# Grapevine Disease Prediction Using Climate Variables from Multi-Sensor Remote Sensing Imagery via a Transformer Model

Weiying Zhao, Natalia Efremova

## Motivation

- Early detection and management of grapevine diseases are important in pursuing sustainable viticulture.
- Solving the complex interdependencies between viticulture disease and multiscale climate variables (Figure 1).
- Integration of advanced machine learning and environmental factors (Figure 2).
- Addressing the challenges of small and imbalanced datasets.

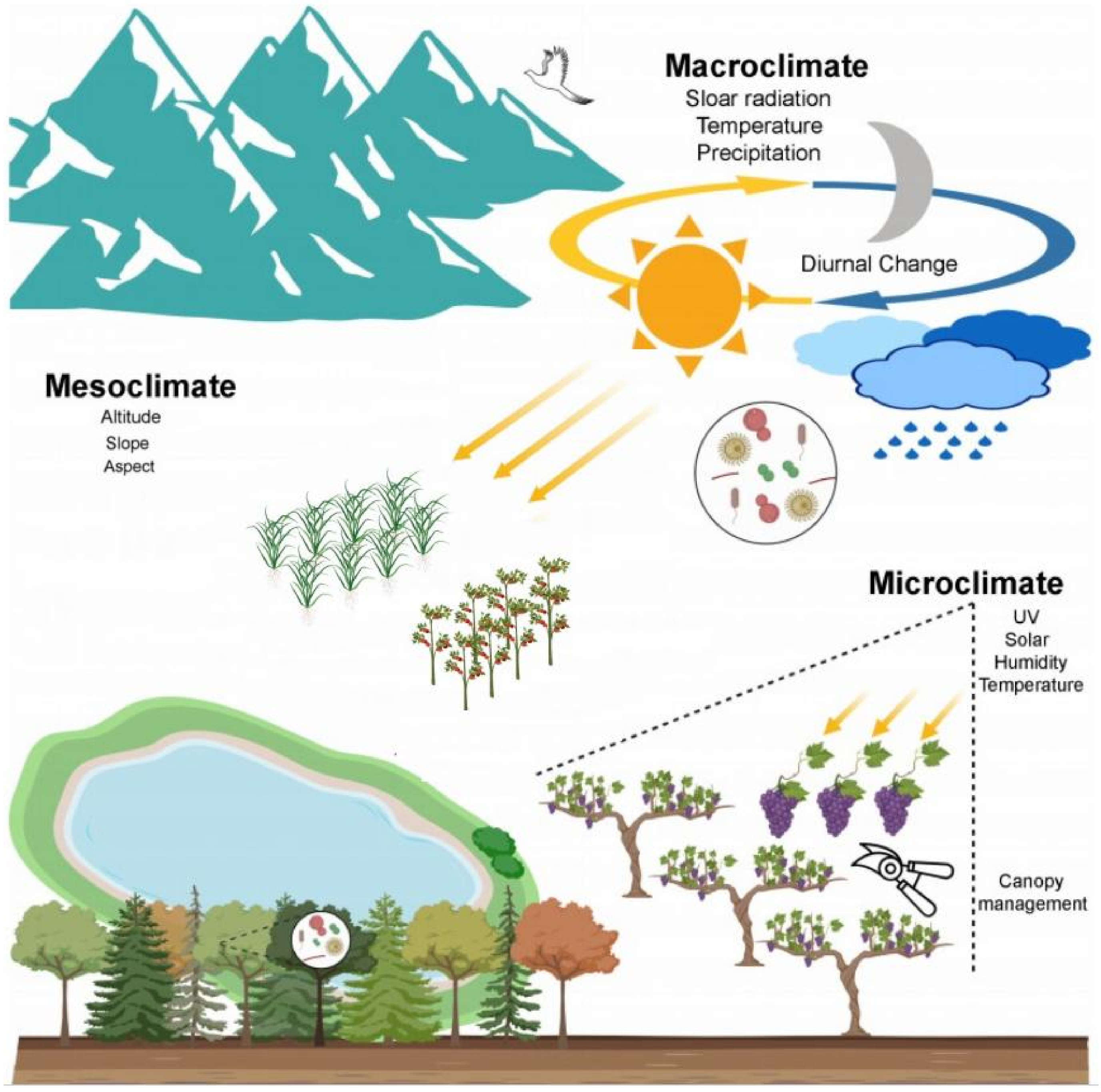


Figure 1. A scenario of microbial biogeography for grapevine. The surrounding ecosystem influences the grapevine microbiota, including soil, air, and local flora. Geographic isolation, driven by reduced gene flow over long distances due to physical barriers and animal vectors like insects, is a key factor. Climate, a critical environmental force, shapes microbial distribution patterns, impacting wine quality. Regional macroclimates affect bacterial and fungal patterns, while microclimate variations at the vineyard level have less effect, particularly on bacteria.[4]

## Methodology

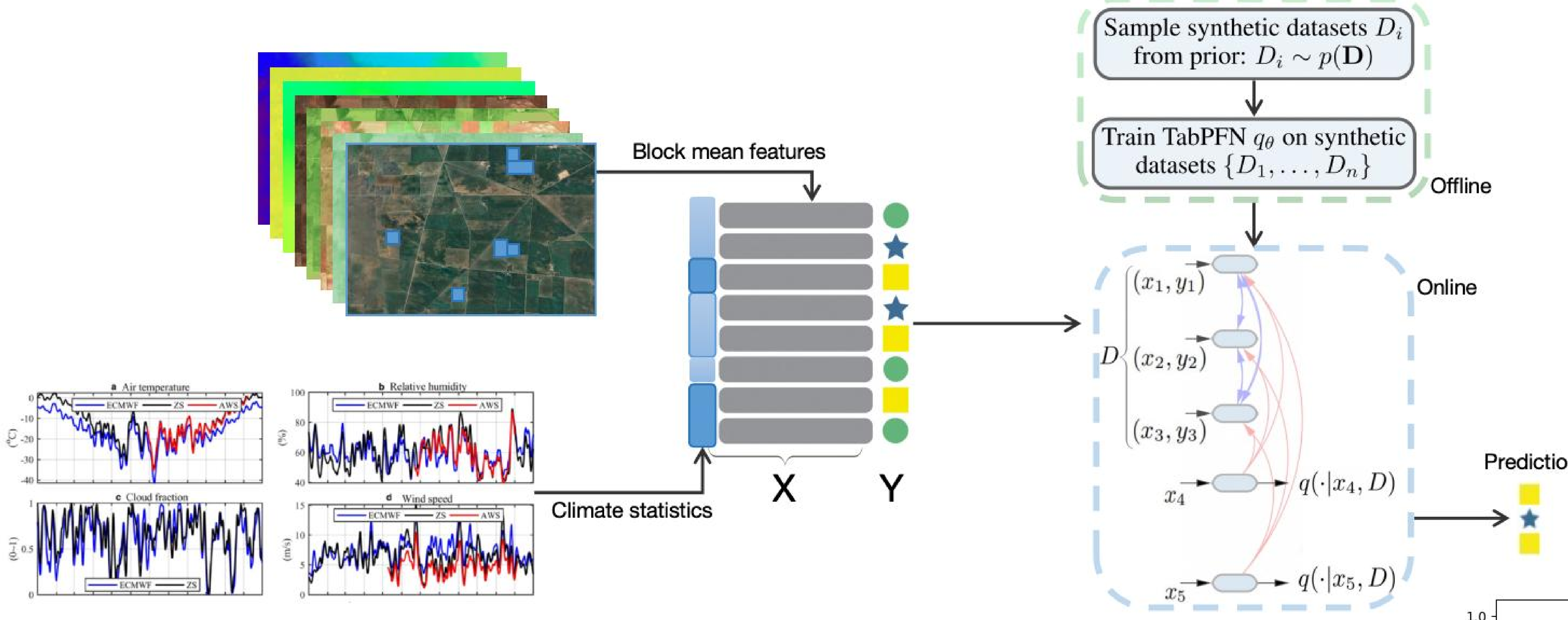
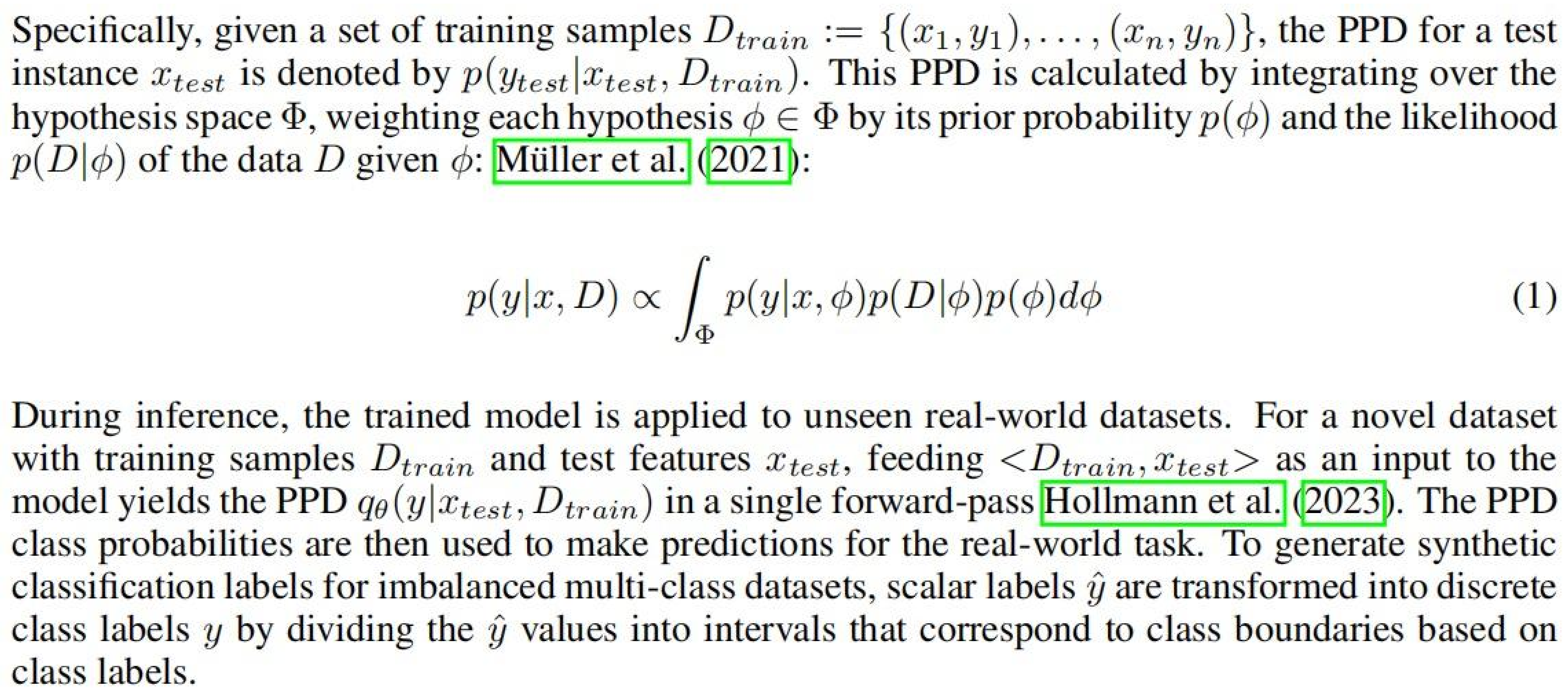


Figure 2. Flowchart of the proposed disease forecasting framework. The TabPFN method [1-2] is used as an example. The TabPFN learns to approximate the PPD of a given prior in the offline stage to yield predictions on a new dataset in a single forward pass in the online stage.[3]



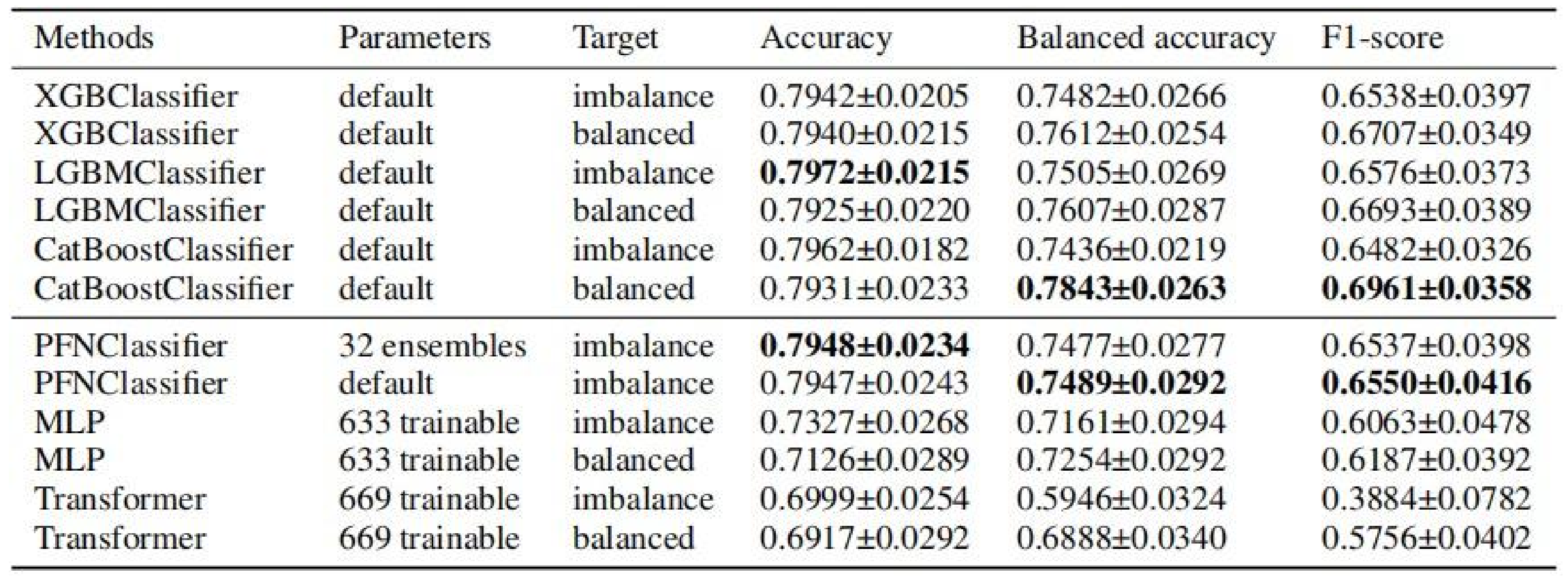
Specifically, given a set of training samples $D_{train} := \{(x_1, y_1), \ldots, (x_n, y_n)\}$, the PPD for a test instance $x_{test}$ is denoted by $p(y_{test}|x_{test}, D_{train})$. This PPD is calculated by integrating over the hypothesis space $\Phi$, weighting each hypothesis $\phi \in \Phi$ by its prior probability $p(\phi)$ and the likelihood $p(D|\phi)$ of the data $D$ given $\phi$: Müller et al. (2021):

$$p(y|x, D) \propto \int_{\Phi} p(y|x, \phi)p(D|\phi)p(\phi)d\phi \tag{1}$$

During inference, the trained model is applied to unseen real-world datasets. For a novel dataset with training samples $D_{train}$ and test features $x_{test}$, feeding $<D_{train}, x_{test}>$ as an input to the model yields the PPD $q_\theta(y|x_{test}, D_{train})$ in a single forward-pass Hollmann et al. (2023). The PPD class probabilities are then used to make predictions for the real-world task. To generate synthetic classification labels for imbalanced multi-class datasets, scalar labels $\hat{y}$ are transformed into discrete class labels $y$ by dividing the $\hat{y}$ values into intervals that correspond to class boundaries based on class labels.

## Experimental results

Table 1. Performance of algorithms over the disease dataset. The assessment includes mean and standard deviation for accuracy, balanced accuracy, and F1-score, computed over 40 iterations with varied seeding for dataset partitioning, to present a comprehensive picture of each model's robustness in balanced and imbalanced classification contexts.



| Methods | Parameters | Target | Accuracy | Balanced accuracy | F1-score |
|---|---|---|---|---|---|
| XGBClassifier | default | imbalance | 0.7942±0.0205 | 0.7482±0.0266 | 0.6538±0.0397 |
| XGBClassifier | default | balanced | 0.7940±0.0215 | 0.7612±0.0254 | 0.6707±0.0349 |
| LGBMClassifier | default | imbalance | **0.7972±0.0215** | 0.7505±0.0269 | 0.6576±0.0373 |
| LGBMClassifier | default | balanced | 0.7925±0.0220 | 0.7607±0.0287 | 0.6693±0.0389 |
| CatBoostClassifier | default | imbalance | 0.7962±0.0182 | 0.7436±0.0219 | 0.6482±0.0326 |
| CatBoostClassifier | default | balanced | 0.7931±0.0233 | **0.7843±0.0263** | **0.6961±0.0358** |
| PFNClassifier | 32 ensembles | imbalance | **0.7948±0.0234** | 0.7477±0.0277 | 0.6537±0.0398 |
| PFNClassifier | default | imbalance | 0.7947±0.0243 | **0.7489±0.0292** | **0.6550±0.0416** |
| MLP | 633 trainable | imbalance | 0.7327±0.0268 | 0.7161±0.0294 | 0.6063±0.0478 |
| MLP | 633 trainable | balanced | 0.7126±0.0289 | 0.7254±0.0292 | 0.6187±0.0392 |
| Transformer | 669 trainable | imbalance | 0.6999±0.0254 | 0.5946±0.0324 | 0.3884±0.0782 |
| Transformer | 669 trainable | balanced | 0.6917±0.0292 | 0.6888±0.0340 | 0.5756±0.0402 |

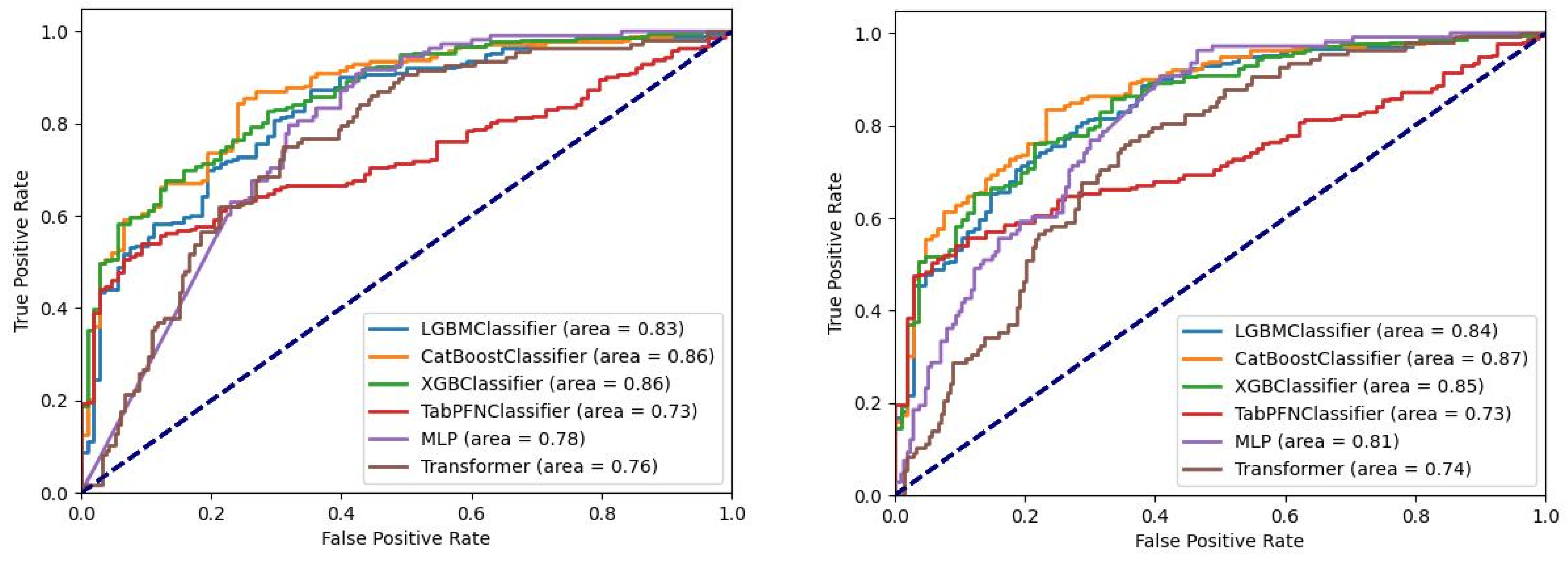


Figure 3. ROC curve results comparison of the methods with imbalance data (left) and with balanced data (right). The comparison are based on the same training and testing databases.

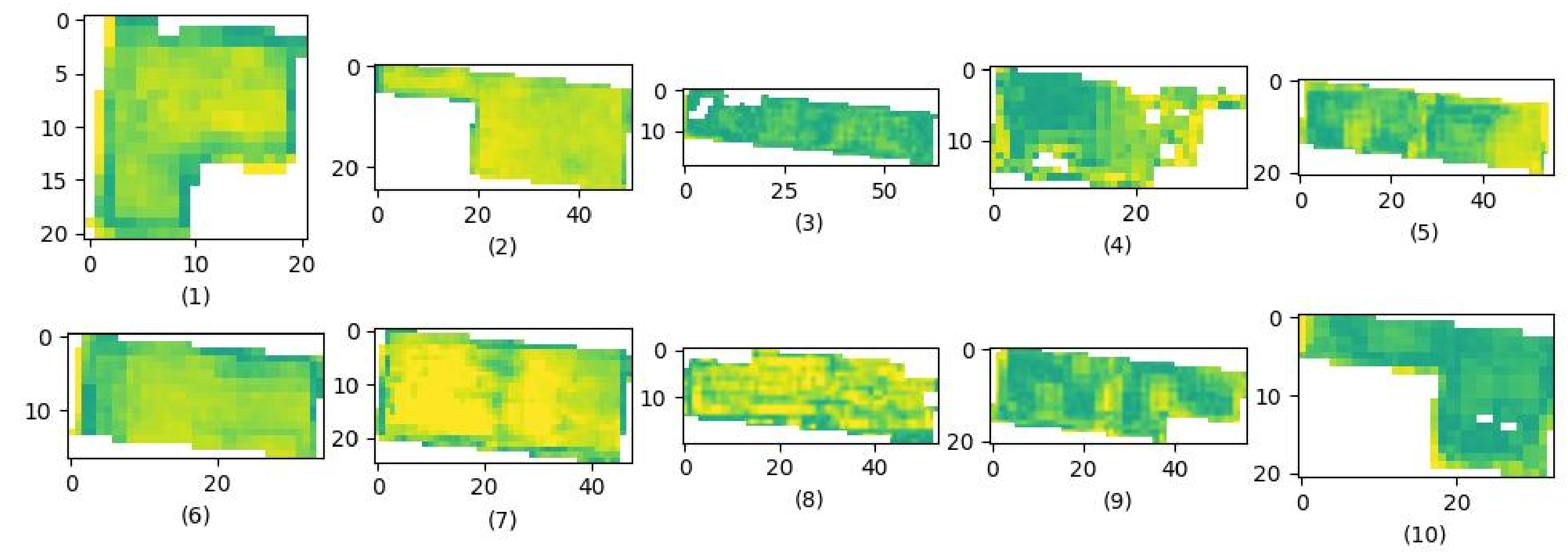


Figure 4. Disease probability maps for 10 blocks in early 2021 in Australia. The probability of disease is displayed as a visual heat map (green = low probability, yellow = high probability).

## References

[1] Samuel Muller, Noah Hollmann, Sebastian Pineda Arango, Josif Grabocka, and Frank Hutter. Transformers can do bayesian inference. arXiv preprint arXiv:2112.10510, 2021.

[2] Noah Hollmann, Samuel Muller, Katharina Eggensperger, and Frank Hutter. TabPFN: A transformer that solves small tabular classification problems in a second. In The Eleventh International Conference on Learning Representations, 2023.

[3] Cyril Picard and Faez Ahmed. Fast and accurate zero-training classification for tabular engineering data. arXiv preprint arXiv:2401.06948, 2024.

[4] Di Liu, Pangzhen Zhang, Deli Chen, and Kate Howell. From the vineyard to the winery: how microbial ecology drives regional distinctiveness of wine. Frontiers in Microbiology, 10:2679,2019.