Leveraging Multi-Color Spaces as a Defense Mechanism Against Model Inversion Attack

Anonymous Authors[1]

Abstract

Privacy is of increasing importance in the world of machine learning in general and in healthcare more specifically due to the nature of patients data. Multiple type of security attacks and mechanisms already exist which allow adversaries to extract sensitive information based only from a high-level interaction with a trained machine learning model. This paper specifically addresses the model inversion attack, which aims to reconstruct input data from a model's output. This paper describes a novel approach of using multi-color spaces as a defense mechanism against this type of attack to strenghten the privacy of open source models trained on image data. The main idea of our approach is to use a combination of those color spaces to create a more generic representation and reduce the quality of the reconstruction coming from a model inversion attack while maintaining a good classification performance. We evaluate the privacy-utility ratio of our proposed security method on retina images.

1. Introduction

In the era of cloud services and open source applications, more deep learning models are being deployed and served in the cloud allowing other parties to benefit and take advantage of them in their own projects. However, despite the black-box deployment of such models, mishonest individuals and entities known as adversaries might still cause and apply harmful actions through leveraging attacks in order to identify sensitive information and contents of the data that were used to train the models. Despite the presence of various data regulations as GDPR (Voigt & Von dem Bussche, 2017) and CCPA (Pardau, 2018), which had a growing positive impact in raising awareness in protecting users data and adding a legal and ethical framework around data-oriented applications, they are unfortunately not enough in stopping the bad intentions of adversaries that still use techniques and methods to infer some information about the training data of the users and patients. An adversary might have 2 types of access: a white-box access, where the adversary knows the model's weights and is aware of the exact architecture of the model, and a black-box access, where the adversary only receives the prediction and output of the model without having any knowledge of how the model is structured. In the latter case, some attacks, such as membership inference and model inversion attacks, can still be applied. Membership inference (Shokri et al., 2017; Carlini et al., 2022) is a type of attack where the adversary builds a binary classification model which verifies and checks if a sample was part of the training data or not. Model inversion attack (Fredrikson et al., 2014) on the other hand is a type of privacy attack that aims to recreate the input data only from the model's output. In the context of privacy-preserving machine learning in healthcare (Xiang et al., 2021; Yadav et al., 2023; Guerra-Manzanares et al., 2023), the highly sensitive nature of patient data emphasizes the importance of ensuring optimal privacy-utility ratios. This consideration is crucial when training and deploying machine learning models in the cloud and open-sourcing them. The objective is to maintain good performance in the original task assigned to the model while protecting patient data as much as possible. In this paper, we explore the impact of combining multimodal autoencoders and color spaces as a way to strengthen representation models applied on images. Representation learning is gaining more in popularity especially with its direct benefit of allowing institutions with limited resources to collaborate with big data companies which offer them Representation-as-a-Service (RaaS) systems defined as models deployed in the cloud and acting as feature extractors on similar data. Model inversion attack can be applied on the learned representations to reverse engineer them and reconstruct the train image data. In our study, We empirically evaluate and experiment on retina images. The presence of personalized patterns created by the blood vessels in a retina directly increases the privacy concern when delivering

[1]Anonymous Institution, Anonymous City, Anonymous Region, Anonymous Country. Correspondence to: Anonymous Author <anon.email@domain.com>.

Preliminary work. Under review by the International Conference on Machine Learning (ICML). Do not distribute.

Artificial Intelligence (AI) solutions on retina images.

We summarize our contributions in this paper as following:

- We propose an encoding strategy based on the use of multimodal autoencoder of various color spaces to reduce the reconstruction quality of model inversion attacks on images.
- We evaluate the privacy-utility ratio of our suggested encoding method on diabetic retina images.
- We show that different combinations of color spaces yield better privacy-utility ratio against model inversion attacks than using the default RGB color format.

2. Background & Related Work

2.1. Model Inversion Attacks

Model inversion attack is a type of privacy threat where an adversary attempts to recover sensitive information about individual training samples or data points by exploiting the model's outputs. It was first introduced by (Fredrikson et al., 2014) where it has been shown that attackers can perform a model inversion attack on a trained model to retrieve patient's genetic markers. Model inversion attacks are defined by two main approaches namely:

- **Optimization-based** (Zhang et al., 2020; Nguyen et al., 2023; Wu et al., 2023): In this approach, the adversary relies on a gradient-based optimization problem for reconstructing the data. This works by finding $\hat{x}$ which approximates the prediction and output of a model $\mathcal{M}$ given an input x. More precisely, the gradient-based optimization work on updating dummy data $\{\hat{x}_1, \hat{x}_2, ..., \hat{x}_n\}$ and labels $\{\hat{y}_1, \hat{y}_2, ..., \hat{y}_n\}$ to match the observed gradients on the original data by minimizing the objective function: defined in Eq 1:

$$\min ||\sum_{i=1}^{n} \nabla L(\mathcal{M}(\hat{x}_i), \hat{y}_i) - \sum_{i=1}^{n} \nabla L(\mathcal{M}(x_i), y_i)||_2^2 \quad (1)$$

  With L being the loss function of the task.
- **Learning-based** (Yang et al., 2019; Zhou et al., 2023): Given a model $\mathcal{M}$ trained on a private dataset $D_{priv} = \{x_i, y_i\}$, the main goal of learning-based inversion attacks is to find an optimal inversion model I, defined as a decoder, and training it to minimize the following objective function:

$$L = R(x, I(\mathcal{M}(x))) \quad (2)$$

  where R is a chosen metric to evaluate the quality of the reconstructed data.

In our paper, we consider the case of training-based model inversion attacks applied on retina image data where we compare the impact of color spaces in reducing the reconstruction attack performance.

2.2. Color Spaces

By definition, a color space is a way to represent how colors are perceived by humans considering different and various angles and parameters. They are mathematical models set to show and encode how colors can be represented as a set of vectors. They were used within the domain of applying and developing machine learning architectures and models in computer vision. ColorNet (Gowda & Yuan, 2019) is a model designed to take an ensemble of 7 different types of color spaces each one linked to a Densenet (Huang et al., 2017) which resulted and yielded better performance. Color spaces were also leveraged in using deep learning for image colorization (Cheng et al., 2015; Yoo et al., 2019; Pucci et al., 2021) where the main goal behind this process is to generate and produce images that appear visually natural. They were also used in the context of health for example *HED* (Haematoxylin-Eosin-DAB) is a color space introduced by (Ruifrok et al., 2001) and was developed with the goal of better analyzing tissues.

2.2.1. RGB

RGB is an additive color model where colors are represented by combinations of red, green, and blue light. *RGB* is known for being labeled as a hardware-friendly type of color space since the vast majority of hardware which stores pixels uses *RGB* as a coloring system which makes it the default and mostly used color space among all.

2.2.2. HSV

HSV is a psychological type of color model and it was primarily designed to order colors which align with human perception and it was modeled on the ways that people consciously break down colors. This model was based more upon how colors are organized and conceptualized in human vision in terms of other color-making attributes, such as hue, lightness, and chroma. HSV as a color space includes the following parameters. Hue (H) represents the type of color, which is intuitively understood by humans. Saturation (S) represents the intensity or purity of the color, which also corresponds well to how humans perceive color. Value (V) represents the lightness or darkness of the color.

2.2.3. L*A*B

The *L*a*b* color space was designed and developed to mimic the function of a human eye. *L* stands for lightness, *a* stands for green-red component and *b* for the yellow-blue one.

3. Methods & Architecture

In this section, we introduce the main components of our proposed privacy solution, aiming to create a more robust

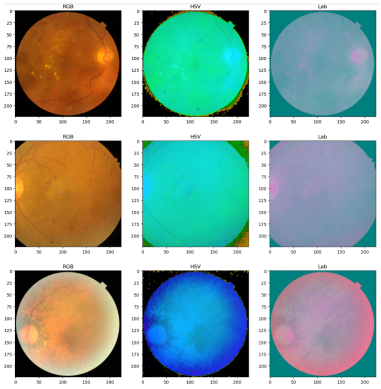

Figure 1. From left to right columns we have examples of RGB, HSV and L*a*b color spaces of retina images respectively

representation against model inversion attacks. Additionally, we provide a step-by-step explanation of the workflow of the overall pipeline.

3.1. Multimodal Autoencoder with Color Spaces

Autoencoders are a type of unsupervised learning neural network designed to learn proper reconstruction mechanisms and patterns. An autoencoder is composed of an encoder $E(.)$ and a decoder $D(.)$ and is trained to minimize the reconstruction loss defined in Eq 3 as the mean squared error between an input x and its reconstructed version x'.

$$L = \sum_{i=1}^{n}(x_i - x_i')^2 = \sum_{i=1}^{n}(x_i - D(E(x_i, \theta_E), \theta_D))^2 \quad (3)$$

Given various modalities, an autoencoder can be extended to a multimodal autoencoder (Jaques et al., 2017; Bachmann et al., 2022) by simultaneously reconstructing multiple modalities in parallel from a common shared latent space and it is trained by optimizing the loss function defined in Eq 4, where n and m represent the number of samples and number of modalities, respectively.

$$L = \sum_{i=1}^{n}\sum_{j=1}^{m}[x_i^{(j)} - x_i'^{(j)}]^2 = \sum_{i=1}^{n}\sum_{j=1}^{m}[x_i^{(j)} - D^{(j)}(E^{(j)}(x_i^{(j)}, \theta_E^{(j)})), \theta_D^{(j)})]^2 \quad (4)$$

In our proposed security method, we compare the security robustness of the representation generated from training multimodal autoencoders with various combinations of the 3 color spaces we previously defined in Section 2.2. We consider here the scenario and use case of cloud services "Representation as a Service" (RaaS) or "Embedding as a Service" (EaaS). These services fall into the category of

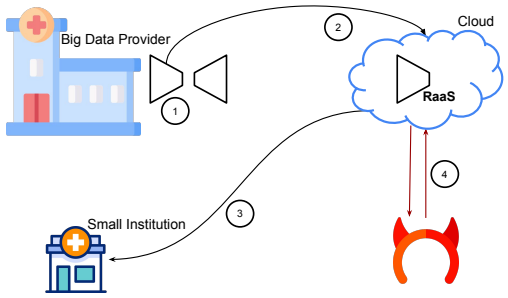

Figure 2. The attack scenario is defined as follows: A hospital with a huge dataset trains an autoencoder locally as shown in ①. Then the encoder part is deployed and open-sourced in the cloud as a feature extractor in ② as a Representation-as-a-Service (RaaS) cloud application. Small clinics and institutions with limited data and computational resources ③ benefits from it by obtaining meaningful representations through querying the RaaS. However, and as illustrated in ④ an adversary, assuming he has some similar data, keeps querying the RaaS and create a decoder D_{inv} which intends to reconstruct data only from the learned representation thus identifying some private and sensitive attributes.

Machine-Learning-as-a-Service (MLaaS). The idea behind RaaS is to allow big data holders and providers to train models which returns a meaningful vector representation for a given datatype and deploy those models into the cloud as a RaaS so that smaller institutions and enterprises with limited training data and capacity can query those models as feature extractors and obtain a representation $\mathcal{R}$ and then fine-tune a predictor for a down-stream task trained on the returned representations provided by the RaaS. In our threat analysis, we assume that an adversary $\mathcal{A}$ aims to apply a model inversion attack on the RaaS. Formerly given a set of representations $\{\mathcal{R}\}$ of a small institution that queried the RaaS in the cloud, the adversary $\mathcal{A}$ using a decoder D_{inv}, as inversion model, aims to reconstruct the samples $X' = D_{inv}(\mathcal{R})$, where $\{X'\}$ is set to be as close as possible to $\{X\}$. In order to obtain D_{inv} we assume that the adversary $\mathcal{A}$ has some data X_{adv} and keeps querying the RaaS until getting a pair dataset $\{\mathcal{R}_{adv}, X_{adv}\}$ and train afterwards the decoder D_{inv} through a learning-based model inversion attack approach as defined in Section 2.1. The overall security/attack scenario and pipeline is summarized in Fig 2.

4. Experiments & Results

4.1. Datasets

4.1.1. RFMiD

The Retinal Fundus Multi-disease Image Dataset (Pachade et al., 2021) consists of 3200 fundus images captured using three different fundus cameras with 46 conditions annotated through adjudicated consensus of two senior retinal experts.

It includes 45 diseases and pathologies. In our paper, we consider the binary classification task of predicting if the retina is healthy or abnormal.

4.1.2. DEEPDRID

Diabetic Retinopathy Grading and Image Quality Estimation Challenge (Liu et al., 2022) is a dataset of retina images with the aim to predict different levels and grades of diabetic retinopathy which is the most common disease caused by diabetes. We consider the task of ordinal regression by predicting the grade of diabetic retinopathy severity scale ranging from 0 (no apparent retinopathy) to 4 (Proliferative Diabetic Retinopathy).

4.1.3. IMPORTANCE OF PRIVACY IN RETINA IMAGES

Because of the complex network of blood vessels in the retina, every eye has a unique pattern where even identical twins retinas are different. In fact, due to those unique and personalized patterns, a biometric method known as retina recognition (Seto, 2009; Choraś, 2012) exists where the distinct patterns found on an individual's retina are utilized to identify them. In addition, retina images include information about the gender of the patient which is considered a private attribute in that context. Previous work has shown that deep learning models can be developed for gender identification on retina images (Korot et al., 2021). Due to these sensitive information and specific patterns present in retina images strengthening RaaS for those type of images is of big importance.

4.2. Experimental Settings

We consider in our experiments X_{train} and X_{test} generated on a 10-fold stratified splits. The experimental pipeline is based on Fig 2. X_{train} is used to train either a standard autoencoder with 1 color space or a multimodal autoencoder with 2 color spaces and more. Those autoencoders are trained for 32 epochs and a batch size of 64. The resulted learned representation to be outputted later on when deployed as a RaaS is of size 64. After being deployed in the cloud as a RaaS, an adversary uses X_{test} and keeps quering the RaaS until having a pair dataset $\{R_{test}, X_{test}\}$, R_{test} being the set of representations relevant to X_{test}. Based on that, the adversary trains a decoder D_{inv} as the model inversion attack for 50 epochs. We assume in this scenario that the adversary is able to access the representation R_{train} relevant to X_{train} and use the learned decoder D_{inv} to obtain the reconstructed version of the train dataset noted as X'_{train}. In our study and experimental setting, the adversary do not know the type of color spaces combination used in the autoencoder and we assume that the adversary recreates the RGB format of the images as it is the default and most used color space in the community. To evaluate how good the inversion attack model performed we compute the image quality difference between X_{train} and X'_{train}. On the other side and to make sure our security method does not discard useful information relevant to diagnostic purposes when trained on the learned representation we use $\{R_{train}, y_{train}\}$ and $\{R_{test}, y_{test}\}$ for classification evaluation. We consider in our study 3 types of machine learning models namely support vector machine, random forest and XGBoost and grid search over them to find the best classification model.

4.3. Evaluation Metrics

Due to the importance of computing the privacy-utility ratio when evaluating a security solution we decided to use macro F1 score as the classification metric and both mean squared error and structural similarity index matching for evaluating the quality of the reconstruction image from the model inversion attack step.

- **Macro F1-Score**: We decided to use macro F1 score as the classification metric due to its robust nature. It is the harmonic mean between the precision and the recall thus it simultaneously considers the ratio of false positive and false negatives. In addition, in the presence of high imbalance in a multi-class problem, macro f1 score gives equal weight to each label which ensures that minority classes are not excluded and are properly evaluated.
- **Mean Squared Error**: Famous metric used in comparing continuous variables, in the case of images it is defined by computing the pixel-by-pixel $l2$ norm. Given 2 colored images of C channels $I, J \in \mathbb{R}^{H \times W \times C}$ their mean squared error is defined as:

$$MSE(I,J) = \frac{1}{H \times W \times C} \sum_{i=1}^{H} \sum_{j=1}^{W} \sum_{k=1}^{C} (I_{i,j,k} - J_{i,j,k})^2 \tag{5}$$

- **Structural Similarity Index Matching (SSIM)**: is a method developed by (Wang et al., 2004) to quantify how similar 2 images are. It takes simultaneously into account luminance and contrast and SSIM is labeled as a perception-based metric. Unlike mean squared error (MSE) which compares pixel by pixel and has no boundaries, SSIM is a normalized metric which ranges between -1 and 1 and evaluate the similarity on the overall perception level rather than pixel-to-pixel level.

$$SSIM(x,y) = \frac{(2\mu_x + \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{6}$$

Where:
- x and y are images we would like to compare.
- $\mu_x$and μ_y are the average pixel values of x and y.
- σ_x^2 and σ_y^2 are the variance of pixel values in x and y.
- σ_{xy} is the covariance of pixel values between x and y.

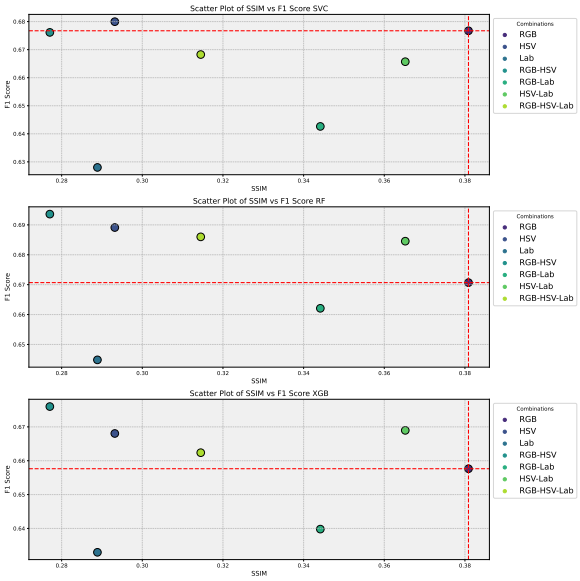

Figure 3. RFMiD

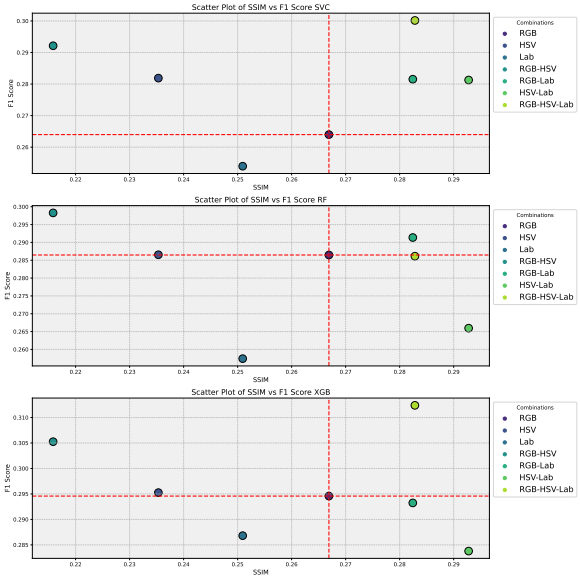

Figure 4. DeepDRID

Figure 5. F1 Score vs SSIM of different color spaces combinations for RFMiD and DeepDRiD Retina Datasets. The best cases to consider is when the F1 score is bigger or equal and the SSIM is smaller than the RGB baseline case thus all datapoints on the top left corner of the interesection of the 2 lines.

– c_1 and c_2 are constants defined to stabilize the division with weak denominator and in the extreme case to avoid the division by zero.

4.4. Results

The results are shown in Table 1 in Appendix A, the best combinations are the ones that give high macro F1 score, high mean squared error and low SSIM, this translates to obtaining a RaaS where the learned representation is useful for the considered diagnostic tasks in addition of being robust enough against a potential model inversion attack. The baseline use case is when the autoencoder is only trained on the RGB format to create the representation. In Fig 5, all color space combination results are plotted in terms of both SSIM and Macro F1 Score. The best combinations are the data points that give an equal or higher macro F1 Score and a smaller SSIM compared to the RGB baseline. In addition, The fact that multiple color spaces combinations performed better than the RGB baseline gives us the flexibility and allows us to not be dependant towards the use of a unique combination which makes it harder to the adversary to potentially know before hand which exact combination we have used.

5. Discussion & Limitations

The main limitation given to our method specifically and to the use of color spaces in general is its constrained application towards only images as a modality. This will for example exclude electronic health records (EHR) and omics data which fall into the category of tabular data.

6. Conclusion

In this paper, we have proposed a method which merges multimodal autencoder and color spaces as a defense mechanism to generate meaningful representations and reduce the reconstruction quality of model inversion attacks.

References

Bachmann, R., Mizrahi, D., Atanov, A., and Zamir, A. Multimae: Multi-modal multi-task masked autoencoders. In *European Conference on Computer Vision*, pp. 348–367. Springer, 2022.

Carlini, N., Chien, S., Nasr, M., Song, S., Terzis, A., and Tramer, F. Membership inference attacks from first principles. In *2022 IEEE Symposium on Security and Privacy (SP)*, pp. 1897–1914. IEEE, 2022.

Cheng, Z., Yang, Q., and Sheng, B. Deep colorization.

In *Proceedings of the IEEE international conference on computer vision*, pp. 415–423, 2015.

Choraś, R. S. Retina recognition for biometrics. In *Seventh International Conference on Digital Information Management (ICDIM 2012)*, pp. 177–180. IEEE, 2012.

Fredrikson, M., Lantz, E., Jha, S., Lin, S., Page, D., and Ristenpart, T. Privacy in pharmacogenetics: An {End-to-End} case study of personalized warfarin dosing. In *23rd USENIX security symposium (USENIX Security 14)*, pp. 17–32, 2014.

Gowda, S. N. and Yuan, C. Colornet: Investigating the importance of color spaces for image classification. In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part IV 14*, pp. 581–596. Springer, 2019.

Guerra-Manzanares, A., Lopez, L. J. L., Maniatakos, M., and Shamout, F. E. Privacy-preserving machine learning for healthcare: open challenges and future perspectives. In *International Workshop on Trustworthy Machine Learning for Healthcare*, pp. 25–40. Springer, 2023.

Huang, G., Liu, Z., Van Der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4700–4708, 2017.

Jaques, N., Taylor, S., Sano, A., and Picard, R. Multimodal autoencoder: A deep learning approach to filling in missing sensor data and enabling better mood prediction. In *2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII)*, pp. 202–208. IEEE, 2017.

Korot, E., Pontikos, N., Liu, X., Wagner, S. K., Faes, L., Huemer, J., Balaskas, K., Denniston, A. K., Khawaja, A., and Keane, P. A. Predicting sex from retinal fundus photographs using automated deep learning. *Scientific reports*, 11(1):10286, 2021.

Liu, R., Wang, X., Wu, Q., Dai, L., Fang, X., Yan, T., Son, J., Tang, S., Li, J., Gao, Z., et al. Deepdrid: Diabetic retinopathy—grading and image quality estimation challenge. *Patterns*, 3(6), 2022.

Nguyen, N.-B., Chandrasegaran, K., Abdollahzadeh, M., and Cheung, N.-M. Re-thinking model inversion attacks against deep neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 16384–16393, 2023.

Pachade, S., Porwal, P., Thulkar, D., Kokare, M., Deshmukh, G., Sahasrabuddhe, V., Giancardo, L., Quellec, G., and Mériaudeau, F. Retinal fundus multi-disease image dataset (rfmid): A dataset for multi-disease detection research. *Data*, 6(2):14, 2021.

Pardau, S. L. The california consumer privacy act: Towards a european-style privacy regime in the united states. *J. Tech. L. & Pol'y*, 23:68, 2018.

Pucci, R., Micheloni, C., and Martinel, N. Collaborative image and object level features for image colourisation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2160–2169, 2021.

Ruifrok, A. C., Johnston, D. A., et al. Quantification of histochemical staining by color deconvolution. *Analytical and quantitative cytology and histology*, 23(4):291–299, 2001.

Seto, Y. *Retina Recognition*, pp. 1128–1130. Springer US, Boston, MA, 2009. ISBN 978-0-387-73003-5. doi: 10.1007/978-0-387-73003-5_132. URL https://doi.org/10.1007/978-0-387-73003-5_132.

Shokri, R., Stronati, M., Song, C., and Shmatikov, V. Membership inference attacks against machine learning models. In *2017 IEEE symposium on security and privacy (SP)*, pp. 3–18. IEEE, 2017.

Voigt, P. and Von dem Bussche, A. The eu general data protection regulation (gdpr). *A Practical Guide, 1st Ed., Cham: Springer International Publishing*, 10(3152676): 10–5555, 2017.

Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Wu, R., Chen, X., Guo, C., and Weinberger, K. Q. Learning to invert: Simple adaptive attacks for gradient inversion in federated learning. In *Uncertainty in Artificial Intelligence*, pp. 2293–2303. PMLR, 2023.

Xiang, D., Cai, W., et al. Privacy protection and secondary use of health data: strategies and methods. *BioMed Research International*, 2021, 2021.

Yadav, N., Pandey, S., Gupta, A., Dudani, P., Gupta, S., and Rangarajan, K. Data privacy in healthcare: In the era of artificial intelligence. *Indian Dermatology Online Journal*, 14(6):788–792, 2023.

Yang, Z., Zhang, J., Chang, E.-C., and Liang, Z. Neural network inversion in adversarial setting via background knowledge alignment. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security*, pp. 225–240, 2019.

Yoo, S., Bahng, H., Chung, S., Lee, J., Chang, J., and Choo, J. Coloring with limited data: Few-shot colorization via memory augmented networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11283–11292, 2019.

Zhang, Y., Jia, R., Pei, H., Wang, W., Li, B., and Song, D. The secret revealer: Generative model-inversion attacks against deep neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 253–261, 2020.

Zhou, S., Zhu, T., Ye, D., Yu, X., and Zhou, W. Boosting model inversion attacks with adversarial examples. *IEEE Transactions on Dependable and Secure Computing*, 2023.

A. Privacy-Utility Evaluation

We summarize privacy and utility performance of all possible color spaces in combinations generated from the set of {RGB, HSV, Lab} in Table 1. The baseline to compare with is RGB since it is the default color space choice within the computer vision community.

The best performing color spaces are the ones that yield similar of better macro F1 score, a higher mean squared error and a smaller structural similarity index matching (SSIM) than the RGB baseline.

We can observe that HSV and RGB-HSV are the color spaces which always verify this criteria with various machine learning models and across both datasets.

Table 1. Privacy-Utility Ratio Results of Multiple Color Space Combinations

	RFMiD				
	F1 SVM	**F1 RF**	**F1 XGB**	**MSE**	**SSIM**
RGB (baseline)	*67.67* ± 3.53	67.06 ± 2.71	65.76 ± 2.96	0.027 ± 0.014	0.380 ± 0.106
HSV	**67.99** ± 3.32	*68.91* ± 2.36	66.80 ± 3.27	0.028 ± 0.012	0.293 ± 0.055
Lab	62.80 ± 4.46	64.48 ± 4.15	63.29 ± 4.11	0.029 ± 0.012	*0.288* ± 0.076
RGB-HSV	67.61 ± 2.46	**69.35** ± 2.28	**67.60** ± 2.32	**0.042** ± 0.013	**0.277** ± 0.081
RGB-Lab	64.26 ± 3.93	66.20 ± 3.50	63.98 ± 2.93	0.027 ± 0.008	0.344 ± 0.081
HSV-Lab	66.57 ± 3.16	68.45 ± 1.80	*66.89* ± 1.96	0.022 ± 0.002	0.365 ± 0.046
RGB-HSV-Lab	66.82 ± 2.61	68.59 ± 3.19	66.23 ± 3.32	*0.031* ± 0.015	0.314 ± 0.075

	DeepDRiD				
	F1 SVM	**F1 RF**	**F1 XGB**	**MSE**	**SSIM**
RGB (baseline)	26.39 ± 3.83	28.64 ± 3.07	29.45 ± 3.40	*0.056* ± 0.021	0.266 ± 0.082
HSV	28.18 ± 3.40	28.65 ± 4.12	29.52 ± 1.92	0.052 ± 0.016	*0.235* ± 0.089
Lab	25.39 ± 4.96	25.74 ± 3.60	28.68 ± 3.65	0.054 ± 0.017	0.250 ± 0.063
RGB-HSV	*29.21* ± 2.73	**29.82** ± 2.44	*30.52* ± 3.39	**0.065** ± 0.018	**0.215** ± 0.072
RGB-Lab	28.15 ± 1.80	*29.13* ± 1.88	29.32 ± 1.50	0.046 ± 0.010	0.282 ± 0.076
HSV-Lab	28.12 ± 4.08	26.59 ± 1.82	28.37 ± 1.95	0.046 ± 0.007	0.292 ± 0.056
RGB-HSV-Lab	**30.01** ± 3.51	28.61 ± 3.65	**31.23** ± 4.20	0.048 ± 0.014	0.282 ± 0.078