Convolutional AutoEncoders for Anomaly Detection in Semiconductor Manufacturing

Document Version
Author Accepted version

Convolutional AutoEncoders for Anomaly Detection in Semiconductor Manufacturing

Mark Gorman
Intelligent System Research Center
Ulster University
Global Factory IT
Seagate Technology, Derry
mark.gorman@seagate.com

Dr Xuemei Ding
Intelligent Systems Research Center
Ulster University, Derry
x.ding@ulster.ac.uk

Prof. Liam Maguire
PVC Research
Intelligent Systems Research Center
Ulster University, Derry
Lp.maguire@ulster.ac.uk

Prof. Damien Coyle
The Bath Institute for the Augmented Human
University of Bath
Intelligent Systems Research Center
Ulster University, Derry
dhc30@bath.ac.uk

Abstract

Semiconductor manufacturing, characterised by its complex processes, demands efficient anomaly detection (AD) systems for quality assurance. This study extends from previous work utilising unsupervised Convolutional AutoEncoders for AD in Semiconductor batch manufacturing by applying the technique to a novel dataset supplied by a local Semiconductor Manufacturer. Our method uses an approach that employs 1-dimensional Convolutional Autoencoders (1d-CAE) to improve AD performance and interpretability through the numerical decomposition of reconstruction errors. Identifying anomalies this way allows engineering resources to explain anomalies more effectively than traditional methods. We validate our approach with experiments, demonstrating its performance in accurately detecting anomalies while providing insights into the nature of these irregularities. The experiments also demonstrate the impact of training setup on detection capability, outlining an efficient framework for determining an optimal hyperparameter set-up in an industrial dataset. The proposed unsupervised learning approach with AE reconstruction error improves model explainability, which is expected to be beneficial for deployment in semiconductor manufacturing, where interpretable and trustworthy results are critical for solution adoption by process engineering teams.

Keywords – Anomaly Detection, Semiconductor Manufacturing, Convolutional AutoEncoder, Explainable Reconstruction Error

I. INTRODUCTION

Semiconductor manufacturing, evolving rapidly, challenges the limits of traditional Machine Learning (ML) in process control [1]. Industrial practitioners increasingly turn to Deep Learning (DL) for more robust performance in managing complex processes. Fault Detection and Classification (FDC), or Anomaly Detection (AD), is pivotal in identifying irregular equipment conditions using time-series sensor data generated during semiconductor wafer processing. In this context, semiconductor processes, characterised by their multi-re-entrant, batch nature and variability due to consumable parts, demand advanced control mechanisms [2].

Batch manufacturing, with its diverse product mixes and recipe variations, makes maintaining numerous process sensors impractical. The shift towards a more automated and self-regulating AI-driven process control system is critical to augment existing engineering expertise. The increasing data volume and complexity in the semiconductor industry outpace the available resources for effective control and monitoring. The definition of AI in this context covers ML and the more recent DL advances in the area. Traditional ML is where data is manually pre-processed, features engineered and optimal features selected by a human in the loop, with established workflows that heavily leverage empirical knowledge and subject matter expertise.

II. RELATED WORK

In [3], we have demonstrated that the 1-dimensional Convolutional Autoencoders (1d-CAE) approach applied to the Tennesse Eastman process and the LAM 9600-Etcher publically available industry datasets is an effective anomaly detector. However, these datasets are limited when attempting to capture the complexity of modern Semiconductor multi-modal batch processes. Comparisons have been made to show [4] that a Traditional ML approach is viable. However, when a representative feature space is created through the aggregation of time-series signals, information loss is likely, therefore decreasing the capacity of an algorithm to detect potential future anomalies within the aggregated feature space [5]. However, in these instances, the modelled domain space is reduced through data aggregation, feature engineering and feature removal through a selection process. DL has been demonstrated to outperform traditional ML approaches and even human-level performance [6]. Although there are several fields of interest within DL, AutoEncoders (AE's) are effective in unsupervised AD [7] and dimensionality reduction [8], [9]. AE's with Dense layers have been applied for AD [10] in batch manufacturing. However, Dense AE's are rigid in their network architecture, potentially reducing the network's ability to capture time-based feature variation, resulting in less accurate models [11]. An alternative approach is a Convolutional AutoEncoder (CAE), created by stacking several non-linear layers that enable the network to extract hierarchical high-level features. Convolutional layers reduce the feature space through Pooling or increased kernel strides [11] and have been shown to perform well as feature extractors [8]. A 2d-convolutional layer is preferred if data is related across the x-axis and y-axis [12]. Conversely, where inputs have only one axis of information dependence, a 1d-convolutional layer is more

suited to the input data shape [11]. 1d-CAEs are proficient extractors of high-level features along a single input axis [13]. For manufacturing use cases, the 1d-convolutional kernel is applied along the time-dependent axis, with the remaining axis containing input sensors [14]. Maggipinto et al. [15] have implemented a hybrid approach where the 1d-CAE is used as a feature extractor for traditional ML AD algorithms. Zhang et al. [16] adopt 1d-convolutional layers for unsupervised feature extraction for other models. It is observed that 1d-convolutional layers achieve high performance for feature extraction [11] in classification frameworks. A significant limitation of AD in Semiconductor manufacturing with high-yield processes is the lack of labelled data. A limitation of implementing the conventional ML or black box DL techniques is the further abstraction from the raw time-series data that decreases the interpretability of predictions.

III. Proposed Approach

In the following section, we outline the challenges in the scenario captured within the novel dataset, accompanied by various experimental results. The objective is to detect abnormal equipment behaviour by analysing data collected during processing. However, the problem's difficulty compounds in the production environment for the following reasons.

A. Data Availability

The access to practical datasets in the public domain from private manufacturers limits the potential for experimentation. The dataset has 26 sensors from 3 separate equipment groups with 3,645 observations. Within the dataset are 54 synthetically generated fails accompanied by failure labels and the sensors within which the fails occur. The synthetic fails were generated using various magnitudes of either Oscillation, Spike or Step functions from the typical sensor operating point. The mean process duration is 480 samples, with variation in process duration introduced through varied consumable age and equipment performance during regular equipment lifecycles. Introducing this novel dataset to open access attempts to remove the obstacle of limited data.

B. Parallel Processing Paths

The occurrence of multiple tools completing the same process can be referred to as multipath processing and is common in Semiconductor manufacturing to ensure stable cycle time and availability. The qualified equipment is 'matched' in the ideal process operation scenario, meaning their processing input parameters are nearly identical. However, this is unrealistic given the inherent variation introduced by incoming materials, consumable ageing and equipment configuration. The challenge of non-stationarity affects algorithm performance over time and can cause the rate of false positives to increase or require the re-training of models. As processes execute in parallel, the number of samples available for training on a given equipment configuration is sometimes small, forcing the implementation to leverage data from the same process recipe on a different equipment set. The study addresses this through the variation across experiments of the training data composition and impact on reconstruction ability.

C. Configuration Management

Furthermore, in cases with significantly different processing equipment, increased configuration effort is needed in the form of more models, as a single model cannot account for all variations observed at the equipment process level. The required model coverage is high so that the maximum operational value can be extracted from the time series data. The training, deploying and managing 1,000's to 100,000's models in production is challenging. A method to automate the deployment standardisation and optimisation is needed to streamline value capture within Semiconductor Fabs. An established hyperparameter optimisation approach is preferred to random or grid search methods to improve deployment efficiency. The study presents the method of hyperparameter optimisation suited to automated production deployment.

D. Multi-Modal Batch Processing

Fig 1. illustrates a sample of the raw time series process data highlighting that each sensor has a prescribed shape outlining the multi-modal process operating scenario. Each sensor measures a separate process parameter and, therefore, has its scale corresponding to the relative unit of measurement (UoM). UoMs are omitted from the dataset in the interest of anonymity. The multivariate example in this dataset is analogous to most batch-processing semiconductor processes where a prescribed recipe has defined duration and changes in operation function throughout. Examples of this modal change can be observed in Fig. 1 through the difference in the magnitude of the sensors during the execution of the process duration. The obstacle to overcome here is that, in some cases, traditional ML techniques require homogeneity in the data populations, which is not aligned with the data in this scenario. A 1d-CAE has been demonstrated in the related work to be an excellent unsupervised feature extractor, thus allowing the entire time series duration to be modelled within the same training schedule. The ability to train the entire trace with a single model decreases the required configuration count and the various modes of operation within which anomalies can be detected.

E. Prediction Explainability

Decreased transparency in ML and DL models is seen as a barrier to adoption within manufacturing; more specifically, when predictions are not interpretable, the intended DL consumer has little context to evaluate or resolve an identified fault. It is important to note that for interpretability in the
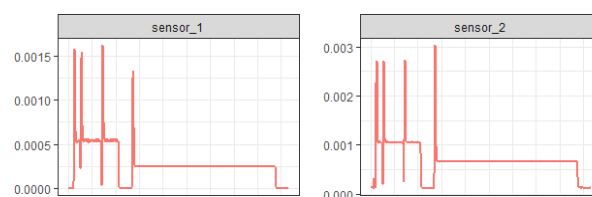

Fig.1 – Raw time series sensor traces from a sample of sensors demonstrating the multi-modal aspect of batch semiconductor processes alongside the various UOMs and shapes that occur

context of Semiconductor FDC, the objective is not to uncover the purpose of each neuron or model parameter concerning the prediction but instead the 'mechanistic explanation' of a network input to output so that users can correlate predictions made onto actions to be taken [17]. The proposed 1d-CAE algorithm implements a reconstruction-based approach where the original input time series trace predictions are point-by-point predictions. The preservation of the actual and predicted time series facilitates the decomposition of error across both the time-dependent and sensor channels to identify sources of error. The decomposition directs engineering resources to potential focus areas within the time series traces that can rectify anomalous behaviour.

IV. EXPERIMENTAL DESIGN

Training the AE on known good process operation, also known as a 'Golden Fingerprint', effectively creates a novelty boundary outside which unknown faults are expected to be detected. The main assumption within this framework is that the 'Golden Fingerprint' is an adequately representative sample that encompasses all 'acceptable'

TABLE I - HYPERPARAMETER SEARCH SPACE FOR THE KERAS-TUNER USING HYPERBAND OPTIMISATION

Parameter	Options
Convolutional Layers	Range 1:5, interval 1
Filter number	Range 5:250, interval 25
Kernel size	Range 5:100, interval 10
Batch Normalisation	Boolean
DropOut	Boolean
DropOut Rate	0.25, 0.5, 0.75
Final Sigmoid Layer Kernel	Range 1:100, Interval 5
Learning Rate	1e-1, 1e-2, 1e-3, 1e-4, 1e-5
Loss Function	MSE, MAE, MSLE
Batch Size	32, 64, 128, 256

process variation. Within future fault situations, an error rate exists large enough to perturb the result outside the defined novelty boundary.

A. Hyperparameter Optimisation

A critical aspect of deploying deep learning models effectively is fine-tuning hyperparameters. Manual hyperparameter tuning can be time-consuming and suboptimal, necessitating the exploration of automated techniques. The Keras Tuner with Hyperband optimisation is a tool for automating the process of hyperparameter tuning in deep learning models [18], specifically designed to work with the Keras deep learning framework. It combines the Keras Tuner, a library for hyperparameter optimisation, with the Hyperband algorithm, an efficient method for exploring hyperparameter configurations [19]. To streamline the model development process in a production manufacturing environment, the Keras-Tuner with Hyperband optimisation is preferred to random or grid search approaches as the process to find optimal configuration setups for large-scale deployments can be automated. Hyperband is an optimisation

algorithm that outperforms traditional methods such as Grid Search and Random Search regarding hyperparameter tuning in machine learning. Table I outlines the search space for the tuner to evaluate. In particular, the parameters were chosen to facilitate the evaluation of filter number, layer number, kernel size and loss function on the modelling architecture given the semiconductor dataset.

B. Experiment Definition

Utilising the Keras-Tuner with Hyperband optimisation pipeline are three experiments within this study. In [3], we have demonstrated that 1d-CAE networks perform well for multivariate manufacturing anomaly detection. The study aims to introduce the dataset and baseline the results using the 1d-CAE framework from a reconstruction perspective. Table II lists the description of the three separate experiments applied to the novel semiconductor dataset. Experiment E0001 is designed to produce results from the standard practice of splitting by time and not considering the impact of equipment mix in the training data. E0002 simulates a scenario where only data is available for one equipment group but predictions are required for other similar, potentially unmatched equipment sets, a form of transfer learning. Finally, E0003 is designed to test the impact of equally distributed training observations by equipment group compared to the mixed distribution seen in E0001. Following the data split, the Hyperband Keras-Tuner is applied to the training data of each given experimental setup. After the tuner trials, the optimal settings of the 10 top hyperparameter setups are selected. These selections are then loaded sequentially to collect training and test data for each modelling setup to determine mean and standard deviation performance for each configuration setup. The reasoning

TABLE II – EXPERIMENTAL DESCRIPTION SHOWING THE BREAKDOWN OF TRAINING AND TEST SPLIT BY EQUIPMENT GROUP

Experiment No.	Observations		Equipment Group	
	Train	Test	Train	Test
E0001	2549	1096	0, 1, 2	0, 1, 2
E0002	1033	1096	0	0, 1, 2
E0003	933	1096	0, 1, 2	0, 1, 2

behind choosing the top 10 is to provide a distribution of performances and the optimal setup parameters for comparison.

C. Reconstruction Error Based Anomaly Detection

Reconstruction error-based anomaly detection is the approach where an algorithm, in this case, an autoencoder, attempts to learn latent representations of the data from which it can reconstruct an output from a given input [3]. The comparison of input and output determines the accuracy of the reconstruction. In cases where the reconstruction error is small, the observation is more likely to be a member of the training data and normal. Conversely, a large reconstruction error suggests that the observation is not similar to the good historical period of operation and is likely anomalous and therefore requiring engineering review. Primarily, the sensor data and associated faults in the main do not have reliable
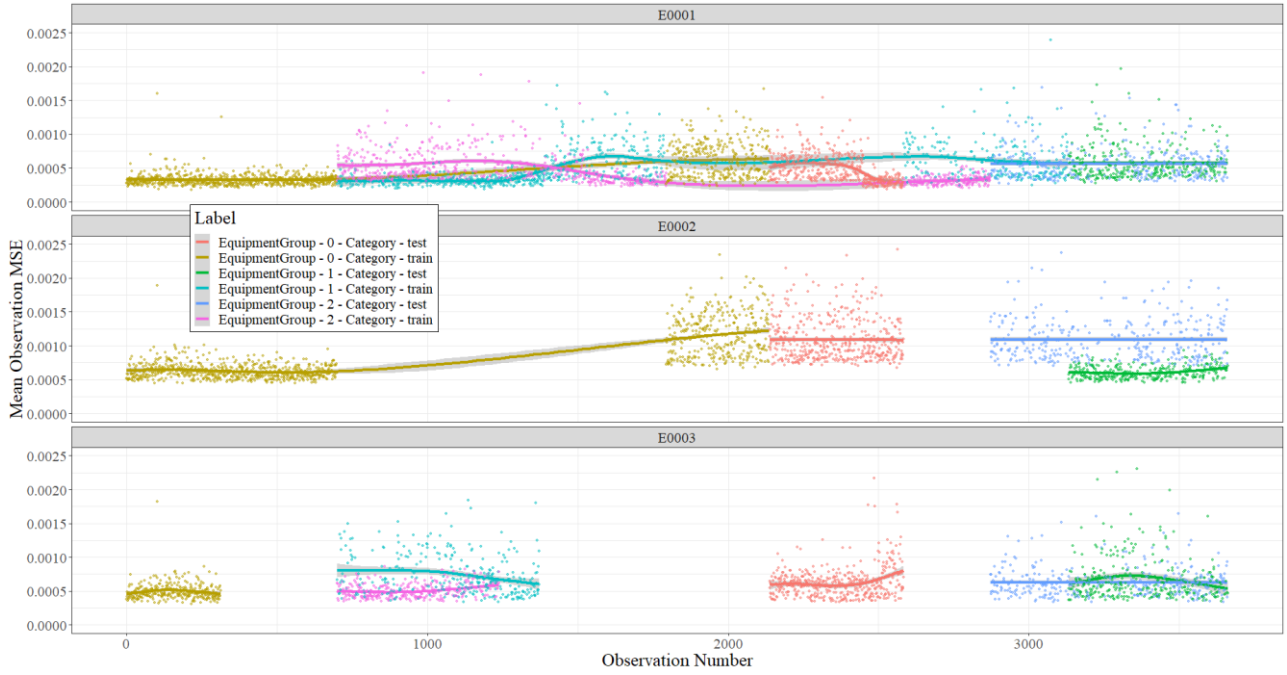
Fig. 2 – Mean Observation MSE by Experiment Number, Equipment Group and Data Category separated by Experiment Number. Smoothed lines of best fit are applied to each group to show a change in Mean Observation MSE by time, indicated by observation number.

labels. The reconstruction error approach allows future unknown fault scenarios to be detected as it would be different from the training data. Secondly, within a manufacturing environment, identifying an anomaly and attempting to provide a reason for the anomalous behavoiour are equally important. Preserving the reconstructions facilitates the error decomposition to determine areas of potential anomalous behaviour.

V. EXPERIMENTAL RESULTS

Table III presents the overall experiment summary of reconstruction ability based on the top 10 hyperparameter configurations identified through the Keras-Tuner Hyperband pipeline. The best-performing reconstruction experiment was E0001, where all three equipment types and all available observations were used for training and testing. Table III demonstrates that E0001 outperformed the other experiments within equipment groups and E0002, with only equipment group 0 generating the lowest reconstruction capability performance. Table III also highlights that E0003

with the balanced equipment group number for training, resulting in a lower standard deviation between the top 10 model hyperparameter configurations. When considering Fig.2, the mean MSE by observation number shows that although E0001 achieves overall lower reconstruction errors, there is an increase in variation in those predictions compared to E0003. Furthermore, the loss in prediction power of the model in E0002 can be seen when comparing the reconstruction error of training and testing observations by the equipment group. Fig.2 also highlights the change in reconstruction error by time and equipment group. E0001 outperforms E0002 and E0003; however, E0002, with the limited training dataset, presents poorer overall results within the equipment groups 0 and 1. As the 1d-CAE approach leverages an unsupervised thresholding mechanism to determine population outliers, assigning a suitable threshold can be challenging. Fig. 3 presents the performance of the top 10 hyperparameter configurations by equipment group, data

TABLE III – THE MEAN AND STANDARD DEVIATION OF TEST DATA MEAN SQUARED ERROR OF RECONSTRUCTION ERROR OF THE TOP 10 HYPERPARAMETER SETTINGS BY EXPERIMENT NUMBER, DATA CATEGORY AND EQUIPMENT GROUP

| Mean of MSE from Reconstruction Errors					Standard Deviation of MSE from Reconstruction Errors			
Equipment Group	**Experiment No**				**Equipment Group**	**Experiment No**		
	E0001	**E0002**	**E0003**			**E0001**	**E0002**	**E0003**
0	0.00048	0.00111	0.00063		**0**	0.00039	0.00026	0.00014
1	0.00057	0.00061	0.00066		**1**	0.00050	0.00007	0.00026
2	0.00056	0.00113	0.00065		**2**	0.00044	0.00024	0.00023
Overall	*0.00054*	0.00095	0.00065		**Overall**	0.00044	0.00019	*0.00021*
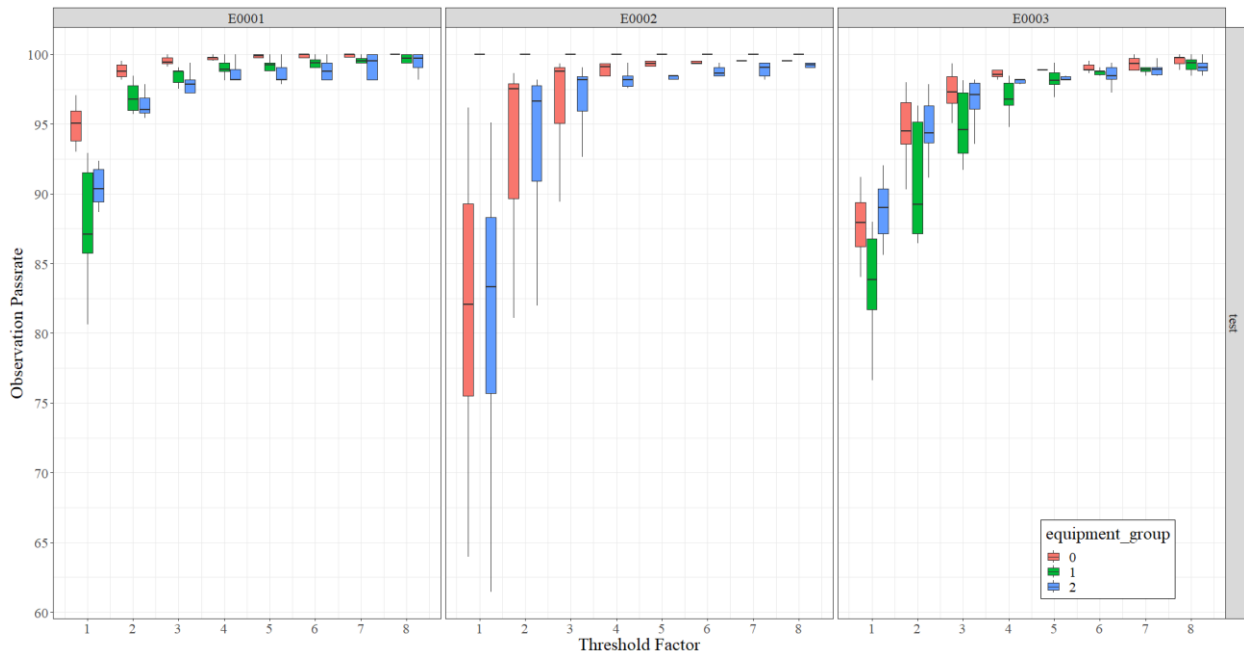
Fig. 3 – Pass rate percentage distribution for top 10 model hyperparameter setups by data category and experiment number, coloured by equipment group.

category and experiment number. The vertical axis corresponds to the pass rate of the observations, with the horizontal axis the thresholding factor applied to the interquartile outlier equation. E0001 outperforms the other setups from a median and standard deviation perspective, with the results from all equipment groups generating results with tighter distributions. Equipment group 1 generates the poorest results for all experimental setups, suggesting an underlying difference between that equipment group population and the others. From an explainability perspective, Fig.4 shows the raw traces of actual input data in green, predicted data in red and the error distribution in black for a sample of observations and sensors. The detailed reconstruction error can be seen in Fig.4, given the multi-modal process for a good observation; in the scenario of anomalous process behaviour, the reconstruction error decomposition can be seen by which sensor and within which process duration. Table IV summarises the optimal hyperparameter setup by percentage occurrence within the top 10 setups by experiment. 90% of optimal models preferred a more shallow network architecture with only a single convolutional layer. Similarly, larger numbers of filters performed better than smaller setups. Both batch normalisation and dropout were not selected as an optimal model architecture, with the most common loss function achieving optimal reconstruction error being MSE. The optimal learning rate was 0.0001 or 0.001 for 100% of the top-performing experimental setups.

VI. Discussion and Conclusion

The results presented in this study offer valuable insights into the performance of various experiments conducted using the 1d-CAE approach for reconstruction and anomaly detection in manufacturing time series data. The experiments were guided by exploring hyperparameter configurations identified through the Keras-Tuner Hyperband pipeline. The

analysis begins by highlighting the best-performing experiment, E0001, which utilised all three equipment types and all available observations for training and testing. E0001 demonstrated superior reconstruction ability compared to other experiments within equipment groups. This underscores the importance of incorporating diverse

TABLE IV – SUMMARY OF OPTIMAL HYPERPARAMETER SETTINGS FOR REPEATED EXPERIMENTS

Setting	Optimal Setting	% of Occurence
Convolutional Layers	1	90
Filter number	105, 155, 205, 355	95
Kernel size	15, 65	75
Batch Normalisation	FALSE	75
Drop Out	FALSE	90
Final Layer Kernel	6, 51	80
Learning Rate	0.0001, 0.001	100
Loss Function	MSE	85
Batch Size	32	85

equipment types and a comprehensive dataset for better reconstruction results. Fig. 3 and Table III further reveals interesting patterns within the experiments. While E0001 outperforms others in terms of reconstruction, E0003, which employed a balanced equipment group for training, exhibited lower standard deviation among the top 10 model hyperparameter configurations. This suggests that achieving consistent results, as seen in E0003, may require careful consideration of equipment group balance in the training dataset. From an explainability perspective, Fig. 4 visually represents the model's raw reconstruction performance. It demonstrates the raw traces of actual input data, predicted data, and error distribution for a sample of observations and sensors. The result allows for a detailed examination of the

reconstruction error, particularly in scenarios involving anomalous process behaviour. Pinpointing errors to specific sensors and process durations is crucial for effectively identifying and diagnosing anomalies. As the 1d-CAE approach employs an unsupervised thresholding mechanism for outlier detection, determining an appropriate threshold is challenging. Fig. 2 and Fig. 3examines the performance of the top 10 hyperparameter configurations based on equipment group, data category, and experiment number. E0001 consistently outperforms other setups in median and standard deviation, suggesting its robustness in outlier detection. However, the poorer results observed in Equipment Group 1 across all experiments hint at underlying differences in this equipment group compared to others, which should be explored further. In Table IV, the optimal network setup demonstrated through the Keras-Tuner is also interesting. The result that fewer layers and more filters yielding improved reconstruction results is counter-intuitive to networks that perform well in image or single mode time series datasets, suggesting that improved feature extraction of increased filter application is preferred to dimensionality reduction of the feature space with deeper networks made up of more layers.
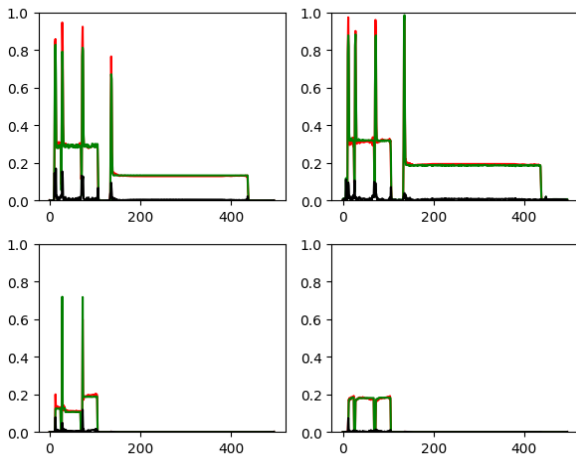

Fig. 4 – Plot of actual (Green) and predicted (Red) with reconstruction error (Black) of time series sensor traces from a sample observation and set of sensors. The y-axis represents the magnitude of the prediction or actual value, and the x-axis

VII. FUTURE WORK

The current experiment setup is limited to the 1d-CAE framework; future experiments should include alternative deep learning architectures such as transformers or Variational Autoencoders (VAE) that could further extract improved representative features from the dataset. Another area of exploration would be the optimisation of reconstruction error thresholding, which, in its current linear form, is not flexible to drift; therefore, a dynamic threshold assignment beyond rigid thresholding should be considered in the future, e.g. an ensemble approach that applies a clustering algorithm to the latent representations of the autoencoder in parallel to the demonstrated thresholding approach.

REFERENCES

[1] Z. Ge, Z. Song, S. X. Ding, and B. Huang, 'Data Mining and Analytics in the Process Industry: The Role of Machine Learning', *IEEE Access*, vol. 5, pp. 20590–20616, Sep. 2017, doi: 10.1109/ACCESS.2017.2756872.

[2] J. Moyne and J. Iskandar, 'Big Data Analytics for Smart Manufacturing: Case Studies in Semiconductor Manufacturing', *Processes*, vol. 5, no. 3, p. 39, 2017, doi: 10.3390/pr5030039.

[3] M. Gorman, X. Ding, L. Maguire, and D. Coyle, 'Anomaly Detection in Batch Manufacturing Processes Using Localized Reconstruction Errors From 1-D Convolutional AutoEncoders', *IEEE Trans. Semicond. Manuf.*, vol. 36, no. 1, pp. 147–150, Feb. 2023, doi: 10.1109/TSM.2022.3216032.

[4] T. Lee and C. O. Kim, 'Statistical comparison of fault detection models for semiconductor manufacturing processes', *IEEE Trans. Semicond. Manuf.*, vol. 28, no. 1, pp. 80–91, 2015, doi: 10.1109/TSM.2014.2378796.

[5] J. Wang, Y. Ma, L. Zhang, R. X. Gao, and D. Wu, 'Deep learning for smart manufacturing: Methods and applications', *J. Manuf. Syst.*, vol. 48, pp. 144–156, Jul. 2018, doi: 10.1016/j.jmsy.2018.01.003.

[6] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, 'Viewpoint: When will ai exceed human performance? Evidence from ai experts', *J. Artif. Intell. Res.*, vol. 62, pp. 729–754, 2018, doi: 10.1613/jair.1.11222.

[7] G. E. Hinton and R. R. Salakhutdinov, 'Reducing the dimensionality of data with neural networks', *Science*, vol. 313, no. 5786, pp. 504–507, Jul. 2006, doi: 10.1126/science.1127647.

[8] J. Y. Xifeng Guo, Xinwang Liu, En Zhu, 'Deep Clustering with Convolutional Autoencoders', *Conf. Pap.*, no. October, pp. 118–125, 2017, doi: https://doi.org/10.1007/978-3-319-70096-0_39.

[9] C. D. Liu Shuo, 'Integrated Autoencoder-Level Set Method Outperforms Autoencoder for Novelty Detection', *Int. Jt. Conf. Neural Netw. IJCNN*, pp. 1–8, 2022, doi: 10.1109/IJCNN55064.2022.9891877.

[10] H. Lee, Y. Kim, and C. O. Kim, 'A deep learning model for robust wafer fault monitoring with sensor measurement noise', *IEEE Trans. Semicond. Manuf.*, vol. 30, no. 1, pp. 23–31, Feb. 2017, doi: 10.1109/TSM.2016.2628865.

[11] S. Chen, J. Yu, and S. Wang, 'One-dimensional convolutional auto-encoder-based feature learning for fault diagnosis of multivariate processes', *J. Process Control*, vol. 87, pp. 54–67, 2020, doi: 10.1016/j.jprocont.2020.01.004.

[12] M. Macas and C. Wu, 'An unsupervised framework for anomaly detection in a water treatment system', *Proc. - 18th IEEE Int. Conf. Mach. Learn. Appl. ICMLA 2019*, no. 15027268, pp. 1298–1305, 2019, doi: 10.1109/ICMLA.2019.00212.

[13] S. Kiranyaz, T. Ince, and M. Gabbouj, 'Real-Time Patient-Specific ECG Classification by 1-D Convolutional Neural Networks', *IEEE Trans. Biomed. Eng.*, vol. 63, no. 3, pp. 664–675, 2016, doi: 10.1109/TBME.2015.2468589.

[14] E. Kim, S. Cho, B. Lee, and M. Cho, 'Fault Detection and Diagnosis Using Self-Attentive Convolutional Neural Networks for Variable-Length Sensor Data in Semiconductor Manufacturing', *IEEE Trans. Semicond. Manuf.*, vol. 32, no. 3, pp. 302–309, Aug. 2019, doi: 10.1109/TSM.2019.2917521.

[15] M. Maggipinto, A. Beghi, and G. A. Susto, 'A deep learning-based approach to anomaly detection with 2-dimensional data in manufacturing', *IEEE Int. Conf. Ind. Inform. INDIN*, vol. 2019-July, pp. 187–192, 2019, doi: 10.1109/INDIN41052.2019.8972027.

[16] C. Zhang, J. Yu, and S. Wang, 'Fault detection and recognition of multivariate process based on feature learning of one-dimensional convolutional neural network and stacked denoised autoencoder', *Int. J. Prod. Res.*, 2020, doi: 10.1080/00207543.2020.1733701.

[17] D. C. Elton, *Self-explaining ai as an alternative to interpretable ai*, vol. 12177 LNAI. Springer International Publishing, 2020. doi: 10.1007/978-3-030-52152-3_10.

[18] K. Team, 'Keras documentation: KerasTuner'. Accessed: Sep. 13, 2023. [Online]. Available: https://keras.io/keras_tuner/#citing-kerastuner

[19] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, 'Hyperband: A Novel Bandit-Based Approach to Hyperparameter Optimization'.