AggMapNet: Enhanced and Explainable Low-Sample Omics Deep Learning with Feature-Aggregated Multi-Channel Networks

Wan Xiang Shen[1,2], Yu Liu[3,4], Yan Chen[1], Xian Zeng[5], Ying Tan[1,6], Yu Yang Jiang[1,7*] and Yu Zong Chen[1,7*]

[1] The State Key Laboratory of Chemical Oncogenomics, Key Laboratory of Chemical Biology, Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, 518055, P. R. China

[2] Bioinformatics and Drug Design Group, Department of Pharmacy, and Center for Computational Science and Engineering, National University of Singapore, 117543, Singapore

[3] Institute for Health Innovation & Technology, National University of Singapore, 117543, Singapore

[4] Department of Biomedical Engineering, Faculty of Engineering, National University of Singapore, 117543, Singapore

[5] Department of Biological Medicines & Shanghai Engineering Research Center of Immunotherapeutics, School of Pharmacy, Fudan University, Shanghai 201203, P. R. China

[6] Shenzhen Kivita Innovative Drug Discovery Institute, Shenzhen, 518110, P. R. China.

[7] Institute of Biomedical Health Technology and Engineering, Shenzhen Bay Laboratory, Shenzhen, 518132 China

*Correspondence:

Yu Zong Chen (phacyz@nus.edu.sg)

Yu Yang Jiang (jiangyy@sz.tsinghua.edu.cn)

Supplementary Tables

Table S1 | The hyperparameters in AggMap feature restructuring

Table S2 | The hyperparameters in AggMapNet

Table S3 | The hyperparameters of the AggMap and AggMapNet for the different datasets

Table S4 | Comparison of classification accuracy of the three feature restructuring methods for each class of 33 cancers in the multi-task TCGA-T dataset.

Table S5 | AggMapNet performances versus three standard ML models on the 18 transcriptome benchmark datasets.

Table S6 | AggMapNet performances versus three standard ML models combined with PCA feature embedding on the 18 transcriptome benchmark datasets.

Table S7 | AggMapNet performances versus three tree-based ML models on the 18 transcriptome benchmark datasets.

Table S8 | AggMapNet performances versus three tree-based ML models combined with feature selection on the 18 transcriptome benchmark datasets.

Table S9 | List of the 6 AggMapNet identified important proteins of Covid-19 severity and the literature reports of the association of these proteins to Covid-19 severity.

Table S1 | The hyperparameters in AggMap feature restructuring

Stage	Parameter	Default	Description
Initialization stage	metric	correlation	Distance metric to measure the similarities between the FPs (**Eqn. 1**)
Fitting stage	cluster_channels	5	Number of the channels / clusters of the feature points
	var_thr	-1	Parameter to remove low-variance features. Feature points with variance lower than this threshold will be removed.
	n_epochs	500	Epochs in in minimization the differences of the two weighed graphs (**Eqn. 11**)
	lr	1.0	Learning rate in minimization the differences of the two weighed graphs (**Eqn. 11**)
	min_dist	0.01	The minimum distance apart that points are allowed to be in the low dimensional representation (**Eqn. 8**)
	n_neighbors	15	K number of nearest neighbours when estimating the manifold structure of the data (**Eqn. 3**)
Transformation stage	scale_method	'minmax'	Data scaling by z-score standard scaling or minmax scaling

Table S2 | The hyperparameters in AggMapNet

Type	Parameter	Default	Description
Network Architecture Parameters (NAPs)	conv1_kernel_size	13	The kernel size of first convolutional layers, should be odd number
	dense_layers	[128]	Number of the pyramidal dense layers and the units per dense layer
	dense_avf	relu'	Activation function in dense layers
	dropout	0	Dropout rate in the dense layers
	batch_norm	FALSE	Whether uses the batch normalization after convolution layers or not
	n_inception	2	Number of the inception layers
Training-Control Parameters (TCPs)	epochs	200	Number of the epochs
	lr	1.00E-04	Learning rate
	batch_size	32	Batch size
	loss	MSE / CE	The loss function
	metric	'ACC'	Evaluation metric during the training, {'ROC', 'ACC', 'PRC'} in classification tasks, {'rmse', 'r2'} in regression task
	monitor	'val_loss'	A monitor for early stopping, can be ''val_loss' and ''val_metric', select the best model by the performance of the validation set
	patience	1000000	Number of epochs with no improvement on the monitor after which training will be stopped

Table S3 | The hyperparameters of the AggMap and AggMapNet for the different datasets

Dataset	MNIST	MNIST	CCTD-U	TCGA-T	TCGA-S	TCGA-G	COV-D	COV-S
Samples (n)	70k	70k,	5	10446,	249~1134	179~554	363	41
Features (p)	784	784	5162	10381	17970	17970	88	1486
HPs	**AggMap feature restructuring**							
Fmap Sizes	28 × 28	28 × 28	72 x 72	102 × 102	135 × 134	135 × 134	10 × 9	39 × 39
Cluster channels	5	5	6	5	5	5	5	5
HPs	**AggMapNet**							
conv1_kernel_size	3	3	-	13	13	13	11	5
dense_layers	(128, 64)	(128, 64)	-	(128)	(128)	(128)	(128)	(128)
epochs	100	100	-	100	30~100	30~100	50	50
lr	1e-4	1e-4	-	1e-3	1e-4	1e-4	1e-4	1e-4
batch_size	64	64	-	64	16	16	4	1

Table S4 | Comparison of classification accuracy of the three feature restructuring methods (Fmaps are generated by Lyu-Reshape(1), Bazgir-REFINED(2), and AggMap) for each class of 33 cancers in the multi-task TCGA-T dataset. The average ten-fold cross-validation accuracy is reported, bold values indicate the better performing model.

Tumor Type	Cohort	Sample size	Lyu-reshape	Bazgir-REFINED	AggMap (C=1)	AggMap (C=5)
Rectum adenocarcinoma	READ	105	0.35238	0.42857	0.41905	**0.46667**
Cholangiocarcinoma	CHOL	45	0.55556	0.64444	0.57778	**0.71111**
Esophageal carcinoma	ESCA	196	0.76531	0.83163	0.82653	**0.86735**
Uterine Carcinosarcoma	UCS	57	0.80702	0.80702	0.82456	**0.82456**
Kidney Chromophobe	KICH	91	0.86813	0.79121	0.87912	**0.91209**
Lung squamous cell carcinoma	LUSC	552	0.90761	0.90036	0.91486	**0.91848**
Cervical and endocervical cancers	CESC	309	0.92880	0.90939	0.90939	0.92880
Kidney renal papillary cell carcinoma	KIRP	323	0.93189	0.89474	0.92260	0.93189
Glioblastoma multiforme	GBM	171	0.94152	0.98246	0.98246	**0.98830**
Mesothelioma	MESO	87	0.94253	0.95402	0.95402	**0.96552**
Adrenocortical carcinoma	ACC	79	0.94937	0.94937	0.97468	**0.97468**
Colon adenocarcinoma	COAD	328	**0.94512**	0.89024	0.88415	0.89634
Kidney renal clear cell carcinoma	KIRC	606	0.95215	0.95545	0.95545	**0.96700**
Lung adenocarcinoma	LUAD	576	0.94792	0.93750	0.93576	0.94792
Stomach adenocarcinoma	STAD	450	0.95556	0.93333	0.92889	**0.95778**
Uterine Corpus Endometrial Carcinoma	UCEC	201	0.95522	0.90547	0.93035	0.95522
Bladder urothelial carcinoma	BLCA	427	**0.96956**	0.94614	0.94379	0.95550
Liver hepatocellular carcinoma	LIHC	423	0.96927	0.96217	0.96927	**0.98109**
Pancreatic adenocarcinoma	PAAD	183	0.96721	0.94536	0.95628	**0.97268**
Sarcoma	SARC	265	0.96604	0.95094	0.98113	**0.98491**
Head and Neck squamous cell carcinoma	HNSC	566	0.97527	0.98763	0.99647	**0.99823**
Brain Lower Grade Glioma	LGG	530	0.98491	0.99434	0.99057	**0.99434**
Skin Cutaneous Melanoma	SKCM	473	0.97886	0.97463	0.98520	**0.98309**
Breast invasive carcinoma	BRCA	1212	0.99422	0.99010	0.99422	**0.99587**
Ovarian serous cystadenocarcinoma	OV	307	0.98697	0.99349	0.99674	**0.99674**
Testicular Germ Cell Tumors	TGCT	156	0.99359	0.98077	0.97436	**1.00000**
Thymoma	THYM	122	**0.99180**	0.98361	0.97541	0.98361
Uveal Melanoma	UVM	80	**0.98750**	0.98750	0.97500	0.97500
Diffuse Large B-cell Lymphoma	DLBC	48	1.00000	0.97917	0.97917	1.00000
Acute Myeloid Leukemia	LAML	173	1.00000	1.00000	1.00000	1.00000
Pheochromocytoma and Paraganglioma	PCPG	187	**1.00000**	0.98930	0.98930	0.99465
Prostate adenocarcinoma	PRAD	550	1.00000	0.99636	0.99818	1.00000
Thyroid carcinoma	THCA	568	1.00000	1.00000	0.99824	1.00000
Total / Average		10446 /317	0.92337	0.92051	0.92494	**0.94029**

Table S5 | AggMapNet performances versus three standard ML models on the 18 transcriptome benchmark datasets. The average ROC-AUCs of five-fold cross-validation are reported, the bold values indicate the better performing model. LGR: L2-regularized multinomial Logistic Regression, RF: Random Forest, kNN: k Nearest Neighbor. * values are taken from the paper of Smith et al.(2020)(3).

	Cancer type	Sample size	Binary task	LGR*	RF*	kNN*	AggMapNet (C=5)
TCGA-S: stage vs. stage	COAD	505	II- vs. III+	0.723	0.689	0.580	**0.724**
	KIRC	544	II- vs. III+	0.774	0.738	0.723	**0.775**
	LIHC	374	I- vs. II+	0.634	0.641	0.561	**0.682**
	LUAD	542	I- vs. II+	0.629	0.649	0.590	**0.656**
	SKCM	249	II- vs. III+	0.619	**0.663**	0.550	0.661
	STAD	416	II- vs. III+	0.617	0.537	0.563	**0.618**
	THCA	513	I- vs. II+	**0.719**	0.644	0.529	0.679
	UCEC	554	I- vs. II+	0.652	0.678	0.638	**0.707**
	LUSC	504	I- vs. II+	**0.662**	0.625	0.557	0.624
	BRCA	1134	II- vs. III+	**0.639**	0.604	0.573	0.629
TCGA-G: grade vs. grade	CESC	306	II- vs. III+	0.633	0.656	0.610	**0.668**
	KIRC	544	II- vs. III+	0.594	0.576	0.559	**0.632**
	LGG	532	II- vs. III+	**0.792**	0.762	0.664	0.774
	LIHC	374	II- vs. III+	0.663	0.670	0.602	**0.689**
	PAAD	179	II- vs. III+	**0.681**	0.621	0.620	0.631
	STAD	416	II- vs. III+	**0.760**	0.720	0.647	0.754
	UCEC	554	II- vs. III+	0.895	0.878	0.815	**0.903**
	HNSC	504	II- vs. III+	0.663	0.717	0.596	**0.758**
Average	-	486	-	0.686	0.670	0.610	**0.698**

Table S6 | AggMapNet performances versus three standard ML models combined with PCA feature embedding on the 18 transcriptome benchmark datasets. The average ROC-AUCs of five-fold cross-validation are reported, the bold values indicate the better performing model. LGR: L2-regularized multinomial Logistic Regression, RF: Random Forest, kNN: k Nearest Neighbor. * values are taken from the paper of Smith et al. (2020)(3)

	Cancer type	Sample size	Binary task	LGR * (PCA)	RF* (PCA)	kNN* (PCA)	AggMapNet (C=5)
TCGA-S: stage vs. stage	COAD	505	II- vs. III+	0.660	0.697	0.582	**0.724**
	KIRC	544	II- vs. III+	0.746	0.746	0.678	**0.775**
	LIHC	374	I- vs. II+	0.624	0.628	0.576	**0.682**
	LUAD	542	I- vs. II+	0.614	0.637	0.569	**0.656**
	SKCM	249	II- vs. III+	0.615	**0.674**	0.544	0.661
	STAD	416	II- vs. III+	0.562	0.601	0.551	**0.618**
	THCA	513	I- vs. II+	0.630	0.652	0.536	**0.679**
	UCEC	554	I- vs. II+	0.696	0.686	0.648	**0.707**
	LUSC	504	I- vs. II+	0.654	**0.658**	0.581	0.624
	BRCA	1134	II- vs. III+	0.620	0.592	0.536	**0.629**
TCGA-G: grade vs. grade	CESC	306	II- vs. III+	0.639	**0.699**	0.561	0.668
	KIRC	544	II- vs. III+	0.580	0.594	0.577	**0.632**
	LGG	532	II- vs. III+	0.730	0.767	0.692	**0.774**
	LIHC	374	II- vs. III+	0.627	0.685	0.602	**0.689**
	PAAD	179	II- vs. III+	0.646	0.584	0.591	**0.631**
	STAD	416	II- vs. III+	0.724	0.748	0.661	**0.754**
	UCEC	554	II- vs. III+	0.877	0.886	0.833	**0.903**
	HNSC	504	II- vs. III+	0.677	0.726	0.602	**0.758**
Average	-	486	-	0.662	0.681	0.607	**0.698**

Table S7 | AggMapNet performances versus three tree-based ML models on the 18 transcriptome benchmark datasets. The average ROC-AUCs of five-fold cross-validation are reported, the bold values indicate the better performing model. RoTF: Rotation Forest, XGB: XGBoost, LGB: LightGBM.

	Cancer type	Sample size	Binary task	RoTF	XGB	LGB	AggMapNet (C=5)
TCGA-S: stage vs. stage	COAD	505	II- vs. III+	0.673	0.720	0.706	**0.724**
	KIRC	544	II- vs. III+	0.725	0.789	**0.794**	0.775
	LIHC	374	I- vs. II+	0.636	0.667	0.650	**0.682**
	LUAD	542	I- vs. II+	0.641	0.648	0.642	**0.656**
	SKCM	249	II- vs. III+	0.586	0.592	0.646	**0.661**
	STAD	416	II- vs. III+	0.532	0.581	0.566	**0.618**
	THCA	513	I- vs. II+	0.640	0.655	0.668	**0.679**
	UCEC	554	I- vs. II+	0.637	0.666	0.661	**0.707**
	LUSC	504	I- vs. II+	0.591	0.606	0.620	**0.624**
	BRCA	1134	II- vs. III+	0.582	0.625	**0.650**	0.629
TCGA-G: grade vs. grade	CESC	306	II- vs. III+	0.640	0.583	0.607	**0.668**
	KIRC	544	II- vs. III+	0.571	0.578	0.594	**0.632**
	LGG	532	II- vs. III+	0.644	0.736	0.745	**0.774**
	LIHC	374	II- vs. III+	0.671	0.651	0.677	**0.689**
	PAAD	179	II- vs. III+	0.599	0.622	0.623	**0.631**
	STAD	416	II- vs. III+	0.708	0.747	**0.766**	0.754
	UCEC	554	II- vs. III+	0.847	0.882	0.892	**0.903**
	HNSC	504	II- vs. III+	0.705	0.752	0.741	**0.758**
Average	-	486	-	0.642	0.670	0.680	**0.698**

Table S8 | AggMapNet performances versus three tree-based ML models combined with feature selection on the 18 transcriptome benchmark datasets. The average ROC-AUCs of five-fold cross-validation are reported, the bold values indicate the better performing model. The feature selection method is performed on the training set, thus the selected features in each fold is different during the cross-validations. RoTF: Rotation Forest, XGB: XGBoost, LGB: LightGBM.

	Cancer type	Sample size	Binary task	Feature Selected	RoTF	XGB	LGB	AggMapNet (C=5)
TCGA-S: stage vs. stage	COAD	505	II- vs. III+	269~700	0.648	0.688	0.689	**0.724**
	KIRC	544	II- vs. III+	2527~3919	0.689	0.766	**0.787**	**0.775**
	LIHC	374	I- vs. II+	1040~1818	0.619	0.655	0.620	**0.682**
	LUAD	542	I- vs. II+	2581~4902	0.622	0.642	0.640	**0.656**
	SKCM	249	II- vs. III+	1163~1976	0.570	0.607	0.632	**0.661**
	STAD	416	II- vs. III+	105~505	0.578	0.529	0.526	**0.618**
	THCA	513	I- vs. II+	137~173	0.590	**0.702**	0.693	**0.679**
	UCEC	554	I- vs. II+	2367~3644	0.633	0.671	0.682	**0.707**
	LUSC	504	I- vs. II+	153~246	0.590	0.610	0.607	**0.624**
	BRCA	1134	II- vs. III+	139~753	0.573	0.597	0.622	**0.629**
TCGA-G: grade vs. grade	CESC	306	II- vs. III+	1470~2420	0.613	0.608	0.604	**0.668**
	KIRC	544	II- vs. III+	173~1027	0.534	0.608	0.584	**0.632**
	LGG	532	II- vs. III+	2390~3233	0.691	0.736	0.736	**0.774**
	LIHC	374	II- vs. III+	1627~2854	0.630	0.653	**0.699**	**0.689**
	PAAD	179	II- vs. III+	1740~2736	0.590	0.593	0.628	**0.631**
	STAD	416	II- vs. III+	1721~5304	0.686	**0.767**	0.765	0.754
	UCEC	554	II- vs. III+	5259~5521	0.791	0.881	0.890	**0.903**
	HNSC	504	II- vs. III+	955~2578	0.687	0.734	0.729	**0.758**
Average	-	486	-	-	0.630	0.669	0.674	**0.698**

Table S9 | List of the 6 AggMapNet identified important proteins of Covid-19 severity and the literature reports of the association of these proteins to Covid-19 severity.

Family / Type	Feature Point	Protein Name	Description
Notch	Q04721	Neurogenic locus notch homolog protein 2, NOTCH2	The expression of NOTCH2 significantly increases the risk of COVID-19 infection(4).
Metalloproteinase	P08253	72 kDa type IV collagenase, MMP2	Although there is no direct literature to support, previous study suggested that its family member MMP9 (92 kDa type IV collagenase) may be an early indicator of respiratory failure in COVID-19 patients(5)
Selenoproteins	P49908	Selenoprotein P, SELENOP	The two markers are highly correlated to each other and have been reported that in viral infection with potential relevance to COVID-19(6), and SELENOP along with Zn and Se as composite biomarker have been used to predict the survival odds in COVID-19(7).
	P22352	Glutathione peroxidase 3, GPX3	
Interleukin-1	Q9NPH3	interleukin-1 receptor accessory protein, IL1RAP	IL1RAP is a coreceptor of type 1 interleukin 1 receptor (IL1R1) and is indispensable for transmission of IL-1 signaling, early IL-1 receptor blockade in severe inflammatory respiratory failure complicating COVID-19(8).
Superoxide dismutase	P08294	Extracellular superoxide dismutase, SOD3	There is evidence of a link between decreased expression of the antioxidant enzyme superoxide dismutase 3 (SOD3) in the lungs of elderly patients with COVID-19 and disease severity(9)

Supplementary Figures

Fig. S1 | AggMapNet architecture with MNIST dataset as input.

Fig. S2 | The full code for AggMap feature restructuring, AggMapNet model learning and AggMapNet model explanations by both Shapley-explainer and Simply-explainer.

Fig. S3 | Various levels of additive Gaussian noise on the test set of the MNIST.

Fig. S4 | The noisy test set generation for the four Fmaps (Org1, OrgRP1, RPAgg1 and RPAgg5).

Fig. S5 | The noise-added Fmaps for TCGA-T dataset.

Fig. S6 | AggMap feature restructuring results on random permutated F-MNIST data.

Fig. S7 | AggMap fitting historical performances and final 2D embedding results on randomly permuted MNIST and F-MNIST.

Fig. S8 | Robustness of AggMapNet classification performance on noise-added test set of MNIST and F-MNIST.

Fig. S9 | A comparison of the Fmaps of 33 cancers of the TCGA-T dataset generated by Lyu and Haque's study(1) and AggMap.

Fig. S10 | Comparison of the Simply-explainer and Shapley-explainer on the noise-free MNIST recognition model explanation.

Fig. S11 | Comparison of the Simply-explainer and Shapley-explainer on the noise MNIST recognition model explanation.

Fig. S12 | Comparison of the Simply-explainer and Shapley-explainer on global explanation of breast cancer diagnostic model trained by WDBC dataset.

Fig. S13 | Comparison of the predicted value on the independent cohorts for the RF and AggMapNet classification of the COV-S dataset.

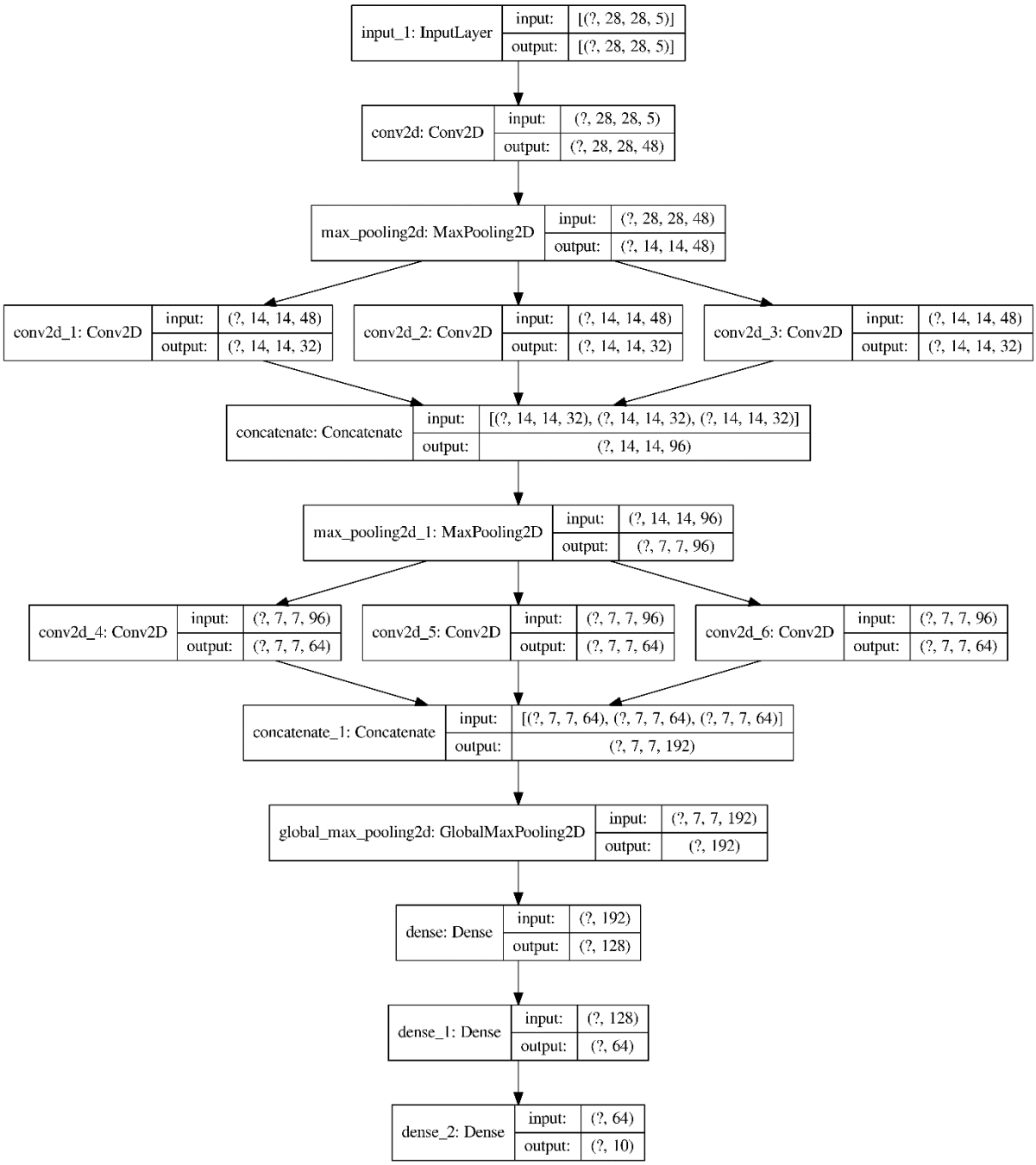


Fig. S1 | AggMapNet architecture with MNIST dataset as input, the inputs are the 4-D tensor: batch size, height, width, channels, where the height, width and channels are 28, 28 and 5, respectively. The number of the trainable parameters in the model is ~0.3M.

```python
# -*- coding: utf-8 -*-
"""
Example pipeline code for
1) AggMap multi-channel Fmaps transformation;
2) AggMapNet model traininng, validation;
3) AggModel explainations using Shap and Simp values.
"""
import pandas as pd
from sklearn.datasets import load_breast_cancer
from aggmap import AggMap, AggMapNet

# Data loading
data = load_breast_cancer()
dfx = pd.DataFrame(data.data, columns=data.feature_names)
dfy = pd.get_dummies(pd.Series(data.target))

# AggMap object definition, fitting, and saving
mp = AggMap(dfx, metric = 'correlation')
mp.fit(cluster_channels=5, emb_method = 'umap', verbose=0)
mp.save('agg.mp')

# AggMap visulizations: Hierarchical tree, embeddng scatter and grid
mp.plot_tree()
mp.plot_scatter()
mp.plot_grid()

# Transoformation of 1d vectors to 3D Fmaps (-1, w, h, c) by AggMap
X = mp.batch_transform(dfx.values, n_jobs=4, scale_method = 'minmax')
y = dfy.values

# AggMapNet training, validation, early stopping and saving
clf = AggMapNet.MultiClassEstimator(epochs=50, gpuid=0)
clf.fit(X, y, X_valid=None, y_valid=None)
clf.save_model(('agg.model'))

# Model explaination by simply-explainer: global, local
simp_explainer = AggMapNet.simply_explainer(clf, mp)
global_simp_importance = simp_explainer.global_explain(clf.X_, clf.y_)
local_simp_importance = simp_explainer.local_explain(clf.X_[[0]], clf.y_[[0]])

# Model explaination by shapley-explainer: global, local
shap_explainer = AggModel.shapley_explainer(clf, mp)
global_shap_importance = shap_explainer.global_explain(clf.X_)
local_shap_importance = shap_explainer.local_explain(clf.X_[[0]])
```

Fig. S2 | The full code for AggMap feature restructuring, AggMapNet model learning and AggMapNet model explanations by both Shapley-explainer and Simply-explainer.

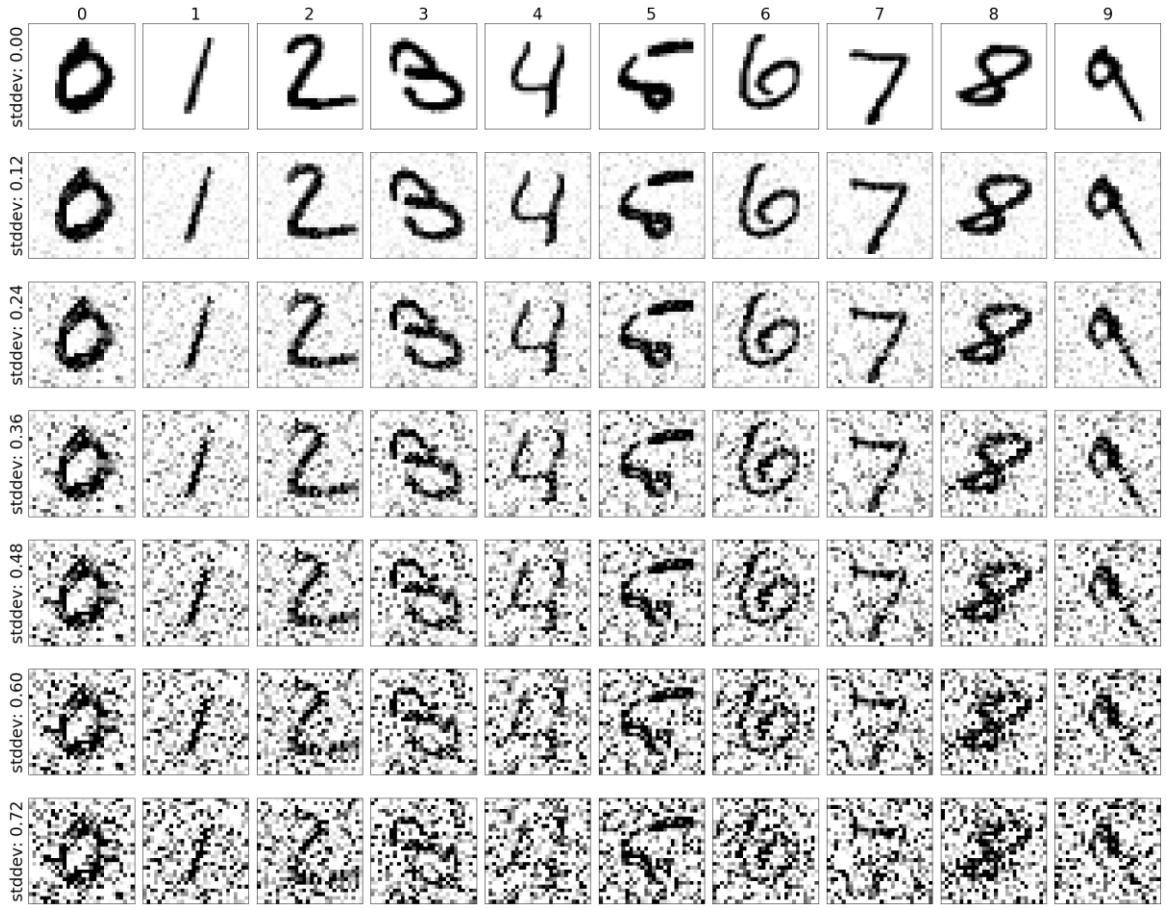


Fig. S3 | Various levels of additive Gaussian noise on the test set of the MNIST. The Gaussian noise is used to simulate the appearance of snow on the test set, the Gaussian noise standard deviation is from 0.12 to 0.72. As noise characterized by a Gaussian distribution is added to examples of different images from the MNIST and F-MNIST dataset, the images become harder to distinguish.

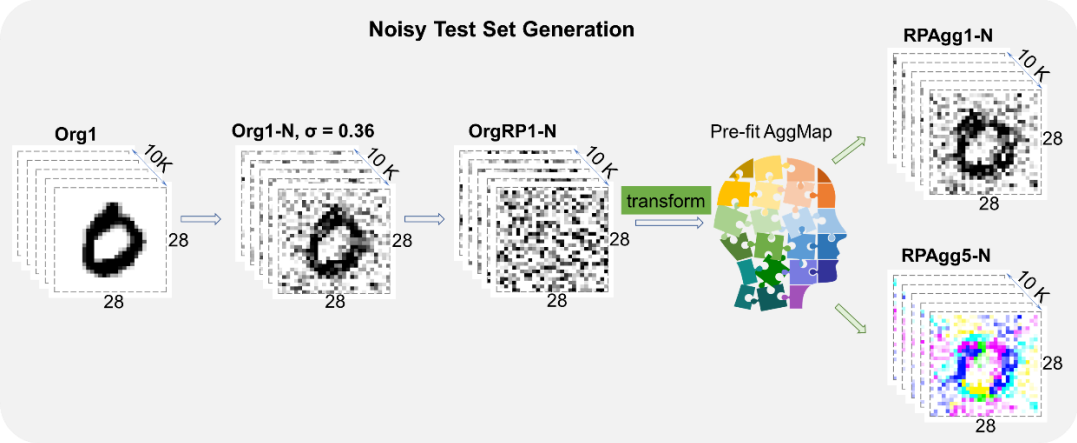


Fig. S4 | The noisy test set generation for the four Fmaps (Org1, OrgRP1, RPAgg1 and RPAgg5). First the various levels of Gaussian noise (the standard deviation 0.00 to 0.72 with a step of 0.12) were added to the **Org1** tests only (The Fmap values are divided by 255 to scale into 0~1), which is to generate the **Org1-N** set, then **Org1-N** Fmaps were further randomly permuted into **OrgRP1-N** using the same random seed as the **OrgRP1** generation. After that, the **OrgRP1-N** Fmaps were transformed into noisy set of **RPAgg1-N** and **RPAgg5-N** by the pre-fit AggMap, the pre-fit AggMap transformation ability is almost not affected by the noise.

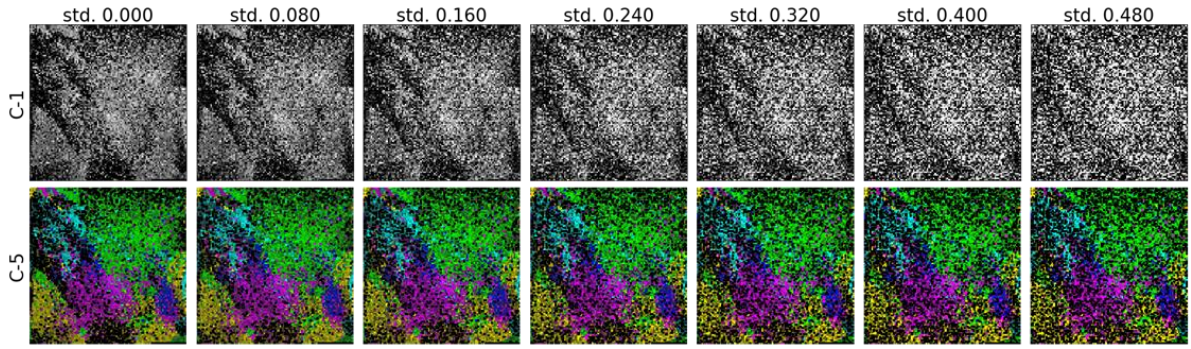


Fig. S5 | The noise-added Fmaps for TCGA-T dataset. The various levels of Gaussian noise (the standard deviation 0.00 to 0.48 with a step of 0.08) were added, C-1 is the Fmap with 1 channel, C-5 is the Fmap with 5 channels.

A

B

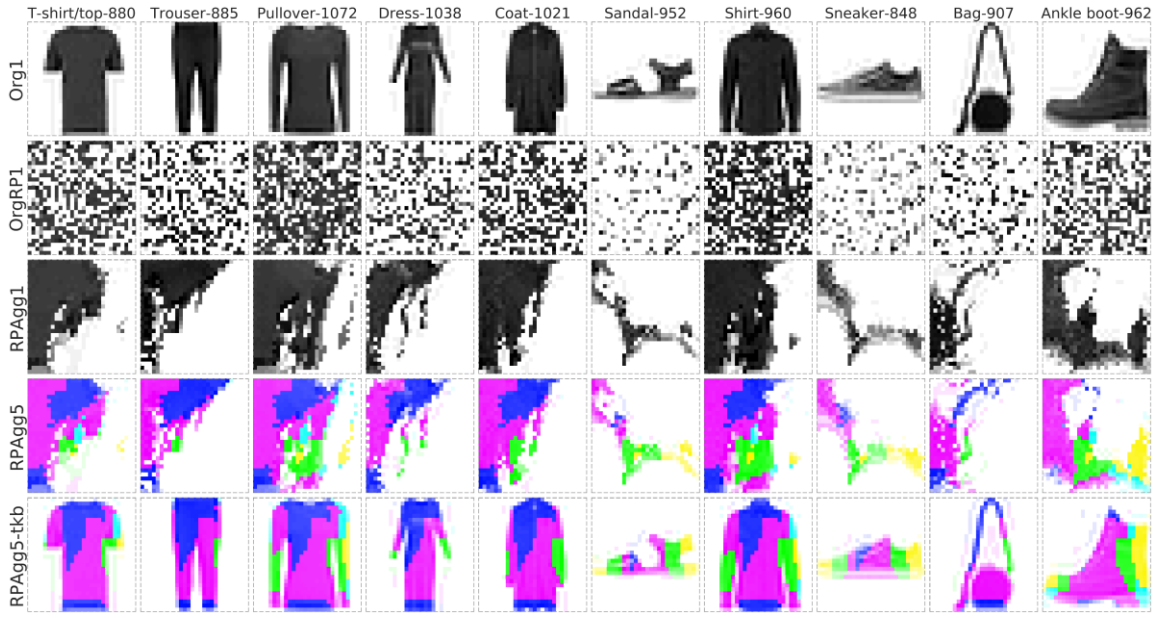


Fig. S6 | AggMap feature restructuring results on random permutated F-MNIST data. A, AggMap pre-fit with a different number of random permuted images to reconstruct the F-MINST images (RPAgg1). The all

(60K), 1/2, 1/5, 1/10, 1/100, and 1/1000 of the randomly permuted F-MNIST training set OrgRP1 were used for pre-fitting by AggMap, which was used for the reconstruction of the randomized F-MNIST test set. **B,** the original (Org1), randomly permuted (OrgRP1), and restructured (RPAgg1 and RPAgg5) F-MNIST data. RPAgg5-tkb: the original images with the pixels divided into 5 groups according to the 5-channels of RPAgg5 and colored in the same way as RPAgg5.

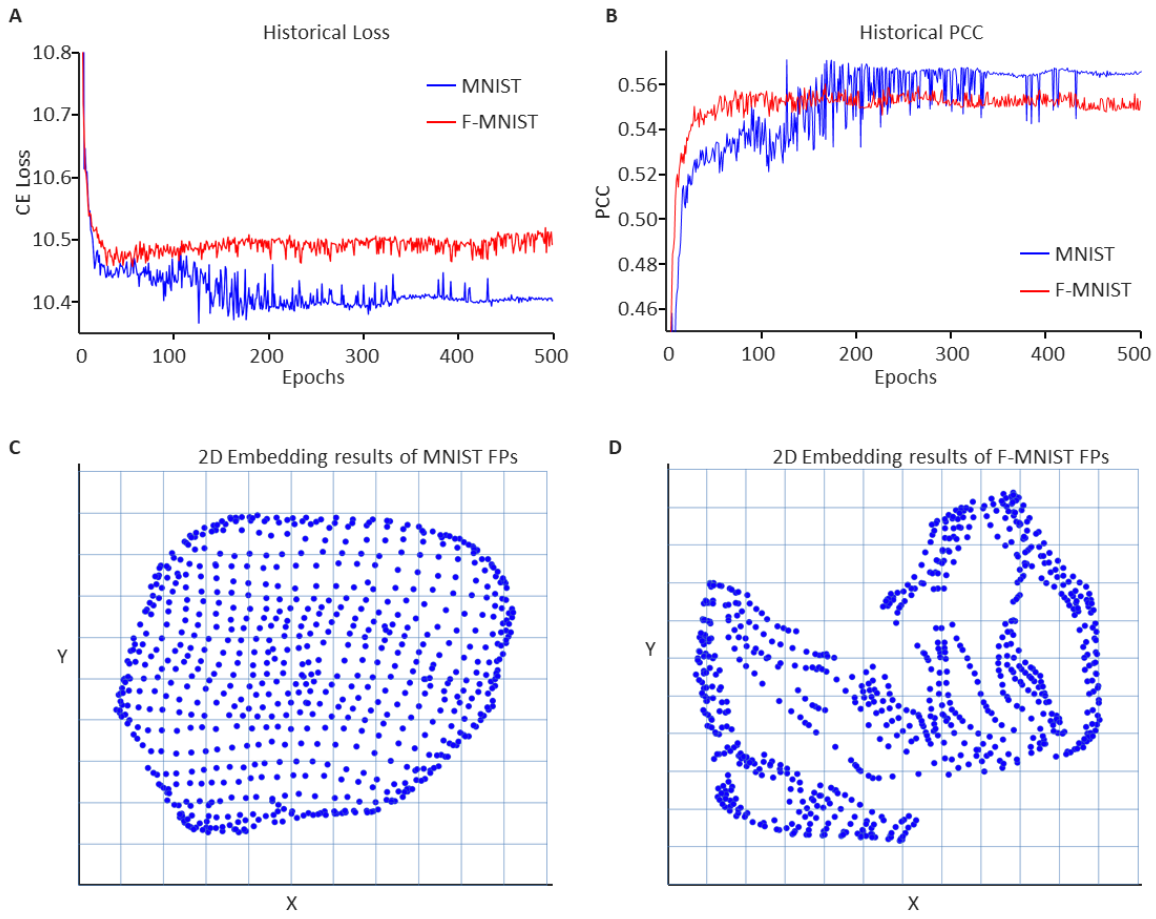
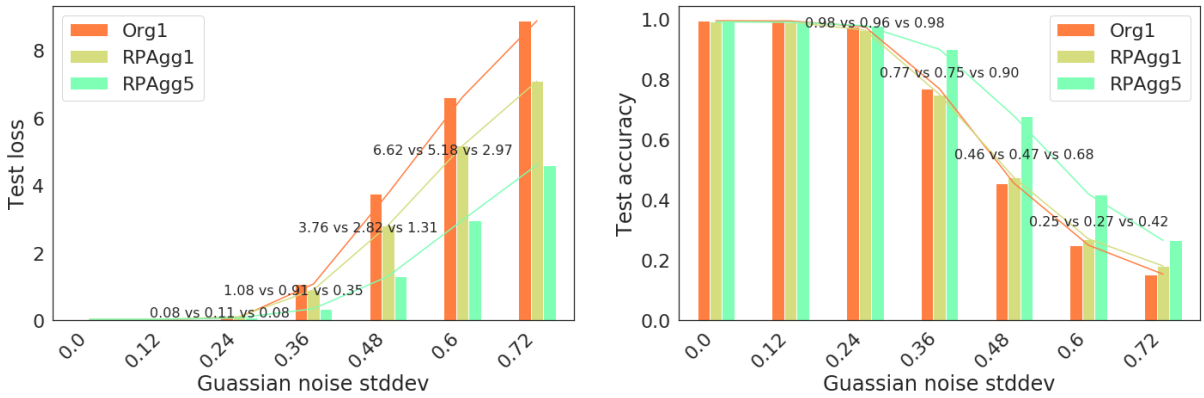


Fig. S7 | AggMap fitting historical performances and final 2D embedding results on randomly permuted MNIST and F-MNIST. A, the historical performance of cross-entropy (CE) loss (**Eqn. 9**) for randomly permuted MNIST and F-MNIST. **B,** the historical performance of PCC metric (**Eqn. 10**) for randomly permuted MNIST and F-MNIST. The dynamic process of MNIST and F-MNIST restructuring from randomly permuted images with 500 epochs is in **Video_MNIST.mp4** and **Video_F-MNIST.mp4**, respectively. **C,** the final 2D embedding results for the randomly permuted MNIST FPs. **D,** the final 2D embedding results for the randomly permuted F-MNIST FPs

A. Robustness of AggMapNet performance on noise-added MNIST



B. Robustness of AggMapNet performance on noise-added F-MNIST

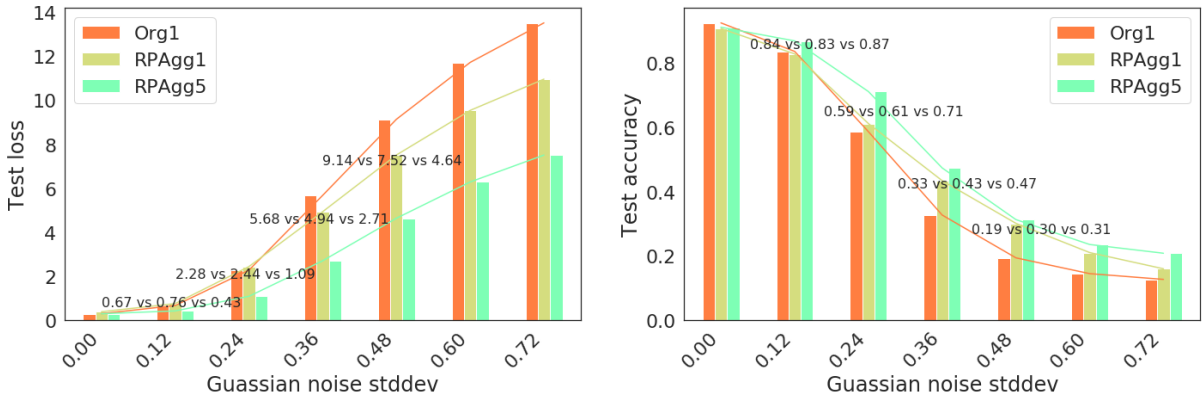


Fig. S8 | Robustness of AggMapNet classification performance on noise-added test set of MNIST and F-MNIST. These models that trained on the Org1, OrgRP1, RPAgg1 and RPAgg5 Fmaps are evaluated on the test set with different noise levels.

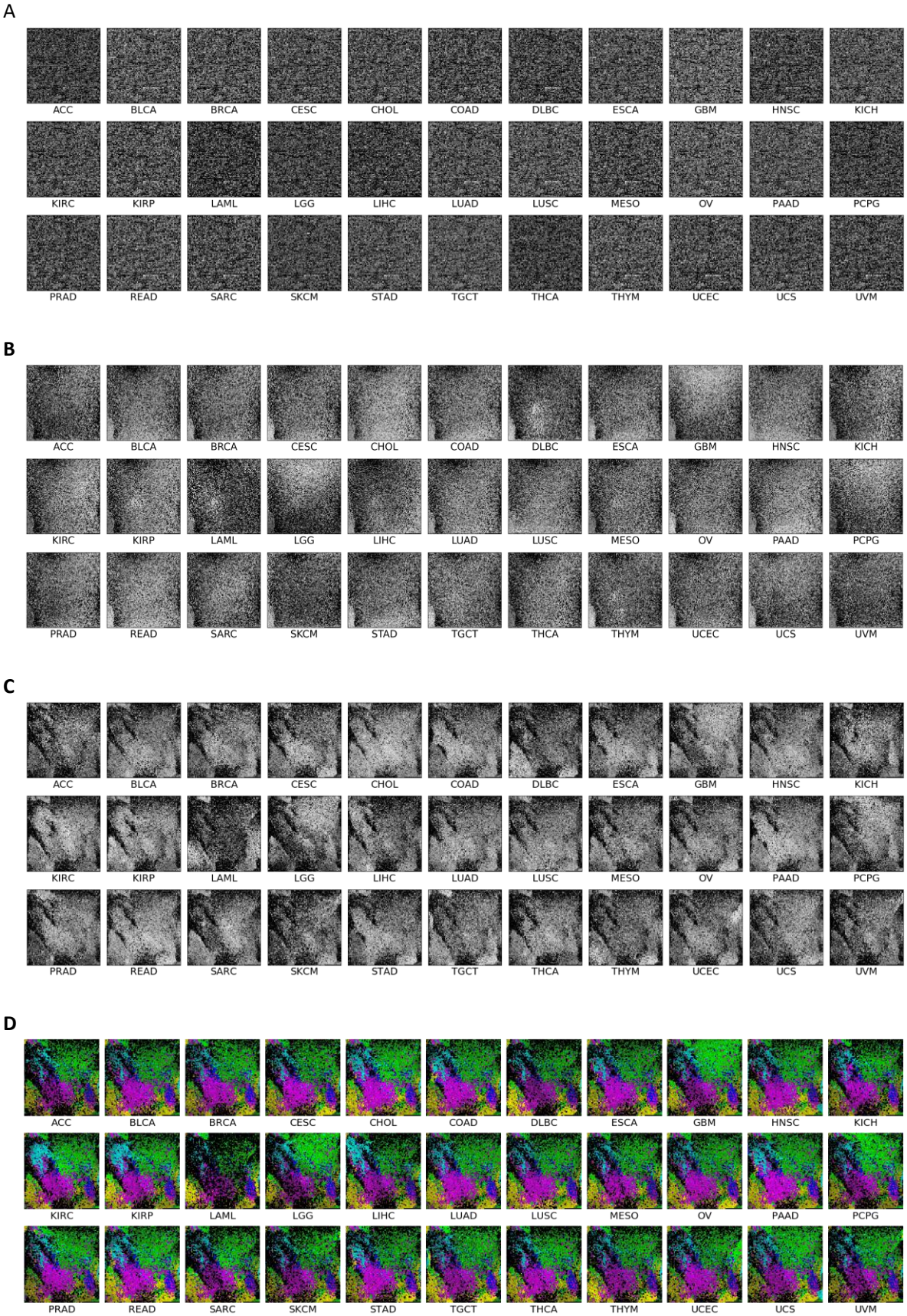


Fig. S9 | A comparison of the Fmaps of 33 cancers of the TCGA-T dataset generated by Lyu and Haque's study(1) and AggMap. A, Example of direct reshaped feature maps, the images are taken from Lyu and

Haque's study(1): https://drive.google.com/file/d/1zUepILj0is71LxPAWAZKmJ7-Kk7L9_XO/view**. B,** Example of REFINED Fmaps for the restructuring of 33 cancers of the TCGA-T dataset (102,102, 1)**. C,** Example of single-channel AggMap Fmaps for the restructuring of 33 cancers of the TCGA-T dataset (102,102, 1)**. D,** Example of multi-channel AggMap Fmaps for the restructuring of 33 cancers of the TCGA-T dataset (102,102, 5)

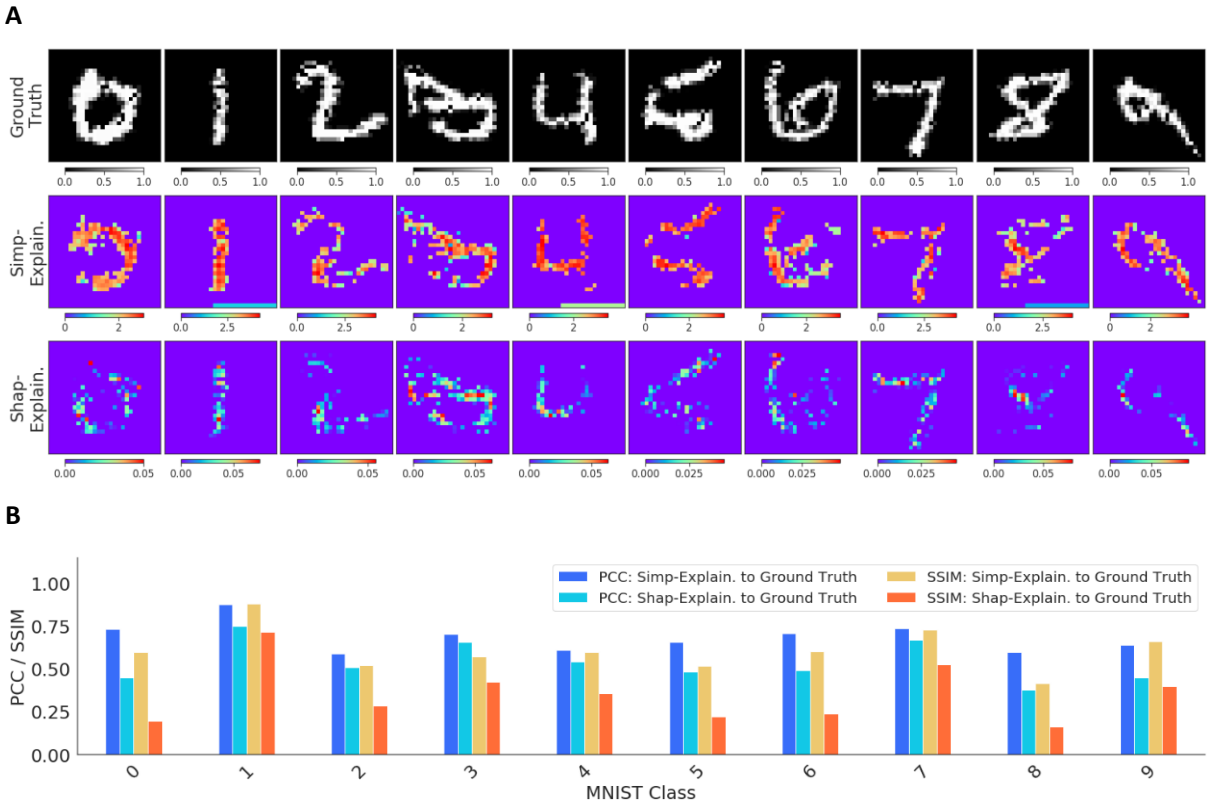


Fig. S10 | Comparison of the Simply-explainer and Shapley-explainer on the noise-free MNIST recognition model explanation. A, the ground truth MNIST images, and the interpretation saliency-map images that are generated by Simply-explainer and Shapley-explainer from ground truth images. **B,** the Pearson's correlation coefficient (PCC) and structure similarity index (SSIM) values between the ground truth images and the interpreted saliency-map images for the two explainers.

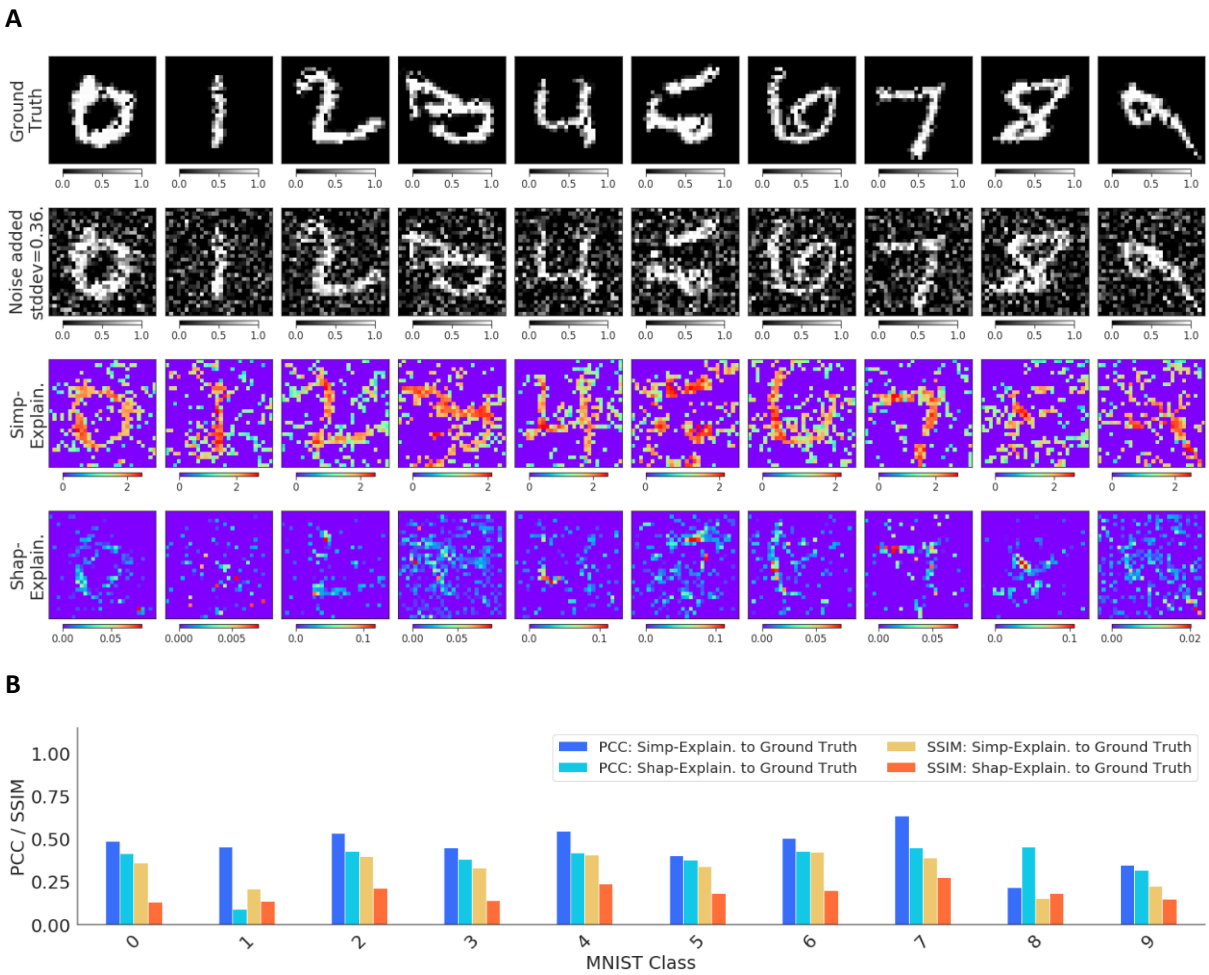


Fig. S11 | Comparison of the Simply-explainer and Shapley-explainer on the noise MNIST recognition model explanation. A, the ground truth MNIST images, the noise-added images (stddev=0.36), and the interpretation saliency-map images that are generated by Simply-explainer and Shapley-explainer from noise-added images. **B,** the Pearson's correlation coefficient (PCC) and structure similarity index (SSIM) values between the ground truth images and the interpreted saliency-map images for the two explainers.

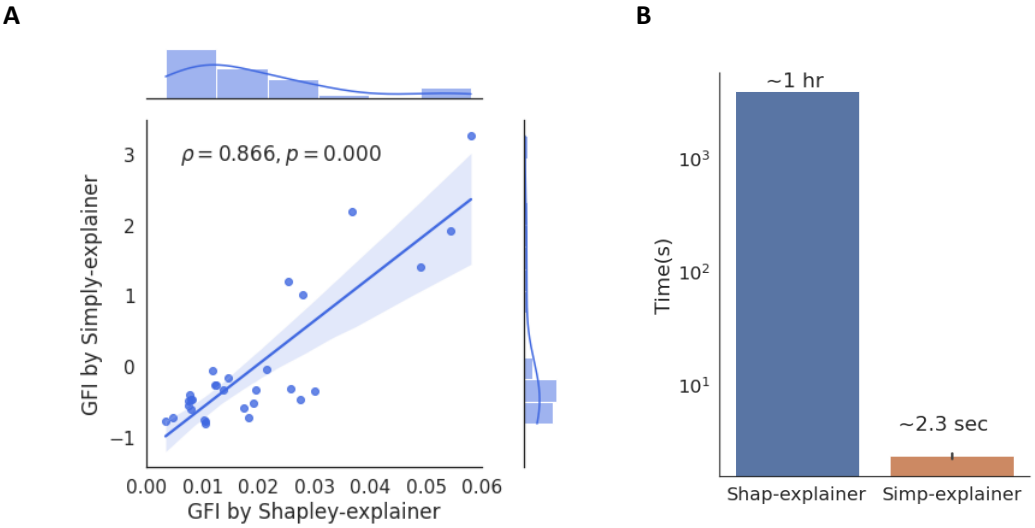


Fig. S12 | Comparison of the Simply-explainer and Shapley-explainer on global explanation of breast cancer diagnostic model trained by WDBC dataset(10). A, the joint scatter plot of the global feature importance (GFI) calculated by Simply-explainer and Shapley-explainer. **B,** the time used for the Simply-explainer and Shapley-explainer in the calculation of GFI. Computational complexity for Simply-explainer is **O(n)**, while the complexity for kernel Shapley-explainer is **O(m*l*(2n+2048))**, where l is the number of background samples, n is number of features and m is number of samples

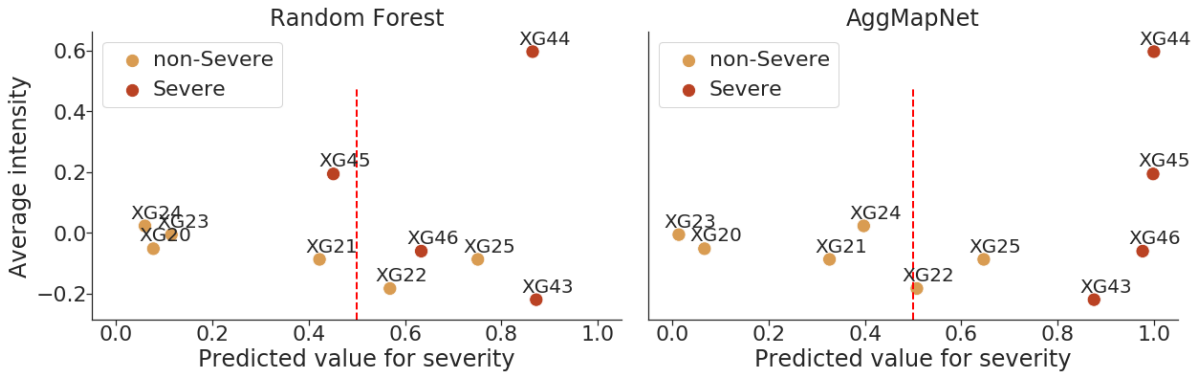


Fig. S13 | Comparison of the predicted value on the independent cohorts for the RF and AggMapNet classification of the COV-S dataset. The RF prediction results are taken from Shen et al., 2020(11). The non-severe patient XG22 had chronic hepatitis B virus (HBV) infection, diabetes, and the longest hospitalization (>50 days) among all non-severe patients, the 43-year-old male non-severe case XG25 was incorrectly classified as severe for reasons unclear(11). Compared with RF classifier, AggMapNet can predict XG22 with relatively lower probability to be severe, and can patient XG45 correctly although he had received traditional Chinese medicines for more than 20 days before admission(11).

Reference

1. Lyu, B. and Haque, A. (2018) Deep learning based tumor type classification using gene expression data. *Proceedings of the 2018 ACM international conference on bioinformatics, computational biology, and health informatics (ACM-BCB)*, 89-96.
2. Bazgir, O., Zhang, R., Dhruba, S.R., Rahman, R., Ghosh, S. and Pal, R. (2020) Representation of features as images with neighborhood dependencies for compatibility with convolutional neural networks. *Nat. Commun.*, **11**, 1-13.
3. Smith, A.M., Walsh, J.R., Long, J., Davis, C.B., Henstock, P., Hodge, M.R., Maciejewski, M., Mu, X.J., Ra, S. and Zhao, S. (2020) Standard machine learning approaches outperform deep representation learning on phenotype prediction from transcriptomics data. *BMC Bioinform.*, **21**, 1-18.
4. Ying, K., Zhai, R., Pyrkov, T.V., Mariotti, M., Fedichev, P.O., Shen, X. and Gladyshev, V.N. (2020) Genetic and phenotypic evidence for the causal relationship between aging and covid-19. *medRxiv*.
5. Ueland, T., Holter, J., Holten, A., Müller, K., Lind, A., Bekken, G., Dudmann, S., Aukrust, P., Dyrhol-Riise, A. and Heggelund, L. (2020) Distinct and early increase in circulating MMP-9 in COVID-19 patients with respiratory failure. *J. Infect.*
6. Zhang, J., Saad, R., Taylor, E.W. and Rayman, M.P. (2020) Selenium and selenoproteins in viral infection with potential relevance to COVID-19. *Redox Biol.*, 101715.
7. Heller, R.A., Sun, Q., Hackler, J., Seelig, J., Seibert, L., Cherkezov, A., Minich, W.B., Seemann, P., Diegmann, J. and Pilz, M. (2020) Prediction of survival odds in COVID-19 by zinc, age and selenoprotein P as composite biomarker. *Redox Biol.*, **38**, 101764.
8. Cauchois, R., Koubi, M., Delarbre, D., Manet, C., Carvelli, J., Blasco, V.B., Jean, R., Fouche, L., Bornet, C. and Pauly, V. (2020) Early IL-1 receptor blockade in severe inflammatory respiratory failure complicating COVID-19. *Proc. Natl. Acad. Sci. U.S.A.*, **117**, 18951-18953.
9. Abouhashem, A.S., Singh, K., Azzazy, H.M. and Sen, C.K. (2020) Is low alveolar type II cell SOD3 in the lungs of elderly linked to the observed severity of COVID-19? *Antioxid. Redox Signal.*, **33**, 59–65.
10. Dua, D. and Graff, C. (2019) UCI machine learning repository, Wisconsin Diagnostic Breast Cancer (WDBC) Data Set, **37**.
11. Shen, B., Yi, X., Sun, Y., Bi, X., Du, J., Zhang, C., Quan, S., Zhang, F., Sun, R. and Qian, L. (2020) Proteomic and metabolomic characterization of COVID-19 patient sera. *Cell*, **182**, 59-72. e15.