines. In further studies we will perform experiments on different combinations of federated algorithms and ResNet50.

## Acknowledgment

The authors gratefully acknowledge the support of the Information Security Education and Awareness (ISEA) program and the Department of Information Technology, Indira Gandhi Delhi Technical University for Women (IGDTUW), for their continuous academic and technical guidance.

## References

- Bayar, B., Stamm, M. C. (2018). Constrained convolutional neural networks: A new approach towards general purpose image manipulation detection. *IEEE Transactions on Information Forensics and Security*, 13(11), 2691-2706.
- Dong, J., Wang, W., Tan, T.: CASIA Image Tampering Detection Evaluation Database. In: *IEEE ChinaSIP*, 2013.
- Farid, H. (2019). Image forensics. *Annual Review of Vision Science*, 5(1), 549-573
- Gautam, V., Singh, A., Kaur, G., Malik, M., Pawar, A., Singh, K. K., Askar, S. S., Abouhawwash, M.: FFDL: Feature Fusion-Based Deep Learning Method Utilizing Federated Learning for Forged Face Detection. *IEEE Access*, 2025.
- He, K., Zhang, X., Ren, S., Sun, J.: Deep Residual Learning for Image Recognition. In: *Proc. CVPR*, 2016.
- Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., et al.: Advances and Open Problems in Federated Learning. *Foundations and Trends® in Machine Learning*, 14(1–2), 1–210, 2021.
- McMahan, H. B., Moore, E., Ramage, D., Hampson, S., Arcas, B. A.: Communication-Efficient Learning of Deep Networks from Decentralized Data. In: *Proc. AISTATS*, 2017.
- Mayer, O., Stamm, M. C. (2019). Forensic similarity for digital images. *IEEE Transactions on Information Forensics and Security*, 15, 1331-1346.
- Piva, A. (2013). An overview on image forensics. *International Scholarly Research Notices*, 2013(1), 496701.
- Qin, Z., Li, G. Y., Ye, H.: Federated Learning and Wireless Communications. *IEEE Wireless Communications*, 28(5), 134–140, 2021.
- Reddi, S. J., Charles, Z., Zaheer, M., Garrett, Z., Rush, K., Kumar, S., McMahan, H. B.: Adaptive Federated Optimization. In: *International Conference on Learning Representations (ICLR)*, 2021.
- Smith, L. N.: A Disciplined Approach to Neural Network Hyper-Parameters: Part 1 – Learning Rate, Batch Size, Momentum, and Weight Decay. *arXiv preprint arXiv:1803.09820*, 2018.
- Wu, Y., AbdAlmageed, W., Natarajan, P. (2019). Mantra-net: Manipulation tracing network for detection and localization of image forgeries with anomalous features. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 9543-9552).
- Yang, Q., Liu, Y., Chen, T., Tong, Y.: Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2), 2019.