# STUNT: Few-shot Tabular Learning with Self-generated Tasks from Unlabeled Tables

Jaehyun Nam[1], Jihoon Tack[1], Kyungmin Lee[1], Hankook Lee[2], Jinwoo Shin[1]

[1]Korea Advanced Institute of Science and Technology (KAIST)
[2]LG AI Research

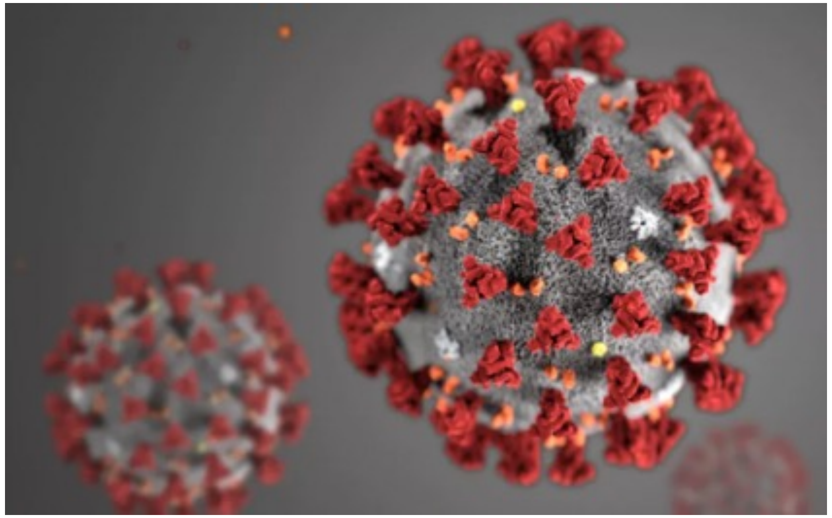
# Importance of Few-shot Tabular Learning in Practical Deployment

**Few-shot semi-supervised tabular learning is a crucial** application:

- **Credit risk in financial datasets**: High labeling costs
- **Early infected patiend of COVID-19**: Difficulties in collecting new samples for novel tasks

- **STUNT**: We propose **a novel framework for few-shot semi-supervised tabular learning**.

💡 Significant performance gain, compared to prior semi- and self-supervised baselines with a simple framework.



| Fever (℃) | Cough | Fatigue | Shortness of breath (breath/min) | Loss of smell |
|---|---|---|---|---|
| 39 | Frequent | Tired | 25 | Yes |
| 38.5 | Occasional | Tired | 22 | No |
| … | … | … | … | … |
| … | … | … | … | … |
| 37.8 | Occasional | Fine | 28 | Yes |

# Tackling Limited Label Issues: Utilize Unlabeled Datasets

**Approaches for few-shot semi-supervised learning across various domains: Utilize unlabeled datasets**

- Learning a generalizable and transferable representation (e.g., images and languages) with self-supervised learning
- Self-supervised learning (e.g., SubTab [Ucar et al., 2021], VIME [Yoon et al., 2020]) are not effective for tabular domains
- Heterogeneous characteristic of tabular data
- Do not bring meaningful performance gain over even a simple kNN **on** few-shot classification accuracy (%)

| Type | Method | income | cmc | karhunen | optdigit | diabetes | semeion | pixel | dna | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| # shot = 1 | | | | | | | | | | |
| Supervised | kNN | 61.22 | 34.99 | 54.42 | 65.58 | 58.56 | 44.35 | 61.48 | 42.67 | 52.82 |
| Self-supervised | SubTab | 61.88 | 35.68 | 50.32 | 67.05 | 58.06 | 40.27 | 60.40 | 45.68 | 52.42 |
| | VIME | 61.99 | 35.30 | 59.62 | 70.52 | 56.95 | 47.20 | 64.17 | 51.36 | 55.89 |
| # shot = 5 | | | | | | | | | | |
| Supervised | kNN | 70.49 | 38.56 | 79.98 | 84.89 | 67.32 | 68.33 | 84.02 | 61.45 | 69.38 |
| Self-supervised | SubTab | 71.91 | 39.51 | 69.56 | 83.60 | 68.79 | 59.87 | 80.13 | 61.57 | 66.87 |
| | VIME | 67.80 | 37.51 | 82.87 | 87.42 | 64.29 | 71.53 | 86.79 | 69.62 | 70.98 |

[Ucar et al., 2021] SubTab: Subsetting Features of Tabular Data for Self-Supervised Representation Learning, NeurIPS 2021
[Yoon et al., 2020] VIME: Extending the Success of Self- and Semi-supervised Learning to Tabular Domain, NeurIPS 2020

# Utilizing Unsupervised Meta-learning?

**Instead, one can utilize the power of unsupervised meta-learning**

- Meta-learning is one of the most effective few-shot learning strategies
- Unsupervised meta-learning (CACTUs [Hsu et al., 2019]) reduces the gap via fast adaptation to unseen few-shot tasks
- **Promising direction** for few-shot tabular learning
- CACTUs outperforms the self-supervised tabular learning methods on few-shot classification accuracy(%)

| Type | Method | income | cmc | karhunen | optdigit | diabetes | semeion | pixel | dna | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | # shot = 1 | | | | | |
| Supervised | kNN | 61.22 | 34.99 | 54.42 | 65.58 | 58.56 | 44.35 | 61.48 | 42.67 | 52.82 |
| Self-supervised | SubTab | 61.88 | 35.68 | 50.32 | 67.05 | 58.06 | 40.27 | 60.40 | 45.68 | 52.42 |
| | VIME | 61.99 | 35.30 | 59.62 | 70.52 | 56.95 | 47.20 | 64.17 | 51.36 | 55.89 |
| Unsupervised Meta. | CACTUs | **64.02** | **36.10** | **65.59** | **71.98** | **58.92** | **48.96** | **67.61** | **65.93** | **59.89** |
| | | | | | # shot = 5 | | | | | |
| Supervised | kNN | 70.49 | 38.56 | 79.98 | 84.89 | 67.32 | 68.33 | 84.02 | 61.45 | 69.38 |
| Self-supervised | SubTab | 71.91 | 39.51 | 69.56 | 83.60 | **68.79** | 59.87 | 80.13 | 61.57 | 66.87 |
| | VIME | 67.80 | 37.51 | **82.87** | **87.42** | 64.29 | **71.53** | **86.79** | 69.62 | 70.98 |
| Unsupervised Meta. | CACTUs | **72.03** | **38.81** | 82.20 | 85.92 | 66.79 | 65.00 | 85.25 | **81.52** | **72.19** |

[Hsu et al., 2019] Unsupervised Learning via Meta-Learning, ICLR 2019

# Unsupervised Meta-Learning for Few-shot Tabular Learning

**Unsupervised meta-learning**: Meta-learn over the self-generated tasks from unlabeled data

- **Generating tasks** for unsupervised meta-learning **is key** to training an effective few-shot learner.

🤔 How should we generate more diverse and effective tasks for tabular unsupervised meta-learning?

💡 **Idea**: We generate a tasks from the unlabeled data by treating the table's column feature as a target.

- The **blood sugar** value can be used as a substituted label for **diabetes**.

| Blood pressure (mmHg) | BMI (kg/m$^2$) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | 26.0 | 137 | 51 | 135 |
| 68 | 33.6 | 197 | 49 | 168 |
| … | … | … | … | … |
| … | … | … | … | … |
| 78 | 35.2 | 89 | 54 | 126 |

| Diabetes |
|---|
| Positive |
| Positive |
| … |
| … |
| Negative |

# Proposed Framework: STUNT

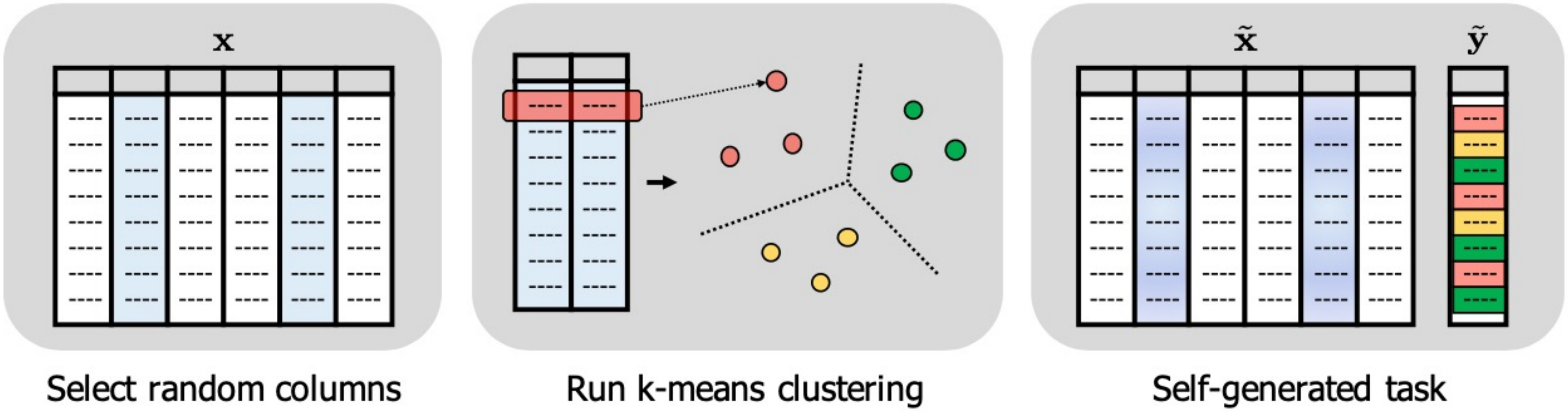
Unsupervised meta-learning with Self-generated Tasks from Unlabeled Tables (STUNT)

- Generate pseudo-labels of the given unlabeled input by k-means clustering on randomly chosen subsets of columns.
- Apply a meta-learning scheme [Snell et al., 2017] to learn generalizable knowledge with the self-generated tasks.

🤔 In the few-shot learning setup, there is no labeled validation set. How can we validate our model?

💡 Utilize STUNT to the unlabeled set for an unsupervised validation!

- Highly effective for hyperparameter searching and early stopping




[Snell et al., 2017] Prototypical Networks for Few-shot Learning, NeurIPS 2017

# Unsupervised Meta-Learning with STUNT: Task Generation

**1. We self-generate diverse tasks from the unlabeled dataset.**

- **Select random columns.**

**Why?**

- Improve the diversity and the possibility of sampling highly correlated columns with the original label.

| Blood pressure (mmHg) | BMI (kg/m$^2$) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | 26.0 | 137 | 51 | 135 |
| 68 | 33.6 | 197 | 49 | 168 |
| … | … | … | … | … |
| … | … | … | … | … |
| 78 | 35.2 | 89 | 54 | 126 |

| BMI (kg/m$^2$) | Blood sugar (mg/dL) |
|---|---|
| 26.0 | 137 |
| 33.6 | 197 |
| … | … |
| … | … |
| 35.2 | 89 |

# Unsupervised Meta-Learning with STUNT: Task Generation

**1. We self-generate diverse tasks from the unlabeled dataset.**

- Select random columns.
- **Randomly replace the chosen column features with a random value sampled from the same column.**

**Why?**

- Prevent generating a trivial task as the task label can be directly inferred by the input columns.

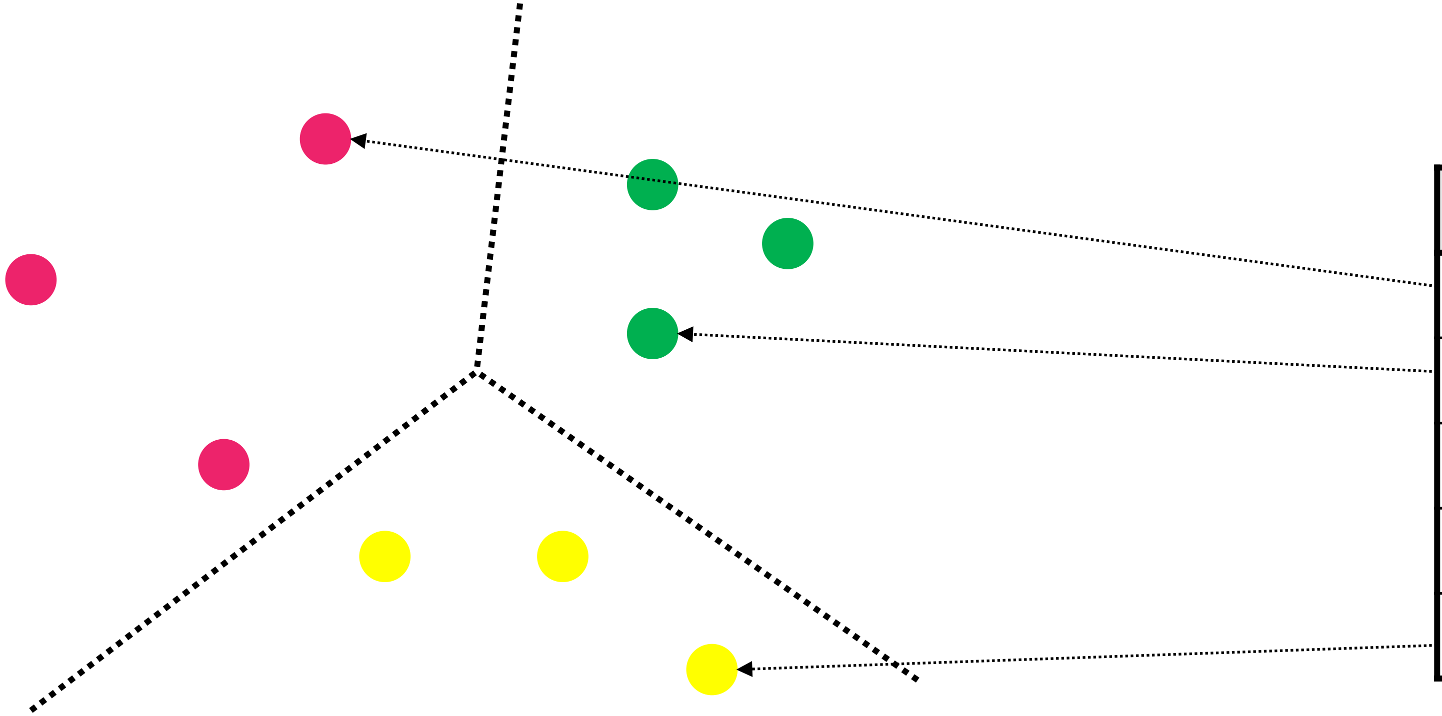
| Blood pressure (mmHg) | BMI ($kg/m^2$) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | **33.6** | **197** | 51 | 135 |
| 68 | **35.2** | **89** | 49 | 168 |
| … | **…** | **…** | … | … |
| … | **…** | **…** | … | … |
| 78 | **26.0** | **137** | 54 | 126 |

| BMI ($kg/m^2$) | Blood sugar (mg/dL) |
|---|---|
| 26.0 | 137 |
| 33.6 | 197 |
| … | … |
| … | … |
| 35.2 | 89 |

# Unsupervised Meta-Learning with STUNT: Task Generation

**1. We self-generate diverse tasks from the unlabeled dataset.**

- Select random columns.
- Randomly replace the chosen column features with a random value sampled from the same column.
- **Generate pseudo-labels by running a k-means clustering over the randomly selected columns.**



| BMI (kg/m$^2$) | Blood sugar (mg/dL) |
|---|---|
| 26.0 | 137 |
| 33.6 | 197 |
| … | … |
| … | … |
| 35.2 | 89 |

# Unsupervised Meta-Learning with STUNT: Task Generation

**1. We self-generate diverse tasks from the unlabeled dataset.**

- Select random columns.
- Randomly replace the chosen column features with a random value sampled from the same column.
- Generate pseudo-labels by running a k-means clustering over the randomly selected columns.
- **The final self-generated task is as below.**

| Blood pressure (mmHg) | BMI (kg/m$^2$) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | **33.6** | **197** | 51 | 135 |
| 68 | **35.2** | **89** | 49 | 168 |
| … | **…** | **…** | … | … |
| … | **…** | **…** | … | … |
| 78 | **26.0** | **137** | 54 | 126 |

| Pseudo-label |
|---|
| |
| |
| … |
| … |
| |

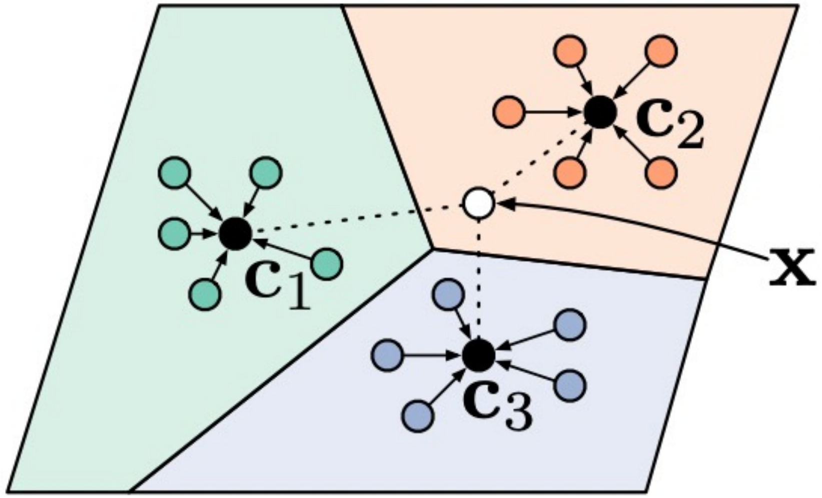
# Unsupervised Meta-Learning with STUNT: Meta-Learning

**2. Based on the generated task, we meta-learn the network by utilizing Prototypical Network.**

- Learn the embedding space in which classification is performed by computing distances to prototypes of each class.

## Why Prototypical Network?

- Allows us to search the effective centroid number rather than fixing it to the class size.
- Model- and data-agnostic.
- Known to outperform advanced meta-learning schemes under various datasets.



# Unsupervised Meta-Learning with STUNT: Meta-Learning

**2. Based on the generated task, we meta-learn the network by utilizing Prototypical Network.**

- Learn the embedding space in which classification is performed by computing distances to prototypes of each class.
- **For a given task generated with STUNT, sample two disjoint sets $\boldsymbol{S}$ and $\boldsymbol{Q}$.**

$\tilde{\mathbf{x}}_u$

| Blood pressure (mmHg) | BMI (kg/m²) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | 33.6 | 197 | 51 | 135 |
| 68 | 35.2 | 89 | 49 | 168 |
| … | … | … | … | … |
| … | … | … | … | … |
| 78 | 26.0 | 137 | 54 | 126 |

$\tilde{\mathbf{y}}_u$

| Pseudo-label | |
|---|---|
| | |
| | $\boldsymbol{S}$ |
| … | |
| … | |
| | $\boldsymbol{Q}$ |

# Unsupervised Meta-Learning with STUNT: Meta-Learning

**2. Based on the generated task, we meta-learn the network by utilizing Prototypical Network.**

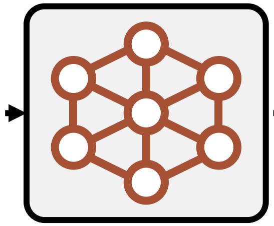
- Learn the embedding space in which classification is performed by computing distances to prototypes of each class.
- For a given task generated with STUNT, sample two disjoint sets $S$ and $Q$.
- **Construct the classifier over the parametrized embedding by using prototypes of each pseudo-class.**

$$f_\theta(y=\tilde{c}|\mathbf{x};\mathcal{S}) = \frac{\exp(-\|z_\theta(\mathbf{x})-\mathbf{p}_{\tilde{c}}\|_2)}{\sum_{\tilde{c}'}\exp(-\|z_\theta(\mathbf{x})-\mathbf{p}_{\tilde{c}'}\|_2)} \qquad \mathbf{p}_{\tilde{c}} := \frac{1}{|\mathcal{S}_{\tilde{c}}|}\sum_{(\tilde{\mathbf{x}}_u,\tilde{\mathbf{y}}_u)\in\mathcal{S}_{\tilde{c}}} z_\theta(\tilde{\mathbf{x}}_u)$$

$S_{\tilde{c}}$: Contains samples with pseudo-class $\tilde{c}$

$\tilde{\mathbf{x}}_u$

| Blood pressure (mmHg) | BMI (kg/m$^2$) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | 33.6 | 197 | 51 | 135 |
| 68 | 35.2 | 89 | 49 | 168 |
| … | … | … | … | … |
| … | … | … | … | … |
| 78 | 26.0 | 137 | 54 | 126 |



Parameterized Embedding

$z_\theta(\tilde{\mathbf{x}}_u)$

# Unsupervised Meta-Learning with STUNT: Meta-Learning

**2. Based on the generated task, we meta-learn the network by utilizing Prototypical Network.**

- Learn the embedding space in which classification is performed by computing distances to prototypes of each class.
- For a given task generated with STUNT, sample two disjoint sets $S$ and $Q$.
- Construct the classifier over the parametrized embedding by using prototypes of each pseudo-class.
- **Train the constructed classifier by computing the cross-entropy loss:** $\mathcal{L}_{\texttt{meta}}(\theta, \mathcal{Q}) := \sum_{(\tilde{\mathbf{x}}_u,\tilde{\mathbf{y}}_u)\in\mathcal{Q}} \mathcal{L}_{\texttt{CE}}\big(f_\theta(\tilde{\mathbf{x}}_u;\mathcal{S}),\tilde{\mathbf{y}}_u\big)$

$\tilde{\mathbf{x}}_u$

| Blood pressure (mmHg) | BMI (kg/m$^2$) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | 33.6 | 197 | 51 | 135 |
| 68 | 35.2 | 89 | 49 | 168 |
| … | … | … | … | … |
| … | … | … | … | … |
| 78 | 26.0 | 137 | 54 | 126 |

$\tilde{\mathbf{y}}_u$

| Pseudo-label | |
|---|---|
| | $S$ |
| | |
| … | |
| … | $Q$ |
| | |

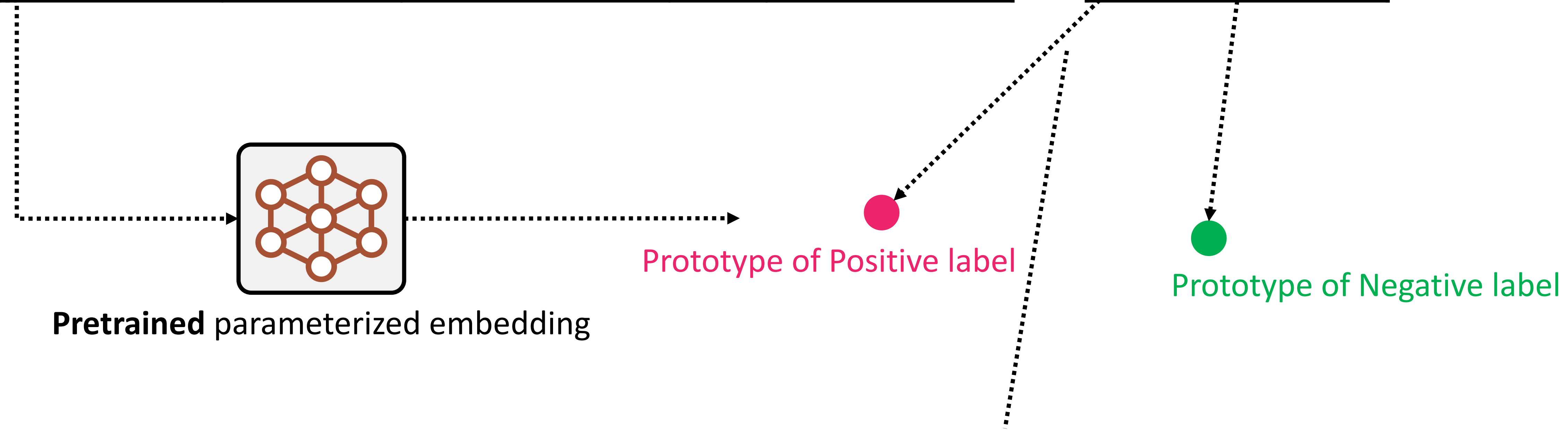
# Unsupervised Meta-Learning with STUNT: Adaptation

**3. Use labeled set to construct the classifier for the few-shot classification.**

- For example, prototypes of a real labels in a 1-shot setup are as below.

| Blood pressure (mmHg) | BMI (kg/m$^2$) | Blood sugar (mg/dL) | Age | Insulin (mmU/ml) |
|---|---|---|---|---|
| 74 | 33.6 | 197 | 51 | 135 |
| 68 | 35.2 | 89 | 49 | 168 |

| Diabetes (Real label) |
|---|
| Positive |
| Negative |




**Pretrained** parameterized embedding

# Experiments: STUNT is effective for 1-shot classification

We report 1-shot test accuracy (%) across various datasets

- We compare to supervised, self-/semi-supervised and unsupervised meta-learning methods.
- Checkmark indicates the use of additional labeled samples for validation.

| Type | Method | Val. | income | cmc | karhunen | optdigit | diabetes | semeion | pixel | dna | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # shot = 1 | | | | | | | | | | | |
| Sup. | CatBoost | ✓ | 57.00 | 34.60 | 55.67 | 61.32 | 60.02 | 43.21 | 59.16 | 41.35 | 52.06 |
| | MLP | ✓ | 60.52 | 35.06 | 48.67 | 61.02 | 57.25 | 40.88 | 55.62 | 44.39 | 50.43 |
| | LR | ✓ | 59.64 | 35.08 | 55.05 | 65.19 | 57.61 | 42.90 | 59.71 | 44.28 | 52.43 |
| | kNN | - | 61.22 | 34.99 | 54.42 | 65.58 | 58.56 | 44.35 | 61.48 | 42.67 | 52.82 |
| Semi-sup. | Mean Teacher | ✓ | 60.63 | 35.58 | 54.57 | 66.10 | 58.05 | 43.56 | 61.02 | 46.58 | 53.26 |
| | ICT | ✓ | 61.83 | 36.53 | 58.37 | 69.12 | 58.08 | 43.48 | 60.88 | 46.55 | 54.36 |
| | Pseudo-Label | ✓ | 60.52 | 34.97 | 49.44 | 61.50 | 57.03 | 41.42 | 56.12 | 44.26 | 50.66 |
| | MPL | ✓ | 60.85 | 35.13 | 47.66 | 61.52 | 57.39 | 41.82 | 56.01 | 44.22 | 50.58 |
| | VIME-Semi | ✓ | 56.40 | 32.97 | 57.40 | 66.85 | 58.16 | 40.43 | 52.86 | 39.18 | 50.53 |
| Self-sup. | SubTab + Fine-tune | ✓ | 59.74 | 35.65 | 41.11 | 49.88 | 59.35 | 30.49 | 42.23 | 40.86 | 44.91 |
| | SubTab + LR | ✓ | 61.88 | 35.68 | 50.32 | 67.05 | 58.06 | 40.27 | 60.40 | 45.68 | 52.42 |
| | SubTab + kNN | - | 61.58 | 35.87 | 48.74 | 66.05 | 59.22 | 39.99 | 61.30 | 44.16 | 52.36 |
| | VIME + Fine-tune | ✓ | 60.50 | 34.98 | 47.50 | 61.31 | 57.23 | 41.09 | 53.79 | 44.30 | 50.09 |
| | VIME + LR | ✓ | 61.99 | 35.30 | 59.62 | 70.52 | 56.95 | 47.20 | 64.17 | 51.36 | 55.89 |
| | VIME + kNN | - | 62.16 | 35.55 | 58.56 | 69.31 | 58.35 | 46.99 | 64.62 | 50.29 | 55.78 |
| Unsup.-Meta. | UMTRA | - | 57.23 | 35.46 | 49.05 | 49.87 | 57.64 | 26.33 | 34.26 | 25.13 | 41.87 |
| | SES | - | 56.39 | 34.59 | 49.19 | 56.30 | 59.97 | 33.73 | 49.19 | 39.56 | 47.37 |
| | CACTUs | - | **64.02** | 36.10 | 65.59 | 71.98 | 58.92 | 48.96 | 67.61 | 65.93 | 59.89 |
| | **STUNT (Ours)** | - | 63.52 | **37.10** | **71.20** | **76.94** | **61.08** | **55.91** | **79.05** | **66.20** | **63.88** |

# Experiments: STUNT is effective for 5-shot classification

We report 5-shot test accuracy (%) across various datasets

- We compare to supervised, self-/semi-supervised and unsupervised meta-learning methods.
- Checkmark indicates the use of additional labeled samples for validation.

| Type | Method | Val. | income | cmc | karhunen | optdigit | diabetes | semeion | pixel | dna | Avg. |
|---|---|---|---|---|---|---|---|---|---|---|---|
| # shot = 5 | | | | | | | | | | | |
| Sup. | CatBoost | ✓ | 64.51 | 39.75 | 82.38 | 84.05 | 65.75 | 68.69 | 84.49 | 63.46 | 69.14 |
| | MLP | ✓ | 66.25 | 37.40 | 77.56 | 83.30 | 64.32 | 66.25 | 81.97 | 59.73 | 67.10 |
| | LR | ✓ | 66.53 | 37.15 | 81.02 | 86.22 | 64.19 | 67.87 | 85.02 | 58.88 | 68.36 |
| | kNN | - | 70.49 | 38.56 | 79.98 | 84.89 | 67.32 | 68.33 | 84.02 | 61.45 | 69.38 |
| Semi-sup. | Mean Teacher | ✓ | 67.05 | 37.73 | 81.08 | 86.66 | 65.45 | 69.67 | 85.24 | 61.47 | 69.29 |
| | ICT | ✓ | 70.13 | 38.09 | 84.58 | 87.01 | 65.47 | 70.26 | 86.12 | 63.37 | 70.63 |
| | Pseudo-Label | ✓ | 66.26 | 37.49 | 78.60 | 83.71 | 64.46 | 67.49 | 82.94 | 60.06 | 67.63 |
| | MPL | ✓ | 67.61 | 37.47 | 77.85 | 83.70 | 64.51 | 67.08 | 82.39 | 59.65 | 67.53 |
| | VIME-Semi | ✓ | 65.13 | 37.32 | 80.53 | 87.13 | 65.39 | 64.80 | 82.83 | 52.08 | 66.90 |
| Self-sup. | SubTab + Fine-tune | ✓ | 66.01 | 37.60 | 67.80 | 75.40 | 66.69 | 56.46 | 75.34 | 55.62 | 62.62 |
| | SubTab + LR | ✓ | 70.12 | 37.67 | 73.25 | 86.07 | 64.92 | 61.34 | 82.14 | 58.90 | 66.80 |
| | SubTab + kNN | - | 71.91 | 39.51 | 69.56 | 83.60 | 68.79 | 59.87 | 80.13 | 61.57 | 66.87 |
| | VIME + Fine-tune | ✓ | 65.97 | 37.25 | 77.82 | 83.13 | 64.40 | 63.63 | 81.01 | 59.58 | 66.60 |
| | VIME + LR | ✓ | 67.80 | 37.51 | 82.87 | 87.42 | 64.29 | 71.53 | 86.79 | 69.62 | 70.98 |
| | VIME + kNN | - | 72.16 | 39.28 | 79.15 | 83.86 | 66.94 | 68.45 | 84.07 | 71.09 | 70.63 |
| Unsup.-Meta. | UMTRA | - | 65.78 | 38.05 | 67.28 | 73.29 | 64.41 | 35.90 | 51.32 | 25.08 | 52.64 |
| | SES | - | 68.27 | 39.04 | 74.80 | 78.46 | 66.61 | 52.74 | 74.80 | 52.25 | 63.37 |
| | CACTUs | - | 72.03 | 38.81 | 82.20 | 85.92 | 66.79 | 65.00 | 85.25 | **81.52** | 72.19 |
| | **STUNT (Ours)** | - | **72.69** | **40.40** | **85.45** | **88.42** | **69.88** | **73.02** | **89.08** | 79.18 | **74.77** |

# Experiments: STUNT is effective for multi-task learning

STUNT can **instantly be adapted to multiple tasks** at test-time without further training the network.

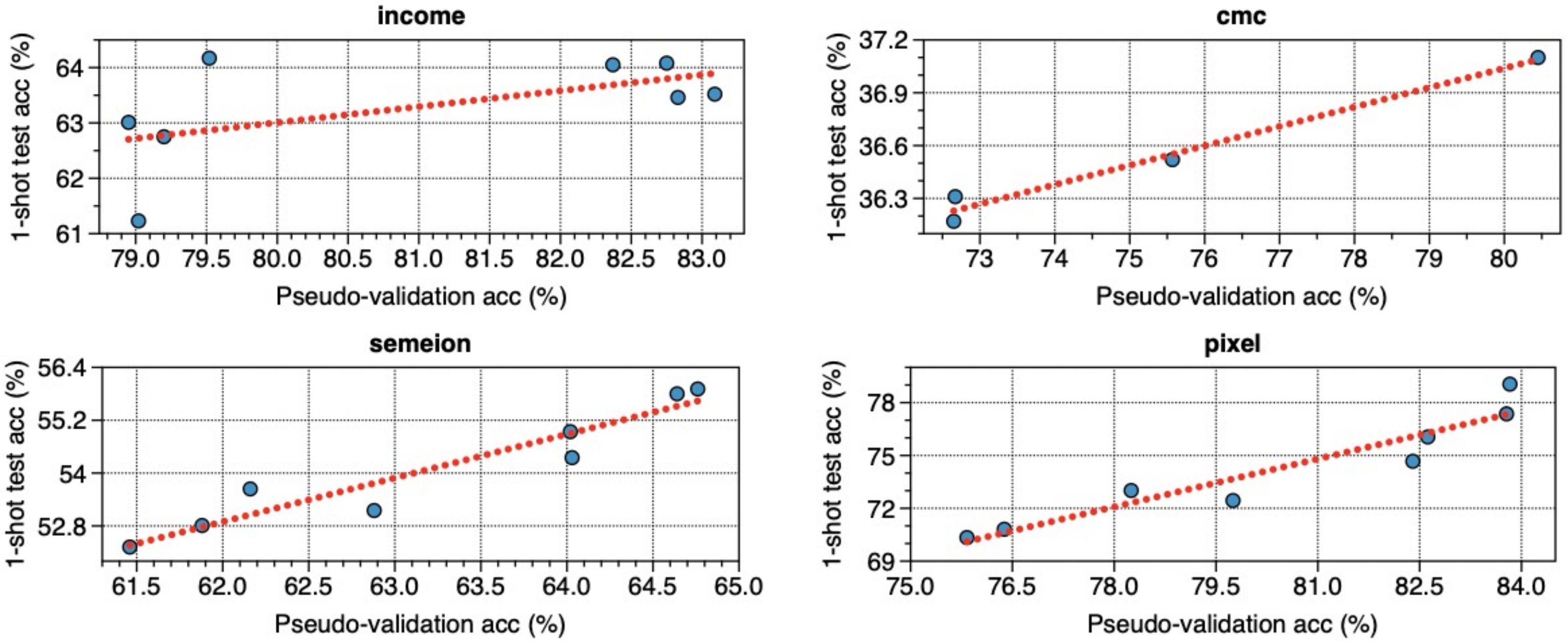
- We report few-shot test accuracy (%) on the emotions dataset consists of 6 binary classification tasks.

| Method \ Task | amazed-surprised | happy-please | relaxing-calm | quiet-still | sad-lonely | angry-aggressive | Avg. |
|---|---|---|---|---|---|---|---|
| # shot = 1 | | | | | | | |
| kNN | 59.04 | 47.14 | 55.77 | 66.86 | 55.96 | 59.47 | 57.37 |
| SubTab + kNN | **63.32** | 48.88 | 56.46 | 62.56 | 54.34 | 57.99 | 57.26 |
| VIME + kNN | 60.07 | 49.51 | 55.62 | 64.74 | 53.95 | 60.29 | 57.36 |
| CACTUs | 61.58 | 50.67 | 55.63 | 63.18 | 55.10 | 59.39 | 57.59 |
| **STUNT (Ours)** | 62.71 | **51.63** | **59.28** | **69.34** | **56.38** | **63.43** | **60.46** |
| # shot = 5 | | | | | | | |
| kNN | 70.71 | 53.48 | 66.34 | 81.03 | 68.51 | 68.07 | 68.02 |
| SubTab + kNN | **74.41** | 52.23 | 64.90 | 72.70 | 62.32 | 63.30 | 64.98 |
| VIME + kNN | 70.71 | 53.10 | 66.24 | 79.54 | 66.34 | 67.76 | 67.28 |
| CACTUs | 71.41 | 53.64 | 65.18 | 77.57 | 64.15 | 66.57 | 66.42 |
| **STUNT (Ours)** | 72.38 | **55.09** | **67.39** | **83.10** | **68.61** | **70.10** | **69.45** |

# Experiments: Effectiveness of Pseudo-Validation with STUNT

Pseudo-validation with STUNT is effective for **hyperparameter searching & early stopping**.

- Pseudo-validation accuracy (%) and test accuracy (%) have a positive correlation.




- Evaluating the early stop model achieves better accuracy (%) than evaluating the model after final train step.

| | income | | cmc | | semeion | | pixel | |
|---|---|---|---|---|---|---|---|---|
| Problem | Last | Early | Last | Early | Last | Early | Last | Early |
| 1-shot | 61.58 | **63.52** | 36.94 | **37.10** | 51.94 | **55.91** | 74.92 | **79.05** |
| 5-shot | 70.84 | **72.69** | **40.43** | 40.40 | 71.55 | **73.02** | 87.60 | **89.08** |

# Further application: Few-shot Tabular Regression with STUNT

🤔 Can STUNT be extended to solve few-shot tabular regression tasks?

💡 Idea: Replace the ProtoNet classifier with a kNN regressor at the adaptation stage.

- STUNT is a competitive approach in few-shot tabular regression task.
- Performance gap is often vacuous because STUNT meta-train networks with classification tasks.
- We report the mean squared error on 5 datasets.

| Input | news | abalone | cholesterol | sarcos | boston |
|---|---|---|---|---|---|
| # shot = 5 | | | | | |
| Raw | 2.74E-04 | 1.75E-02 | 1.37E-02 | **1.05E-02** | 3.65E-02 |
| VIME | 2.69E-04 | 1.70E-02 | 1.37E-02 | 1.06E-02 | **3.53E-02** |
| CACTUs | 2.75E-04 | 1.72E-02 | 1.46E-02 | 1.06E-02 | 3.76E-02 |
| **STUNT (Ours)** | **2.68E-04** | **1.66E-02** | **1.35E-02** | 1.06E-02 | 3.70E-02 |
| # shot = 10 | | | | | |
| Raw | **2.53E-04** | 1.49E-02 | 1.13E-02 | 9.21E-03 | 2.88E-02 |
| VIME | **2.53E-04** | 1.49E-02 | 1.13E-02 | 9.24E-03 | **2.78E-02** |
| CACTUs | 2.54E-04 | 1.51E-02 | 1.21E-02 | **9.16E-03** | 2.94E-02 |
| **STUNT (Ours)** | **2.53E-04** | **1.46E-02** | **1.12E-02** | **9.16E-03** | 2.90E-02 |

# STUNT: Simple & Effective Framework for Few-shot Tabular Learning

Summary: We propose a simple yet effective framework for few-shot semi-supervised tabular learning.

We propose STUNT = Self-generated Tasks from Unlabeled Table for unsupervised meta-learning

1. Generate a diverse tasks from the unlabeled data by treating a column feature as a useful pseudo-label
2. Few-shot classification: Effective without using a labeled validation set
3. Multi-task learning: Can be instantly adapted to multiple tasks without further training
4. Pseudo-validation with STUNT is effective for hyperparameter searching and early stopping


| Signal | Material | AOD |
| --- | --- | --- |
| … | … | … |
| … | … | … |


| Blood Pressure | BMI | Age |
| --- | --- | --- |
| … | … | … |
| … | … | … |

| Credit history | Job | Housing |
| --- | --- | --- |
| … | … | … |
| … | … | … |