**1. Literature Summary (Approaches)**

* Two major approaches for energy prediction:
  + **Physical (white-box) models:** Rely on physics and domain knowledge; highly accurate but time-consuming.
  + **Data-driven models:** Use machine learning on real operational data; more scalable and practical.
* **Challenge identified:**  
  Most previous studies focused on **supervised ML** but **neglected feature engineering**, which is crucial for:
  + Reducing dimensionality
  + Improving model generalization
  + Handling noisy or missing data
* **Goal of this study:**  
  Develop **automated feature extraction** using **deep learning** to improve prediction performance.

**🤖 2. Algorithms Used**

**🔧 Feature Engineering Methods:**

* **Traditional (Baseline) Methods:**
  + Principal Component Analysis (PCA)
  + Statistical Summarization (e.g., means over time windows)
* **Deep Learning-based Methods:**
  + **Autoencoder (AE):** Learns compressed representations via unsupervised learning.
  + **Convolutional Autoencoder (CAE):** Captures temporal patterns from time series data.
  + **Generative Adversarial Networks (GAN):** Extracts high-level features by training a generator and discriminator.

**📈 Supervised Learning Models Tested:**

Used to predict cooling load based on the extracted features:

* **Multiple Linear Regression (MLR)** (baseline)
* **Artificial Neural Network (ANN)**
* **Support Vector Regression (SVR)**
* **Extreme Gradient Boosting (XGB)**

**🧪 3. Key Findings**

| **Feature Method** | **Best Model** | **CV-RMSE (↓)** | **Key Result** |
| --- | --- | --- | --- |
| **PCA** | XGB | 19.3% | Linear features, poor with MLR |
| **Statistical** | XGB | 19.3% | Poor temporal representation |
| **AE (Autoencoder)** | XGB | 19.0% | Good nonlinear compression |
| **CAE** | XGB | 18.3% | Better for time-series data |
| **GAN** | **XGB** | **17.7%** | **Best performance across all metrics** |

**Key Insight:**

* **GAN-based features consistently yielded the highest accuracy**, especially for advanced models like XGB and ANN.
* Linear models like MLR only performed well when paired with GAN features.

**💻 4. Software & Tools**

* **Language:** R
* **Deep Learning Library:** Keras (via R interface)
* **Machine Learning:** Implemented using custom scripts in R and Keras
* **Data:** One-year operational data (17,040 samples, 30-min intervals) from a university building in Hong Kong.

**💡 5. Practical Implications**

* **Deep learning-based feature engineering** is:
  + Robust to noise and missing data
  + Effective for automatic feature generation
  + Scalable for large or complex buildings
* GANs not only extract features but can **generate synthetic energy profiles** for simulation and benchmarking.

**✅ Conclusion**

This paper demonstrates that **deep learning models—especially GANs—are superior** in creating high-quality features for building energy prediction tasks. These approaches **automate the prediction pipeline** and outperform traditional methods both in accuracy and resilience to poor-quality data.

**عنوان مقاله:**

**روش‌های مهندسی ویژگی مبتنی بر یادگیری عمیق برای بهبود پیش‌بینی مصرف انرژی ساختمان**

**🔍 ۱. خلاصه ادبیات (رویکردها)**

* دو رویکرد اصلی در پیش‌بینی مصرف انرژی ساختمان:
  1. **مدل‌های فیزیکی (White-box):** بر اساس قوانین فیزیکی و دانش مهندسی؛ دقیق ولی بسیار وقت‌گیر و پیچیده.
  2. **مدل‌های داده‌محور (Data-driven):** بر اساس داده‌های واقعی بهره‌برداری؛ سریع‌تر و مقیاس‌پذیرتر.
* **مشکل شناسایی‌شده:**  
  در اکثر پژوهش‌های قبلی تمرکز روی **مدل‌های یادگیری نظارت‌شده** بوده و به **مهندسی ویژگی (Feature Engineering)** توجه کافی نشده است، در حالی که مهندسی ویژگی برای:
  1. کاهش ابعاد داده
  2. ساده‌تر شدن مدل
  3. مقابله با نویز و داده ناقص  
     بسیار حیاتی است.
* **هدف این مطالعه:**  
  استفاده از **یادگیری عمیق بدون نظارت** برای **استخراج خودکار ویژگی‌های با کیفیت** جهت بهبود پیش‌بینی مصرف انرژی.

**🤖 ۲. الگوریتم‌ها و روش‌ها**

**🔧 روش‌های مهندسی ویژگی:**

* **روش‌های سنتی (پایه):**
  + **تحلیل مؤلفه‌های اصلی (PCA)**
  + **آمار خلاصه‌سازی ساده** (میانگین در بازه‌های زمانی)
* **روش‌های مبتنی بر یادگیری عمیق:**
  + **خودرمزگذار (Autoencoder - AE):** فشرده‌سازی داده‌ها با حفظ اطلاعات مهم.
  + **خودرمزگذار کانولوشنی (CAE):** درک وابستگی زمانی در داده‌های سری زمانی.
  + **شبکه‌های مولد خصمانه (GAN):** استخراج ویژگی‌های سطح بالا با تولید داده مصنوعی و رقابت بین مولد و تفکیک‌گر.

**📈 الگوریتم‌های پیش‌بینی (یادگیری نظارت‌شده):**

* **رگرسیون خطی چندگانه (MLR)** – مدل پایه
* **شبکه عصبی مصنوعی (ANN)**
* **رگرسیون بردار پشتیبان (SVR)**
* **XGBoost (مدل ترکیبی قدرتمند مبتنی بر درخت تصمیم)**

**🧪 ۳. نتایج کلیدی و یافته‌ها**

| **نوع ویژگی** | **بهترین مدل** | **خطای CV-RMSE** | **نتیجه** |
| --- | --- | --- | --- |
| **PCA** | XGB | ۱۹٫۳٪ | عملکرد ضعیف در مدل‌های خطی |
| **آماری (میانگین)** | XGB | ۱۹٫۳٪ | ضعیف در نمایش زمان‌بندی |
| **AE (خودرمزگذار)** | XGB | ۱۹٫۰٪ | بهتر از روش‌های خطی |
| **CAE** | XGB | ۱۸٫۳٪ | قوی در تشخیص الگوهای زمانی |
| **GAN** | **XGB** | **۱۷٫۷٪** | **بهترین عملکرد کلی** ✅ |

* ویژگی‌های استخراج‌شده با GAN دقیق‌ترین پیش‌بینی‌ها را در همه مدل‌ها داشتند.
* مدل خطی (MLR) فقط در صورت استفاده از ویژگی‌های GAN عملکرد قابل‌قبول داشت.

**💻 ۴. نرم‌افزارها و ابزارهای استفاده‌شده**

* زبان برنامه‌نویسی: **R**
* کتابخانه یادگیری عمیق: **Keras (نسخه R)**
* داده‌ها: داده‌های عملیاتی سالانه از یک ساختمان دانشگاهی در هنگ‌کنگ (17040 نمونه، فاصله ۳۰ دقیقه‌ای)

**💡 ۵. کاربردهای عملی**

* روش‌های یادگیری عمیق در مهندسی ویژگی:
  + به خوبی **نویز و داده‌های ناقص** را مدیریت می‌کنند.
  + **ویژگی‌های سطح بالا** را بدون دخالت انسانی استخراج می‌کنند.
  + می‌توانند **پروفایل‌های مصنوعی مصرف انرژی** تولید کنند (برای شبیه‌سازی و بنچمارک).

**✅ ۶. جمع‌بندی**

* روش‌های مبتنی بر یادگیری عمیق به‌ویژه **GANها** در استخراج ویژگی‌های مفید از داده‌های انرژی ساختمان بسیار مؤثر هستند.
* این روش‌ها فرآیند مدل‌سازی را **خودکار و مقیاس‌پذیر** کرده و دقت پیش‌بینی را **افزایش چشم‌گیر** می‌دهند.
* در صورت استفاده از داده‌های با کیفیت پایین یا ناقص نیز **دقیق و مقاوم** عمل می‌کنند.
* این مقاله پایه‌ای برای توسعه ابزارهای هوشمندتر در مدیریت انرژی ساختمان‌ها ارائه می‌دهد.