A Hybrid Machine Learning Approach for Energy Consumption Prediction in Additive Manufacturing

Yixin Li[1], Fu Hu[1], Jian Qin[2], Michael Ryan[1], Ray Wang[3], Ying Liu[1]

[1]High-Value Manufacturing Research Group
School of Engineering, Cardiff University
Cardiff, UK
{liy248, huf4, ryanm6, LiuY81} @cardiff.ac.uk
[2]The Welding Engineering and Laser Processing Centre
School of Aerospace, Transport and Manufacturing, Cranfield
j.qin@cranfield.ac.uk
[3]Unicmicro (Guangzhou) Co., Ltd, China
ray.wang@unicmicro.com

Abstract. Additive manufacturing (AM), as a fast-developing technology for rapid manufacturing, offers a paradigm shift in terms of process flexibility and product customisation, showing great potential for widespread adoption in the industry. In recent years, energy consumption has increasingly attracted attention in both academia and industry due to the increasing demands and applications of AM systems in production. However, AM systems are considered highly complex, consisting of several subsystems, where energy consumption is related to various correlated factors. These factors stem from different sources and typically contain features with various types and dimensions, posing challenges for integration for analysing and modelling. To tackle this issue, a hybrid machine learning (ML) approach that integrates extreme gradient boosting (XGBoost) decision tree and density-based spatial clustering of applications with noise (DBSCAN) technique, is proposed to handle such multi-source data with different granularities and structures for energy consumption prediction. In this paper, four different sources, including design, process, working environment, and material, are taken into account. The unstructured data is clustered by DBSCAN so to reduce data dimensionality and combined with handcrafted features into the XGBoost model for energy consumption prediction. A case study was conducted, focusing on the real-world SLS system to demonstrate the effectiveness of the proposed method.

Keywords: Additive Manufacturing, Energy Consumption, Modelling, Machine Learning, Multi-Source Data

1   Introduction

AM is often referred to as a 3D printing technology and defined as a process of adding materials layer by layer to fabricate products based on 3D model data [1]. Compared with conventional manufacturing techniques, AM provides the feasibility for complex-shaped parts [2]. Besides, AM allows a short time to fabricate products from a concept,

and it shows the significance of processing and improving material properties of products with low material consumption [3]. Therefore, AM is applied in the variety of industrial and medical usages, such as aerospace, automobile and dental equipment [4].

However, the growing environmental concerns on sustainability, particularly on energy consumption, have been considered [5]. Therefore, improving energy management in AM systems is urgent. So far, more and more researchers and manufacturers have increased their attention towards this aspect. AM has the potential to achieve a larger yield of product, resulting in an increasing amount of energy consumed. The eco-design is necessary at an early stage in the manufacturing system, supporting designers and manufacturers in energy management, decision-making and improvement in the process [6]. However, it is challenging to improve energy management, as the subsystems of AM will generate numerous data during the entire process. As investigated by Ahuett-Garza and Kurfess [7], the production process of an AM system consists of six stages, including 1) conversion, 2) positioning and orientation, 3) adding support structure, 4) slicing, 5) fabrication and 6) post-processing. The variety of data sources leads to the complexity of processing data due to the different dimensions and structures of the data in energy consumption prediction. According to Qin et al. [8], four categories of data related to design, process, working environment and material are considered, which are also known as the multi-source data. The multi-source data contains valuable information that uncovers the correlations between the selected features and energy consumption. Compared with traditional manufacturing techniques, the complexity of AM systems is very challenging for energy consumption analysis. However, the application of IoT technology can perform real-time monitoring from multiple processes. This makes a single model not suitable for this situation. Therefore, this paper proposes a hybrid ML method for predictive modelling of energy consumption.

The method is integrated by using supervised and unsupervised learning, which are proposed as XGBoost and DBSCAN, respectively. The data were obtained from an SLS system with different sources, where energy consumption is affected. The primary function of the DBSCAN algorithm is to get informative data and simultaneously reduce data dimensionality, combining selected features into the XGBoost model for energy consumption prediction. Root mean squared error (RMSE) and the model correlation coefficient (MCC) is used to assess the model performance.

Section II reviews the factors that influence energy consumption in AM systems and existing studies of ML techniques to establish predictive modelling in various AM system under different scenarios. Section III describes the detailed framework supporting the adoption of DBSCAN and XGBoost. Section IV demonstrates the outcome of the proposed methodology in the specific SLS system. Section V concludes the paper.

2 Literature Review

2.1 Analysis of Energy Consumption in AM Systems

AM has promoted a new manufacturing pattern that is involved in small-batch manufacturing with customisation, satisfying customers' demand [9]. According to different working principle and material supplies, some typical AM techniques include electron

beam melting (EBM) [10], fused deposition modelling (FDM) [6], selective laser sintering (SLS) [11] and selective laser melting (SLM) [12].

Table 1. The List of Different AM Systems in terms of Working Principles, Material Supply and Energy Consumption

AM System	Working Principles	Material Supply	Unit Energy Consumption (kW/h)
EBM	Utilisation of high-intensive electron beam to melt the material powder	Ti-6Al-4V, 316L stainless steel	17-49.2
FDM	The nozzle of printer extrudes fused thermoplastic material	Acrylonitrile Butadiene Styrene (ABS), Polycarbonate (PC)	23.1-346.4
SLS	Laser sinters the material powder	Polyamide, nylon	14.2-40.0
SLM	Laser melts the material powder	Ti-6A-4, 316L stainless steel	23.1-163.3

Noticeably, different working principles and types of supplied material used in AM systems play a vital role in affecting the energy consumption according to Table I. Other impact factors from process, design, working environment and material are also considered. Many researchers have investigated those impact factors in different AM systems. For instance, an investigation was given by Paul and Anand [13], who conducted numerical studies, demonstrating the impact factors, including layer thickness and part orientation. Baumers et al. [14, 15] conducted experiments to determine the four relevant factors during the AM process by comparing two different working platform. In their research, the processing stage, scanning, recoating, and Z-height consumed energy. Peng [16] focused on the process of 3D printing and broadened the analysis in terms of primary and secondary energy. Primary energy referred to direct energy consumption such as material form and properties, while secondary energy highly depended on in-process energy consumption. Watson and Taminger [2] computed an energy consumption model to illustrate the flow of energy consumption, considering the life cycle from feedstock to the end of product life. Differently from other authors, they highlighted that the transportation distance influenced the energy consumed in the AM system. Liu et al. [17] investigated the machine and process level in AM system that could have a significant impact on energy consumption. Furthermore, they also concentrated on material characteristics from the micro-level, which has an indirect impact on energy consumption.

For the investigations and studies above, it is found that various factors from different sources have a significant impact on the energy consumption of the AM systems. These contribute to a better understanding of applying the predictive model in practice. With regards to different scenarios, it is essential to set up specific models which are progressive to the specific task. The following contents in this section briefly review the machine learning (ML) for predictive modelling in different AM systems.

2.2 Machine Learning for Predictive Modelling in AM System

In general, AM systems are complicated manufacturing systems, including many sub-systems and the sensors and processors which generate massive amounts of data with various features and types under the IoT environment [9]. Hence, the time cost of pre-processing often makes up the largest proportion of time used during the entire prediction tasks. With consideration of the redundancy and irrelevance of some data, it is crucial to extract the relevant information by advanced data analytics. Machine learning (ML) and deep learning (DL) show merit to tackle the issue.

In the experiments by Bhinge et al. [18], data were collected from a FANUC controller and using a high-speed power meter to handle process data and power time-series data, respectively. The work applied the Gaussian Process regression model to cope with the small amount of data with high dimensionality. The model showed the significance of illustrating the correlation between features and energy usages in machine tools, which helps with depicting the tool path of the target machine for energy management.

Another application of ML and DL in anomaly detection was achieved by Zhang et al. [19]. A high-speed camera captured the images from the SLM system. Firstly, the conducted work was to apply principal component analysis on feature selection, reducing data dimensionality. Secondly, the image data were combined into the support vector machine and convolutional neural network (CNN) to classify the image data according to features with 90.1% and 92.7% of accuracy, respectively.

Other researchers benefited from ML techniques for the feature selection and extraction in AM systems. For instance, Wu et al. [20] conducted a numerical study to apply data integration method at the feature level, to process signals received by monitoring as the input, which was applied in ML models, followed by predicting the surface roughness of builds in the AM system. Some contributions were investigated in [8] and [21]. In [8], Qin et al. proposed a hybrid ML and DL approaches to deal with unstructured data with different features, types and dimensions in a complex AM system. The advantage is that it achieves considerable information compression. In [21], the focus of this paper was to integrate heterogeneous data to uncover the hidden knowledge with correlations between different features to help designers make decisions.

3 Methodology

This section will elaborate on the proposed methodology, targeting the multi-source data from an SLS system by using DBSCAN to cluster the unstructured data, which is then integrated into the XGBoost model to predict the energy consumption. The main stages of the experiment consist of three main stages, i.e. 1) data sensing and collection, 2) the hybrid ML-based approach for predictive modelling and 3) model validation and evaluation.

3.1   Multi-Source Data Sensing and Collection

The original data is collected from the target AM system, where the data can be categorised into four different types. They are operation process, material, working environment and design. In Fig. 1, process data stem from the parameter settings collected from the SLS machine, such as the measured values from the dispenser, recoater speed and the laser power used in sintering, which relies on the experience and knowledge of the operators. With regards to material data, it depends on the material itself. In this case, the type of material is known, referring to two kinds of nylon powder. Design data is the data collected from computer-aided design (CAD) models created by designers, often including design parameters for each layer [5], which are often determined at the beginning stage of the entire process. The working environment can be monitored by sensors and data stored in the conditional monitoring files for the illustration. This kind of data source collected from the working environment by an IoT platform is considered as the layer level from real-time monitoring. Some monitoring files can demonstrate these data to better comprehend the structure of the data.
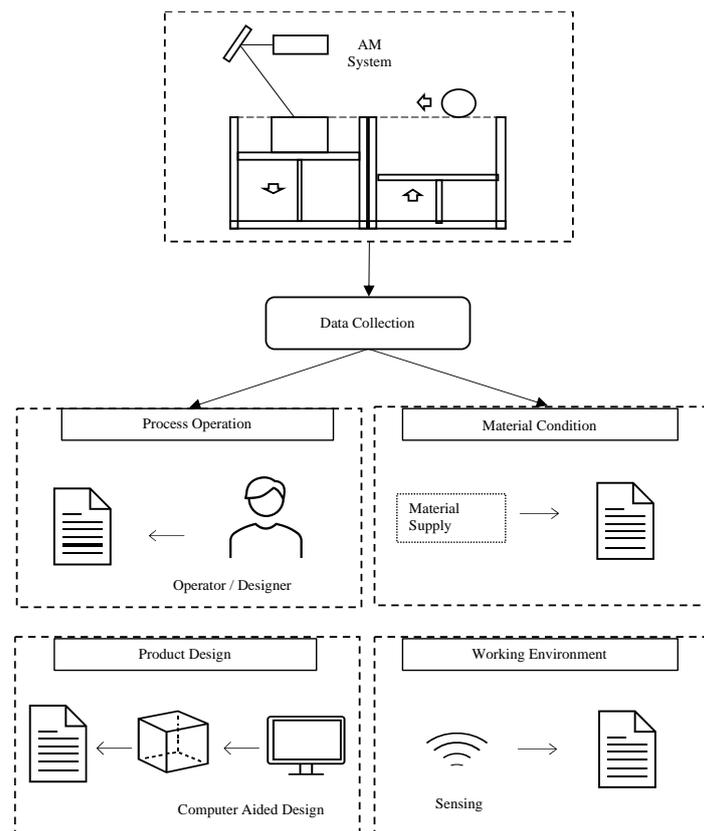
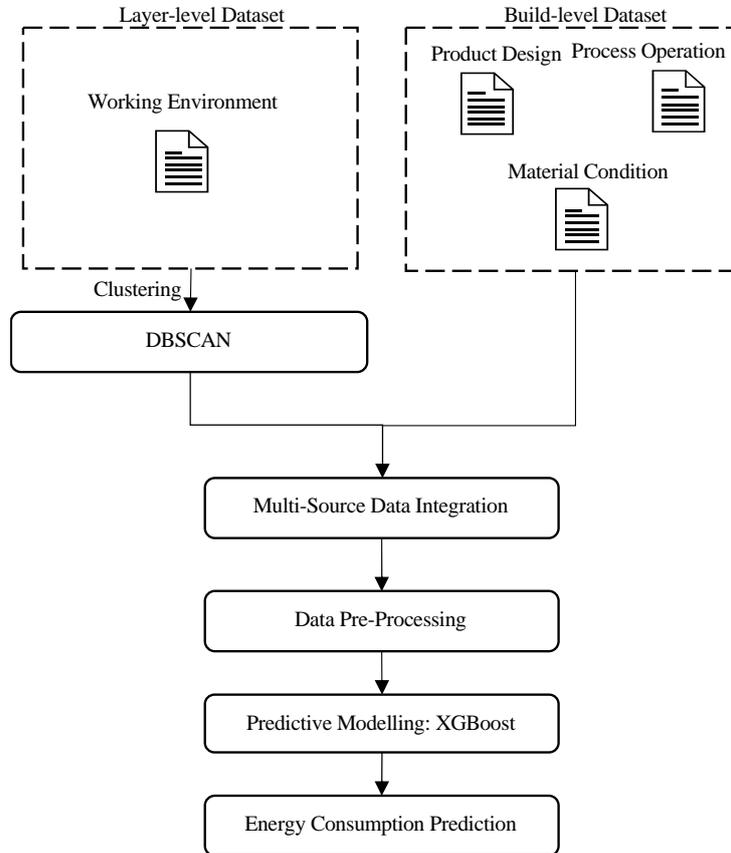**Fig. 1.** The Multi-Source Data Collection from AM System.

Fig. 2. The Framework of Proposed Methodology for Energy Consumption Prediction

Fig. 2 demonstrates the framework of the proposed methodology in pre-processing and modelling. The entire process is divided into three stages, corresponding to their respective roles. At the first stage, the input data are collected from the SLS system and categorised into four types of datasets according to their sources. The working environment data contains different quantities in each separate file, which is essential for integrating these data with using DBSCAN to unify the structure of layer-level data. Secondly, the dimensionality of the integrated dataset is reduced through DBSCAN clustering, and this dataset is fed into the XGBoost decision tree. Finally, the energy consumption is obtained, and the performance of the XGBoost model is evaluated using RMSE and MCC.

3.2 The Hybrid ML Approach for Energy Consumption Prediction

Advanced data analysis and ML methods show the ability to predict the model. ML is usually divided into supervised learning, unsupervised learning and reinforcement

learning according to the learning method [22]. This work uses a hybrid method that combined unsupervised learning and supervised learning, i.e. unsupervised learning aims to integrate different dimensional datasets, while supervised learning is utilised to predictive the energy consumed in the SLS system.

The clustering problem is related to an unsupervised learning problem. According to predefined rules, the clustering problem is used to find the uncovered patterns to be classified with similar characteristics between data [23]. DBSCAN is a data clustering algorithm targeting unstructured data. Specifically, DBSCAN used a density-based clustering approach, which is the most commonly used in clustering the spatial data. This algorithm adopts the concept of density-based clustering, which requires the number of points lied in a specific region of the clustering space, with minimum numbers of objects ($MinPts$) and it should exceed the given threshold. The following equations demonstrate the nature of DBSCAN and the random point p in its neighbourhood is defined in equation (1)

$$N_{Eps} = \{q \in D / dist\,(p, q) < Eps\} \tag{1}$$

$$N_{Eps}(P) > MinPts \tag{2}$$

where Eps is the neighbourhood of the radius, given the collection of objects D. The core point P is defined in equation (2) if it contains minimal numbers of points.

In other words, a core point, a boundary point or an outlier is determined by two indicators: $MinPts$ and Eps, and the outlier is removed. The algorithm connects core points under the condition of equation (2), allocating the boundary point to the closest core point and finally obtaining the clustering results [24].

When comparing to k-means clustering, DBSCAN is faster in terms of clustering speed and more effective in processing noise points, handling abnormal data, and in exploring spatial clusters of random shapes. Besides, the unbiased-shaped clusters do not need to divide the number of clusters [25, 26]. A satisfactory clustering algorithm needs to have the following characteristic: 1) to determine knowledge from inputs, especially for the large datasets, 2) capable of finding arbitrary shaped clusters and 3) efficient to handle large datasets [27]. The working environment data is collected layer by layer over thousands of data in separate files with various types from the entire process, as a consequence of large data volume and heterogeneity. Therefore, DBSCAN is expected to tackle the issues. Furthermore, this algorithm was applied at the beginning, and it demonstrates the mean values which can be a representative of the entire cluster. These values can be combined into design-relevant datasets on the build-level, in order to unify the format of each working environment data file.

XGBoost refers to a tree-based ensemble learning using tree algorithm, proposed by Chen and Guestrin [28]. This boosting method is an effective ML method. XGBoost uses regression trees ensembles which have the same decision rules as the decision tree (DT), containing one score in each leaf value. Two aspects allow it to be distinguished from other tree boosting machine. Firstly, XGBoost has a different objective function. For each regression tree, this ensemble method accumulates the sum of scores as the prediction value for all tree. Assuming there are k trees, the output for tree ensemble is defined as follow:

$$\hat{y}_i = \sum_{k=1}^{K} f_x(x_i), f_x \in \mathcal{F} \qquad (3)$$

Moreover, the objective function, which is the sum of training loss and complexity of the trees to control overfitting, is defined as follow:

$$Obj = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{k}^{K} \Omega(f_k), f_x \in \mathcal{F} \qquad (4)$$

$$\Omega(f_k) = \gamma T + \frac{1}{2}\lambda\|w\|^2 \qquad (5)$$

where $\hat{y}_i$ is the predicted value of the model, y_i stands for the ith feature label, f_k represents the kth tree model, T is the number of nodes and w is the collection of score combinations. In the reduction of the objective function, predicted value adds a new function f in each iteration. This additive training defines a new objective function to optimise and search for a new tree model.

Another difference is the division of nodes. There are four proposed splitting algorithms from Chen and Guestrin's work. XGBoost adopts (1) basic exact greedy algorithm, (2) approximate algorithm, (3) weighted quantile sketch and (4) sparsity-aware split finding methods. Among these four split finding algorithms, algorithm (2) and (3) is to solve the problem that the data fails to load into memory at once or algorithm (1) is not distributed efficiently. The XGBoost approach calculates the gain of each feature in parallel and chooses the feature with the largest information gain to split.

XGBoost provides an idea for processing sparse data and enables the handling of instance weights in tree learning. Compared with the traditional tree model, it shows the merits of regularisation in controlling the model complexity and reducing the variance of the model to avoid overfitting. This model is used to predict the energy consumption in the SLS system. By targeting this specific task, XGBoost integrates the weak learner to form a stronger learner to increase the accuracy. In addition, the sparsity-aware split finding method of XGBoost can process the missing values in the combined datasets. Also, it increases the learning rate effectively by controlling the model complexity, which is important when dealing with large datasets.

3.3 Validation of Prediction Model

Various subsystems consume electric power in the SLS system [29]. The consumption is noted as E_U, referring to the unit energy consumption in kWh/kg and it is:

$$E_U = \frac{E_T}{M_T} \qquad (6)$$

where E_T is the total energy consumed in the AM system and M_T means the total weight of fabricated products.

The performance of the predictive model XGBoost decision tree can be evaluated by RMSE and MCC [29-31]. RMSE identifies the actual value (a_t) and the predicted value (p_t). The low value of RMSE determines the high accuracy of the model, which is given by:

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(p_t - a_t)^2} \tag{7}$$

$$MCC = \frac{S_{PA}}{\sqrt{S_P S_A}} \tag{8}$$

$$S_{PA} = \frac{\sum_i (p_i - \bar{p})(a_i - \bar{a})}{n-1}; S_P = \frac{\sum_i (p_i - \bar{p})^2}{n-1}; S_A = \frac{\sum_i (a_i - \bar{a})^2}{n-1} \tag{9}$$

In equation (8) and (9), MCC reveals the correlations between the predicted and actual data obtained from the model, where $\bar{p}$ is the mean value of predicted data, and $\bar{a}$ is the mean value of the entire data.

4 Case Study

The case study was based on an SLS machine (EOS P700) using nylon powder (PA2200 and PA3200GF) to create builds. The data was collected from 2016 to 2018. The working environment data has different quantities, and pre-processing is considered.

4.1 Experimental Setup

Data description – The data collected from the SLS system are divided into four categories, including process, material, working environment and design. These four datasets obtained from different sources can be distinguished from their data structure.

By inspecting the datasets, the variety of data types is taken into accounts, such as time-series data (build date), nominal data (material types) and numeric data (data collected from the four described sources). Different types of data sources make it complicated to analyse data directly, while data pre-processing is critical before predictive modelling. In details, the layer- and build-level data are allocated based on four kinds of datasets.

For instance, in the working environment dataset, data has different dimensions because each parameter or feature was collected layer by layer, indicating the different height of prints. With regards to the design data, combining with information of material supply and operation process, this combined dataset consists of build date, process parameters, material supply and unit energy consumption of each build. Fig.3 illustrates the statistical distribution of energy consumption. It can be observed that most of the values are located in the range of 200 to 400 kWh/kg. However, the energy consumption of each build showed the difference.

The multi-source data makes it complicated to model directly with these heterogeneous data. By observing the collected datasets, 31 attributes are recorded in the corresponding datasets. These datasets are allocated into two new classes: layer-level and build-level, which layer-level datasets include working environment data and build-level data constitute design, process operation and material. For parameters that remain constant during the process, such as scan speed, the material type and the height of builds, they can be classified as build-level data, corresponding to process, material and

product design data, respectively. Conversely, the working environment data are collected from each layer where the information tends to be different during the manufacturing process. In addition, the informative data monitored and collected from the working environment are placed into layer-level datasets, such as the chamber temperature and the oxygen content.

Table 2. Data Categories in terms of Sources and Types.

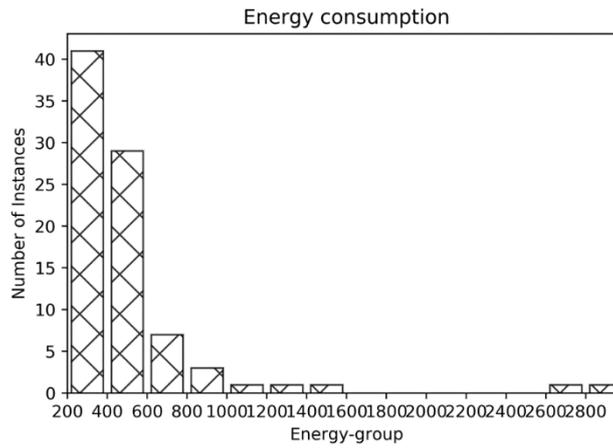| Data Types	Data Sources	Data Attributes
Build-level data	Process operation	Dispenser values, Scan speed, Recoater speed, etc.
	Product design	Number of builds, Build height, Filling degree, etc.
	Material	Type of material supply
Layer-level data	Working environment	Chamber temperature, Frame temperature, Oxygen level, scanner temperature, etc.

Fig. 3. The Distribution of Unit Energy Consumption from Datasets.

Data pre-processing – Prior to the establishment of the predictive model, this stage is used to minimise the data noise and outliers from the datasets. Some instances or features containing massive missing values, they were replaced by mean values or removed from the whole dataset. Considering the complexity and heterogeneity of working environment data, DBSCAN provides a clustering method to select a specific value which can represent the cluster, targeting the working environment data. This is utilised to select the valuable feature to conduct training from existing features, simultaneously reducing the dimensionality to unify the datasets. Differently from DL techniques, ML techniques need labelled and handcrafted features. This method aims to extract a series

of the most representative data point from the clusters, which is also constructive to dimensionality reduction.

Model setups – The proposed methodology is to use XGBoost decision tree to predict energy consumption. Three other algorithms are adopted to the prediction task as benchmarks to compare the performance. These are SVR with linear kernel, gradient boosting regression tree (GBRT) and convolutional neural network (CNN). By comparison, the prediction performance of the proposed model is more convincing.

Model validation – The model adopts five-cross validation for testing. XGBoost and benchmarks can be evaluated by RMSE and MCC.

4.2 Results and Discussion

There are three ML and one DL algorithms adopted in energy consumption prediction of the case study. The working environment data was trained then combined with other types of datasets into the proposed ML technique and three other benchmarks. After applying XGBoost, RMSE, and MCC to determine the performance of the model, the effectiveness of the proposed approach is demonstrated. Fig. 4 illustrates the comparison between the four algorithms with using layer-level, build-level and combined datasets. It can be observed that XGBoost appears the highest value of MCC, 0.708 when applying combined datasets, which shows the best degree of fitting to the experimental data when DBSCAN is applied. In general, this coefficient lies in the range of -1 to 1, representing negative correlation and positive correlation, respectively. It can be seen that all variables are positively correlated with the output value, that is, energy consumption. Followed by the MCC of XGBoost, that of SVR and GBRT's MCC obtains similar values (0.669 and 0.676, respectively), indicating that these two models are suitable for prediction. Regarding CNN, this DL technique often processes image data and performs classification tasks. The application of regression is uncommon. In addition, when applying the multi-source data into the predictive models, the results show that some of the performance of algorithms (GBRT and XGBoost, known as ensemble methods, based on the tree learner) will be optimised, while other methods yield a slight decrease in MCC. SVM yields the best results when applying the layer-level datasets, while it obtains the lowest MCC when only using build-level datasets.

RMSE describes the error in the models more intuitively and refers to the loss function from the regression analysis. Fig. 5 demonstrates the comparison of the RMSE for each model. For the RMSE of XGBoost (130.783 kWh/kg), which is within an acceptable range, it measures the deviation between actual and predicted values. This value shows the lowest RMSE when applying the entire datasets. After that, the RMSE of CNN changes dramatically and has the largest value at 231.958 kWh/kg. For this neural net-based algorithm, the best application of industry is associated with pattern recognition or classification. The adoption of multi-source data will affect the performance of predictive models. Fig. 6 is the comparison between test data and predicted data, which shows a similar trend for energy consumption. As a result, it can be observed that some outliers affected the final results. When utilising more heterogeneous data and combining them into XGBoost regression tree, the pattern of data fluctuates. However, the trend of predicted data and original still demonstrate a big gap as more data enters,

which may be caused by irrelevant features from the datasets and can be solved by collecting data to create new features.
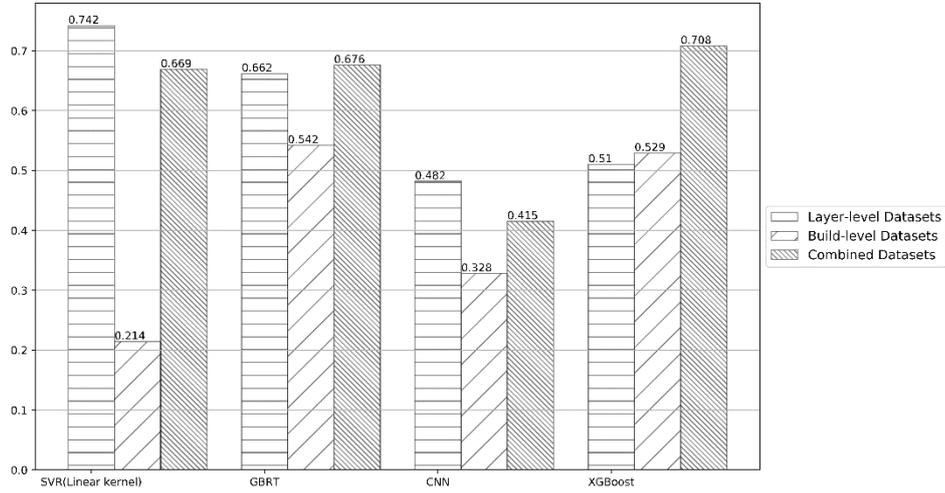

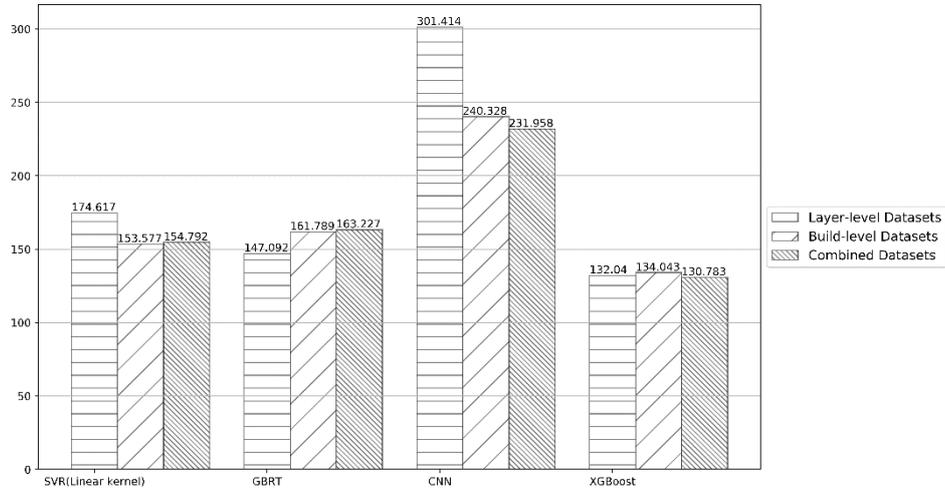**Fig. 4.** Comparison of MCC of XGBoost and Benchmarks.

Fig. 5. Comparison of RMSE of XGBoost and Benchmarks.

By analysing the nature of different predictive models, ensemble methods (GBRT and XGBoost) show the merits in prediction, and statistic-based algorithm (SVM), also has good learning performance. CNN, as the neural-net-based algorithm, has applicability in processing pattern and mere applications in the regression or prediction. Consequently, it has the lowest outcome in MCC and the highest values in RMSE. Another

pattern can be observed by the integration of datasets. The combined dataset is determined to influence the prediction performance of each model. The consideration of the overfitting problem is the most essential. Complicated training models may lead to the overfitting of the training data. Furthermore, as one of the challenges of ML, the quality of data will influence the performance of models. If training data contains many noise and outliers, the performance of the model will be affected. In some situation, it depends on the size of samples, which requires the representative features. Thus, before employing heterogeneous data into predictive models, data pre-processing is necessary.
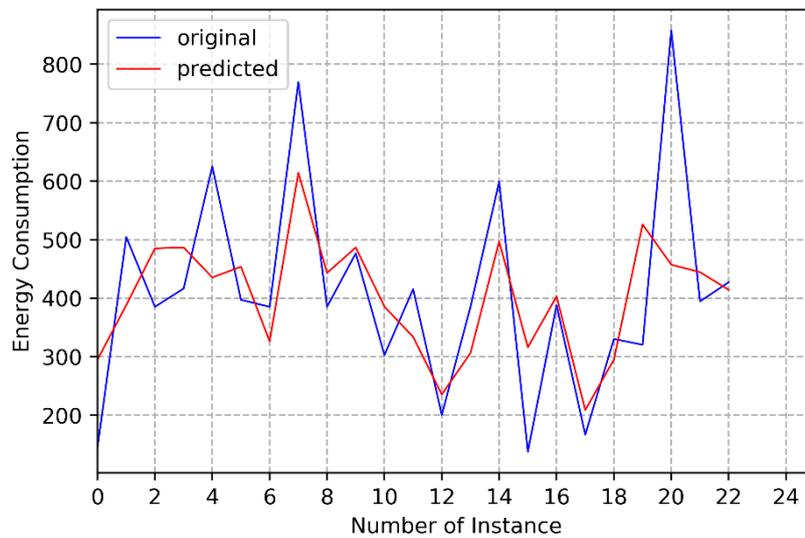

Fig. 6. The Prediction Result between Predicted Values and Original Values.

This hybrid ML approach has presented better performance among the other three algorithms and connected the target and input with high dimensionality when using combined datasets. For tackling the real-world issue regarding handling heterogeneous data, a single learner cannot be adopted, while the integration of DBSCAN and XGBoost suits the case.

5   Conclusions

AM is a comprehensive manufacturing technique embedded with various technologies, which currently employed across many different industries. Manufacturers and researchers are beginning to focus on the sustainability of AM and have started to optimise the process in terms of the energy aspect through data-driven approaches. However, different subsystems generate this data which makes it complicated to handle data,

meaning a single predictive model does not adapt to this complicated situation. Hence, a hybrid ML-based approach is proposed.

The experiment firstly adopted the DBSCAN method to select informative and representative data, simultaneously reducing data dimensionality, where the data was clustered and combined into the predictive model. Secondly, this paper applied a tree-based ensemble learning technique, XGBoost, and used it to predict the energy consumption of the SLS system. The evaluation matrix of XGBoost demonstrated a significance in the performance in dealing with heterogeneous data in a complex SLS system. The performance of XGBoost outperformed the other benchmarks, demonstrating the highest MCC and lowest RMSE compared with other algorithms, which shows that XGBoost greatly improves the degree of fitting and accuracy of the model. XGBoost, as an ensemble learning algorithm, provides a feasible approach to predict energy consumption in a complex AM system.

However, the predicted and original value show a significant gap after prediction. This may be affected by interfering and irrelevant information from the raw datasets. Feature extraction will be considered in future work, which is expected to be implemented in the data pre-processing stage. This could improve the modelling performance when the model is built on the data from a large volume.

References

1. ASTM International, Standard Terminology for Additive Manufacturing Technologies. PA: ASTM International. West Conshohocken, 2012.
2. J. K. Watson and K. M. B. Taminger, "A decision-support model for selecting additive manufacturing versus subtractive manufacturing based on energy consumption," J. Clean. Prod., vol. 176, pp. 1316–1322, 2018.
3. L. Wang and C. A. Alexander, "Additive manufacturing and big data," Int. J. Math. Eng. Manag. Sci., vol. 1, no. 3, pp. 107–121, 2016.
4. K. Kellens, M. Baumers, T. G. Gutowski, W. Flanagan, R. Lifset, and J. R. Duflou, "Environmental Dimensions of Additive Manufacturing: Mapping Application Domains and Their Environmental Implications," J. Ind. Ecol., vol. 21, pp. S49–S68, 2017.
5. Y. Yang, L. Li, Y. Pan, and Z. Sun, "Energy Consumption Modeling of Stereolithography-Based Additive Manufacturing Toward Environmental Sustainability," J. Ind. Ecol., vol. 21, pp. S168–S178, 2017.
6. D. Freitas, H. A. Almeida, H. Bártolo, and P. J. Bártolo, "Sustainability in extrusion-based additive manufacturing technologies," Prog. Addit. Manuf., vol. 1, no. 1–2, pp. 65–78, Jun. 2016.
7. H. Ahuett-Garza and T. Kurfess, "A brief discussion on the trends of habilitating technologies for Industry 4.0 and Smart manufacturing," Manuf. Lett., vol. 15, pp. 60–63, 2018.
8. J. Qin, Y. Liu, and R. Grosvenor, "Multi-source data analytics for AM energy consumption prediction," Adv. Eng. Informatics, vol. 38, no. October, pp. 840–850, 2018.
9. Y. Wang, Y. Lin, R. Y. Zhong, and X. Xu, "IoT-enabled cloud-based additive manufacturing platform to support rapid product development," Int. J. Prod. Res., vol. 57, no. 12, pp. 3975–3991, Jun. 2019.
10. W. E. Frazier, "Metal Additive Manufacturing: A Review," J. Mater. Eng. Perform., vol. 23, no. 6, pp. 1917–1928, Jun. 2014.

11. F. Ma, H. Zhang, K. K. B. Hon, and Q. Gong, "An optimisation approach of selective laser sintering considering energy consumption and material cost," J. Clean. Prod., vol. 199, pp. 529–537, 2018.
12. S. L. Sing et al., "Direct selective laser sintering and melting of ceramics: A review," Rapid Prototyp. J., vol. 23, no. 3, pp. 611–623, 2017.
13. R. Paul and S. Anand, "Process energy analysis and optimisation in selective laser sintering," J. Manuf. Syst., vol. 31, no. 4, pp. 429–437, 2012.
14. M. Baumers, C. Tuck, D. L. Bourell, R. Sreenivasan, and R. Hague, "Sustainability of additive manufacturing: measuring the energy consumption of the laser sintering process," Proc. Inst. Mech. Eng. Part B J. Eng. Manuf., vol. 225, no. 12, pp. 2228–2239, Dec. 2011.
15. M. Baumers, C. Tuck, R. Wildman, I. Ashcroft, E. Rosamond, and R. Hague, "Transparency Built-in," J. Ind. Ecol., vol. 17, no. 3, pp. 418–431, Jun. 2013.
16. T. Peng, "Analysis of Energy Utilization in 3D Printing Processes," Procedia CIRP, vol. 40, pp. 62–67, 2016.
17. Z. Y. Liu, C. Li, X. Y. Fang, and Y. B. Guo, "Energy Consumption in Additive Manufacturing of Metal Parts," Procedia Manuf., vol. 26, pp. 834–845, 2018.
18. R. Bhinge, J. Park, K. H. Law, D. A. Dornfeld, M. Helu, and S. Rachuri, "Toward a Generalised Energy Prediction Model for Machine Tools," J. Manuf. Sci. Eng., vol. 139, no. 4, pp. 1–12, Apr. 2017.
19. Y. Zhang, G. S. Hong, D. Ye, K. Zhu, and J. Y. H. Fuh, "Extraction and evaluation of melt pool, plume and spatter information for powder-bed fusion AM process monitoring," Mater. Des., vol. 156, pp. 458–469, 2018.
20. D. Wu, Y. Wei, and J. Terpenny, "Predictive modelling of surface roughness in fused deposition modelling using data fusion," Int. J. Prod. Res., vol. 57, no. 12, pp. 3992–4006, 2019.
21. F. Hu, Y. Liu, J. Qin, X. Sun, and P. Witherell, "Feature-level Data Fusion for Energy Consumption Analytics in Additive Manufacturing," in 2020 IEEE 16th International Conference on Automation Science and Engineering (CASE), 2020, pp. 612–617.
22. J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, "A survey of machine learning for big data processing," EURASIP J. Adv. Signal Process., vol. 2016, no. 1, 2016.
23. H. Jiang, J. Li, S. Yi, X. Wang, and X. Hu, "A new hybrid method based on partitioning-based DBSCAN and ant clustering," Expert Syst. Appl., vol. 38, no. 8, pp. 9373–9381, 2011.
24. S. Fong, S. U. Rehman, K. Aziz, and I. Science, "DBSCAN : Past , Present and Future," pp. 232–238, 2014.
25. D. Birant and A. Kut, "ST-DBSCAN: An algorithm for clustering spatial-temporal data," Data Knowl. Eng., vol. 60, no. 1, pp. 208–221, 2007.
26. J. Sander, M. Ester, H. P. Kriegel, and X. Xu, "Density-based clustering in spatial databases: The algorithm GDBSCAN and its applications," Data Min. Knowl. Discov., vol. 2, no. 2, pp. 169–194, 1998.
27. T. N. Tran, K. Drab, and M. Daszykowski, "Revised DBSCAN algorithm to cluster data with dense adjacent clusters," *Chemom. Intell. Lab. Syst.*, vol. 120, pp. 92–96, 2013.
28. T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discov. Data Min., pp. 785–794, Mar. 2016.
29. J. Qin, Y. Liu and R. Grosvenor, "Data analytics for energy consumption of digital manufacturing systems using Internet of Things method," 2017 13th IEEE Conference on Automation Science and Engineering (CASE), Xi'an, 2017, pp. 482-487.
30. J. Han, J. Pei, M. Kamber, Data mining: concepts and techniques, Elsevier2011.
31. J. Taylor, Introduction to error analysis, the study of uncertainties in physical measurements, 1997.