Differentially private federated learning for high-accuracy carbon footprint prediction

Vijay Narasimhan
Tingting Ou
Hanna Jarlaczynska

EMD Electronics

Using Machine Learning in Carbon Footprint Prediction

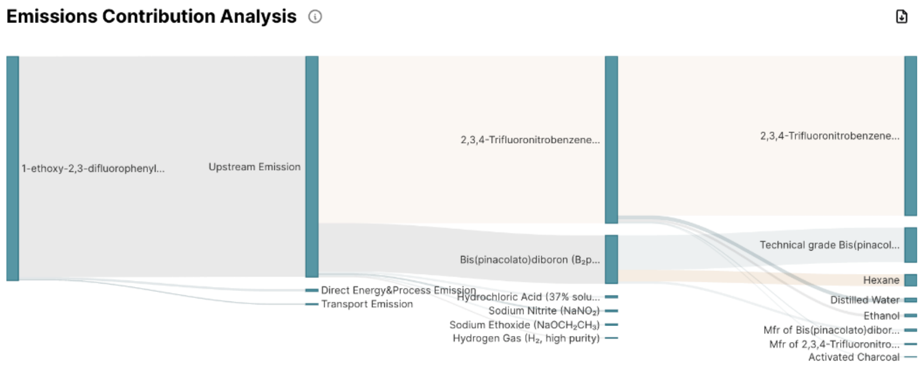
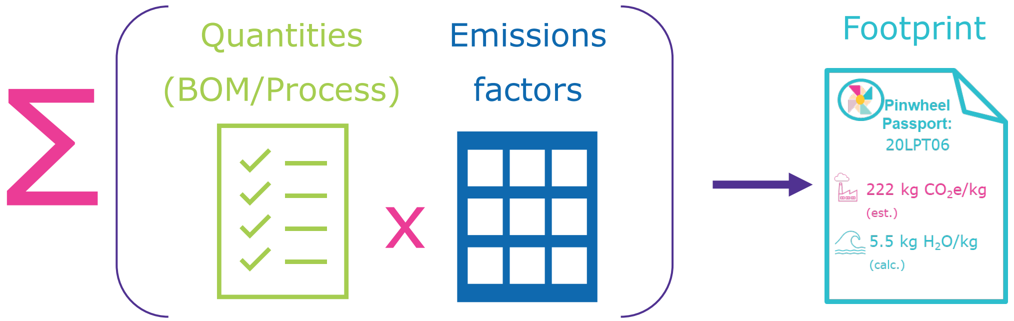
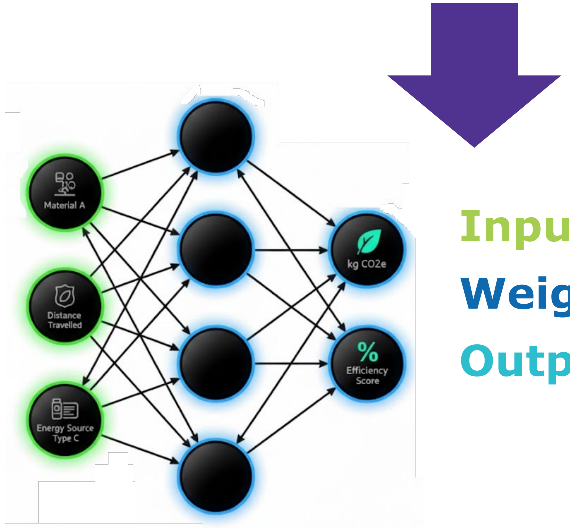


image generated from Muir AI PCF analysis

Recasting **footprint calculation as a deep learning problem** enables accurate, scalable, and privacy-preserving predictions.





Inputs: Quantities

Weights: Factors

Outputs: Footprints

Environmental footprint reporting and sustainable design requires **tracing inputs across complex supply chains**.

Currently, a lack of data hampers accurate calculations:

- Most emissions **databases cover only a small fraction** of known materials.
- **Proprietary processes** and inputs restrict data availability.

Differentially Private Federated Learning (DPFL) Framework

Federated learning

Aggregate the <u>models</u> , <u>not</u> <u>the data</u> from different clients

Differential privacy

<u>Add a little noise</u> to the model updates to hide the true updates in the FL framework

Algorithm 1 Differentially Private Federated Learning (DPFL)

Input: Datasets $\{D_i\}_{i\in[m]}$ belonging to m clients, initial baseline model θ_b, target privacy budget (ϵ, δ), number of local epochs E, learning rate η, the number of round T
Output: Differentially private global model $\bar{\theta}$
for each round $t = 1, 2, \ldots, T$ **do**
  for each client $i \in [m]$ in parallel **do**
    Receive global model θ_b from the server
    Compute noisy local gradient updates:
    $\tilde{g}_i = \mathsf{LocalTrain}(\theta_b, D_i, E, \eta, \frac{\epsilon}{Tm}, \frac{\delta}{Tm})$
    Send $\tilde{g}_i$ to the server
  end for
  Aggregate noisy gradients at the server:
  $\Delta = \frac{1}{m}\sum_{i\in\mathcal{S}} \tilde{g}_i$
  Update global model: $\theta_b \leftarrow \theta_b + \Delta$
end for
Return the final global model $\bar{\theta} = \theta_b$

Experimental Results

- **Data source:** Real-world processes from **TianGong** (10k+ LCA unit processes).
- **Emission factors:** Matched with **OpenLCA Nexus** database (kg CO_2 eq/unit).
- **Filtering:** Kept **508 processes** with full material–factor alignment (**97 materials** total).
- **Label computation:** Total footprint = Σ(input × emission factors)
- **Split: 80% training (3 clients)**, **20% testing**.

ϵ	$R^2(1)$	$R^2(2)$	$R^2(3)$	$R^2(agg)$	$R^2(baseline)$
1.5	0.9775	0.8150	0.4637	0.7709	0.9039
3	0.6969	0.7240	0.8734	0.8976	0.9852
15	0.9952	0.9464	0.9917	0.9608	0.9947
30	0.9990	0.9853	0.9910	0.9789	0.9954

At ε = 15, DPFL predicts footprints within 5% of the non-private baseline, striking a strong **balance between privacy and accuracy**.

Together, we can enable **collaborative sustainability.**

For more information, contact:

Vijay K. Narasimhan, PhD
Director, R&D Collaborations
Vijay.Narasimhan@emdgroup.com