Transforming GPP Estimation in Terrestrial Ecosystems using Remote Sensing and Transformers

Yanting Zheng and Ryan Rad
Khoury College of Computer Science, Northeastern University
Vancouver, Canada
{zheng.yant, r.rad}@northeastern.edu

***Abstract*—Gross Primary Productivity (GPP) is a critical measure of carbon uptake by terrestrial vegetation, essential for understanding the global carbon cycle and developing climate mitigation strategies. This study introduces a novel approach to estimating annual GPP in Europe and North America using a deep tabular model that integrates remote sensing data from Google Earth Engine with the FT-Transformer architecture. We utilized transfer learning to pre-train the model on extensive MOD17A2H/A3H (MOD17) data and fine-tuned it with FLUXNET station data, addressing the scarcity of in-situ measurements. Our results demonstrate that the proposed model provides more accurate GPP estimates compared to traditional MOD17 values, reflecting a closer alignment with FLUXNET observations. This enhanced accuracy highlights the model's ability to capture complex ecological and climatic interactions, offering a promising tool for advancing our understanding of terrestrial carbon dynamics. The full code repository is available[1].**

***Index Terms*—terrestrial carbon sequestration, gross primary productivity, Google Earth engine, transformer, transfer learning**

I. Introduction

Gross Primary Production (GPP) represents the total carbon dioxide captured by land plants over time through the photosynthetic conversion of into organic compounds [1]. The terrestrial ecosystem, constituting the most complex carbon reservoir, stores 25%-30% of anthropogenic emissions [2], playing a pivotal role in upholding the global carbon cycle and mitigating the impacts of climate change [3]. Accurate quantification of GPP is essential for evaluating ecosystem carbon balance and conducting climate change research. Additionally, a thorough understanding of global carbon exchange facilitates the assessment of the support capacities of terrestrial ecosystems for achieving the sustainable development of human society.

Currently, it is impossible to measure GPP directly at a scale beyond the leaf level. Eddy Covariance (EC) methods are recognized as a standard method for indirectly estimating GPP by measuring the exchange between land surface and atmosphere at a larger scale. However, despite the establishment of a global flux observation network system, its sparse distribution across vast land areas still necessitates the development of simulation methods for estimating GPP at a larger scale. Among the simulation methods, Light Use Efficiency (LUE) models, which consider physiological and ecological processes involved in photosynthesis, have been widely used for estimating GPP at regional and global scales when combined with remote sensing data. The MODIS team's MOD17 global GPP/NPP product is derived from such an approach. However, LUE models can be limited since they rely on human understanding to select constant values, model parameterization schemes, and model structures, which can be potentially biased and limited within the known domain. In contrast, machine learning models, which limit the uncertainties associated with traditional empirical process models, have gained more attention recently to uncover more generalized relationships between influential factors and the GPP values.

The contributions of this work can be summarized as follows:

- *Enhanced Model Objectivity*: Our approach minimizes bias by using point samples and a transformer-like architecture, moving away from traditional time-series CNN methods, and allowing for more objective data interpretation.
- *Superior Accuracy in GPP Estimations*: The integration of advanced deep tabular architecture and transfer learning significantly improves GPP estimates compared to Gradient Boosted Decision Trees (GBDT) methods and the MOD17 product. Notably, our fine-tuned FT-Transformer model, even without pre-training, outperforms some GBDT methods, demonstrating the robustness and effectiveness of this architecture.

II. Related Work

Many studies have used machine learning methods to estimate GPP. Sarkar et al. (2022) and Hao et al. (2023) both used random forest (RF), while the former also compared it with Support Vector Machine (SVM) and EXtreme Gradient Boosting (XGBoost), stating that RF outperformed the two as well as the estimations of MOD17 [4] [5]. Lee et al. (2020) experimented with SVM, RF, artificial neural network (ANN), and Deep Neural Network (DNN) with data sampled from Korea and concluded that DNN outperformed the others and gave a stable result under abnormal climatic conditions [6]. Wu et al. (2019) used a Convolutional Neural Network (CNN) on time-series data to estimate global forest GPP and got a highly consistent result with the ground observation [7]. Many of them utilized remote sensing data from Google Earth

[1] https://github.com/MiaZhengLS/Estimating-Gross-Primary-Product

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00262

TABLE I: Input Bands: NDVI: normalized difference vegetation index; EVI: the enhanced vegetation index; ET: evapotranspiration; PR: precipitation; LC: landcover; LSTD: land surface temperature (day).

Feature	Dataset	Availability	Frequency	Resolution
GPP	MOD17A3HGF.061	2001-2022.1	Annual	500m
NDVI	MODIS Combined 16-Day NDVI	2000-present	16 days	500m
EVI	MOD13A1.061	2000-present	16 days	500m
ET	TerraClimate	1958-2022.12	Monthly	5000m
PR	TerraClimate	1958-2022.12	Monthly	5000m
LC	MCD12Q1.061	2001-2022.1	Annual	500m
LSTD	MOD11A1.061	2000.2-present	Daily	1000m

TABLE II: Datasets Summary

Stage	Dataset	Time	Amount
Upstream tuning	MOD17 training MOD17 validation	2015-2020 2021	141623 samples 23624 samples
Downstream training	FLUXNET training (4/5) FLUXNET validation (1/5)	2001-2020	996 samples
Test	FLUXNET test	2021-2022	160 samples

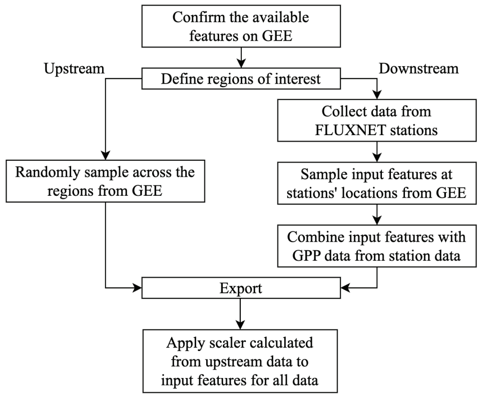


Fig. 1: Pre-processing Workflow

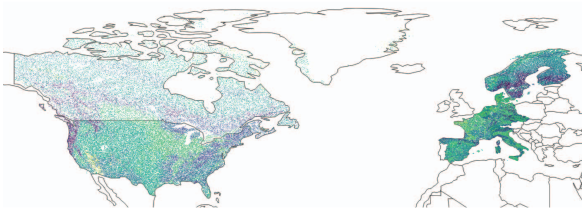


Fig. 2: Upstream MOD17 Sample Locations

Engine (GEE) to obtain features due to its easy access [5] [7]. Some also used MOD17 as a benchmark and conducted a comparative analysis to prove that their estimations outperform those of MOD17.

Recently, Transformer, believed to be the state-of-the-art deep architecture, has shown competitive performance on tabular data problems [8]–[12]. Transformer-like models possess remarkable feature-capturing capabilities, yet they typically demand a substantial volume of data for effective training and satisfactory results. In our study, we harnessed the potential of Feature Tokenizer + Transformer (FT-Transformer) [13], integrating it with transfer learning to leverage extensive upstream data and apply the knowledge to a more limited downstream dataset. We chose MOD17 as the source of upstream data since it is readily accessible from GEE. Our downstream data is from FLUXNET stations. This strategy demonstrated improved prediction results compared to direct training on the downstream dataset, and it also exhibited superior performance over traditionally dominant gradient-boosted decision trees (GBDT) such as XGBoost or CatBoost. Notably, the outcomes are also more consistent with FLUXNET data compared to that from MOD17, emphasizing the effectiveness of our chosen approach.

The data source and methodology are introduced in Section II. The results and discussion are described in Section III. Conclusions and future work are demonstrated in Section IV.

III. Data Source and Methodology

A. Data Source

Our study utilizes upstream and downstream data categories, with input features sourced from Google Earth Engine (GEE) and target features from FLUXNET stations for downstream data. Six input features, detailed in Table I, were selected based on ecological and climatic considerations.

The regions of interest are North America and Europe, chosen for their comprehensive FLUXNET data coverage of 2001-2022 and the diversity of ecosystem types. Our goal is to estimate annual GPP, and we harmonized the temporal resolution of input features accordingly, using medians for each year and annual sums for precipitation.

For effective training of all machine learning models considered in this study, input features were scaled using scikit-learn's RobustScaler based on upstream training data. The overall workflow is illustrated in Fig. 1.

1) Upstream Datasets: Over 140k training and 23k validation samples were collected for upstream training of the FT-Transformer, detailed in Table II. The geographical distribution of upstream samples is shown in Fig. 2 with different colors representing different landcover types.

2) Downstream Datasets: The downstream data comprises $1,156$ site-year samples from 142 stations in North America and 64 stations in Europe, as shown in Table II. The geolocations of these stations are plotted in Fig. 3, with different colors denoting various landcover types. For comparative analysis with MOD17, we also sampled MOD17 values aligned with the site-year information of the test dataset. The unit of FLUXNET data is $g \times C/m^2$, which is different from $kg \times C/m^2$ used by MOD17. So all the downstream target values are converted to make the unit consistent.

B. Methodology

We implemented the FT-Transformer model, following the guidelines of the original study [13], and compared its performance with fine-tuned CatBoost, XGBoost, SVR, SGD, and RF regression models to assess the efficacy of transfer learning.

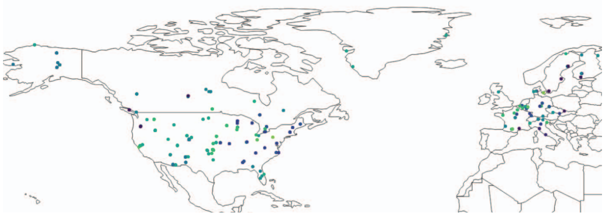


Fig. 3: Downstream FLUXNET Station Locations

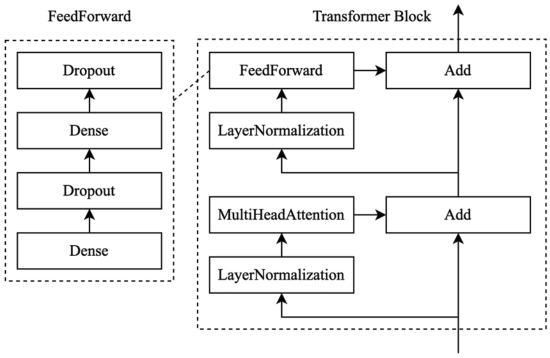


Fig. 4: Transformer Block Structure

1) Model Development: FT-Transformer is a deep learning (DL) model specifically designed for tabular data. It consists of feature tokenizers, multiple transformer layers, and an output layer. The numeric feature tokenizer applies a randomized weight vector and a bias for the numeric input vector. The categorical feature tokenizer works similarly to the numeric feature tokenizer except that it uses a category look-up table, thus the categorical feature number may not be consistent with the output dimension of the categorical tokenizer. The transformer layers are a stack of the same structured layers as shown in Fig. 4. According to [13], the first normalization in the first Transformer layer should be removed to achieve better performance.

Initial experiments were conducted using default hyperparameters, followed by fine-tuning to optimize performance.

2) Transfer Learning Strategy: As shown in [14], FT-Transformer, with transfer learning strategies, can outperform GBDT methods when downstream data is limited and there is a strong correlation between upstream and downstream data. In our case, MOD17 provides a consistent global estimation of GPP that makes it possible to utilize massive upstream data for pre-training. To evaluate the effect of the transfer learning strategy on our problem, we also trained FT-Transformer entirely on downstream data for comparison. For pre-trained models, we experimented with two types of tuning: end-to-end tuning and output layer tuning.

TABLE III: Quantitative Performance Comparison. FT-Transformer(full): Pre-trained, fully trained on downstream; FT-Transformer(head): Pre-trained, only output layer trained on downstream; FT-Transformer(down): Trained solely on downstream, no upstream data.

Model	**RMSE**	**R^2**	**Max Error**
MOD17	4553.657	0.502	15910.398
CatBoost	**3873.608**	**0.640**	13520.283
RF	4415.322	0.532	**11570.286**
SVR	4613.222	0.489	14607.598
XGBoost	4613.490	0.489	13605.111
SGDR	4624.181	0.486	13526.716
FT-Transformer (head)	4090.135	0.598	12817.128
FT-Transformer (down)	4378.650	0.539	19443.531
FT-Transformer (full)	**3760.305**	**0.660**	**11849.963**

IV. Experimental Results and Analysis

A. General Results

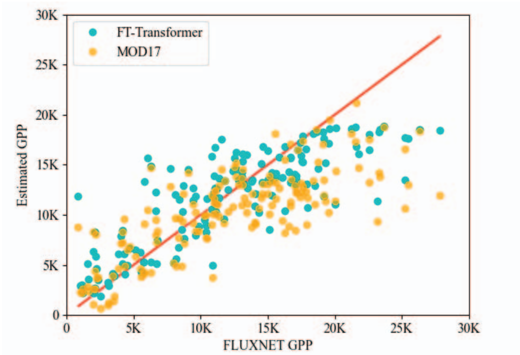
We evaluated the model performance with Root Mean Squared Error (RMSE), coefficient of determination R^2, and Maximum Error (ME). RMSE measures the average error between prediction values and actual values. R^2 indicates the level of correlation between the actual values and the predicted values. ME is the maximum distance among all distances between predicted values and the corresponding actual values, which can reflect the extent of errors. The results are listed in Table III.

Upon analysis of RMSE and R^2, we can observe that end-to-end FT-Transformer and CatBoost gain significantly better performance than other models, including the benchmark MOD17. Notably, CatBoost demonstrates a larger ME, indicating a greater worst-case error in comparison to end-to-end FT-Transformer. Besides the two top models, pre-trained FT-Transformer with tuned output layer, FT-Transformer tuned solely on downstream data and RF also outperform MOD17 across all three metrics. We can also conclude that transfer learning and end-to-end downstream tuning help boost the performance of FT-Transformer by comparing the results of the three FT-Transformer models.

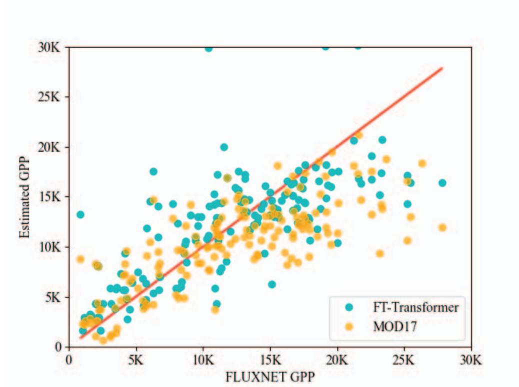
We also visualized the predicted GPP values, MOD17 GPP values, and FLUXNET GPP values on a 2D plane to view the consistency among the three. We presented the results for the FT-Transformer model in Fig. 5a. Results for other models are depicted in Figs. 5b to 5h. On the plotted plane, a closer alignment of data points with the diagonal line indicates higher accuracy. This visual inspection reveals an improved accuracy in matching FLUXNET GPP for the FT-Transformer model when compared to MOD17.

B. Performance for Different Landcover Types

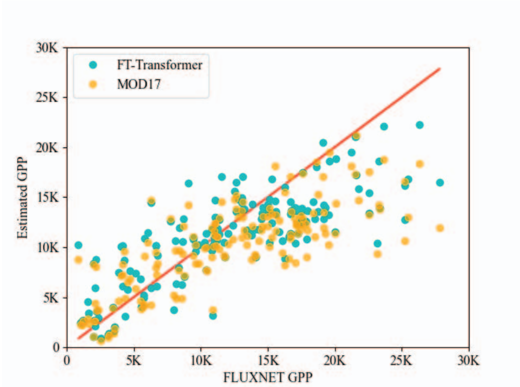
We analyzed the models' performance for different landcover types involved in the test data and visualized the results in Fig. 6. We can see that the estimated GPP of both FT-Transformer and CatBoost are well aligned with FLUXNET GPP in most landcover types. But FT-Transformer predicted a much higher value for EBF type than the observed value. This may be explained by the imbalanced landcover types in the upstream training data and the dominating role of



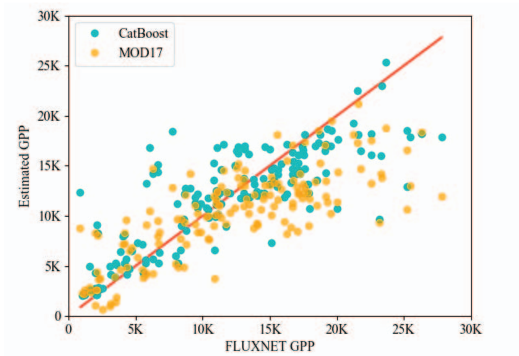
(a) FT-Transformer (end-to-end) vs. MOD17



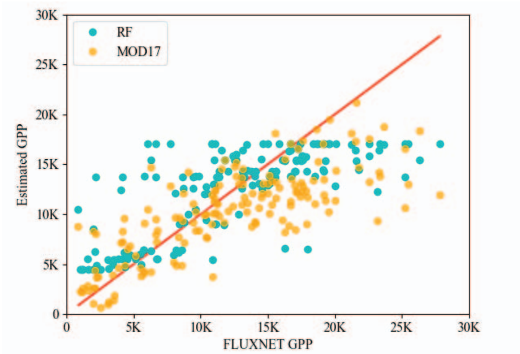
(b) FT-Transformer (down) vs. MOD17



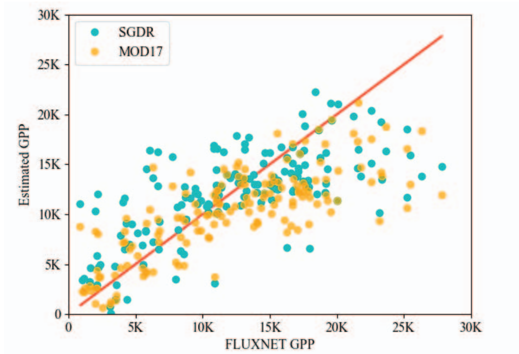
(c) FT-Transformer (head) vs. MOD17



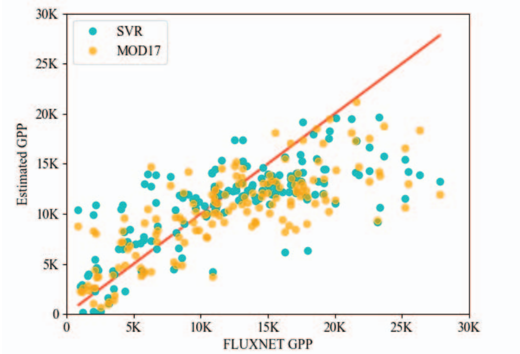
(d) CatBoost vs. MOD17



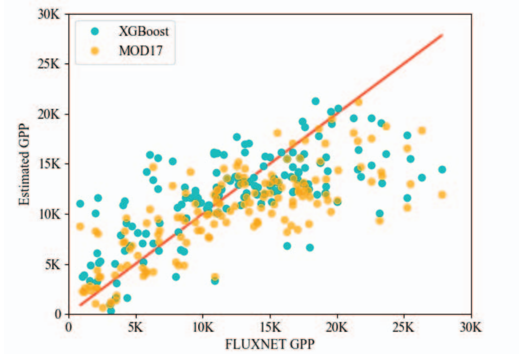
(e) RF vs. MOD17



(f) SGDR vs. MOD17



(g) SVR vs. MOD17



(h) XGBoost vs. MOD17

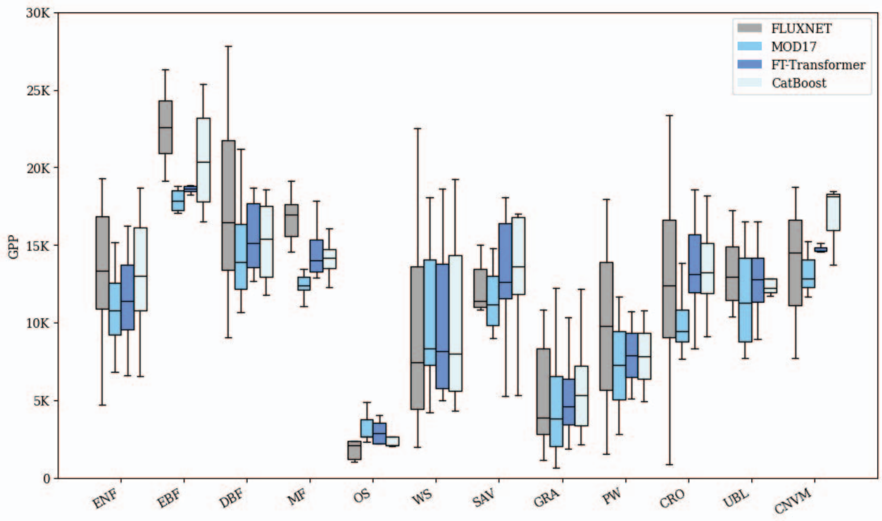


Fig. 6: Performance Comparison based on Landcover Types

ENF: Evergreen Needleleaf Forests; EBF: Evergreen Broadleaf Forests; DBF: Deciduous Broadleaf Forests; MF: Mixed Forests; OS: Open Shrublands; WS: Woody Savannas; SAV: Savannas; GRA: Grasslands; PW: Permanent Wetlands; CRO: Croplands; UBL: Urban and Built-up Lands; CNVM: Cropland/Natural Vegetation Mosaics

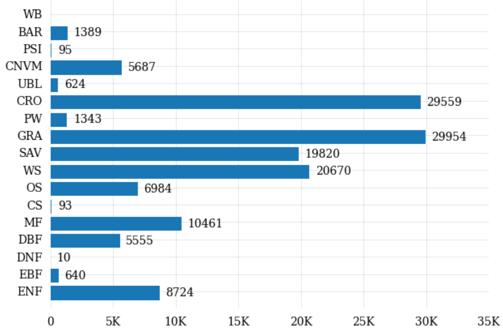


Fig. 7: Distribution of landcover in upstream data

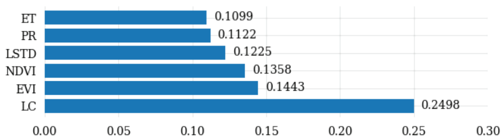


Fig. 8: Feature Importance Analysis

landcover types in estimating GPP values. As displayed in Fig. 7, the upstream training samples with EBF type are very limited, which makes it more difficult for FT-Transformer to learn enough information between this landcover type and the corresponding GPP values. Besides this, we also visualized the feature importance for the FT-Transformer model in Fig. 8 and discovered that landcover type is the most important feature. This means that less landcover information can significantly impact the accuracy of GPP estimation.

C. Discussion

This section evaluates the MOD17 GPP methodology alongside the advantages and drawbacks of the FT-Transformer model in GPP estimation.

MOD17 computes daily GPP with the equation

$$GPP = \epsilon \times APAR$$

where ϵ is the radiation use efficiency coefficient, and APAR is the absorbed photosynthetically active radiation. The efficiency coefficient ϵ is derived from vegetation types, daily minimum temperature (T_{min}), and Vapour Pressure Deficit (VPD). APAR is determined using the equation

$$APAR = IPAR \times FPAR$$

combining 8-day estimates of the Fraction of Photosynthetically Active Radiation (FPAR) from MOD15 with daily Incident Photosynthetically Active Radiation (IPAR) from GMAO/NASA. These calculations are based on specific assumptions, potentially limiting their ability to fully capture complex environmental dynamics.

The FT-Transformer model, with fewer intrinsic assumptions, offers a more nuanced understanding of ecological data, revealing patterns not readily apparent with traditional methods. However, it faces its own set of challenges:

- *Interpretability:* Despite improvements in DNN model transparency, fully interpreting the relationships between input factors and target values is still challenging.
- *Data Quality and Volume:* The model's performance significantly depends on the quality and quantity of upstream training data, as elaborated in Section IV-B.
- *Resource Intensiveness:* The FT-Transformer demands substantial computational resources, which may contradict environmental sustainability goals due to increased CO2 emissions.

These considerations highlight the FT-Transformer's potential in ecological modeling and underscore areas for future refinement.

V. Conclusion

This study marks a significant advancement in estimating terrestrial GPP by effectively combining remote sensing data with deep learning. Our approach, leveraging the FT-Transformer model and transfer learning, demonstrates superior performance over traditional GBDT methods and the widely used MOD17 product in predicting annual GPP. The model's strength lies in its ability to process complex, multi-dimensional data, capturing the nuanced interplay between various ecological and climatic factors influencing GPP. While our results underscore the potential of deep learning in environmental modeling, we also acknowledge the challenges in interpretability and the reliance on extensive training data. Future work will focus on enhancing data quality, especially for underrepresented landcover types, and exploring alternative deep learning architectures for improved efficiency and accuracy. Our findings contribute to a more nuanced understanding of the terrestrial carbon cycle, offering valuable insights for climate change research and policy-making.

References

[1] C. Gough, "Terrestrial primary production: Fuel for," *Nature Education Knowledge*, vol. 3, no. 10, p. 28, 2011.

[2] K. Qiu, "Estimating regional vegetation gross primary productivity (gpp), evapotranspiration (et), water use efficiency (wue) and their spatial and temporal distribution across china," *Beijing Forestry University*, 2015.

[3] W. Cramer, A. Bondeau, F. I. Woodward, I. C. Prentice, R. A. Betts, V. Brovkin, P. M. Cox, V. Fisher, J. A. Foley, A. D. Friend *et al.*, "Global response of terrestrial ecosystem structure and function to co2 and climate change: results from six dynamic global vegetation models," *Global change biology*, vol. 7, no. 4, pp. 357–373, 2001.

[4] H. Wang, W. Shao, Y. Hu, W. Cao, and Y. Zhang, "Assessment of six machine learning methods for predicting gross primary productivity in grassland," *Remote Sensing*, vol. 15, no. 14, p. 3475, 2023.

[5] D. P. Sarkar, B. U. Shankar, and B. R. Parida, "Machine learning approach to predict terrestrial gross primary productivity using topographical and remote sensing data," *Ecological Informatics*, vol. 70, p. 101697, 2022.

[6] B. Lee, N. Kim, E.-S. Kim, K. Jang, M. Kang, J.-H. Lim, J. Cho, and Y. Lee, "An artificial intelligence approach to predict gross primary productivity in the forests of south korea using satellite remote sensing data," *Forests*, vol. 11, no. 9, p. 1000, 2020.

[7] W. Wu, C. Gong, X. Li, H. Guo, and L. Zhang, "An online deep convolutional model of gross primary productivity and net ecosystem exchange estimation for global forests," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 12, no. 12, pp. 5178–5188, 2019.

[8] V. Borisov, K. Seßler, T. Leemann, M. Pawelczyk, and G. Kasneci, "Language models are realistic tabular data generators," *arXiv preprint arXiv:2210.06280*, 2022.

[9] N. Hollmann, S. Müller, K. Eggensperger, and F. Hutter, "Tabpfn: A transformer that solves small tabular classification problems in a second," *arXiv preprint arXiv:2207.01848*, 2022.

[10] X. Huang, A. Khetan, M. Cvitkovic, and Z. Karnin, "Tabtransformer: Tabular data modeling using contextual embeddings. arxiv 2020," *arXiv preprint arXiv:2012.06678*, 2012.

[11] J. Kossen, N. Band, C. Lyle, A. N. Gomez, T. Rainforth, and Y. Gal, "Self-attention between datapoints: Going beyond individual input-output pairs in deep learning," *Advances in Neural Information Processing Systems*, vol. 34, pp. 28 742–28 756, 2021.

[12] G. Somepalli, M. Goldblum, A. Schwarzschild, C. B. Bruss, and T. Goldstein, "Saint: Improved neural networks for tabular data via row attention and contrastive pre-training," *arXiv preprint arXiv:2106.01342*, 2021.

[13] Y. Gorishniy, I. Rubachev, V. Khrulkov, and A. Babenko, "Revisiting deep learning models for tabular data," *Advances in Neural Information Processing Systems*, vol. 34, pp. 18 932–18 943, 2021.

[14] R. Levin, V. Cherepanova, A. Schwarzschild, A. Bansal, C. B. Bruss, T. Goldstein, A. G. Wilson, and M. Goldblum, "Transfer learning with deep tabular models," *arXiv preprint arXiv:2206.15306*, 2022.