Machine Learning-Based Surrogate Models for Performance Prediction of Triboelectric Nanogenerators

Shashank Mishra [1] **Subramanian Ramamoorthy** [2] **Jacqueline M. Cole** [3] **Themis Prodromakis** [1]

[1]*Centre for Electronic Frontiers, School of Engineering, University of Edinburgh, Edinburgh EH9 3BF, UK* [2]*Institute of Perception, Action and Behaviour, School of Informatics, University of Edinburgh, Edinburgh EH8 9AB, UK* [3]*Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge CB3 0HE, UK*. Correspondence to: Shashank Mishra smishra2@ed.ac.uk.

1. Introduction

Triboelectric nanogenerators (TENGs) are widely employed for energy harvesting from sources such as rain, wind, ocean waves, and human body motion[1, 2]. They convert mechanical or vibrational energy into electrical energy through coupled contact electrification and electrostatic induction processes, offering a promising pathway toward self-powered sensing electronics[3, 4]. A typical contact-separation TENG structure (as shown in Fig. 1) consists of a pair of active layers (generally fabricated from polymer nanocomposites that either generate or trap electrical charge) and conductive electrodes. When subjected to external mechanical stimuli, the electrical output of these devices exhibits pronounced sensitivity as a result of interactions between dissimilar materials and their surface characteristics. This sensitivity is governed by a range of factors, including the selection of materials for the active layers and electrodes, the micro- and nanostructures of the active layer surfaces, the effective contact area and separation distance between interacting layers, the frequency and amplitude of the applied mechanical excitation, environmental conditions such as humidity, and the load impedance[4, 5]. These nonlinear, multiphysics interactions render first-principles modeling and brute-force experimental approaches costly and time-consuming, thereby hindering rapid design iteration and application-specific optimization. Recently, machine learning (ML)–based surrogate models have been developed to describe the behavior of TENG devices[6, 7, 8]. However, these models typically consider only a limited set of parameters, often assuming nearly constant material compositions, which restricts their applicability.

To address these challenges, we propose a machine learning–based framework that predicts TENG performance by linking atomic-scale material properties to device-level behavior. The framework is built on a curated literature dataset comprising composition of active layer and electrode materials, few device features, and operating conditions. It also incorporates atomic-level properties from databases such as Materials Project [9] and Polymer Genome [10]. By integrating atomic-scale descriptors with device and operation-level variables, the model learns cross-scale relationships governing key performance metrics, such as open-circuit voltage,and hence, enabling rapid design screening and accelerated development of next generation TENG devices.

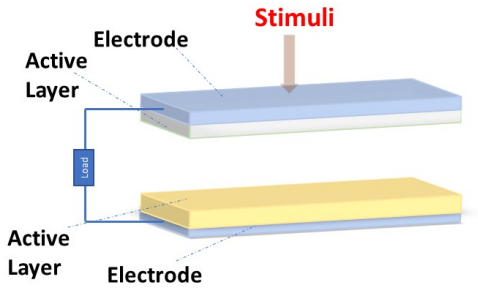

Fig. 1: Structure and components of a TENG device

2. Methodology

Our framework consisted of a typical ML workflow (as depicted in Fig. 2) where we first curated a dataset from the literature, followed by featurization of data and finally ML models were developed and validated over the featurized dataset.

2.1 Data extraction and curation

We manually extracted data from the literature related to TENG devices and their experimental characterisation. Data related to composition of active layers and electrodes, surface area, thickness, surface roughness and distance between active layers, the amplitude and frequency of external stimulus, external humidity and open-circuit voltage were extracted from text and plots of these research articles. A total of 403 data points were extracted from 30 research articles, with open-circuit voltages ranging from 0 to 1700 V.

2.2 Data processing and feature engineering

The dataset was pre-processed to get machine readable features for all the factors. Descriptors for the active materials, reflecting the chemical compo-

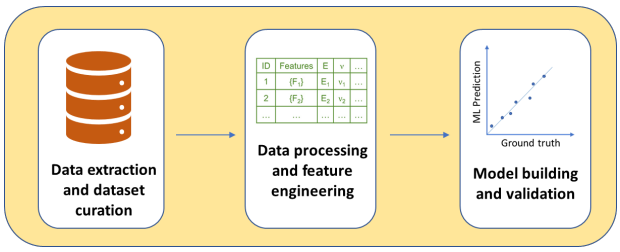

Fig. 2: Workflow of the ML-based prediction framework

sition of the layers, were decomposed into matrix and filler identities along with their loading percentages. The matrix and filler identities were further featurised to include their band gap/HOMO–LUMO gap values. These values were obtained from the *Materials Project*[9] for inorganic materials and from the *Polymer Genome*[10] tool for polymers. Similarly, we obtained resistivity values for corresonding electrode materials from the resistivity dataset[11]. We selected these particular electronic descriptors for the active layer and electrode components because prior studies have shown that they significantly influence triboelectric nanogenerator (TENG) performance[5]. Finally, the featurized dataset comprised of 20 key features categorizing all material, device, operational and performance parameters.

2.3 Model building and validation

Following dataset featurization, ML–based surrogate models were constructed to learn a mapping from 19 material, device, and operation related input features to the output performance metric, namely the open-circuit voltage. Three explainable gradient boosting algorithms—XGBoost, LightGBM, and CatBoost were employed for model development using identical hyperparameter settings. Model performance was evaluated using the coefficient of determination (R^2), root mean squared error (*RMSE*), and mean absolute error (*MAE*). To assess generalization capability and mitigate overfitting, a rigorous 5-fold cross-validation strategy was adopted, augmented with a bootstrap ensemble approach to enable uncertainty quantification and robust performance estimation across 25 models per algorithm.

3. Results

Table 1 presents a comparison of the predictive performance of all the evaluated models. Overall, the models demonstrate comparable accuracy, with R^2, *MAE*, and *RMSE* values ranging from 0.85–0.86, 63.10–65.93, and 123.52–128.83, respectively. Given that the dataset was curated from diverse literature sources with inherent variability in materials, device configurations, and testing conditions, this level of agreement indicates that the selected features effectively capture the dominant factors governing the open-circuit voltage.

Further, the CatBoost model shows stable and consistent behavior, and its parity plot (Fig.3) demonstrates strong agreement between predicted and reported values across the full voltage range. Importantly, the parity plot is augmented with uncertainty estimates obtained using a bootstrap ensemble approach, where the color distribution of data points reflects prediction confidence. Lower uncertainty is observed for densely sampled regions of the data space (output voltage < 700 V), while higher uncertainty appears in sparsely populated regimes (output voltage > 700 V), providing valuable insight into model reliability and highlighting areas where additional data would be most beneficial. This behavior suggests that the model is able to generalize well across a broad range of operating conditions and voltage magnitudes.

Table 1: Model comparison on various metrices.

Algorithm	R^2	*MAE (V)*	*RMSE (V)*
XgBoost	0.85 ± 0.04	65.21 ± 9.38	128.83 ± 18.57
LightGBM	0.86 ± 0.02	65.93 ± 7.97	126.08 ± 14.58
CatBoost	0.86 ± 0.03	63.10 ± 7.23	123.52 ± 22.54

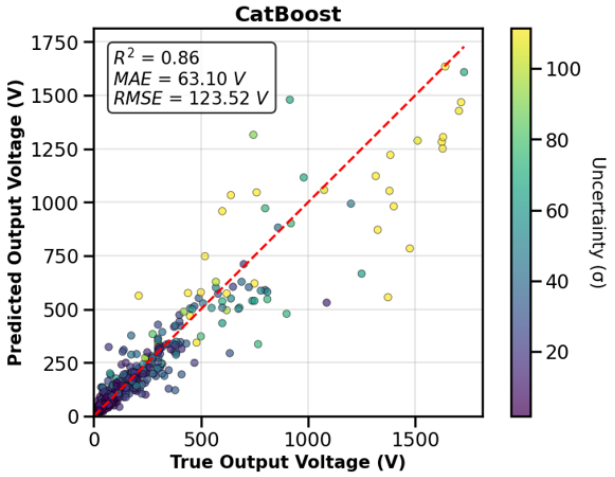

Fig. 3: Parity plot showing predicted versus experimental open-circuit voltage for the CatBoost model. Point colors indicate prediction uncertainty estimated using a bootstrap ensemble, with darker colors representing higher confidence and lighter colors indicating greater uncertainty.

4. Conclusion

In this work, we present a ML–based framework that predicts output performance of TENG by linking atomic-scale material descriptors with device-level behavior. By integrating curated experimental data with features from open-source materials databases and predictors, the multiscale approach captures the key factors governing output of TENG devices. The models achieved consistent R^2 values around 0.85, demonstrating robustness across variable experimental conditions. Future work will expand the dataset and explore additional material descriptors to further improve accuracy and applicability for data-driven TENG design.

Acknowledgments

We acknowledge support of the UKRI AI Hub programme for the APRIL AI Hub [Grant number EP/Y029763/1].

References

[1] Dongwhi Choi, Younghoon Lee, Zong-Hong Lin, Sumin Cho, Miso Kim, Chi Kit Ao, Siowling Soh, Changwan Sohn, Chang Kyu Jeong, Jeongwan Lee, et al. Recent advances in triboelectric nanogenerators: from technological progress to commercial applications. *ACS nano*, 17(12):11087–11219, 2023.

[2] Guanbo Min, Gaurav Khandelwal, Abhishek Singh Dahiya, Shashank Mishra, Wei Tang, and Ravinder Dahiya. Multisource energy harvester on textile and plants for clean energy generation from wind and rainwater droplets. *ACS Sustainable Chemistry & Engineering*, 12(2):695–705, 2024.

[3] Shashank Mishra, Nitheesh M Nair, Gaurav Khandelwal, Beena Rai, and Vihar Georgiev. Capacitive-triboelectric-based hybrid sensor system for human-like tactile perception. *IEEE Sensors Letters*, 8(2):1–4, 2024.

[4] Tinghai Cheng, Jiajia Shao, and Zhong Lin Wang. Triboelectric nanogenerators. *Nature Reviews Methods Primers*, 3(1):39, 2023.

[5] Chen Cao, Zhongjie Li, Fan Shen, Qin Zhang, Ying Gong, Hengyu Guo, Yan Peng, and Zhong Lin Wang. Progress in techniques for improving the output performance of triboelectric nanogenerators. *Energy & Environmental Science*, 17(3):885–924, 2024.

[6] Mohammad Abrar Uddin, Myeongju Lim, Rubiga Kim, Barrett London Burgess, Ken Roberts, Junghyun Kim, and Taeil Kim. Machine learning-driven surrogate modeling for optimization of triboelectric nanogenerator design parameters. *Advanced Electronic Materials*, page 2400771, 2025.

[7] Chi Han, Mingyu Jin, Fuying Dong, Pengchong Xu, Xinnian Jiang, Sheling T Cai, Yuanwen Jiang, Yongfeng Zhang, Yin Fang, and Simiao Niu. Interpretable machine learning for evaluating nanogenerators' structural design. *ACS nano*, 19(14):14456–14466, 2025.

[8] Abdulkerim Okbaz, Adem Yar, Geng-Sheng Lin, and Zhaohui Tong. Numerical and machine learning-based triboelectric nanogenerator simulators: Contact-separation mode. *Energy Technology*, page 2402101, 2025.

[9] Anubhav Jain, Shyue Ping Ong, Geoffroy Hautier, Wei Chen, William Davidson Richards, Stephen Dacek, Shreyas Cholia, Dan Gunter, David Skinner, Gerbrand Ceder, et al. Commentary: The materials project: A materials genome approach to accelerating materials innovation. *APL materials*, 1(1), 2013.

[10] Chiho Kim, Anand Chandrasekaran, Tran Doan Huan, Deya Das, and Rampi Ramprasad. Polymer genome: a data-powered polymer informatics platform for property predictions. *The Journal of Physical Chemistry C*, 122(31):17575–17585, 2018.

[11] Resistivity of the elements. `https://periodictable.com/Properties/A/Resistivity.v.log.html`. Accessed: 2025-09-30.