The International Journal of Advanced Manufacturing Technology (2026) 142:431–441
https://doi.org/10.1007/s00170-025-17052-y

ORIGINAL ARTICLE

Monitoring and prediction of the process energy in multi-stage cold forging using recurrent and self-attention based neural networks

Papdo Tchasse[1] · **Mathias Liewald**[1]

Received: 24 January 2025 / Accepted: 15 November 2025 / Published online: 4 December 2025

Abstract

The effects of greenhouse gases on the global warming and the natural limitations of manufacturing resources oblige the cold forging manufacturers to pay more attention to their energy consumption. Although the material production entails the higher energy demand for the part manufacturing, understanding the energy requirement of single processes in a press shop still remains crucial for fully determining the life cycle assessment of cold forging parts. The application of classical energy measurement tools to monitor cold forging processes is rather challenging due to the complexity of the manufacturing environment and the resulting expense for the measuring equipment. For this reason, this study explored the processing of the forming forces for the real time monitoring and prediction of the required forming energy. For this purpose, a two-stage cold forging process was considered and two tasks were designed. The first task was the prediction of the cumulated energy during the forming stroke based on the recent in-stroke energy history and the second task focused on understanding the effects of one forming stroke on the energy evolution of the future ones. For both tasks, the performance of recurrent and self-attention based neural networks was compared and the results showed high predictions accuracy for the different model architectures. Being able to predict the forming energy would allow to identify the process variations due to the fluctuation of the part geometry, material properties and surface quality, the tool wear state and the machine downtimes.

Keywords Cold forging · Energy prediction · Soft sensor · Deep learning · Sustainability

1 Introduction

Understanding the energy load of the process is an essential prerequisite for cold forging manufacturers in order to establish their CO2-footprint [1]. Among the applicable strategies for energy efficiency of manufacturing processes, Allwood et. al. identified four major approaches: designing longer-lasting products, modularisation and remanufacturing, component re-use and designing products with less material [2]. Considering metal forming tools, the design of longer-lasting products is the standard approach for energy saving in the production. However, the online tool-life assessment remains a challenge for industrial manufacturers because of the multiple influencing factors, the highly dynamical environment of metal forming processes and the related cost for measuring equipment.

Against this background, an essential step towards energy efficiency would involve the online measurement of the forming energy, which, unfortunately, is usually not directly accessible in many use cases. A solution to solve this issue consists in developing soft sensors that rely on accessible data in order to indirectly deduce the relevant aspects of the energy requirement of the process. The further development of the industrial hardware enables the processing of sensor data at different complexity level. Furthermore, it allows the application and exploitation of the versatility of machine learning (ML) solutions in order to solve use case specific challenges. Therefore, this study focused on processing force and stroke data in order to monitor and predict the energy requirement during and after a cold forging stroke. The energy deployed in order to form a part depends on the current process state including the tool and the machine and is like a signature of the current state of the process and the workpiece. So being able to understand it and how it affects

✉ Papdo Tchasse
papdo.tchasse@ifu.uni-stuttgart.de

1 Institute for Metal Forming Technology, University of Stuttgart, Holzgartenstraße 17, 70174 Stuttgart, Germany

the future strokes is not only a way of better understanding the energy flow of the process but it is also an indirect way to evaluate the state of the workpiece and the forming operation. For this reason, this paper explores ML methods able to compute the required process energy and to predict its future evolution. Such methods would allow to better evaluate how the current state of the manufacturing system, including the part, the tool and the machine, influences the energy load of a process. This would permit, for example, an early assessment of the tool life, product quality and machine downtimes.

2 State of the art

Following the global trends of sustainability, many researches focused in the past years on the energy efficiency of different metal forming processes. To this end, some works compared for example the potential energy saving of applicable processes for manufacturing the same part geometry. To this end, Ingarao et. al. for instance compared single point incremental forming (SPIF) and deep drawing for the manufacture of a pyramidal part [3], which results showed that SPIF requires more energy than deep drawing for the specific use case. In this work, although the workpiece material was the main factor influencing the energy expenditure of the process, the part and the tooling happened to play a non-negligible role. Generally, the state of the manufacturing tool and machine are known to highly influence the energy efficiency of a process [4], as shown for machining, injection moulding and coating processes in [5, 6]. Besides, another aspect to consider is the residual materials and emissions due, for example to lubrication and cleaning agents [7]. For this reason, Avram et. al. showed in their work that the first step for energy saving is to match the manufacturing system with the productivity goals [5].

Optimizing the energy consumption related to the manufacture of single products also require the monitoring of every step involved in the production. This is still a considerable challenge for manufacturing companies due, partially, to the cost of the measuring equipment [8], the complexity of the manufacturing systems and the number of data sources [9]. For many processes however, the energy consumption is directly related to the forces involved into the process. So, an essential step for assessing the energy load of a process, which would allow to implement an appropriate energy saving strategy, would be to monitor the evolution of the process forces [10–12]. In their work, Hohmann et. al. identified that monitoring the maximum forces of each stamping stroke might not be sufficient to access the wear state and hence the tool-life of the stamping punches [13]. Rather, the integral of the force, which corresponds to the process energy represents a more reliable alternative. Also, given that a high part of the energy consumption in the industry is spent as idle machine intake [14], a tool able to monitor and predict the future development of the required process energy would also allow to predict machine downtimes. As beside the quality of the manufacturing equipment, the energy monitoring is the essential prerequisite for improving the energy efficiency [15, 16].

Different works in the field of metal forming focused on the prediction or monitoring of different process factors. For the energy estimation for example, Mirandola et al. [17] explored the application of different ML methods in order to determine the required energy to form a part in a radial-axial ring rolling process. This use case was implemented in a numerical environment and the model inputs were the initial ring geometry, the final ring geometry, the ring material properties, the process settings and the initial ring temperature. Artificial neural networks, linear, kernel and ensemble models were considered in the study and trained to determine the force integral over time, which is directly related to the required forming energy for the process. In the field of quality monitoring or quality prediction, Zhao et al. attempted to predict the part flatness in the cold rolling process [18]. This quality prediction was the base for a feedback control in order to optimize the same feature. Also, Rasche et al. focused on the prediction of the forming forces and the lack of volume in the flange forging process [19]. For this purpose, different ML models were considered like artificial neural networks, support vector machines, the linear and polynomial regression models. This study was carried out in a numerical environment. In the domain of tool life and wear prediction, Kong et al. [20] proposed a method, based on artificial neural networks, for the prediction of the wear state of the tool in cold forging processes. In this theoretical study, the recommended model inputs were the forces, the acoustic emissions and other process data like temperatures, stroke rates and the surface lubrication of the in-feed material. Furthermore, Glaeser et al. explored in [21] the faults detection using a convolutional neural network and a decision tree. In that use case, the authors considered and induced different die and tool damages. Then, the experimental data were collected and processed and a high accuracy in determining the different damages could be achieved.

According to the state of the art, the most ML applications in the bulk metal forming were implemented either in a numerical environment or focused on the tool life or wear monitoring [22]. The study presented in this paper however addresses the process energy monitoring and prediction in a way that would enable a real time industrial application. The present work assumes that the internal correlation between the different phases of the process depends on the state of the manufacturing system with the part and the tool as major players. So, the goal of the data processing is to extract patterns from partial process phases and use the knowledge in order to extrapolate on the short- and long-term process evolution.

3 Methodical approach

For this use case, an initially designed two-stage cold forging process was considered. For this process, the first stage consisted in a cold forward extrusion and the second stage an upsetting step. The objective of the process is to form a cylindrical billet of diameter 15 mm and height 20 mm into a screw-like part without thread, drive and tip and with head diameter 20±0.15 mm, head height 6.25±0.1 mm, body diameter 10 mm and body height 20 mm. So, the first stage involved forming the part body and the second stage the part head. The shape of the part during the process is shown in Fig. 1.

The billet material for the cold forging process was 28B2 (1.5510) KGK. the billet was cut from a rod, that was drawn, annealed and trailed and which surface was coated with phosphate and was soaped. The formability study was mainly performed using finite element simulations. For the simulation to be started, compression tests were carried out in order to acquire the billet material properties. The process tool was mounted in a servo mechanical knuckle joint press from the Schuler company with a ram upsetting capacity of 5000 kN and a stroke rate range of 3 to 45 min^{-1}. Figure 2 shows the process tool and the press.

3.1 Process parameters

The quality of interest in this study is the energy needed to form and eject one part. Given that the process was designed with two stages, the forming energy of each stage was considered at every stroke. However, the energy itself could not be recorded directly. For this reason, it was measured and then calculated via the force and stroke measurements, based on the formula

$$W = \int F ds,$$

With F being the force, s the stroke and W the energy. Figure 3 shows, for the forward extrusion stage, the relationship between energy, force and stroke for the forming operation and the later part ejection. Thus, combining force and stroke measurements makes it possible to determine not only the energy evolution during the forming operation but also the cumulated or maximum energy expended at every stroke.

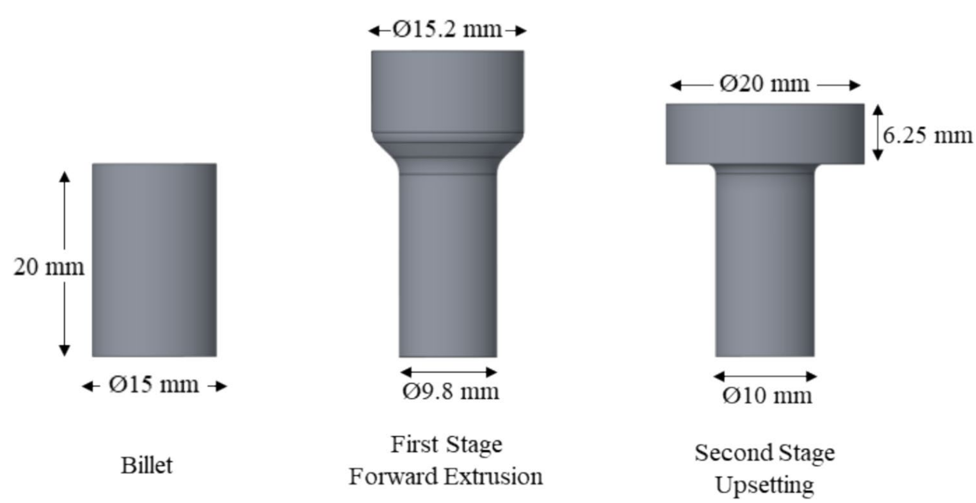
3.2 Sensor concept

Two sensors were used to record the forces during the forming operations and two were used to record the part ejection forces, which means two sensors for each process step and a total of four force sensors. Furthermore, the ram and ejector strokes were recorded with two displacement sensors. Figure 4 shows how the sensors were disposed in the tool.

3.3 Data collection

The different sensors were connected to an amplifier, which was also used to record the data. For this study a total of 500 strokes was recorded. For every stroke, the data were recorded within the ram angle range of 148° to 270°. The forces and displacements could be combined online using the recording hardware or calculated offline for the ML model design in order to determine the process energy. For both cases the energy was calculated using the trapezoidal rule for function integration. This rule approximates a function integration for digital values as follows:

Fig. 1 Billet and part shape after the two forming stages



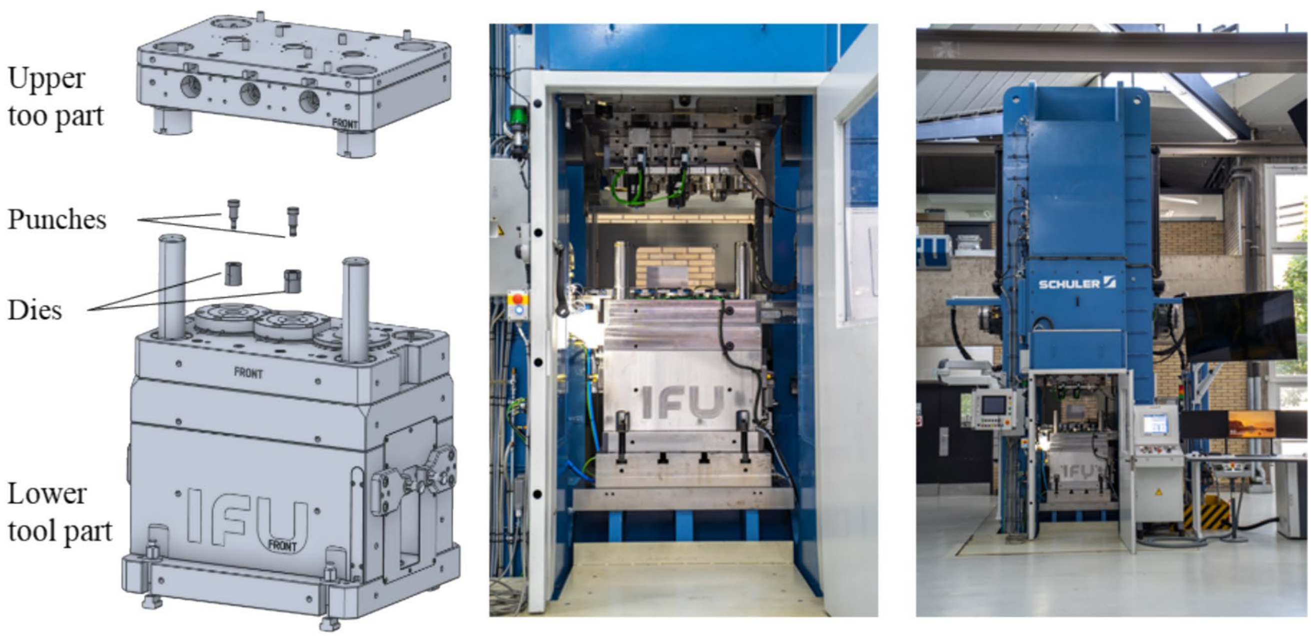
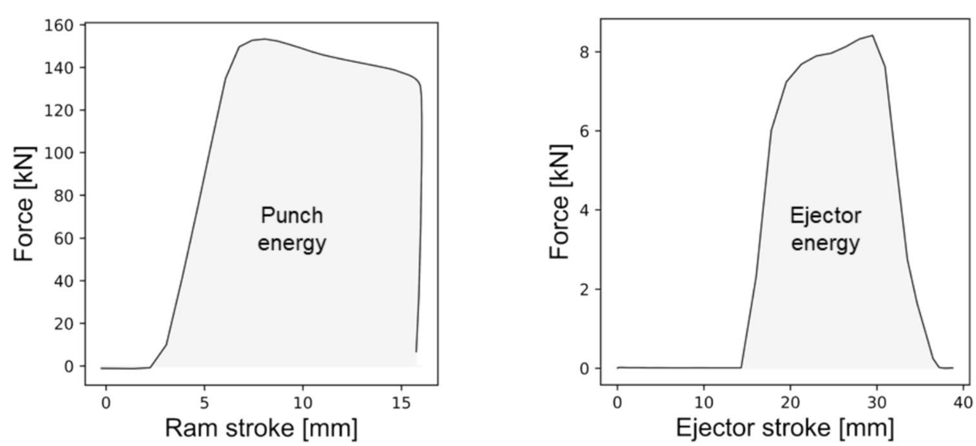


Fig. 2 Experimental setup for the process monitoring and data acquisition

Fig. 3 Forming and ejection energy for the extrusion stage as the integral of the force over stroke



$$W = \int_a^b F ds = \sum_{k=1}^{N} \frac{F_{k-1} + F_k}{2} (s_k - s_{k-1})$$

With

$F = [F_1, .., F_N]$

$s = [s_1, .., s_N]$

$s_1 = a, s_N = b$

Figure 5 shows for one stroke the evolution of the maximum forming and the ejection energy relative to the forming and ejection force.

3.4 Offline machine learning modelling

The ML model design in this study was driven by two principal interrogations. The first interrogation was, how does the cumulated or maximum energy evolve during the forming operation and how this evolution impacts the rest of the operation? And the second interrogation was, how does the energy deployed for one stroke affect the future process development? To answer these interrogations, two tasks were designed and carried out. The first task concentrated on predicting the cumulated energy requirement during a stroke based on the recent energy history within that same stroke and the second task focused on predicting the required energy for the 20 future strokes based on the energy evolution within one forming stroke. Thus, the input

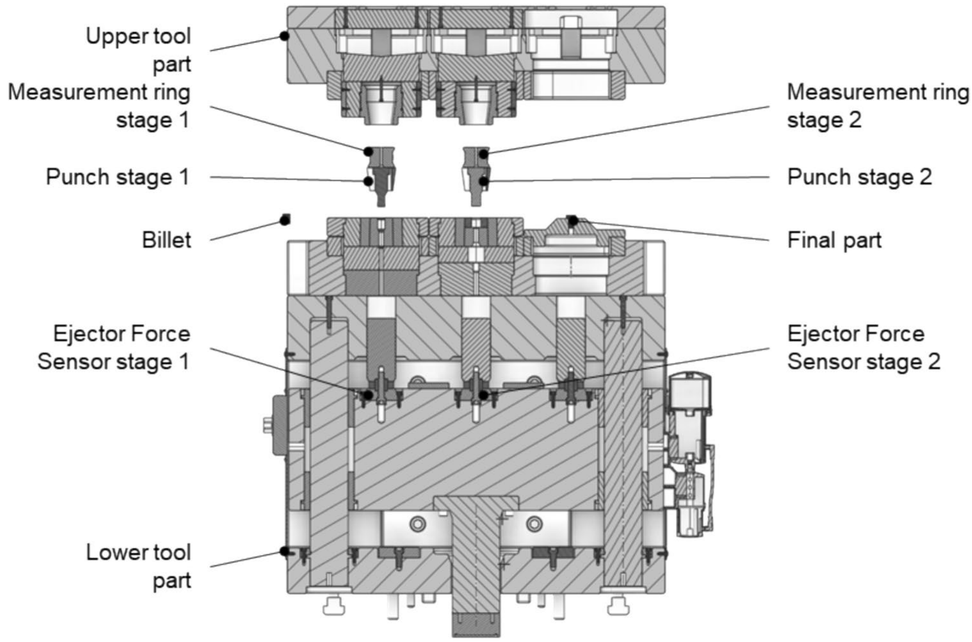


Fig. 4 Sensor concept of the two process stages for the data acquisition

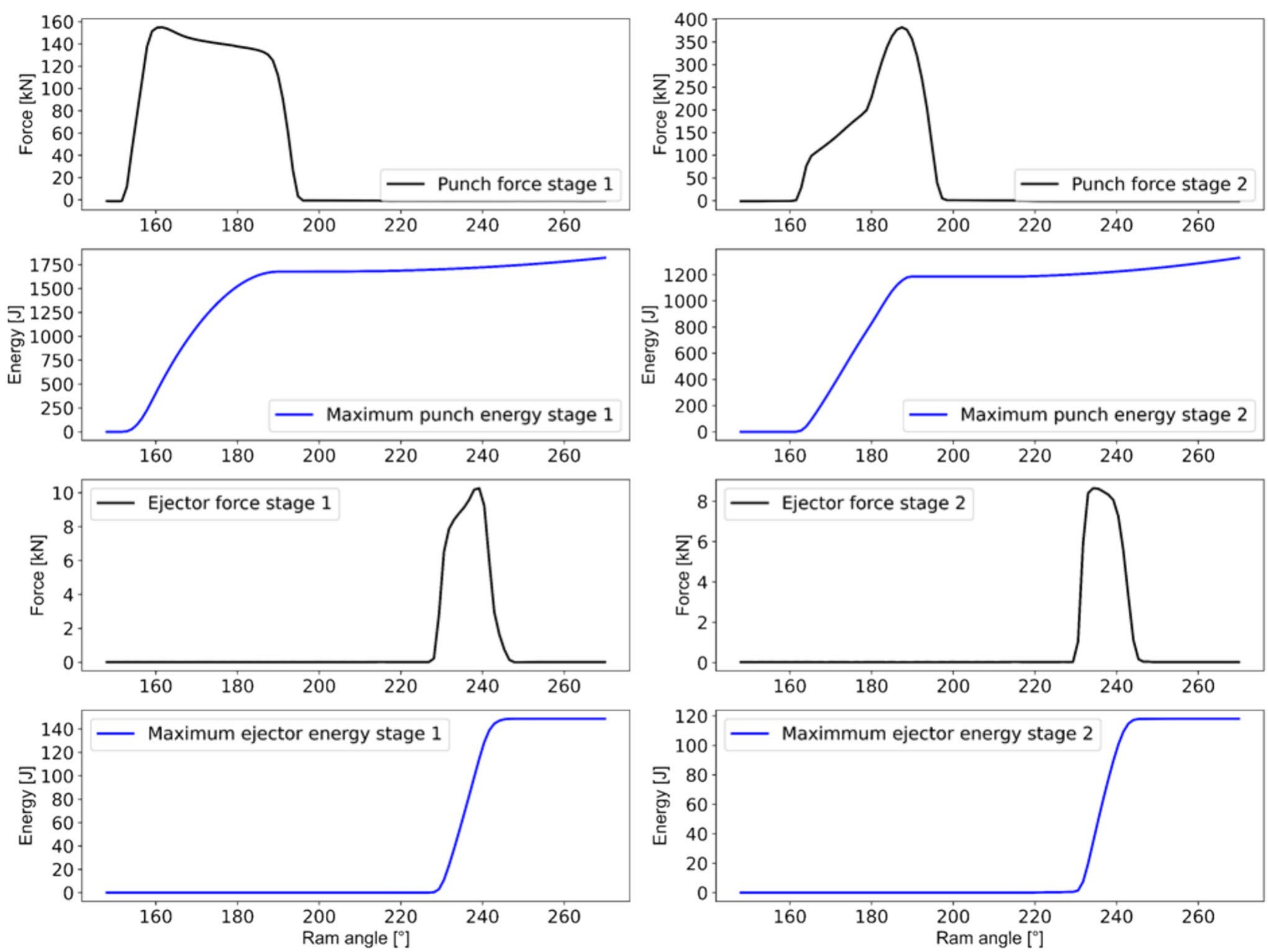


Fig. 5 Force and energy evolution during a stroke for the two process stages

and output features for both tasks were the process energies as listed in Table 1, the goal being the estimation of the short- and long-term energy evolution of the process based only on partial stroke measurements.

Two different ML or more specifically deep learning approaches were considered and compared for the data processing. The first approach was based on recurrent neural networks (RNN) and the second was based on the self-attention mechanism.

3.4.1 Recurrent neural networks

RNN are ANN with neurons that, in addition to the forward connections, also have a connection to themselves or to other neurons of the same or previous layers [23]. This ability of having a connection to themselves makes RNN more suitable for the processing of time series or more generally data, for which the sequential order is essential. Two variants of RNN have captured a lot of attentions the recent years, the long short-term memory (LSTM) Networks and the gated recurrent unit (GRU) networks [24]. The issue with the standard, also called vanilla, RNN is that the gradients of the loss function can vary exponentially or even disappear, when the model weights are respectively bigger or smaller than one, which makes the model particularly instable for the processing long sequences [23]. These issues are known as the exploding and the vanishing gradients. LSTM units were introduced [25] in order to solve those issues of the vanilla RNN. However, LSTM have many weights that have to be optimized, which makes the training slow. For this reason, gated recurrent units (GRU) were developed as a simplification of the LSTM in order to reduce the number of weights of the LSTM and make the training faster [26]. In this study, both the LSTM and the GRU were considered for the data processing.

3.4.2 Transformer encoder

The self-attention mechanism used in this study is the scaled dot-product attention introduced in [27], which is the fundamental component of the transformer architecture. The general principle of attention is to compute importance scores and assign them to different features in a representation. So, some features become “more important” or can get “more attention” than others. Self-attention is a particular use case of attention mechanisms, where elements of a sequence are replaced by a weighted sum of all the elements of the same sequence, with the weights representing how related that element is to the other sequence elements. This representation of the sequence allows to contextualise every element depending on others. So, the attention would compute how much an element of the target sequence is related to every element of the input sequence, and then use the computed score to weight a sum of the values of the input sequence. Multiple layers of scaled dot attention can be combined for an independent parallel information processing. Such a combination is called a multi-head attention. The multi-head attention computes different values for the same weights, then these values are concatenated. The advantage of the self-attention introduced with the transformer architecture is that, for many use cases, it appears to learn faster than other comparable models. In this study, the transformer encoder was rebuilt and adapted to fit the input and output sequences of the two regression tasks.

Table 1 Features considered for the data acquisition

Input–Output features	Unit
Forming energy step 1	J
Forming energy step 2	J
Ejection energy step 1	J
Ejection energy step 2	J

3.4.3 Data preprocessing

For this use-case the learning strategies for time series processing were considered. These strategies are one-to-one, one-to-many, many-to-one, many-to-many and they are named according to the length of input sequence-to-the length of the output sequence for the different models [28]. For the first task, the applied learning strategy was many-to-many and for the second task it was many-to-one, as illustrated in Figure 6. .

Both the first and the second task are multioutput regression tasks, as multiple real values are to be predicted. The inputs and the targets for these tasks are the same features but of different time steps within a press stroke and different strokes. For every stroke, the energy curves were divided into 100 time steps. While for the first task, 10 time steps during a stroke are used in order to predict the next 10, for the second task, the whole energy curves of one stroke, which corresponds to 100 timesteps, is used in order to predict the maximum energy of the 20 future strokes. The number of timesteps considered for one input–output pair is called a window. For this study, the ram angular positions were used as the time steps in the time series data processing and the window for the first task was set to 20 timesteps, with one timestep corresponding to an angle difference of 1.22°. Accordingly, the window for the second tasks contained 100 timesteps.

After collecting the data, they were shuffled and split into training, validation and test set. The training and test set represented 80% and 20% of the total data, which contained 500 strokes of 100 time steps each. 20% of the training data was also considered for the model validation during the training. Then, the data were standardized using the mean

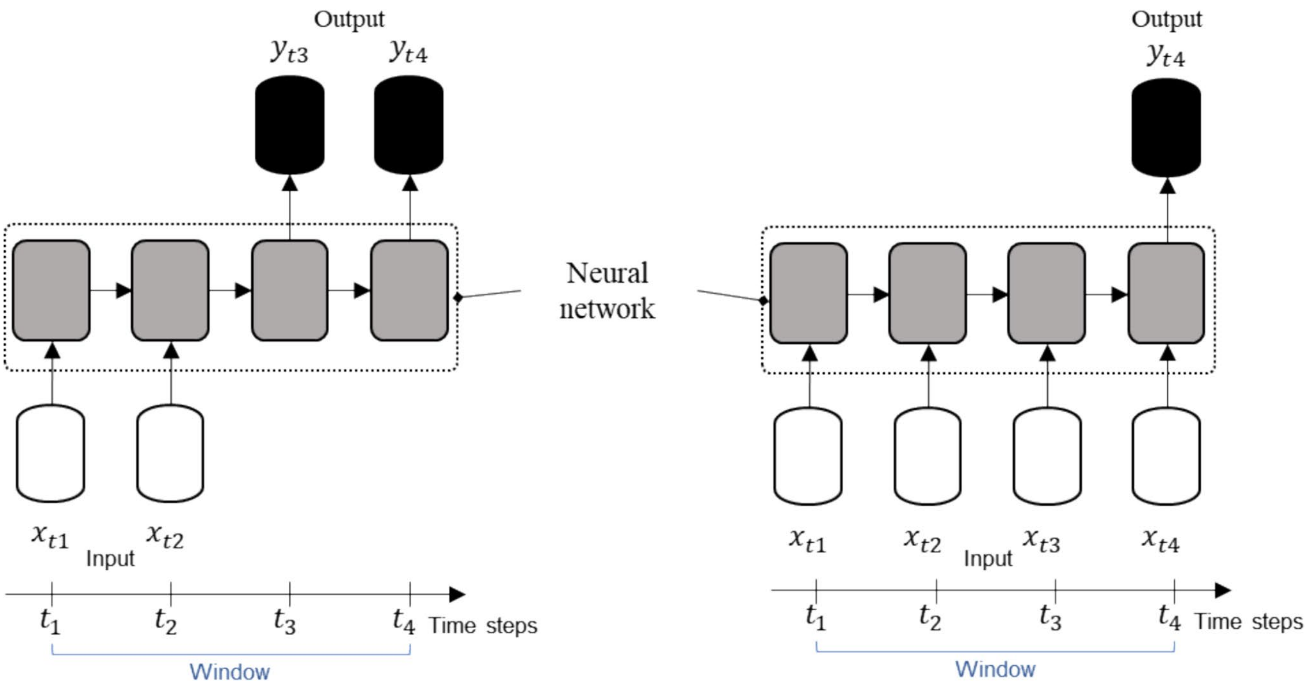


Fig. 6 Operating principle of the different models for the two tasks, left: many-to-many, right: many-to-one

and standard deviation of the training data [29]. After Data standardization, the windows and pair of input and output were generated.

3.4.4 Model building

In order to build the LSTM, GRU and self-attention models considered in this study, an automatic hyperparameter optimization with the hyperband search algorithm was used [30]. Hyperparameters are parameters that control the learning behaviour of a model and are not learned. They are set before the learning iterations and they strongly influence the prediction accuracy of the different models. For the two tasks, the same hyperparameters were considered for the different models. The activation function for the RNN layers was the hyperbolic tangent and for the transformer encoder, it was the rectified linear unit. The hyperparameters and their ranges were the same for the two RNN networks.

During the hyperparameter optimization an average of 20 models were trained for a maximum of 20 epochs with different values of the hyperparameters. After the optimization procedure, the best LSTM, GRU and self-attention networks were trained once again for a maximum of 100 epochs. The resources allocated to the model training were reduced due to the limited hardware capacity. However, this didn't prevent the different models from achieving good results as shown in the Sect. 4. During the model training, the early stopping regularisation was used in order to keep the validation loss minimal. The hyperparameters as well as their values range and best values for the LSTM, GRU and self-attention networks are summarized in the Table 2.

4 Results

The three models for this study were trained and could achieve high regression performances. The goodness of their fit was evaluated with the coefficient of determination or R^2-score and the root mean squared error (RMSE).

The objective of the first task was to predict, during the stroke, the energy development 10 timesteps ahead based on the 10 previous timesteps. The performance of the LSTM GRU and self-attention networks for this task are shown in Fig. 7.

For this first task, all three models showed an R^2-score of around 0.9 and the best performing model was the LSTM network, followed by the GRU network. Considering one stroke, Fig. 8 illustrates the predictions of the different models on a selected sample of the test set. For this example, the forming energy of the first process stage was considered and the results show that at different moment during the stroke, the models can reliably predict the further evolution of the energy.

For the second task, the targets were the maximum energy in the next 20 strokes. The idea of this task was to evaluate

Table 2 Hyperparameters of the different deep learning models

Hyperparameters	Value range	Best value
LSTM and GRU networks		
Number of hidden layers	[2, 10]	1
Number of neurons per hidden layer	[1, 100]	20
Layer Normalization per hidden layer	(yes, no)	No
Learning rate	(1e-2, 1e-3, 1e-4, 1e-5)	1e-3
Loss function	(Log-cosh, MSE)	MSE
Transformer encoder		
Number of head of the multi-head attention	[1, 8]	8
Number of encoder blocks	[1, 8]	8
Number of feed forward layers	[1, 10]	1
Number of Neurons per feed forward layer	[0, 100]	75
Dropout rate	[0.2, 0.5]	0.2
Learning rate	(1e-2, 1e-3, 1e-4, 1e-5)	1e.-3
Loss function	(Log-cosh, MSE)	MSE

the effect of a stroke on the further process development. For this task also, all the models reached an R^2-score of around 0.9 and an RMSE of max 0.319, as shown in Fig. 9.

Considering the same stroke as in Fig. 8, Fig. 10 shows the predictions of the maximum energy of the 20 next strokes in comparison with the true values. The results in this case show that the model can predict not only the trend but also the energy amplitudes with a certain accuracy.

Generally, the required energy for each stroke is mostly influenced by the workpiece and the tool. Aspects of the workpiece that might affect the process energy are its micro- and macro-geometry and structure, the mechanical properties of its material and its surface treatment. This is also valid for the tool, whereby there is usually no high fluctuation of the mechanical properties of the tool material during the production in contrast to the part. The variations related to the part

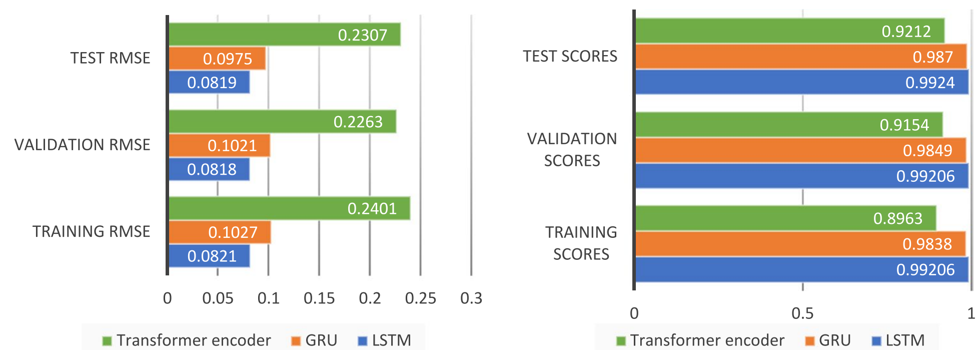


Fig. 7 Model performances for the maximum prediction during the stroke

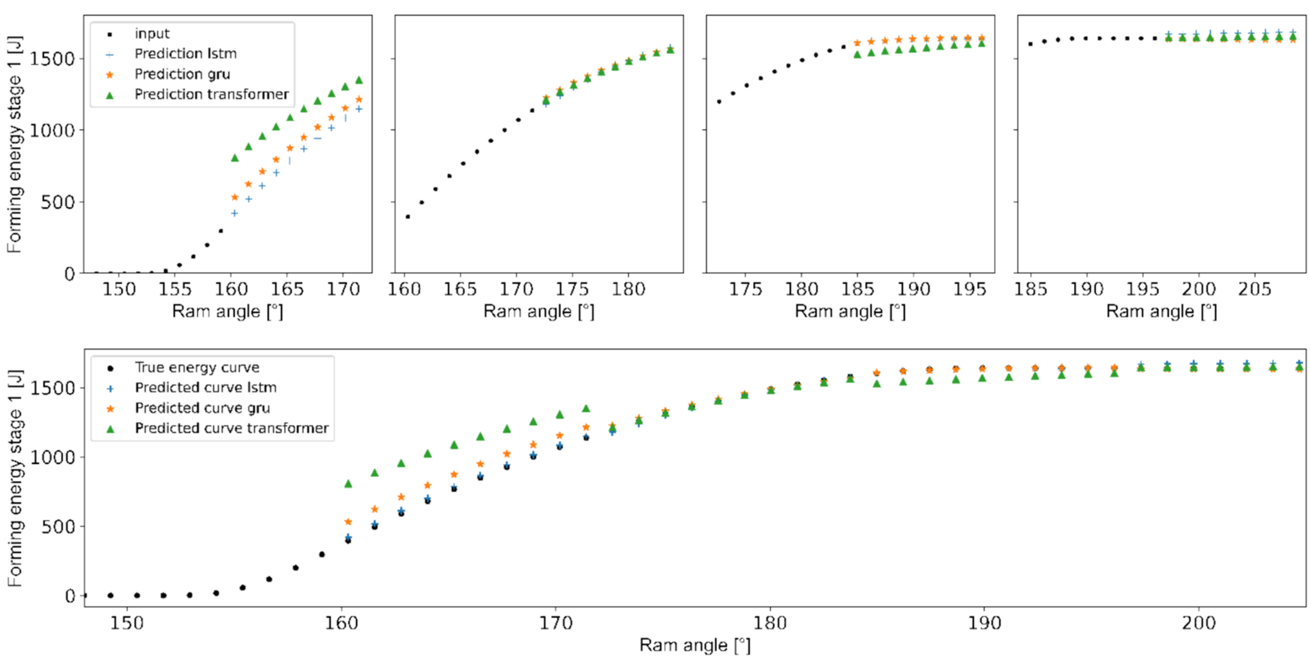


Fig. 8 Model predictions for the maximum forming energy prediction during the stroke for the first process stage

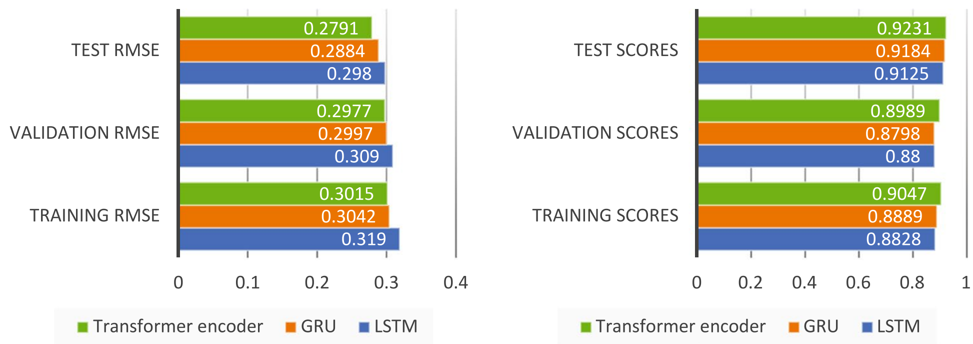


Fig. 9 Model performances for the maximum energy prediction of 20 strokes ahead

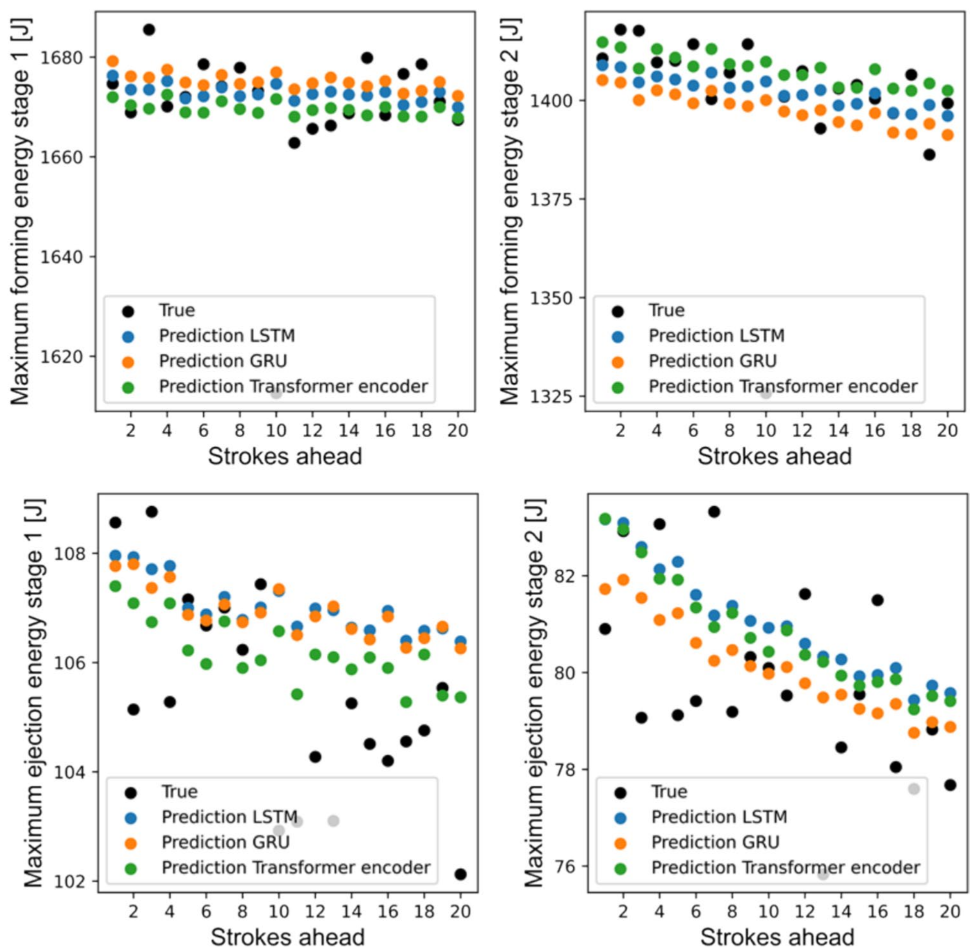


Fig. 10 Model predictions of the maximum energy 20 strokes ahead for one stroke

and the tool can be assigned to different temporalities regarding the fluctuations in the energy requirement of the process. While the short-term energy variations might be related to the fluctuations of the part geometry and surface quality, the long-term might be related to the tool. The prediction for the next 20 strokes as shown in Fig. 10 might represent a long-term prediction with reference to the actual stroke. However, with regard to the tool life, 20 strokes still represent a short-term scenario. Against this background the scattering observed in Fig. 10 rather exposes the fluctuations in the material, geometry and surface properties of the workpiece.

5 Conclusion and outlook

In a context of limited natural resources, improving the energy efficiency of metal forming processes becomes an imperative. If the high quality of the manufacturing infrastructure and the high production rate are key aspects for optimizing the energy expenditure of production systems, the energy monitoring also remains essential for a realistic assessment of the tool- and machine life. Against this background, this study focused on the energy monitoring in a multi-stage cold forging process and presented methods to determine and predict the energy requirement of the forming process using deep learning models for time series processing. The fundamental hypothesis for this study was that the different phases of a forming stroke and generally of the process correlate with each other, as they are related to unique workpieces and a particular state of the tool and the machine. Therefore, partial information of the process can be exploited in order to predict its further evolution. To confirm this hypothesis, LSTM, GRU and self-attention neural networks were built to predict the in-stroke and the future energy requirement for cold forging operations. The results showed that the energy load of a forming stroke as well as the energy requirements of future press strokes can be predicted with high accuracy. This work showed that RNN and self-attention-based architectures, which are well suited for time series processing, can both be leveraged for determining the energy expenditure of cold forging manufacturing systems. Being able to monitor and predict the energy consumption in a process allows to understand the effects of the current state of the manufacturing components on the further development of the process. Because the energy required for every stroke depends on the part, the tool and the machine, predicting its evolution allows to indirectly monitor the state of these essential components of the production system. Therefore, aspects such as the part quality, tool wear and machine downtimes can also be inferred. Furthermore, it would enable to better understand the process and optimally orient control strategies in order to be more sustainable. These solutions can be well integrated online and could help manufacturing companies better manage their energy consumption. Besides, the high prediction scores achieved by the times series models indicate that there might be a common factor that can explain the relationship between different workpieces from the same batch, although they were randomly feeded into the process. In this context, methods of explainable artificial intelligence can be used to better highlight such effects in order to understand the distribution of the required forces and energy for the part manufacturing.

Funding Open Access funding enabled and organized by Projekt DEAL.

Data availability Due to the sensitive nature of the research supporting data and the ongoing works, data sharing is not applicable to this article.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

1. Singh S, Ramakrishna S, Gupta MK (2017) Towards zero waste manufacturing: a multidisciplinary review. J Clean Prod 168:1230–1243
2. Allwood JM, Ashby MF, Gutowski TG, Worrell E (2011) Material efficiency: a white paper. Resour Conserv Recycl 55(3):362–381
3. Ingarao G, Ambrogio G, Gagliardi F, Di Lorenzo R (2012) A sustainability point of view on sheet metal forming operations: material wasting and energy consumption in incremental forming and stamping processes. J Clean Prod 29–30:255–268
4. Ingarao G, Di Lorenzo R, Micari F (2011) Sustainability issues in sheet metal forming processes: an overview. J Clean Prod 19(4):337–347
5. Avram OI, Xirouchakis P (2011) Evaluating the use phase energy requirements of a machine tool system. J Clean Prod 19(6–7):699–711
6. T. Gutowski, J. Dahmus and A. Thiriez, "Electrical Energy Requirements f or Manuf acturing Processes 生产过程的电能要求", *13th CIRP Int. Conf. LIFE CYCLE Eng.*, pp. 623–628, 2006.
7. Geiger M (1995) Towards clean forming techniques. CIRP Ann 44(2):581–588
8. Rahimifard S, Seow Y, Childs T (2010) Minimising embodied product energy to support energy efficient manufacturing. CIRP Ann-Manuf Technol 59(1):25–28
9. Vijayaraghavan A, Dornfeld D (2010) Automated energy monitoring of machine tools. CIRP Ann-Manuf Technol 59(1):21–24
10. Rajemi MF, Mativenga PT, Aramcharoen A (2010) Sustainable machining: selection of optimum turning conditions based on minimum energy considerations. J Clean Prod 18(10–11):1059–1065
11. Filice L, Ambrogio G, Micari F (2006) On-line control of single point incremental forming operations through punch force monitoring. CIRP Ann 55(1):245–248
12. Zhao GY, Liu ZY, He Y, Cao HJ, Guo YB (2017) Energy consumption in machining: classification, prediction, and reduction strategy. Energy 133:142–157
13. J. Hohmann, T. Schatz and P. Groche, "Intelligent wear identification based on sensory inline information for a stamping process", *Lect. Notes Mech. Eng.*, pp. 285–295, 2017.
14. Fysikopoulos A, Pastras G, Alexopoulos T, Chryssolouris G (2014) On a generalized approach to manufacturing energy efficiency. Int J Adv Manuf Technol 73(9–12):1437–1452
15. Gao M, He K, Li L, Wang Q, Liu C (2019) A review on energy consumption, energy efficiency and energy saving of metal forming processes from different hierarchies. Processes. https://doi.org/10.3390/pr7060357
16. Seow Y, Rahimifard S (2011) A framework for modelling energy consumption within manufacturing systems. CIRP J Manuf Sci Technol 4(3):258–264
17. Mirandola I, Berti GA, Caracciolo R, Lee S, Kim N, Quagliato L (2021) Machine learning-based models for the estimation of the

energy consumption in metal forming processes. Metals. https://doi.org/10.3390/met11050833
18. Zhao J et al (2023) A novel paradigm of flatness prediction and optimization for strip tandem cold rolling by cloud-edge collaboration. J Mater Process Technol 316:117947
19. Rasche NF, Langner J, Stonis M, Behrens B-A (2017) Predicting forming forces and lack of volume with data mining methods for a flange forging process. Int J Model Optim 7(6):363–369
20. Kong LX, Nahavandi S (2002) On-line tool condition monitoring and control system in forging processes. J Mater Process Technol 125–126:464–470
21. Glaeser A et al (2021) Applications of deep learning for fault detection in industrial cold forging. Int J Prod Res 59(16):4826–4835
22. Cao J, Bambach M, Merklein M, Mozaffar M, Xue T (2024) Artificial intelligence in metal forming. CIRP Ann 00:1–27
23. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
24. F. Chollet, *Deep Learning with Python*, 2nd ed., 2021.
25. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Comput 9(8):1735–1780
26. J. Chung, C. Gulcehre, K. Cho and Y. Bengio, "Empirical Evaluation of Gated Recurrent Neural Networks on Sequence Modeling", pp. 1–9, 2014.
27. A. Vaswani *et al.*, "Attention is all you need", *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Nips, pp. 5999–6009, 2017.
28. Géron A (2019) Hands-on Machine Learning whith Scikit-Learing, Keras and Tensorfow. O'Reilly Media
29. Y. A. LeCun, L. Bottou, G. B. Orr and K.-R. Müller, "Efficient BackProp BT - Neural Networks: Tricks of the Trade", *Neural Networks: Tricks of the Trade*, vol. 7700, no. Chapter 3, pp. 9–48, 2012.
30. Li L, Jamieson K, DeSalvo G, Rostamizadeh A, Talwalkar A (2018) Hyperband: a novel bandit-based approach to hyperparameter optimization. J Mach Learn Res 18:1–52