Machine Learning-Based Prediction of Building Energy Consumption in Residential Structures

Introduction & Relevance

Buildings account for a significant portion of global energy usage and carbon emissions. Poor energy efficiency in the design phase often results in long-term environmental and economic consequences. Traditional energy modeling methods are input-heavy and unsuitable for early-stage design evaluation. This project aims to leverage machine learning (ML) and deep learning (DL) techniques to predict annual building energy consumption using minimal design inputs and local climate data. The goal is to empower architects and engineers with accurate tools to evaluate and optimize energy efficiency at the conceptual stage of residential building design.

3. Research Question & Objectives Main Question:

- How accurately can we predict the annual energy consumption of a residential building based on design-stage metadata and meteorological conditions?

Sub-questions:

1. Which features (e.g., floor area, room count, wall/roof/window types, local climate) contribute most to energy use prediction?
2. Do deep learning approaches (e.g., DNNs) outperform traditional ML algorithms (e.g., RF, SVM, LR) in prediction accuracy?
3. How does dataset size affect the performance of different ML models?

Objectives:

- Collect and preprocess metadata from 5000 residential buildings and corresponding meteorological data.
- Engineer relevant features for model training.
- Train and compare various supervised learning models (ANN, DNN, SVM, RF, GB, etc.).
- Evaluate models using MAE, RMSE, MSE, R^2.
- Analyze feature importance and conduct sensitivity tests by building type and data size.

4. **State of the Art**

- Forecasting energy consumption using ML has gained momentum over the past decade. Key insights include:

- ANN models consistently yield high accuracy, especially with large datasets (Runge & Zmeureanu, 2019).

- SVM models are suitable for small datasets and perform competitively in hourly load prediction (Dong et al., 2005).
- Ensemble models like Gradient Boosting and Random Forest offer robust performance and interpretability.
- DNNs, with their deep architecture, have shown superior accuracy in recent energy prediction studies but remain underutilized in residential design-phase forecasting.
- Comparative studies often lack a unified dataset, limiting fair model evaluations.

Gap: There is limited research applying and benchmarking multiple ML algorithms on the same, large residential dataset during the design phase.

5. Methodology

Phase	**Methods & Tools**	**Output**
Data Collection	UK building metadata, Meteostat API	5000 residential instances
Preprocessing	Python (pandas, sklearn)	Cleaned, normalized, merged dataset
Feature Selection	Random Forest, ExtraTreesClassifier	Top 10 predictive features
Model Training & Evaluation	scikit-learn, Keras, TensorFlow	Metrics: RMSE, MAE, R^2, MSE
Sensitivity Analysis	Subsets by building type and data size	Performance variation insights

5. Evaluation Criteria

To assess the accuracy and reliability of the machine learning models used for predicting annual building energy consumption, the following evaluation metrics are applied:

- **RMSE (Root Mean Squared Error):** Measures prediction accuracy with emphasis on large errors.
- **MAE (Mean Absolute Error):** Average magnitude of prediction errors.
- **MSE (Mean Squared Error):** Penalizes large errors, useful for model comparison.
- **R^2 (Coefficient of Determination):** Proportion of variance explained by the model (closer to 1 = better).
- **Training Time:** To assess computational efficiency.

7. Work Plan & Timeline

Weeks	Task
1	Literature review and framework design
2	Dataset acquisition (building + weather)
3	Data cleaning and integration
4	Feature selection and exploratory analysis

5-6	Model development: traditional ML (RF, LR, SVM, DT, etc.)
7-8	Deep learning model development (ANN, DNN)
9	Model evaluation and comparison
10	Sensitivity analysis (clusters and data size)
11	Final visualization and result consolidation
12	Report preparation and submission

8. Expected Results & Outlook

- **Quantitative:** Identification of the most accurate model (DNN expected to perform best with RMSE ~1.16 and R^2 ~0.95).
- **Qualitative:** Insights into the most influential features for design-stage prediction.
- **Deliverables:**
    1. Python notebooks and scripts for full pipeline.
    2. Comparative performance report with visualizations.
    3. Final exposé and project presentation.

References

[1] A.-D. Pham, et al., Predicting energy consumption in multiple buildings using machine learning for improving energy efficiency and sustainability, J. Clean. Prod. 260 (2020) 121082, https://doi.org/10.1016/j.jclepro.2020.121082.

[2] B. Dandotiya, Climate-Change-and-Its-Impact-on-Terrestrial-Ecosystems, 2020, https://doi.org/10.4018/978-1-7998-3343-7.ch007.

[3] P. Aversa, et al., Improved thermal transmittance measurement with HFM technique on building envelopes in the mediterranean area, Sel. Sci. Pap. J. Civ. Eng. 11 (2016), https://doi.org/10.1515/sspjce-2016-0017.

[4] Soheil Fathi, et al., Machine learning applications in urban building energy performance forecasting: a systematic review, Renew. Sustain. Energy Rev. 133 (2020) 110287, https://doi.org/10.1016/j.rser.2020.110287.

[5] G. Serale, M. Fiorentini, M. Noussan, 11 - development of algorithms for building energy efficiency, in: F. Pacheco-Torgal, et al. (Eds.), Start-Up Creation, second ed., Woodhead Publishing (Woodhead Publishing Series in Civil and Structural Engineering), 2020, pp. 267–290, https://doi.org/10.1016/B978-0-12-819946-6.00011-4.

[6] L. Li, et al., Impact of natural and social environmental factors on building energy consumption: based on bibliometrics, J. Build. Eng. 37 (2021) 102136, https://doi.org/10.1016/j.jobe.2020.102136.

[7] M. Hamed, S. Nada, Statistical analysis for economics of the energy development in North Zone of Cairo, Int. J. Finance Econ. 5 (2019) 140–160