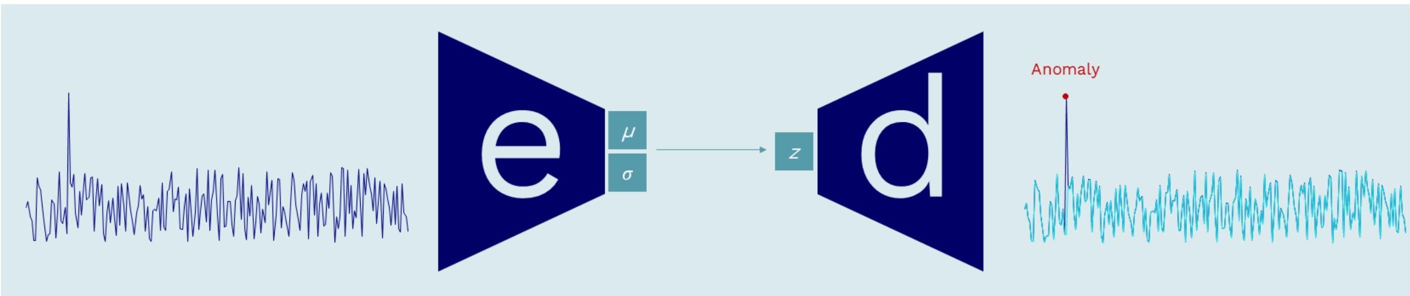
A real-time semi-supervised anomaly detection framework for fault diagnosis of marine machinery systems

Christian Velasco-Gallego[1] (SM), **Iraklis Lazakis**[1] (M)
1. Department of Naval Architecture, Ocean and Marine Engineering, University of Strathclyde, 100 Montrose Street, G4 0LZ, Glasgow, United Kingdom



Maritime companies are currently working to ensure a digital revolution within the maritime industry. Smart maintenance is pivotal in leading this transition, the aim of which is to employ internet of ships to perform real-time data collection through the utilisation of smart sensors, reliable communications, and seamless integration in order to apply predictive maintenance with the application of artificial intelligence and provision of relevant information. Therefore, regular diagnosis and prognosis can be performed to assess the current and future health of machinery to assist in decision-making processes. To enhance the current practices in this area, an innovative anomaly detection framework implementing LSTM-based VAE is proposed to address the challenges identified within this sector. A case study of a diesel generator of a tanker ship is introduced to assess the proposed methodology. Results demonstrated the capability of identifying anomalous instances under various simulated scenarios, thus achieving the maximum precision and recall when the context considers significant anomaly dimensions.

KEY WORDS: anomaly detection, smart maintenance, intelligent real-time systems, LSTM-based VAE, deep learning, marine machinery

NOMENCLATURE

AAKR	Auto Associative Kernel Regression
CBM	Condition-Based Maintenance
CNN	Convolutional Neural Network
DBN	Deep Belief Network
D/Gen	Diesel Generator
DL	Deep Learning
DLSTM	Deep Long Short-Term Memory
EB	Expected Behaviour
FPGA	Field Programmable Gate Array
FWC	Fresh Water Cooling
GAN	Generative Adversarial Network
HPC	High Performance Computing
IoS	Internet of Ships
LO	Lube Oil
LSTM	Long Short-Term Memory
MA	Maintenance Analytics
NN	Neural Network
OCSVM	One-Class Support Vector Machine
RMSE	Root Mean Square Error
RNN	Recurrent Neural Network
SOM	Self-Organizing Map
SPRT	Sequential Probability Ratio Tests
SVM	Support Vector Machine
TPR	True Positive Rate
VAE	Variational Autoencoder
XAI	eXplainable Artificial Intelligence

INTRODUCTION

The prosperity of Condition-Based Maintenance (CBM) in the maritime industry is facilitating data accessibility to enable innovative data-driven strategies, thus enhancing current practices in relation to Operations and Maintenance (O&M) activities within this industrial sector. Accordingly, analysis can empower such strategies by the implementation of Maintenance Analytics (MA) frameworks. As outlined by Karim et al. (2016), and Jasiulewicz-Kaczmarek and Gola (2019), MA is constituted by four interconnected time-line phases (maintenance descriptive, maintenance diagnostic analytics, maintenance predictive analytics, and maintenance prescriptive analytics), the aim of which is the promotion of maintenance actions by improving the understanding of data and information. This inquiry is centred in the maintenance diagnostic analytics phase, which aims to determine the current health of marine machinery. Such a phase is usually established by three consecutive steps: 1) fault detection, 2) fault isolation, and 3) fault identification. The scope of this study targets the first step, fault detection, through the implementation of anomaly detection.

Anomaly detection aims to detect data patterns that deviate significantly from normal operation behaviour. Its implementation has been identified as being of paramount importance due to its extensive application domains (Erhan et al., 2021), such as in manufacturing (Ducharlet et al., 2020; Alaoui-Belghiti et al., 2019; Morariu et al. 2020), railway (Oliveira et al., 2019; Xue et al., 2019; Shi et al., 2019), and aerospace (Roy et al., 2018; Li et al., 2019; Imbassahy et al., 2020). If the maritime domain is analysed, a total of 5 publications related to fault diagnosis of marine machinery can be identified.

Although Deep Learning (DL) methodologies are considered a current trend in such a matter, they are not yet widely established in the maritime industry. An indicator of this is the lack of publications in relation to these techniques, as only 1 of the identified 5 considered DL approaches. To contribute to the implementation of smart intelligent systems within the maritime sector, this inquiry presents an analysis of Deep Long Short-Term Memory (DLSTM)-based Variational Autoencoder (VAE) Neural Network (NN) for anomaly detection performance. Due to the challenges in obtaining a balanced data set with fault data, diverse scenarios are simulated. Accordingly, a case study is introduced to assess the model performance. Specifically, the parameter winding temperature collected from a diesel generator (D/Gen) of a tanker ship is considered.

The following paragraphs are structured as follows. Section 2 presents a literature review in the form of a bibliometric analysis. Section 3 describes the proposed methodology. Section 4 reflects on the results obtained after implementing the proposed methodology through a case study. To finish, in Section 5, the conclusions are presented.

LITERATURE REVIEW

Anomaly detection is of the utmost importance in various sectors due to its capability to identify data patterns that deviate significantly from normal operation conditions. It is not surprising therefore that there are more than 40,000 results when searching papers about anomaly detection in Scopus database. Indeed, if trends are analysed, it is expected that academic publications pertaining to anomaly detection will only continue to increase, and swiftly. Hence, the analysis of each publication individually to acquire knowledge about this topic is unfeasible, which indicates the necessity of complementing regular literature reviews with data-driven review approaches. In this study bibliometrics is implemented, as it has the capability of introducing systematic, transparent, and reproducible review processes that hinge on statistical measurements (Aria and Cuccurullo, 2017). Specifically, the bibliometric analysis of this study is performed to identify trends in research developed in relation to anomaly detection in the maintenance analytics context.

A total of 90 publications are objects of study. The analysis of the most cited articles demonstrated that there is no clear algorithm that outperforms all possible scenarios that can be considered in relation to the system and the type of fault being analysed. Deep Learning (DL) methodologies are gaining attention within recent years, thus indicating their applicability as anomaly detection approaches. However, there are various challenges that are yet to be tackled in relation to these algorithms, an example of which is the lack of trust in "black-box" models by the industry. Although known for their undeniable lack of transparency, there is no doubt whatsoever about their extraordinarily accurate predictions. Therefore, the incorporation of explainable artificial intelligence models is of paramount importance to guarantee the enhancement of O&M activities while ensuring transparency, interpretability, and explainability. Another relevant challenge is the lack of available data, as the industry is exceedingly reserved due to the sensitive information that can be extracted. Therefore, some data, such as fault data, are deeply laborious to obtain, significantly slowing the research process. Consequently, the cooperation between industry and academia is of preeminent importance to make data available, thus facilitating the implementation of DL methodologies. In addition, as Internet of Ships (IoS) is in its infancy, there is a lack of data quality due to unreliable outcomes caused by certain anomalies and missing values originating from device failure, network collapse, and human error (Balakrishnan and Sangaiah, 2018; Izonin et al., 2019; Noor et al., 2014). Accordingly, the adequate implementation of data pre-processing steps, such as data imputation, is essential to guarantee reliable data-driven models. Although data imputation is a compelling pre-processing step that has gained recent popularity, there is a lack of formalisation and analysis, thus far, within the maritime industry (Velasco-Gallego and Lazakis, 2020; Cheliotis et al., 2019). In accordance with this aspect, the deployment of these novel models within the maritime industry is also yet to be adequately formalised.

Regarding the methodologies implemented, One-Support Vector Machine, Decision Trees, Neural Networks, Auto Encoders, Feature Extraction, and clustering algorithms were the most applied within the studies. Specifically, deep learning and support vector machines prevail among other unsupervised and semi-supervised algorithms. Bearings' datasets, Structural Health Monitoring (SHM) data, industrial and manufacturing equipment sensor data are the major typologies of datasets in which state-of-the-art anomaly detection approaches have been modelled for fault diagnosis.

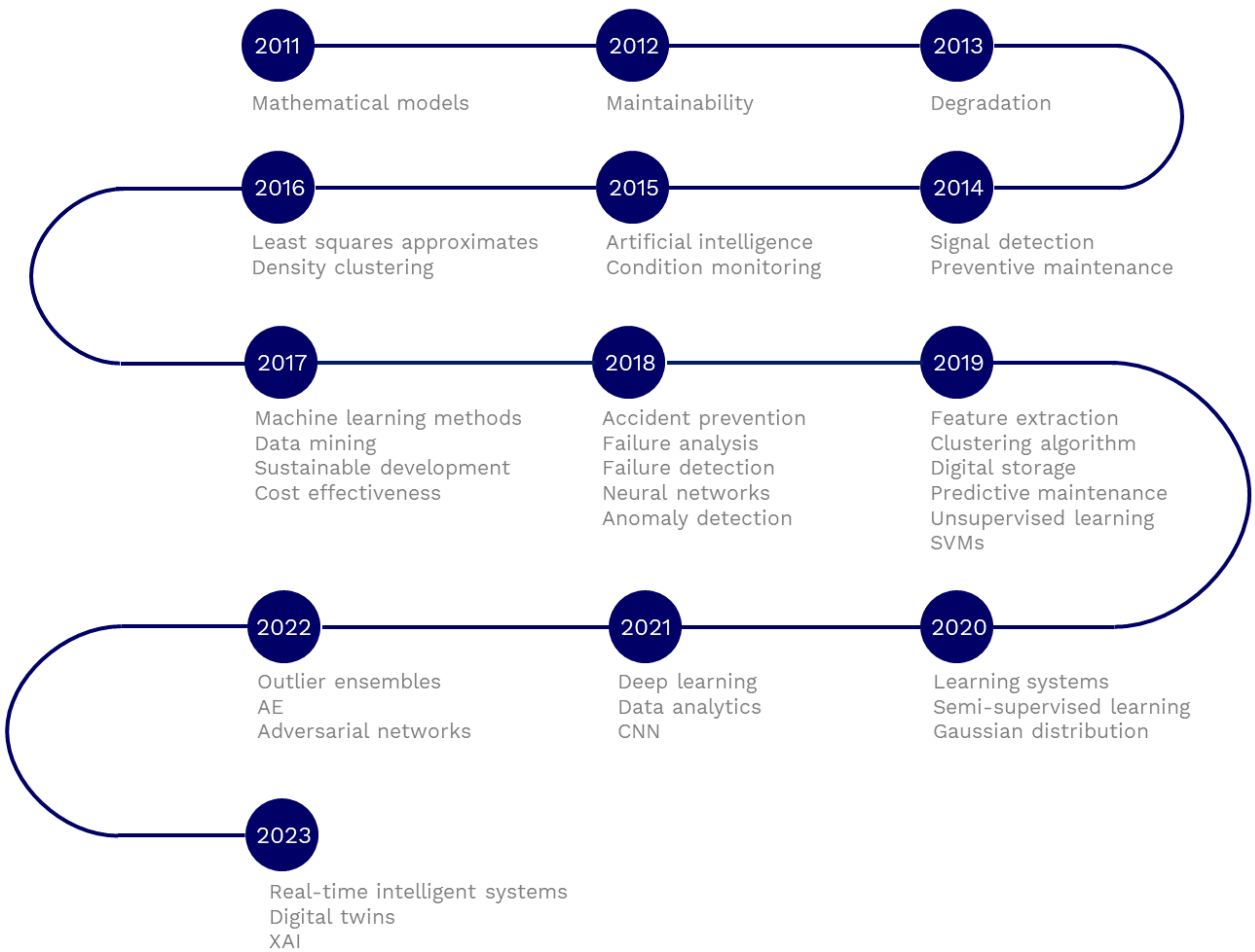


Fig. 1. Past and present trends and future perspectives of anomaly detection.

Past and present trends as well as future perspectives are outlined in Fig. 1. based on the results obtained from applying the bibliometric analysis. Although predictive maintenance is not a new topic, its importance has not gained significant attention from 2017 onwards. The application of innovative technologies in the maintenance context has demonstrated its capabilities to enhance operations and maintenance practices, reduce costs, extend equipment lifetime, improve safety, protect the environment, and ensure quality.

Prior investigations referred to model methods based on the understanding of the system dynamics. However, due to an increase of data availability and quality, data-driven models, which focus on automatically learning the characteristics of the analysed systems from the monitored data, were introduced as promising approaches to be implemented for fault diagnosis. Therefore, machine learning methodologies started to be the focus of research. Least squares approximation, clustering algorithms, one-class support vector machines, and Gaussian distribution-based machine learning models were highlighted. Although studies in relation to machine learning models are still consistent, a transition from such models to deep learning methodologies can be perceived over the last few years.

Deep learning methodologies have demonstrated their capability to automatically adapt to different typologies and complex datasets, and the provision of accurate predictions. Furthermore, feature engineering and data labelling, which are concepts especially complex to address in the maritime industry, are not required. Despite their undeniable results, various challenges are yet to be addressed, including: the need for large volumes of data, the lack of flexibility, the risks of obtaining an overfitted model, and the consideration of such methodologies as "black-box" models due to their lack of transparency. Therefore, it is expected that further research will encompass strategies that address such challenges and promote the stabilisation of deep learning methodologies within the maintenance analytics context. Specifically, eXplainable Artificial Intelligence (XAI) is one of the concepts that is expected to provide promising results in relation to fault diagnosis to facilitate more explainable models whilst also ensuring high performance and provide further understanding in relation to the models' behaviour.

Additionally, while Digital Twins is not a new concept, it is expected to achieve its analysis to address fault diagnosis challenges within the next years. By developing a digital replica of the system being analysed, it is possible to further assess the possible degradation scenarios and the performance of anomaly detection models without the necessity of creating physical prototypes to reproduce these anomalous scenarios. The integration of real-time intelligent systems into digital twins is expected to revolutionise smart maintenance.

If the maritime sector is only considered, a total of 5 articles were identified. Although all the methodologies demonstrated their accuracy in the case studies implemented, most of them do not provide evidence that such methodologies can be implemented in real time. In addition, most of the analysed frameworks do not consider the characteristics of time-series data, albeit the features considered in the case studies are indexed by time. Consequently, methodologies that apply distance-based approaches may yield inaccurate results, as anomalous data points in time-series data may be considered as normal data points since its numerical value is within normal operational thresholds.

In relation to the leading industrial sectors that provide comprehensive studies of anomaly detection techniques, such as the wind energy and manufacturing sectors, it can be determined that the maritime industry is far behind in this respect. As previously stated, DL methodologies are established as the current technological trends for guaranteeing the adequate detection of anomalies and implementation of fault detection. However, only one study implemented these types of methodologies within the maritime industry. Furthermore, some authors even stated that the implementation of such approaches is not recommended due to its level of complexity, computational time, and transparency. Therefore, further studies in this context need to be applied to incorporate explainable artificial intelligence models to ensure the enhancement of operations and maintenance activities within the maritime industry.

Analogous to Fig. 1, past and present trends along with future perspectives in relation to anomaly detection techniques applied within the maritime industry based on the results obtained in the bibliometric analysis are outlined in Fig 2. As previously stated, the studies in relation to unsupervised and semi-supervised anomaly detection techniques for fault diagnosis within the maritime industry is not significant. However, a marked increase in the number of publications has been perceived recently due to promising opportunities coming to the fore in this industrial sector. These opportunities include: 1) the consideration of digital data standards, 2) the increase of data quality and accessibility due to the implementation of smart sensors and condition monitoring, 3) the development and operations of autonomous ships, 4) the increased interest in system health monitoring and management tools, 5) the possibility of implementing real-time intelligent systems due to advancements in communication technologies in ships, and 6) the necessity to guarantee environmental management and resource efficiency. The evolution perceived within the maritime industry is analogous to the one identified from the bibliometric analysis. Initially, model-based methods were widely applied. As a result of an increase in data accessibility, data-driven models started to be further analysed. More precisely, the following machine learning approaches were highlighted: 1) Auto Associative Kernel Regression, 2) Kernel Ridge Regression, and 3) One-class support vector machine. Subsequently, due to the pre-processing complexities of raw sensor data obtained from marine machinery (e.g., detection of operational states, unlabelled data, data imbalance, and feature engineering) unsupervised learning approaches started to be analysed extensively. Therefore, a transition from machine learning, which requires further pre-processing, to deep learning methodologies was distinguished, as such methodologies demonstrated promising results in terms of high performance in analogous industrials sectors, examples of which are the railway, manufacturing, and aerospace sectors. However, the implementation of deep learning methodologies for fault diagnosis in the maritime industry is still inconsistent. An indicator of which is the lack of analysis in relation to such approaches, as only variational autoencoders have been analysed for this matter. Further analysis is required in other methodologies, such as Convolutional Neural Networks (CNNs), Long Short-Term Memory (LSTM) networks, Deep Belief Networks (DBNs), Generative Adversarial Networks (GANs), and Self-Organizing Maps (SOMs). Moreover, challenges arise with the implementation of deep learning. Examples of which include 1) the computational costs, 2) the need of implementing High Performance Computing (HPC) platforms, 3) the development of cost-efficient deep learning frameworks, and 4) the requirement of ensuring high performance while providing further understanding in relation to the models' behaviour. Consequently, concepts such as explainable artificial intelligence models are expected to also be examined. In addition, due to the provision of communication advancements and the necessity to guarantee the security of autonomous and unmanned ships, real-time intelligent systems will be critical in this context, which will derive from the integration of these systems into digital twins.

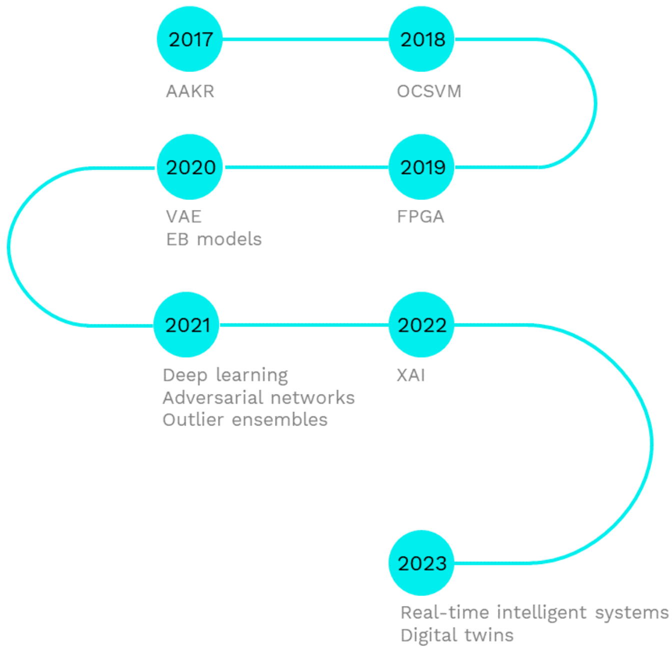


Fig. 2. Past and present trends and future perspectives in relation to anomaly detection techniques applied within the maritime industry.

To collaborate in the development of a real-time MA tool within the maritime industry, this study presents a LSTM-based VAE to address the anomaly detection task of marine systems. Although analogous studies have been performed in this context, there is no evidence that such an approach has been implemented in the maritime industry to the best of the author's knowledge. This novelty is complemented by the possibility of the methodology to be implement in real-time to assist instant decision-making strategies.

METHODOLOGY

The proposed methodology is graphically represented in Fig. 3. The first step refers to data pre-processing, which is of preeminent importance due to the characteristics of the data set. Subsequently, LSTM-based VAE is applied as a deep generative model. Accordingly, the reconstruction error between the input data and the generated data is estimated to determine the anomalousness of each instance, thus labelling the behaviour identified at each time step. To adequately assess the performance efficiency of such a methodology, an evaluation phase is also applied.

Pre-processing

To adequately apply pre-processing, a previous data understanding phase needs to be applied to establish the steps required based on the characteristics of the data set.

A frequent challenge to be addressed when dealing with sensor data of marine systems is the identification of steady states, as various typologies of operational and non-operational states may be consistently present. Therefore, those states that differ from the operational steady states need to be adequately identified and discarded. For this inquiry expert knowledge is applied. Subsequently, data imputation is performed by applying the methodology expressed in Velasco-Gallego and Lazakis (2020).

The successive step applied is data denoising, as time series data are susceptible to containing high noise. Moving average is implemented as the denoising step, as this pre-processing step has demonstrated its capability to enhance the model performance in the analysis performed by Velasco-Gallego and Lazakis (2021). Additionally, input data normalisation is applied to yield values between -1 and 1. The time series is successively sectioned into sequences by applying the sliding window algorithm. To finalise the pre-processing step, the data are split into training, validation, and test sets to avoid model overfitting.

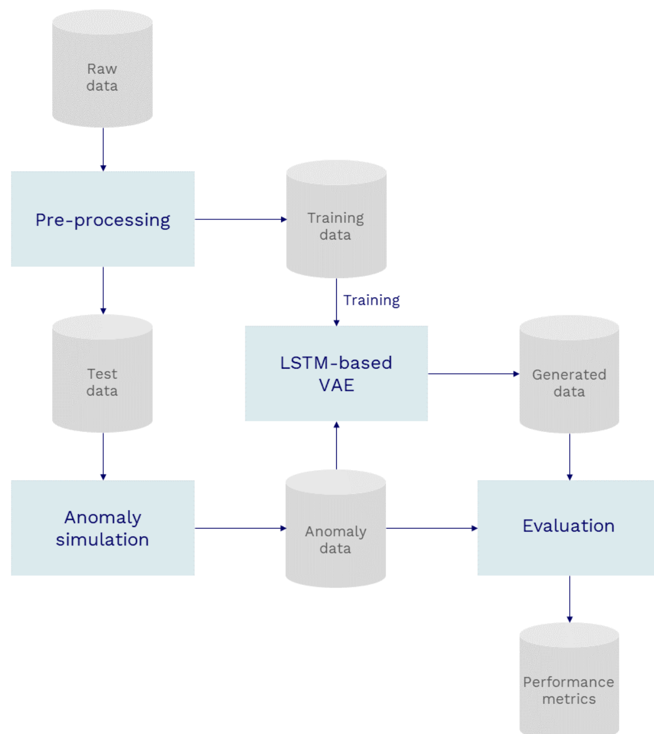


Fig. 3. Graphical presentation of the proposed methodology.

LSTM-based VAE

VAE, developed by Kingma and Welling (2013), is a generative algorithm capable of modelling the distribution of the data. This is a modification of an autoencoder that learns the parameters of a probability distribution. The model is constituted by a probabilistic encoder, which aims to learn both how to reduce the input dimensions and compress the inputs into an encoded representation. This compressed state, a.k.a. latent space representation, presents the lowest possible dimensions of the inputs. Subsequently, the decoder is utilised to learn how to reconstruct the data contained in the latent space representation to reproduce the inputs as analogously as possible. The architecture of such a model is described in Fig. 4.

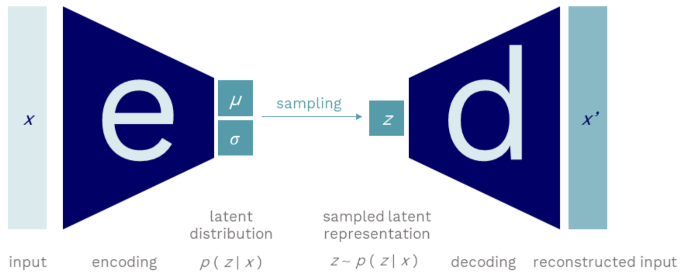


Fig. 4. VAE architecture.

The loss function being minimised is constituted by the reconstruction loss, which aims to ensure the efficient performance of the encoder-decoder arrangement, and the regularisation loss. The latter is determined by estimating the Kullback-Leiber divergence between the approximate posterior and prior latent variable *z*.

To consider the temporal dependences of sensor data of marine systems, the VAE approach is combined with LSTM in both the encoder and the decoder. LSTM is a type of Recurrent Neural Network (RNN) introduced by Hochreiter and Schmidhuber (1997) that learns long-term dependencies. As described in Fig. 5, the core component of such a network is the memory cell, which consists of a cell state vector and gating units, the latter regulating the information flow into and out of the memory, to maintain its state over time.

This step is performed by the implementation of the Python libraries Tensorflow and Keras.

Anomaly simulation

Assuming that anomalous instances are rarely presented and differ significantly from normal operational steady states, these anomalies are poorly reconstructed, thus being accessible to distinguish as anomalous instances.

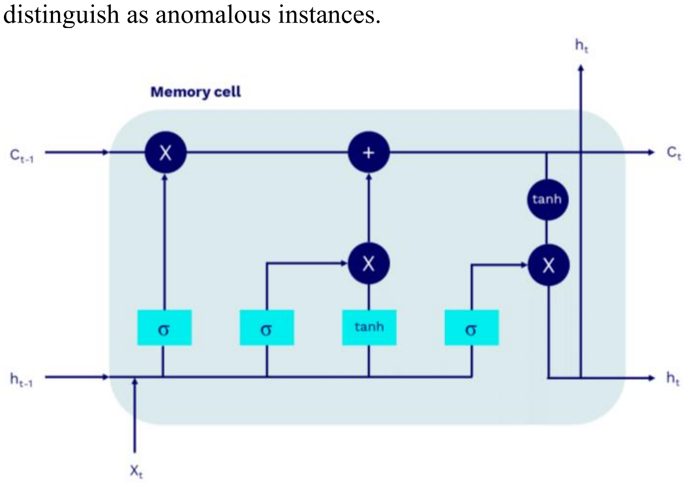


Fig. 5. LSTM cell architecture.

Furthermore, it is considered that all the raw data set refers to normal operational instances, and thus a semi-supervised anomaly detection approach is being examined in this inquiry. Therefore, to adequately assess the proposed methodology for performing anomaly detection tasks, anomaly instances are simulated at random. When expressly analysing time series, the anomalies observed can be categorised in three distinct types. The anomaly is identified as a point anomaly when a unique instance is anomalous in relation to the remaining instances of the data set. Conversely, contextual anomalies are considered when the abnormality is context specific. The final group is noted as collective anomalies when the anomalies are detected by considering a set of instances within the data set.

In this inquiry the first typology, point anomalies, are assessed. To that end, these are simulated by specifying the percentage of anomalies to be included in the data set and the dimension of such anomalies.

Evaluation

Two metrics are utilised to assess the performance of the proposed methodology as a semi-supervised anomaly detection technique.

The first estimated metric is the recall or True Positive Rate (TPR), which determines the actual percentage of anomalies adequately identified (Eq. 1).

$$TPR = \frac{TP}{TP + FN}, \qquad \text{(Eq. 1)}$$

where *TP* are the anomalies adequately classified, whereas *FN* are those anomalies that have not been identified as actual anomalies.

The second metric estimated is the precision or Positive Predictive Value (PPV), which determines the percentage of anomalies identified in relation to the total of instances identified as anomalies (Eq. 2).

$$PPV = \frac{TP}{TP + FP}, \qquad \text{(Eq. 2)}$$

where *FP* is the total of normal instances identified as anomalies.

RESULTS

Having explored the methodology being analysed as a univariate semi-supervised anomaly detection technique, a case study is introduced to assess its performance. As such, the winding temperature parameter collected from a diesel generator of a tanker ship is considered.

The winding temperature is being analysed due to the possible overheating problems that may occur. If the alternator overheats, for instance, the windings could burn out, thus altering their insulating properties. Moreover, other anomalous behaviours may occur in relation to the sensor being utilised, such as problems with the communications, sensor's failure, or human error. Consequently, the monitoring of such a parameter is considered of crucial importance. However, although this inquiry only analyses such a parameter due to the papers' limitations, this approach can be applied for any parameter (e.g., diesel generator power, turbocharger exhaust gas outlet temperature, and cooling air temperature).

The analysed parameter has been collected in a 1-minute frequency and includes a total of 10,000 instances. Fig. 6 represents graphically the time series data of such a parameter prior to and post the application of denoising. The descriptive statistics is also presented in Table 1.

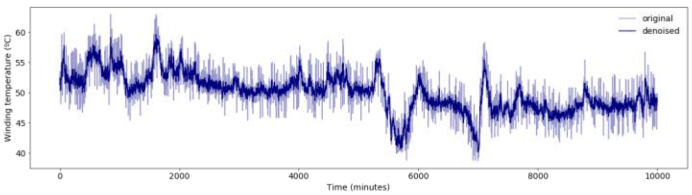


Fig. 6. Time series plot of the diesel generator winding temperature.

Table 1. Descriptive statistics of the monitored parameter.

Count	10,000
Mean	50.0
Std.	3.3
Min.	38.7
25%	47.8
50%	50.2
75%	51.9
Max.	62.9

Once the pre-processing step is finalised, the LSTM-based VAE is applied. After testing various architectures, the LSTM of the encoder is formed by 2 layers (128, 64) and `tanh` activation function. Analogously, the decoder is constituted by 2 layers (64, 128). The ratio of the training set has been set to 0.8 (0.2 of which refers to the validation set) and the test set to 0.2. Adam optimizer has been applied to compile such a model. Subsequently, the model has been trained, setting the number of epochs to 100 and the batch size to 5.

To assess the anomaly detection performance of the model, four scenarios have been considered based on the dimensions of the point anomalies. The value of these, which were selected at random, have been increased by a) 3%, b) 5%, c) 7.5%, and d) 10% of their actual values. Fig. 7 describes both the simulated time series and the reconstructed time series for each scenario. As observed, the reconstruction error is higher when the instances are more anomalous. Consequently, the reconstruction error is greater when the actual value has been increased by 10% rather than the remaining analysed scenarios. Such an outcome can be also perceived in Fig. 8, in which the maximum RMSE error is achieved when the context *d* (anomaly increment ratio of 10%) is considered. Likewise, the reconstruction performance can also be observed in Fig. 9, in which the anomalous instances can be easily clustered for contexts *b*, *c*, and *d*. However, if context *a* (anomaly increment ratio of 3%) is analysed, it can be perceived that the reconstruction error of the normal and anomalous instances does not differ significantly. Therefore, for such a context it is possible that the anomaly detection framework performance is not as efficient as it is expressed in the remaining contexts.

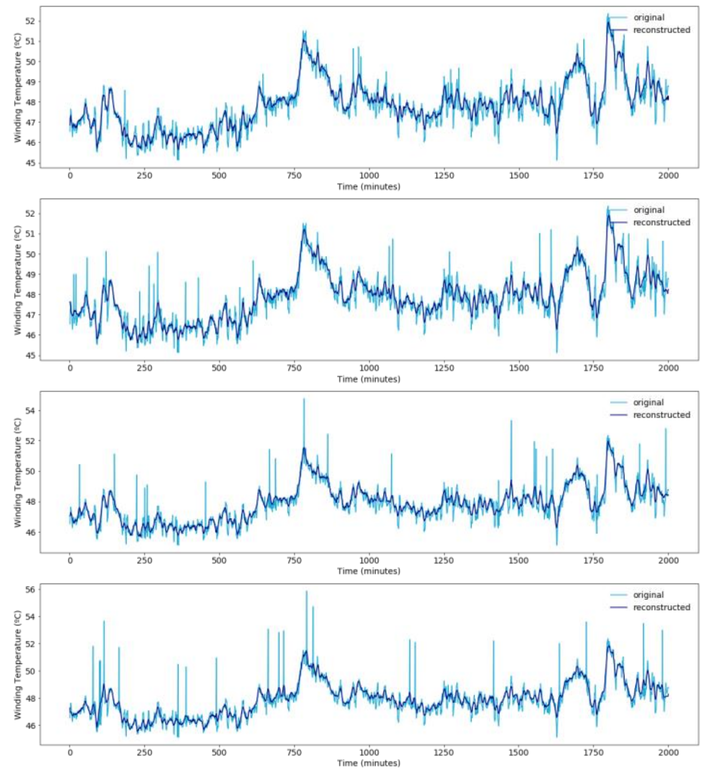
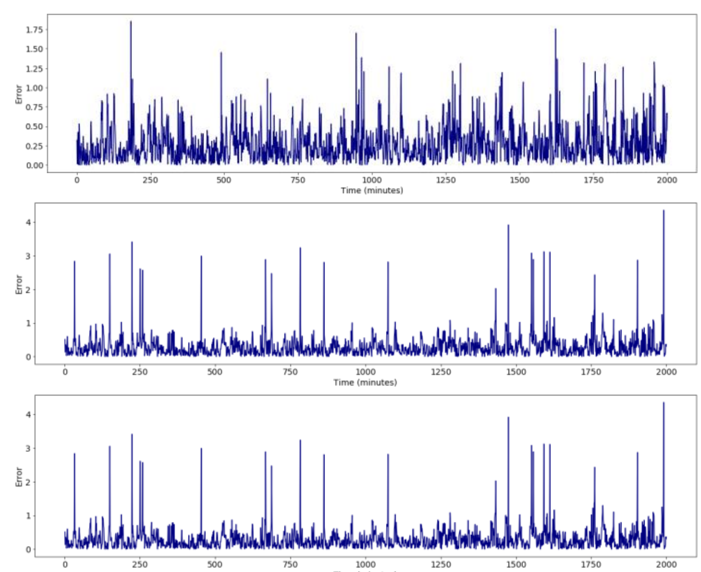


Fig. 7. Reconstructed and original time series with anomaly increments of a) 3%, b) 5%, c) 7.5%, and d) 10%.



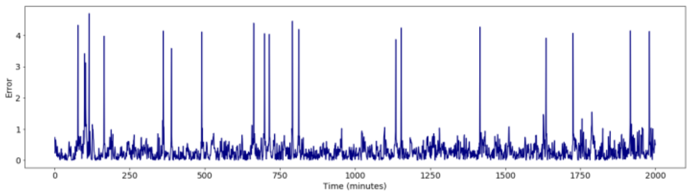


Fig. 8. Individual RMSE for each instance of the simulated time series with anomaly increments of a) 3%, b) 5%, c) 7.5%, and d)10%.

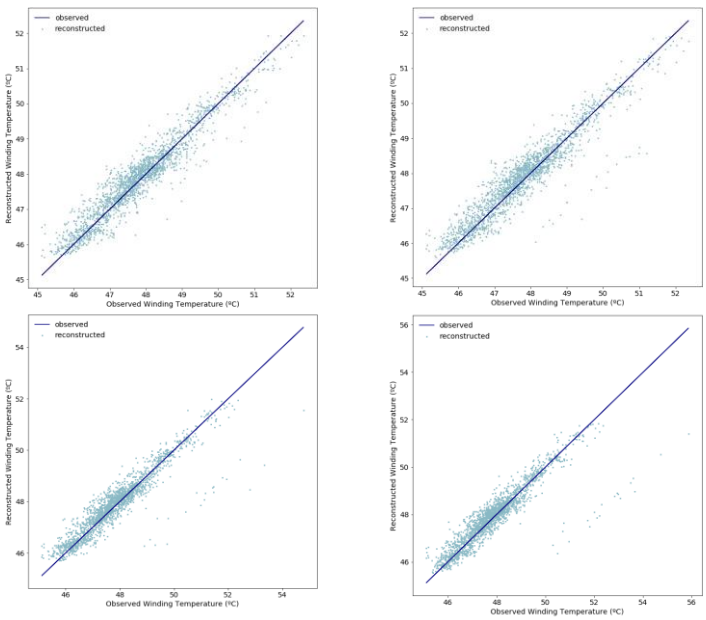


Fig. 9. Observed and reconstructed time series instances with anomaly increments of a) 3%, b) 5%, c) 7.5%, and d) 10%.

To adequately determine the performance of the proposed methodology, the precision and recall metrics have been estimated for a range of anomaly thresholds. These thresholds have been determined based on the minimum and maximum error identified in each context. The results are described in Fig. 10.

As can be perceived in the preceding results, the performance of the proposed model depends on the dimension of the anomaly instance. Therefore, when context *d* is considered, the different anomalies can be identified more efficiently than when the context *a* is analysed. Precisely, if context *a* and *b* are analysed, the anomalies cannot be fully determined unless either diverse false negatives or false positives are recognised. This may indicate that the instances have not been properly labelled, thus utilising anomalous instances in the training process. Additionally, it is also possible that the model overfits and was able to learn the high noise that time series data usually contains. However, the assumption more accepted for this inquiry is that the incremented ratio utilised in some of the contexts is not significant, and therefore these 'anomalous' values do not deviate reasonably from normal operation behaviour. Hence, the proposed model can effectively identify point anomalies that sensor data of marine machinery systems contain, although the performance of such a model relies on the dimensions of the anomalies to be distinguished.

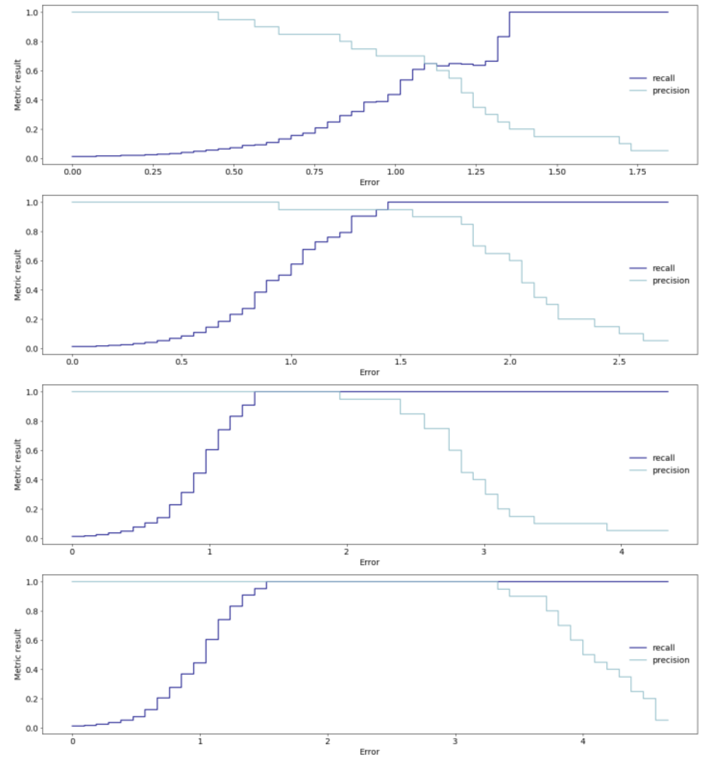


Fig. 10. Recall and precision estimates in relation to the selected threshold for the time series with anomalies of a) 3%, b) 5%, c) 7.5%, and d) 10%.

To guarantee the enhancement opportunities based on the results obtained from this inquiry, the following work guidelines are considered for further research: 1) the analysis of optimisation techniques to adequately select the architecture of the deep neural network, 2) the study of XAI to provide further understanding in relation to the models' behaviour, 3) the addition of more sophisticated methods to estimate the anomaly score, and 4) the consideration of further anomaly contexts (e.g., simulation of contextual anomalies).

CONCLUSIONS

By enhancing data accessibility, the implementation of data-driven models has been possible to empower strategies in relation to O&M activities. The implementation of anomaly detection frameworks is but one example. Anomaly detection techniques, the aim of which is the detection of data patterns that deviate significantly from normal operation behaviour, have been fundamental in sectors such as manufacturing, railway, and aerospace. However, when the maritime industry is considered, anomaly detection is, as yet, a nascent practice. Therefore, further efforts need to be applied to implement innovative and collaborative tools that will facilitate the application of smart maintenance in this industrial sector.

Accordingly, an analysis of DLSTM-based VAE was applied to assess its performance in semi-supervised anomaly detection tasks. In particular, the developed methodology was implemented as a univariate technique. To highlight the adequate performance of such an approach, a case study has been performed, in which the winding temperature parameter collected from a D/Gen of a tanker ship was considered. Results demonstrated the capability of the DLSTM-based VAE model to identify point anomalies that sensor data of marine machinery may contain. However, as expected, the performance of such an approach hinges on the dimensions of the anomalies to be identified, the utilisation of adequate pre-processing steps to ensure data quality, and the optimal definition of both the architecture and the training process of the model.

To continue collaborating in the development of a maintenance analytics tool to establish smart maintenance within the maritime industry, various enhancement opportunities have been identified and will be addressed in further research. This may be rooted in adding more sophisticated method to estimate anomaly scores, and the consideration of further anomaly contexts.

REFERENCES

Alaoui-Belghiti A., Chevallier S., Monacelli E. Unsupervised Anomaly Detection Using Optimal Transport for Predictive Maintenance. Artificial Neural Networks and Machine Learning - ICANN 2019: Text and Time Series. ICANN 2019. Lecture Notes in Computer Science 11730 (2019): 686-697, doi: https://doi.org/10.1007/978-3-030-30490-4_54.

Aria M., Cuccurullo C. Bibliometrix: An R-tool for comprehensive science mapping analysis. Journal of Informetrics 11, (2017): 959-975, doi: https://doi.org/10.1016/j.joi.2017.08.007.

Balakrishnan S. M., Sangaiah A. K. Chapter 6 - aspect oriented modeling of missing data imputation for Internet of Things (IoT) based healthcare infrastructure. Intelligent Data-Centric Systems (2018): 135-145, doi: https://doi.org/10.1016/B978-0-12-813314-9.00006-2.

Cheliotis M., Gkerekos C., Lazakis I., Theotokatos G. A novel data condition and performance hybrid imputation method for energy efficient operations of marine systems. Ocean Engineering 188 (2019): 1-14, doi: https://doi.org/10.1016/j.oceaneng.2019.106220.

Ducharlet K., Travé-Massuyès L., Le Lann M. V., Miloudi Y. A Multi-phase Iterative Approach for Anomaly Detection and Its Agnostic Evaluation. Trends in Artificial Intelligence Theory and Applications. Artificial Intelligence Practices. IEA/AIE, (2020): 505-517, doi: https://doi.org/10.1007/978-3-030-55789-8\44.

Erhan L., Ndubuaku M., Di Mauro M., Song W., Chen M., Fortino G., Bagdasar O., Liotta A. Smart anomaly detection in sensor systems: A multi-perspective view. Information Fusion 67 (2021): 64-79, doi: https://doi.org/10.1016/j.inffus.2020.10.001.

Hochreiter S., Schmidhuber, J. Long short-term memory 9(8) (1997): 1735-1780, doi: https://doi.org/10.1162/neco.1997.9.8.1735.

Imbassahy D. W. D., Marques H. C., Rocha G. C., Martinetti A. Empowering Predictive Maintenance: A Hybrid Method to Diagnose Abnormal Situations. Appl. Sci. 10 (2020): 1-27, doi: https://doi.org/10.3390/app10196929.

Izonin I., Kryvinska N., Tkachenko R., Zub K. An approach towards missing data recovery within IoT smart system. Procedia Computer Science 155 (2019): 11-18, doi: https://doi.org/10.1016/j.procs.2019.08.006.

Jasiulewicz-Kaczmarek M, Gola A. Maintenance 4.0 Technologies for Sustainable Manufacturing - an Overview. IFAC-PapersOnLine 52:10 (2019): 91-96, doi: https://doi.org/10.1016/j.ifacol.2019.10.005.

Karim R, Westerberg J, Galar D, Kumar U. Maintenance Analytics - The New Know in Maintenance. IFAC-PapersOnLine 49:28 (2016): 214-219, doi: https://doi.org/10.1016/j.ifacol.2016.11.037.

Kingma D. P., Welling M. Auto-encoding variational Bayes. arXiv.org (2013): 1-14, https://arxiv.org/abs/1312.6114.

Li X., Zhang T., Liu Y. Detection of Voltage Anomalies in Spacecraft Storage Batteries Based on a Deep Belief Network. Sensors 19 (2019): 1-21, doi: https://doi.org/10.3390/s19214702.

Morariu C., Morariu O., Răileanu S., Borangiu T. Machine learning for predictive scheduling and resource allocation in large scale manufacturing systems. Computers in Industry, (2020): 1-13 doi: https://doi.org/10.1016/j.compind.2020.103244.

Noor N. M., Abdullah M. M. A. B., Yahaya A. S., Ramli N. A. Comparison of linear interpolation method and mean method to replace the missing values in environmental data set. Materials Science Forum 803 (2014): 278-281, doi: https://doi.org/10.4028/www.scientific.net/MSF.803.278.

Oliveira D. F. N., Vismari L. F., de Almeida J. R., Cugnasca P. S., Camargo J. B., Marreto E., Doimo D. R., de Almeida L. P. F, Gripp R., Neves M. M. Evaluating Unsupervised Anomaly Detection Models to Detect Faults in Heavy Haul Railway Operations. 2019 18th IEEE International Conference On Machine Learning and Applications (ICMLA), (2019): 1016-1022, doi: https://doi.org/10.1109/ICMLA.2019.00172.

Roy M. Bose S. K., Kar B., Gopalakrishnan P. K., Basu A. A Stacked Autoencoder Neural Network based Automated Feature Extraction Method for Anomaly detection in On-line Condition Monitoring. 2018 IEEE Symposium Series on Computational Intelligence (SSCI), (2018), doi: https://doi.org/10.1109/SSCI.2018.8628810.

Shi W., Lu N., Jiang B., Zhi Y., Xu Z. An Unsupervised Anomaly Detection Method Based on Density Peak Clustering for Rail Vehicle Door System. 2019 Chinese Control and Decision Conference (CCDC), (2019): 1954.1959, doi: https://doi.org/10.1109/CCDC.2019.8833427.

Velasco-Gallego C., Lazakis I. Real-time data-driven missing data imputation for short-term sensor data of marine systems. A comparative study. Ocean Engineering 218

(2020): 1-23., doi: https://doi.org/10.1016/j.oceaneng.2020.108261.

Velasco-Gallego C., Lazakis I. Data imputation of missing values from marine machinery systems sensor data. Evaluation, visualisation, and sensor failure detection. Proceedings of the RINA Maritime Innovation and Emerging Technologies (2021): 15-23.

Xu Z., Yu D. A Bibliometrics analysis on big data research (2009-2018). Journal of Data Information and Management 1 (2019): 1-13, doi: https://doi.org/10.1007/s42488-019-00001-2.

Xue L., Gao S. Unsupervised anomaly detection system for railway turnout based on GAN. Journal of Physics: Conference Series 1345 (2019): 1-5, doi: https://doi.org/10.1088/1742-6596/1345/3/032069.