

Article

Short-Term Firm-Level Energy Consumption Forecasting for Energy-Intensive Manufacturing: A Comparison of Machine Learning and Deep Learning Models

Andrea Maria N. C. Ribeiro¹, Pedro Rafael X. do Carmo², Iago Rodrigues³, Djamel Sadok⁴, Theo Lynn⁵, Patricia Takako Endo⁶

¹ Universidade Federal de Pernambuco; amnc@cin.ufpe.br

² Universidade Federal de Pernambuco; prxc@cin.ufpe.br

³ Universidade Federal de Pernambuco; irrs@cin.ufpe.br

⁴ Universidade Federal de Pernambuco; jamel@cin.ufpe.br

⁵ Irish Institute of Digital Business, Dublin City University; theo.lynn@dcu.ie

⁶ Universidade de Pernambuco; patricia.endo@upe.br

Abstract: To minimise environmental impact, avoid regulatory penalties, and improve competitiveness, energy-intensive manufacturing firms require accurate forecasts of their energy consumption so that precautionary and mitigation measures can be taken. Deep learning is widely touted as a superior analytical technique to traditional artificial neural networks, machine learning, and other classical time series models due to its high dimensionality and problem solving capabilities. Despite this, research on its application in demand-side energy forecasting is limited. We compare two benchmarks (Autoregressive Integrated Moving Average (ARIMA), and an existing manual technique used at the case site) against three deep learning models (simple Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Unit (GRU)) and three machine learning models (Support Vector Regression (SVM), Random Forest, and K-Nearest Neighbors (KNN)) for short term load forecasting (STLF) using data from a Brazilian thermoplastic resin manufacturing plant. We use the grid search method to identify the best configurations for each model, and then use Diebold-Mariano testing to confirm the results. Results suggests that the legacy approach used at the case site is the worst performing, and that the GRU model outperformed all other models tested.

Keywords: Short term load forecasting; STLF; deep learning; RNN; LSTM; GRU; machine learning; SVR; random forest; KNN; energy consumption; energy-intensive manufacturing; time series prediction; industry.

1. Introduction

The industrial sector is the largest consumer of delivered energy worldwide and energy-intensive manufacturing is the largest component in that sector [1]. Energy-intensive manufacturing includes the manufacture of food, beverage, and tobacco products, pulp and paper, basic chemicals, refining, iron and steel, non-ferrous metals and non-metallic metals [2]. Energy-intensity is driven by the mix of activity in these sectors including basic chemical feedstocks, process (including heating and cooling) and assembly, steam and cogeneration, and building-related energy consumption e.g. lighting, heating, and air conditioning [2]. World industrial energy consumption is forecast to grow from c. 242 quadrillion British thermal units (Btu) in 2018 to about 315 quadrillion Btu in 2050; the proportion of energy-intensive manufacturing is forecast to remain at approx. 50% during that period [1].

28 It is widely held that energy consumption tends to be positively associated with a higher rate of
29 economic growth [3], however against this backdrop is the environmental impact of this consumption.
30 In 2017, direct industrial carbon dioxide (CO₂) emissions rose 0.3% to 8.5 gigatonnes representing
31 approx. 24% of total global emissions[4]. Industrial contribution is much higher once indirect emissions
32 related to electricity use are included. As such, the industrial sector, in general, and energy-intensive
33 manufacturing in particular, are a significant negative contributor to climate change. As part of the
34 United Nations Sustainable Development Goals (SDG), SDG 9 has industrial targets to reduce CO₂
35 emission per unit of value added through increased resource-use efficiency and greater adoption of
36 clean and environmentally sound technologies and industrial processes¹.

37 To meet these targets, governments worldwide are imposing regulations and taxes to reduce
38 the environmental impact of industrial energy consumption. To minimise the environmental impact
39 and reduce costs, firms are increasingly adopting environmentally responsible manufacturing (ERM)
40 practices [5]². ERM is an economically driven, system-wide and integrated approach for reducing
41 and eliminating waste streams associated with the design, manufacture, use and/or disposal of
42 products and materials [6]; in this context waste also includes any type of pollution and energy
43 waste including greenhouse gas (GHG) emissions. Central to ERM is firm-level monitoring of
44 environmental performance based on objective measures [7]. This not only aids in planning and
45 decision making by management in relation to environmental performance and energy cost control
46 [8], but is effective in demonstrating the value of environmental efforts and promoting environmental
47 responsibility to middle management and individual employees [9]. Firm-level forecasts of energy
48 consumption are essential for precautionary and mitigation measures decision making by firms to
49 minimise environmental impact, manage cashflow, and reduce or eliminate risk [8]. Accurate firm-level
50 monitoring can also improve the data quality available to policymakers for local, regional, and national
51 policies and actions.

52 Extant literature has typically (i) focused on supply-side perspectives, (ii) aggregated energy costs,
53 and (iii) failed to recognise the idiosyncrasies of the energy-intensive manufacturing sector and the
54 associated centrality of energy management in production planning. There is a paucity of studies in
55 demand-side process-related short term load forecasting (STLF) using deep learning and machine
56 learning for energy-intensive manufacturing. The limited studies that have been published do not
57 compare deep learning performance against widely used machine learning models, classical time
58 series models, or approaches used in practice. In addition to proposing prediction models, we also
59 address this gap.

60 In this paper, we focus on performance analyses of deep learning and machine learning models
61 for firm-level STLF considering energy-intensive manufacturing process of a thermoplastic resin
62 manufacturer. We use energy consumption and production flow data from a real Brazilian industrial
63 site. In addition to the energy consumption time series, we use data from two different stages of the
64 thermoplastic resin production process – polymerisation and solid-state polymerisation. We propose
65 three deep learning models - simple Recurrent Neural Networks (RNN), Long Short-Term Memory
66 (LSTM) networks, and Gated Recurrent Unit (GRU) - and three machine learning models - Support
67 Vector Regression (SVR), Random Forest, and K-Nearest Neighbors (KNN)- for predicting daily energy
68 consumption. We use the grid search method to identify the best model configurations. We compare
69 the performance of the deep learning and machine learning models against (i) a classical time series
70 model, Autoregressive Integrated Moving Average (ARIMA), and (ii) the current manual approach at
71 the case site.

72 The remainder of this paper is organised as follows. Section 2 provides a brief overview of the
73 deep learning and machine learning techniques used in the study. Section 3 presents a discussion

¹ <https://sustainabledevelopment.un.org/sdg9>

² Sometimes referred to as Environmentally Conscious Manufacturing (ECM)

74 about the related works. This is followed by a description of the data, pre-processing, and proposed
75 evaluation metrics in Section 4. Section 5 presents the models identified for evaluation. Our results
76 and most relevant findings are discussed in Section 6. The paper concludes with a summary of the
77 paper and future avenues for research in Section 7.

78 2. Background

79 Artificial neural networks (ANN) are computational models that comprise many simple process
80 units called neurons which may be activated by inputs from the external environment, and likewise,
81 other neurons may trigger actions in the external environment [10]. The learning process involves
82 identifying weights that result in an ANN exhibiting desired behaviour [10]. ANNs have been used to
83 solve a wide range of regression and classification tasks for problems including object recognition [11],
84 natural language processing [12], and understanding signals [13].

85 Until recently, deep neural networks (DNN), while producing better results than shallow ANNs,
86 received comparatively less attention due to the time and effort required to train such neural
87 networks [10,14]. The introduction of hardware accelerators such as Graphics Processing Units (GPUs),
88 advancement in deep learning architectures and techniques, the availability of big data, and practical
89 success in high-impact problems renewed interest in deep learning in the last decade [10,14,15]. The
90 high dimensionality of DNNs enables them outperform traditional machine learning and shallow
91 ANNs particularly for modelling highly complex data or non-linear relationships between variables
92 [15]. Energy consumption data is typically modelled and analysed using time series. Many DNNs,
93 such as multi-layer perceptrons (MLP) and convolutional neural networks (CNNs) assume all inputs
94 are independent of each other, and as a result they may not be suited to modelling sequential data [15].
95 RNNs were specifically designed to target sequential data such as like time series data.

96 2.1. Deep Learning

97 2.1.1. Recurrent neural networks

98 RNNs are a type of ANN designed to recognise patterns in sequential data streams. In RNNs, the
99 decision, classification, or learning done at a given moment $t-1$ influences the decision, classification,
100 or learning at a subsequent time t in the time series. RNNs contain two sources of input, the present
101 and the recent past. These data are combined to determine how new data is predicted. RNNs have a
102 memory, that, for example, MLP and CNN do not. As such, RNNs use information in the sequence
103 itself to perform tasks that other ANNs are unable to do. RNNs have limitations, the most significant
104 of which are difficulties in training RNNs to capture long-term dependencies due to vanishing and
105 exploding gradient problems [16,17]. LSTM and GRU are variations of RNN that overcome such
106 problems.

107 2.1.2. Long short-term memory

108 LSTM [18] is a variation of RNN. LSTM overcomes gradient problems through the use of a chain
109 structure containing four neural networks and different blocks of memory [19]. LSTM updates its unit
110 states using three gates - a forget gate, input gate, and output gate. The forget gate deletes information
111 that is no longer useful in the unit [19]. The current input x_t and the output from the previous unit h_{t-1}
112 are multiplied by the weight matrix. The result is passed through an activation function that provides
113 a binary output that causes the data to be forgotten. The input gate performs the addition of useful
114 information to the unit's status. First, the information is adjusted using a sigmoid function. Then the
115 \tanh function is used to create a vector that produces -1 to $+1$. Finally, the output gate completes the
116 task of extracting useful information from the current state of the unit to be displayed as an output. In
117 order to do so, a vector is generated by applying a \tanh function to a cell. Due to its structure, LSTMs
118 can predict time series with time intervals of unknown duration [20], a significant advantage over
119 traditional RNNs. Notwithstanding this, long training times are a significant limitation [21].

120 2.1.3. Gated recurrent unit

121 GRUs reduce the complexity of LSTMs by only utilizing an update gate and a reset gate to
122 determine how values in the hidden states are computed [17]. In GRUs, only one hidden state
123 is transferred between the time steps [17]. This state is capable of maintaining long and short term
124 dependencies at the same time. GRU gates are trained to selectively filter out any irrelevant information
125 while maintaining what is useful. These gates are vectors containing binary values, as in LSTM, and
126 determine the importance of the information. Crucially, research suggests that GRUs have significantly
127 faster training times with comparable performance to LSTM [17,21].

128 2.2. Machine Learning

129 2.2.1. Support vector regression

130 Support Vector Machines (SVM) has been proposed as an alternative to traditional ANNs for
131 classification and regression tasks. In particular, SVM provides better support for forecasting time
132 series from non-linear systems [22]. SVM is a machine learning technique based on statistical learning
133 theory [23]. Extant literature suggests that SVM performs well in forecasting time series [22,24,25].
134 Support Vector Regression (SVR) is a regression technique based on SVM [26]. The main differences
135 relate to the formats and types of input and output. Kernel functions are used to map the data through
136 nonlinear functions in an n-dimensional space. In this way, it is possible to transform nonlinear
137 problems into linear problems. Research suggests that SVR presents accurate results for predicting
138 energy consumption and as such is commonly used in the field [27]. Despite its advantages, the lack of
139 predetermined heuristics for both the design and parameterisation of SVR models is a major drawback
140 in using SVR [22]. As such, studies tend to be application-specific and lack generalisability [22].

141 2.2.2. Random Forest

142 Random Forest is a machine learning technique based on different decision trees. Random Forests'
143 implementation involves the random selection of features based on the position of the root node. The
144 model output consists of the average of the results for all trees. When compared to a single decision
145 tree, Random Forest presents a better performance [28–30]. The greater the number of trees, normally
146 the better the performance of the model, but makes the model slower and inefficient for real-time
147 predictions. It is one of the most popular machine learning techniques used for classification and
148 regression problems [31–33]. Random Forests' popularity is often attributed to its higher accuracy
149 when compared with ANNs and SVR [34].

150 2.2.3. K-nearest Neighbors

151 K-nearest neighbors (KNN) is a supervised and non-parametric machine learning technique
152 [35,36]. KNN is based on the the measurement of a point's similarity to a training set with relevant
153 patterns and class labels. As the classification is made by aggregating values proximate to a K centroid,
154 the definition of the parameter, K , and an appropriate number of K neighbours is critical. The simplicity,
155 easy of use and wide applicability for solving classification and regression problems has resulted in
156 widespread adoption and use [37,38]. When using high values of N prediction stage might be slow.
157 This technique is sensitive to irrelevant features and the scale of the data.

158 2.3. Statistical Technique

159 2.3.1. ARIMA

160 The Autoregressive Integrated Moving Average (ARIMA) is a generalisation of ARMA [39,40],
161 and this is a technique applied to forecasting time series and presents good results for short term
162 forecasting [41]. The non-seasonal ARMA technique is represented as ARIMA (p, d, q), where p is the

163 order of the autoregressive, d is the degree of differencing, and q is the order of the moving average.
164 ARIMA is used in applications where the time series follows a non-stationary trend. The model applies
165 an initial stage of differentiation to eliminate non-stationary behaviour, transforming the time series
166 into a stationary one for learning [42]. Normally for ARIMA to perform well, it needs to work with a
167 long historical dataset [40].

168 3. Related work

169 Short term load forecasting using deep learning and machine learning has been examined from
170 a variety of perspectives. For example, there is a well-established literature on supply-side energy
171 consumption and demand forecasting using deep learning from the perspective of the management
172 and optimisation of power systems and electricity grids. These include studies using deep neural
173 networks [43], deep belief networks [44], CNNs [45], Autoencoder and LSTM [46], SVM and Random
174 Forest [33] amongst others. The focus of this paper is demand-side. STLF for grids and utility
175 companies have a fundamentally different motivation and context than manufacturing firms, not least
176 the public interest aspect of energy systems, as opposed to profit maximisation, operational efficiencies
177 and other business objectives.

178 Similarly, there has been a number of studies on the use of deep learning for forecasting
179 load prediction for different energy consumer types – residential [47–49], commercial [50,51], and
180 industrial [52,53]. While there are certainly learnings from these works, their focus is overwhelmingly
181 on load forecasting for utility companies and grids. Residential and commercial use cases have
182 fundamentally different energy consumption patterns than industry in terms of decision-making time
183 horizon, building code standards, population density, building design, and response to regional
184 climate, amongst others [54]. As discussed in Section 1, energy-intensive manufacturing has a
185 significantly different energy consumption profile than other industry sectors, leaving aside the obvious
186 differences with residential and commercial use. As a result, the motivation for load forecasting is
187 substantially different than other industrial use cases. In particular, these operations tend to have high
188 process-related energy requirements, are not subject to climate changes, and energy management is
189 core capability in their business. As production is central to manufacturing, the demand for energy is
190 derived from production planning and energy forecasting and optimization based on the over-riding
191 demands of production [55].

192 Ryu et al. [43] explore demand-side STLF for variety of industry categories including
193 manufacturing. Using data sourced from a Korean utility company, they propose a DNN-based
194 STLF framework based on industry category, temporal patterns, location and weather conditions.
195 A comparative analysis was performed with three different forecasting techniques - shallow neural
196 network (SNN), double seasonal Holt–Winters (DSHW) and autoregressive integrated moving average
197 (ARIMA). Using mean absolute percent error (MAPE) and relative root mean squared error (RRMSE).
198 The results suggest that the DNN-based STLF model achieved the best performance when compared
199 to the other models with lower MAPE (2.19%) and lower RRMSE (2.76%). Our approach differs in
200 three important ways. Firstly, we adopt RNNs that have significant advantages in terms of time series
201 data. Second, in [43], because the data comes from the power company, all energy consumption at a
202 firm level is bundled up and it is not possible to distinguish different sources of energy consumption
203 and their impact from within the firm e.g. buildings vs process-related consumption. Third, we focus
204 on energy-intensive manufacturing. It is not clear whether energy-intensive manufacturing is included
205 in [43].

206 Mawson & Hughes [56] explore the use of a deep feedforward neural networks (DFNN) and a
207 deep RNN (DRNN) to predict STLF for a medium-sized manufacturing facility. Inputs to the DNNs
208 included weather conditions and machining schedules. Results suggest that both models performed
209 well but that the DRNN outperformed the DFNN for predictions of building energy consumption,
210 achieving an accuracy of 96.8% compared to 92.4% for DFNN. The focus of [56] was optimising heating,
211 ventilation and air conditioning (HVAC). As such, the impact of the production process was not a focus

212 *per se*. While data for boiler energy, cooling energy and machine scheduling were taken in to account,
213 again, unlike our work, specific process-related energy consumption was not considered and the focus
214 was not energy-intensive manufacturing. Additionally, Mawson & Hughes [56] use simulated data
215 whereas we use ground truth data to train and validate the models.

216 In contrast to [43,56], Chen et al. [57] study the use of DNN for STLF in an energy-intensive
217 manufacturing use case. Using data from the melt shop of steel plant, they sought to use DNN to
218 predict energy consumption for one specific process, the electric arc furnace (EAF) for different types
219 of scrap. The performance of the DNN was compared with linear regression (LR), SVM, and decision
220 tree (DT) based on the model correlation coefficient and mean absolute error (MAE). The proposed
221 DNN outperformed other models with the highest correlation index, at 0.854, and the lowest MAE,
222 at 1.5%. While [57] is the closest use case to our paper, it focuses exclusively on one process and does
223 not seek to calculate the overall plant energy consumption. While they identify the potential of deep
224 learning over traditional statistical and machine learning approaches, they do not evaluate the relative
225 performance against other deep learning architectures.

226 Yeom & Choi [58] describe a platform, E-IoT, for collecting a wide range of data (over 1,556
227 variables for one process) at a Korean manufacturing plant. From the data collected by E-IoT, they
228 use a least absolute shrinkage and selection operator (LASSO) technique, based on machine learning,
229 to extract relevant variables to predict plant-level STLF based on the first stage of one process, using
230 LSTM. The proposed LSTM model achieved an MAE of 0.07 and an accuracy of 79%. The paper suffers
231 from a significant lack of detail. For example, while the energy consumption profile for the process
232 presented in [58] appears energy-intensive when compared to total plant energy consumption, it is
233 unclear from the paper whether the manufacturing plant was energy-intensive or not. It is also unclear
234 why only the first stage of the manufacturing process was used, and how many other processes are
235 involved. Furthermore, no detail is provided on how the LSTM model configuration was selected, and
236 it is not compared with existing techniques or other deep learning models.

237 Li et al. [33] explore STLF for industrial customers in China and source data from a cable factory
238 and a lithium factory located in Chongqing. They propose two short-term (20 day) energy consumption
239 forecasting models using SVM and Random Forest based on historical consumption data as well as
240 seasonal factors (holidays) and upstream value chain data i.e. the price of non-ferrous metals and
241 raw material consumption at each factory. Both models accurately predicted the electricity loads for
242 both factories. The mean absolute percentage error (MAPE) for both the SVM and Random Forest
243 model were similar for each factory - 5% for the cable factory and 2% for the lithium factory. The study
244 highlighted the need for research using additional industry- and firm-specific variables to increase
245 accuracy.

246 As can be seen from the above, there is a paucity of research on demand-side STLF for
247 energy-intensive manufacturing using deep learning and machine learning models. Decision making
248 for utility companies has little in common with manufacturers. Similarly, STLF for residential and
249 commercial use has little relevance to industrial use cases, and within industrial energy consumption,
250 energy intensive manufacturing is idiosyncratic. The few similar studies lack detail on the degree
251 to which they are energy-intensive manufacturers, aggregate all energy consumption, or focus on
252 one process alone. Furthermore, where proposed deep learning and machine learning models were
253 compared, they were either evaluated against only traditional techniques or only other deep learning
254 models, or not at all. We address all of these shortcomings in our paper.

255 4. Material and method

256 4.1. Dataset

257 The data used in this study was sourced from the Brazilian subsidiary of an international
258 thermoplastic resin manufacturer, an energy-intensive manufacturing plant. The plant size is c.
259 55,000 m^2 with a production capacity of approximately 500,000 tons per annum. Currently, the case site

260 calculates energy consumption forecasts manually and prepares a technical energy consumption index
261 (TECI) as a proxy for energy efficiency. Five years of data from 1 January 2015 to 31 December 2019 for
262 daily total energy consumption at the plant (ENERGY dataset), as well as process-related data for two
263 different stages of the manufacturing process - polymerization (POLY_PRODUCTION dataset) and
264 solid-state polymerization (SSPOLY_PRODUCTION dataset) were provided. Each dataset comprised
265 1,826 values. Figure 1 presents the time series of the three datasets used in this work.

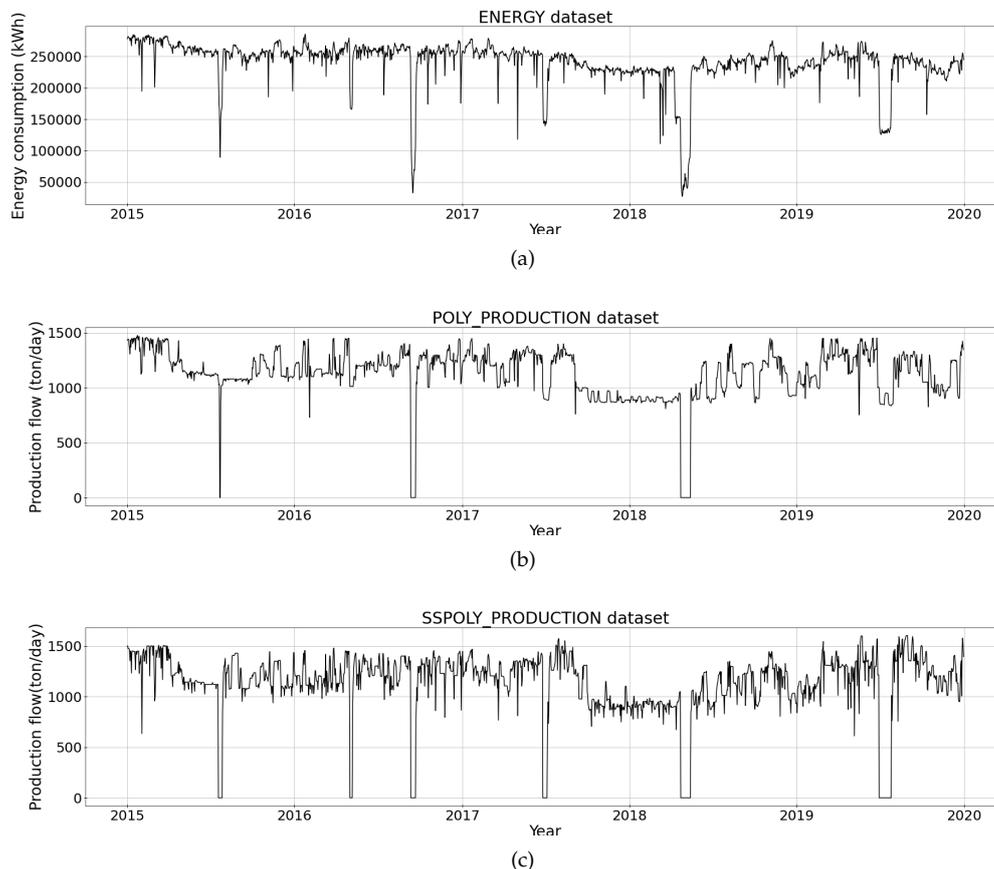


Figure 1. Daily dataset of: (a) Energy consumption, (b) Production flow for polymerization stage, and (c) Production flow for solid-state polymerization stage.

266 To identify the impact of the variations in production flow data on total plant energy consumption,
267 we performed a Pearson correlation analysis. It showed a moderate positive correlation between
268 the ENERGY dataset and the combined production flow dataset (POLY_PRODUCTION dataset and
269 SSPOLY_PRODUCTION dataset). The correlation coefficient between the ENERGY dataset and the
270 POLY_PRODUCTION dataset was 0.71, while the relationship between the ENERGY dataset and the
271 SSPOLY_PRODUCTION dataset was 0.75. Once this analysis was completed, a decision was made to
272 include all three time series (ENERGY, POLY_PRODUCTION, SSPOLY_PRODUCTION) as input in to
273 the deep learning models.

274 4.2. Data preprocessing

275 Missing values and measurement errors can lead to unpredictable results. To avoid removing
276 these, the anomalous data was filled with imputed data and then normalized. For the former, we
277 replaced missing data with the average of the data from the previous seven days as per [59–61]. We
278 then normalised the data so that all model inputs had equal weights, and the sigmoid activation
279 function could be applied in the deep learning models as per [52,62,63]. Normalization reduces the

280 data range to zero and one [0, 1]. Sklearn's MinMaxScaler function was used to normalize the data in
 281 this study, based on Equation 1.

$$X'_i = \frac{X_i - \min(x)}{\max(x) - \min(x)} \quad (1)$$

282 Where X'_i is the rescaled value; X_i is the original value; $\min(x)$ is the minimum value in feature;
 283 $\max(x)$ is the maximum value in feature.

284 4.3. Evaluation metric

285 Root mean squared error (RMSE), MAPE, and Mean Absolute Error (MAE) are the most commonly
 286 metrics used in the evaluation the accuracy of energy consumption models [64], and in particular,
 287 studies related to STLF using deep learning [43,57,58].

288 RMSE is defined as the square root of the mean squared error (MSE) [65] (Equation 2), that is,
 289 the root of the mean square error of the difference between the prediction (P_i) and the real value (R_i),
 290 where n is sample size. As RMSE is more sensitive to more significant errors (outliers) because, it
 291 squares the difference between the predicted value and the real value. RMSE presents error values in
 292 the same dimensions as the analysed variable [65]. It is widely applied in models that use time series
 293 [66].

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n |P_i - R_i|^2} \quad (2)$$

294 MAPE is widely used for evaluating prediction models particularly where the quality of the
 295 forecast is required and is used in numerous energy consumption forecasting studies [43,47,52,62,63].
 296 MAPE is defined in Equation 3 [43,52,67], and expresses the accuracy of the error as a percentage. It
 297 can be applied in a wide range of contexts, as it is relatively intuitive interpretation of relative errors
 298 however it can only be used if the values in the dataset do not equal zero [68].

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{P_i - R_i}{R_i} \right| \quad (3)$$

299 MAE is defined in Equation 4 [62,67]. Unlike the other metrics presented, the MAE depends on
 300 the scale of the data. It is not sensitive to outliers, as it treats all errors in the same way. We use it to
 301 quantify a model's ability to predict energy consumption.

$$MAE = \frac{1}{n} \sum_{i=1}^n |P_i - R_i| \quad (4)$$

302 While one or more of RMSE, MAPE, and MAE have featured in related studies [43,57], we have
 303 chosen to measure all of them for better comprehensiveness and analysis of different aspects of what is
 304 being studied [33].

305 5. Finding models to predict energy consumption

306 Due to the non-linear characteristics of the datasets used in this research, the need for both
 307 accuracy and fast run-times, and the promising results obtained in other works that used deep learning
 308 [47,49,62,69], three deep learning techniques were selected for STLF in this study - simple RNN, LSTM,
 309 and GRU. In addition to these techniques, three different machine learning techniques were selected
 310 for the purpose of comparison - SVR, Random Forest, and KNN. These were selected as they feature in
 311 related works on STLF for demand side energy consumption [43,57].

312 To determine the most suitable configuration for each selected architecture (within a specific
 313 predefined range), we use the grid search method to determine hyperparameters for learning

314 algorithms [70–74]. It is used widely as it is quick to implement, trivial to parallelize, and intuitively
 315 allows an entire search space to be explored [75].

316 5.1. Deep Learning

317 To perform the grid search, the dataset was separated into a training set consisting of 80% of
 318 the original dataset (from 1 January 2015 to 30 December 2018), and a test set comprising 20% of the
 319 original dataset (from 31 December 2018 to 31 December 2019). The hyperparameters evaluated by the
 320 grid search for deep learning techniques were (i) the number of layers, and (ii) the number of nodes in
 321 each layer (see Table 1).

Table 1. Parameter and levels of deep learning technique.

Parameter	Levels
Number of nodes	From 10 to 90, step 20
Number of layers	From 1 to 4, step 1

322 For deep learning models the following parameters were fixed: 100 epochs, batch size of 16,
 323 a Sigmoid [76] as the activation function, MSE as the loss function, and a method for stochastic
 324 optimization (Adam) as the optimizer. These parameters were chosen empirically through a previous
 325 analysis. Due to the stochastic nature of the optimization process, the grid search was performed 30
 326 times, and the average of the RMSE, MAPE, and MAE were calculated. Figures 2, 3 and 4 present
 327 the grid search normalized results for RMSE, MAPE, and MAE, respectively. The configuration with
 328 one layer and 30 nodes for deep learning models generated the best models - RNN-1-30, LSTM-1-30,
 329 and GRU-1-30 - for all deep learning techniques, except in one instance where the average MAPE
 330 presented the best result for RNN with a configuration of four layers and 30 nodes i.e. RNN-4-30.

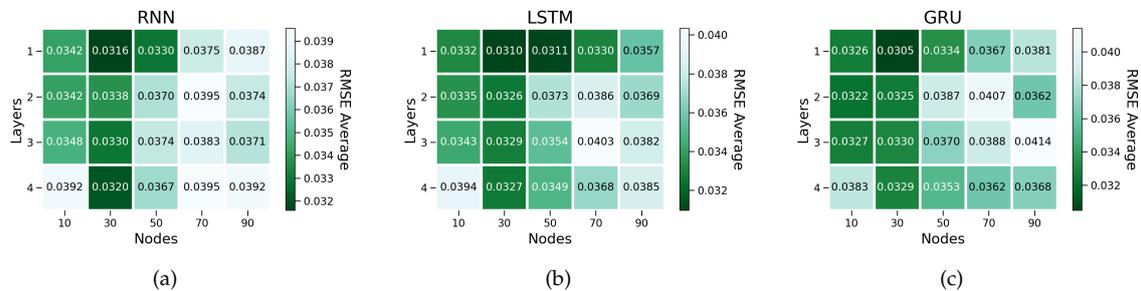


Figure 2. RMSE grid search result for (a) RNN, (b) LSTM, and (c) GRU.

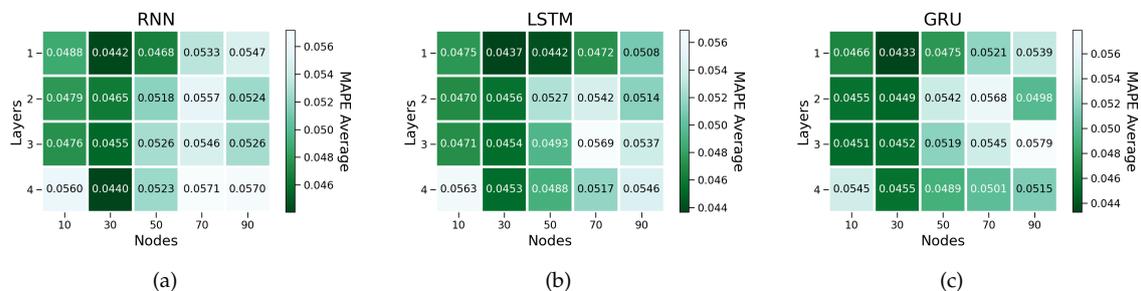


Figure 3. MAPE grid search result for (a) RNN, (b) LSTM, and (c) GRU.

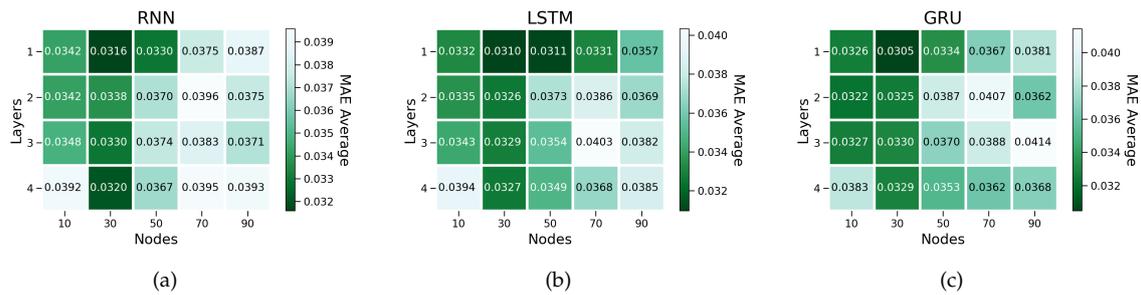


Figure 4. MAE grid search result for (a) RNN, (b) LSTM, and (c) GRU.

331 5.2. Machine Learning

332 To perform the grid search for the machine learning techniques, the dataset was separated as per
 333 the deep learning models with a training set comprising 80% of the original dataset, and a test set
 334 comprising 20% of the original dataset. The hyperparameters used vary according to the technique
 335 (see Table 2). For SVR, cost and the type of kernel were used whereas the maximum depth of trees and
 336 the number of trees were used for Random Forest. For KNN, the K parameter was used.

Table 2. Machine learning parameters and levels.

Technique	Parameter	Levels
SVR	Number of C	0.1, 1 and 10
SVR	Type of kernel	Linear, polinomial and RBF
Random Forest	Number of max. depth	From 3 to 6, step 1
Random Forest	Number of trees	From 50 to 200, step 50
KNN	Number of K	From 1 to 9, step 2

337 Figures 5, 6 and 7 present the grid search results for RMSE, MAPE, and MAE, respectively.

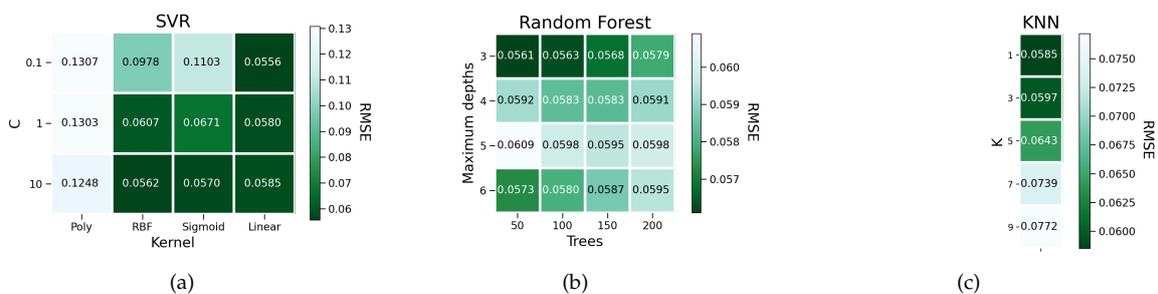


Figure 5. RMSE grid search result for (a) SVR, (b) Random Forest, and (c) KNN.

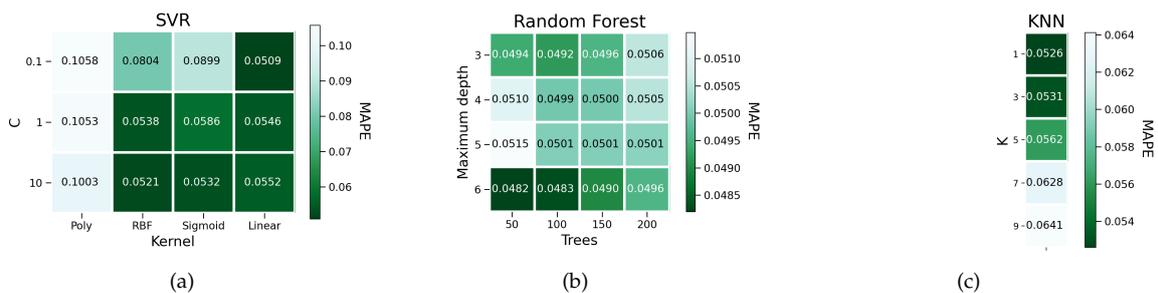


Figure 6. MAPE grid search result for (a) SVR, (b) Random Forest, and (c) KNN.

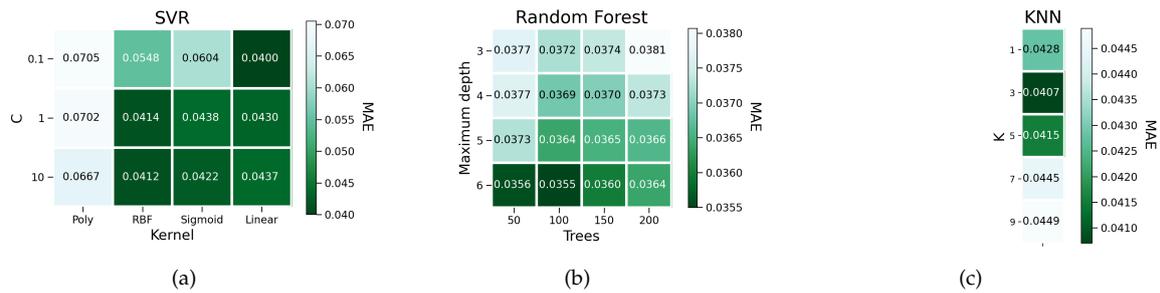


Figure 7. MAE grid search result for (a) SVR, (b) Random Forest, and (c) KNN.

338 For SVR, the best configuration across the three metrics used - RMSE, MAPE, and MAE - is
 339 represented by SVR-0.1-linear, whose C value is 0.1 and used the linear kernel. For Random Forest, the
 340 models with configurations with (a) a maximum depth of three with 50 trees (Random Forest-3-50),
 341 (b) maximum depth of six with 50 trees (Random Forest-6-50), and (c) a maximum depth of six with
 342 100 trees (Random Forest-6-100), generated the best results for RMSE, MAPE, and MAE, respectively.
 343 For KNN, the configuration with K equals 1 (KNN-1) presented the best result for RMSE and MAPE.
 344 Based on MAE, the best configuration is where K equals 3 (KNN-3).

345 5.3. Benchmarks

346 In addition to comparing the performance of deep learning and machine learning models, two
 347 additional benchmarks were selected for comparison purposes. The first benchmark is the manual
 348 technique used by the case site providing the dataset for this study. The second benchmark is an
 349 ARIMA model. ARIMA was selected because of its widespread use in energy forecasting and in
 350 particular in related works [43].

351 The manual technique used by the case site is performed by a simple calculation as per Equation
 352 5, where the energy consumption of a given day, $C_{predicted}$, is the planned production flow ($F_{planned}$)
 353 and the TECI ($n_{previous}$) based on measured from data collected on the previous day.

$$C_{predicted} = F_{planned} \cdot n_{previous} \quad (5)$$

354 With respect to the ARIMA model, the order of the AR ($p = 1$), is the degree of differences ($d = 0$),
 355 and the order of MA ($q = 1$). For the ARIMA models, the parameters were empirically selected.

356 6. Results and discussions

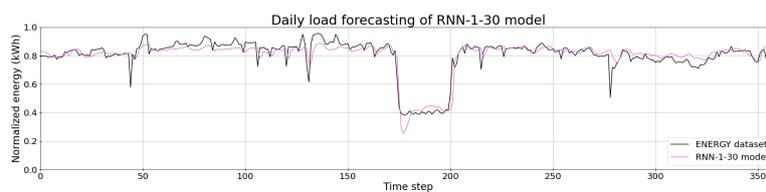
357 Table 3 presents the RMSE, MAPE, and MAE results for the four deep learning models (RNN-1-30,
 358 RNN-4-30, LSTM-1-30, and GRU-1-30), and six machine learning models (SVR-0.1-linear, Random
 359 Forest-3-50, Random Forest-6-50, Random Forest-3-100, KNN-1, and KNN-3) identified by the grid
 360 search method, as well as the manual and ARIMA benchmarks.

361 Based on the RMSE metric, the deep learning models outperformed the machine learning models,
 362 and the manual and ARIMA benchmarks. This behavior can be explained by the ability of deep
 363 learning models have to achieve insights outside of the domain of training data. The GRU model
 364 presented the best performance of all models tested as well as reducing the complexity inherent in the
 365 other deep learning models analyzed; the simple RNN models presented the worst performance. In
 366 contrast, based on MAPE and MAE, the ARIMA model outperformed the deep learning models, the
 367 machine learning models, and the legacy manual approach. Although ARIMA presents good results,
 368 its performance will not improve over time and the processing time is high, unlike deep learning
 369 models, being a significant limitation for practical use.

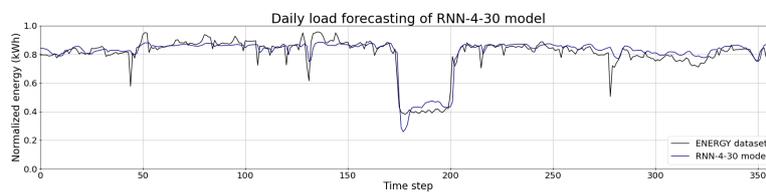
370 Figures 8, 9, and 10 illustrates the daily load forecasts for the deep learning and machine learning
 371 models, the manual and ARIMA benchmarks compared against the ground truth data. These clearly

Table 3. RMSE, MAPE, and MAE results for selected deep learning models, machine learning models, and benchmarks.

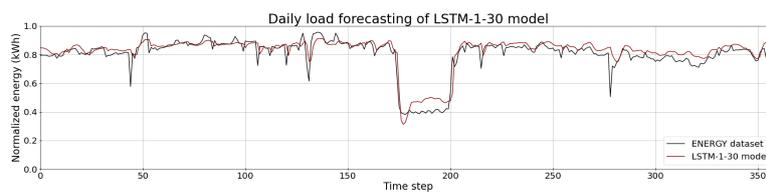
Models	RMSE	MAPE	MAE
ARIMA	0.0471	0.0352	0.0249
RNN-1-30	0.0316	0.0442	0.0316
RNN-4-30	0.0320	0.0440	0.0320
LSTM-1-30	0.0310	0.0437	0.0310
GRU-1-30	0.0305	0.0433	0.0305
SVR-0.1-linear	0.0556	0.0509	0.0400
Random Forest-3-50	0.0561	0.0494	0.0377
Random Forest-6-50	0.0573	0.0482	0.0356
Random Forest-6-100	0.0580	0.0483	0.0355
KNN-1	0.0585	0.0526	0.0428
KNN-3	0.0597	0.0531	0.0407
Case site technique	0.4119	0.5161	0.4039



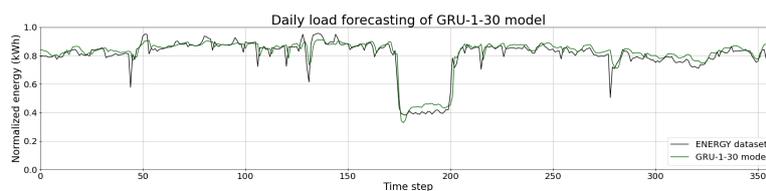
(a)



(b)



(c)



(d)

Figure 8. Daily load forecasting using (a) RNN-1-30 model, (b) RNN-4-30 model, (c) LSTM-1-30 model, and (d) GRU-1-30 model.

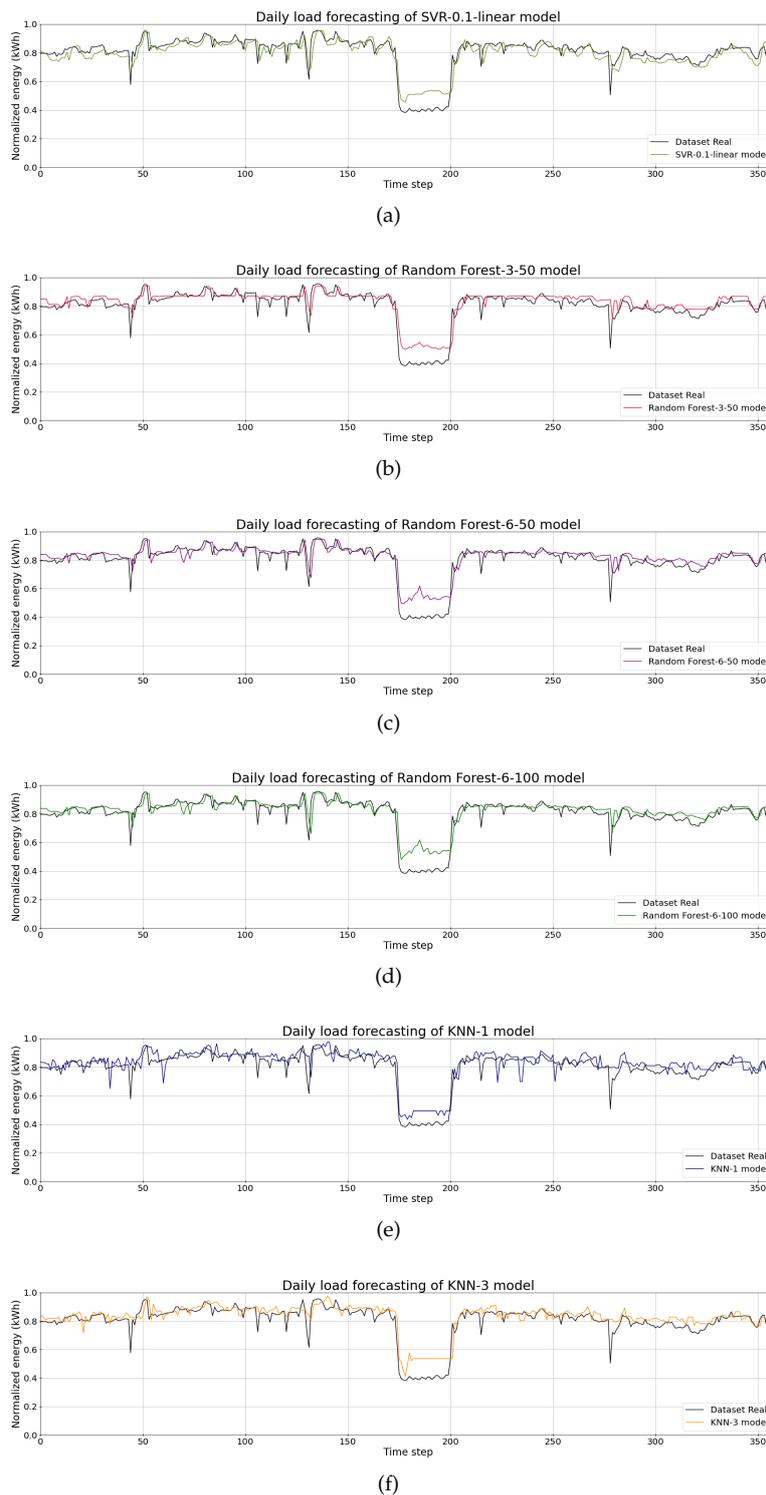


Figure 9. Daily load forecasting using (a) SVR-0.1-linear model, (b) Random Forest-3-50 model, (c) Random Forest-6-50 model, (d) Random Forest-6-100 model, (e) KNN-1 model, and (f) KNN-3 model.

372 show that the proposed deep learning models (Figure 8 (a-d)) are very similar to the ground truth data
 373 compared to the existing manual technique used at the case site (Figure 10 (b)).

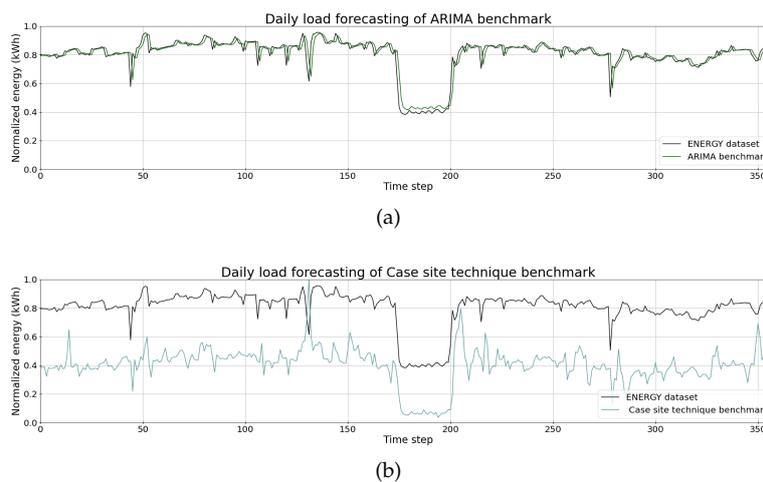


Figure 10. Daily load forecasting using (a) case site technique, and (b) ARIMA model.

374 6.1. Diebold-Mariano statistical test

375 Table 3 suggests that GRU-1-30 and ARIMA achieved the best results for RMSE, MAPE, and
 376 MAE metrics, respectively. As the values of the RMSE, MAPE, and MAE are very similar, we used
 377 the Diebold-Mariano [77] test to confirm the results. The Diebold-Mariano test is a hypothesis test of
 378 two samples used compare the predictions of two predicted time series. Table 4 presents the results
 379 obtained.

Table 4. Two-sample Diebold-Mariano test results for the deep learning and machine learning models, the manual and ARIMA approaches.

	ARIMA	RNN-1-30	RNN-4-30	LSTM-1-30	GRU-1-30	SVR-0.1-linear	Random Forest-3-50	Random Forest-6-50	Random Forest-6-100	KNN-1	KNN-3
Case site technique	51.38	51.55	51.27	50.79	51.30	51.06	50.20	49.76	49.70	50.32	50.13
ARIMA	-	-0.54	-0.83	-1.46	0.45	-2.96	-3.36	-2.95	-3.18	-3.25	-3.73
RNN-1-30	-	-	-1.39	-1.34	1.99	-2.43	-2.76	-2.41	-2.56	-2.69	-3.74
RNN-4-30	-	-	-	-0.99	3.04	-2.07	-2.57	-2.25	-2.40	-2.55	-3.73
LSTM-1-30	-	-	-	-	6.56	-1.96	-2.81	-2.37	-2.54	-2.66	-4.12
GRU-1-30	-	-	-	-	-	-3.87	-4.47	-3.61	-3.71	-4.10	-5.44
SVR-0.1-linear	-	-	-	-	-	-	-0.24	-0.78	-1.08	-0.99	-1.79
Random Forest-3-50	-	-	-	-	-	-	-	-0.76	-1.15	-0.90	-1.96
Random Forest-6-50	-	-	-	-	-	-	-	-	-1.83	-0.42	-1.15
Random Forest-6-100	-	-	-	-	-	-	-	-	-	-0.19	-0.81
KNN-1	-	-	-	-	-	-	-	-	-	-	-0.50

380 The Diebold-Mariano test result equals zero when the techniques being tested are equal; negative
 381 values present when the left technique obtains a better performance and vice-versa. If the absolute
 382 Diebold-Mariano result values are high, the tested techniques have significantly different prediction
 383 values. The first line of Table 4 shows the comparison between the case site technique compared to all
 384 other models.

385 It is clear that the existing manual technique used at the case site has the worst performance in
 386 comparison to the all models examined. The high statistical values obtained for this technique confirms
 387 that it is sub-optimal for STLF in this case. The only model that outperformed the ARIMA model was
 388 the GRU-1-30 model. All deep learning models outperform the machine learning models. However
 389 the variance in Diebold-Mariano values are not as significant. While the Diebold-Mariano test result
 390 values for deep learning models are similar, the GRU-1-30 model achieved the best prediction indexes
 391 when compared to all models tested. As such, the initial hypothesis from the grid search results are
 392 confirmed.

393 These results suggest a significant improvement in the accuracy of the STLF for this
 394 energy-intensive manufacturer. This can be used to provide more accurate energy management
 395 to meet production demands, improve cashflow, reduce environmental impact, and mitigate risks
 396 associated with energy inefficiencies. Accurate STLF results can be used for anticipatory optimisation

397 and remediation. For example, anomaly detection can be used to identify possible machine degradation
398 or failure from anomalous loads at different stages in the manufacturing process. This would enable
399 predictive maintenance and avoid production downtime.

400 7. Conclusion and future work

401 This paper is one of the first papers to compare the efficacy of deep learning and machine learning
402 models for short term load forecasting for energy-intensive manufacturing plants. In addition, we
403 benchmark these models against the incumbent manual prediction technique, and a classic time
404 series forecasting technique, ARIMA. Unlike existing studies, we consider multi-year ground truth
405 data including total plant energy consumption data, and data from two stages in a complex energy
406 intensive manufacturing process. The use of production data contributed significantly to improved
407 STLTF accuracy by reducing the RMSE.

408 Based on grid search results and Diebold-Mariano test results, we found that the all deep learning
409 and machine learning models outperformed the incumbent manual technique. Furthermore, GRU
410 (GRU-1-30) outperformed the basic RNN and LSTM models based, achieving the best performance
411 for RMSE (0.0305), MAPE (0.0433), and MAE (0.0305). The model that predicted short term industrial
412 energy consumption the best was the GRU-1-30.

413 Accurate STLTF can be used in a variety manufacturing processes to achieve energy efficiencies
414 and can be used as an input in to a range of operational decisions including energy management
415 (e.g. heat storage and cooling), anomaly detection, predictive machine maintenance, and proactive
416 plant and machine management amongst others. The reduction of machine idle-times would seem
417 to be particularly attractive to such manufacturers. Given the dearth of research on this topic in
418 energy-intensive manufacturing, there are many avenues for future research. As the Industrial Internet
419 of Things matures, a significantly larger volume of time series data will be available to further refine
420 the accuracy of the models but also extend the use of deep learning beyond prediction to actuation.

421 For near real-time prediction, very short-term load forecasting (VSTLTF) may be needed. In such
422 use cases, rapid training times will be required. While GRUs may meet this criteria, further research
423 is required. Furthermore, medium-term load forecasting may prove fruitful, deep learning training
424 models may need to be augmented with historic trend data to account for longer seasonal cycles or
425 predictable events. Medium-term load forecasting may enable new use cases including switches to
426 more sustainable or lower cost power supply. Similarly, as production planning is prioritised over
427 energy management in energy-intensive manufacturing, multi-step forecasting strategy may be more
428 appropriate or preferable. This may require ensemble solutions and is worthy of exploration.

429 This paper highlights the potential of deep learning and ARIMA in energy-intensive
430 manufacturing. The adoption of deep learning, like all data science technologies, requires overcoming
431 human, organisational and technological challenges however against intense rivalry, firms may not
432 have a choice.

433 **Acknowledgments:** This research has been partially financial supported by Fundação de Amparo à Ciência e
434 Tecnologia de Pernambuco (FACEPE) and by Irish Institute of Digital Business.

435 Abbreviations

436 The following abbreviations were used in this manuscript:

ANN	Artificial Neural Networks
ARIMA	Autoregressive Integrated Moving Average
Btu	British Thermal Units
CNN	Convolutional Neural Network
CO ₂	Carbon Dioxide
D-M	Diebold-Mariano
DFNN	Deep Feed Forward Neural Networks
DNN	Deep Neural Network
DRNN	Deep Recurrent Neural Networks
DSHW	Double Seasonal Holt–Winters
DT	Decision Tree
EAF	Electric Arc Furnace
ECM	Environmentally Conscious Manufacturing
ERM	Environmentally Responsible Manufacturing
GHG	Greenhouse Gas
GPU	Graphic Processing Unit
GRU	Gated Recurrent Unit
HVAC	Heating, Ventilation and Air Conditioning
437 KNN	K-Nearest Neighbor
LASSO	Least Absolute Shrinkage and Selection Operator
LR	Linear Regression
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MAPE	Mean Absolute Percent Error
MLP	Multi-Layer Perceptrons
MSE	Mean Squared Error
MTLF	Medium-Term Load Forecasting
RMSE	Root Mean Squared Error
RNN	Recurrent Neural Networks
RRMSE	Relative Root Mean Square Error
SDG	Sustainable Development Goals
SNN	Shallow Neural Network
STLF	Short-Term Load Forecasting
SVM	Support Vector Machines
SVR	Support Vector Machine Regression
TECI	Technical Energy Consumption Index
VSTLF	Very Short-Term Load Forecasting

438 References

- 439 1. EIA. International Energy Outlook 2019, 2019. <https://www.eia.gov/outlooks/ieo/>, Last accessed on
440 2020-04-24.
- 441 2. EIA. *International Energy Outlook 2016, with Projections to 2040*; Government Printing Office, 2016.
- 442 3. Gozgor, G.; Lau, C.K.M.; Lu, Z. Energy consumption and economic growth: New evidence from the OECD
443 countries. *Energy* **2018**, *153*, 27–34.
- 444 4. IEA. Global Energy & CO₂ Status Report 2019, 2019. [https://www.iea.org/reports/global-energy-co2-
445 status-report-2019/emissions](https://www.iea.org/reports/global-energy-co2-status-report-2019/emissions), Last accessed on 2020-04-24.
- 446 5. De Marchi, V. Environmental innovation and R&D cooperation: Empirical evidence from Spanish
447 manufacturing firms. *Research policy* **2012**, *41*, 614–623.
- 448 6. Curkovic, S. Environmentally responsible manufacturing: the development and validation of a
449 measurement model. *European Journal of Operational Research* **2003**, *146*, 130–155.
- 450 7. Johannson, L. The power of IT: How can information technology support TQEM? *Environmental Quality
451 Management* **1993**, *2*, 339–346.

- 452 8. Sundarakani, B.; De Souza, R.; Goh, M.; Wagner, S.M.; Manikandan, S. Modeling carbon footprints across
453 the supply chain. *International Journal of Production Economics* **2010**, *128*, 43–50.
- 454 9. Cramer, J.; Roes, B. Total employee involvement: measures for success. *Environmental Quality Management*
455 **1993**, *3*, 39–52.
- 456 10. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural networks* **2015**, *61*, 85–117.
- 457 11. Zhiqiang, W.; Jun, L. A review of object detection based on convolutional neural network. 2017 36th
458 Chinese Control Conference (CCC). IEEE, 2017, pp. 11104–11109.
- 459 12. Young, T.; Hazarika, D.; Poria, S.; Cambria, E. Recent trends in deep learning based natural language
460 processing. *IEEE Computational Intelligence Magazine* **2018**, *13*, 55–75.
- 461 13. Xie, D.; Zhang, L.; Bai, L. Deep learning in visual computing and signal processing. *Applied Computational*
462 *Intelligence and Soft Computing* **2017**, 2017.
- 463 14. LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *Nature* **2015**, *521*, 436–444.
- 464 15. Kraus, M.; Feuerriegel, S.; Oztekin, A. Deep learning in business analytics and operations research: Models,
465 applications and managerial implications. *European Journal of Operational Research* **2020**, *281*, 628–641.
- 466 16. Bengio, Y.; Simard, P.; Frasconi, P. Learning long-term dependencies with gradient descent is difficult.
467 *IEEE transactions on neural networks* **1994**, *5*, 157–166.
- 468 17. Chung, J.; Gulcehre, C.; Cho, K.; Bengio, Y. Empirical evaluation of gated recurrent neural networks on
469 sequence modeling. *arXiv preprint arXiv:1412.3555* **2014**.
- 470 18. Sundermeyer, M.; Schlüter, R.; Ney, H. LSTM neural networks for language modeling. Thirteenth annual
471 conference of the international speech communication association, 2012.
- 472 19. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural computation* **1997**, *9*, 1735–1780.
- 473 20. Gers, F.A.; Schmidhuber, J.; Cummins, F. Learning to forget: continual prediction with LSTM. 1999 Ninth
474 International Conference on Artificial Neural Networks ICANN 99. (Conf. Publ. No. 470), 1999, Vol. 2, pp.
475 850–855 vol.2. doi:10.1049/cp:19991218.
- 476 21. Jozefowicz, R.; Zaremba, W.; Sutskever, I. An empirical exploration of recurrent network architectures.
477 International conference on machine learning, 2015, pp. 2342–2350.
- 478 22. Sapankevych, N.I.; Sankar, R. Time series prediction using support vector machines: a survey. *IEEE*
479 *Computational Intelligence Magazine* **2009**, *4*, 24–38.
- 480 23. Vapnik, V. The nature of statistical learning theory (p. 189). *New York: Springer-Verlag*. doi **1995**, *10*, 978–1.
- 481 24. Müller, K.R.; Smola, A.J.; Rätsch, G.; Schölkopf, B.; Kohlmorgen, J.; Vapnik, V. Predicting time series with
482 support vector machines. International Conference on Artificial Neural Networks. Springer, 1997, pp.
483 999–1004.
- 484 25. Simon, H. *Neural networks: a comprehensive foundation*; Prentice hall, 1999.
- 485 26. Drucker, H.; Burges, C.J.; Kaufman, L.; Smola, A.J.; Vapnik, V. Support vector regression machines.
486 Advances in neural information processing systems, 1997, pp. 155–161.
- 487 27. Golkarnarenji, G.; Naebe, M.; Badii, K.; Milani, A.S.; Jazar, R.N.; Khayyam, H. Support vector regression
488 modelling and optimization of energy consumption in carbon fiber production line. *Computers & Chemical*
489 *Engineering* **2018**, *109*, 276–288.
- 490 28. Breiman, L. Random forests. *Machine learning* **2001**, *45*, 5–32.
- 491 29. Cutler, D.R.; Edwards Jr, T.C.; Beard, K.H.; Cutler, A.; Hess, K.T.; Gibson, J.; Lawler, J.J. Random forests for
492 classification in ecology. *Ecology* **2007**, *88*, 2783–2792.
- 493 30. Liaw, A.; Wiener, M.; others. Classification and regression by randomForest. *R news* **2002**, *2*, 18–22.
- 494 31. Lahouar, A.; Slama, J.B.H. Day-ahead load forecast using random forest and expert input selection. *Energy*
495 *Conversion and Management* **2015**, *103*, 1040–1051.
- 496 32. Li, C.; Tao, Y.; Ao, W.; Yang, S.; Bai, Y. Improving forecasting accuracy of daily enterprise electricity
497 consumption using a random forest based on ensemble empirical mode decomposition. *Energy* **2018**,
498 *165*, 1220–1227.
- 499 33. Li, Q.; Zhang, L.; Xiang, F. Short-term Load Forecasting: A Case Study in Chongqing Factories. 2019
500 6th International Conference on Information Science and Control Engineering (ICISCE). IEEE, 2019, pp.
501 892–897.
- 502 34. Caruana, R.; Karampatziakis, N.; Yessenalina, A. An empirical evaluation of supervised learning in high
503 dimensions. Proceedings of the 25th international conference on Machine learning, 2008, pp. 96–103.

- 504 35. Loftsgaarden, D.O.; Quesenberry, C.P.; others. A nonparametric estimate of a multivariate density function. *The Annals of Mathematical Statistics* **1965**, *36*, 1049–1051.
- 505
- 506 36. Dritsas, E.; Kanavos, A.; Trigka, M.; Vonitsanos, G.; Sioutas, S.; Tsakalidis, A. Trajectory Clustering and k-NN for Robust Privacy Preserving k-NN Query Processing in GeoSpark. *Algorithms* **2020**, *13*, 182.
- 507
- 508 37. Zhang, S.; Li, X.; Zong, M.; Zhu, X.; Cheng, D. Learning k for knn classification. *ACM Transactions on Intelligent Systems and Technology (TIST)* **2017**, *8*, 1–19.
- 509
- 510 38. Altman, N.S. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician* **1992**, *46*, 175–185.
- 511
- 512 39. Nason, G.P. Stationary and non-stationary time series. *Statistics in Volcanology. Special Publications of IAVCEI* **2006**, *1*, 000–000.
- 513
- 514 40. Sen, P.; Roy, M.; Pal, P. Application of ARIMA for forecasting energy consumption and GHG emission: A case study of an Indian pig iron manufacturing organization. *Energy* **2016**, *116*, 1031–1038.
- 515
- 516 41. Al-Musaylh, M.S.; Deo, R.C.; Adamowski, J.F.; Li, Y. Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia. *Advanced Engineering Informatics* **2018**, *35*, 1–16.
- 517
- 518
- 519 42. Yuan, C.; Liu, S.; Fang, Z. Comparison of China's primary energy consumption forecasting by using ARIMA (the autoregressive integrated moving average) model and GM (1, 1) model. *Energy* **2016**, *100*, 384–390.
- 520
- 521 43. Ryu, S.; Noh, J.; Kim, H. Deep neural network based demand side short term load forecasting. *Energies* **2017**, *10*, 3.
- 522
- 523 44. Qiu, X.; Ren, Y.; Suganthan, P.N.; Amaratunga, G.A. Empirical mode decomposition based ensemble deep learning for load demand time series forecasting. *Applied Soft Computing* **2017**, *54*, 246–255.
- 524
- 525 45. Dong, X.; Qian, L.; Huang, L. Short-term load forecasting in smart grid: A combined CNN and K-means clustering approach. 2017 IEEE International Conference on Big Data and Smart Computing (BigComp). IEEE, 2017, pp. 119–125.
- 526
- 527
- 528 46. Gensler, A.; Henze, J.; Sick, B.; Raabe, N. Deep Learning for solar power forecasting—An approach using AutoEncoder and LSTM Neural Networks. 2016 IEEE international conference on systems, man, and cybernetics (SMC). IEEE, 2016, pp. 002858–002865.
- 529
- 530
- 531 47. Kong, W.; Dong, Z.Y.; Jia, Y.; Hill, D.J.; Xu, Y.; Zhang, Y. Short-term residential load forecasting based on LSTM recurrent neural network. *IEEE Transactions on Smart Grid* **2017**, *10*, 841–851.
- 532
- 533 48. Shi, H.; Xu, M.; Li, R. Deep learning for household load forecasting—A novel pooling deep RNN. *IEEE Transactions on Smart Grid* **2017**, *9*, 5271–5280.
- 534
- 535 49. Güngör, O.; Akşanlı, B.; Aydoğan, R. Algorithm selection and combining multiple learners for residential energy prediction. *Future Generation Computer Systems* **2019**, *99*, 391–400.
- 536
- 537 50. Xypolytou, E.; Meisel, M.; Sauter, T. Short-term electricity consumption forecast with artificial neural networks—A case study of office buildings. 2017 IEEE Manchester PowerTech. IEEE, 2017, pp. 1–6.
- 538
- 539 51. Petri, I.; Li, H.; Rezgui, Y.; Chunfeng, Y.; Yuce, B.; Jayan, B. A modular optimisation model for reducing energy consumption in large scale building facilities. *Renewable and Sustainable Energy Reviews* **2014**, *38*, 990–1002.
- 540
- 541
- 542 52. Azadeh, A.; Ghaderi, S.; Sohrabkhani, S. Annual electricity consumption forecasting by neural network in high energy consuming industrial sectors. *Energy Conversion and Management* **2008**, *49*, 2272–2278.
- 543
- 544 53. Olanrewaju, O.A. Predicting Industrial Sector's Energy Consumption: Application of Support Vector Machine. 2019 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM). IEEE, 2019, pp. 1597–1600.
- 545
- 546
- 547 54. Hobby, J.D.; Tucci, G.H. Analysis of the residential, commercial and industrial electricity consumption. 2011 IEEE PES Innovative Smart Grid Technologies. IEEE, 2011, pp. 1–7.
- 548
- 549 55. Hadera, H.; Labrik, R.; Mäntysaari, J.; Sand, G.; Harjunkoski, I.; Engell, S. Integration of energy-cost optimization and production scheduling using multiparametric programming. In *Computer aided chemical engineering*; Elsevier, 2016; Vol. 38, pp. 559–564.
- 550
- 551
- 552 56. Mawson, V.J.; Hughes, B.R. Deep Learning techniques for energy forecasting and condition monitoring in the manufacturing sector. *Energy and Buildings* **2020**, p. 109966.
- 553
- 554 57. Chen, C.; Liu, Y.; Kumar, M.; Qin, J. Energy consumption modelling using deep learning technique—a case study of EAF. *Procedia CIRP* **2018**, *72*, 1063–1068.
- 555

- 556 58. Yeom, K.R.; Choi, H.S. Prediction of Manufacturing Plant's Electric Power Using Machine Learning. 2018
557 Tenth International Conference on Ubiquitous and Future Networks (ICUFN). IEEE, 2018, pp. 814–816.
- 558 59. Demirhan, H.; Renwick, Z. Missing value imputation for short to mid-term horizontal solar irradiance
559 data. *Applied Energy* **2018**, *225*, 998–1012.
- 560 60. Andiojaya, A.; Demirhan, H. A bagging algorithm for the imputation of missing values in time series.
561 *Expert Systems with Applications* **2019**, *129*, 10–26.
- 562 61. Peppanen, J.; Zhang, X.; Grijalva, S.; Reno, M.J. Handling bad or missing smart meter data through
563 advanced data imputation. 2016 IEEE Power & Energy Society Innovative Smart Grid Technologies
564 Conference (ISGT). IEEE, 2016, pp. 1–5.
- 565 62. Berriel, R.F.; Lopes, A.T.; Rodrigues, A.; Varejao, F.M.; Oliveira-Santos, T. Monthly energy consumption
566 forecast: A deep learning approach. 2017 International Joint Conference on Neural Networks (IJCNN).
567 IEEE, 2017, pp. 4283–4290.
- 568 63. Kuo, P.H.; Huang, C.J. A high precision artificial neural networks model for short-term energy load
569 forecasting. *Energies* **2018**, *11*, 213.
- 570 64. Debnath, K.B.; Mourshed, M. Forecasting methods in energy planning models. *Renewable and Sustainable*
571 *Energy Reviews* **2018**, *88*, 297–325.
- 572 65. Willmott, C.J.; Matsuura, K. Advantages of the mean absolute error (MAE) over the root mean square error
573 (RMSE) in assessing average model performance. *Climate research* **2005**, *30*, 79–82.
- 574 66. Kolomvatsos, K.; Papadopoulou, P.; Anagnostopoulos, C.; Hadjiefthymiades, S. A Spatio-Temporal Data
575 Imputation Model for Supporting Analytics at the Edge. Conference on e-Business, e-Services and e-Society.
576 Springer, 2019, pp. 138–150.
- 577 67. Hsieh, T.J.; Hsiao, H.F.; Yeh, W.C. Forecasting stock markets using wavelet transforms and recurrent neural
578 networks: An integrated system based on artificial bee colony algorithm. *Applied soft computing* **2011**,
579 *11*, 2510–2525.
- 580 68. De Myttenaere, A.; Golden, B.; Le Grand, B.; Rossi, F. Mean absolute percentage error for regression
581 models. *Neurocomputing* **2016**, *192*, 38–48.
- 582 69. Lago, J.; De Ridder, F.; De Schutter, B. Forecasting spot electricity prices: Deep learning approaches and
583 empirical comparison of traditional algorithms. *Applied Energy* **2018**, *221*, 386–405.
- 584 70. Bergstra, J.; Bengio, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning*
585 *Research* **2012**, *13*, 281–305.
- 586 71. Liao, J.M.; Chang, M.J.; Chang, L.M. Prediction of Air-Conditioning Energy Consumption in R&D Building
587 Using Multiple Machine Learning Techniques. *Energies* **2020**, *13*, 1847.
- 588 72. Yoon, H.; Kim, Y.; Ha, K.; Lee, S.H.; Kim, G.P. Comparative evaluation of ANN-and SVM-time series
589 models for predicting freshwater-saltwater interface fluctuations. *Water* **2017**, *9*, 323.
- 590 73. Kavaklioglu, K. Modeling and prediction of Turkey's electricity consumption using Support Vector
591 Regression. *Applied Energy* **2011**, *88*, 368–375.
- 592 74. Samsudin, R.; Shabri, A.; Saad, P. A comparison of time series forecasting using support vector machine
593 and artificial neural network model. *Journal of applied sciences* **2010**, *10*, 950–958.
- 594 75. Young, S.R.; Rose, D.C.; Karnowski, T.P.; Lim, S.H.; Patton, R.M. Optimizing deep learning
595 hyper-parameters through an evolutionary algorithm. Proceedings of the Workshop on Machine Learning
596 in High-Performance Computing Environments, 2015, pp. 1–5.
- 597 76. Han, J.; Moraga, C. The influence of the sigmoid function parameters on the speed of backpropagation
598 learning. From Natural to Artificial Neural Computation; Mira, J.; Sandoval, F., Eds.; Springer Berlin
599 Heidelberg: Berlin, Heidelberg, 1995; pp. 195–201.
- 600 77. Diebold, F.X. Comparing predictive accuracy, twenty years later: A personal perspective on the use and
601 abuse of Diebold–Mariano tests. *Journal of Business & Economic Statistics* **2015**, *33*, 1–1.