Based on the paper **"Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques"**, here is a **summary of the literature** including the **approaches**, **best algorithms**, **findings**, and **software used**:

**🔍 1. Approaches Reviewed in the Literature**

* **Two categories of models**:
  + **Forward (physics-based)** models: Require detailed building parameters (e.g., HVAC, insulation), often simulated using tools like **EnergyPlus**, **TRNSYS**, **DOE-2**.
  + **Data-driven (ML-based)** models: Use historical energy and environmental data. Less demanding in terms of required input details.
* **Data-driven approaches** favored due to:
  + Flexibility
  + Lower input requirements
  + Better scalability
  + Faster execution time
* **Focus of this study**: Forecasting **annual energy consumption at the early design phase**, which is **underexplored** in existing studies.

**🤖 2. Machine Learning Algorithms Compared**

Nine ML models were analyzed **on the same dataset** (5000 residential buildings), making the results directly comparable:

| **Model** | **R²** | **RMSE** | **Training Time** | **Performance Summary** |
| --- | --- | --- | --- | --- |
| **DNN (Deep Neural Network)** | **0.95** | **1.16** | 5.2s | ✅ **Best overall performance** |
| ANN (Artificial Neural Network) | 0.94 | 1.20 | 3.7s | Strong results |
| GB (Gradient Boosting) | 0.92 | 1.40 | 1.8s | Fast, promising |
| SVM (Support Vector Machine) | 0.90 | 1.61 | 2.0s | Good for small data |
| RF (Random Forest) | 0.89 | 1.69 | 2.5s | Good, less tuning |
| KNN | 0.77 | 2.40 | 1.4s | Lower performance |
| DT (Decision Tree) | 0.74 | 2.55 | 1.2s | Fastest, but weak |
| LR (Linear Regression) | 0.73 | 2.59 | 1.4s | Weak baseline |
| Stacking | 0.73 | 2.60 | 1.3s | Lowest performance |

**Best Model**: **Deep Neural Network (DNN)**  
**Worst Model**: **Stacking Ensemble & Linear Regression**  
**Most efficient (fastest)**: **Decision Tree**, but low accuracy

**🧪 Key Findings**

* **DNN** outperforms all models in all evaluation metrics (RMSE, MAE, R², etc.).
* **ANN** and **Gradient Boosting** are also strong contenders.
* **Model performance improves significantly with larger datasets**.
* **Building type** (e.g., house vs. apartment) **does not significantly affect** model accuracy or feature selection.
* Feature importance was consistent across building types and data sizes.

**💻 Software and Tools Used**

* **Programming language**: Python
* **IDE**: Sublime Text
* **Libraries**:
  + scikit-learn for model implementation, normalization, and feature selection
  + pandas for data handling
  + Keras/TensorFlow likely used for DNN (although not explicitly stated)
* **Hardware**: Apple MacBook Air (M1 chip, 16 GB RAM)

**📊 Data Summary**

* **Total buildings**: 5000 residential buildings (10 UK postcode areas)
* **Target**: Annual energy consumption (kWh/m²)
* **Input features**:
  + Building design parameters (walls, windows, floors, etc.)
  + Weather data (temperature, wind speed, pressure)

**📌 Conclusion**

This paper provides **strong empirical evidence** that **DNNs are the most effective method** for early-stage prediction of residential building energy consumption. It also highlights the **need for more comparative studies using large, consistent datasets**, and suggests future work on **larger datasets (>40,000 instances)** and **ensemble learning methods**.

### 🔍 ****۱. مرور ادبیات (رویکردها، الگوریتم‌ها و یافته‌ها)****

#### ✅ ****رویکردها:****

مطالعات پیشین مدل‌های پیش‌بینی انرژی ساختمان را به دو دسته اصلی تقسیم کرده‌اند:

1. **مدل‌های فیزیکی (Forward Models)**
   * نیاز به ورودی‌های دقیق و زیاد دارند (مانند سیستم‌های HVAC، ضخامت عایق، نوع دیوار و غیره).
   * ابزارهایی مانند **EnergyPlus**، **TRNSYS** و **DOE-2** استفاده می‌شوند.
   * زمان‌بر و اغلب پیچیده هستند.
2. **مدل‌های داده‌محور (Data-driven Models)**
   * مبتنی بر یادگیری ماشین و داده‌های واقعی (مثلاً داده‌های شمارش‌گرهای هوشمند)
   * نیاز کمتری به ورودی‌های پیچیده دارند.
   * سریع‌تر، مقیاس‌پذیرتر و ساده‌تر در اجرا هستند.
   * به‌ویژه در پیش‌بینی انرژی مصرفی در **مرحله طراحی اولیه** بسیار کاربردی هستند.

### 🤖 ****۲. الگوریتم‌های یادگیری ماشین مقایسه‌شده****

نویسندگان ۹ الگوریتم یادگیری ماشین را **روی یک دیتاست یکسان (۵۰۰۰ ساختمان)** آزمایش کرده‌اند تا مقایسه منصفانه‌ای صورت گیرد:

| **الگوریتم** | **R² (ضریب تعیین)** | **RMSE** | **زمان آموزش** | **خلاصه عملکرد** |
| --- | --- | --- | --- | --- |
| **DNN (شبکه عصبی عمیق)** | **0.95** | **1.16** | ۵.۲ ثانیه | ✅ **بهترین عملکرد کلی** |
| ANN (شبکه عصبی مصنوعی) | 0.94 | 1.20 | ۳.۷ ثانیه | بسیار خوب |
| GB (گرادیان بوستینگ) | 0.92 | 1.40 | ۱.۸ ثانیه | سریع و امیدوارکننده |
| SVM (ماشین بردار پشتیبان) | 0.90 | 1.61 | ۲.۰ ثانیه | مناسب داده کم |
| RF (جنگل تصادفی) | 0.89 | 1.69 | ۲.۵ ثانیه | بدون نیاز به تنظیمات زیاد |
| KNN (نزدیک‌ترین همسایه) | 0.77 | 2.40 | ۱.۴ ثانیه | عملکرد ضعیف‌تر |
| DT (درخت تصمیم‌گیری) | 0.74 | 2.55 | **۱.۲ ثانیه** | سریع‌ترین اما کم‌دقت |
| LR (رگرسیون خطی) | 0.73 | 2.59 | ۱.۴ ثانیه | عملکرد پایه ضعیف |
| Stacking (مدل ترکیبی) | 0.73 | 2.60 | ۱.۳ ثانیه | ضعیف‌ترین عملکرد |

**نتیجه نهایی**: مدل **شبکه عصبی عمیق (DNN)** بهترین دقت را ارائه داد.  
📉 مدل‌های **Stacking و رگرسیون خطی** کمترین دقت را داشتند.  
⚡ درخت تصمیم‌گیری سریع‌ترین مدل بود ولی دقت پایینی داشت.

**🧪 ۳. یافته‌های کلیدی پژوهش**

* **DNN بهترین نتایج را از نظر MAE، RMSE، R² و MSE داشت.**
* عملکرد مدل‌ها با **افزایش حجم داده‌ها بهتر می‌شود**.
* **نوع ساختمان (مثلاً خانه یا آپارتمان)** تأثیر زیادی روی نتایج یا انتخاب ویژگی‌ها ندارد.
* ویژگی‌های مهم انتخاب‌شده شامل:
  + مساحت طبقات، تعداد اتاق‌ها، توصیف دیوار و سقف، دمای هوا، سرعت باد و فشار هوا هستند.

**💻 ۴. نرم‌افزار و ابزارهای مورد استفاده**

* زبان برنامه‌نویسی: **Python**
* محیط توسعه: **Sublime Text**
* کتابخانه‌ها:
  + scikit-learn برای مدل‌سازی، انتخاب ویژگی و نرمال‌سازی
  + pandas برای پردازش داده
  + احتمالاً Keras یا TensorFlow برای پیاده‌سازی DNN
* سخت‌افزار: لپ‌تاپ **MacBook Air با چیپ M1 و ۱۶ گیگ رم**

**📊 ۵. ویژگی‌های دیتاست**

* ۵۰۰۰ ساختمان مسکونی از ۱۰ ناحیه پستی در انگلستان
* ویژگی‌ها شامل:
  + متغیرهای ساختمانی (نوع دیوار، پنجره، سقف، تعداد اتاق‌ها و ...)
  + متغیرهای هواشناسی (دما، سرعت باد، فشار)
* هدف مدل: پیش‌بینی مصرف انرژی سالیانه (kWh/m²)

**✅ ۶. نتیجه‌گیری**

* مدل **DNN** بهترین عملکرد را داشت و برای استفاده در **مرحله طراحی اولیه ساختمان‌ها** مناسب است.
* الگوریتم‌های ANN، GB و SVM نیز عملکرد قابل قبولی داشتند.
* **اندازه دیتاست و کیفیت داده‌ها نقش کلیدی در دقت مدل دارند**.
* مدل DNN می‌تواند ابزاری مؤثر برای معماران و طراحان ساختمان باشد تا در مراحل ابتدایی تصمیم‌گیری بهینه‌ای انجام دهند.