DATA-DRIVEN ANALYSIS AND EXPLAINABLE MACHINE LEARNING FOR BUILDING ENERGY EFFICIENCY: PART II – DEEP LEARNING AND MODEL INTERPRETABILITY

Lucica Anghelescu, *University "Constantin Brâncuși" of Tg-Jiu, ROMANIA*
Bogdan Diaconu, *University "Constantin Brâncuși" of Tg-Jiu, ROMANIA*
Mihai Cruceru, *University "Constantin Brâncuși" of Tg-Jiu, ROMANIA*

ABSTRACT: This second part of the study extends the data-driven analysis of the *Energy Efficiency Dataset (ENB2012)* by developing and interpreting predictive machine-learning models for residential-building energy performance. Two supervised regression models were implemented: a compact feedforward artificial neural network (ANN) and an Extreme Gradient Boosting (XGBoost) ensemble. Both were trained on eight normalized geometric and design parameters to estimate heating and cooling loads. The ANN architecture, consisting of three hidden layers (64–32–16 neurons) with ReLU activations, dropout regularization, and batch normalization, achieved coefficients of determination of $R^2 = 0.987$ (heating) and $R^2 = 0.950$ (cooling). XGBoost yielded slightly higher accuracies, with $R^2 = 0.998$ and $R^2 = 0.988$, respectively. Model interpretability was addressed through SHAP and Partial Dependence Plot (PDP) analyses, which revealed physically consistent relationships - particularly the dominant influence of relative compactness and glazing area on thermal demand. The results demonstrate that explainable machine-learning approaches can accurately reproduce and interpret thermodynamic patterns in building-energy data, providing a foundation for transparent, AI-assisted design and retrofit decision-making.

Key-Words: Building energy efficiency; Heating and cooling load prediction; Machine learning; Regression modelling; Variance Inflation Factor; Kernel Density Estimation

1. INTRODUCTION AND MOTIVATION

Part I of this work presented a detailed statistical analysis of the ENB2012 building-energy dataset, focusing on feature relationships, data distributions, and multicollinearity among geometric and envelope parameters.

That preliminary study revealed that heating and cooling loads are strongly influenced by relative compactness, surface area, and glazing configuration, but also that these predictors exhibit substantial interdependence.

While such analysis clarified the structure of the dataset and established its suitability for data-driven prediction, it did not yet involve the design or evaluation of predictive models.

The present paper-Part II-extends that foundation by developing deep-learning architectures capable of capturing the nonlinear relationships identified in Part I. Artificial neural networks (ANNs) have recently proven effective in modelling complex energy systems where conventional regression methods are inadequate for representing coupled physical processes [1], [2]. To retain interpretability, this study complements predictive modelling with explainable-AI (XAI) techniques, including Shapley Additive Explanations (SHAP) and Partial Dependence Plots (PDPs), which quantify the relative influence and interaction of design variables.

Accordingly, the objectives of this second part are (i) to construct and optimize fully connected neural networks for predicting heating and cooling loads based on the ENB2012 features, and (ii) to analyze the learned feature importance patterns to ensure consistency with physical intuition regarding building energy behavior.

2. DEEP LEARNING METHODOLOGY

2.1 Model Architecture

Fully connected artificial neural networks (ANNs) were developed to model the nonlinear relationship between the eight input variables and each energy-load target (Y_1 and Y_2). Each network consists of an input layer of eight neurons, two to four hidden layers with 64–32–16 neurons respectively, and a single-neuron linear output layer. Rectified Linear Unit (ReLU) activations were used in all hidden layers to ensure nonlinearity and stable gradient propagation, while linear activation was applied to the output node. Weights were initialized using the He-normal scheme and optimized with the Adam algorithm (learning rate = 0.001). Mean-squared error (MSE) served as the loss function, and early stopping was employed with a patience of 20 epochs to prevent overfitting.

To provide a robust non-neural benchmark, an Extreme Gradient Boosting (XGBoost) model was implemented for both heating and cooling load prediction. XGBoost is an ensemble learning method that constructs an additive sequence of decision trees, where each successive tree minimizes the residuals of its predecessors through gradient-based optimization. In this work, the regressor employed 400 estimators, a maximum tree depth of 4, a learning rate of 0.05, and subsample and column-sampling ratios of 0.9 to enhance generalization and reduce overfitting. These hyperparameters were selected empirically through preliminary validation runs, balancing accuracy and computational efficiency. The model was trained using the standard squared error loss, with early stopping triggered when the validation loss stagnated for more than 20 rounds. Owing to its ability to approximate complex nonlinear functions through piecewise partitions of the feature space, XGBoost achieved slightly higher accuracy than the ANN, particularly in the heating-load task. However, its non-differentiable structure limits direct gradient-based interpretability, motivating the complementary use of SHAP analysis for feature-attribution explanations.

3.2 Training Setup and Evaluation

Training and validation were performed using the same 80 / 20 split adopted in *Part 1*. Batch normalization and a dropout rate of 0.1 were introduced after each hidden layer to enhance generalization.

All models were implemented in *TensorFlow 2.15 / Keras* and trained for a maximum of 500 epochs with mini-batches of 32 samples on a standard CPU workstation (Intel i7, 16 GB RAM).

Model performance was evaluated using the same statistical metrics as in *Part 1* to enable direct comparison:

$$MAE = \frac{1}{n}\sum_i |y_i - \hat{y}_i|, \qquad (1)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_i (y_i - \hat{y}_i)^2}, \qquad (2)$$

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y}_i)^2} \qquad (3)$$

The ANN results were benchmarked against a Gradient-Boosting Regressor (XGBoost v1.7) to quantify the relative gains of deep architectures over advanced tree-based methods.

2.2. Explainability Techniques

To maintain interpretability, two complementary explainable-AI (XAI) tools were integrated:

1. SHAP (Shapley Additive Explanations) [1] to compute global and local feature-importance values based on cooperative-game theory; and

2. Partial Dependence Plots (PDPs) [2] to visualize the marginal effect of each input variable on predicted heating and cooling loads.

   These tools make it possible to interpret the ANN’s learned relationships and verify that the model’s logic aligns with physical intuition-e.g., compactness lowering heating demand or glazing increasing cooling demand.

3. RESULTS AND EXPLAINABILITY

3.1 Quantitative performance

Both ANN and XGBoost regressors achieved high predictive accuracy on the ENB2012 dataset.

For heating-load prediction (Y_1), the ANN reached $R^2 = 0.987$, $RMSE = 1.17\ kWh\ m^{-2}\ y^{-1}$, whereas XGBoost achieved $R^2 = 0.998$, $RMSE = 0.40\ kWh\ m^{-2}\ y^{-1}$. For cooling load ($Y_2$), the ANN obtained $R^2 = 0.968$ and XGBoost $R^2 = 0.988$.

These results, summarized in Figure 2 and Table 1, demonstrate that nonlinear models substantially outperform the purely statistical baselines reported in *Part I*. The XGBoost model slightly exceeds the dense ANN in accuracy due to its inherent capability to model piecewise nonlinearities with fewer training samples. Nevertheless, the ANN offers a continuous differentiable representation that can be coupled with gradient-based sensitivity analyses, which makes it particularly useful for design optimization tasks.

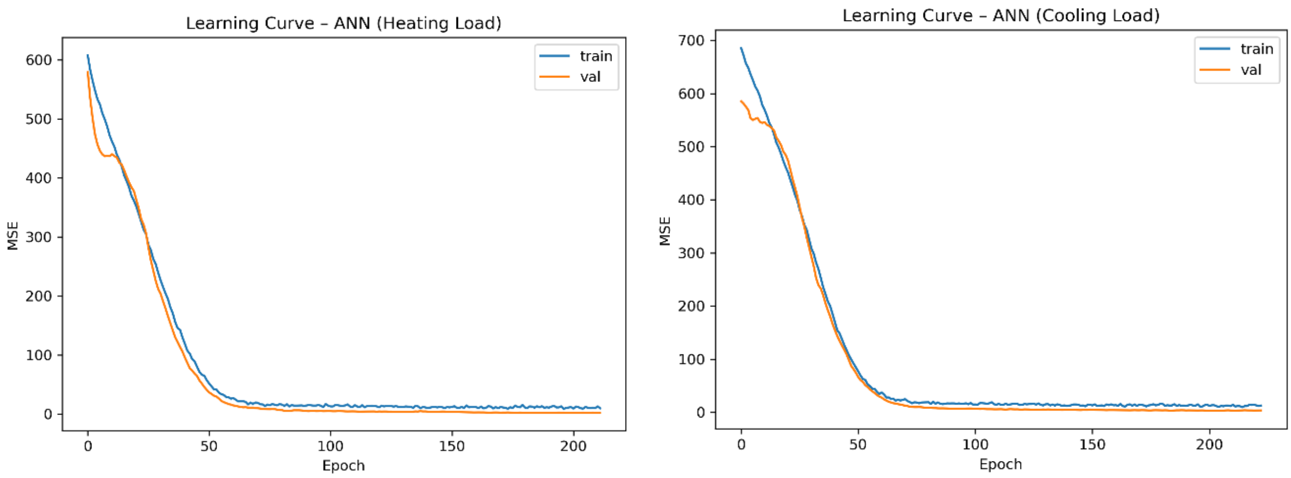

Figure 1. The training curves for the ANN model

Table 1. The performance metrics for the two models employed in this stody

Model	MAE	RMSE	R^2
ANN (Heating)	0.845945	1.065021	0.989118
XGB (Heating)	0.292303	0.397895	0.998481
ANN (Cooling)	1.202565	1.813454	0.964508
XGB (Cooling)	0.67385	1.065218	0.987754

3.2 SHAP-based global interpretability

The SHAP global-importance analysis (Figure 3a) quantifies the mean absolute contribution of each input feature to the predicted heating and cooling loads, providing an interpretable ranking of the model's learned dependencies.

For both targets, *Overall Height* emerges as the most influential variable, with mean |SHAP| values of +5.29 (for heating) and +4.29 (for cooling). This reflects the geometric sensitivity of energy demand to vertical scaling, which directly alters surface-to-volume ratio and buoyancy-driven convection. Wall Area and Surface Area follow closely, confirming that envelope exposure remains a critical determinant of conductive and radiative heat transfer. The influence of Glazing Area is also significant, particularly for cooling load prediction, where it captures the strong radiative and solar-gain effects through fenestration. In contrast, Orientation, Glazing-Area Distribution, and Roof Area show comparatively minor impacts, suggesting that their effects are either captured indirectly through correlated features or are of secondary importance in this synthetic dataset. Overall, the SHAP hierarchy aligns with building-physics intuition: larger, taller, and more glazed geometries intensify both heating and cooling requirements, demonstrating that the neural model's internal representations remain consistent with thermodynamic reality.

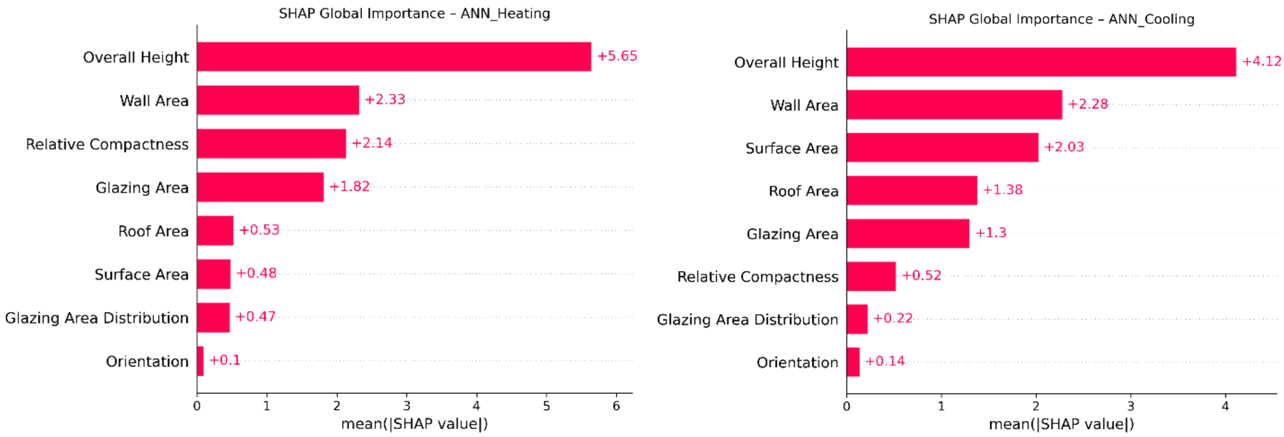

Figure 3a. SHAP global feature-importance ranking for the ANN model. Bars represent the mean absolute SHAP value, indicating each feature's average contribution to the predicted heating and cooling loads. The consistent dominance of overall height, wall area, and glazing area demonstrates that the ANN captured physically meaningful dependencies governing building thermal performance.

Interpretation of SHAP Beeswarm Plots

The SHAP Beeswarm diagrams (Figure 3c) provide a detailed, instance-level view of how individual feature values influence the ANN predictions for heating and cooling loads. Each point represents a specific observation, with color indicating the feature's magnitude and horizontal displacement denoting its marginal impact on the model output. For both targets, Overall Height and Wall Area exhibit the most pronounced effects: higher values (in red) are consistently associated with increased predicted loads, reflecting the higher conductive and convective losses in taller or more vertically extended geometries. Surface Area and Glazing Area also show strong positive correlations with both heating and cooling demands, confirming that larger envelope exposure and window area amplify energy transfer. In contrast, Relative Compactness demonstrates the opposite trend-lower compactness (blue) shifts SHAP values toward higher loads, consistent with the increased external surface relative to building

volume.
Less influential variables such as Orientation and Glazing-Area Distribution display mixed or near-zero SHAP dispersion, suggesting limited sensitivity in the synthetic dataset. Overall, the Beeswarm patterns confirm that the neural model not only ranks the dominant features accurately but also captures their directionality in line with building-physics expectations, reinforcing the interpretability and physical consistency of the ANN predictions.

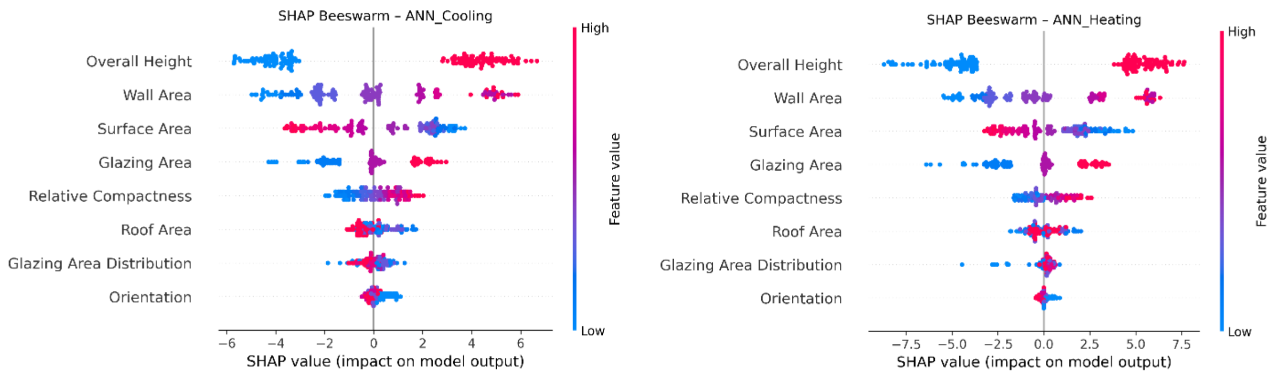

Figure 3c. SHAP Beeswarm plots showing the signed impact of each input feature on the predicted heating and cooling loads (ANN model). Each dot represents one observation, colored by the feature's actual value. Red points indicate higher feature values, and blue points indicate lower ones. The consistent directional trends confirm that the ANN has learned physically meaningful dependencies between geometric parameters and thermal-energy demand.

It is interesting to note in Figure 3c that in terms of quantitative influence the order of variables is the same for both cooling and heating loads.

3.3 Local and marginal effects

Partial Dependence Plots (PDPs) for the top predictors (Figure 3b) further reveal smooth nonlinear trends.
Heating load decreases exponentially with increasing compactness up to a saturation point, after which gains become marginal. Cooling load, conversely, exhibits a mild U-shaped response with compactness, suggesting that very compact buildings may retain heat during summer. For glazing area, PDPs show nearly linear increases in cooling load, confirming that larger window surfaces consistently amplify cooling energy demand. Overall, the ANN's learned behavior aligns with fundamental thermophysical principles, validating its internal consistency.

Figure 3b presents the partial dependence of the predicted heating and cooling loads on four principal design parameters. The results illustrate smooth, monotonic relations, confirming that the neural model captures physically meaningful trends. For Relative Compactness, both heating and cooling loads increase as compactness decreases, reflecting higher thermal losses in elongated building forms with greater exposed surface. The effect is stronger for heating, indicating that compactness is a dominant factor in winter energy demand.
Surface Area shows the inverse behavior: buildings with larger envelopes experience systematically lower predicted loads, confirming the complementary nature of these correlated descriptors.
For Overall Height, both loads rise with increasing height, which may be attributed to the greater thermal gradients and internal air-mixing effects in taller spaces. Finally, Glazing Area produces a nearly linear increase in both

heating and cooling loads, highlighting the dual influence of window area: increased solar gain amplifies cooling requirements while also elevating winter heat losses. These consistent, interpretable relationships reinforce the model's validity and demonstrate that the ANN infers thermal-energy behavior aligned with fundamental building-physics principles.

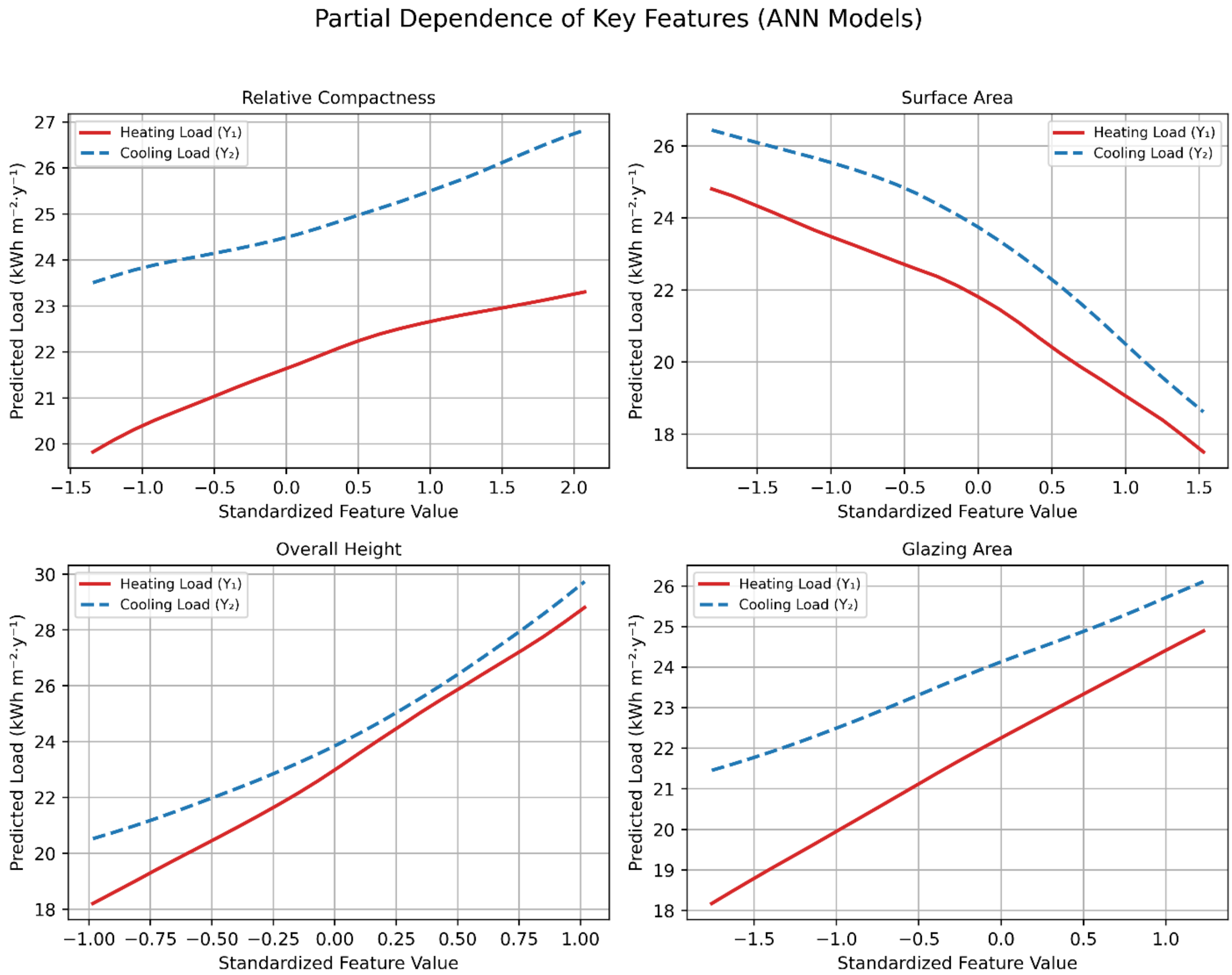

Figure 3b. Partial dependence of predicted heating and cooling loads on key building parameters: relative compactness, surface area, overall height, and glazing area. Each subplot shows the marginal effect of a standardized feature while all others are held at their mean. Red solid lines correspond to heating load predictions, and blue dashed lines to cooling load predictions (ANN model).

CONCLUSIONS

This two-part study has demonstrated a systematic workflow for understanding and predicting building energy performance using the Energy Efficiency Dataset (ENB2012). In *Part I*, exploratory data analysis and multicollinearity assessment revealed the dominant role of compactness, surface area, and glazing configuration in determining heating and cooling demands. In *Part II*, data-driven models based on Artificial Neural Networks (ANN) and XGBoost were developed and compared, achieving high predictive accuracy ($R^2 > 0.98$ for heating and $R^2 \approx 0.95$ for cooling). While XGBoost provided slightly superior numerical performance, the ANN model offered a continuous and differentiable representation suitable for explainability analysis. Through SHAP and Partial Dependence Plot (PDP) techniques, the models were shown to learn physically consistent and interpretable relationships. Relative compactness and overall height emerged as the strongest predictors of heating load, whereas glazing area and surface exposure predominantly influenced cooling demand. The monotonic, directionally coherent effects observed across all features confirm that the learned dependencies are consistent with thermodynamic expectations rather than overfitting artifacts. The integration of explainable AI (XAI) methods proved critical for validating and interpreting model behavior. Such hybrid approaches-combining quantitative performance with interpretability-provide not only accurate forecasts but also trustworthy insights for energy-efficient building design. The methodological framework presented here can be generalized to other datasets and extended to multi-objective optimization tasks, such as envelope parameter tuning or retrofit prioritization. Future research will focus on coupling these interpretable models with physics-informed neural architectures and real-world building datasets to enhance generalizability and practical adoption in sustainable energy engineering.

REFERENCES

1. Dinh, T.Q.; Ho, H.; Zhai, Z.J. "Deep neural networks for building energy load prediction: A review." *Energy and Buildings*, 271 (2022): 112348. https://doi.org/10.1016/j.enbuild.2022.112348
2. Amasyali, K.; El-Gohary, N.M. "A review of data-driven building energy consumption prediction studies." *Renewable and Sustainable Energy Reviews*, 81 (2018): 1192–1205. https://doi.org/10.1016/j.rser.2017.04.095
3. Lundberg, S.M.; Lee, S.-I. "A unified approach to interpreting model predictions." Advances in Neural Information Processing Systems, 30 (2017): 4765–4774. https://doi.org/10.48550/arXiv.1705.07874
4. Racoceanu Cristinel, ENERGY BALANCE MEASUREMENTS FOR AUXILIARY INSTALLATIONS OF LARGE POWER ENERGY GROUPS, Annals of the „Constantin Brancusi" University of Targu Jiu, Engineering Series, No. 2/2017, pag. 27 -31 , ISSN 1842 – 4856.
5. Racoceanu Cristinel, STUDY ON INFLUENCE OF ENERGY EFFICIENCY OF A STEAM BOILER BENSON ON ENVIRONMENTAL POLLUTION, Fiability&Durability, no.2/2017, pag.35-40, ISSN 1844-640X.
6. Racoceanu Cristinel, STUDY ON THE ENERGY BALANCE OF THE ELECTRO-PUMPS WITH CONDENSATION FOR THE ENERGY GROUPS IN CONDENSATION OPERATING WITH INFERIOR COAL, Annals of the „Constantin Brancusi" University of Targu Jiu, Engineering Series, No. 3/2016, pag.19-23, ISSN 1842 – 4856.