Elevator Condition Monitoring: A Lightweight Receptive Field and Temporal Correlations Enhanced Deep Learning Model to Tackle Imbalanced Dataset

.

ABSTRACT

It has been witnessed with the thriving popularity of elevators in contemporary society during modernization while posing an unavoidable challenge to consistently maintain their safety and reliability in routine operations. The conventional monitoring mechanisms prove insufficient to consistently monitor and evaluate the health condition of elevators in actual deployment. Under a plethora of research utilizing data-driven approaches to measure the degradation status of elevators, there also exist multiple industrial pain points while constraining their broad applications in the elevator industry. For instance, the imbalanced dataset with rare abnormal samples in actual life would lead to incomplete and inadequate representative feature learning during the model training process; the excessive reliance on synthetic laboratory data with proprietary application domains in experiment would impair the model's generalization capability; whilst the complex architecture as stacked in pursuit of functionality comprehensiveness would by no means contribute to effectiveness enhancement at the cost of efficiency decrease. This paper established a lightweight deep learning model with streamlined elegant architecture to resolve the aforementioned issues. The real-life imbalanced data have been collected and utilized directly under the devised algorithmic structure and training strategy without any need for laboratory synthesis on their compositions, yet largely alleviating the impact of imbalanced dataset phenomenon in effect. With the proposed effective feature extraction mechanisms, the receptive field and temporal correlations have been enhanced to further facilitate the monitoring efficiency and accuracy. As validated by the systematic experimental results, the receptive fields have been enlarged by 75% in comparison with the vanilla model structure and the parameter quantity have also been reduced by 21% while boosting the model's training and inferencing efficiency during the actual deployment. Meanwhile, in order to rigorously evaluate the model's robustness, the results have been thoroughly validated after splitting the dataset for multiple times with comprehensive ablation study scenarios to explore the optimal model configuration parameters. The proposed model demonstrated its superior performance by the outstripped average detection precision ratios under asymmetrical data allocations while shedding light on its practical application by industrial practitioners for real-life elevator condition monitoring.

Keywords — Anomaly detection, condition monitoring, deep learning, elevator system, Internet of Things (IoT), real-observed dataset

Managerial Relevance Statement

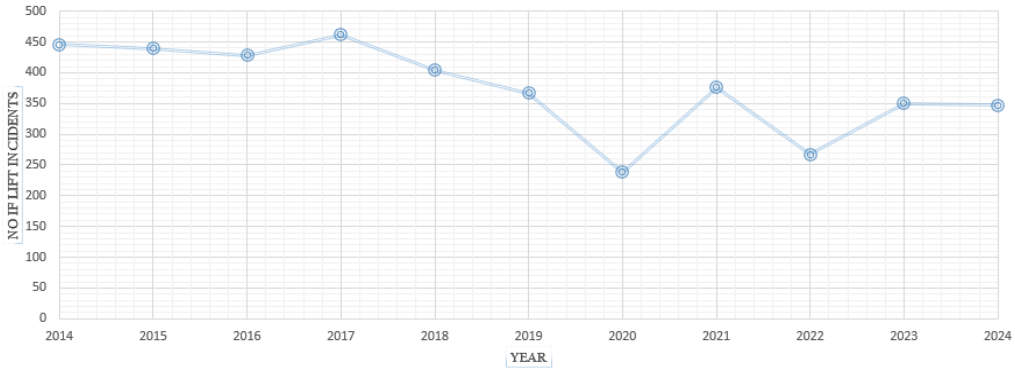
This paper conducts innovative research with feasible data-driven solutions to facilitate the condition monitoring capability on elevators in actual life, which sheds light on the managerial decision-making in adoption of such deep learning-oriented data-driven methods during the real elevator operation process for engineering managers and policy makers. As well acknowledged that there exist considerable deficiencies by the conventional way of monitoring for elevator anomaly detection, which might cause unavoidable latencies or induce hazardous accidents. By the prevailing condition monitoring practice, such existing issues as the imbalanced dataset, architectural complexity and insufficient feature extraction have impaired the generalization capability of deep learning models in their applications to the elevator condition monitoring domain. This paper proposes a lightweight streamlined deep learning model to boost the condition monitoring effectiveness on elevators without exerting additional computational burden, which could be broadly applied based on the real-life imbalanced dataset without the need for excessive data manipulation or feature engineering. Overall speaking, the important managerial insights can also be summarized from the following three-layer statement: 1). Problem Layer: There exist deficiencies and shortcomings in the conventional way of elevator condition monitoring mechanisms, whilst likely to induce hazardous events during the elevator operation process; 2). Solution Layer: The proposed lightweight deep learning model can effectively resolve the existing issues while improving the elevator condition monitoring efficiency and tackling the imbalanced dataset issue; 3). Deliverable Layer: By enhancing the temporal feature correlations with enlarged receptive fields, the proposed data-driven approach achieves superior performance compared with prevailing baseline models, while shedding light on the utilization of feasible AI solutions for elevator condition monitoring with improved monitoring efficiency and effectiveness. Concretely speaking, through the smart monitoring of elevators by the proposed solution, the elevator's safety and reliability shall be leveraged on a large scale, which is crucial for assisting the engineering managers in elevator manufacturing and maintaining enterprises to improve their working efficiency and effectiveness. Meanwhile, the adoption of such a data-driven approach for elevator condition monitoring also sheds light on the execution of practical ways with the employment of artificial intelligence in the safety-critical domains, while enlightening the policy makers in the relevant government bureaus in enacting the relevant guidelines for regulating the utilization of AI in elevator-relevant safety domains. The proposed solution shall assist the engineering managers and policy makers to help with their strategic decision making while embracing the artificial intelligence-oriented digital transformation technologies, thus providing a meaningful and impactful solution to enhance the elevator safety monitoring efficiency and effectiveness for practical applications during the elevator modernization process. Additionally, the captioned study shall contribute to a couple of SDGs: SDG 9 and SDG 11, while fostering sustainable industrialization with technological innovation with more resilient and reliable infrastructure under the Smart City as facilitated.

1. Introduction

With the thriving development of technology and rapid improvement of living standards, people put much more emphasis than ever on the reliability and safety of their routinely used instruments and apparatus, including elevators assimilated into people's everyday life in metropolitan areas. Consequently, how to enable the smart monitoring of elevator conditions by the adoption of cutting-edge deep learning algorithms to integrate the non-invasive Internet-of-things (IoT) (Aheleroff et al., 2020; Wang et al., 2021) sensors under Industry 4.0 (Giovanni et al., 2024) with real-life data collection has become a top concern for academia and industrial practitioners during the digital transformation process (Gao et al., 2025; Lakemond et al., 2024).

The elevator (or referred as "Lift" in this paper), as an indispensable part of contemporary society, plays an irreplaceable role in ensuring the convenient haulage and smooth transportation of goods and passengers. With the increasingly extensive utilization of lift systems in commercial skyscrapers, residential mansions, industrial facilities, and office buildings, it demands high-level attentive maintenance for elevators to ensure the safety of pedestrians and passengers. However, it is discovered that most lift accidents are incurred by the mechanical pitfalls or the degradation of lift components that probably being installed decades ago, while posing great challenges for the lift registered contractors (RC) to conduct the systematic modernization in prevention of hazardous events. According to the latest statistics as released by the Hong Kong Electrical and Mechanical Services Department (EMSD) in Fig.1, there were totally 375 reported lift accidents during 2021, in comparison of 367 reported accidents in 2019 and 238 accidental events as recorded in 2020 (EMSD, 2022c). Such accidents come thick and fast consecutively each year, hence it proves imperative for lift RCs to abide by the compulsory requirement with effective condition monitoring to ensure the reliability of lift systems. In order to optimize the existing situation with safety assurance and risk mitigation, EMSD has launched the "Lift and Escalators Ordinance (Chapter 618)" (EMSD, 2022b) and "Code of Practice (CoP)" (EMSD, 2022a) with concrete measures and statutory requirements as legislated to regulate the lift construction and maintenance work.

Nevertheless, in view of the complexity of elevator systems, the existing industrial practice proves insufficient to ensure the smooth operation of lifts with hazard prevention. The maintenance work is normally conducted either periodically or after the failure of lift systems with passive condition evaluation and excessive overheads. The prevailing countermeasures are mainly focused on the installation of destructive transducers interlinking the subcomponents with the speed encoder, weighbridge, and other relevant components. Despite of the professionalism and expert knowledge (Xiahou et al., 2021) needed for installation of such intrusive sensors, the maintenance also demands the service interruption with repetitive data analysis and extra workmanship for condition evaluation through the specialized equipment and proprietary software designed under a specific lift brand.



Year	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024
Number of Lift Incidents	446	439	428	462	404	367	238	376	267	350	347

Fig. 1. The recorded elevator incidents in Hong Kong during the past 10 years (EMSD, 2025)

Embraced by technological advancement, there are also emerging trends for installing non-intrusive sensors (de Oliveira et al., 2022) with collaborative maintenance (García et al., 2022) in combination with conventional monitoring manners. Meanwhile, more industrial practitioners and academia are investigating the possibility for non-intrusive elevator condition monitoring by the adoption of data-driven oriented artificial intelligence *(AI)* (Medghalchi & Vogel, 2024; Xu & Lin, 2025; Yu et al., 2024) without interference on the intrinsic circuitry of elevators. Whereas there exist industrial bottlenecks for mass implementation after consulting the domain experts in lift industry: (i) lack of abundant real abnormal data and real-observed dataset; (ii) lack of effective AI model for adoption in such safety critical systems; (iii) lack of the holistic framework with the integration of non-invasive IoT sensors (Rajnoha & Hadač, 2024) for condition monitoring of complex systems in real applications.

In order to fundamentally tackle the aforementioned problems, we established a deep learning (DL) oriented data-driven approach to construct the smart condition monitoring system for elevators (Mishra & Huhtala, 2019). By means of the well-established model, no sophisticated feature engineering needed, while the multivariant current signals as abstracted via non-invasive sensors were fed into the architecture with automated feature extraction and real-time anomaly detection. By the proposed model architecture, the receptive fields have been enlarged dramatically with enhanced temporal correlations, which could leverage the feature extraction capability of the DL model in tackling the real-case data patterns during elevator condition monitoring process after commissioning. The main contributions are listed in the following:

i) Establish the holistic elevator condition monitoring framework that integrates a deep learning-based health index predictor and an anomaly detector, tailored for application on non-intrusive one-dimensional sensor signals.

ii) Propose the state-of-the-art deep learning-based model comprising a one-dimensional dilated convolutional neural network (1D-Dilated CNN) with expanded receptive fields, followed by a concatenated gated recurrent unit (GRU) architecture to further fetch the temporal correlations within

the multi-variant data, while enhancing the feature extraction capability for efficient elevator condition monitoring.

iii) Propel the theoretical advancement with the proposed model architecture to tackle the imbalanced dataset during the elevator condition monitoring process in real life by creatively designing the devised training strategy and loss function.

iv) Construct the real-world elevator condition monitoring dataset reflecting actual multi-variant data patterns under the realistic industrial operating conditions and validate the model robustness with multiple dataset splitting through the systematic ablation study.

The remainder of this paper is organized as follows: Section 2 briefly overviews the prevailing methods for condition monitoring of E&M systems; Section 3 formulates the problem with the defined objective function; Section 4 illustrates the architecture of the established 1D-DiCNN-GRU DL model; Section 5 introduces the data pre-processing procedure with data concatenation and segmentation process; Section 6 demonstrates the experiment results after splitting the dataset and Section 7 concludes this paper with a brief summary.

2. Related works

2.1 Conventional Ways of Elevator Condition Monitoring

Condition monitoring for the electrical and mechanical (E&M) system has been conducted with plethora of research focusing on the conventional approaches with physics-guided or knowledge-based empirical methods for anomaly detection based on embedded sensors (Dai & Gao, 2013; Khan et al., 2024; Yin et al., 2014). The Physics of Failure (PoF) approach has been integrated into reliability engineering for complex systems as evolved with significant attention on the prediction accuracy amongst environmental and operational influential factors (Zeng et al., 2016). Generally speaking, PoF method has been regarded as one of the conventional but invaluable techniques that being widely utilized for prevention of failure occurrence while encompassing the intricate practices of reliability design in mechanical systems (Wileman et al., 2021). As for the model-based methods that proposed in (Esteban et al., 2014), they devised the approach for detecting the anomalies of E&M sub-systems while employing the Linear Parameter Varying (LPV) model for joint estimation on the elevator performance. As explored by (González et al., 2019), the model-based approach for evaluating the guiding rails was utilized in the vertical transportation systems with extended Kalman filter (EKF) for quantitative diagnosis. However, both the conventional PoF method and model-based method would require the prior predecessor knowledge about the mechanical condition or maintenance histories about a specific brand of elevator with process mining (Tang et al., 2023) to investigate the root causes against the system failures. Nevertheless, such previous experience or prior knowledge proved difficult to acquire

across different brands of elevators, which posed the restriction for application with such vanilla methods in real practice especially under the Industry 4.0 processes (Mazzoleni et al., 2022).

2.2 Emerging Trends of Elevator Condition Monitoring with Machine Learning Approaches

In view of the aforementioned shortcomings, there has been an emerging approach with AI-based (Koutsoupakis et al., 2023) or specifically machine learning (ML)-based (Arora et al., 2022) data-driven methods that has given rise across different industrial domains (Hennig et al., 2021; Kumar Sharma et al., 2022; Moura et al., 2019). Some previous work was focused on the variants of support vector machines (SVM) to diagnose the vibration signals from the elevator doors for detecting the malfunctioning components using the least squared SVM (LS-SVM) (Wan et al., 2015). Under the complex mechanical system of elevators, there emerged another approach with fault detection based on Bayesian network algorithm for the motor, gearbox, brake, sheave, suspension rope and lift car (Liu et al., 2017). Under the fusion approach, the model combing the particle swarm optimization (PSO) algorithm and back propagation (BP) neural network were employed for lift door fault prediction and diagnosis (Wen et al., 2016). Whereas the mentioned approaches inevitably need the redundant process for feature engineering in order to delineate the feature profiles and inter-dependencies, thus burdening the computation complexity with unnecessary offloading, especially under the multi-variant high-dimensional dataset (Fu et al., 2021).

2.3 IoT Integrated Monitoring with Deep Learning Methods

Some of the latest studies investigated the installation of IoT sensor nodes (Ilango et al., 2022; Skog et al., 2017) to build the Internet-of-Elevators with non-invasive fault monitoring. For instance, the work as mentioned in (Chai et al., 2021) employed the architecture with long short-term memory (LSTM) and convolutional neural network (CNN) model for extracting the temporal and spatial correlations. Whereas the paper failed to address the imbalanced dataset problem, and the proposed model demonstrated complex architecture with excessive parameters.

To further streamline the DL model with reduced parameters, the dilated CNN model architecture was initially proposed by (Koltun, 2016), mainly utilized in semantic segmentation, natural language processing (NLP) and computer vision (CV) relevant domains. The notable improvement of Dilated CNN in comparison with conventional CNN architecture lies in its architecture with the devised dilated rate to enlarge the receptive field without adding additional model parameters, nor expanding the kernel size (Li et al., 2023; Xu et al., 2022). The Gated Recurrent Unit (GRU) architecture was proposed in (Chung, 2014), the invention of which was to remedy the vanishing gradient issues encountered with LSTM model with streamlined architecture. By reducing the original three gates of LSTM to the two gates in GRU, namely the “update gate” and “reset gate”, it is validated that the learnable parameters and kernel computations are dramatically reduced without performance deprivation to enhance the efficiency.

Despite the work conducted by predecessors with fault diagnosis of industrial processes and machinery (Feng et al., 2021; Yan et al., 2022), the systematic research for elevator health monitoring (Chen et al., 2021) via IoT sensor nodes using the advanced DL methods has been scarcely conducted (Nguyen et al., 2022). Moreover, from the perspective of data-driven approaches with AI utilization for Human-Machine Collaboration (HMC) (Lingxiao et al., 2025), it has been discovered that in the domain of elevator condition monitoring, the previous work is either reliant on the synthetic laboratory data or seldomly focused on the feasible solutions to tackle the real-life imbalanced dataset issue, which is quite commonly encountered in real syndromic surveillance of elevators during actual operations.

2.4 Attempts in Tackling Imbalanced Dataset and Limitations for the Problem Domain

Under the scope of DL for alleviating the impact of imbalanced datasets, there exist previous research works attempting to adopt measures from different aspects aiming to leverage fault diagnosis accuracy. For instance, there exists the One-class Classification (OCC) algorithm to address the limited data sample issues in distributed systems (Lin et al., 2018). Thereafter as a typical example, the one-class SVM-based fault diagnosis method was explored based on the feature space-located distance metrics (Mahadevan & Shah, 2009). Whereas such methods would be overshadowed by complex datasets, especially when extracting the dominant features from the elevator's heterogeneous multi-variant real-monitoring data. In the presence of sample generation techniques, there also exist shallow oversampling algorithms such as ADASYN (Haibo et al., 2008), SMOTE (Huo et al., 2015), Geometric SMOTE (Douzas & Bacao, 2019), Random-SMOTE (Dong & Wang, 2011) etc. By adding the synthesized minority class of samples, though such methods could help alleviate the asymmetry data allocation issues, the generated instances might fail to represent the underlying density against the elevator's dynamic operation environment during the synthesization process, thus falling into the overlapping noisy areas and weakening the actual performance (Analide et al., 2020). Similar to the principle for data augmentation, the latest generative AI (GAI) technology has given rise to the utilization of deep generative models (DGM) (Tomczak, 2024) in domains of CV and semantic segmentation etc. However, in its adaptation to industrial applications for 1D data facilitation, the current literature has scarcely been covered. Meanwhile, the training complexity of DGM might impact the inferencing efficiency for real-time monitoring during the actual operation process. The evaluation metrics against the generated anomalous samples with appropriate manifestations remain insufficiently explored while arousing hesitation from industrial practitioners for practical deployment.

2.5 Niche Statement

The prevailing industrialized approaches are mainly focused on the utilization of conventional physics-guided empirical methods while lacking the generalization capability across the diversified application scenarios. Whilst the ML-based methods cannot alleviate the redundant process for feature engineering with the burdened computational complexity. Regarding the DL-integrated data-driven

approach with real-case IoT sensor nodes, it is discovered that seldom research has entered such a sphere in actual operation, while mainly relying on synthetic laboratory data for analysis. Despite of efforts devoted to tackling the imbalanced dataset issue, there exists a lack of feasible methods for real-life adoption during the elevator's real operation process. It thus requests the model with high inferencing efficiency applicable to real-time monitoring scenarios, without a complex algorithmic structure nor overshadowed performance against the multi-variant accumulative data during the elevator's routine monitoring process, which becomes the research niche of our work. Hence under such conditions, in the domain of elevator condition monitoring, how to design a feasible and executable DL model with leveraged efficiency while tackling the imbalanced dataset issue with a devised lightweight but elegant architecture for real-case efficient application could be a worthwhile direction carried with valuable research impacts.

3. Problem formulation

3.1. Holistic elevator condition monitoring framework

By the hierarchical design, there are four primary non-intrusive IoT sensors installed to extract the current signals from the core components of the elevator, namely the traction motor, brake, safety circuit and lift door (as shown in Fig. 2). The holistic framework is composed of multiple IoT sensor nodes for real-life data acquisition without intervening the normal operation process of the elevator. The devised DL model is employed for real-time fault detection and prognostics with broad application potential during the elevator modernization or retrofitting process. The workflow is summarized by below stages:

Stage 1). Data capturing and acquisition. The non-intrusive clamp-type current sensors are installed to fetch the electric current signals flowing through the core components during real-life observation.

Stage 2). Data transference and storage. The micro-controller is configured with Raspberry Pi 4i for data transfer under 4G LTE router and upload to the cloud-based data storage hub empowered by Amazon Web Services (AWS).

Stage 3). Data analysis and anomaly detection. The DL-based model is employed to analyse the real-life data with condition monitoring and health status evaluation. The anomaly detection results will be provided to the lift maintenance crew to support the predictive maintenance (Jaenal et al., 2024) with prognostic health management (PHM).

3.2. Monitored components

The measured components are explicitly demonstrated in Fig. 3, with electric current signature analysis for real-time condition monitoring. The distance sensor and reflection plate are installed with the aim of capturing the distance between the roof of elevator car and the ceiling of shaft for velocity calculation.

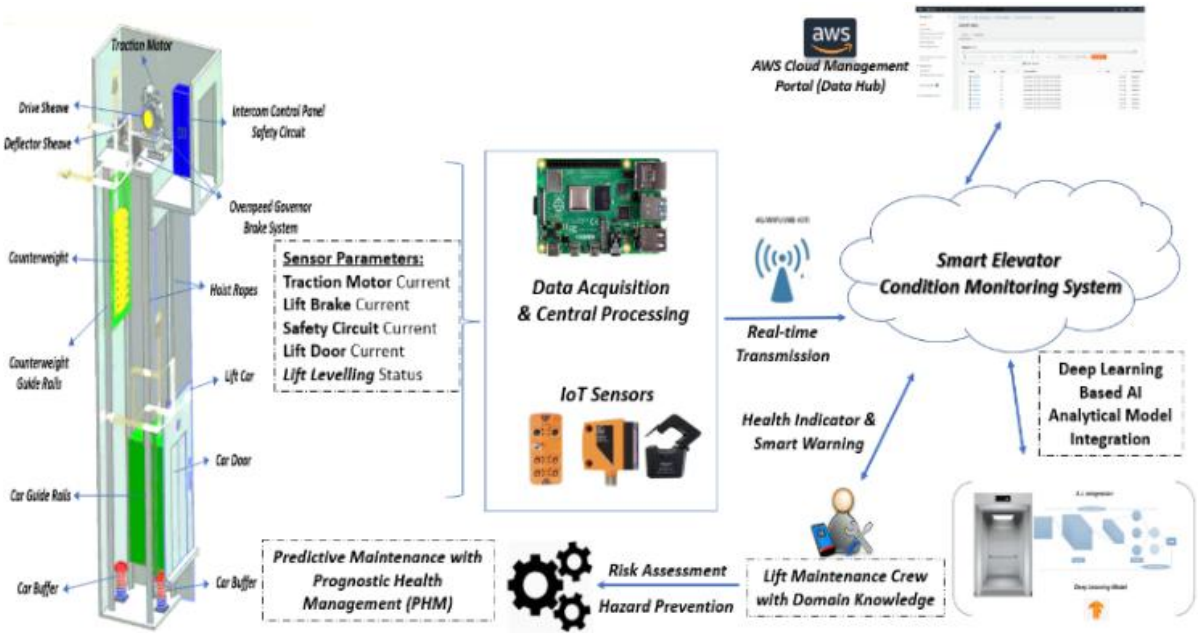


Fig. 2. Overview of the non-intrusive smart elevator condition monitoring framework.

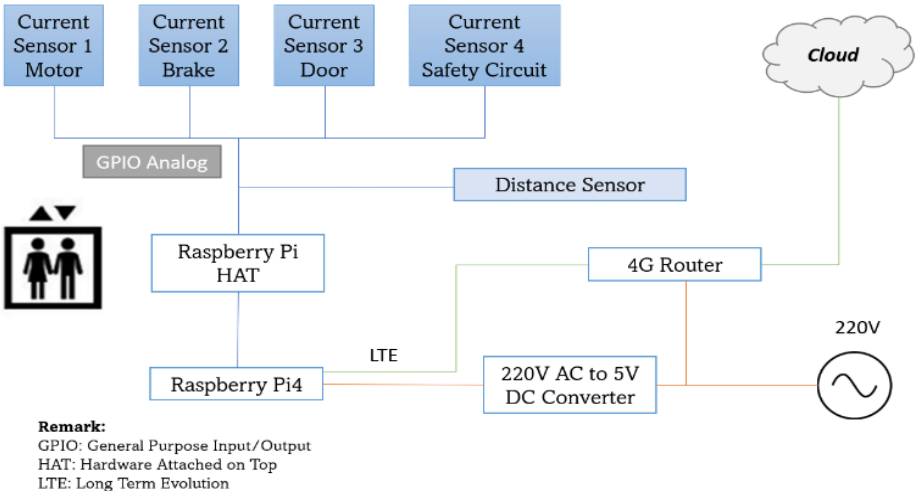


Fig. 3. Overview of the monitored components.

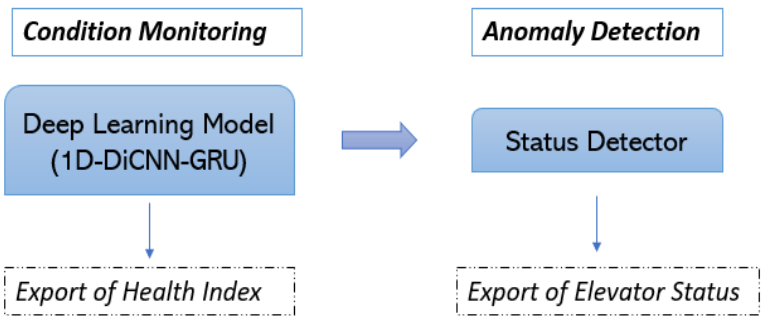


Fig. 4. Workflow components of the proposed model framework.

1) Traction Motor

The asynchronous AC (alternating current) motors are widely equipped on elevators to empower the magnetic ropes with driving force. The magnetic and electric circuits are interlinked with the inner components as the stator and rotor, generating the electromagnetic induction to convert into mechanical energy. Hence the electric current signals passing through the motor is one of the critical factors to ensure the haulage and kinetic power supply for the elevators.

2) Brake

The elevator is normally equipped with the magnetic power-off brake system. The brake coil shall be demagnetized when lift is halted with the power supply being cut off. The pressure plate and cover plate are forced by the tensioned springs to get the friction disc tightly locked, while impeding the rotation of hub and shaft to stop the elevator.

3) Safety Circuit

The safety circuit acts as the core component interlinking the safety relays with hazards prevention by lift notch-off, while reducing the risk to the acceptable level.

4) Door

The door of elevator serves as the indispensable part to ensure the safe travel of passengers. The lift door motor is energized to open and close the lift gates shortly after the lift parked at the designated floor with consecutive spikes of door current, acting as the indicator to reflect the lift motion.

3.3. Problem statement

Under the holistic condition monitoring framework as illustrated, it has been recognized as the industrial bottleneck encountered with imbalanced dataset during lift observation. Hence the aim of this paper is to devise an elegant DL model to analyse such multi-variant signals from the four measured components with imbalanced dataset, in the aim of monitoring the condition of elevators with anomaly detection (Fang et al., 2023).

Let X_{im} be the input of multi-variant electric current signals under real-life imbalanced dataset with X_{im}=$\{x_1, x_2, x_3, x_4\}$(Aheleroff et al., 2020) and let Y_h=$\{1,0\}$ be the health index reflecting the status of elevator, where Y_h equals "1" means the elevator in good health condition.

Under the task of condition monitoring with estimation on the elevator's health condition, our aim is to minimize the variation between the predicted health index and the true condition of elevators as interpreted by the overall objective function depicted in the following equation:

$$min_{\theta} \| f_D(X_{im}; \theta) - Y_h \|, \quad (1)$$

where f_D denotes the lightweight DL network with learnable parameter Θ, Y_h represents the elevator's health index. The objective is to minimize the difference and deviation between the predicted health index and the actual health condition of elevators under the heterogeneous imbalanced data distribution, such that the situation of elevators could be determined with enhanced accuracy and efficiency,

4. Model architecture

In this section, we shall give a detailed introduction to the DL solution for the problem stated in *Section 3*. The proposed DL framework is elaborated with the framework (as shown in Fig. 4 comprising of two parts. One is the one-dimensional stacked dilated convolutional neural network (1D-Dilated CNN) followed by a GRU regression head (1D-DiCNN-GRU) for the task of condition monitoring; the other as the "status detector" with binary classification to determine whether any anomaly is detected based on the health index value as output from the previous part.

4.1. 1D-DiCNN-GRU model

As shown in Fig. 5, the model basically consists of a series of vanilla or dilated 1D convolutional layers, and a GRU. More specifically, in this paper, the multi-variant 1D signal will first go through a vanilla 1D convolutional layer with kernel size equal to "8", and then pass a cascade of three stacked 1D dilated convolutional layers with the same kernel size 3 but different dilated rates – "1, 2, 5", respectively. We set the dilated rates as "1, 2, 5" sequentially, so as to exponentially expand the receptive field without loss of any coverage, in other words, getting rid of the gridding effect with minimized artifacts caused by the dilated convolution. This combined dilation rates configuration can ensure a dense and contiguous coverage span traversing the input data, which also helps the model better capture the long-term dependencies existing in the elevator's real-monitoring data.

Among the three stacked 1D dilated convolutional layers with kernel size of "3", assume that there are M filters with input dimension as "1 × N", the feature map in the l-th layer can be represented by Equation (2).

$$x_j^l = f\left(\sum_{i=1,\dots,N} x_i^{l-1} \times k_{ij}^l + b_j^l\right), j = 1, \dots, M \tag{2}$$

where *f* denotes the activation functions with *x* representing the convolutional operations; x_i^{l-1} and x_j^l denote the *l*-th input feature map and the output map on the *j*-th layer; k_{ij}^l and b_j^l denote the kernel and the corresponding bias of the *j*-th layer respectively.

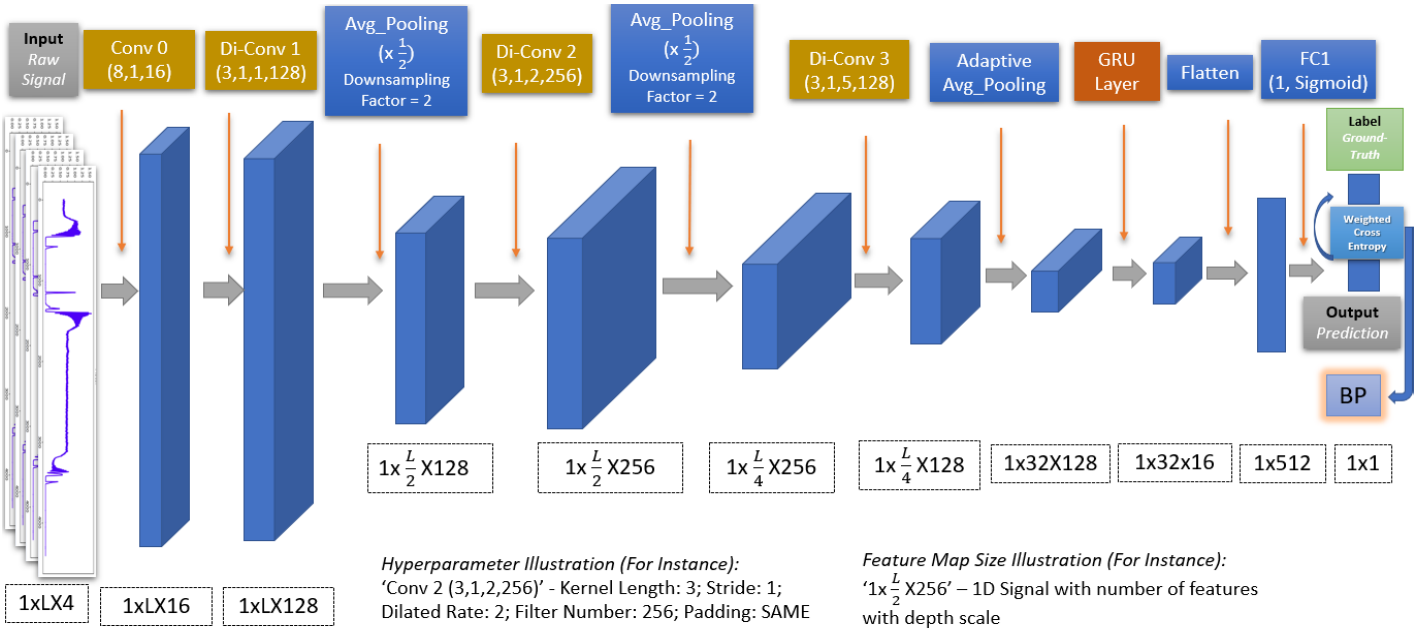


Fig. 5. Workflow components of the proposed model framework.

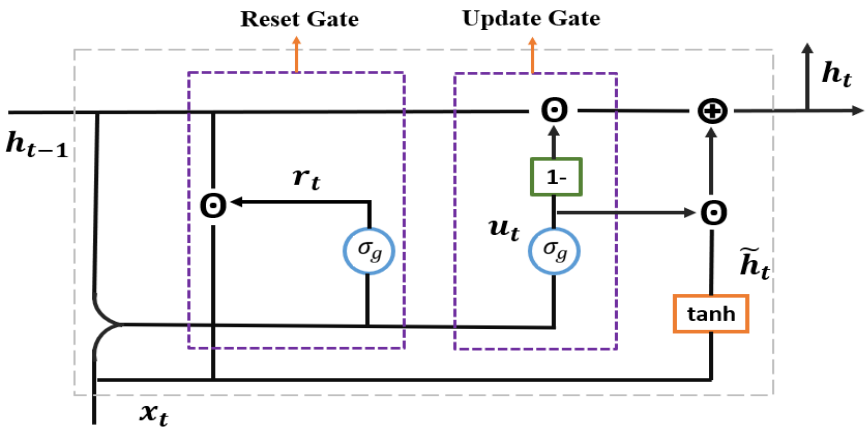


Fig. 6. GRU cell structure as concatenated in the proposed model.

It should be noted that an additional vanilla convolutional layer with large kernel size "8" is added before the three dilated convolutional layers to further enhance the capability of the network to investigate temporal correlation within the input signal at low level. It is very crucial to enlarge the receptive field of the neuron, especially for our case as the input sequence is extremely long, containing thousands of timestamps. The temporal correlation between far neighbors will not be grasped if the receptive field is not big enough. And the proposed combination of one vanilla convolutional layer and three stacked dilated convolutional layers can take care of this issue effectively with less parameters compared to the scenario with all vanilla convolutional layers. Out of the four convolutional layers is the feature of size "$1 \times L/8 \times 128$". An adaptive average pooling layer is then applied to uniform the length of the extracted feature as "32".

As the convolutional kernel will slide towards one direction only, from left to right, for example, in the 1D convolutional layer, the feature we obtained after the adaptive average pooling layer basically remains the chronological order as the raw input. Based on this characteristic, on top of the CNN module we adopt a GRU (as shown in Fig. 6) to capture the inherent temporal correlation within extracted features and then fully connect to the final output, instead of using the trivial multilayer perceptron (MLP) model (Bisong, 2019) as the regression head. Experiments in *Section 6* proved the superiority of our proposed network.

The operation inside of GRU cell can be depicted by following formula:

$$u_t = \sigma_g(W_{zx}x_t + W_{zh}h_{t-1}) \tag{3}$$

$$r_t = \sigma_g(W_{rx}x_t + W_{rh}h_{t-1}) \tag{4}$$

$$\tilde{h}_t = \tanh\left(Wx_t + r_t \odot W_{hh}h_{t-1}\right) \tag{5}$$

$$h_t = (1 - u_t) \odot \tilde{h} + u_t \odot h_{t-1} \tag{6}$$

where u_t denotes the update gate, r_t represents the reset gate, $\tilde{h}_t$ describes the candidate state of the hidden node, h_t represents current hidden state and h_{t-1} denotes the previous hidden state; x_t is the model input and W is the weight matrices; $\odot$ denotes the element-wise dot product, whilst σ_g and tanh denote the Sigmoid and tanh activation functions respectively.

The 1D-DiCNN-GRU model takes the regression task with estimated health index valued between "0" and "1" as the output. Based on the health index, the second part of framework serves as the anomaly detector to do the conditional classification to determine whether the lift system is in good condition or not by comparing the predicted health index to the anomalous threshold.

4.2. Characteristics of adaptive pooling layer

The devised 1D-Dilated CNN backbone includes three stacked blocks composed of one 1D-Dilated convolutional layer and an average pooling. With the aim to reduce the feature dimensions, pooling layers are critical to alleviate computational complexity. Considering the multi-variant signals in different lengths under multi-instance learning (MIL) (Liu et al., 2021), the adaptive pooling was utilized to trim and standardize the signal length to "32" before fed into the GRU. Below equation demonstrated the operation of the pooling layer with feature map calculation for the *l*-th layer.

$$x_j^l = f\left(\beta_j^l \text{Đ}\left(x_j^{l-1}\right) + b_j^l\right), j = 1, \ldots, M \tag{7}$$

where M denoted the number of maps, f stands for the activation function and $Đ$ represents the pooling (down-sampling) function; β_j^l and b_j^l denote the bias for multiplication and addition accordingly of the j-th filter, while x_j^l & x_j^{l-1} stand for the output and input maps of the j-th filter after and before the pooling operation.

The adaptive pooling layer as concatenated before the GRU ensures the uniform length of input for the last fully connected layer. Consequently, the model is capable of processing the arbitrary length of signals for inference. It is the adaptive selection of kernel size that enables the adaptive pooling layer to produce outputs with same length regardless of the input length of raw signals to accommodate our application scenarios with lift condition monitoring.

4.3. Employment of weighted cross entropy loss

It has been one of the primary goals to tackle the imbalanced dataset of lift while lacking abnormal data (Zhou et al., 2021). The countermeasures proposed by this paper are categorized into two approaches, namely the devised training strategy and the refined loss function.

The loss function was employed in light of cross-entropy that being utilized to measure the difference between two probability distributions as follows,

$$H\,(q,p) = -\sum_{x\in\chi} q(x)\,log(p(x))\,, \tag{8}$$

where q and p represent sample distribution of the ground truth and prediction respectively, χ represents the support. To further tackle the problem with imbalanced dataset to improve the convergence speed and training accuracy during evaluation, we have come up with the weighted cross entropy as for weighting the loss function to achieve the optimal balance with improved performance. The corresponding formula as obtained by Equation (3).

$$H\,(q,p) = -\sum_{x\in\chi} \omega(x)q(x)\,log(p(x))\,, \tag{9}$$

where $\omega(\cdot)$ is the weight function. Particularly, in this paper, $q = \{y, 1-y\}$, $p = \{\hat{y}, 1-\hat{y}\}$, and the loss function can be rewritten as follows explicitly,

$$L_\theta = -\omega \cdot ylog\hat{y} - (1-y)\log\,(1-\hat{y})\,, \tag{10}$$

where $\hat{y} = f_D(X;\theta)$, and ω denotes the loss weight.

The ablation study was conducted accordingly in *Section 6 – Experiments*, while further discussing how the weight selection could affect the testing results. The optimum weight was selected afterwards to adjust and balance the portion of normal and abnormal dataset. By the refined allocation, the weighted cross entropy loss function as affiliated with MLP dense layer could help tackle the issue under imbalanced dataset. The trend of the exported mean absolute error (MAE) values was observed so as to continuously monitor the elevator's health condition and to strategically integrate with the anomaly detector for consecutive status evaluation.

5. Data preprocessing

In this section, the data preprocessing details are provided with elaboration on the data concatenation and segmentation mechanism for training sample generation.

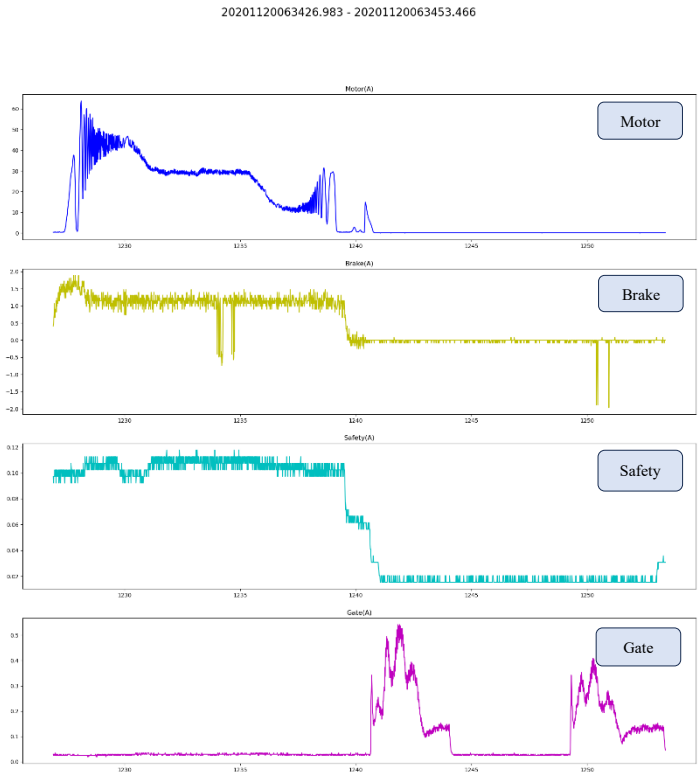


Fig. 7. A typical cycle with homogeneous segmentation.

The real data were collected from a lift with model of EXDN/VVVF as manufactured by "Fujitech", which was installed at one of routinely used office buildings in Hong Kong. Specifically, the four multi-variant current signals were collected during the period from 1 Nov 2020 till 31 Dec 2020 with sampling frequency of 70 Hz. All collected data have been pre-processed with totally 39367 segments prepared.

5.1. Velocity calculation

The increase and decrease of the monitored distance indicate the motion of lift in descending and ascending status. Generally, the fluctuation on distance and velocity are compared with the current variations of two consecutive timestamps to determine the lift movement with predefined threshold as encoded in our algorithm. The aim of the calculation is to capture the starting and ending point for the lift movement as reflected in the levelling/velocity variations, while further discovering the correlations with multi-variant current data.

5.2. Data homogeneous segmentation

For explicit display of the multi-variant current signals to discover their correlations, the input signals were partitioned into a set of homogeneous operational segments. Each segment starts when the velocity variations were detected at the initiation point and ended when the elevator halted, followed by two consecutive spikes of door current, while indicating the open and closure of elevator gates for passenger offboarding and onboarding.

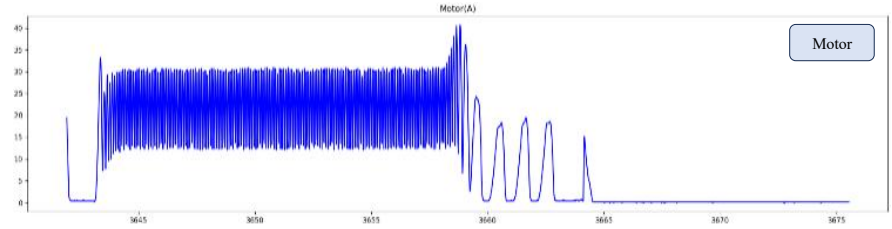


Fig. 8. Observed abnormal pattern with condensed waveform for traction motor.

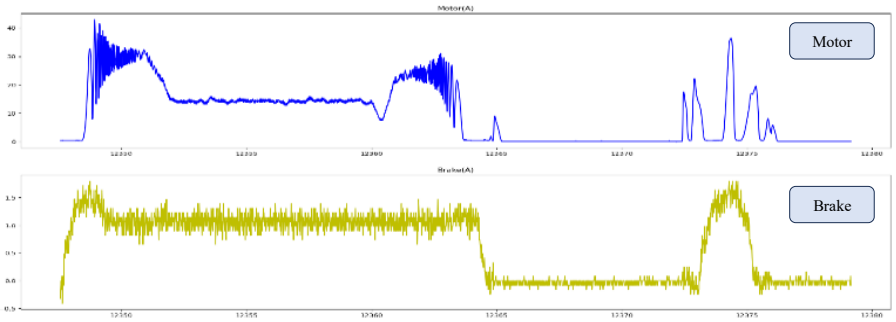


Fig. 9. Observed abnormal pattern with abrupt current increasement.

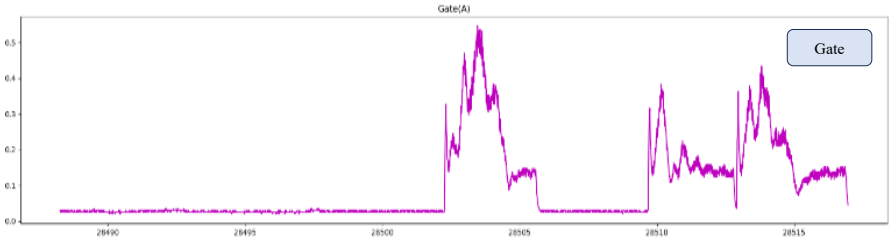


Fig. 10. Observed abnormal pattern with door closing issue.

The concatenated data were processed with automated homogeneous segmentation while including *i)* elevator car motion and *ii)* door motion (open and close). The normal patterns of current signals of the four measured components within a typical lift movement cycle were depicted by Fig. 7.

5.3. Observation of abnormal sequence patterns

During the data pre-processing process, we discovered three characteristics of the potential anomalies as reflected in the current patterns of the traction motor, brake and the lift door as elaborated respectively in the following section.

1) Condensed Motor Current

The induction motor is functioned to drive the magnetic rope with haulage and kinetic energy. The normal motor current starts with peak spike waveform at initialization to empower the lift car while triggering the sub-peak when prepared for a halt at certain floors. During the observation, the condensed waveform pattern of motor current was discovered in Fig. 8. Under advice from the domain experts, one of the potential root causes might be related with the gradual degradation of traction motor, as located in rotor bar or shaft, that inducing more current drawn.

Table 1.
Number of samples preparation in dataset.

Dataset	Sequence Quantity	Normal Condition	Abnormal Condition
Training	8049	7896	153
Validation	1012	996	16
Test	1023	1008	15

2) Abrupt Surging of Motor & Brake Current

The electric current flowing through the brake and traction motor shall be synergized to power the lift car with movement. This abnormal situation was observed via the pattern of unstable current signals with abrupt increasement as shown in Fig. 9, which could be induced by the potential degradation of subcomponents in the brake. Under riffle effect, the conjunct motor current also increased simultaneously thus indicating the needs for comprehensive checking with status evaluation.

3) Door Closing Issues

The lift door anomalies occurred when odd number of consecutive spikes of current signals were spotted. As shown in Fig. 10, the detected three spikes of gate current indicated an additional third gate motion after opening/closure. As discussed with domain experts, it might be induced by the malfunctioning of door sensor or obstruction on the trackway, with the gate bounced back in consecutive movements.

6. Experiments

In this section, the training strategy and model hyperparameters shall be introduced and justified accordingly, together with the anomaly detection threshold and evaluation metrics configuration. Under *Subsection 6.2*, the systematic *Ablation Study* shall be conducted while exploring and substantiating the model concatenation rationale, the loss function design and the batch size selection. The model compression experiment results and performance benchmark against other state-of-the-art baseline models shall be conducted as well with results unveiled since *Table 4* till *Table 6* after splitting the dataset multiple times for objective comparison. The model's efficiency and convergence performance were validated as well thereafter.

As illustrated in *Section 5*, the whole dataset as prepared incorporates 10084 multivariant signals with three types of abnormal samples acquired through real observation. From the dataset summary as listed in Table 1, the allocation for training dataset, validation dataset and testing dataset takes around 8:1:1 under appropriate proportion. The experiment was carried out via computer with Max. 5.1 GHz 8-core i7 10875H CPU (16GB RAM), equipped under Nvidia GeForce GTX-2060 GPU using TensorFlow framework in Python 3.6.

Table 2.
Details of dilated CNN+GRU.

Conv 0 (8,1,16) *Kernel Size*: 8; Stride: 1; *Padding*: Same (Same input and output length) *Filter Number*: 16 (16 Kernels with length as *L*) Conv 0 Output: 1* *L* *16 (Feature Map Size)
Dilated Conv 1 (3,1,1,128) *Kernel Size*: 3; Stride: 1; *Dilated Rate: 1; Filter Number*: 128 (128 Kernels with length as *L*) *Padding*: Same Dilated Conv 1 Output: 1* *L/2* *128 (Feature Map Size) Avg_Pooling with *Down-sample factor: 2*; Length turned to *L/2*.
Dilated Conv 2 (3,1,2,256) Increased dilated rate to 2 for larger receptive field. *Padding*: Same Dilated Conv 2 Output: 1* *L/4* * 256 (Feature Map Size) Avg_Pooling with *Down-sample factor: 2*; Length reduced to *L/4*.
Dilated Conv 1 (3,1,5,128) Increased dilated rate to 5 for larger receptive field. *Padding*: Same Dilated Conv 3 Output: 1**L/8**128 (Feature Map Size)
Adaptive Avg_Pooling: Uniform the length of the output of previous dilated conv layer to 32. Adaptive Avg_Pooling Output: 1*32*128 (Feature Map Size)
GRU cell (16) Number of hidden neurons in the gate: 16 GRU cell output: 1*32*16 (Feature Map Size)
FC1 (1, Sigmoid): Fully Connected Layer with predicted health index (predicted label). FC1 Output: 1 * 1.

Table 3
Details of vanilla CNN+MLP

Conv 1 (8,1,128) *Kernel Size*: 8; Stride: 1; *Padding*: Same (Same input and output length) *Filter Number*: 128 (128 Kernels with length as *L*) *Conv 1 Output*: 1* *L/2* *128 (Feature Map Size) Avg_Pooling with *Down-sample factor: 2*; Length turned to *L/2*.
Conv 2 (5,1,256) Reduced Kernel Size with *deepened* filter (filter number added to 256) Conv 2 Output: 1* *L/2* * 256 (Feature Map Size) Avg_Pooling with *Down-sample factor: 2*; Length reduced to *L/4*.
Conv 3 (3,1,128) Reduced Kernel Size with *less deepened* filter (number reduced to 128) Conv 3 Output: 1**L/2*128 (Feature Map Size)
Adaptive Avg_Pooling: Uniform the length with 32 extracted features. Depth GAP: Global Average Pooling through depth as reduced to 1. Depth GAP Output: 1 * 32
FC1 (32, ReLU): Fully Connected Layer with input and output with 32 neurons. Activation function: ReLU. FC1 Output: 1 * 32.
FC2 (1, Sigmoid): Fully Connected Layer with predicted health index (predicted label). FC2 Output: 1 * 1.

6.1. Training strategy and model hyperparameter

During the training process, with the aim of balancing the normal and abnormal data allocation to deal with the imbalanced dataset issue, it was particularly configured with the devised algorithm to force the allocation of abnormal data (negative data) with "1/8 portion" among the data in each mini batch. For instance, the batch size was set as "8", among which there were 7 normal samples with one abnormal sample. The model hyperparameters are outlined in Table 2 with batch size of "8" over 100 epochs in Adam optimizer and "1.00e-4" learning rate.

6.1.1 Anomaly detection threshold selection

The back propagation (BP) is conducted to consecutively update the parameters with iterations and export the predicted health index value (with Sigmoid activation), as represented by *h(x)* in Equation (11).

$$f(x) = \begin{cases} normal \ \ , h(x) > \lambda \\ anomaly, h(x) \leq \ \lambda \end{cases} \tag{11}$$

The anomaly detection classifier is illustrated with λ as the threshold, the value of which has been acquired while traversing all the samples in validation dataset with the best performance. According to our experiment, the optimal threshold λ was acquired with the value as "0.8". Consequently, when *h(x)* > "0.8", the system was determined as healthy; when *h(x)* ≤ "0.8", the condition was judged as abnormal with anomaly detected.

6.1.2 Evaluation metrics

The overall results were further examined via below metric by the "Precision" score, "Recall" score and the "F1-Score" accordingly since Equation (12) till Equation (14) for performance evaluation.

$$\text{Precision} = \frac{\boldsymbol{TruePositive\ (TP)}}{\boldsymbol{TruePositive\ (TP)+FaulsePositive\ (FP)}} \tag{12}$$

$$\text{Recall} = \frac{\boldsymbol{TruePositive\ (TP)}}{\boldsymbol{TruePositive\ (TP)+FaulseNegtive\ (FN)}} \tag{13}$$

$$\text{F1-Score} = 2 * \frac{\boldsymbol{Precision*Recall}}{\boldsymbol{Precision + Recall}} \tag{14}$$

The confusion matrix was employed to measure the overall performance. Except for the critical FP ratio that we were intended to reduce, the FN condition should also be one of the consideration factors in order to improve the maintenance efficiency to avoid the laborious attentive operations with extra unnecessary field checking especially when confronted with shortage of elevator technicians.

6.2. Ablation study

A series of experiments have been conducted to get the final 1D-DiCNN-GRU model as specified in Table 2. In this section, we shall analyze all the experimental results in detail, showing advantages of

each component in the model, while verifying the effectiveness of the training strategy described in Section 6.2.

Specifically, the capability of dilated CNN feature extractor followed by the GRU regression head will be addressed through the comparison with the vanilla CNN-MLP framework. And then the selection of hyperparameters such as the loss weight in weighted binary cross entropy loss, and the batch size for training would be investigated. Two approaches to reduce the number of parameters of the network and strengthen potential generalizability of the model were studied, and we finally concluded with our proposed model in Table 2 and Table 3. In order to enhance the rigorousness of the proposed model for comprehensive validation, we have split the dataset and summarized the representative results after splitting and testing multiple times. Afterwards, the averaged performance for anomaly detection was evaluated via two metrics, namely the mean value - *Ab. M. (μ)* and standard deviation value - *Ab. S. (σ)* correspondingly. The results were shown in Table 4, Table 5 and Table 6 accordingly to justify the model concatenation and hyperparameter selection configurations.

6.2.1 Reason for dilated CNN in concatenation with GRU

In this section, we compared the model performance between the proposed model to the vanilla CNN-MLP framework as described in Table 3. From the collection of confusion metrics results in Table 4 after splitting the dataset, it was found that the dilated CNN + GRU (32) performed the best with mean values *(μ)* of Precision/Recall and F1-Score reaching the highest standard deviation *(σ)* as the least after splitting the dataset. The model performance could be qualitatively evaluated as well by analysing the confusion matrix results. Compared to the best model dilated CNN + GRU (32), by replacing the dilated CNN feature extractor with the vanilla CNN, more defects were not identified with smaller mean and larger deviations on average. Likewise, by substituting the GRU with the trivial MLP, some normal samples were mis-classified as anomalous, which also demonstrated the suboptimal performance in comparison. As a result, the dilated CNN+GRU shows stronger capability in extracting features and capturing their internal temporal correlation.

6.2.2 Reason for weighted binary cross entropy loss function

In order to alleviate the influence of imbalanced dataset, we have applied weighted cross entropy (WCE) as the loss function for training. As the ablation study, we selected different weights ω imposed on the normal samples to test the model performance. When the weight $\omega <$ "1", one can expect that less attention will be paid to the normal samples, and in turn, more will be exerted on the contribution of abnormal samples to the final loss. While $\omega =$ "1" means that no extra attention is paid to either negative or positive samples. From Table 5, it can be found when $\omega =$ "1" in dataset *Splitting No. 1*, two out of fifteen defects escaped from the detection and one normal sample was mis-identified as abnormal. While after applying the weighted binary cross entropy loss with weight $\omega =$ "0.1", the number of both false negative and false positive decreased, which can be represented by the increased mean value and reduced

standard deviation after splitting the dataset, thus proving the effectiveness of the WCE loss function in dealing with this imbalanced dataset issue.

6.2.3 Selection of batch size

Fixing the structure of network - the dilated CNN+GRU (32), and the loss weight ω = "0.1" for training, we conducted controlled experiments to explore the effect of the training batch size on the model performance. By setting the batch size as "4, 8" respectively as in Table 5, it was observed with gradual improvement of the model in terms of averaged means and standard deviations of key performance metrics, while the accuracy deviated in larger extent as the batch size further increased from "8" to "16". Based on these findings, we trained our final proposed model with batch size as "8" to make a balance between the training speed and the memory consumption. It should be noted that when batch size is smaller than "8", such as "4", we randomly add one negative sample to each mini-batch, which means that there would be "3" normal samples together with 1 abnormal sample in a mini-batch. When batch size is equal or larger than "8", we followed the "1/8 portion" training strategy described in Section 6.2.

6.2.4. Model compression

In previous sections, we have verified the effectiveness of the dilated CNN+GRU framework in solving the condition monitoring problem that we formulated in Section 3. It is valuable to explore the optimal model configurations with leveraged efficiency and accuracy. Consequently, we made attempts regarding the two main components of the network – the dilated CNN and the GRU respectively. To be more specific, we first tried to halve the channel number of the last three dilated convolutional layers, from (128, 256, 128) to (64, 128, 64). We then halved the number of neurons in GRU cell from "32" to "16". The performance of the models has then been evaluated after splitting the dataset, the result of which will be discussed in the next session.

6.2.5. Model performance comparison after splitting dataset

As mentioned, in order to thoroughly evaluate the robustness of the model in different compression scenarios, we have split the training and evaluation sets for multiple times while testing on the identical test set correspondingly. To systematically quantify the model performance under different parameter configurations, we calculated the mean *(μ)* and standard deviation *(σ)* values for "Precision", "Recall" and "F1-Score" values accordingly under five typically split dataset. As shown in Table 6, among the three models as M_1= Dilated CNN (256) + GRU (32), M_2= Dilated CNN (128) + GRU (32) and M_3= Dilated CNN (256) + GRU (16), the μ of M_1 achieved the largest value, indicating the model achieved higher accuracy and precision for predicting the health status of elevator under both normal and abnormal conditions. After comparing the standard deviation among the three models, it was found that the σ values of M_1 also demonstrated the least variation among the three candidate models, thus showing the best stability and robustness of this model.

Table 4

Performance of models with different architectures.

	Splitting		Confusion Matrix		Precision	Recall	F1-Score	Support
CNN+MLP	*0*	*Abnormal*	13(TP)	2 (FN)	0.93	0.87	0.90	15
		Normal	1 (FP)	1007(TN)	1.00	1.00	1.00	1008
	1	*Abnormal*	13	2	0.81	0.87	0.84	15
		Normal	3	1005	1.00	1.00	1.00	1008
	4	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
	5	*Abnormal*	10	5	0.77	0.67	0.71	15
		Normal	3	1005	1.00	1.00	1.00	1008
	9	*Abnormal*	12	3	0.86	0.80	0.83	15
		Normal	2	1006	1.00	1.00	1.00	1008
				*Ab. M. (**μ**)*	**0.87**	**0.83**	**0.85**	
				*Ab. S. (**σ**)*	**0.09**	**0.10**	**0.10**	
CNN+GRU	*0*	*Abnormal*	13	2	1.00	0.87	0.93	15
		Normal	0	1008	1.00	1.00	1.00	1008
	1	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	.0	1008	1.00	1.00	1.00	1008
	4	*Abnormal*	13	2	1.00	0.87	0.93	15
		Normal	0	1008	1.00	1.00	1.00	1008
	5	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
	9	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
				*Ab. M. (**μ**)*	**1.00**	**0.91**	**0.95**	
				*Ab. S. (**σ**)*	**0**	**0.03**	**0.02**	
Dilated CNN+MLP	*0*	*Abnormal*	14	1	0.82	0.93	0.87	15
		Normal	3	1005	1.00	1.00	1.00	1008
	1	*Abnormal*	11	4	0.73	0.73	0.73	15
		Normal	4	1004	1.00	1.00	1.00	1008
	4	*Abnormal*	14	1	0.88	0.93	0.90	15
		Normal	2	1006	1.00	1.00	1.00	1008
	5	*Abnormal*	15	0	0.79	1.00	0.88	15
		Normal	4	1004	1.00	1.00	1.00	1008
	9	*Abnormal*	12	3	0.71	0.80	0.75	15
		Normal	5	1003	1.00	1.00	1.00	1008
				*Ab. M. (**μ**)*	**0.79**	**0.88**	**0.83**	
				*Ab. S. (**σ**)*	**0.07**	**0.11**	**0.08**	
Dilated CNN+GRU	*0*	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
	1	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
	4	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
	5	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
	9	*Abnormal*	14	1	1.00	0.93	0.97	15
		Normal	0	1008	1.00	1.00	1.00	1008
				*Ab. M. (**μ**)*	**1.00**	**0.93**	**0.97**	
				*Ab. S. (**σ**)*	**0**	**0**	**0**	

Note: 1). The GRU cell used here are of 32 neurons in the gate; 2). Ab. M. & Ab. S. represent the mean and standard deviation for fault detection under abnormal conditions.

Table 5

Performance of models with different training hyperparameters.

	(batch size, loss weight)	Splitting	Confusion matrix			Precision	Recall	F1-Score	Support
	(4, **0.1**)	*0*	*Abnormal*	14(TP)	1 (FN)	1.00	0.93	0.97	15
			Normal	0 (FP)	1008(TN)	1.00	1.00	1.00	1008
		1	*Abnormal*	15	0	1.00	1.00	1.00	15
			Normal	0	1008	1.00	1.00	1.00	1008
		4	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		5	*Abnormal*	15	0	0.06	1.00	0.11	15
			Normal	237	771	1.00	0.76	0.87	1008
		9	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
					*Ab. M. (**μ**)*	**0.81**	**0.96**	**0.80**	
					*Ab. S. (**σ**)*	**0.42**	**0**	**0**	
	(16, **0.1**)	*0*	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		1	*Abnormal*	14	1	0.93	0.93	0.93	15
			Normal	1	1007	1.00	1.00	1.00	1008
		4	*Abnormal*	13	2	1	0.87	0.93	15
			Normal	0	1008	1.00	1.00	1.00	1008
		5	*Abnormal*	14	1	0.93	0.93	0.93	15
			Normal	1	1007	1.00	1.00	1.00	1008
		9	*Abnormal*	15	0	0.83	1.00	0.91	15
			Normal	3	1005	1.00	1.00	1.00	1008
					*Ab. M. (**μ**)*	**0.94**	**0.93**	**0.93**	
					*Ab. S. (**σ**)*	**0.07**	**0**	**0**	
Dilated CNN +GRU	(8, **1**)	*0*	*Abnormal*	13	2	0.93	0.87	0.90	15
			Normal	1	1007	1.00	1.00	1.00	1008
		1	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		4	*Abnormal*	14	1	0.88	0.93	0.90	15
			Normal	2	1006	1.00	1.00	1.00	1008
		5	*Abnormal*	14	1	0.93	0.93	0.93	15
			Normal	1	1007	1.00	1.00	1.00	1008
		9	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
					*Ab. M. (**μ**)*	**0.95**	**0.92**	**0.93**	
					*Ab. S. (**σ**)*	**0.05**	**0**	**0**	
	(8, **0.1**)	*0*	*Abnormal*	14	1	1	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		1	*Abnormal*	14	1	1	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		4	*Abnormal*	14	1	1	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		5	*Abnormal*	14	1	1	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		9	*Abnormal*	14	1	1	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
					*Ab. M. (**μ**)*	**1**	**0.93**	**0.97**	
					*Ab. S. (**σ**)*	**0**	**0**	**0**	

Note: 1). The GRU cell used here are of 32 neurons in the gate; 2). Ab. M. & Ab. S. represent the mean and standard deviation for fault detection under abnormal conditions.

Table 6

Performance evaluation after compression with different model configurations.

	Model Configuration	Splitting		Confusion Matrix		Precision	Recall	F1-Score	Support
Dilated CNN +GRU	$\boldsymbol{M_1}$= Dilated_CNN_256_GRU_32 (16, 128, **256**, 128, **32**)	*0*	*Abnormal*	14(TP)	1 (FN)	1.00	0.93	0.97	15
			Normal	0 (FP)	1008(TN)	1.00	1.00	1.00	1008
		1	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		4	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		5	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		9	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
					*Ab. M. (**μ**)*	**1.00**	**0.93**	**0.97**	
					*Ab. S. (**σ**)*	**0**	**0**	**0**	
	$\boldsymbol{M_2}$= Dilated CNN_128_GRU_32 (16, 64, **128**, 64, **32**)	*0*	*Abnormal*	14	1	0.88	0.93	0.90	15
			Normal	2	1006	1.00	1.00	1.00	1008
		1	*Abnormal*	14	1	0.82	0.93	0.87	15
			Normal	3	1005	1.00	1.00	1.00	1008
		4	*Abnormal*	14	1	0.93	0.93	0.93	15
			Normal	1	1007	1.00	1.00	1.00	1008
		5	*Abnormal*	13	2	1.00	0.87	0.93	15
			Normal	0	1008	1.00	1.00	1.00	1008
		9	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
					*Ab. M. (**μ**)*	**0.93**	**0.92**	**0.92**	
					*Ab. S. (**σ**)*	**0.08**	**0.03**	**0.04**	
	$\boldsymbol{M_3}$= Dilated_CNN_256_GRU_16 (16, 128, **256**, 128, **16**)	*0*	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
		1	*Abnormal*	13	2	1.00	0.87	0.93	15
			Normal	0	1008	1.00	1.00	1.00	1008
		4	*Abnormal*	13	2	0.93	0.87	0.90	15
			Normal	1	1007	1.00	1.00	1.00	1008
		5	*Abnormal*	15	0	1.00	1.00	1.00	15
			Normal	0	1008	1.00	1.00	1.00	1008
		9	*Abnormal*	14	1	1.00	0.93	0.97	15
			Normal	0	1008	1.00	1.00	1.00	1008
					*Ab. M. (**μ**)*	**0.99**	**0.92**	**0.95**	
					*Ab. S. (**σ**)*	**0.03**	**0.05**	**0.04**	

Note: 1). Parameters (a,b,c,d,e) in the column of Model configuration represent the number of filters for the four convolutional layers and the number of neurons in the gate of GRU cell; 2). Ab. M. & Ab. S. represent the mean and standard deviation for fault detection under abnormal conditions.

It was thus concluded that the more complex the model, the more robust and stable the model appeared in the mentioned dataset for different configuration scenarios. Under the case of M_3, though the larger σ value was witnessed while representing larger performance deviations in comparison with M_1, whereas the *mean value* of precision metric (μ=0.99) almost scored the same high with M_1(μ=1.00). Consequently, considering the demand with real-case applications of the proposed DL model into the lift sites in normal operation, the streamlined architecture (in M_3) would help boost the inferencing and computational speed under leaner architectures. Hence, considering the minor difference in accuracy performance deviations, together with the higher demand for robustness on the elevator sites, we would choose M_3= Dilated CNN (256) + GRU (16) as the appropriate model for practical application.

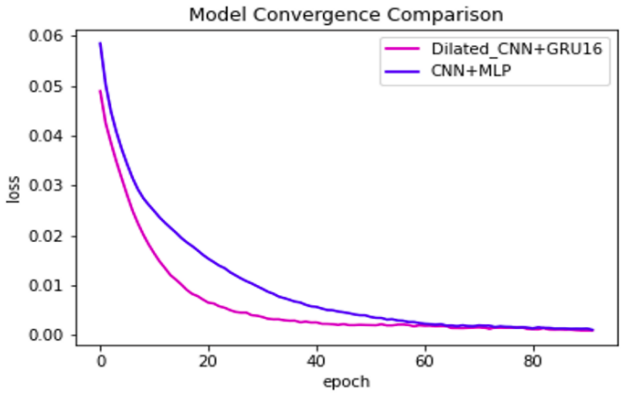


Fig. 11. Comparison of loss reduction speed and convergence performance.

6.2.6. Model efficiency and convergence performance

Regarding the model efficiency, it was proved that the quantity of parameters were reduced by 21% from around "268K" to "211K" for the proposed Dilated CNN+GRU (16) in comparison with the vanilla CNN+MLP model, thus largely improving the model performance with lessened computation complexity and lightweight architecture. Meanwhile, the receptive field was also enlarged by "75%" from the original (126×32) in the vanilla CNN+MLP structure to the current (220×32) under the optimal architecture with the Dilated CNN+GRU (16) model. Regarding the model convergence performance during the training process, it could be validated by Fig. 11 that the proposed Dilated CNN+GRU (16) model demonstrated comparatively faster convergence speed under fewer epochs' iteration in terms of loss reduction compared with the vanilla CNN+MLP model. Consequently, the optimal model of Dilated CNN+GRU (16) was derived with more lightweight architecture and enhanced efficiency in simplicity and elegance.

The model effectiveness could be validated by the following perspectives, on the one hand, by replacing the vanilla convolutional layer in the backbone with dilated convolutional layer, the receptive field was expanded dramatically with fewer parameters; on the other hand, the concatenation of GRU layer instead of the trivial MLP could enhance the temporal correlation among the extracted feature representations to enhance the overall efficiency.

By specifying model structure as the Dilated CNN+GRU (16) and setting batch size equal to "8" with loss weight as "0.1" in the weighted cross-entropy loss, the testing results were leveraged among the best evaluation matrix outcomes, which demonstrated the superior performance of the proposed model. Meanwhile, the core metrics of precision, recall and F1-Score for normal conditions also reached "1.00" with good capability in recognizing and distinguishing the normal situations during elevator condition monitoring. Furthermore, the anomaly detector conducted the role of the classifier with the optimal threshold learned through the training process to determine the elevator status.

7. Managerial implications

The routine safety surveillance on elevators during the practical operation process proves to be a crucial topic for corporate managerial and executive personnel in relevant elevator manufacturers and maintenance-responsible enterprises. By the prevailing measures taken for anomaly detection on elevator's electrical and mechanical (E&M) systems, the most commonly used post-hoc routine inspection manner would inevitably cause unexpected delays for detecting the faults with gradual degradation of safety-critical components. Moreover, such conventional methods are mostly restricted by the proprietary intrusive equipment while lacking applicability on elevators of different brands in a broad scope. With the emerging artificial intelligence technology, under a wide plethora of research utilizing data-driven methods for condition monitoring, it has been recognized as a critical topic to tackle the influence of the imbalanced dataset issue. Especially for deep learning-based data-driven approaches, the imbalanced dataset would cause the inadequate extraction of multi-level and cross-hierarchical feature dependencies, while leading to the insufficient learning of data characteristics with suboptimal fault detection accuracy. Though there exist devious workaround routes such as the data manipulation with laboratory synthetic experiments, or the stacking of complicated deep learning model architectures trying to increase the complexity of models with the presumably better feature learning capabilities, the side effects could not be neglected as well while impairing the monitoring effectiveness. For instance, laboratory synthetic data manipulation would lead to the real data generated in the highly dynamic actual elevator operation conditions not being captured and learnt, while decreasing the model's generalization capability during practical deployment; whilst the complexity increasement would also lead to excessive model parameters as generated, thus affecting the model's inferencing speed and computation efficiency. Especially for real deployment to actual elevator operation sites, much greater demands are put forward to the proposed solutions with streamlined lean architectures without excessive computation power consumption while handling a wide variety of real-case imbalanced dataset issues with good generalization capability. Hence, a delicate balance point might be required for the proposed feasible solutions customized for real-case applications to the elevator condition monitoring domain.

Anchored in the practical requirement as mentioned above, this paper proposes a novel deep learning model with a lightweight architecture without adding excessive computation burdens, while the devised training strategy and other embedded algorithmic designs could effectively alleviate the impacts of imbalanced dataset as well during the actual implementation process. There are a couple of manifold implications for the managerial level of enterprise from the following perspectives by the proposed work:

1). Innovative Monitoring Manner. Distinct from the conventional prevailing measure with routine site inspection and maintenance on specific components using proprietary equipment, the proposed solution was established based on a non-invasive framework while monitoring the electric current signals flowing through the core components, namely the traction motor, brake system, safety circuit and lift door. By observing the patterns of electric current signals, the health condition of the critical safety-

relevant components can be monitored remotely with anomaly detection performed by the proposed deep learning model. Whilst the data are collected via Internet-of-Things (IoT) sensors and transferred via a Cloud-based data hub with constant surveillance on the functionality of elevators. Such innovative monitoring manner with the adoption of IoT sensors and AI deep learning analytical models can provide an efficient smart way of monitoring in comparison with conventional methods, while eliminating the cumbersome efforts of onsite regular checking and post-hoc maintenance after the occurrence of elevator accidents, thus improving the overall safety level of elevators in routine operations.

2). Enhanced Feature Extraction Capability. The proposed model can enhance the temporal feature correlations by the devised gated recurrent unit (GRU) architecture, especially capable of learning the multi-scale and cross-hierarchical timeseries data patterns. Due to the electric current signals are collected constantly under high sampling rates by the IoT sensors, hence the temporal feature correlations play a decisive role in analysing the multi-channel current signals for fault detection. Meanwhile, the proposed model has concatenated the dilated convolutional neural network (DiCNN), while largely increasing the receptive fields by 75% in comparison with the state-of-the-art baseline model, thus stimulating the neuronal response with evoked feature correlations and enhanced learning effectiveness on the multi-variant data characteristics.

3). Improved Inferencing Speed. By the experimental validation, it is proved that the proposed Dilated CNN+GRU (16) model demonstrated comparatively faster convergence speed under fewer epochs' iteration in terms of loss reduction compared with the vanilla CNN+MLP baseline model. Such performance factors as emphasized by the managerial level of companies could affect the model's training efficiency and inference speed with positive correlations. For elevators installed under the highly dynamic real operational environment, the improved training and inference speed shall be a crucial evaluation point for managers while considering the applicability of the proposed model to cater the real-time elevator monitoring circumstances.

4). Model Elegancy with Non-excessive Complexity. By the proposed model, the parameter quantity has been dramatically reduced by 21% in comparison with the CNN+MLP baseline model. Rather than stacking the complex deep learning architectures, the proposed lightweight model presents a streamlined architecture with largely reduced model parameters, which could provide a useful implication for managerial level in selection of the appropriate models with reduced complexity and enhanced efficiency.

5). Delicate Balance Between Model Efficiency and Ingenious Imbalanced Dataset Tackling Methods. There exist various methods for alleviating the impacts of the imbalanced dataset issue, yet could be overshadowed with inferior performance against the complex dataset as composed of the elevator's real-operational heterogeneous electric current signal data. Alternatively, the laboratory synthesis of minority class samples would impair the model's generalization capability, whilst the stacking of complicated redundant model structures in remedy of the insufficient minority-class feature learning could affect the

model's inferencing speed during actual deployment. Hence, under the real-case high demands for real-time monitoring, how to delicately implement a feasible way for tackling the imbalanced dataset issue while retaining the model's structural elegance without adding additional computation burdens could be one of the top concerns for managerial levels. By the proposed devised training strategy and customized loss function, the model has achieved superior performance with the high precision value after the systematic ablation studies, thus justifying the effectiveness of the proposed methods for addressing and alleviating the imbalanced dataset issue for elevator condition monitoring based on the real-collected dataset during the actual elevator operation process.

For the corporate managers or investors in elevator manufacturing or maintenance responsible enterprises, all the aforementioned implications could be taken into consideration to support their strategic decision-making while aiming to construct the holistic non-intrusive elevator monitoring framework during the digital transformation process. On the one hand, this paper provides one of the pioneering works for utilizing the deep learning-based data-driven models in the domain of non-intrusive elevator condition monitoring, the framework of which brings considerable reference value for managerial-level personnel in such companies. On the other hand, this paper proposes a relatively lean architectural framework in adoption of the streamlined deep learning model with reduced parameters and enhanced feature extraction capability to boost the lean digitalization (Zhang et al., 2025) for such companies, while laying a solid foundation to facilitate further in-depth research and application in relevant domains. Moreover, the proposed model is customized with wide applicability using real-case elevator operation data and feasible measures to tackle the imbalanced dataset issue, thus largely enhancing the model's generalization capability during the practical commissioning process by the managerial-level personnel. Consequently, the work proposed in this paper has brought manifold managerial implications from different perspectives and paved the way for constructing the AI-empowered smart elevator monitoring framework with advanced engineering management technologies for broad practical applications.

8. Conclusions

This paper has proposed a novel deep learning-based solution for the condition monitoring and anomaly detection of elevators under imbalanced dataset to foster practical applications with diversified environmental dynamics. With the model architecture composing the customised Dilated CNN and GRU under the explored optimal configurations through the ablation study, the proposed 1D-DiCNN-GRU model demonstrated supreme robustness and accuracy in fault detection by analysing the elevator's electric current signals as non-intrusively captured in real life. The proposed solution generated the optimum performance to tackle the industrial bottleneck under imbalanced dataset scenarios with a relatively lightweight but elegant architecture while capturing both the intricate spatial and temporal data

correlations from multi-variant time-series 1D signals directly, so as to seek broader applications by industrial practitioners.

Expect for the research rigour evaluation via the devised testing scenarios and systematic ablation studies, the model's robustness was also validated through real practice by field experts in the elevator industry. As an effective supplementary system to the prevailing conventional lift monitoring mechanisms, the proposed model and system can provide a rather convenient retrofitting method by non-intrusively installing the clamp-type electric current sensors with embedded AI algorithms for routine elevator status monitoring. Such a smart approach can help industrial practitioners effectively measure the degradation status of elevators with relatively simple but elegant model architecture to propel the utilization of AI technology in safety-critical systems, which can increase their efficiency with the integrated capability for predictive maintenance.

As well acknowledged that the collection of abnormal signals proves very challenging and laborious in real life. Deemed as one of the major contributions, we also established an industrial dataset based on the non-invasive signals as captured and measured during the real observation process for lift condition monitoring. Regarding future improvement, a more robust algorithmic mechanism to tackle the imbalanced dataset issue shall be explored and refined, while facilitating the construction of a more rigorous dataset architecture with enhanced integrity for real-life data observation. More advanced data-driven analytical model by integrating with state-of-the-art DL methodologies shall be further investigated and validated to leverage the sustainability of model performance with more accurate anomaly detection and fault diagnosis results.

Acknowledgements

The work presented in this paper is supported by Centre for Advances in Reliability and Safety (CAiRS) admitted under AIR@InnoHK Research Cluster. Real-world electric current data on the elevator system provided by a CAiRS partner is also acknowledged. This paper also contributes to the PhD thesis of author Ye Wang for acknowledgement.

References

Aheleroff, S., Xu, X., Lu, Y., Aristizabal, M., Pablo Velásquez, J., Joa, B., & Valencia, Y. (2020). IoT-enabled smart appliances under industry 4.0: A case study. *Advanced Engineering Informatics*, *43*, 101043.

Analide, C., Novais, P., Camacho, D., & Yin, H. (2020). A Preprocessing Approach for Class-Imbalanced Data Using SMOTE and Belief Function Theory. In (Vol. 12490). Springer International Publishing AG.

Arora, V., Yin-Kwee Ng, E., & Singh, A. (2022). Chapter One - Machine learning and its applications. In R. Sehgal, N. Gupta, A. Tomar, M. D. Sharma, & V. Kumaran (Eds.), *Smart Electrical and Mechanical Systems* (pp. 1-37). Academic Press.

Bisong, E. (2019). *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*. Apress L. P.

Chai, S., Li, X. I., Jia, Y., He, Y., Yip, C. H., Cheung, K. K., & Wang, M. (2021). A Non-Intrusive Deep Learning Based Diagnosis System for Elevators. *IEEE Access*, *9*, 20993-21003.

Chen, Z., Guo, R., Lin, Z., Peng, T., & Peng, X. (2021). A Data-Driven Health Monitoring Method Using Multiobjective Optimization and Stacked Autoencoder Based Health Indicator. *IEEE Transactions on Industrial Informatics*, *17*(9), 6379-6389.

Chung, J., Gulcehre, C., Cho, K., & Bengio, Y. (2014, December 2014). Empirical evaluation of gated recurrent neural networks on sequence modeling. NIPS 2014 Workshop on Deep Learning,

Dai, X., & Gao, Z. (2013). From Model, Signal to Knowledge: A Data-Driven Perspective of Fault Detection and Diagnosis. *IEEE Transactions on Industrial Informatics*, *9*(4), 2226-2238.

de Oliveira, J. P. G., J.A. Bastos-Filho, C., & Oliveira, S. C. (2022). Non-invasive embedded system hardware/firmware anomaly detection based on the electric current signature. *Advanced Engineering Informatics*, *51*, 101519.

Dong, Y., & Wang, X. (2011). A New Over-Sampling Approach: Random-SMOTE for Learning from Imbalanced Data Sets. Berlin, Heidelberg.

Douzas, G., & Bacao, F. (2019). Geometric SMOTE a geometrically enhanced drop-in replacement for SMOTE. *Information sciences*, *501*, 118-135.

EMSD. (2022a). *Code of Practice on the Design and Construction of Lifts and Escalators*. Retrieved from https://www.emsd.gov.hk/en/lifts_and_escalators_safety/publications/code_of_practice/index.html

EMSD. (2022b). *Lifts and Escalators Ordinance (Cap. 618).* https://www.elegislation.gov.hk/hk/cap618?tab=m&xpid=ID_1438403530602_001

EMSD. (2022c). *The summaries of the reported lift incidents.* https://www.emsd.gov.hk/en/lifts_and_escalators_safety/publications/contractors_performance_rating/reported_lift_incident_records/

Esteban, E., Salgado, O., Iturrospe, A., & Isasa, I. (2014). Model-based estimation of elevator rail friction forces. International Conference on Condition Monitoring of Machinery in Non-Stationary Operation,

Fang, W., Shao, Y., Love, P. E. D., Hartmann, T., & Liu, W. (2023). Detecting anomalies and de-noising monitoring data from sensors: A smart data approach. *Advanced Engineering Informatics*, *55*, 101870.

Feng, J., Yao, Y., Lu, S., & Liu, Y. (2021). Domain Knowledge-Based Deep-Broad Learning Framework for Fault Diagnosis. *IEEE transactions on industrial electronics (1982)*, *68*(4), 3454-3464.

Fu, Y., Liu, X., Sarkar, S., & Wu, T. (2021). Gaussian mixture model with feature selection: An embedded approach. *Computers & Industrial Engineering*, *152*, 107000.

Gao, Y., Zhang, H., Ding, L., Zhang, S., & Zeng, C. (2025). Overcoming the Digital Dilemmas: How Do Enterprises Develop Digital Transformative Capability? *IEEE Transactions on Engineering Management*, *72*, 1653-1667.

García, Á., Bregon, A., & Martínez-Prieto, M. A. (2022). A non-intrusive Industry 4.0 retrofitting approach for collaborative maintenance in traditional manufacturing. *Computers & Industrial Engineering*, *164*, 107896.

Giovanni, P. D., Belvedere, V., & Grando, A. (2024). The Selection of Industry 4.0 Technologies Through Bayesian Networks: An Operational Perspective. *IEEE Transactions on Engineering Management*, *71*, 2921-2936.

González, M., Salgado, O., Hernandez, X., Croes, J., Pluymers, B., & Desmet, W. (2019). Model-based condition monitoring of guiding rails in electro-mechanical systems. *Mechanical Systems and Signal Processing*, *120*, 630-641.

Haibo, H., Yang, B., Garcia, E. A., & Shutao, L. (2008, 2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning.

Hennig, M., Grafinger, M., Hofmann, R., Gerhard, D., Dumss, S., & Rosenberger, P. (2021). Introduction of a time series machine learning methodology for the application in a production system. *Advanced Engineering Informatics*, *47*, 101197.

Huo, Y., Gu, Q., Cai, Z., & Yuan, L. (2015). Classification method for imbalance dataset based on genetic algorithm improved synthetic minority over-sampling technique. *Jisuanji Yingyong / Journal of Computer Applications*, *35*(1), 121-139.

Ilango, H. S., Ma, M., & Su, R. (2022). A FeedForward–Convolutional Neural Network to Detect Low-Rate DoS in IoT. *Engineering Applications of Artificial Intelligence*, *114*, 105059.

Jaenal, A., Ruiz-Sarmiento, J.-R., & Gonzalez-Jimenez, J. (2024). MachNet, a general Deep Learning architecture for Predictive Maintenance within the industry 4.0 paradigm. *Engineering Applications of Artificial Intelligence*, *127*, 107365.

Khan, A. Q., El Jaouhari, S., Tamani, N., & Mroueh, L. (2024). Knowledge-based anomaly detection: Survey, challenges, and future directions. *Engineering Applications of Artificial Intelligence*, *136*, 108996.

Koltun, F. Y. V. (2016, 30 Apr 2016). *Multi-scale Context Aggregation by Dilated Convolutions* International Conference on Learning Representations,

Koutsoupakis, J., Giagopoulos, D., & Chatziparasidis, I. (2023). AI-based condition monitoring on mechanical systems using multibody dynamics models. *Engineering Applications of Artificial Intelligence*, *123*, 106467.

Kumar Sharma, D., Brahmachari, S., Singhal, K., & Gupta, D. (2022). Data driven predictive maintenance applications for industrial systems with temporal convolutional networks. *Computers & Industrial Engineering*, *169*, 108213.

Lakemond, N., Holmberg, G., & Pettersson, A. (2024). Digital Transformation in Complex Systems. *IEEE Transactions on Engineering Management*, *71*, 192-204.

Li, H., Cao, P., Wang, X., Yi, B., Huang, M., Sun, Q., & Zhang, Y. (2023). Multi-task spatio-temporal augmented net for industry equipment remaining useful life prediction. *Advanced Engineering Informatics*, *55*, 101898.

Lin, Z., Duan, D., Yang, Q., Cheng, X., Yang, L., & Cui, S. (2018, 26-29 Nov. 2018). One-class Clasifier Based Fault Detection in Distribution Systems with Distributed Energy Resources. 2018 IEEE Global Conference on Signal and Information Processing (GlobalSIP),

Lingxiao, Q., Ximing, Y., Yi, W., & Jin, C. (2025). Deciphering AI's Role in Corporate Innovation: A Holistic Framework of AI Resources, Capability, and Performance. *IEEE Engineering Management Review*, *53*(2), 85-95.

Liu, C., Zhang, X., Liu, X., & Chen, C. (2017). The research of elevator health diagnosis method based on Bayesian network. IOP Conference Series: Earth and Environmental Science,

Liu, Y., Chen, H., Wang, Y., & Zhang, P. (2021). Power pooling: An adaptive pooling function for weakly labelled sound event detection. 2021 International Joint Conference on Neural Networks (IJCNN),

Mahadevan, S., & Shah, S. (2009). Fault Detection and Diagnosis in Process Data Using One-Class Support Vector Machines. *Journal of Process Control*, *19*, 1627-1639.

Mazzoleni, M., Sarda, K., Acernese, A., Russo, L., Manfredi, L., Glielmo, L., & Del Vecchio, C. (2022). A fuzzy logic-based approach for fault diagnosis and condition monitoring of industry 4.0 manufacturing processes. *Engineering Applications of Artificial Intelligence*, *115*, 105317.

Medghalchi, B., & Vogel, A. (2024). A machine learning-based algorithm for automated detection of frequency-based events in recorded time series of sensor data. *Engineering Applications of Artificial Intelligence*, *133*, 108536.

Mishra, K. M., & Huhtala, K. (2019). Elevator Fault Detection Using Profile Extraction and Deep Autoencoder Feature Extraction for Acceleration and Magnetic Signals. *Applied Sciences*, *9*(15).

Moura, J., Batista, F., Cardoso, E., & Nunes, L. (2019). Intelligent management and efficient operation of Big Data. In *Web Services: Concepts, Methodologies, Tools, and Applications* (pp. 1991-2016). IGI Global.

Nguyen, T. N., Ngo, Q.-D., Nguyen, H.-T., & Giang, N. L. (2022). An Advanced Computing Approach for IoT-Botnet Detection in Industrial Internet of Things. *IEEE Transactions on Industrial Informatics*, 1-1.

Rajnoha, R., & Hadač, J. (2024). Strategic Key Elements in Big Data Analytics as Driving Forces of IoT Manufacturing Value Creation: A Challenge for Research Framework. *IEEE Transactions on Engineering Management*, *71*, 90-105.

Skog, I., Karagiannis, I., Bergsten, A. B., Harden, J., Gustafsson, L., & Handel, P. (2017). A Smart Sensor Node for the Internet-of-Elevators—Non-Invasive Condition and Fault Monitoring. *IEEE Sensors Journal*, *17*(16), 5198-5208.

Tang, J., Liu, Y., Lin, K.-y., & Li, L. (2023). Process bottlenecks identification and its root cause analysis using fusion-based clustering and knowledge graph. *Advanced Engineering Informatics*, *55*, 101862.

Tomczak, J. M. (2024). *Deep Generative Modeling* (1 ed.). Springer International Publishing AG.

Wan, Z., Yi, S., Li, K., Tao, R., Gou, M., Li, X., & Guo, S. (2015). Diagnosis of Elevator Faults with LS-SVM Based on Optimization by K-CV. *Journal of Electrical and Computer Engineering*, *2015*, 1-8.

Wang, J., Lim, M. K., Wang, C., & Tseng, M.-L. (2021). The evolution of the Internet of Things (IoT) over the past 20 years. *Computers & Industrial Engineering*, *155*, 107174.

Wen, P., Zhi, M., Zhang, G., & Li, S. (2016). Fault Prediction of Elevator Door System Based on PSO-BP Neural Network. *Engineering*, *08*(11), 761-766.
Wileman, A., Perinpanayagam, S., & Aslam, S. (2021). Physics of Failure (PoF) Based Lifetime Prediction of Power Electronics at the Printed Circuit Board Level. *Applied Sciences*, *11*(6).
Xiahou, T., Zeng, Z., & Liu, Y. (2021). Remaining Useful Life Prediction by Fusing Expert Knowledge and Condition Monitoring Information. *IEEE Transactions on Industrial Informatics*, *17*(4), 2653-2663.
Xu, C., & Lin, B. (2025). Embracing Artificial Intelligence: How Does Intelligent Transformation Affect the Technological Innovation of New Energy Enterprises? *IEEE Transactions on Engineering Management*, *72*, 703-716.
Xu, X., Li, X., Ming, W., & Chen, M. (2022). A novel multi-scale CNN and attention mechanism method with multi-sensor signal for remaining useful life prediction. *Computers & Industrial Engineering*, *169*, 108204.
Yan, J., He, Z., & He, S. (2022). A deep learning framework for sensor-equipped machine health indicator construction and remaining useful life prediction. *Computers & Industrial Engineering*, *172*, 108559.
Yin, S., Ding, S. X., Xie, X., & Luo, H. (2014). A Review on Basic Data-Driven Approaches for Industrial Process Monitoring. *IEEE Transactions on Industrial Electronics*, *61*(11), 6418-6428.
Yu, Y., Lakemond, N., & Holmberg, G. (2024). AI in the Context of Complex Intelligent Systems: Engineering Management Consequences. *IEEE Transactions on Engineering Management*, *71*, 6512-6525.
Zeng, Z., Kang, R., & Chen, Y. (2016). Using PoF models to predict system reliability considering failure collaboration. *Chinese journal of aeronautics*, *29*(5), 1294-1301.
Zhang, S., Bi, X., Tian, R., Chen, T., Ding, K., & Yu, Z. (2025). Lean Digitalization Solutions for SMEs in China. *IEEE Transactions on Engineering Management*, *72*, 1546-1558.
Zhou, Z., Huang, H., & Fang, B. (2021). Application of Weighted Cross-Entropy Loss Function in Intrusion Detection. *Journal of Computer and Communications*, *09*(11), 1-21.