

Collaborative Colorectal Cancer Classification on Highly Class Imbalanced Data Setting via Federated Neural Style Transfer Based Data Augmentation



Mehmet Nergiz

Department of Computer Engineering, Dicle University, Diyarbakır 21280, Turkey

Corresponding Author Email: mnergiz@dicle.edu.tr

<https://doi.org/10.18280/ts.390620>

ABSTRACT

Received: 23 October 2022

Accepted: 10 December 2022

Keywords:

federated learning, neural style transfer, convolutional neural network, colorectal cancer, computational pathology

The deep learning algorithms achieved promising results in the computational pathology in recent decade but the high data demand of the deep learning algorithms get stuck in the multi-institutional data collaborations. The federated learning is a novel concept, which proposes to train the models of the different sites collaboratively via an orchestrating server without leaking private data. However, the imbalanced data distributions are challenging for federated learning and result in performance decrease and destabilization. In this study, the federated version of the neural style transfer algorithm, which was offered by Gatys et al. is proposed as a data augmentation method on the highly class imbalanced configuration of Chaoyang colorectal cancer imaging dataset. The proposed method works by firstly selecting characteristic style images and then generating the gram style matrices on the local sites and then transferring them to the other imbalanced sites by not leaking any private data. The proposed method contributed the ACC, F1 Score and AUC results of pure FL by 22.07%, 42.51% and 9.65% using only 20 images for content and 5 images for style. Additionally, the experiments having different content and style numbers achieved the satisfactory and consisting results.

1. INTRODUCTION

Colorectal cancer (CRC) is the fourth prevailing reason of death related to cancer and third most prevalent cancer type [1]. The Western countries suffer from CRC relatively more than the other countries and the current incidence of CRC is approximately 5% and tending to increase every year [1]. Because of the high frequency of CRC, the colonoscopy is a widespread imaging technique and the task of colorectal cancer classification is a heavy burden over pathologists [2]. Another important challenge of CRC is that variance among the pathologists is high and the colorectal polyps which could not be diagnosed and treated may quickly turn into CRC [3].

The manual analysis process of histopathological tissue samples has been shifted to the era of the digital pathology with the innovation of the digital scanners enabling the whole-slide images (WSI) which capture the image from glass slides as a whole [4]. More and more, the digitally acquired tissue images and the rise of the machine learning (ML) and deep learning (DL) algorithms give birth to a novel field called computational pathology (CP) [4-6]. In the ML field, as one of the most vibrant fields in academia and the industry, a new sub-branch called federated learning (FL) has been initiated by Google researchers in 2016 [7]. The FL is offered as a new concept of ML by providing a collaborative technology for the institutions having their own local private data [7]. As a well-known fact, the DL models need too many well-annotated diverse data to reach a robust and high performance level but the commercial and privacy restrictions prohibit a relax inter-institutional data sharing process [7]. In order to get rid of this

obstacle, the FL comes up with a strategy in which the local data holders do not share their own private data but only their locally trained models via a dedicated model aggregator server. Thus, applying FL strategies to the CP field is relatively a new and promising field, which may give rise to a burst of inter-institutional collaboration about various histopathological researches. Different medical institutions have different kind and amount of data and the best performing as well as hopefully clinically applicable ML models are only likely to be obtained by bringing the power of data together. The FL is proposed also as a distributed architecture solution to the storage and computational load for processing the big imaging data [7].

The FL algorithms show different performances under different distribution scenarios. The imbalanced data case in which each collaborator has diverse non-independent and non-identical (NonIID) distribution is a hot research area in FL field [8]. In the FL scope, the imbalanced data challenge is handled as local and global imbalance of data and accordingly the dataset used in this study is configured and analyzed for a local and highly class imbalanced case. Even if the FL methods gives promising results for most of the balanced cases, the standard FL algorithms do not converge to a monotonically improving, successful and generalizable aggregated model for the local imbalanced cases [8]. The local models in each collaborator client over fit to such diverse feature spaces and thus classical model aggregation causes the catastrophic forgetting for the next local training round [8]. There are some recently published or preprinted researches to overcome the imbalance data issue in FL via miscellaneous algorithms.

2. RELATED WORKS

2.1 FL applications on CP

Lu et al. [9] applied FL based weakly-supervised attention and multiple instance learning methods on the WSIs of breast invasive carcinoma and renal cell carcinoma dealt on four collaborator clients. Their findings showed that the weakly supervised DL methods can be applied by FL architectures and accurate results can be obtained. The case study of Adnan et al. [10] was about applying FL on the independent and identical (IID) and Non-IID data obtained from the Cancer Genome Atlas (TCGA) imaging dataset.

Andreux and Terrail [11] offered a genuine DL layer named as local-statistic batch normalization (BN) which trains the DL models collaboratively but ends up with collaborator-specific models. They also prevent data leakage by not sharing the statistics of the collaborator-specific layer activations and benchmarked the model on Camelyon 16 and Camelyon 17 breast cancer datasets [11]. Gunesli et al. [12] proposed FedDropoutAvg which is a novel FL model aggregation method inspired from the well-known DL regularization DropOut method and benchmarked this proposed method on the CRC dataset of TCGA portal. It is reported that the FedDropOutAvg method reached to the closest level of centralized training compared to the FedProx and FedAvg [12].

Wagner et al. [13] proposed a generative BottleGAN supported with an unsupervised FL training architecture to standardize the staining styles of the histopathological images of the different collaborators and benchmarked the model on a epithelial tissue H&E stained prostatectomy imaging dataset. Chen et al. [14] preprinted the FL version of a genuine domain generalization method which exchanges the styles of collaborators.

2.2 FL applications on imbalanced data

Sarkar et al. [15] proposed the FL version of the focal loss which adaptively down-weights the cross-entropy loss assigned for easily classified images of the sampled collaborators from the last round. Ran et al. [16] proposed the Dynamic Margin for FL method for imbalanced datasets by enlarging the margin of the local classifier by adding a dynamic term in the favor of the minority class. Zhang et al. [17] offered FedSens as a FL framework to solve the challenges of imbalanced data and resource restrictions of the collaborating edge devices. FedSense operates via an extrinsic-intrinsic deep reinforcement-learning model, which decides for the best timing for the global and local updates in order to obtain maximal accuracy [17].

Duan et al. [18] proposed the method called Astraea which is also defined as a self-balancing FL benchmarked on the globally imbalanced mobile data. The Astraea offers to solve the global imbalance problem by z-score-based data augmentation and down sampling as well as a mediator for rescheduling the attendance of collaborators by exploiting the Kullback–Leibler divergence of their distributions [18]. Shuai et al. [19] offered BalanceFL method which aims to solve the local and global data imbalance by updating the locally trained models as it is trained on uniform data. In the BalanceFL study, the local imbalance problems are defined as causing from two reasons such as data amount and class missing, which are solved respectively by inter-class balancing and knowledge inheritance techniques [19].

Wu et al. [20] offered FedRare which is a novel method for class imbalance problem especially for rare disease cases. FedRare firstly trains each local data via supervised contrastive learning and then extracts the separable latent features from the obtained local models and sends to the server [20]. The server selects the strongest latent features and send them back to the collaborator clients [20]. Finally, each collaborator is trained together using the inter-client contrastive loss [20].

The main contributions of this study are listed as follows:

- 1- The FL version of the Neural Style Transfer (NST) algorithm is proposed without violating data privacy.
- 2- The proposed Federated NST (FNST) algorithm achieves promising performance contributions on the highly class imbalanced settings of Chaoyang CRC imaging dataset.
- 3- The effect of FNST algorithm is measured on different number of content and style image scenarios.
- 4- The FNST algorithm smoothens the fluctuating F1 score curve of training on the class imbalanced the dataset.

3. MATERIAL

Zhu et al. [21] is a publicly available CRC imaging dataset curated in Chaoyang hospital, which contains 842 serrated, 664 adenoma, 1404 adenocarcinoma and 1111 normal histopathological images for training phase as well as 321 serrated, 273 adenoma, 840 adenocarcinoma and 705 normal histopathological images for testing phase. The dataset is prepared as patches having 512x512 pixels and then 3 specialist pathologists annotate each patch based on 4 different classes such as normal, serrated, adenoma and adenocarcinoma. The sample selection criteria for training set does not look for a consensus among the pathologists and a random label is selected among the decision of one of three pathologists. More and more, this dataset is named as a noisy dataset since the annotators label only 60% of the images of the training dataset as the same class [21]. However, the sample selection criteria of testing set requires a complete labeling agreement among three pathologists. The sample images for normal, serrated, adenoma and adenocarcinoma classes are shown in Figure 1 a-d.

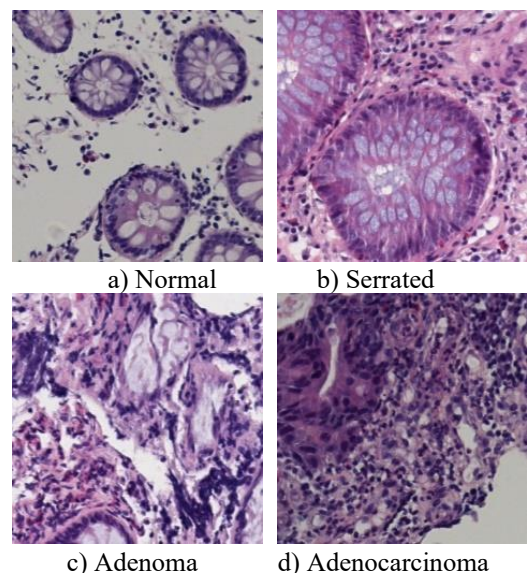


Figure 1. The sampled images for normal, serrated, adenoma and adenocarcinoma classes

4. METHOD

4.1 NST

NST algorithm is a novel method proposed by Gatys et al. [22] which separates content and style information of CNN representations and recombines them in a different and artistic way. A new image x is generated from a white noise image and iteratively converges its content information, main objects and scenes to the content image p and its style information like color tones, painting tricks and illuminance to the style image a . This iterative convergence procedure is achieved by decreasing the $L_{total}()$ cost which is the weighted sum of the $L_{content}()$ and $L_{style}()$ cost functions as shown in Eq. (1). The α and β weights are reciprocally multiplied with $L_{content}()$ and $L_{style}()$ cost functions are set as 1000 and 0.01 respectively. $L_{content}()$ is calculated for only one layer whereas $L_{style}()$ is calculated as a sum for more than one layers.

$$L_{total}(p, a, x) = \alpha L_{content}(p, x) + \beta L_{style}(a, x) \quad (1)$$

$$L_{content}(p, x, l) = \frac{1}{2} \sum_{i,j} (F_{ij}^l - P_{ij}^l)^2 \quad (2)$$

$L_{content}()$ cost function is calculated as the sum of the square of the difference between the F_{ij}^l and P_{ij}^l which are the feature maps of the image x which is aimed to be generated and image p which is the content image respectively as given in Eq. (2). The F_{ij}^l and P_{ij}^l feature maps are obtained by applying x and p images to the pretrained VGG-19 model respectively. The subscripts like i and j denote the i^{th} filter at position j whereas l symbolizes the layer number.

$L_{style}()$ cost function is calculated as the sum of the weighted sum of the layer specific style cost functions of the E_l as shown in Eq. (6). Each E_l cost is calculated as the sum of the square of the difference of G_{ij}^l and A_{ij}^l which are ‘‘Gram Style Matrices’’ (GSM) of the image x that is aimed to be generated and image a that is the style image respectively as given in Eq. (5). The F_{ij}^l and AF_{ij}^l feature maps are obtained by applying x and a images to the pretrained VGG-19 model respectively. The N_l and M_l values in the initial fractional element in the Eq. (5) are the number of the filters in layer l and the size of the filters calculated as the width times the height of the feature maps. The GSM is the key mathematical notion, which digitizes the style concept into the solid numbers by applying the dot product among each of the filter channels of each feature maps such as AF^l and F^l as shown in Eq. (3) and (4). The A_{ij}^l and G_{ij}^l matrices are the GSMs of the images a and x respectively. The height and width sizes of the A_{ij}^l and G_{ij}^l matrices are $N_l \times N_l$ and their elements having higher values show the more correlated filter channels. In other words, the higher values of A_{ij}^l GSM matrix are marker for the high correlation among the i^{th} and j^{th} filter channels at the k coordinates. The NST algorithm proposed by Gatys et al. [22] assumes that two different filter channels both showing high activation at the same k coordinates are more likely to have similar styles because of the effect of dot product.

$$A_{ij}^l = \sum_k AF_{ik}^l \cdot AF_{jk}^l \quad (3)$$

$$G_{ij}^l = \sum_k F_{ik}^l \cdot F_{jk}^l \quad (4)$$

$$E_l = \frac{1}{4N_l^2 M_l^2} \sum_{i,j} (G_{ij}^l - A_{ij}^l)^2 \quad (5)$$

$$L_{style}(a, x) = \sum_{l=0}^L w_l E_l \quad (6)$$

4.2 FL

The DL methods are data hungry algorithms but the data privacy issues are also a big concern. In order to get rid of this problem, MacMahan et al. offered the FL concept which is a collaborative distributed machine learning notion supported by some data privacy algorithms like differential privacy and homomorphic encryption [7]. The basic procedures of FL are summarized as the following steps [7]:

1. Each collaborator downloads the global DL model from the server.
2. Each collaborator trains their downloaded global model using its custom data. Thus, each collaborator obtains an updated version of the downloaded global model at this round.
3. The model parameter updates of each collaborator are sent back to the server.
4. The server aggregates and obtains a new global model. This new global model is also distributed back to the collaborators for the next round. The whole process is executed again from step 2 to step 4 iteratively until the planned round number is reached.

4.3 FNST

In this study, the novel FNST algorithm is proposed and its main procedure steps are described in Algorithm 1 and Figure 2. The novelty of the FNST algorithm is provided by separating the style and content loss calculation steps from each other, which are depicted in Eq. (1). At the step ‘‘1’’ of the algorithm, the main trick of FNST is managed by firstly applying the Eq. (3) in the collaborators’ site by giving selected style images a to the pretrained VGG-19 model and get the A_{ij}^l GSMs from the ‘conv1_1’, ‘conv2_1’, ‘conv3_1’, ‘conv4_1’ and ‘conv5_1’ layers. Then, the obtained A_{ij}^l GSM files for these 5 layers are loaded to the server. At the step ‘‘2’’ of the algorithm, the server loads only the A_{ij}^l GSMs belonging to the rare classes to the certain collaborators and act only like a dispatcher. The critical point here is that only the A_{ij}^l GSMs are transferred to the other collaborators not the raw data or the direct feature maps of VGG-19 model. At the step ‘‘3’’ the local content images are prepared for the NST algorithm. At the step ‘‘4’’ of the FNST algorithm, the $L_{style}()$ is calculated by using GSMs of (G_{ij}^l) and the A_{ij}^l as defined in Eq. (4), (5) and (6). Additionally, $L_{content}()$ is calculated using locally existing content image p and the image x which is to be generated as defined in Eq. (2). The final cost function $L_{total}()$ is calculated as the sum of the $L_{content}()$ and $L_{style}()$ as defined in Eq. (1) and the synthetic style transferred x image is generated iteratively. The FNST algorithm is executed only once and just before the classical FL algorithm. The effect of FNST algorithm is shown in Figure 3. It is worth to highlight that the proposed novel FNST algorithm does not manipulate general mathematical representation of NST algorithm proposed by

Gatys et al. but exploits the Eq. (3) and (4) by executing the steps of these equations separately and remotely on each private site. The steps of the Eq. (3) and (4) are remotely executed only once and A_{ij}^l and G_{ij}^l matrices as GSMs are generated on each private site. Thus, this trick of FNST algorithm takes away the necessity of simultaneous coexistence of content and style images at the same physical site for the original NST.

Algorithm 1: The FNST Algorithm

define:

- 1.1: C_i , $1 \leq i, z \leq N$ $i \neq z$ // C_i : i^{th} Collaborator
// N = Number of Collaborators
- 1.2: $1 \leq j \leq M$ // j^{th} Class M = Number of Classes
- 1.3: S_i , // S_i : the number of selected
// Style images in the i^{th} Collaborator
- 1.4: $LS_{i,j}$ // $LS_{i,j}$: the selected
// Local Style images of C_i for Class j
- 1.5: C_z , // C_z : the number of selected
// Content images in the z^{th} Collaborator
- 1.6: $LC_{z,j}$ // $LC_{z,j}$: the selected Local
// Content images of C_z for Class j

Start:

do

- 1.7: **for each** C_i and Class j **do**
StyleTransfer($LS_{i,j}$) \rightarrow ($GramMatrices_i$)_{server}
- 2: Load($GramMatrices_i$)
- 3: Load($LC_{z,j}$)
- 4: NST($LC_{z,j}$, $GramMatrices_i$) \rightarrow ($SyntheticImages_i$)

End

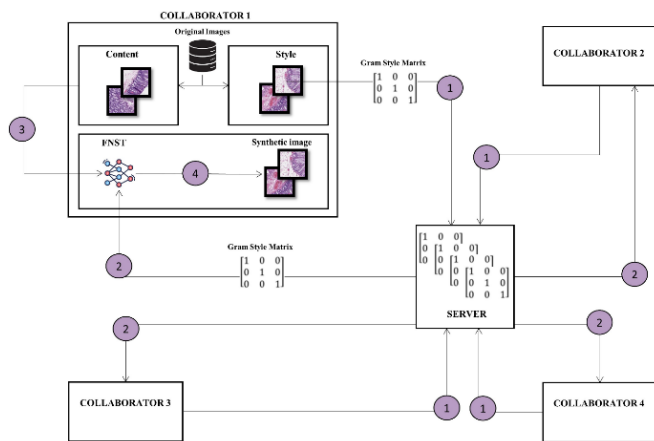


Figure 2. The flowchart of FNST architecture depicting the relations of server and clients. The detailed 4 steps are visualized only for the Collaborator 1

4.4 Architecture designs for benchmarking

In this research, the ResNet-18 model is benchmarked on mainly three different architectures such as Single Learning (SL), Centralized Learning (CL) and FL. The SL is the case when all the attending collaborators locally train their model only on their own data and no communication is performed

among the collaborators. The SL case can be seen as the most common situation currently happening across the institutions by using only the local private data and no collaboration is established. The CL is the case where some intuitions collaborate within certain restrictions and after wasting energy and time for reaching a common point. In the CL, all the data of the collaborators are gathered to a single machine and the DL model is trained on the whole data and it is expected to obtain the highest performance.

The CL architecture is the ideal case but the data privacy, strategical and commercial restrictions mostly prohibit this consensus. More and more, in the CL case all the computational power needs to be gathered to a single machine, which is also technically risky. In the FL case, four collaborators are proposed. The highly imbalanced data settings are tested on SL and FL cases. The last but not least, the FL architecture is benchmarked both on the raw imbalanced data setting and the synthetic data augmented settings which are generated by FNST algorithm.

In this research, the ResNet-18 model is implemented in Python using the PyTorch AI library [23]. The Flower platform is used for converting classical ResNet model to the federated version [24]. The data privacy issues of FL are handled via SecAgg and SecAgg+ secure model aggregation algorithms by the Flower platform [25, 26]. The FaultTolerantFedAvg is preferred as the model aggregation algorithm over the FedYogi, FedAVG and FedAVGM [27]. The local epoch numbers for FL scenarios are set as 10 whereas the global round/epoch numbers for SL, CL and FL are set as 100,100 and 50. All the settings and scenarios are executed on a workstation having 64 GB RAM, 2xNvidia RTX A4000 16GB GPUs and i7-11700F 3.6 GHz CPU.

5. RESULTS

In this study, the data augmentation settings are proposed based on 5 different combinations of the content and style image counts such as “Style 5 – Content 5 (S5-C5)”, “Style 5 – Content 10 (S5-C10)”, “Style 5 – Content 20 (S5-C20)”, “Style 10 – Content 10 (S10-C10)” and “Style 20 – Content 10 (S20-C10)” for Chaoyang dataset. The experiments are made by using ResNet-18 model for SL, CL, FL architectures. The proposed FNST method is also exploited to generate synthetic images and the contribution of these images on the performance is observed by comparing pure FL results, which are only based on the original images. The obtained ACC, F1 Score and AUC results are shown in Tables 1-4. Each row of the results is obtained by selecting the best AUC test results. The AUC values are calculated as “one-vs-rest” for multiclass image classification case but only the macro average of the obtained four different results are noted on the result tables.

The proposed FNST method is also exploited to generate synthetic images and the contribution of these images on the performance is observed by comparing pure FL results, which are only based on the original images. The obtained ACC, F1 Score and AUC results are shown in Tables 1-4. Each row of the results is obtained by selecting the best AUC test results. The AUC values are calculated as “one-vs-rest” for multiclass image classification case but only the macro average of the obtained four different results are noted on the result tables.

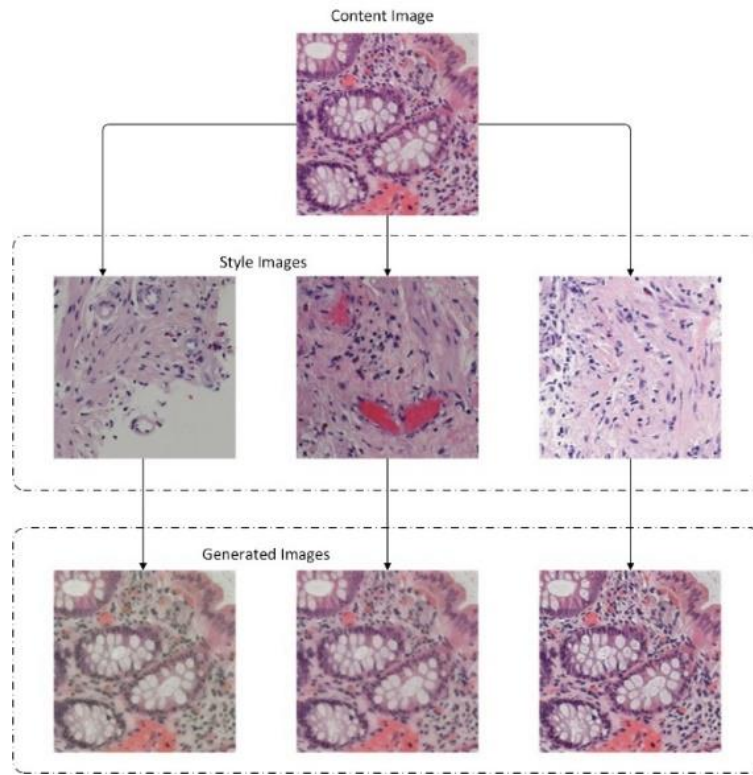


Figure 3. The effect of FNST algorithm on sample content and style images which are labeled as “Normal”

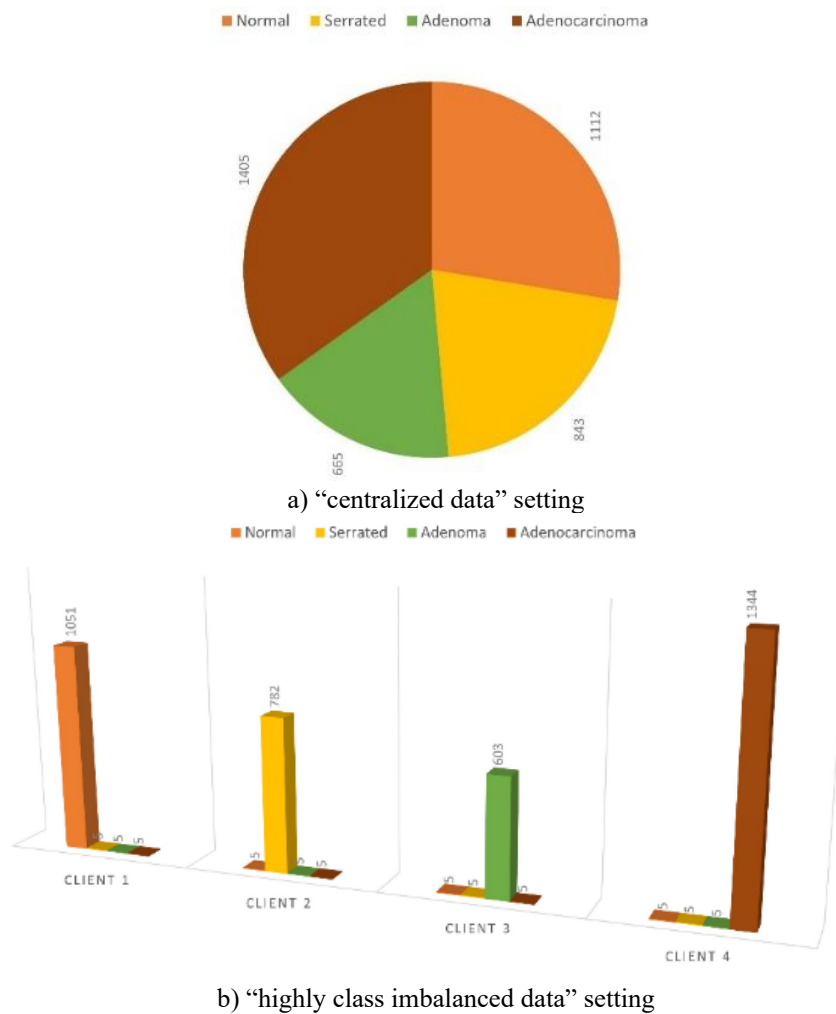


Figure 4. “Centralized data” setting for CL architecture and “highly class imbalanced data” setting having 5 images per rare classes for SL and FL architectures

Table 1. The results for CL architecture and “centralized data” setting

CL			
Local Epochs	ACC	F1_Score	MAX ROC-AUC
100	0.7447	0.6714	0.8906

Table 2. The results of SL and FL architectures on “highly class imbalanced data” settings having 5, 10 and 20 images per rare classes before FNST based Data Augmentation

Rare Label Counts	Metric	Client 1	Client 2	Client 3	Client 4	
		SL	SL	SL	SL	FL
5	ACC	0.3295	0.1500	0.1276	0.3927	0.3356
	F1_Score	0.1239	0.0652	0.0565	0.1409	0.1504
	AUC	0.5524	0.5435	0.6320	0.6717	0.7878
10	ACC	0.3295	0.1500	0.1276	0.3983	0.3356
	F1_Score	0.1239	0.0652	0.0565	0.1579	0.1504
	AUC	0.6110	0.5251	0.6420	0.6968	0.7878
20	ACC	0.3342	0.1500	0.1337	0.4296	0.4698
	F1_Score	0.1335	0.0652	0.0640	0.2308	0.2888
	AUC	0.7325	0.5826	0.6862	0.7582	0.8228

Table 3. The results of FL architecture on “highly class imbalanced data” settings having 5, 10 and 20 images per rare classes after FNST based Data Augmentation

Local Epochs	Metric	Style-5			Style-10	Style-20
		Content-5	Content-10	Content-20	Content-10	Content-10
50	ACC	0.4763	0.5736	0.6503	0.5568	0.5806
	F1_Score	0.3246	0.4831	0.5699	0.4703	0.5035
	AUC	0.7986	0.8136	0.8547	0.8045	0.8127

**Figure 5.** The F1 Score comparison of the SL and CL results as well as the FL results obtained after FNST for 5 rare images case

The CL architecture, which uses “centralized data” setting, defines the standard machine learning cases where all the dataset is located and executed at a single central location. The class distribution of the “centralized data” setting of Chaoyang dataset is shown in Figure 4-a. Additionally, the results for CL architecture are shown in Table 1.

In order to measure the effect of the extent of the imbalance ratio, 3 different “highly imbalanced data” settings which have 5, 10 and 20 images per rare classes are proposed. This imbalanced setting which is visualized in Figure 4-b are distributed of 4 clients and then the SL and FL architectures are applied on this configuration. The number of epochs for training of FL and SL are set as 50 and 100 respectively. However, the trained models are continuously tested just after 1 or 2 epoch steps for FL and SL respectively and the test results having the highest AUC values are reported in tables.

The novel FNST model proposed in this study is used to generate different synthetic images for 5 different configurations such as (S5-C5), (S5-C10), (S5-C20), (S10-C10) and (S20-C10) which are carefully selected from the original Chaoyang dataset. Different data augmentation configurations generate not only different images but also different number of images. The (S5-C5) generates $5 \times 5 = 25$

images whereas (S5-C20) generates $5 \times 20 = 100$ different extra images. The results of FL architecture on “highly class imbalanced data” settings after applying FNST based Data Augmentation are shown in Table 3.

The FL results from Table 3 which are obtained after applying FNST and the CL results from Table 1 as well as the SL results from Table 2 are visualized comparatively for 5 rare images case in Figure 5. All the FL architectures running on FNST based data augmented data exceed the SL values of the best performing clients dramatically.

The FL results from Table 3 which are obtained after applying FNST and the FL results from Table 2 which are applied only on the original dataset are visualized comparatively for 5, 10 and 20 rare images case in Figure 6. All the FL architectures running on FNST based data augmented data exceed the standard FL values dramatically.

6. DISCUSSION

The idea of using NST for the purpose of data augmentation is not new and was applied on different fields like skin melanoma diagnosis, histopathological images, x-ray and

magnetic resonance images [28-30]. The NST method was observed as contributing to the overall performance of the machine learning models. However, the pure NST algorithm requires having the content and style images at the same computational location simultaneously. In terms of data privacy issues, this requirement acts like a barrier to the different institutions to apply NST among their confidential images. Even if the fact that FL algorithm is proposed for enabling the collaboration among the private data of the institutions, it is observed that FL does not converge to a stable reasonable federated model in the case of “highly class imbalanced data” settings as shown in Figure 6, 7 and 8. The FNST algorithm is proposed as a novel method for converting the classical NST algorithm to its federated version. The FNST is achieved by detaching and moving the GSM calculation step to the collaborator client side and dispatching the GSM files of different collaborators to each other and finally resuming the rest of the cost calculation and optimization steps at the site of the other collaborator clients.

The before and after effect of FNST on the “highly class imbalanced data” settings with only 5 and 20 content images for the locally rare classes can be observed on Figure 7 and 8. Before applying FNST, the F1 Score values are monitored as relatively low and fluctuating more frequently like in Figure 7-b or staying monotonically like in Figure 7-a during the training time. After applying FNST, it is monitored that the monotonic structure of F1 Score of Figure 7-a shows relatively higher, fluctuating and slowly increasing behavior in Figure 8-a. The real effect of FNST is observed on Figure 8-b by having

higher, smoother F1 scores for all label classes.

Even if the fact that proposed FNST method is benchmarked on a CRC imaging dataset and the shape of the structures in the tissue is the primary concern, it may also be very beneficial for medical cases in which the histological tissue stain analysis is very important. Antibody Mediated Rejection (AMR) complication which may occur after kidney transplantation is an example in which the stain analysis a critical issue and has low inter-observer compliance [30]. Each biopsy has specific color and texture features and if the datasets are small then the generalization capacity of the DL models will decrease dramatically [30]. Thus, the proposed novel FNST method provides both the FL and NST algorithms together and enables secure collaborative data augmentation among different medical institutions especially for rare medical images.

As the main paradigm of FL is not disturbing the data privacy, the FNST is also designed as not violating the data privacy. A_{ij}^l , which is the GSM of style image “a” is computed by the dot product of the feature maps AF_{ik}^l and AF_{jk}^l after applying style image “a” to the VGG-19 model as given in Eq. (3). Even if one assumes that the AF_{ik}^l and AF_{jk}^l feature maps are exploited for reverse engineering to estimate style image “a”, the GSM A_{ij}^l cannot be reverse engineered by the adverse parties since the fact that having only the result of a dot product does not give any information about the multiplied feature maps and hence style image “a”.



Figure 6. The F1 Score comparison of the FL results applied only on the original data and the FL results obtained after FNST for 5, 10 and 20 rare images cases

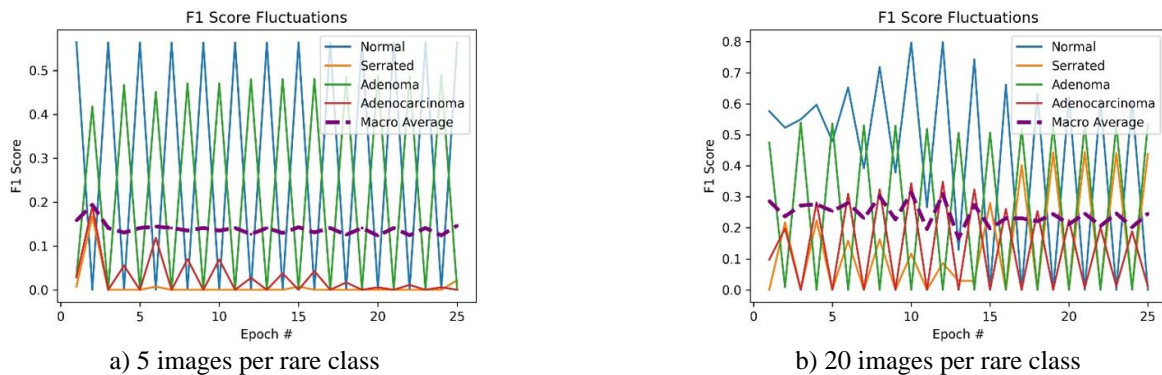


Figure 7. “Highly class imbalanced data” setting with only 5 and 20 content images for the locally rare classes before FNST

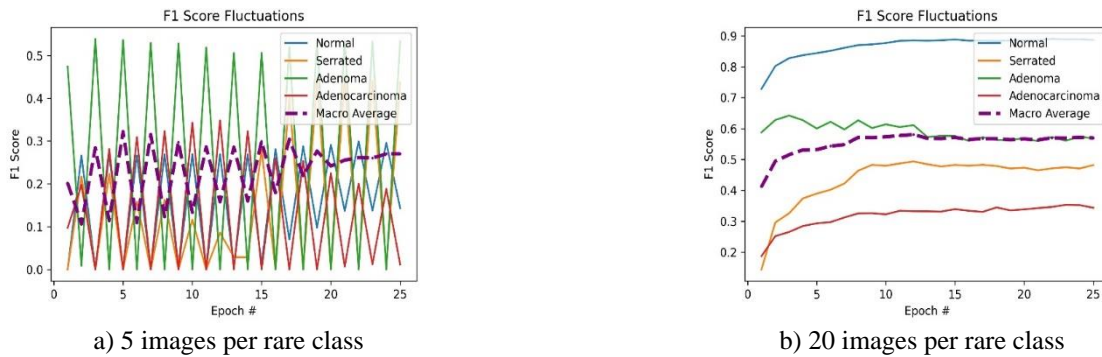


Figure 8. “Highly imbalanced data” setting with only 5 and 20 content images for the locally rare classes after FNST

Table 4. Comparison table of the contributions of FNST with respect to the FL and SL architectures for all settings

Setting	Contribution of FNST wrt. FL			Contribution of FNST wrt. Max SL		
	ACC	F1 Score	AUC	ACC	F1 Score	AUC
S5 – C5	0.1407	0.1504	0.0108	0.0836	0.1837	0.1269
S5 – C10	0.2380	0.3327	0.0258	0.1753	0.3252	0.1168
S5 – C20	0.1805	0.2811	0.0319	0.2207	0.4251	0.0965
S10 – C10	0.2212	0.3199	0.0167	0.1585	0.3124	0.1077
S20 – C10	0.2450	0.3531	0.0249	0.1823	0.3531	0.0249

The quantitative analysis of the contribution of FNST is tabularized with respect to the SL and FL architectures for all the style and content settings as shown in Table 4. In this study, the F1 Score is considered as the most meaningful metric since the proposed data settings are highly imbalanced. As can be seen from the Table 4, the most effective performance parameter is the number of the content images rather than the style images. This result probably stems from that transferred style only carry the information like illuminance, color and texture whereas the content images contain structure information, which is more valuable for the colorectal cancer classification task.

Nevertheless, the FNST is likely to contribute more dramatically for the cases in which the collaborator clients have different datasets having diverse styles such as different staining, lighting and image acquisition techniques. In this study, the content and style images are selected manually as expecting them to be maximally diverse. Different content and style images may give better results and using different layers of VGG-19 during computing $L_{content}()$ and $L_{style}()$ cost functions may give better results.

7. CONCLUSION

In this study, the ResNet-18 model is used for benchmarking of its SL, CL and FL versions on the “highly class imbalanced data” setting. The classical FL is observed as not being able to converge the local models to a reasonable collaborative model in the highly biased dataset. Thus, a novel FNST method which converts the classical NST algorithm to the federated version is proposed. The FNST increased the ACC, F1 Score and AUC results of pure FL by 24.5%, 35.31% and 2.49% using 10 images for content and 20 images for style. More and more, the FNST increased the ACC, F1 Score and AUC results of pure FL by 22.07%, 42.51% and 9.65% using 20 images for content and 5 images for style. The obtained results show that FNST algorithm can be used successfully by different medical institutions, which work especially on rare diseases or medical cases.

ACKNOWLEDGMENT

The use and release of Chaoyang dataset was made publicly available by Chaoyang Hospital and required citation is done for Zhu et al. [21]. This work has been supported by funding from Dicle University (DÜBAP Project No: MÜHENDİSLİK.22.001).

REFERENCES

- [1] Mármol, I., Sánchez-de-Diego, C., Pradilla Dieste, A., Cerrada, E., Rodriguez Yoldi, M.J. (2017). Colorectal carcinoma: a general overview and future perspectives in colorectal cancer. *International Journal of Molecular Sciences*, 18(1): 197-197. <https://doi.org/10.3390/ijms18010197>
- [2] Wong, N.A., Hunt, L.P., Novelli, M.R., Shepherd, N.A., Warren, B.F. (2009). Observer agreement in the diagnosis of serrated polyps of the large bowel. *Histopathology*, 55(1): 63-66. <https://doi.org/10.1111/j.1365-2559.2009.03329.x>
- [3] Farris, A. B., Misraji, J., Srivastava, A., Muzikansky, A., Deshpande, V., Lauwers, G.Y., Mino-Kenudson, M. (2008). Sessile serrated adenoma: challenging discrimination from other serrated colonic polyps. *The American journal of surgical pathology*, 32(1): 30-35. <https://doi.org/10.1097/PAS.0B013E318093E40A>
- [4] Ciompi, F., Veta, M., Van Der Laak, J., Rajpoot, N. (2021). Editorial Computational Pathology. *IEEE Journal of Biomedical and Health Informatics*, 25(2): 303-306. <https://doi.org/10.1109/JBHI.2021.3052029>
- [5] Hassan, L., Saleh, A., Abdel-Nasser, M., Omer, O.A., Puig, D. (2021). Efficient multi-organ multi-center cell nuclei segmentation method based on deep learnable aggregation network. *Traitement du Signal*, 38(3): 653-661. <https://doi.org/10.18280/ts.380312>
- [6] Burçak, K. C., Uğuz, H. (2022). A new hybrid breast cancer diagnosis model using deep learning model and

- relief. *Traitement du Signal*, 39(2): 521-529. <https://doi.org/10.18280/ts.390214>
- [7] Brendan McMahan, H., Moore, E., Ramage, D., Hampson, S., Agüera y Arcas, B. (2016). Communication-efficient learning of deep networks from decentralized data. *arXiv e-prints*, arXiv-1602.
- [8] Chakraborty, D., Ghosh, A. (2022). Improving the robustness of federated learning for severely imbalanced datasets. *arXiv preprint arXiv:2204.13414*.
- [9] Lu, M.Y., Kong, D., Lipková, J., Chen, R.J., Singh, R., Williamson, D.F., Chen, T.Y., Mahmood, F. (2022). Federated learning for computational pathology on gigapixel whole slide images. *Medical image analysis*, 76: 102298-102298. <https://doi.org/10.1016/j.media.2021.102298>
- [10] Adnan, M., Kalra, S., Cresswell, J.C., Taylor, G.W., Tizhoosh, H.R. (2022). Federated learning and differential privacy for medical image analysis. *Scientific reports*, 12(1): 1-10. <https://doi.org/10.1038/s41598-022-05539-7>
- [11] Andreux, M., Terrail, J.O.D., Beguier, C., Tramel, E.W. (2020). Siloed federated learning for multi-centric histopathology datasets. In *Domain Adaptation and Representation Transfer, and Distributed and Collaborative Learning*, 129-139. https://doi.org/10.1007/978-3-030-60548-3_13
- [12] Gunesli, G.N., Bilal, M., Raza, S.E. A., Rajpoot, N.M. (2021). Feddropoutavg: Generalizable federated learning for histopathology image classification. *arXiv preprint arXiv:2111.13230*.
- [13] Wagner, N., Fuchs, M., Tolkach, Y., Mukhopadhyay, A. (2022). Federated Stain Normalization for Computational Pathology. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 14-23. https://doi.org/10.1007/978-3-031-16434-7_2
- [14] Chen, J., Jiang, M., Dou, Q., Chen, Q. (2022). Federated Domain Generalization for Image Recognition via Cross-Client Style Transfer. *arXiv preprint arXiv:2210.00912*.
- [15] Sarkar, D., Narang, A., Rai, S. (2020). Fed-focal loss for imbalanced data classification in federated learning. *arXiv preprint arXiv:2011.06283*.
- [16] Ran, X., Ge, L., Zhong, L. (2021). Dynamic margin for federated learning with imbalanced data. In *2021 International Joint Conference on Neural Networks (IJCNN)*, 1-8. <https://doi.org/10.1109/IJCNN52387.2021.9534372>
- [17] Zhang, D.Y., Kou, Z., Wang, D. (2021). FedSens: A federated learning approach for smart health sensing with class imbalance in resource constrained edge computing. In *IEEE INFOCOM 2021-IEEE Conference on Computer Communications*, 1-10. <https://doi.org/10.1109/INFOCOM42981.2021.9488776>
- [18] Duan, M., Liu, D., Chen, X., Liu, R., Tan, Y., Liang, L. (2020). Self-balancing federated learning with global imbalanced data in mobile systems. *IEEE Transactions on Parallel and Distributed Systems*, 32(1): 59-71. <https://doi.org/10.1109/TPDS.2020.3009406>
- [19] Shuai, X., Shen, Y., Jiang, S., Zhao, Z., Yan, Z., Xing, G. (2022). BalanceFL: Addressing Class Imbalance in Long-Tail Federated Learning. *Proc - 21st ACM/IEEE Int Conf Inf Process Sens Networks*, 271-284. <https://doi.org/10.1109/IPSN54338.2022.00029>
- [20] Wu, N., Yu, L., Yang, X., Cheng, K.T., Yan, Z. (2022). FedRare: Federated Learning with Intra-and Inter-Client Contrast for Effective Rare Disease Classification. *arXiv preprint arXiv:2206.13803*.
- [21] Zhu, C., Chen, W., Peng, T., Wang, Y., Jin, M. (2021). Hard Sample Aware Noise Robust Learning for Histopathology Image Classification. *IEEE Transactions on Medical Imaging*, 41(4): 881-894. <https://doi.org/10.1109/TMI.2021.3125459>
- [22] Gatys, L.A., Ecker, A.S., Bethge, M. (2015). A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*. <https://doi.org/10.1167/16.12.326>
- [23] Inkawhich, N. (2021). Finetuning Torchvision Models-PyTorch Tutorials 1.2. 0 documentation. https://pytorch.org/tutorials/beginner/finetuning_torchvision_models_tutorial.html.
- [24] Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., de Gusmão, P.P., Lane, N.D. (2020). Flower: A friendly federated learning research framework. *arXiv preprint arXiv:2007.14390*.
- [25] Bonawitz, K., Ivanov, V., Kreuter, B., et al. (2017). Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 1175-1191. <https://doi.org/10.1145/3133956.3133982>
- [26] Bell, J.H., Bonawitz, K.A., Gascón, A., Lepoint, T., Raykova, M. (2020). Secure single-server aggregation with (poly) logarithmic overhead. In *Proceedings of the 2020 ACM SIGSAC Conference on Computer and Communications Security*, 1253-1269. <https://doi.org/10.1145/3372297.3417885>
- [27] Beutel, D.J., Topal, T., Mathur, A., Qiu, X., Parcollet, T., Gusmão, P.P., Lane, N.D. (2022). Flower: A friendly federated learning framework. *HAL Science ouverte*. <https://hal.archives-ouvertes.fr/hal-03601230>.
- [28] Hernandez-Cruz, N., Cato, D., Favela, J. (2021). Neural Style Transfer as Data Augmentation for Improving COVID-19 Diagnosis Classification. *SN Computer Science*, 2(5): 1-12. <https://doi.org/10.1007/s42979-021-00795-2>
- [29] Mikołajczyk, A., Grochowski, M. (2018). Data augmentation for improving deep learning in image classification problem. In *2018 International Interdisciplinary PhD workshop (IIPHDW)*, pp. 117-122. <https://doi.org/10.1109/IIPHDW.2018.8388338>
- [30] Cicalese, P.A., Mobiny, A., Yuan, P., Becker, J., Mohan, C., Nguyen, H.V. (2020). StyPath: Style-transfer data augmentation for robust histology image classification. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 351-361. https://doi.org/10.1007/978-3-030-59722-1_34

NOMENCLATURE

L	loss function
A, AF, F, P	feature map
a, S	style image
p, C	content image
x, G	generated image
w	weight of the neural network

Greek symbols

α content loss weight coefficient
 β style loss weight coefficient

Subscripts

i, z collaborator client index
 j class index
 l layer id