Remote Sensing Time Series as Covariance Matrices for Crop Classification

Master Thesis Defense

Khizer Zakir

26/06/2024

With the support of the Erasmus+ Programme of the European Union

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

04 DATA/DATA-PREPARATION

05 EXPERIMENTAL SETTING

06 RESULTS

07 CONCLUSION & FUTURE PERSPECTIVES

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

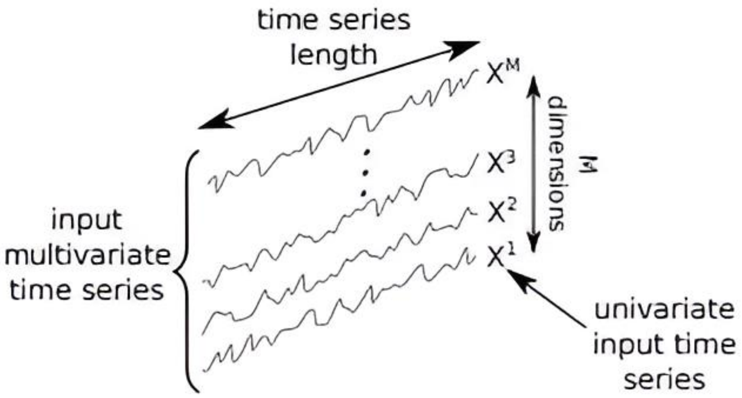
04 DATA/DATA-PREPARATION

05 EXPERIMENTAL SETTING

06 RESULTS

07 CONCLUSION & FUTURE PERSPECTIVES

Introduction



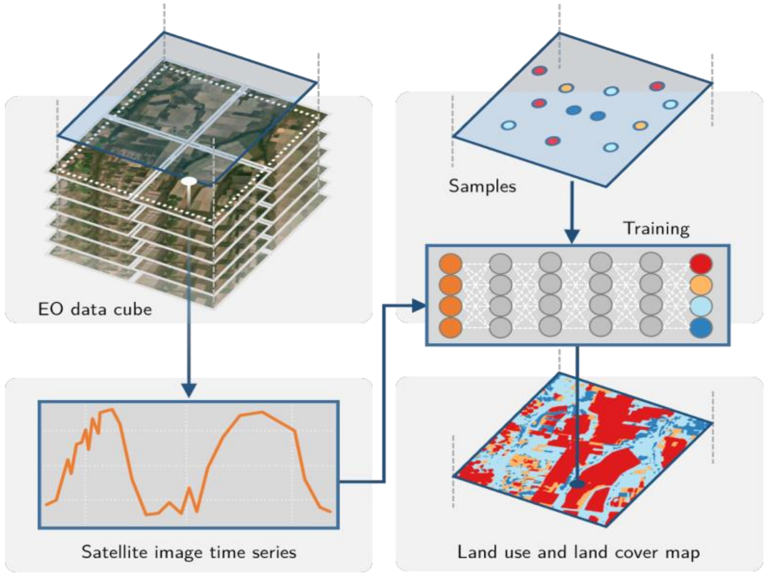
Representation of multi-variate TS. Source: Akodad, S., 2021

$$\{(\boldsymbol{X}_1, y_1), (\boldsymbol{X}_2, y_2), \ldots, (\boldsymbol{X}_n, y_n)\}$$

$$X_k = (x_i^1, x_i^2, \ldots, x_i^T)_{i=1}^q$$

- Large timeseries (TS) of **multi-spectral** images.
- Captures **spectral** information across multiple wavelengths across a finite **temporal** sequence.

Supervised learning task for crop classification using remote sensing TS data through second-order statistics, specifically **covariance matrices**.



Example of SITS. Source: Intro to SITS

Motivations

In remote sensing TS, we have **high-dimensional data.**

With long sequences, we get **large data sizes.**

Limited labeled sample, and **irregular temporal sampling**, leading to **missing data**, particularly due to **cloud cover**.

Reliance on **first-order statistics**, which address the **global structure** but fail to capture **local intricacies**.

Objectives

Following the classical manifold learning theory, **learning** or **preserving** the original data structure can enhance classification performance (Huang, Z. and Gool, L. V., 2016). We focus on the following research objectives:

1. Explore the unique **covariance representation** of remote sensing timeseries data to capture the **spectral-temporal** intricacies and dependencies.
2. Leveraging the underlying **Riemannian structure** of the data and engineer a state-of-the-art supervised learning algorithm.
3. Explore possible **variants** of the model with combinations of data representation.
4. Utilize **benchmark** datasets to evaluate the performance of the proposed models.

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

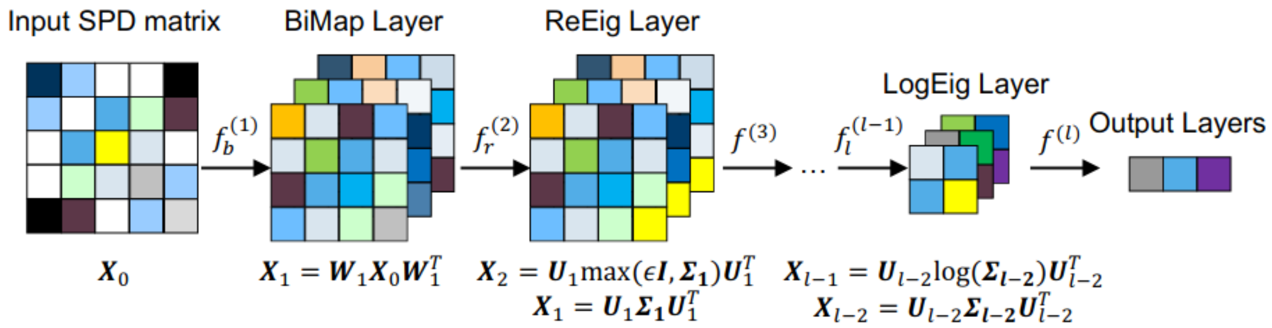
04 DATA/DATA-PREPARATION

05 EXPERIMENTAL SETTING

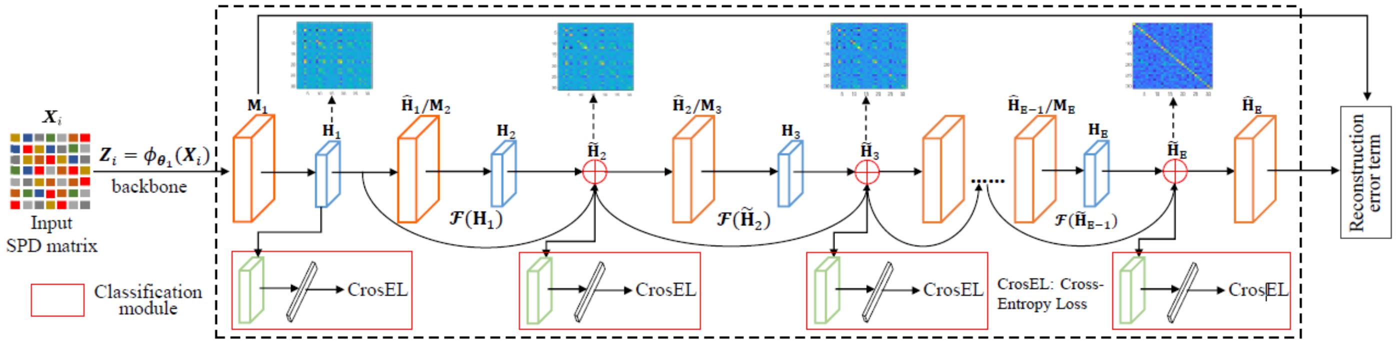
06 RESULTS

07 CONCLUSION & FUTURE PERSPECTIVES

Riemannian Networks



Source: (Zhiwu Huang and Luc Van Gool, 2017)



Source: (Wang et.al, 2022)

SPDNet: A Riemannian Network

- SPDNet works on SPD matrices directly and exploits multiple layers tailored for SPD matrix deep learning.
- Our implementation is based on the foundational concept of this model.

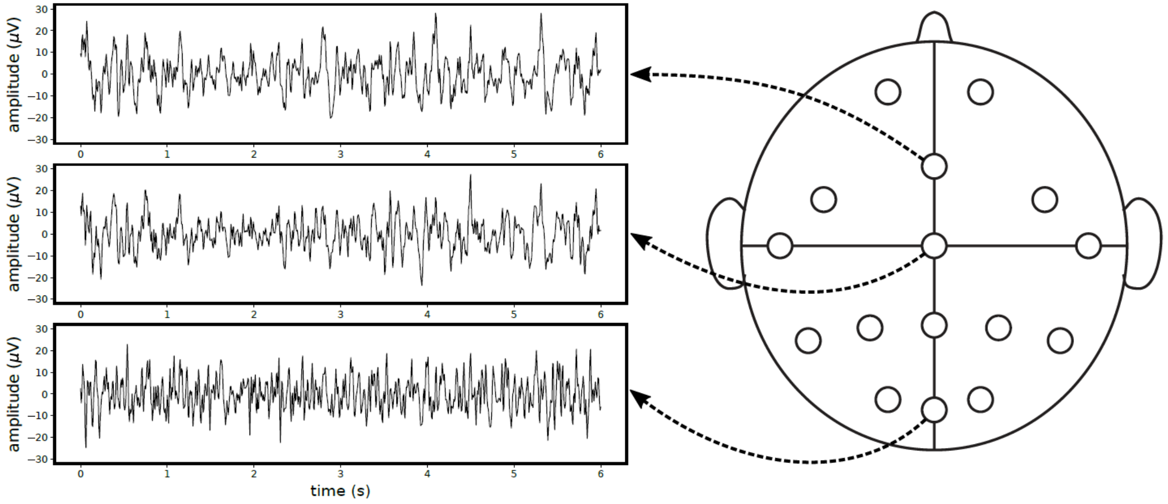
DreamNet: A Stacked Riemannian Autoencoder (SRAE) Network

- DreamNet implants several residual-like blocks using shortcut connections to augment the representational capacity of SRAE.

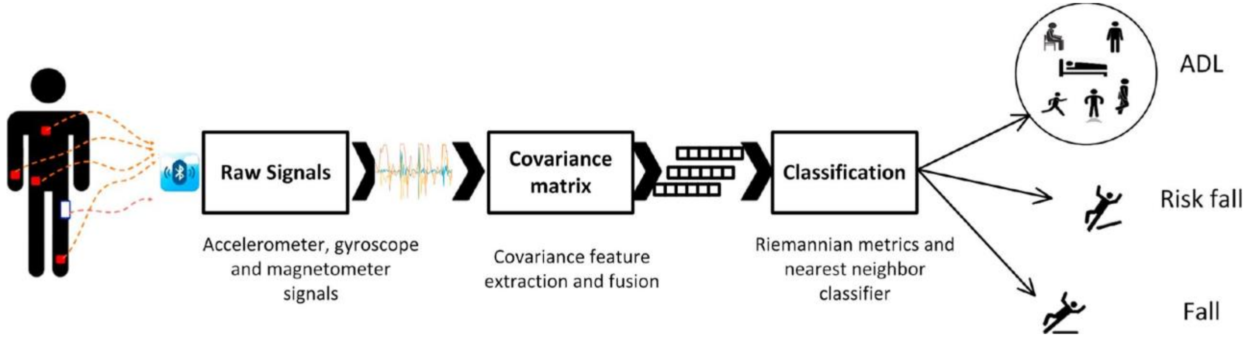
02

LITERATURE

Applications of Riemannian Approach



EEG classification. Source: (Pedro L. C. Rodrigues, 2019)



Fall detection from multiple wearable sensors. Source: (Bouetella, 2019)

- Brain Computer Interface (BCI) classification [1]
- Radar signals classification [2]
- Text, audio, and video classification [3]
- Stock market-state classification [4]

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

04 DATA/DATA-PREPARATION

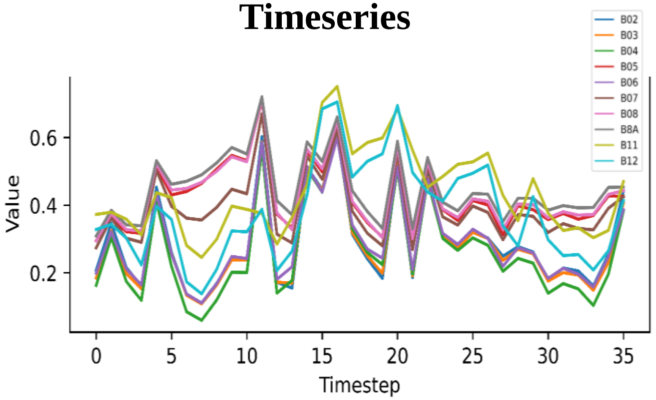
05 EXPERIMENTAL SETTING

06 RESULTS

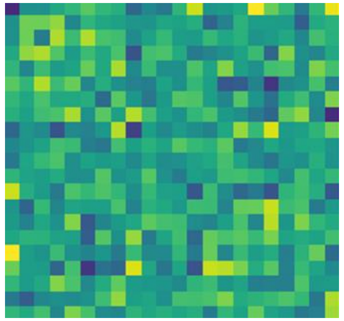
07 CONCLUSION & FUTURE PERSPECTIVES

Methodology

Timeseries



Covariance



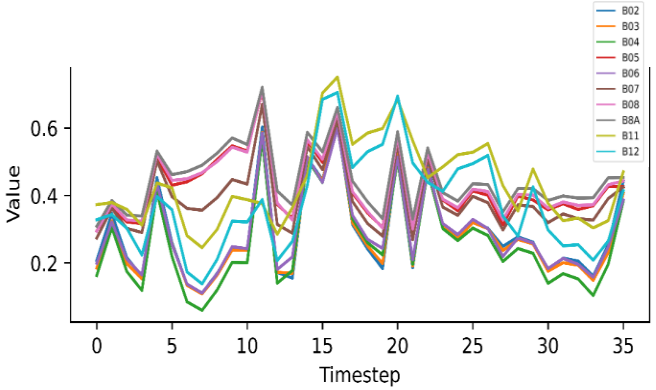
Processing

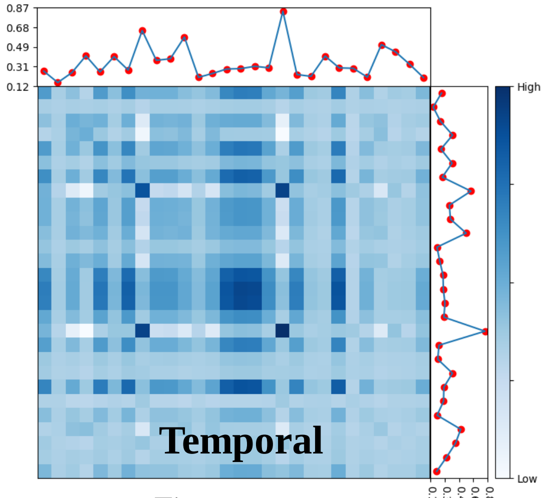
LERN MODEL

Crop class

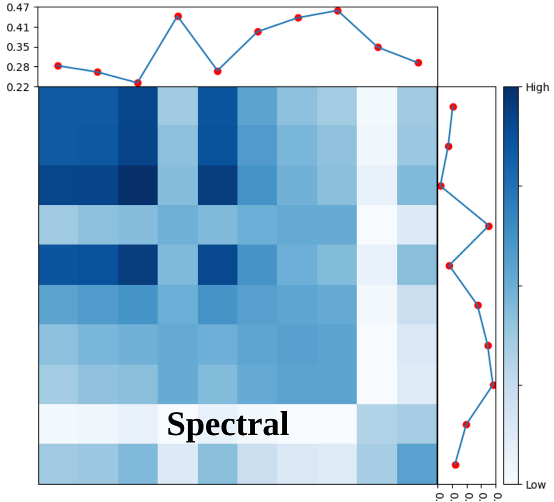
Winter Wheat
Source: [Dr. Erick J. Larson, 2013]

Covariance Estimation





Time sequence



Spectral bands

Sample Covariance Matrix

$$\boldsymbol{P} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i - \bar{\boldsymbol{X}})(\boldsymbol{X}_i - \bar{\boldsymbol{X}})^T$$

$$\boldsymbol{P} = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{X}_i)(\boldsymbol{X}_i)^T$$

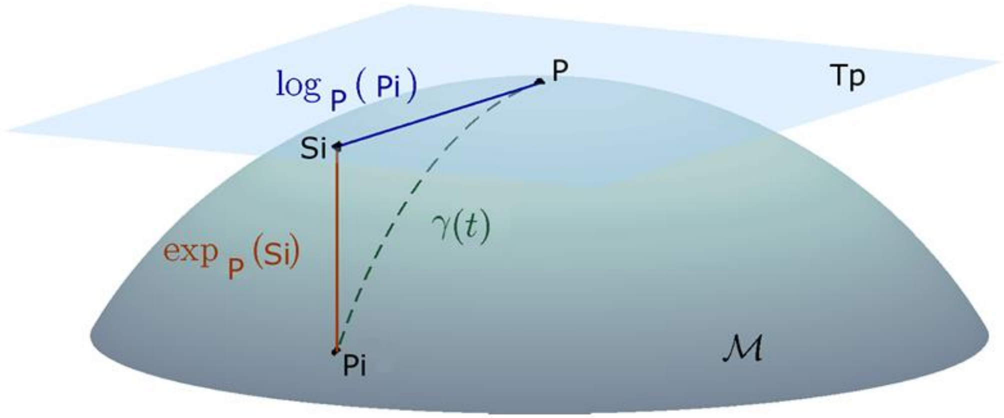
Regularization

$$\boldsymbol{P}_{reg} = \boldsymbol{P} + \epsilon \boldsymbol{I}$$

$$\boldsymbol{P}_{shrink} = (1-\alpha)\boldsymbol{P} + \alpha \boldsymbol{I}$$

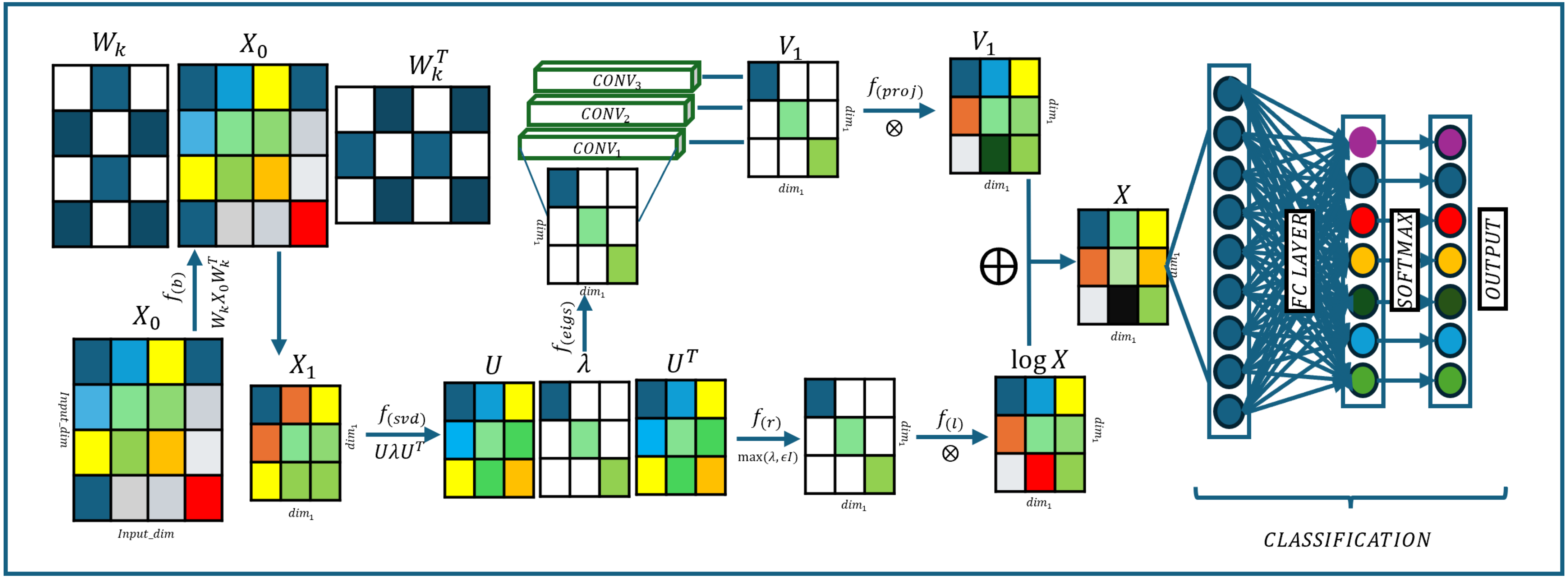
Covariance Matrix → SPD Matrix

- **Symmetric Positive Definite (SPD)**
  - $\boldsymbol{P} = \boldsymbol{P}^T$
  - Each $\boldsymbol{P}$ satisfies $x^T\boldsymbol{P}x > 0$ for all non-zero vectors $x \in \mathbb{R}^n$
- A **Riemannian manifold** $\mathcal{M}$ is a differentiable manifold, where the **tangent space**⇒ $T_P\mathcal{M}$ at each point P is a finite-dimensional Euclidean space.
- **SPD manifold** $S_+^d \in \mathcal{M}$ → $\boldsymbol{tangent\ space}$ ⇒ $T_P S_+^d$
- **Riemannian metric** ⇒ inner product: $g_p: T_p S_+^d \times T_p S_+^d \to \mathbb{R}$



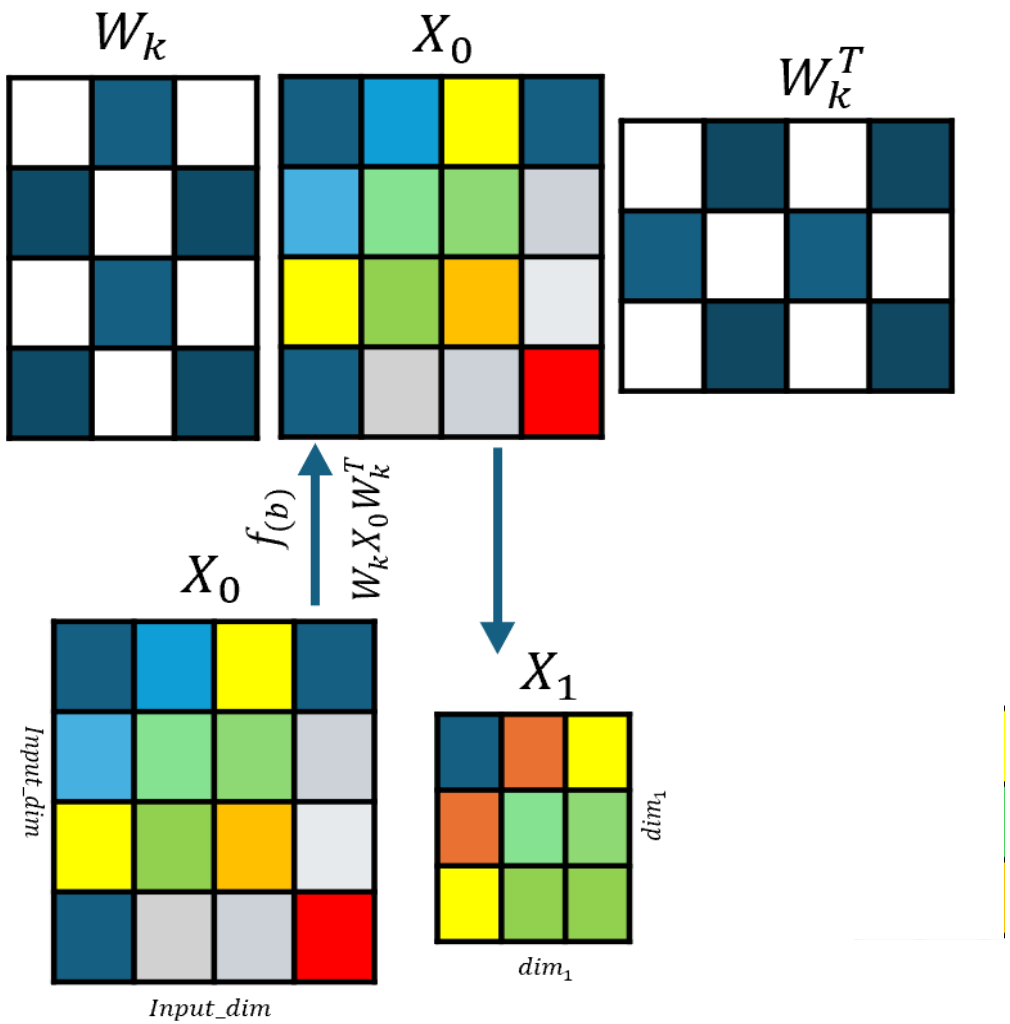
This represents the tangent space of the SPD manifold S_+^d at point P, where S_i the tangent vector of P_i and $\gamma(t)$ is the geodesic between P and P_i. Source: Barachant, A. et.al.

LogEucResNet (LERN) Model



Conceptual Diagram for LERN

First Block



BiMap Layer

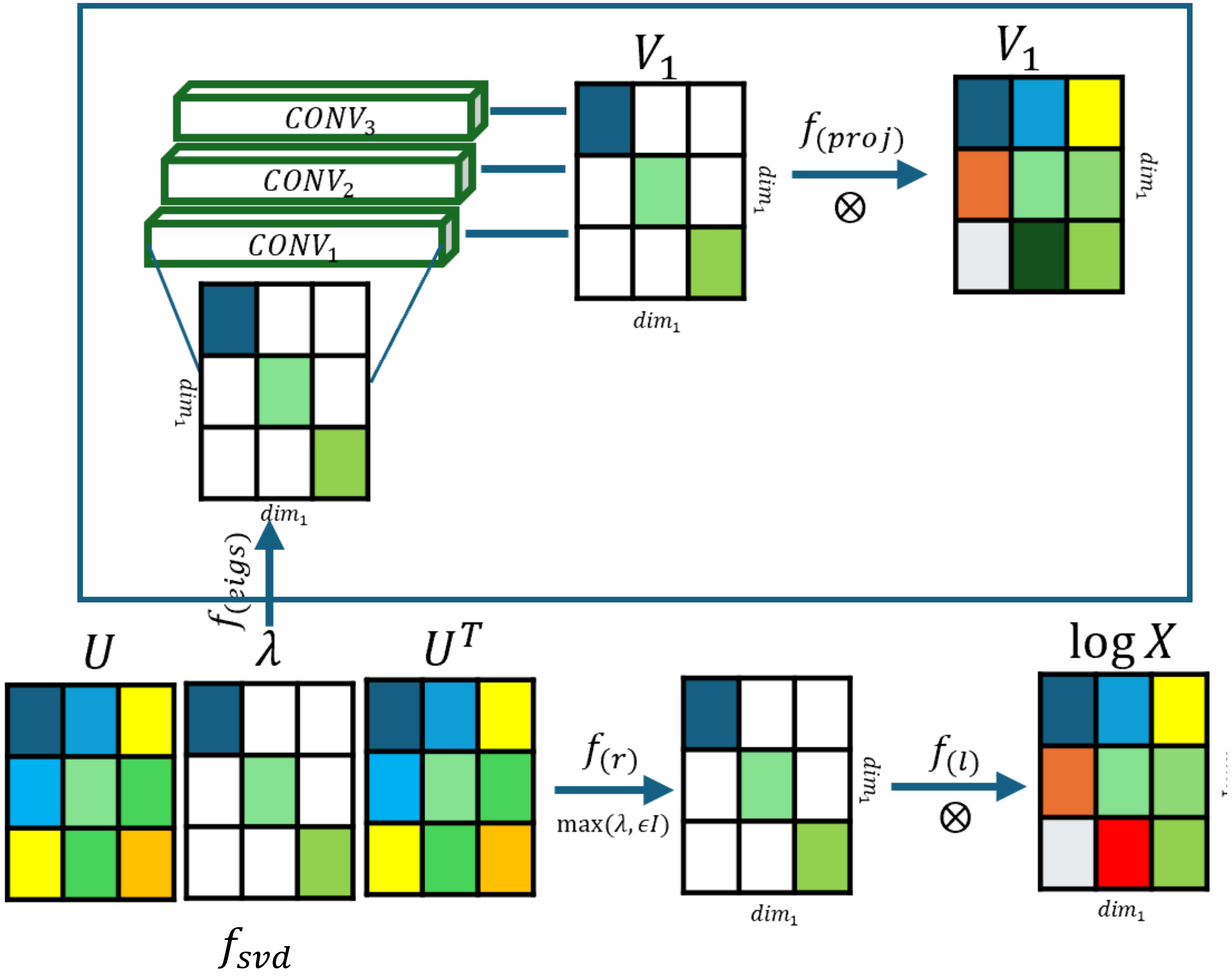
$$\boldsymbol{X}_k = f_{(b)}^{k}(\boldsymbol{X}_{k-1}, \boldsymbol{W}_k) = \boldsymbol{W}_k \boldsymbol{X}_{k-1} \boldsymbol{W}_k^T$$

- Input SPD matrices to new SPD matrices in a **sub-space.**
- **Reduce** dimension while preserving the **geometric structure** of the data.

Model Hyperparameter ⚠

$$\mathrm{dim}_1 = Bimap\ dimension$$

Second Block (a)



Single-Value Decomposition (SVD)

$$\boldsymbol{X}_k = f_{(svd)}^k (\boldsymbol{X}_{k-1}) = \boldsymbol{U}_{k-1} \boldsymbol{\lambda}_{k-1} \boldsymbol{U}_{k-1}^T$$

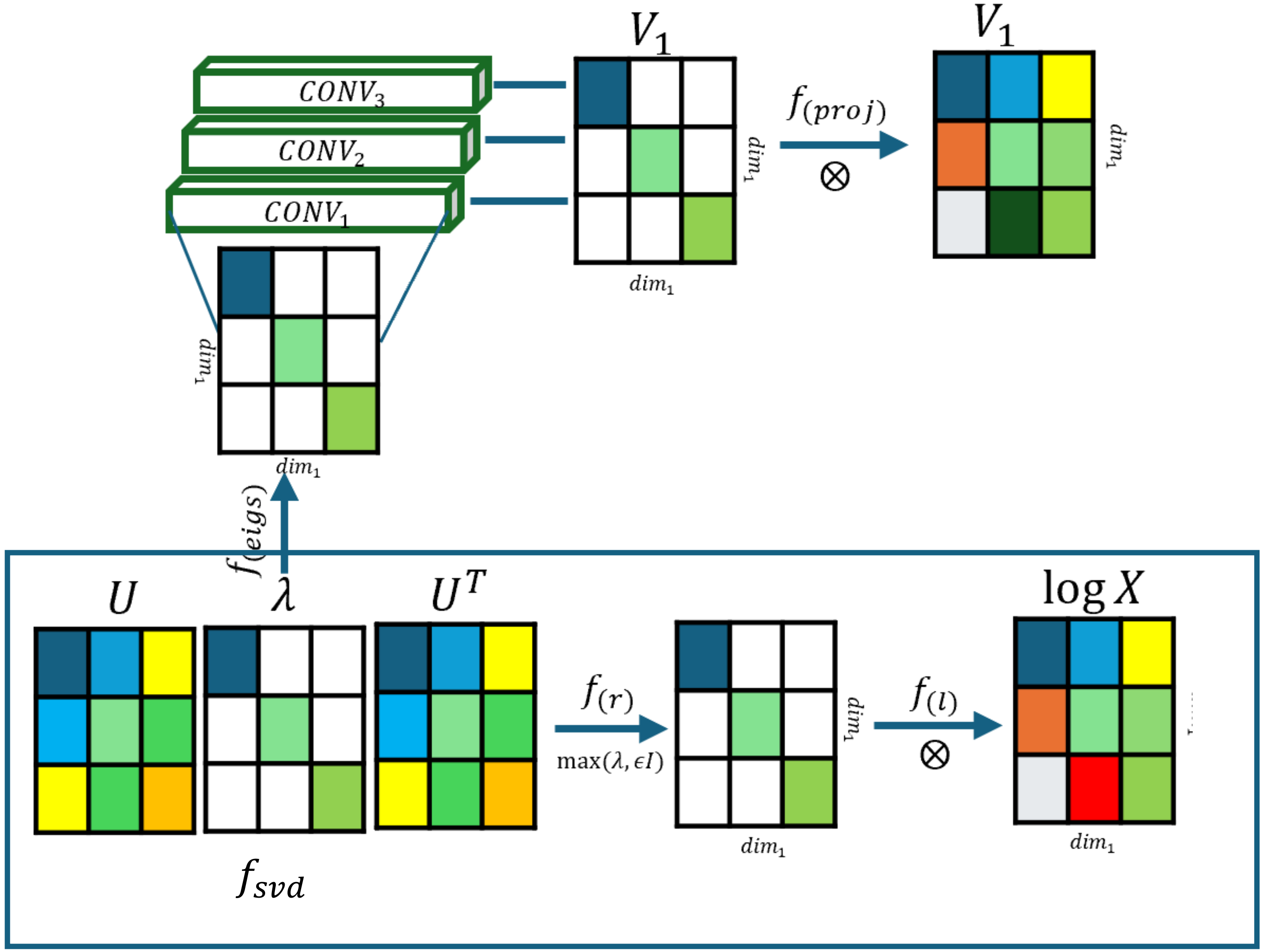
Spectrum Map

$$\widehat{\boldsymbol{v}} = f_{(eigs)}^k(\boldsymbol{\lambda}_{k-1})$$

Projection

$$\boldsymbol{v} = f_{(proj)}^k(\widehat{\boldsymbol{v}}) = \boldsymbol{P}\, \widehat{\boldsymbol{v}}\, \boldsymbol{P}^T$$

Second Block (b)



Single-Value Decomposition (SVD)

$$\boldsymbol{X}_k = f_{(svd)}^{k}(\boldsymbol{X}_{k-1}) = \boldsymbol{U}_{k-1}\,\boldsymbol{\lambda}_{k-1}\,\boldsymbol{U}_{k-1}^{T}$$

ReEig Layer

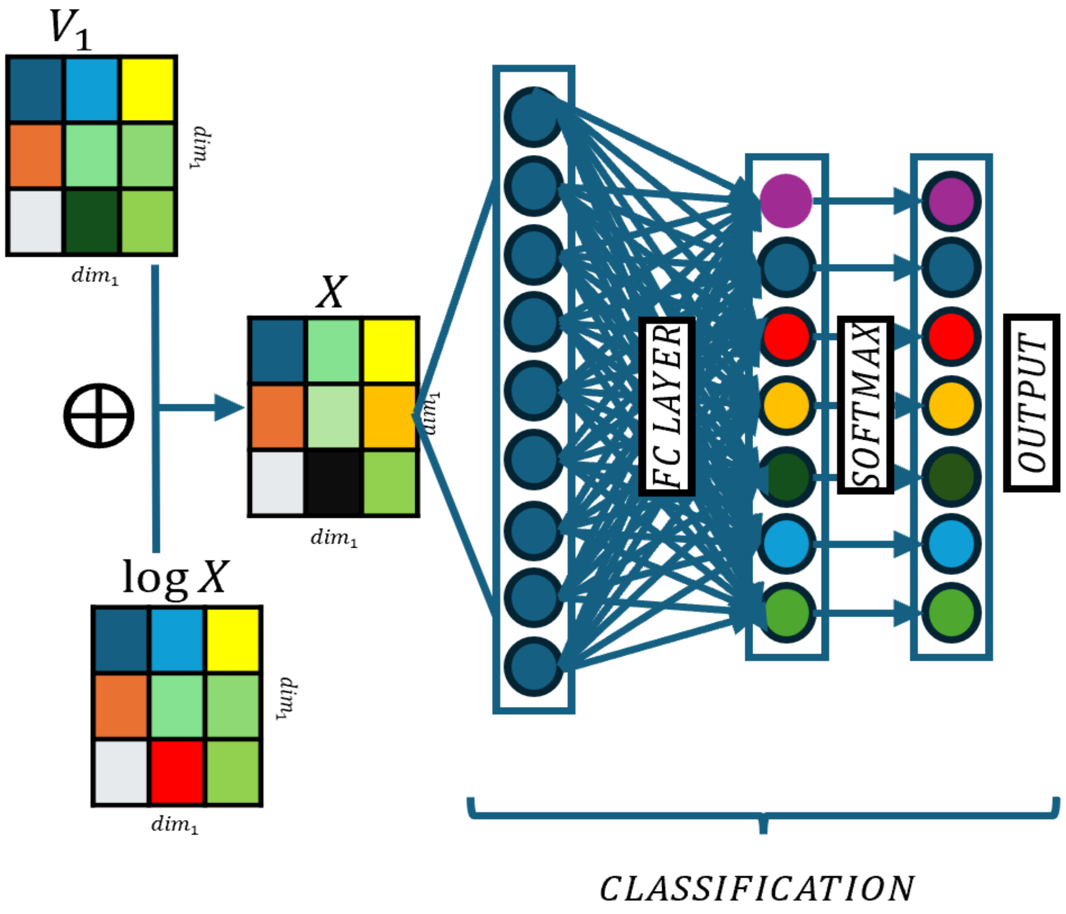
$$\boldsymbol{X}_k = f_{(r)}^{k}(\boldsymbol{X}_{k-1}) = \boldsymbol{U}_{k-1}\,\boldsymbol{max}(\boldsymbol{\lambda}_{k-1}, \epsilon\boldsymbol{I})\,\boldsymbol{U}_{k-1}^{T}$$

$$\boldsymbol{max}(\boldsymbol{\lambda}_{k-1}, \epsilon\boldsymbol{I}) = \begin{cases} \boldsymbol{\lambda}_{k-1}(i,i) & if \quad \boldsymbol{\lambda}_{k-1}(i,i) > \epsilon \\ \epsilon & if \quad \boldsymbol{\lambda}_{k-1}(i,i) \leq \epsilon \end{cases}$$

LogEig

$$\boldsymbol{X}_k = f_{(l)}^{k}(\boldsymbol{X}_{k-1}) = \boldsymbol{U}_{k-1}\,log(\boldsymbol{\lambda}_{k-1})\,\boldsymbol{U}_{k-1}^{T}$$

Third Block



Residual Block

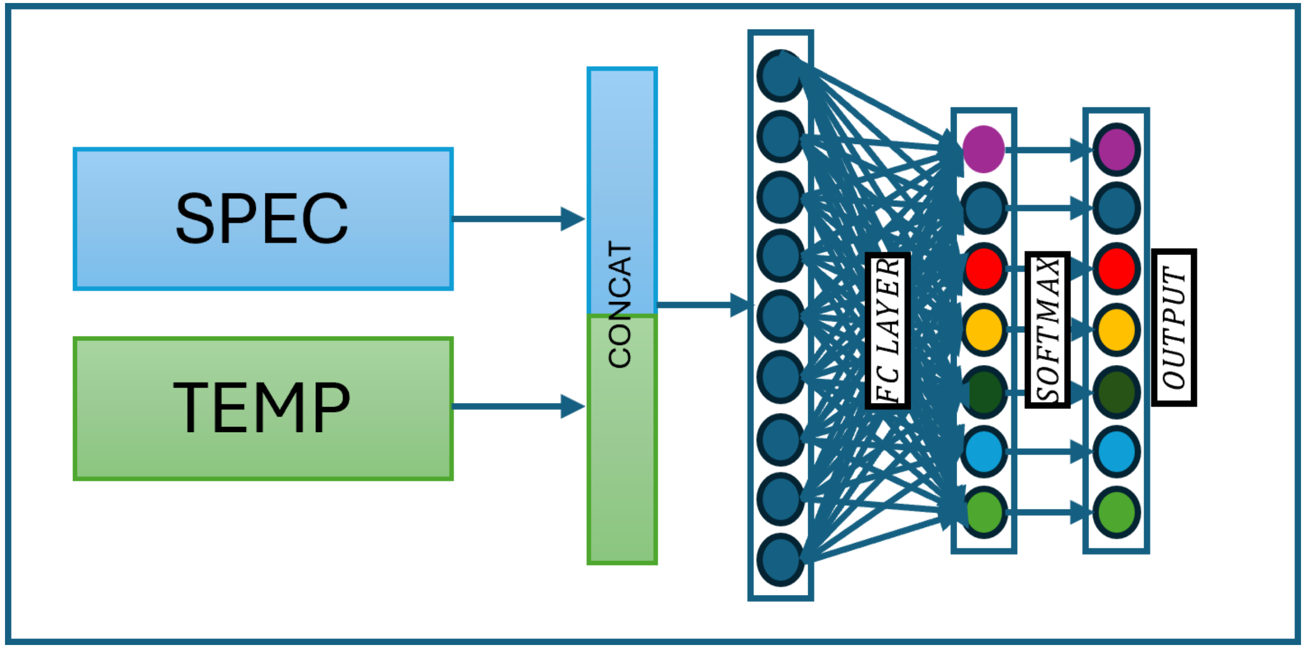
$$\boldsymbol{X} = \boldsymbol{X}_k \oplus \boldsymbol{v}$$

Classification Block

$$S_i = \log \boldsymbol{X} \cdot \text{proto}_i^T$$

$$D = \begin{bmatrix} S_1 \\ S_2 \\ \vdots \\ S_{n_{\text{proto}}} \end{bmatrix}$$

LERN - Combo



Conceptual Diagram for LERN-Combo

- This framework draws inspiration from the **Duplo** network by Interdonato, R., et.al (2019).
- **Fusing** two chains of networks OR sub-models
- Handle spectral and temporal covariance representations and captures distinct **patterns** and **features**

Loss Function

- **Focal Loss**

It was introduced by Lin et al. (2018), primarily designed for object detection to address class imbalance

$$\mathcal{L}_{\text{focal}} = -\alpha(1 - p_t)^{\gamma} \log(p_t)$$

α is **class weights**

γ is the **focusing parameter**

p_t is the model's **predicted probability**

Performance Metric

- **Weighted F-1 Score**

$$\text{F1-weighted} = \frac{1}{\sum_{i=1}^{n} N_i} \sum_{i=1}^{n} N_i \cdot \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

- **Macro F-1 Score**

$$\text{F1-macro} = \frac{1}{n} \sum_{i=1}^{n} \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

04 DATA/DATA-PREPARATION

05 EXPERIMENTAL SETTING

06 RESULTS

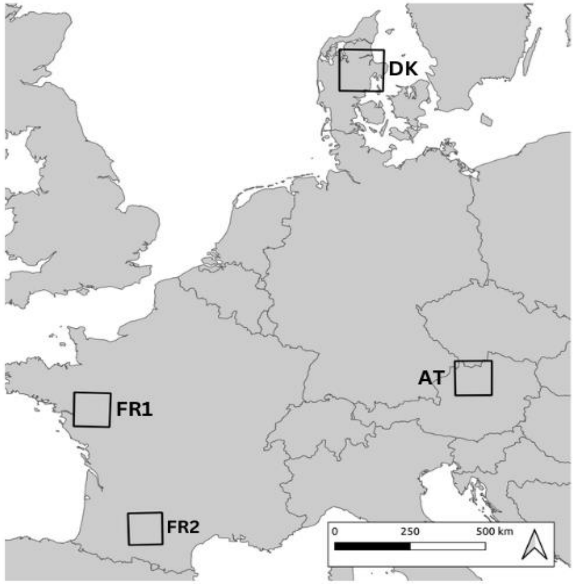
07 CONCLUSION & FUTURE PERSPECTIVES

Datasets

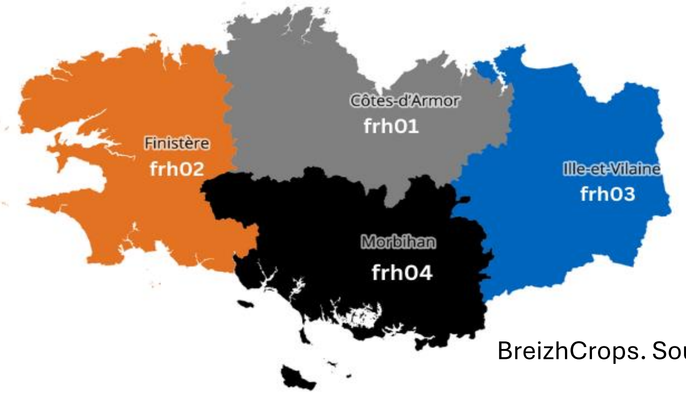
MiniTimeMatch → SITS for 4 regions across Europe
BreizhCrops → SITS for *Breizh* = *Bretagne* region

Table 1. Characteristics of the datasets

Region	# Observations	Bands	Sequence Length	# Classes
		MiniTimeMatch		
FR1	106,028	10	36	11
FR2	74,156	10	38	12
DK	53,823	10	28	12
AT	43,809	10	28	10
Total	**277,816**			
		BreizhCrops		
frh01	178,602	10	71	7
frh02	140,621	10	71	7
frh03	166,371	10	71	7
frh04	122,601	10	71	7
Total	**608,195**			



MiniTimeMatch. Source: link



BreizhCrops. Source: link

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

04 DATA/DATA-PREPARATION

05 EXPERIMENTAL SETTING

06 RESULTS

07 CONCLUSION & FUTURE PERSPECTIVES

Experiments

- **Default settings** with spectral-temporal covariance representations
- Manual **empirical hyperparameter tuning**
  - Bimap Layer Dimension
  - Gamma (γ) value
- Model evaluation with **optimal hyperparameters**
- Testing on **other regions**
- **Larger dataset** experimentation

Table 3. Initial configurations for the LERN models

Hyperparameter	**Value**
BiMap dim	$0.7 \times input_{dim}$
Number of Classes	11 (FR1 dataset)
Learning Rate (η)	10^{-2}
Batch Size	64
Epochs	200
Optimizer	RiemannianAdam
Scheduler	StepLr with step-size = 30
Loss function ($\mathcal{L}$)	Focal loss
Focusing parameter (γ)	2
Spectrum Map Parameters	Conv filter 10, 5, 1

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

04 DATA/DATA-PREPARATION

05 EXPERIMENTAL SETTING

06 RESULTS

07 CONCLUSION & FUTURE PERSPECTIVES

Results: MiniTimeMatch

MACRO F1-SCORE

Region	1-nn Log-Euc	1-nn Riem	CNN	LERN-Spec	LERN-Temp	LERN-Combo
AT	0.66	0.68	0.70	0.68	0.80	**0.81**
DK	0.56	0.64	0.51	0.54	0.61	**0.69**
FR1	0.63	0.65	0.78	0.71	0.78	**0.84**
FR2	0.59	0.64	0.61	0.63	0.67	**0.73**

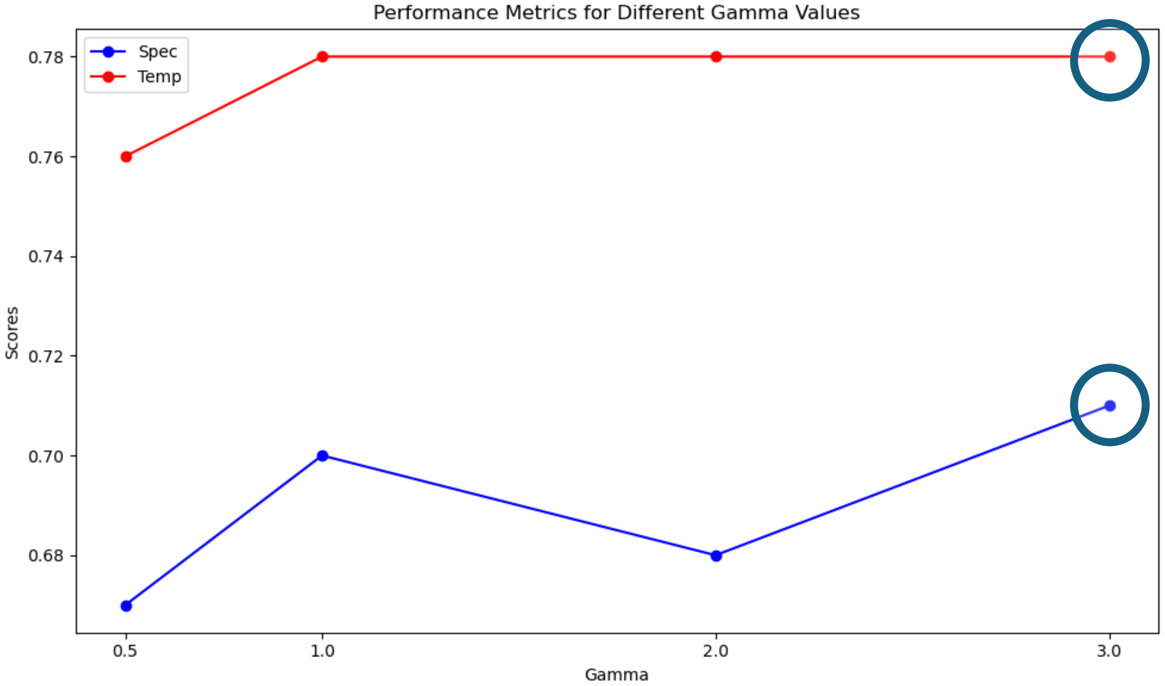
WEIGHTED F1-SCORE

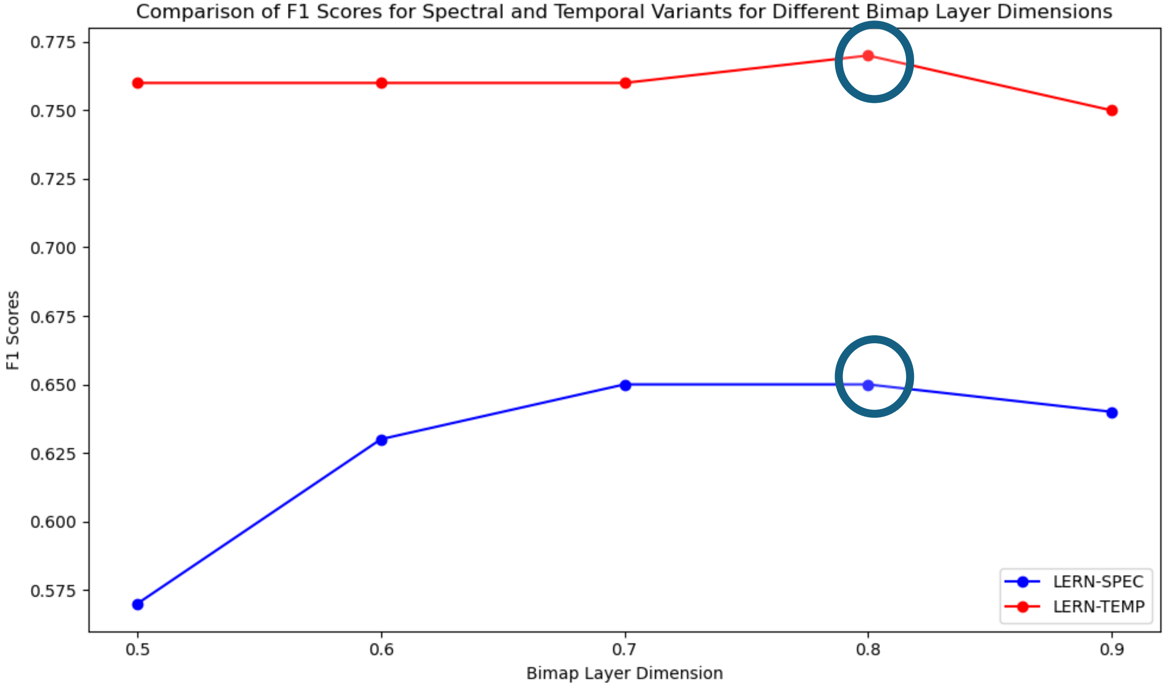
Region	1-nn Log-Euc	1-nn Riem	CNN	LERN-Spec	LERN-Temp	LERN-Combo
AT	0.90	0.91	**0.95**	0.93	0.94	**0.95**
DK	0.75	0.79	0.86	0.80	0.85	**0.89**
FR1	0.89	0.90	0.96	0.93	0.95	**0.97**
FR2	0.89	0.90	**0.95**	0.94	0.94	0.94

06

RESULTS

Optimal hyperparameters





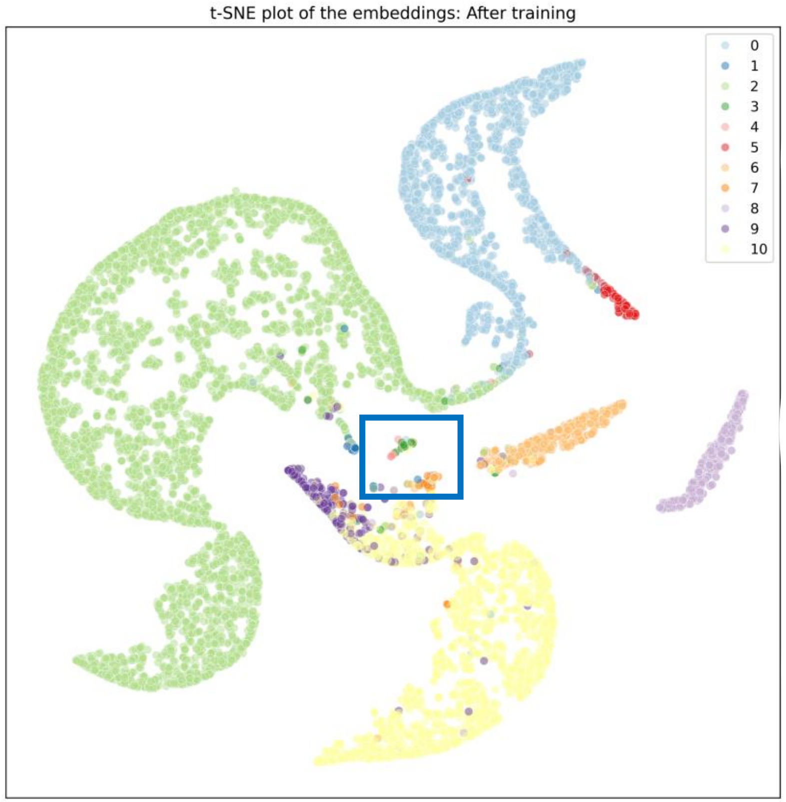
Sensitivity of the model to its main hyperparameters

Optimal value

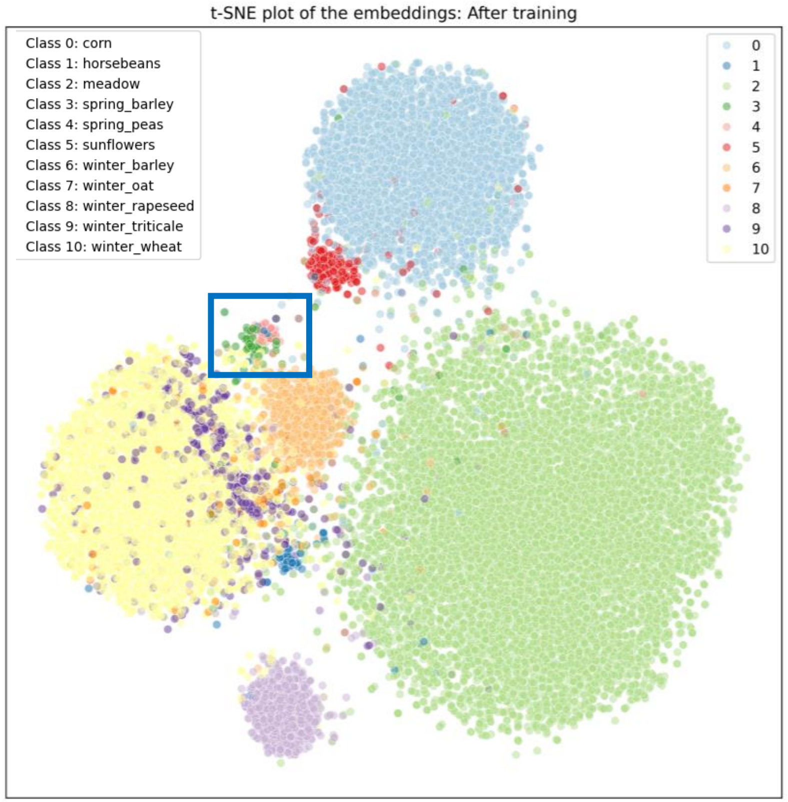
06

RESULTS

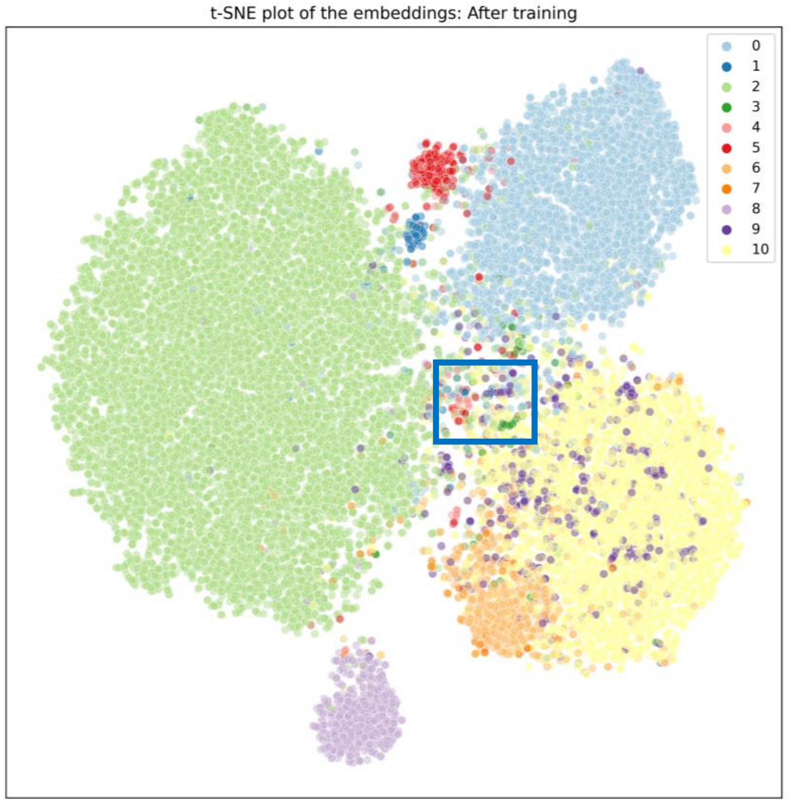
t-SNE plots for FR1



Combo output



Temp output



Spec output

Results: BreizhCrops

MACRO F1-SCORE

Region	1-nn Log-Euc	1-nn Riem	CNN	LERN-Spec	LERN-Temp	LERN-Combo
frh01	-	-	-	0.70	**0.76**	0.75
frh02	-	-	-	0.72	**0.77**	0.76
frh03	-	-	-	0.68	0.73	**0.74**
frh04	-	-	-	0.69	**0.74**	0.73

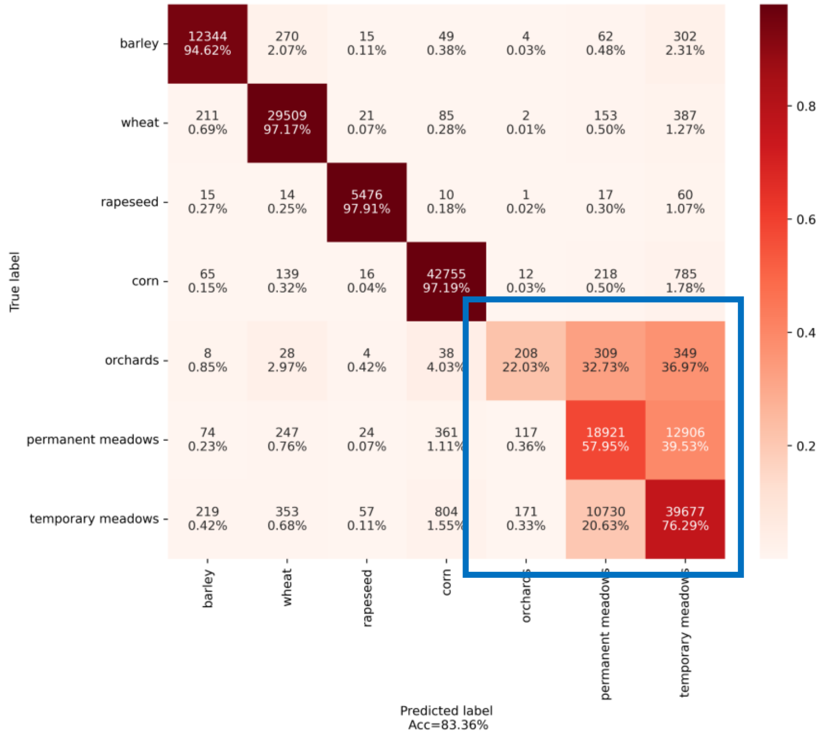
WEIGHTED F1-SCORE

Region	1-nn Log-Euc	1-nn Riem	CNN	LERN-Spec	LERN-Temp	LERN-Combo
frh01	-	-	-	0.79	**0.82**	**0.82**
frh02	-	-	-	0.77	0.80	**0.81**
frh03	-	-	-	0.76	0.78	**0.79**
frh04	-	-	-	0.75	0.78	**0.79**

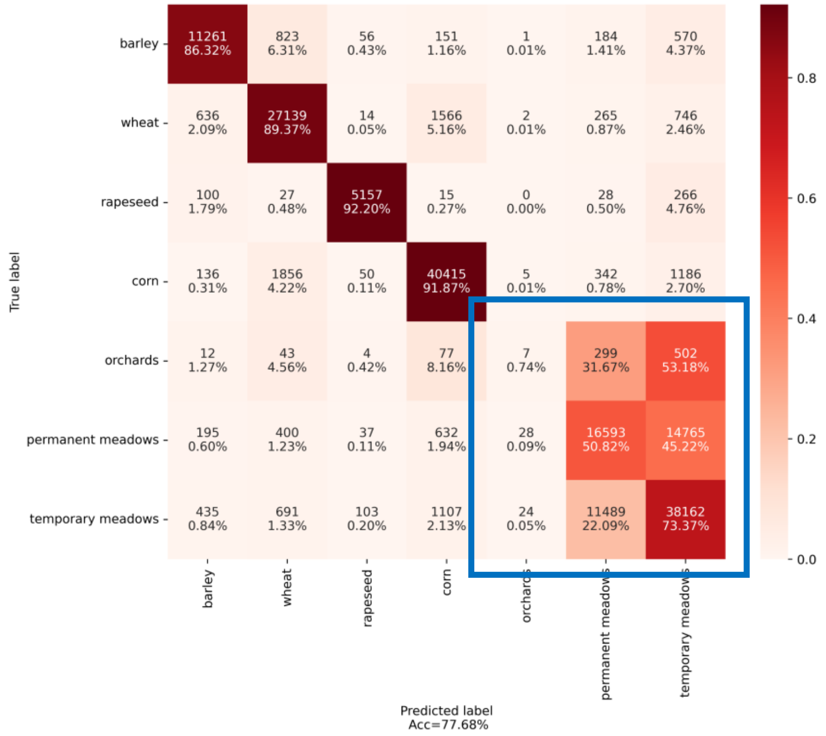
06

RESULTS

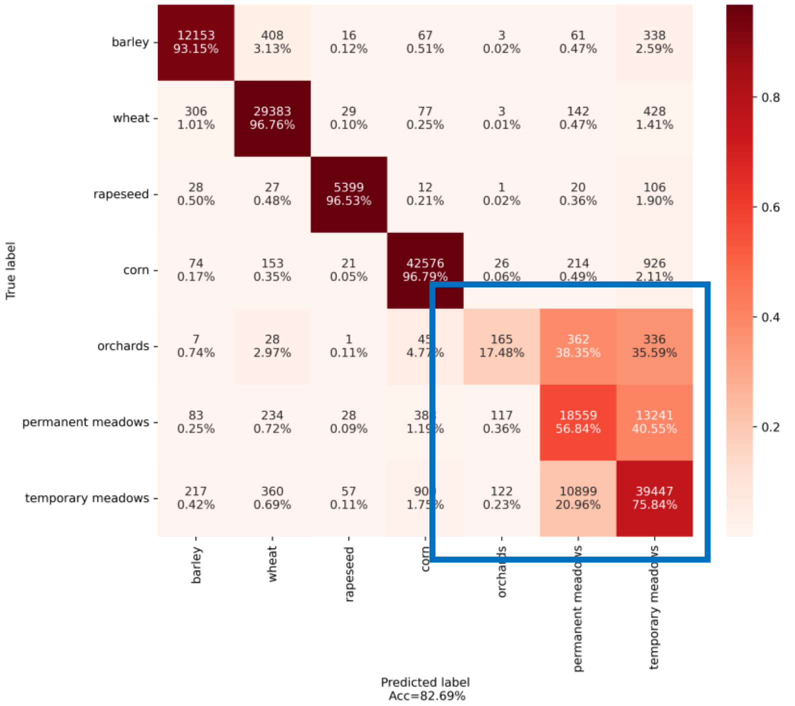
Breizhcrops



Combo output



Spec output



Temp output

Content

01 INTRODUCTION

02 EXISTING LITERATURE

03 METHODOLOGY

04 DATA/DATA-PREPARATION

05 EXPERIMENTAL SETTING

06 RESULTS

07 CONCLUSION & FUTURE PERSPECTIVES

Conclusion

LERN-Combo > LERN-Temp > LERN-Spec

Introduced LERN & LERN-Combo: second-order geometry-aware neural network.

Utilized covariance representation for remote sensing timeseries data.

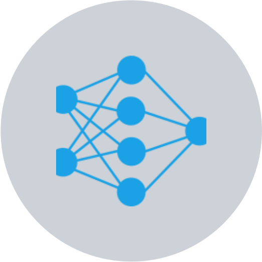
Embedded covariance matrices onto a Riemannian manifold using the LERN network.

Enhanced understanding and classification of complex timeseries patterns for crop classification.

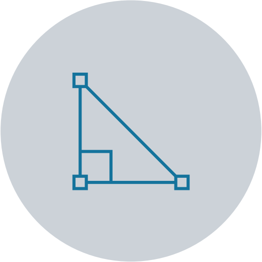
Challenges

- Sampling inhomogeneity
- SVD on high-dimensional matrices often computed on CPU only, making parallelization difficult.
- Dataset imbalances
- Spectral similarities
- High computational demands, particularly for matrix operations

Future Perspectives



EXPLORE ARCHITECTURAL HYPERPARAMETERS



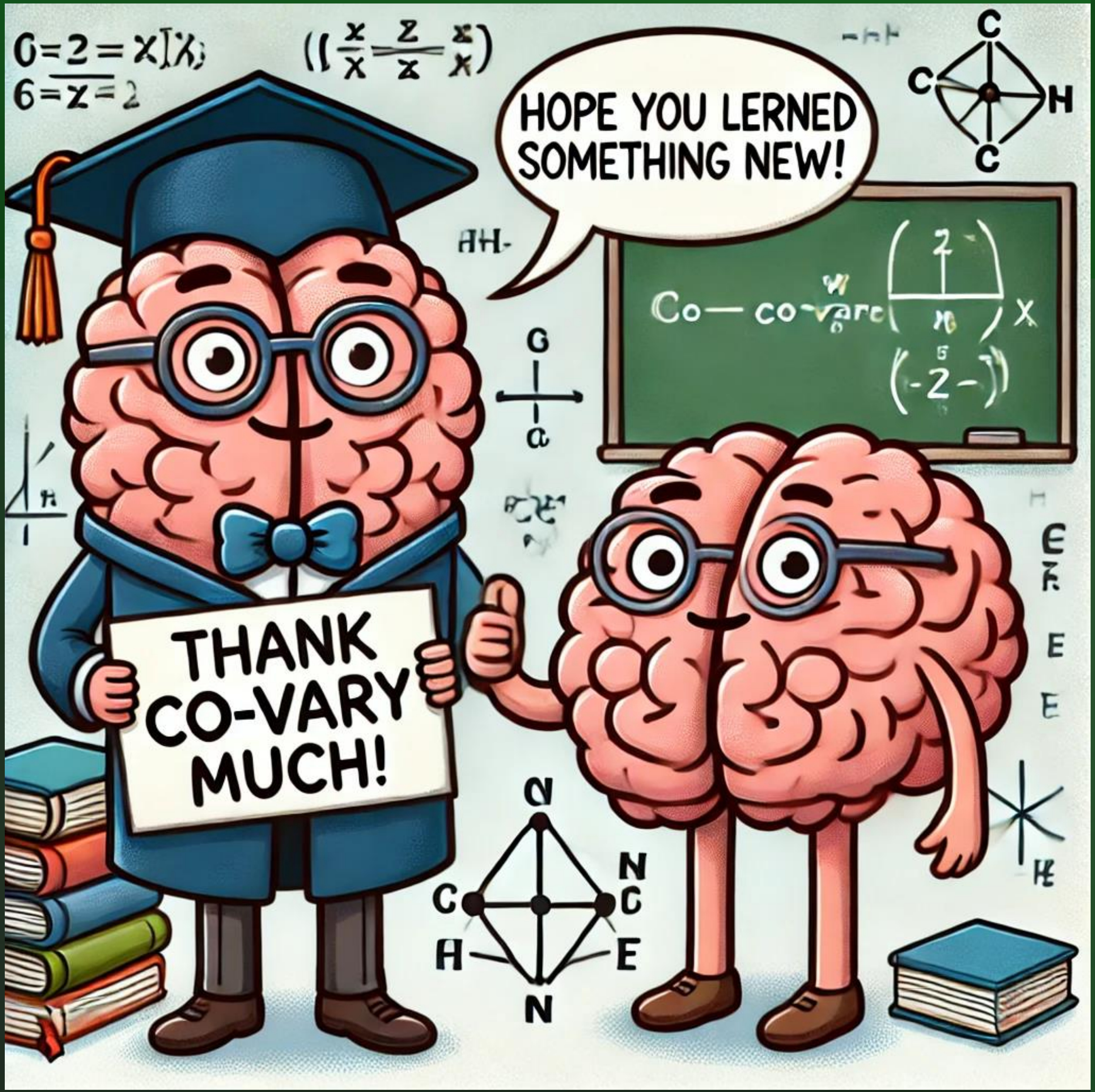
RIEMANNIAN BATCH NORMALIZATION (RBN)

STATISTICAL SIGNIFICANCE TESTS

CLASS-SPECIFIC ANALYSIS

References

- Akodad, S. (2021). Ensemble learning methods on the space of covariance matrices: application to remote sensing scene and multivariate time series classification. Theses, Universite de Bordeaux.
- Barachant, A., Bonnet, S., Congedo, M., and Jutten, C. (2010). Riemannian geometry applied to bci classification. In Vigneron, V., Zarzoso, V., Moreau, E., Gribonval, R., and Vincent, E., editors, Latent Variable Analysis and Signal Separation, pages 629–636, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Brooks, D. (2020). Deep Learning and Information Geometry for Time-Series Classification. Theses, Sorbonne Universite.
- Huang, Z., and Gool, L. V. (2016). A riemannian network for SPD matrix learning.
- Interdonato, R., Ienco, D., Gaetano, R., and Ose, K. (2019). Duplo: A dual view point deep learning architecture for time series classification. ISPRS Journal of Photogrammetry and Remote Sensing, 149:91–104.
- Lin, T.-Y., Goyal, P., Girshick, R., He, K., and Dollar, P. (2018). Focal loss for dense object detection.
- Painblanc, F., Chapel, L., Courty, N., Friguet, C., Pelletier, C., and Tavenard, R. (2023b). Match-and-deform: Time series domain adaptation through optimal transport and temporal alignment. https://github.com/rtavenar/MatchAndDeform.git.
- Ruswurm, M., Pelletier, C., Zollner, M., Lefevre, S., and Korner, M. (2020). Breizhcrops: A time series dataset for crop type mapping. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences ISPRS (2020).
- Wang, R., Wu, X.-J., Chen, Z., Xu, T., and Kittler, J. (2022). Dreamnet: A deep Riemannian network based on spd manifold learning for visual classification.



Made with ChatGPT

Appendix

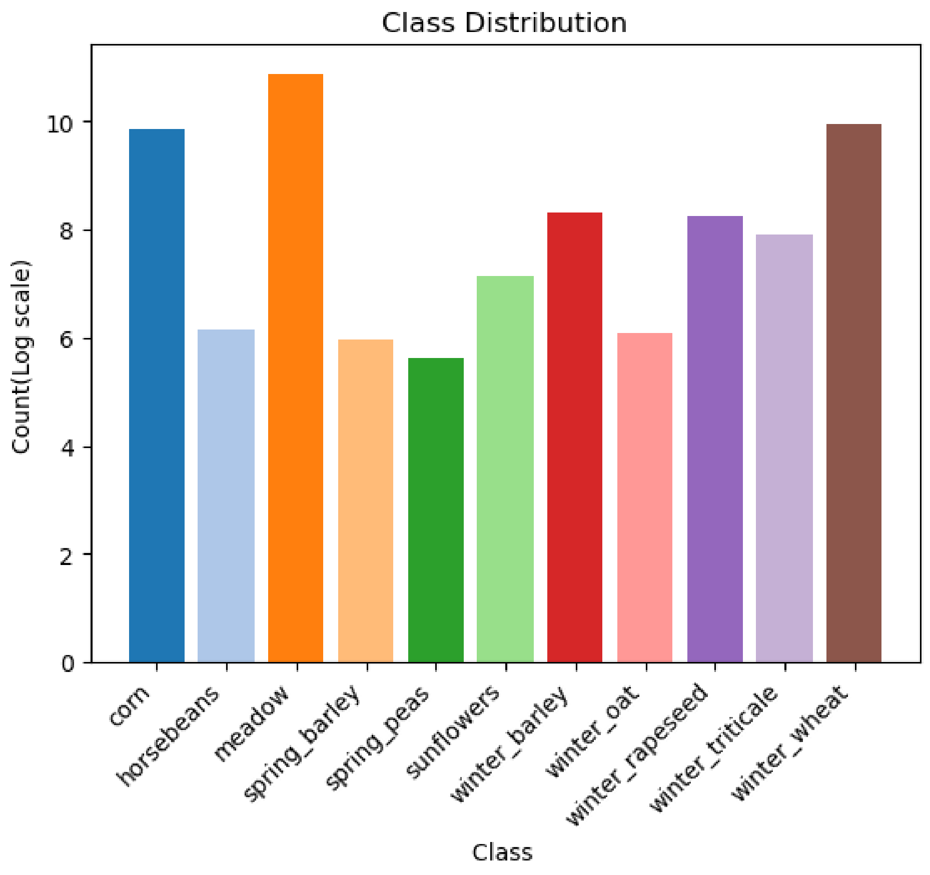
Experiments

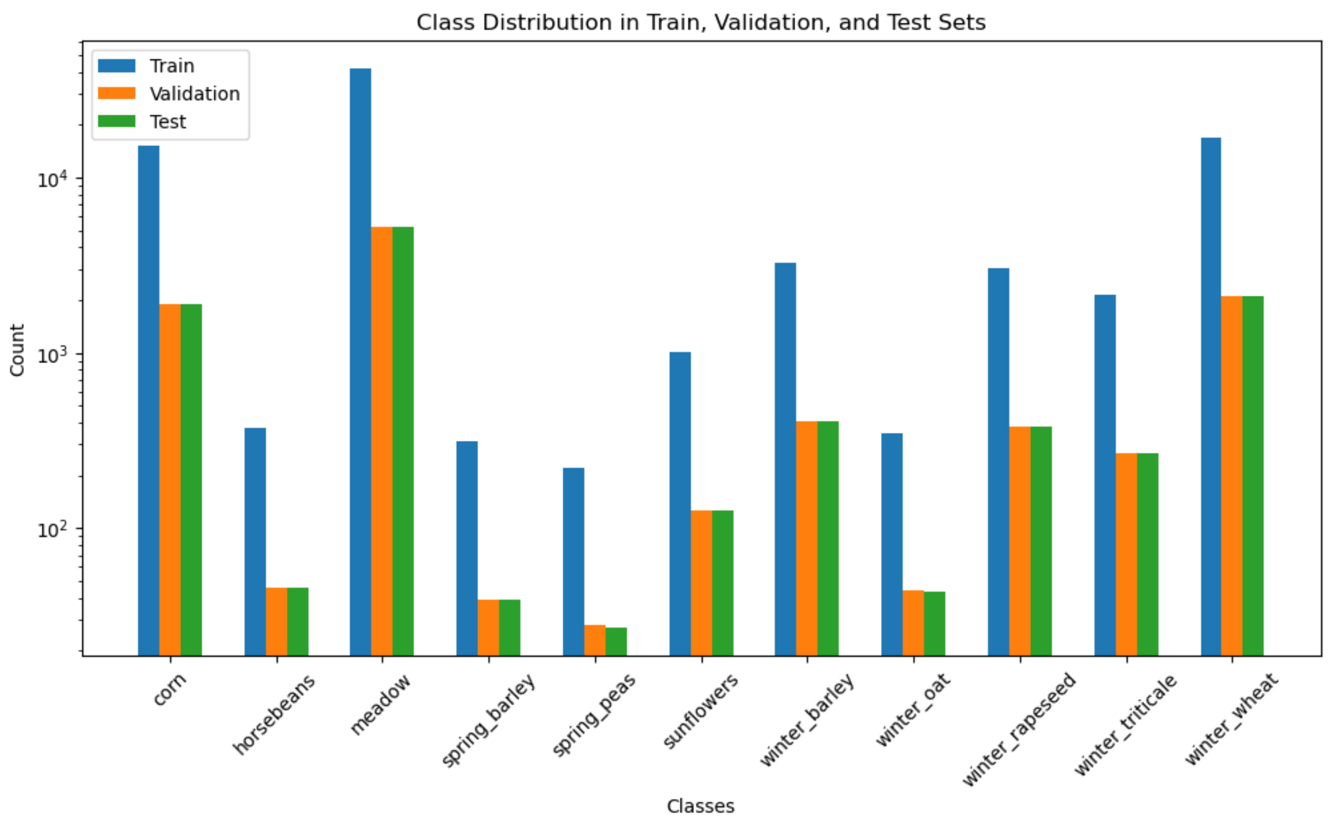
- Primary basline model – 1-nn
- Secondary baseline model – 1D-CNN
- Sampling inhomogeneity - linear interpolation

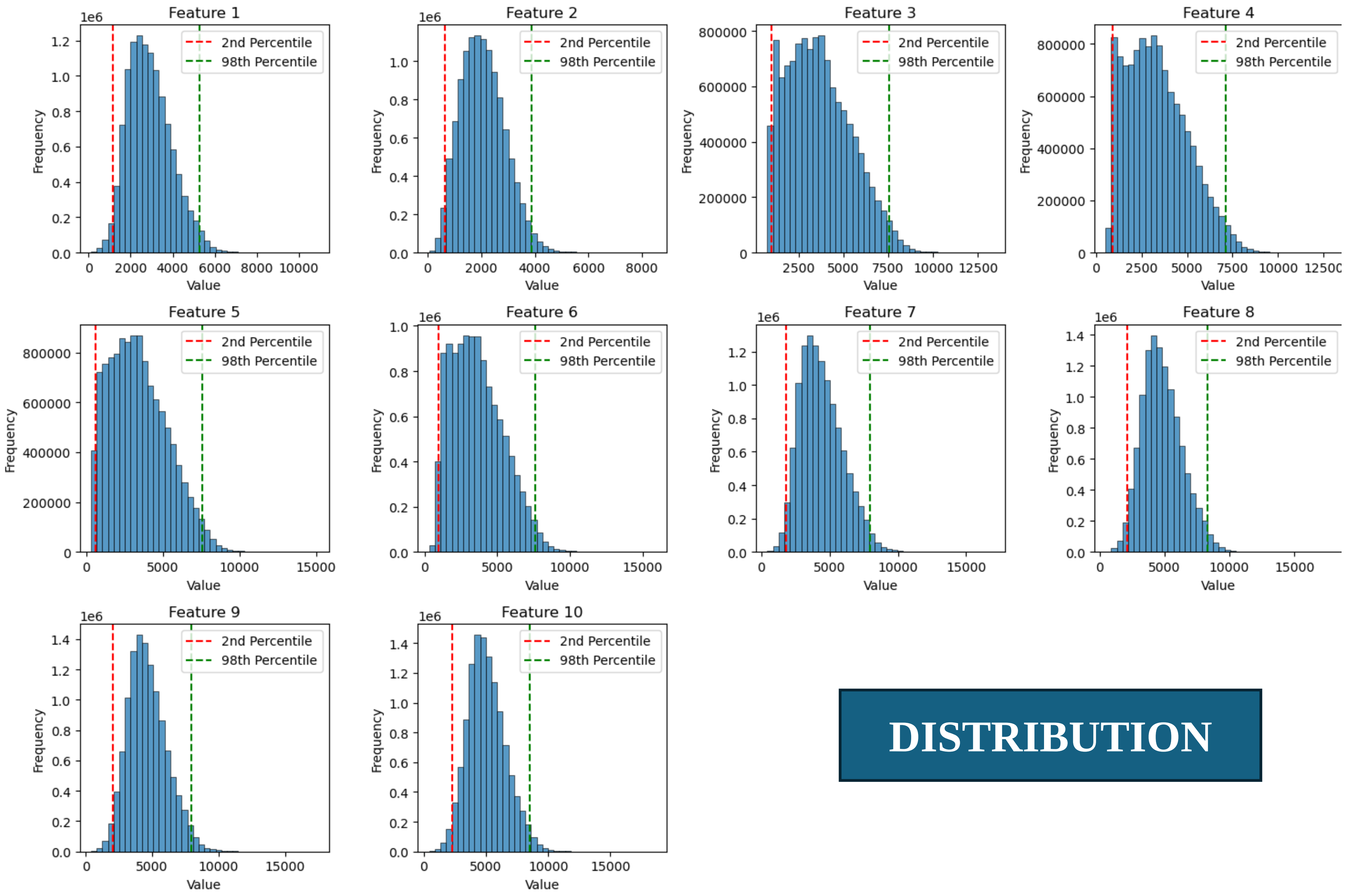
Table 2. Trainable Parameters and Time per Epoch

	Model			
Metric	**1D- CNN**	**LERN-Spec**	**LERN-Temp**	**LERN-Combo**
Trainable Parameters	41,323	1,412	15,391	31,965
Time per Epoch (secs)	9.21	12.6	17.78	32.3

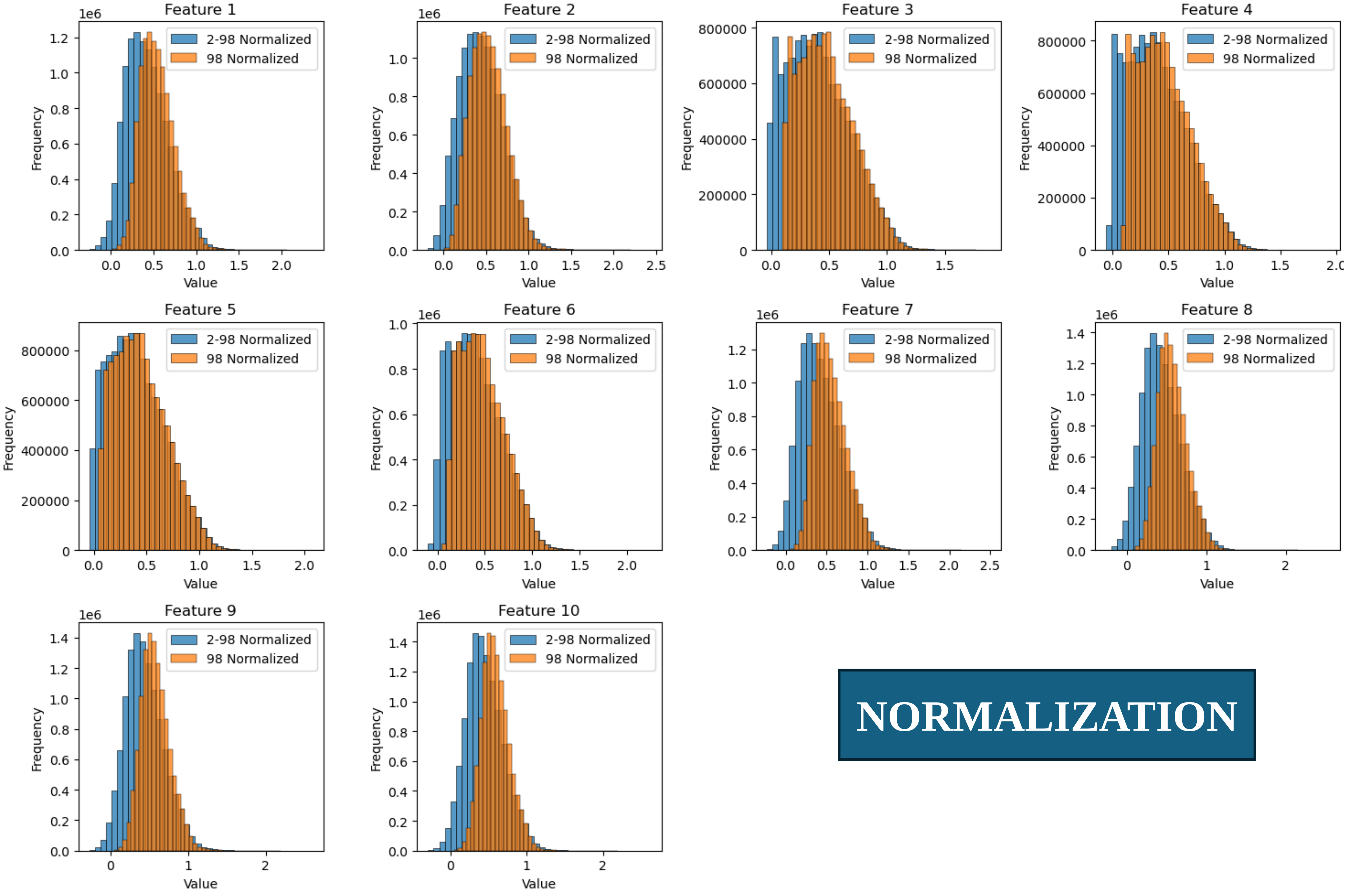
Class Distribution - Split







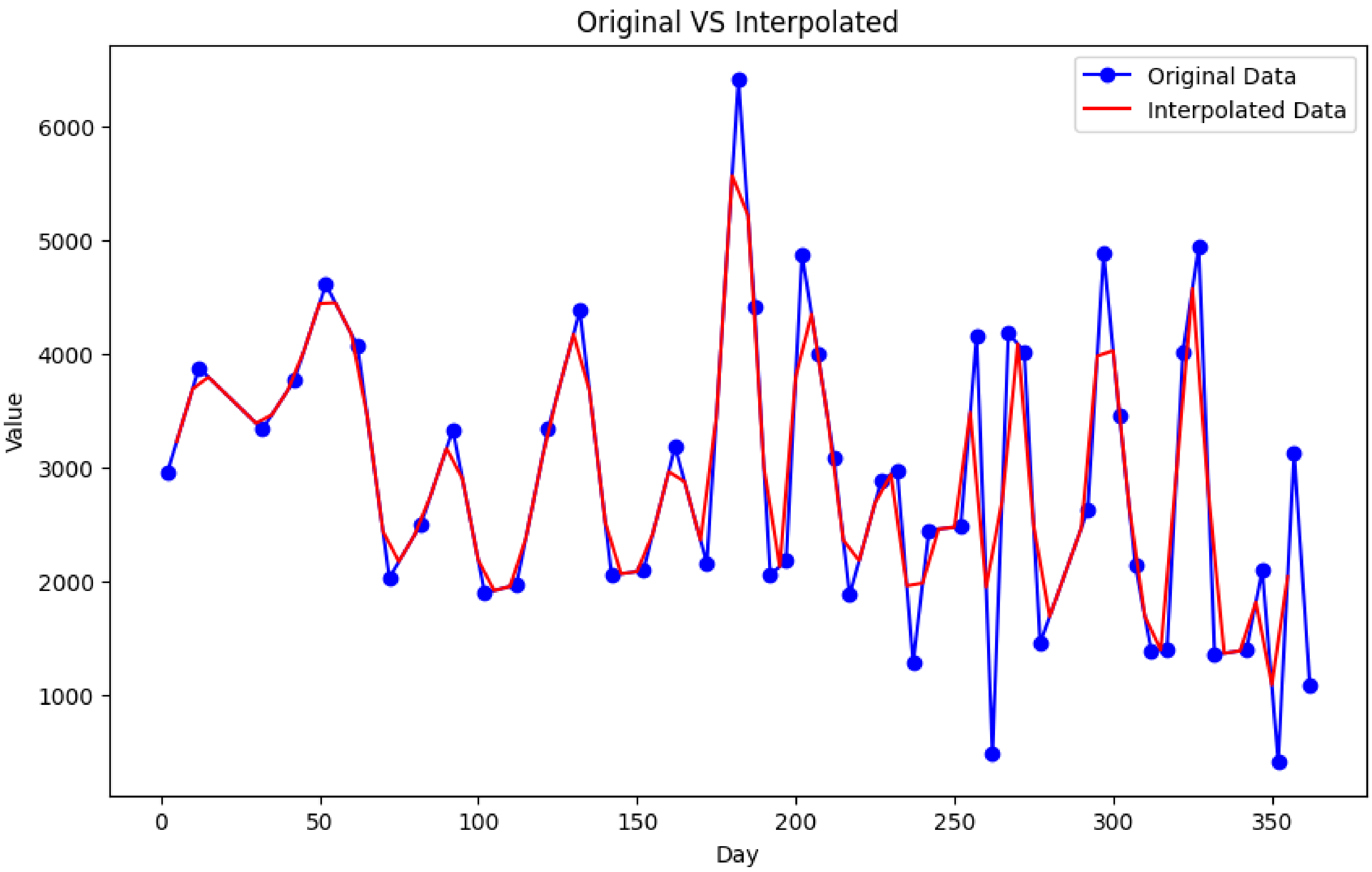
DISTRIBUTION



NORMALIZATION

Table 4. Outputs from combined datasets - BreizhCrops

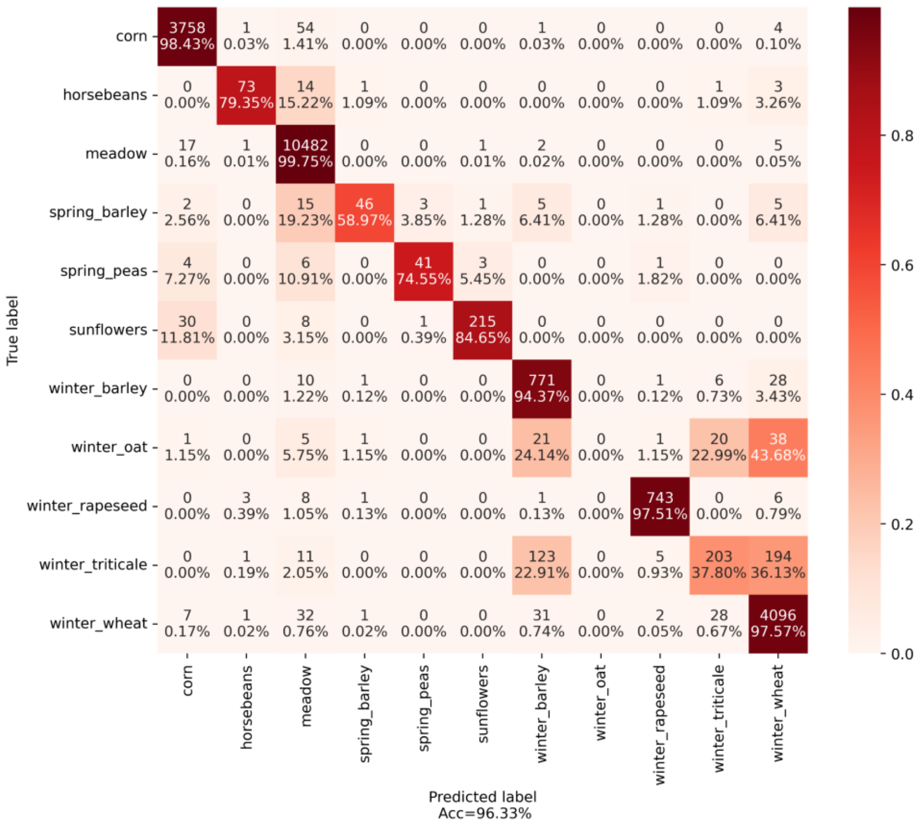
	Model		
Region	**LERN-Spec**	**LERN-Temp**	**LERN-Combo**
	Macro F1 Score		
frh01	0.68	-	-
frh02	0.70	-	-
frh03	0.65	-	-
frh04	0.70	0.77	**0.79**
	Weighted F1 Score		
frh01	0.76	-	-
frh02	0.77	-	-
frh03	0.74	-	-
frh04	0.78	**0.83**	**0.83**


Original VS Interpolated
Original Data
Interpolated Data
6000
5000
4000
3000
2000
1000
Value
0
50
100
150
200
250
300
350
Day

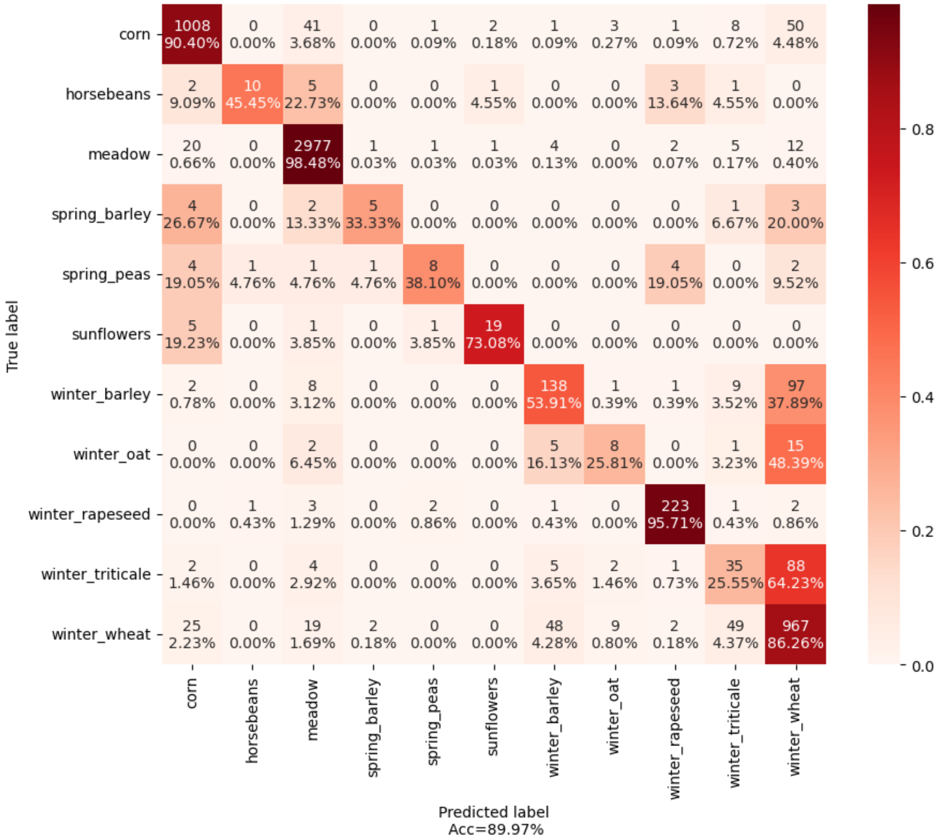
06

RESULTS

Baseline

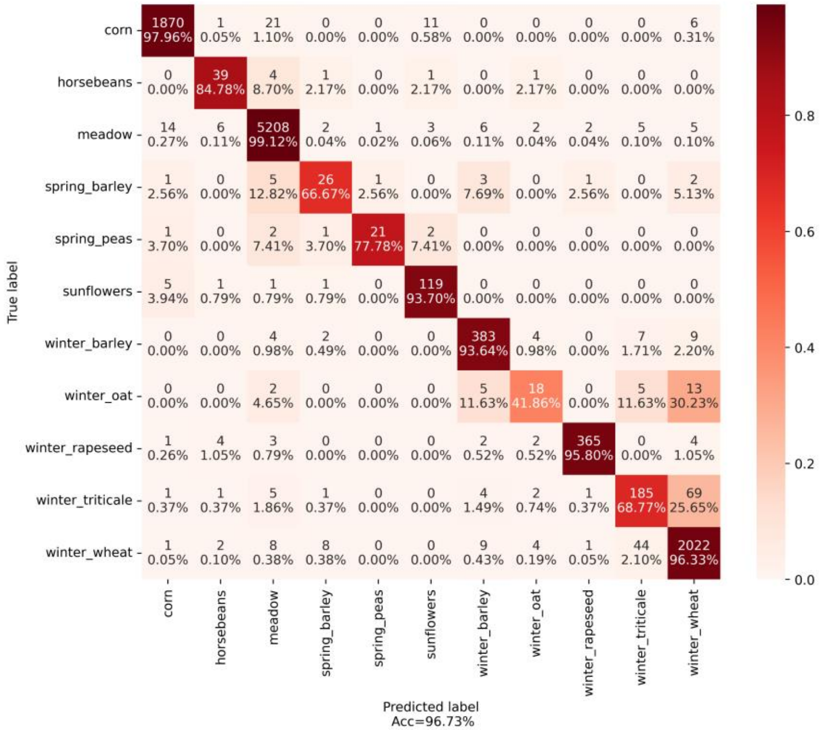


Output from CNN

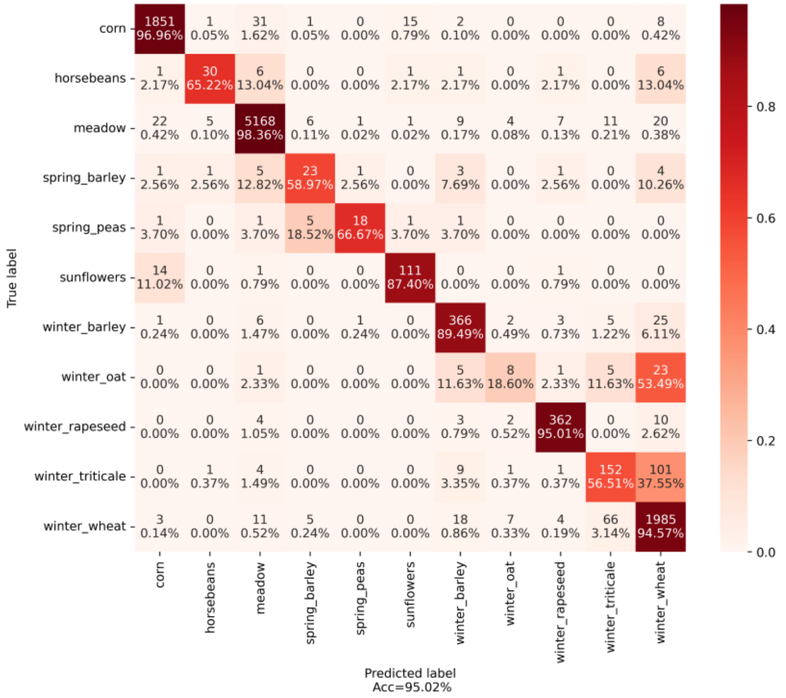


Output from 1-nn

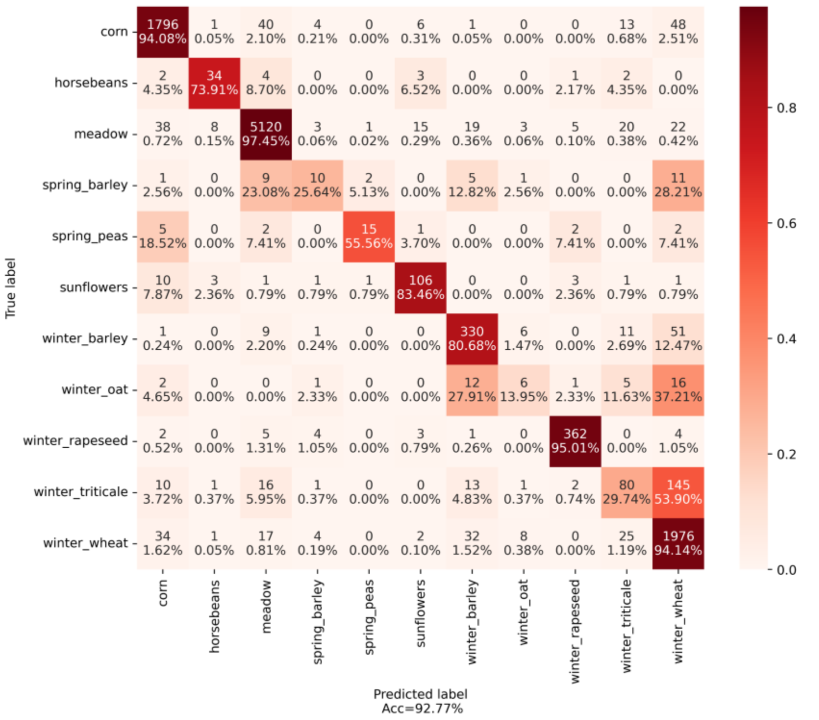
FR1-MiniTimeMatch



Combo output

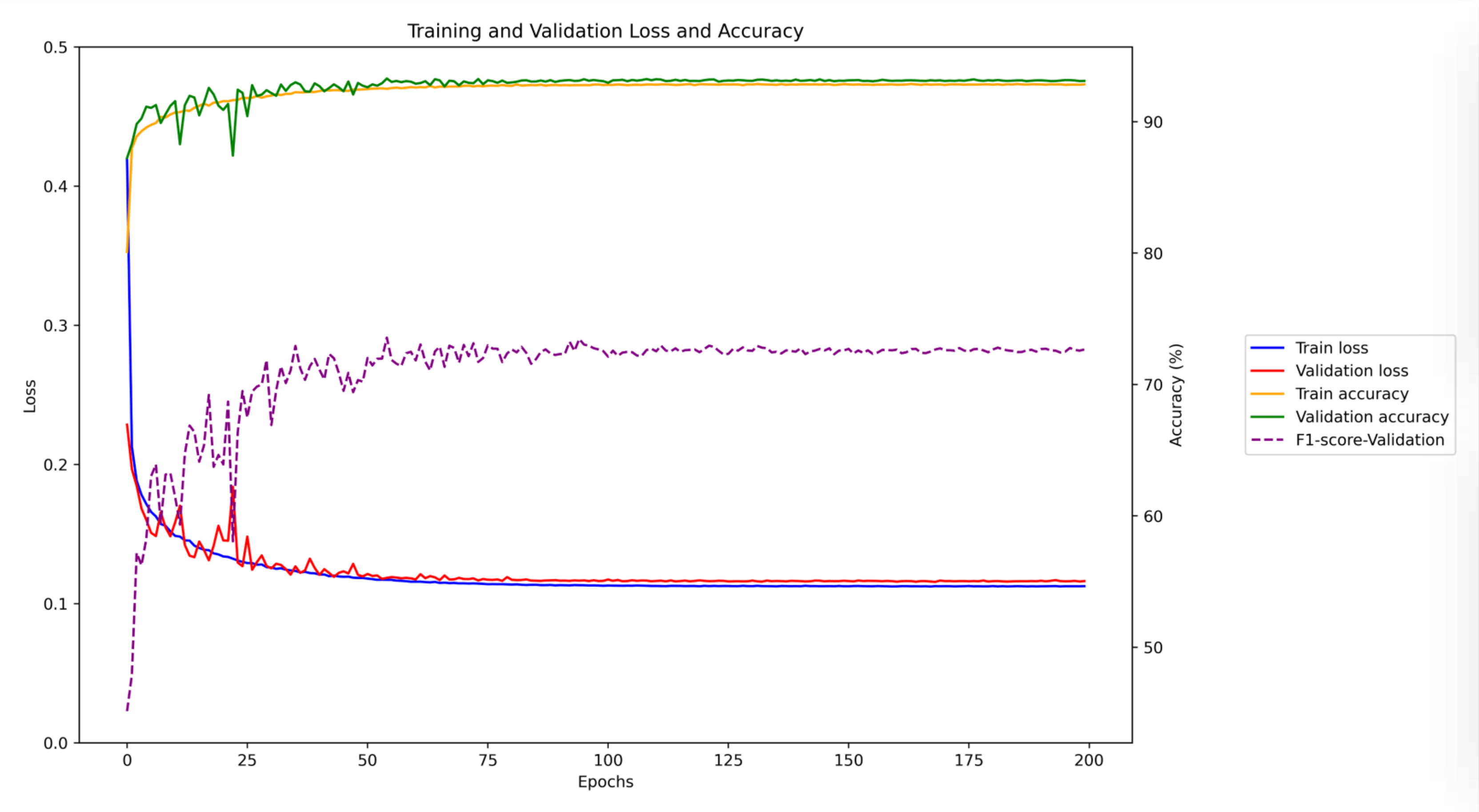


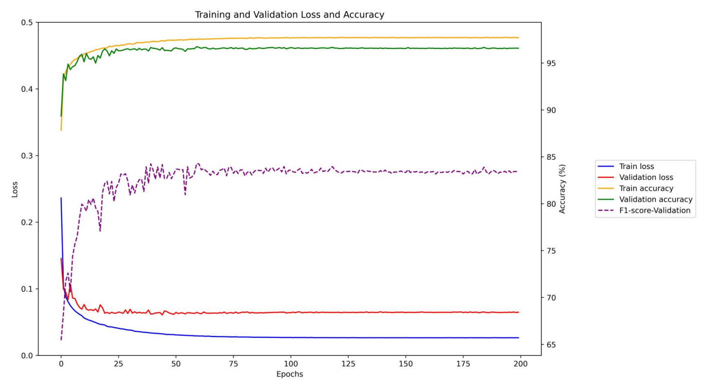
Temp output

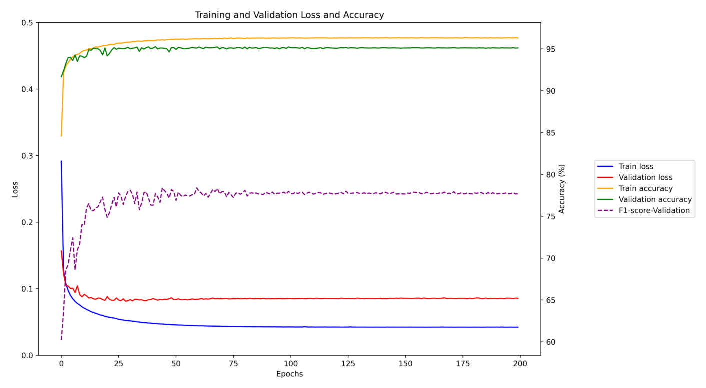


Spec output

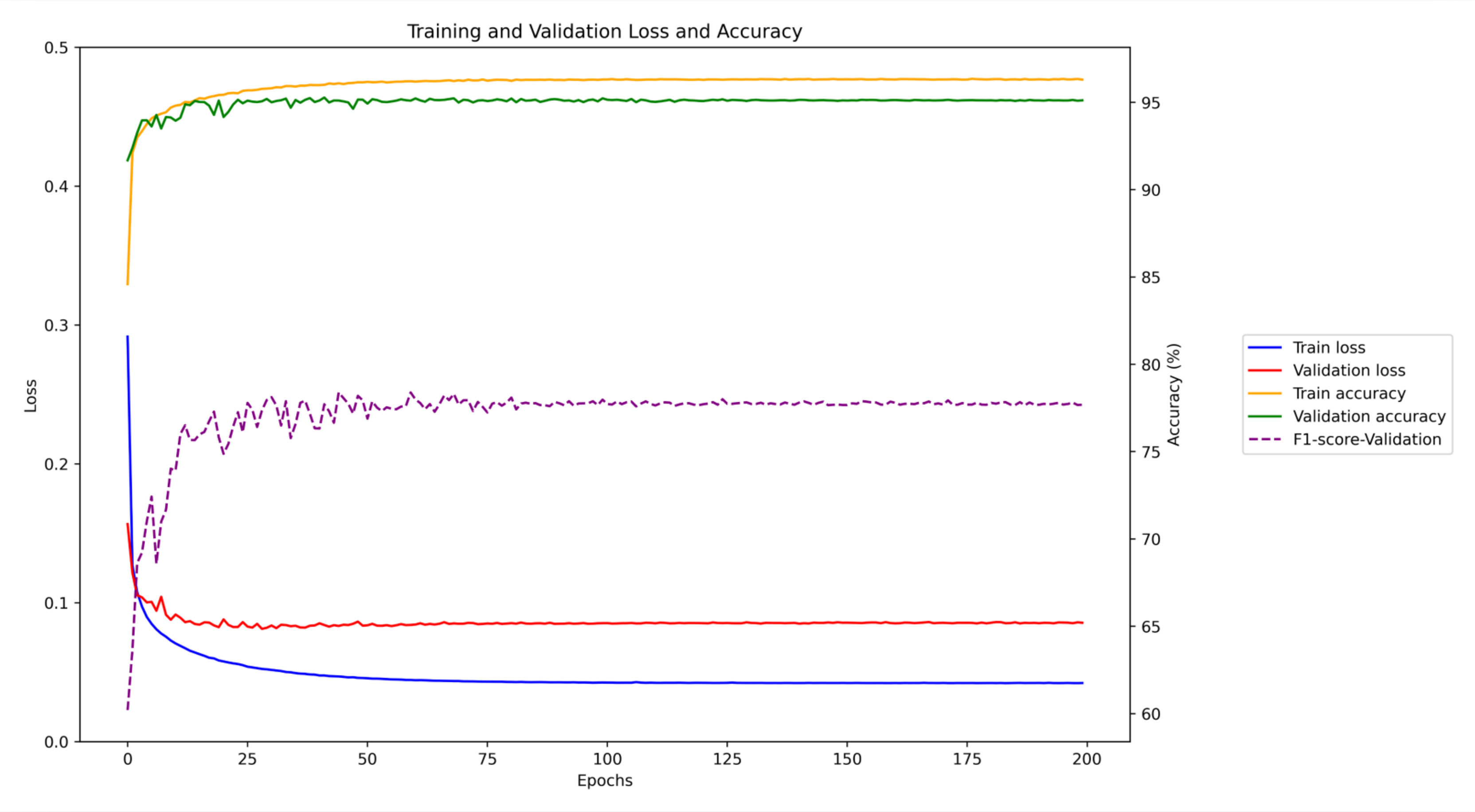
Curves for FR1



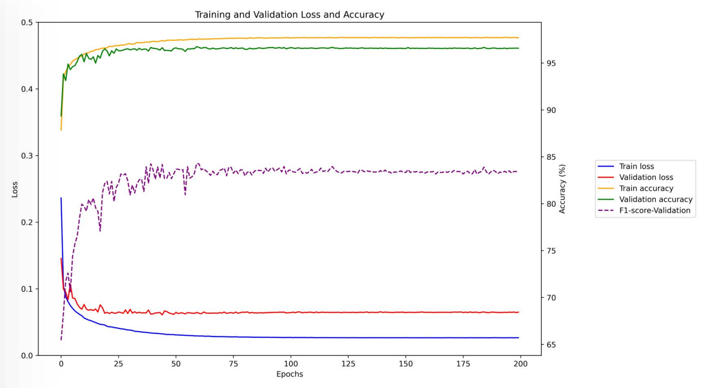




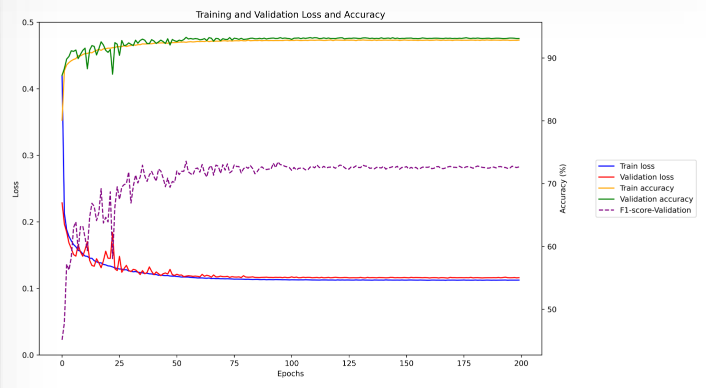
Cont…



Training and Validation Loss and Accuracy
Loss
Accuracy (%)
Epochs
Train loss
Validation loss
Train accuracy
Validation accuracy
F1-score-Validation

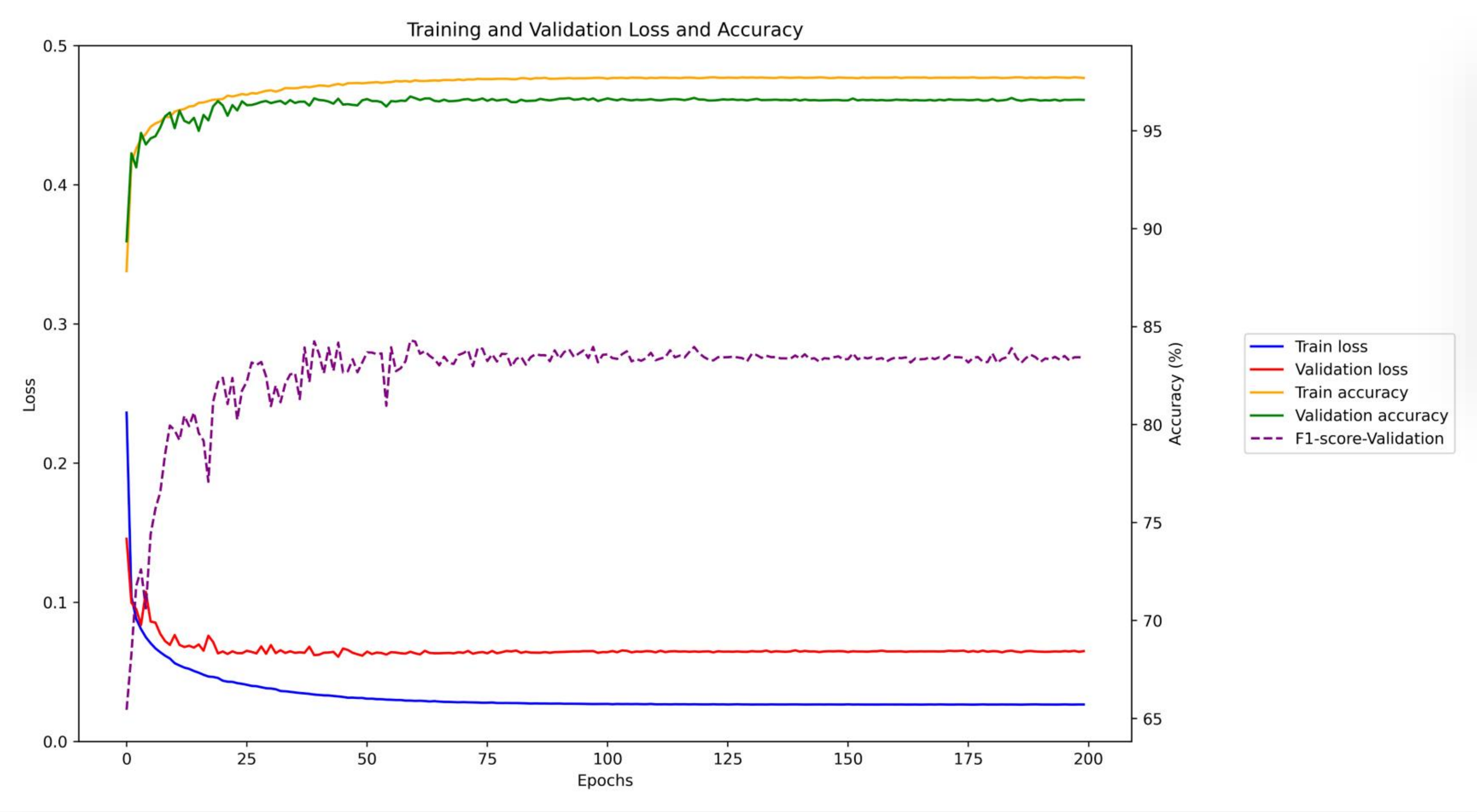


Training and Validation Loss and Accuracy

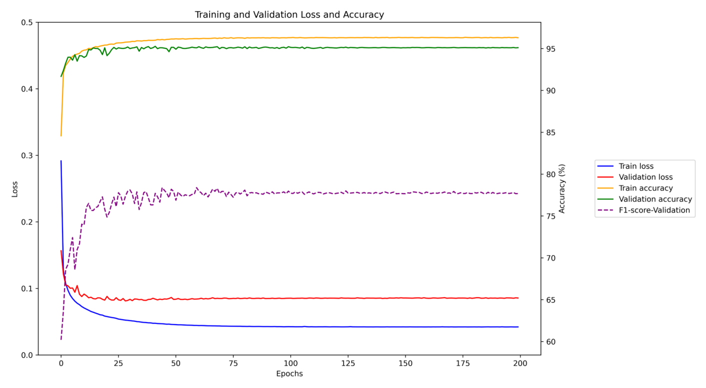


Training and Validation Loss and Accuracy

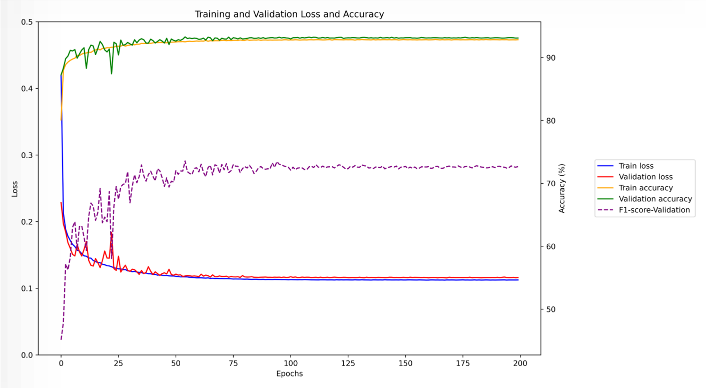
Cont…



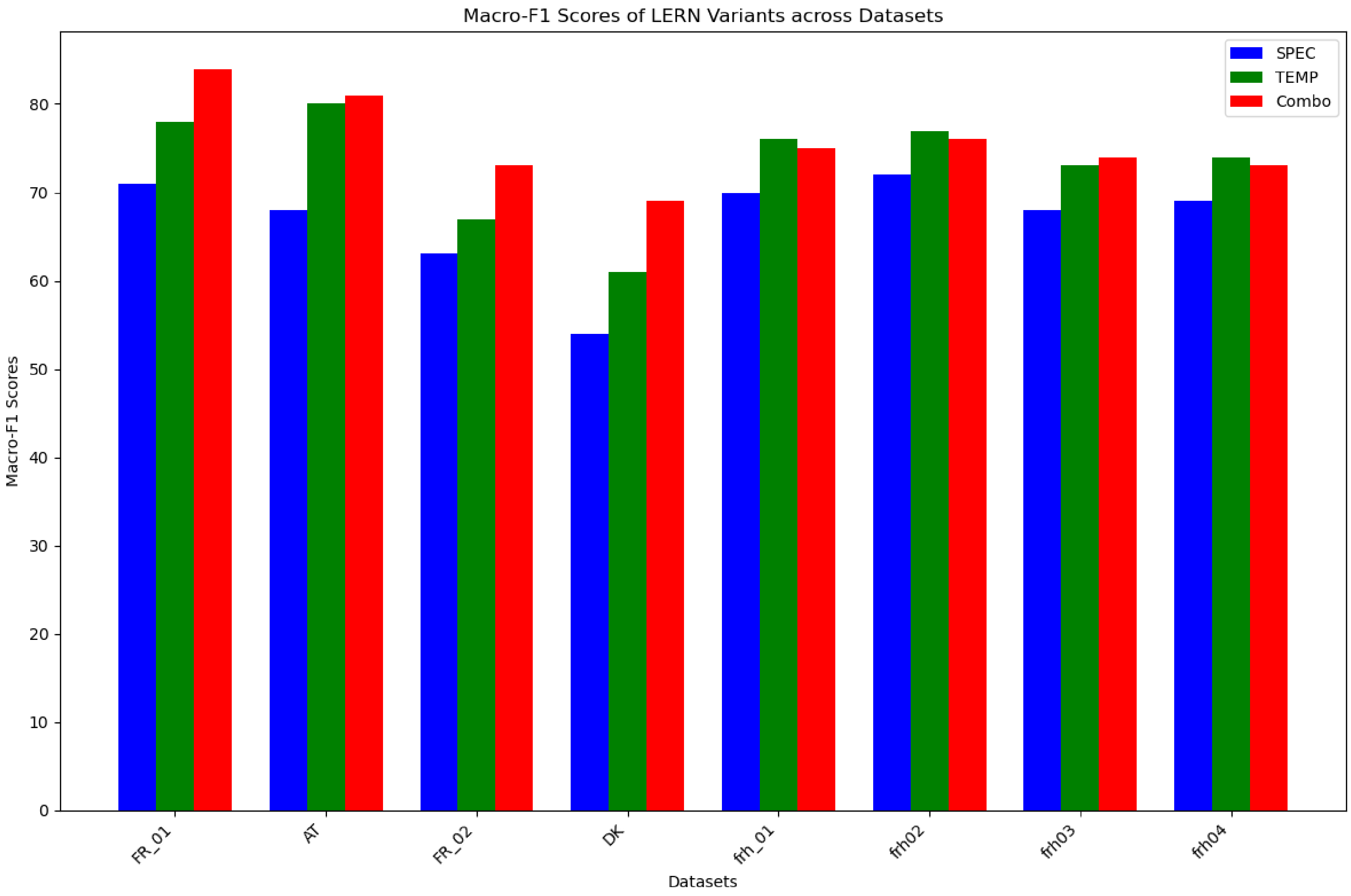
Training and Validation Loss and Accuracy
Loss
Accuracy (%)
Epochs
Train loss
Validation loss
Train accuracy
Validation accuracy
F1-score-Validation



Training and Validation Loss and Accuracy
Epochs



Training and Validation Loss and Accuracy
Epochs



Macro-F1 Scores of LERN Variants across Datasets
SPEC
TEMP
Combo
Macro-F1 Scores
0
10
20
30
40
50
60
70
80
FR_01
AT
FR_02
DK
frh_01
frh02
frh03
frh04
Datasets