Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 220 (2023) 576–583

Procedia Computer Science

www.elsevier.com/locate/procedia

The 6th International Conference on Emerging Data and Industry (EDI40), March 15-17, 2023, Leuven, Belgium

Predictive maintenance in pharmaceutical manufacturing lines using deep transformers

Isaak Kavasidis[a,*], Efthimios Lallas[b], Vassilis C. Gerogiannis[b], Theodosia Charitou[c], Anthony Karageorgos[b]

[a]*South-East European Research Centre (SEERC), Thessaloniki and University of Thessaly, Larissa, Greece*
[b]*School of Technology, University of Thessaly, Larissa, Greece*
[c]*Department of Computer Science and Biomedical Informatics, University of Thessaly, Larissa, Greece*

Abstract

Inherent complexities in pharmaceutical manufacturing lines of modern industrial facilities make precise and timely detection of malfunction occurrences necessary. In fact, unpredicted malfunctions in a production line can often provoke a cascade of adverse effects that can occur everywhere in the production chain bringing the manufacturing line to a halt for undefined time periods. Such events can have unfortunate consequences that are not always confined to the damaged part itself but propagate throughout the production line. Nevertheless, modern production lines are equipped with a multitude of data sensors that enable the real-time and fine-grained monitoring of each constituent part of the production process providing a richness of information that can be exploited by intelligent data processing methods.

In this work, we present ManuTrans, a deep learning-based model for monitoring real-time raw sensor data, deriving the condition of a pharmaceutical manufacturing line and predicting the next moment in time when a malfunction can occur. The model is further able to predict the severity of the next malfunction and can contribute adjunct information in corporate decision-making. The suggested approach exploits the capacity of deep transformer models for extracting both long- and short-term correlations as well as patterns in sequential data and, combined with a linear output layer, conducts both classification and regression. The proposed approach was tested on a real dataset comprising raw data from two manufacturing lines, and it achieved promising results.

Keywords: Deep Learning; Industry 4.0; Smart Manufacturing

* Corresponding author.
E-mail address: ikavasidis@seerc.org

1877-0509
10.1016/j.procs.2023.03.073

Nomenclature

AI Artificial Intelligence
ARIMA Autoregressive Integrated Moving Average
CE Cross Entropy
LSTM Long Short Term Memory
MSE Mean Squared Error
SGD Stochastic Gradient Descent
SVM Support Vector Machine
XAI Explainable Artificial Intelligence

1. Introduction

The pharmaceutical industry relies on rigorous regulations regarding pharma manufacturing data which must abide by specific quality standards to be auditable whenever government institutions need to. For this reason, pharmaceutical production industries must monitor each step of their production processes at a very fine-grained level. Nowadays, pharmaceutical manufacturing lines include various sensors designed to monitor many discreet and continuous physical variables registered and stored for an extended period after production, such as data about temperatures, rotations, pressures, and chemical compositions.

Nowadays, there is an increased need to deliver high-quality drug products in a short time frame as possible. For this reason, it is necessary to decrease the level of manual intervention in manufacturing processes and move toward modular manufacturing where continuous adjustments and improvements can be made based on in-line sensors and process monitoring [4]. Furthermore, latest innovative pharmaceuticals manufacturing technologies require novel control methods and increased regulatory checks. Therefore, given the richness of information produced during the manufacturing process, it is imperative to ensure that manufacturing organizations remain agile and able to adopt new capabilities to optimize manufacturing operations and deliver high quality products [11].

Considering the complexity of such manufacturing lines, there is a higher probability for malfunctions to occur. Pharmaceutical industries invest many resources in physically monitoring and maintaining their production lines to minimize the risk of a malfunction occurring. In this respect, accuracy and high quality of manufacturing equipment are vital, since even small deviations in the functioning of pharmaceutical equipment may result in serious consequences involving financial losses, delays, and even affecting human lives. Hence, a critical challenge faced in pharmaceutical manufacturing is to ensure that the equipment functions in optimal conditions, and thus reduce downtime and prevent bottlenecks.

All these needs for optimized performance have led the pharmaceutical industry to adopt technologies and solutions that can help prevent delays and maintain efficient operations. Along this line industry professionals have steered towards predictive maintenance, for proactive asset maintenance and management [3]. Predictive maintenance can be considered a data-driven approach that collects and analyses machine health and performance data, to predict when an asset will fail. Production process-related data are typically collected by industrial IoT sensors and by further data analysis, meaningful and actionable insights are provided.

To this end, in this work, we present ManuTrans, a model based on deep learning transformers that can evaluate the health condition of a production line and also predict the moment in time of the next malfunction and its severity. The model is trainable in an end-to-end manner, with a focus on reducing the complexity of the training process. The proposed approach was trained on raw data gathered by two operating manufacturing lines and labeled manually by two line operators in three years of a real pharmaceutical manufacturing process. The performance of the model was further tested and compared against three different approaches and the obtained results are reported.

The remainder of the paper is structured as follows: Section 2 presents related work which has been carried out in predictive maintenance, with a particular focus on pharmaceutical production lines, while Section 3 presents the model itself. Section 4 reports the performance evaluation of the presented method and in the last section of the paper conclusions are drawn and future directions are given.

2. Related work

It is common practice in data-driven approaches, which are very popular in Industry 4.0 and predictive maintenance, to use large amounts of data to predict the system residual operational lifetime, mostly based on raw data from sensor signals [5].

Both supervised and unsupervised learning methods are employed, with each category having its respective advantages and drawbacks. While unsupervised learning algorithms are easier to train and they are being used for unlabeled data classification, the natural application domain of this type of algorithms lies in the detection of anomalies and they can be applied in cases where the generation or acquisition of labelled data is resource-intensive [2].

Conventional data-driven predictive maintenance approaches generally require too much prior knowledge [15, 16]. Furthermore, in most cases, data processing and analysis is performed manually by members of the operating staff, which may be time and cost consuming, especially in the case of big data automatized processing. On the other hand, modern data-driven predictive maintenance approaches based on deep learning methods exploit the minimum required signal processing knowledge, completely automatically, and without the need for human intervention [6].

Unsupervised approaches in predictive maintenance usually employ statistical analysis [18], cluster analysis [1] and deep learning [13] and, in the specific field of pharmaceutical manufacturing sector, they have used autoencoders [8]. However, when data labels are available, supervised learning is the preferred way to deal with both regression and classification tasks.

Established deep learning methods such as the LSTMs [24, 21, 17], or even more sophisticated ones, such as the Convolutional Bi-directional LSTM [25], have been extensively employed to exploit time-series raw sensor data and achieve residual life prediction and predictive maintenance of utilized equipment. However, these methods often do not provide any information about the severity of the next production failure.

Furthermore, Deep Transformer Models constitute a drastic solution in industrial systems with complex and dynamically evolved patterns and multivariate time series data, such as the ones used for time-series forecasting [23]. Even transformer-based models that were not originally built for the predictive maintenance task specifically (e.g., language modelling [19]) are being repurposed.

However, to the best of our knowledge, the existing deep learning approaches proposed in literature for predictive maintenance do not deal with the aspect of malfunction severity and manufacturing line condition estimation using one single model. To this end, in the next section we present ManuTrans, a transformer-based method that aims at providing such information by using a single model trained in an end-to-end manner.

3. Method

The model presented herein (ManuTrans) exploits the capacity of transformers in modeling relationships and patterns in sequential data. Uncovering long and/or short-term patterns in data can provide manufacturing facilities with the ability to predict when the next production line malfunction is going to occur and consequently take appropriate measures. From the “vanilla” transformer’s architecture [20], our model borrows the encoder module but instead of the decoder, our approach employs a set of three linear networks as output to predict the line’s health condition, the next instant of a malfunction and its severity, respectively.

In particular, the proposed model features three outputs for each predicted variable, as shown in Fig. 1. Preprocessing starts by normalizing the values in the signals per channel and then sequences of the signals are created and fed into the encoder part of the architecture together with a classification token. The multi-attention heads of each encoder are then trained to identify interdependencies in the incoming data, both short- and long-term.

After processing the input by the encoder, the resulting intermediate representation along with the classification token is then fed into a cascade of linear layers. Two of the layers operate in classification mode and, as such, the SoftMax function is employed, while the last linear layer operates in regression mode.

To train the network, a set of losses is used which is equal to the number of predictions the model is designed to perform. To train the two classification layers, the Categorical Cross-entropy loss is used:

$$L_{CE_M} = -\sum_{0}^{2}(y(x)log(p(x)) \qquad (1)$$

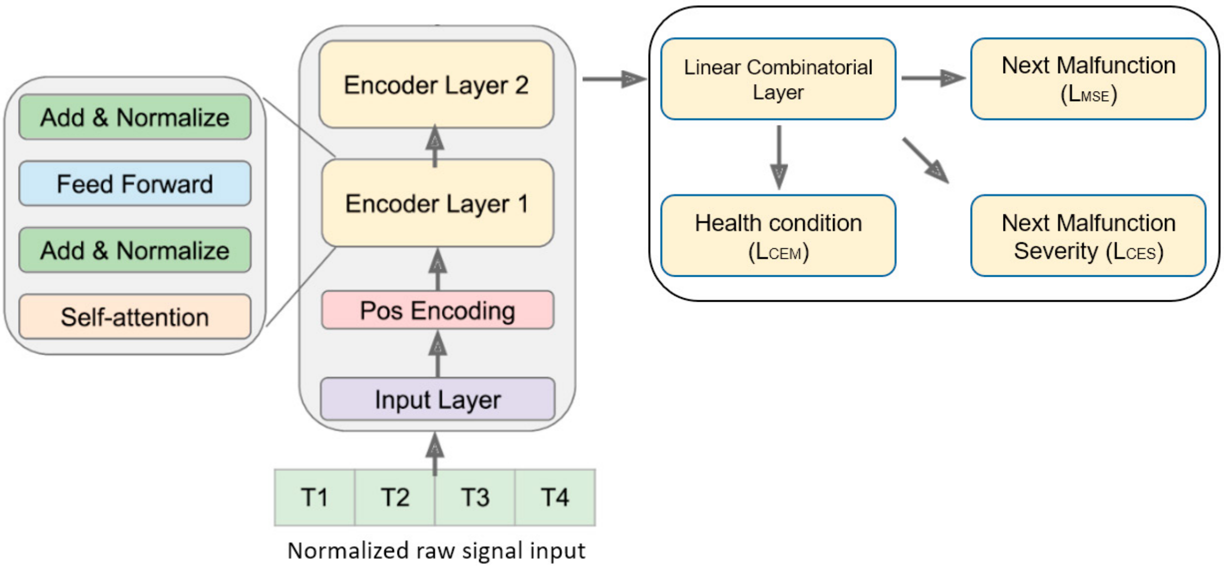

Fig. 1. The ManuTrans model for checking the manufacturing line's health and predicting malfunctions. The input signals are fed into the encoder of the transformer where the multi-attention head derives long- and short-term interdependencies in data. The output of the encoder is then fed to a cascade of linear networks with three outputs, each one trained against its corresponding loss (in parentheses) to solve the predictive maintenance task.

and

$$L_{CE_S} = -\sum_{0}^{2}(y(x)log(p(x))) \tag{2}$$

where, for each input sample x, $y(x)$ denotes the actual classification target of x and $p(x)$ denotes the probability of x to belong to the target distribution. Both classification tasks are posed as three-class problems: in the case of the manufacturing line's health condition, the targets are 0 (*Bad*), 1 (*Intermediate*) and 2 (*Good*), while, when performing the severity prediction task, the targets are 0 (*Mild*), 1 (*Moderate*) and 2 (*Severe*).

The output of the regression layer reports a real number instead which is expressed in manufacturing cycles when the next malfunction is predicted to take place. For each raw data input i, the Mean Squared Error between the output p_i and the target time p_i is computed:

$$L_{MSE} = \frac{1}{N}\sum_{i}^{n}(p_i - y_i)^2 \tag{3}$$

Finally, the total network loss, L is given by the sum of the three aforementioned losses:

$$L = L_{MSE} + L_{CE_S} + L_{CE_M} \tag{4}$$

.

In this way, the network is trained in an end-to-end fashion and no intermediate intervention is needed. More details on the mathematical background on which the above mentioned approach is based can be found in [20].

4. Performance evaluation

4.1. Dataset and hyperparameters

The data for this work was provided by a pharmaceutical company and obtained from two manufacturing line sensors. The values contained within the dataset include mainly physical time-series variables such as temperature, humidity, speed, pressure, etc. The first production line data (I1000) contains 176 batches of independent drug batches, the second production line data (I600) contains 296 batches. Each batch is treated as a new cycle of production process. When a batch ends, it starts a new circle of drug production, producing chronologically consecutive batch production orders.

For each data sample (i.e., data related to a single production batch) labels were provided by two manufacturing line operators independently. The labels include the following three pieces of information:

1. The global "health" condition of the manufacturing line as a whole (0,1,2)
2. The remaining time before the next programmed maintenance intervention (in numbers of batches), and
3. The severity of the malfunction that would occur, in case the next maintenance intervention did not take place (0,1,2).

In the case of a discordance between the evaluation of the two operators, then the manufacturing line's supervisor decides about the correct label.

All data values were normalized in the range $[0, 1]$ on a per-channel manner. Data was divided in chunks each one containing a production batch related data. The hyperparameters used in the experiments presented in the next subsection are shown in Table 1. The results reported in the next subsection are the performance values obtained by testing the model on the test set, with the lowest loss of the model's performance on the validation set.

Table 1. Hyperparameters for the ManuTrans evaluation. When deciding the values of the hyperparameters marked with an asterisk (*), the values reported were assigned after explorative evaluation (see Table 3).

Hyper parameter	Value
Learning rate	0.0001
Optimizer	Adam [10]
Beta1	0.9
Beta2	0.999
Batch size	64
ManuTrans Encoder layers *	2
ManuTrans Attention heads *	8

4.2. Experiments

In order to evaluate the performance of our method, the aforementioned data was also used to test how conventional recurrent neural networks and traditional machine learning methods work. For this reason, we compare our method against [22], where an LSTM-based neural network was suggested to predict malfunctions for the purpose of predictive maintenance. We also compare our method with the SVM-based classifier presented in [12] and the ARIMA-based statistical classifier described in [9].

For the model suggested in the current paper and the LSTM-based one, one network only was trained for all three tasks while for the ARIMA and the SVM classifiers one model was trained on the whole data for each task. In the SVM case for the task of predictive maintenance an SGD Regressor was used.

The results obtained after testing the aforementioned methods can be seen in Table 2. For the task of classifying the manufacturing line's health condition, almost all methods performed well with the exception of the SVM classifier. Deep-learning methods performed marginally better when compared against the ARIMA classifier (1% and 2% for ManuTrans and the LSTM-based one, respectively).

The inferior performance of the model presented in the present paper with respect to the LSTM-based one could be explained by two correlated facts: i) transformer-based architectures are notoriously known for the large volumes of data needed to train them in order to achieve good performance, and ii) the simplicity of the performed classification task. For such a simple independent task, the LSTM model managed to better exploit patterns in data that possess informational content with respect to ManuTrans. However, the experiments that follow demonstrate that even with

Table 2. Performance of the tested models in the three tasks. In the cases of ARIMA and SVM, three different models were trained, one for each task. The best obtained values are shown in bold.

Method	Health condition	Predictive maintenance (error, in cycles)	Malfunction severity
ARIMA [9]	91%	2.96	84%
SVM [12]	84%	1.87	84%
LSTM [22]	**93%**	0.76	72%
ManuTrans	92%	**0.41**	**91%**

such low volumes of data samples, transformers perform better in the case of complex and correlated tasks (i.e., patterns carrying information about predicting the next failure in the production line contain also informational content about the failure's severity).

In the predictive maintenance task, the proposed model performed the best with a large margin (0.41 vs 0.76, 1.87 and 2.96 for ManuTrans, LSTM, SVM and ARIMA, respectively) highlighting the capacity of transformers for modelling better and more effectively complex long- and short-term relationships in time-series data.

Finally, for the task of predicting the severity of the next malfunction in the production line, again, ManuTrans outperforms with a large margin the other methods and this can also be explained by the higher performance obtained in the previous task. While the two statistical methods (SVM and ARIMA) performed identically, the presented model outperformed these methods by a solid 12% in classification accuracy.

In addition to the previously described experiments, we have further conducted some hyperparameter tuning experiments regarding the complexity of the transformer encoder network. In particular, we have studied how the number of the used encoder layers and the number of attention heads are influencing the results. The results of the hyperparameter analysis can be seen in Table 3.

Table 3. Analysis results with varying number of encoder layers and attention heads . In bold, the configuration with the highest obtained performance.

Number of encoder layers	Attention heads	Health condition	Predictive maintenance (error, in cycles)	Malfunction severity
1	4	71%	4.32	63%
1	8	74%	3.28	71%
1	16	88%	2.01	77%
2	4	91%	0.74	88 %
2	**8**	**92%**	**0.41**	**91%**
2	16	90%	0.76	87%
3	4	89%	1.14	83%
3	8	84%	1.37	80%
3	16	84%	1.46	80%
4	4	73%	2.12	74%
4	8	73%	2.12	74%
4	16	73%	2.13	74%

The results show that the best performing configuration of the transformer for the tasks tested in this work is an encoder with 2 encoder layers and 8 attention heads. Using less than that results in the model underperforming due to underfitting and using more layers and/or heads resulted, again, in the model underperforming due to overfitting. In fact, the results show that while there is a steep and steady increase in performance in all three tasks when both encoder layers and attention head number increase, it peaks at 2 layers and 8 attention heads. Afterwards, as the complexity of the network increases, the resulting performance gradually deteriorates, until reaching a plateau. When the number of encoder layers was set equal to 4, the model performed virtually the same, no matter what the number of attention heads was.

5. Conclusion

Optimized configuration and maintaining the best possible production process conditions are paramount requirements in the pharmaceuticals industry in order to not only avoid possible malfunctions that can compromise the correct functioning of the facility itself, but also to guarantee the highest possible product quality.

Thus, in this work, we presented a model for dealing with the task of preventive maintenance in pharmaceutical manufacturing lines. The model is able to evaluate the manufacturing line's health condition, predict the instant in time of the next malfunction occurrence (in production cycles) and also predict the severity of such malfunction. The model is based on the transformers and in particular the encoder part and works on normalized raw signal data.

The model was tested and compared against LSTM, SVM and ARIMA based classifiers and regressors overperforming all of them in almost all tasks. The only exception occurred in the manufacturing line's health condition evaluation task where the LSTM-based classifier outperformed our model marginally (by 1%).

Future work is needed to integrate more automated methods into our model such as XAI methods [7, 14] as well as to pinpoint the exact location of the predicted function failure, providing further advantages in maintaining production lines.

Acknowledgements

This work has been developed under the auspice of the "Smart Pharmaceutical Manufacturing (SPuMoNI)" research project funded from 2019 to 2022 by CHIST-ERA, the Horizon 2020 Future and Emerging Technologies programme of the European Union through the ERA-NET Co-fund funding scheme (CHIST-ERA BDSI Call 2017).

References

[1] Amruthnath, N., Gupta, T., 2018a. Fault class prediction in unsupervised learning using model-based clustering approach, in: 2018 International Conference on Information and Computer Technologies (ICICT), IEEE. pp. 5–12.
[2] Amruthnath, N., Gupta, T., 2018b. A research study on unsupervised machine learning algorithms for early fault detection in predictive maintenance, in: 2018 5th international conference on industrial engineering and applications (ICIEA), IEEE. pp. 355–361.
[3] Carvalho, T.P., Soares, F.A., Vita, R., Francisco, R.d.P., Basto, J.P., Alcalá, S.G., 2019. A systematic literature review of machine learning methods applied to predictive maintenance. Computers & Industrial Engineering 137, 106024.
[4] Çınar, Z.M., Abdussalam Nuhu, A., Zeeshan, Q., Korhan, O., Asmael, M., Safaei, B., 2020. Machine learning in predictive maintenance towards sustainable smart manufacturing in industry 4.0. Sustainability 12, 8211.
[5] Daily, J., Peterson, J., 2017. Predictive maintenance: How big data analysis can improve maintenance, in: Supply chain integration challenges in commercial aerospace. Springer, pp. 267–278.
[6] Deutsch, J., He, D., 2017. Using deep learning-based approach to predict remaining useful life of rotating components. IEEE Transactions on Systems, Man, and Cybernetics: Systems 48, 11–20.
[7] Gade, K., Geyik, S.C., Kenthapadi, K., Mithal, V., Taly, A., 2019. Explainable ai in industry, in: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining, pp. 3203–3204.
[8] Juodelyte, D., Cheplygina, V., Graversen, T., Bonnet, P., 2022. Predicting bearings' degradation stages for predictive maintenance in the pharmaceutical industry. arXiv preprint arXiv:2203.03259 .
[9] Kanawaday, A., Sane, A., 2017. Machine learning for predictive maintenance of industrial machines using iot sensor data, in: 2017 8th IEEE international conference on software engineering and service science (ICSESS), IEEE. pp. 87–90.
[10] Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 .
[11] Kumar, S.H., Talasila, D., Gowrav, M., Gangadharappa, H., 2020. Adaptations of pharma 4.0 from industry 4.0. Drug Invention Today 14.
[12] Li, H., Parikh, D., He, Q., Qian, B., Li, Z., Fang, D., Hampapur, A., 2014. Improving rail network velocity: A machine learning approach to predictive maintenance. Transportation Research Part C: Emerging Technologies 45, 17–26.
[13] Lygren, S., Piantanida, M., Amendola, A., 2019. Unsupervised, deep learning-based detection of failures in industrial equipments: The future of predictive maintenance, in: Abu Dhabi International Petroleum Exhibition & Conference, OnePetro.
[14] Matzka, S., 2020. Explainable artificial intelligence for predictive maintenance applications, in: 2020 Third International Conference on Artificial Intelligence for Industries (AI4I), IEEE. pp. 69–74.
[15] Medjaher, K., Tobon-Mejia, D.A., Zerhouni, N., 2012. Remaining useful life estimation of critical components with application to bearings. IEEE Transactions on Reliability 61, 292–302.
[16] Mosallam, A., Medjaher, K., Zerhouni, N., 2013. Nonparametric time series modelling for industrial prognostics and health management. The International Journal of Advanced Manufacturing Technology 69, 1685–1699.
[17] Nguyen, K.T., Medjaher, K., 2019. A new dynamic predictive maintenance framework using deep learning for failure prognostics. Reliability Engineering & System Safety 188, 251–262.

[18] da Silva Arantes, J., da Silva Arantes, M., Fröhlich, H.B., Siret, L., Bonnard, R., 2021. A novel unsupervised method for anomaly detection in time series based on statistical features for industrial predictive maintenance. International Journal of Data Science and Analytics 12, 383–404.

[19] Usuga-Cadavid, J.P., Lamouri, S., Grabot, B., Fortin, A., 2022. Using deep learning to value free-form text data for predictive maintenance. International Journal of Production Research 60, 4548–4575.

[20] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. Advances in neural information processing systems 30.

[21] Wu, Y., Yuan, M., Dong, S., Lin, L., Liu, Y., 2018. Remaining useful life estimation of engineered systems using vanilla lstm neural networks. Neurocomputing 275, 167–179.

[22] Yuan, M., Wu, Y., Lin, L., 2016. Fault diagnosis and remaining useful life estimation of aero engine using lstm neural network, in: 2016 IEEE international conference on aircraft utility systems (AUS), IEEE. pp. 135–140.

[23] Zerveas, G., Jayaraman, S., Patel, D., Bhamidipaty, A., Eickhoff, C., 2021. A transformer-based framework for multivariate time series representation learning, in: Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining, pp. 2114–2124.

[24] Zhang, J., Wang, P., Yan, R., Gao, R.X., 2018. Long short-term memory for machine remaining life prediction. Journal of manufacturing systems 48, 78–86.

[25] Zhao, R., Yan, R., Wang, J., Mao, K., 2017. Learning to monitor machine health with convolutional bi-directional lstm networks. Sensors 17, 273.