The Gaussian-Head OFL Family

One-shot federated learning from client global statistics

Data-free | One-shot | Strong non-IID robustness

Main idea

Federate class statistics once.
Reconstruct strong Gaussian heads centrally.
No raw data, no gradients, no public proxy set.

Fabio Turazza ■ Marco Picone ■ Marco Mamei

Why another one-shot FL method?

What existing OFL still struggles with

- Many methods are not truly **data-free**: they rely on **public/proxy data** or server-side distillation.
- Under strong **non-IID** heterogeneity, one-shot aggregation often becomes unstable and accuracy degrades.
- Several pipelines depend on heavy **knowledge distillation**, **ensembles**, or model/head sharing.

Design principle. Share only additive class statistics, then reconstruct the global head centrally.

No proxy data ■ No model exchange ■ Better aligned with non-IID one-shot aggregation

Typical OFL recipe

public/proxy data
knowledge distillation
ensemble aggregation
model or head sharing

GH-OFL:

class counts +
first/second-order statistics

Federate statistics, not models.

GH-OFL in one slide

1 Frozen encoder on each client $\Longrightarrow$ **2** Per-class sufficient statistics $\Longrightarrow$ **3** Secure aggregation $\Longrightarrow$ **4** Gaussian + synthetic heads

What clients send once

- N_c: class counts
- A_c: summed embeddings
- B, S_c, D_c: second-order moments
- Optional public random projection for compression/privacy

Closed-form heads NB-diag, LDA, QDA are computed directly from moments.

What the server does

- Estimate means, priors, pooled/class covariances
- Build a Fisher subspace
- Sample **synthetic Fisher-space features**
- Train **FisherMix** and **Proto-Hyper**

Why it is new the entire family is data-free, partition-invariant in expectation, and server-centric.

A family of heads, not a single classifier

Closed-form GH heads

NBdiag / LDA / QDA

Built directly from moments.

Strength: zero training, minimal overhead.

FisherMix

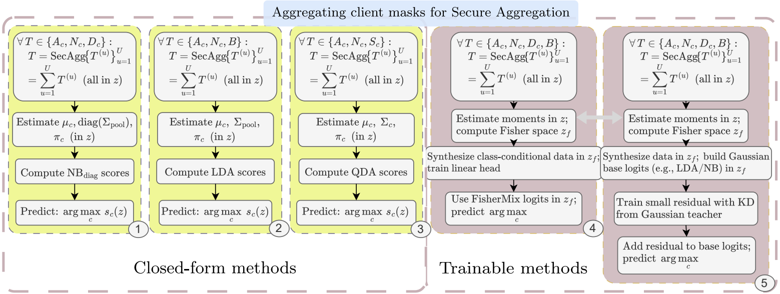
Linear head in the Fisher subspace.

Strength: improves margins beyond purely Gaussian rules.

Proto-Hyper

Low-rank residual over a Gaussian base.

Strength: tiny trainable correction, still one-shot and data-free.



Empirical takeaways

Method	CIFAR-10			CIFAR-100			SVHN		
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.50$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.50$	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.50$
FL baselines (multi-round, $R \in \{1, 10, 50\}$)									
FedAvg (1 round)	27.38	24.53	31.17	2.15	2.03	1.50	19.26	11.94	19.25
FedAvg (10 rounds)	64.13	73.08	82.62	29.19	29.92	33.41	54.94	74.16	80.73
FedAvg (50 rounds)	77.42	84.60	91.52	62.46	63.58	68.55	<u>78.79</u>	87.71	92.13
FedProx (1 round)	27.49	32.24	35.65	2.24	2.00	1.47	12.99	11.97	19.23
FedProx (10 rounds)	64.18	81.82	89.02	29.51	29.79	33.71	57.84	74.15	81.01
FedProx (50 rounds)	77.54	85.68	<u>91.74</u>	62.76	63.68	<u>68.61</u>	76.30	<u>87.88</u>	<u>92.17</u>
SCAFFOLD (1 round)	20.36	27.71	31.81	1.93	1.63	1.50	12.87	12.05	18.79
SCAFFOLD (10 rounds)	64.58	79.04	82.69	28.87	30.96	33.28	57.45	71.45	80.80
SCAFFOLD (50 rounds)	73.78	<u>88.39</u>	91.66	61.63	64.09	68.60	75.99	87.45	92.00
OFL baselines									
DENSE	31.26	56.21	62.42	14.31	17.21	26.49	37.49	51.53	**77.44**
Co-Boost.	44.37	60.41	67.43	20.30	24.63	34.43	41.90	57.13	**84.65**
FedPFT	56.08	56.43	56.80	36.79	37.16	37.95	42.55	43.03	43.84
FedCGS	63.95	63.95	63.95	39.95	39.95	39.95	57.77	57.77	57.77
GH-OFL (ours)									
GH-NB$_{diag}$	78.84	78.84	78.84	55.51	55.51	55.51	39.24	39.24	39.24
GH-LDA	<u>**86.05**</u>	**86.05**	**86.05**	63.92	63.92	63.92	**62.16**	**62.16**	62.16
GH-QDA$_{full}$	84.40	84.40	84.40	**66.52**	**66.52**	**66.52**	55.30	55.30	55.30
FisherMix	84.74	84.74	84.74	<u>**66.99**</u>	<u>**66.99**</u>	**66.99**	57.79	57.79	57.79
Proto-Hyper	**85.74**	**85.74**	**85.74**	64.05	64.05	64.05	**61.97**	**61.97**	61.97

What stands out

LDA is a strong default.

QDA helps when class covariance matters.

FisherMix / **Proto-Hyper** recover accuracy beyond closed-form rules.

Takeaway

Richer covariance modeling pays off when feature geometry becomes class-dependent.

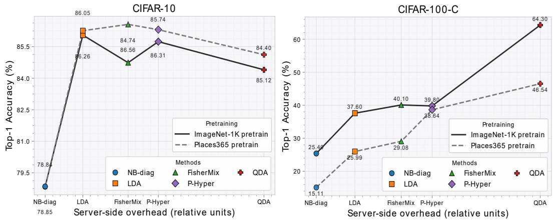
Why QDA matters on hard shift

Interpretation

- Corruptions distort classes **differently**.
- A shared covariance model (**LDA**) can become too rigid.
- **QDA** captures class-specific spread and correlations.
- Fisher-space trainable heads are a lighter compromise.

Take-home. On easy regimes, simple Gaussian heads are near the Pareto frontier; under stronger shift, richer covariance or synthetic refinements restore robustness.

	CIFAR-100-C	
Method	**Shared Stats**	**Acc.**
FedCGS	A, B, N	24.4%
GH-NB$_{diag}$	A, D, N	25.4%
GH-LDA	A, B, N	37.6%
FisherMix	A, B, N, D	40.1%
ProtoHyper	A, B, N, D	39.8%
GH-QDA$_{full}$	A, N, S	**64.3%**



Federate statistics, not models.

One-shot. Data-free. Principled. Practical.

GH-OFL shows that additive client statistics are enough to build strong global heads: closed-form when simplicity is enough, synthetic refinements when richer geometry matters.

One upload

No raw data

Non-IID aware