Energy 221 (2021) 119775

Contents lists available at ScienceDirect

Energy

journal homepage: www.elsevier.com/locate/energy

AI-assistance for predictive maintenance of renewable energy systems

Won Shin [a,1], Jeongyun Han [b,1], Wonjong Rhee [b,*]

[a] ONYX InSight Analytics Ltd, Nottingham, NG7 2TU, United Kingdom
[b] Seoul National University, Department of Transdisciplinary Studies, 1 Gwanak-ro, Gwanak-gu, Seoul, 08826, Republic of Korea

ARTICLE INFO

Article history:
Received 4 March 2020
Received in revised form
30 December 2020
Accepted 3 January 2021
Available online 6 January 2021

Keywords:
AI-Assistance
Predictive maintenance
Image-based diagnostics
Wind energy
Bearing fault

ABSTRACT

Although promising results of high-performance AI algorithms have been reported in recent predictive maintenance researches, most of the existing studies merely deal with AI-only solutions and do not consider the interaction between humans and AI. In this study, we explicitly focus on the benefits of interactions where a human inspector is assisted by AI solutions. A case study is conducted for predictive maintenance of wind farms, where endoscopic images were used for bearing fault detection. The experiment consisted of 54 technical inspectors and 2301 images collected over 138 wind turbines, and each inspector was shown images and asked to identify bearing faults in the absence and presence of AI-assistance. The results showed that AI-assistance had a statistically significant impact on improving the technical inspector's specificity and time efficiency. The level of improvement was dependent on the level of expertise, where the generalist group showed greater improvements in specificity and time efficiency (24.6% and 25.3%, respectively) when compared with the specialist group (4.7% and 6.4%, respectively). Both groups responded positively on the reuse intention and usefulness of AI-assistance, and the change in cognitive load was not statistically significant.

1. Introduction

In the renewable energy industry, operation and maintenance (O&M) costs are an essential part of commercial success and play a crucial role in important measures such as levelized cost of energy (LCOE) [1,2]. To minimize O&M cost, it is critical to identify faults in the *early stages*. Once a fault progresses to a more severe stage, the cost for repair can become exponentially larger and the time until full repair can become extensively longer. Compared with other machinery industries, the impact can be excessive in the renewable energy industry. For instance, consider an offshore wind farm. If the damage to a major component is found late, manufacturing and transportation of a massive component and scheduling of a crane vessel for lifting will be required. Because the logistics can be complicated, the wind turbine might need to be kept off for several months, thus resulting in a large loss of revenue in addition to the repair cost. Therefore, it is important to predict a failure in advance such that catastrophic equipment damage can be prevented and time-consuming repair scheduling can be done in time. Generally, such activities for in-service equipment are called *predictive maintenance*.

For predictive maintenance, the industry has actively embraced new technologies that utilize sensors and imaging devices [3–8]. In this study, we focus on *image-based inspection*, where visual images collected from various devices such as endoscopes and thermal imaging cameras are utilized [3,4]. With the latest imaging devices, inspectors can examine critical mechanical parts that were previously difficult to access (e.g., internal components of a gearbox). Therefore, image-based inspection is becoming a core part of routine maintenance. As an example, a leading wind turbine manufacturer has included endoscopic inspection of a generator in its scheduled service list [6].

The basic mode of image-based inspection is the *human-only* approach where the human inspectors in the field determine if a fault is present. With the recent advances in deep learning techniques, numerous studies have analyzed the potential of the *AI-only* approach and demonstrated promising results. For example, a previous study [9] applied convolution neural networks (CNNs) for fault detection of photovoltaic cell defects and achieved 93.02% accuracy. Another study [10] employed region-based CNN (R–CNN) to automate the training of object detection in a real-time manner, while one study [11] used R–CNN to detect five types of structural surface damages on bridges and achieved a mean average precision

* Corresponding author.
E-mail addresses: won.shin@onyxinsight.com (W. Shin), hanjy@snu.ac.kr (J. Han), wrhee@snu.ac.kr (W. Rhee).
[1] These authors contributed equally.

https://doi.org/10.1016/j.energy.2021.119775

(mAP) of 87.8%. A recent survey indicated that predictive maintenance is expected to be one of the first fields where AI-based technologies will be successfully put into practice [12].

Despite the impressive performance and increasing interest, the AI-only approach has its own limitations. First, a sufficiently large-scale dataset is required for the training of an AI-only solution, and it takes time and effort to collect the labeled dataset [13]. Furthermore, data collection might be extremely difficult, if not impossible, for some failure modes that occur rarely. For such rare cases, AI-only solutions will be helpless, whereas we can expect experienced inspectors to be additionally cautious toward addressing such situations. Second, the current AI-only solutions can typically detect five or fewer failure modes where a real in-field inspection often requires handling of a large number of failure modes. This issue is pronounced for in-situ inspections where all the inspections have to be performed during normal operation. For instance, the ISO 15243 standard specifies 16 classes of failure modes for bearing fault diagnosis [14].

Because AI-only solutions are unlikely to be able to perform fully automated fault diagnosis on their own, employing the *AI-assistance* approach has garnered attention in cases where human inspectors are assisted by AI algorithms [13,15]. In this integrated approach, AI provides diagnostic suggestions to human inspectors to assist them in making a final decision. Such assistance can be expected to lead to better predictive maintenance by overcoming several limitations of human inspection, such as fatigue and variability among inspectors [16,17]. In this aided decision-making process, measuring the performance of AI-assistance systems in isolation is not sufficient to understand its full effect. Thus, it is crucial to investigate the *comprehensive effect* to actual users. In the machinery field, however, previous studies have mainly focused on the isolated performance of AI-only solutions (e.g., Refs. [18–20]). Especially, in the field of renewable energy, the concept of AI-assistance has not been investigated sufficiently, and it is the main subject of this study.

When investigating the comprehensive effect of AI, task proficiency of a human inspector should be considered. We can expect generalists with relatively low task proficiency to benefit from AI suggestions, but its influence to the specialists with professional inspection skills might or might not be positive [21]. In addition, the effects of AI-assistance need to be investigated not only in terms of performance but also in terms of perception. Perceptions such as perceived cognitive load, usefulness, and intention to reuse are crucial human factors to be considered when a new technology has to be introduced [22,23]. Even when a new technology has many proven technical benefits for improving the performance of users, its adoption is often unsuccessful when there is a lack of acceptance because of the users' poor perception of it. Therefore, investigating the perception on AI-assistance is an essential research topic for its effective application.

In this regard, this study aims to investigate the effect of AI-assistance on both performance and perception of human inspectors by considering their task proficiency. For this purpose, the research questions were as follows:

- How does AI-assistance affect the performance and perception of human inspectors?
- How do the effects differ between the generalist group and specialist group?

To address these questions, we developed an AI model based on a deep learning algorithm, specifically convolutional neural network (CNN), utilizing the labeled endoscope image dataset. The AI model assisted human inspectors' in the tasks of detecting bearing faults for treatment in this study. The effects of AI models were evaluated by two-way ANOVA based on two factors—existence of AI-assistance and level of task proficiency. The evaluation was conducted on both performance and perception of human inspectors; performance was evaluated based on the results of the inspection tasks in terms of specificity, sensitivity, and time efficiency, and perceptions were evaluated using questionnaires in terms of cognitive load, intention to reuse, and usefulness. Even though the experiment was a classification task, area under the receiver operating characteristics (AUROC) were not considered because they could not be calculated for human involved experiments. Based on this comprehensive investigation, we discuss the expected impacts of AI-assisted solutions when it is deployed to the industrial fields.

2. Background

In human-only inspection scenarios, where human inspectors make decisions manually, several issues have to be considered. First, the inspector's performance is highly dependent on their expertise level. In the experiment by Smith-Bindman R et al. [24], the rate of false-positive examinations by less experienced physicians was about 50% higher than that of more experienced physicians in screening mammography. This issue is particularly problematic for the renewable energy industry, which lacks specialists who are well trained and experienced. Second, it is difficult to ensure consistency of diagnostic decisions among multiple inspectors. In the experiment by Gulshan et al. [25], seven US certified ophthalmologists were asked to grade the severity levels of diabetic retinopathy based on photographs of the retinal fundus. The results confirmed that the consistency among ophthalmologists was very poor. Out of 683 cases, only 20% cases showed complete agreement among the ophthalmologists. Lastly, human inspectors who analyze a large number of images for a long time can make more mistakes and work inefficiently because of mental fatigue and cognitive overload [17]. With the recent advance of imaging devices, high-quality image data allow human inspectors to perform fault diagnosis on a wider range of critical parts of machines that were inaccessible before [5]. However, there is a definite limit for human inspectors to handle this growing size of image data adequately.

To address the limitations of human-only inspection, researchers have applied machine vision approaches to develop AI-based solutions for image-based fault diagnosis. CNN is the most common type of deep learning algorithm proven to be effective for image-based diagnosis in the medical field [25], and it is certainly applicable to the image-based machine fault diagnosis as well. This algorithm automatically extracts important features from an image which are required for fault diagnosis. Multiple failure modes can be trained with a sufficient amount of training data, and the AI model can determine whether there is a fault on a previously unseen image. For example, Miaoyiquan et al. [26] used an AI model based on CNN to detect three types of faults in power cables and achieved 81%–87% accuracy. Cha et al. [11] used the CNN model to detect structural surface damages on bridges and achieved a mean average precision of 87.8%. By employing AI-based solutions, technical inspection can be performed in a fully automated manner based on the image dataset with high performance.

Although the performance of AI-based models has significantly improved, the AI-only approach still has several shortcomings. The first is regarding the datasets for building these AI models. The models require a sufficient amount of data to learn important patterns for fault inspection, which means that they perform poorly when the dataset is not available. Therefore, Brynjolfsson and Mitchell [13] suggested "large data sets exist or can be created containing input-output pair" as one of the seven criteria for assessing whether AI can automate a particular task. However, because a large amount of cost and time are required to collect and maintain such large datasets, it is not easy to develop and implement these AI-based solutions to the real field in a short period of time. In particular, in some cases of fault modes that occur very rarely, collecting sufficient data with high quality labels to build AI models is challenging.

The next shortcoming is regarding the number of fault types that can be detected by AI models. AI-based solutions developed in previous studies have been focused on a limited number of fault types, while there are more types of faults required in the real technical inspection; the ISO 15243 standard classifies the types of bearing faults into 16 categories [14]. For example, the AI model in Ref. [11] was trained to detect merely five types of fault modes, and [26] used three types of fault modes to train its AI model. This situation indicates that there are still fault modes that the AI model cannot handle, which makes it difficult for the AI-only approach to perform a fully automated inspection.

Therefore, employing the solution as an AI assistant to human users can make a meaningful contribution to fault diagnosis. In this approach, the AI model assists human inspectors in making a final decision on technical diagnostic investigations by providing additional suggestions or information [15]. In this aided decision-making process, measuring the AI models' performance in isolation is not sufficient to completely understand its effect. Instead, it is crucial to investigate the effects on human inspectors in a comprehensive manner. However, to the best of our knowledge, no published work in the field of predictive maintenance has evaluated the effects of AI approaches on human inspectors. Instead, recent studies mainly focused on the performance improvement of AI-only approach (e.g., Refs. [18–20]). To complement and address the shortcomings in both AI-only and human-only approaches, it is necessary to comprehensively investigate the effects of the AI-assistance approach in which the two approaches are integrated.

In the medical field, the integrated approach has become well established where image data are actively used for diagnostic tasks. In this field, the development of a stand-alone expert system that provides definitive diagnosis was attempted before the 1990s [27]. Recently, the role of the system has been shifted from decision-maker to task-assistance which provides objective and reliable decision support for human physicians. This process is called computer-assisted diagnostics (CAD) [27,28]. CAD has been practically applied in the clinic since 1998 [29], and many studies have shown that CAD significantly improves radiologists' task performance by overcoming the shortcomings of the human-only fault diagnosis [30–35]. In addition, researchers in the medical field have examined the system's effect considering the radiologist's proficiency level for diagnosis tasks because radiologists may have different rooms for performance improvement being affected by the assistance system [21,36]. The effect of incorrect detection by CAD has been investigated at more specific levels, such as false positives and false negatives because the system cannot provide 100% correct answers [37,38]. Also, the effects of the system on user perception have to be considered. Performance gains using a new technological tool are often obstructed by the users' unwillingness to accept and use the available system. Studies on the technology acceptance model revealed that both the system's diagnostic performance and the user's perception of the system are important factors driving the user's willingness to accept the system [40]. Therefore, it is necessary to measure the user's perception to comprehensively assess the effect on users.

While many essential lessons have been learned in the medical field, there are fundamental differences between medical field and energy field. First, energy field can be more receptive to the AI technology. In the medical field, diagnosis accuracy can directly affect human's life and thus ethics and regulation immediately become a crucial part. This aspect can potentially slow down the process of adopting AI technology. For instance, CAD has been developed and proven to be effective since 1990s, but its uptake in the clinical practice has been limited [39]. In contrast, adopting AI in the energy field is primarily decided based on the economic benefits, and not so much on ethics or regulation. Therefore, the adoption can be much faster with an immediate impact. Second, energy field has both generalists and specialists while medical field has only specialists. In the medical field, stringent regulations are in place and only specialists such as licensed radiologists are qualified to work on diagnosis. In the energy field, however, both skilled specialists and less skilled generalists can work on diagnosis. In fact, more and more diagnostic tasks are being handled by generalists as predictive maintenance becomes a common and repeated practice. The benefit of assistance system is known to be dependent on the proficiency level [21,36], and a thorough investigation is required to understand the effects of AI assistance to the specialists and generalists in the energy field.

3. Dataset and AI model

To investigate the effects of AI-assistance on human inspectors, we focused on bearing fault diagnostics of wind turbines. Wind turbines are designed to convert wind's kinetic energy into electric energy through rotating components and an electrical generator. Naturally, mechanical bearings are used to constrain the motions and to reduce friction between the rotating parts. In this section, we explain our endoscope image dataset and how we have constructed the AI model for the AI-assisted system.

3.1. Dataset of endoscope bearing images

A total of 3073 endoscope images of rolling-element bearing were collected from 138 wind turbine gearboxes and main bearings. All data collections were performed using endoscopy as a part of in-situ inspection of wind turbines in operation. Because of the nature of endoscopy, the position and orientation of the bearings in the images varied. The inspected wind turbines consisted of various models manufactured by different manufacturers, and their operating age ranged from just deployed to over ten years. The endoscopes used for inspection were also of various models, and the resulting images different in brightness, resolution, and image quality. The rolling element bearing and three examples of endoscopic images in normal and abnormal conditions are shown in Fig. 1.

Among the collected images, some were not appropriate for this study. Therefore, the inappropriate images were removed from the dataset based on the following rules: (1) if the rolling bearing elements took up less than 20% of the image, (2) if the image was out of focus, and (3) if the surface of the bearing was not visible enough due to excessive oil or grease. After the filtering procedure, 772 images were removed, and the remaining 2301 images were used for this study.

3.2. Ground truth labels and data division

Ground truth labeling was required to build the AI model and conduct the experiments in this study. Ideally, longitudinal observational records of the wind turbines or independent and more accurate results based on vibration test or disassembly inspection would be used for creating the ground truth labels. In our study, however, such information was unavailable. To address the issue, we referred to the rigorous labeling process developed in the medical field [25] based on a *majority voting* over a group of expert

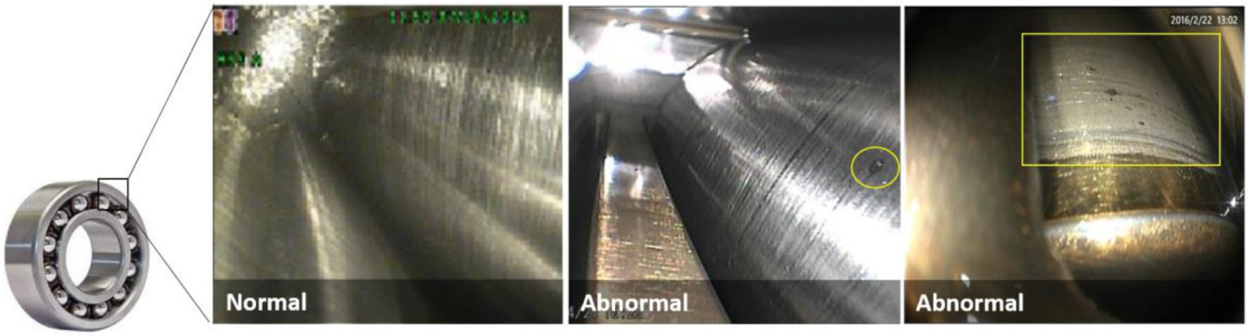


Fig. 1. Image dataset example: Rolling element bearing (far left); an endoscope image of the rolling element in the normal condition (left); and two endoscopic images in abnormal conditions (right) are shown. For the images in abnormal conditions, the fault areas are marked with a yellow circle and a yellow box. The markings are only for illustration purposes, and they are not available in the dataset. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

examiners. Following the process, images were labeled by three expert inspectors, and majority voting was applied to determine the ground truth labels. Before starting the labeling process, the inspectors were provided with a technical standard including the criteria for abnormality. After the labeling process, 1415 images were labeled as normal, and 886 images were labeled as abnormal.

The labeled dataset was divided into training and test datasets. Among the 2301 bearing images, 2101 images were randomly selected and assigned to the training dataset and the remaining 210 images were assigned to the test dataset. The training dataset was used to build a classification model based on a deep learning algorithm, CNN (further details are provided in Section 3.3). The test dataset was used to evaluate the classification model's performance.

3.3. AI model

For the AI-assistance experiment, a high-performance deep learning model based on CNN was developed for bearing fault detection. The input image was resized to $150 \times 150 \times 3$ pixels (height $\times$ width $\times$ RGB channel). Our CNN model had four convolutional layers and two fully connected layers. Batch normalization and dropout layers were adopted to improve model performance and a standard softmax layer was used to classify whether each input image was normal or abnormal. For training, 2101 training images were used where data augmentations (rotation by 90°, 180°, and 270° and left-right flip) were applied to the training dataset to enhance the generalization performance. To handle the imbalance between the number of normal images and the number of abnormal images, data augmentation was applied more extensively to the abnormal samples. The AI model showed high performance when evaluated over the test dataset; the specificity was 0.91 and the sensitivity was 0.89.

4. Experimental settings

We designed an experiment using the AI model to measure the effects of AI-assistance on the human inspector. The experiment was designed to have two main factors. The first factor was the existence of AI-assistance in the inspection task—*unassisted or assisted*. The second factor was human inspectors' level of task proficiency—*generalist or specialist*. Throughout the experimental procedure, the two groups of inspectors were asked to perform bearing fault identification tasks with and without AI-assistance. The effects of AI-assistance were analyzed in terms of both performance and perception. For performance, we focused on specificity, sensitivity, and time efficiency. For perception, we focused on the inspector's perceived cognitive load, reuse intention, and usefulness (details of the metrics will be further explained in subsection 4.3). The data measured by these metrics were analyzed to investigate the overall effect of the AI-assisted approach.

4.1. Participants

A total of 54 engineers participated in this study. All participants were field engineers in the wind energy industry, but not all participants had specialty areas directly relevant to bearing fault inspection. Depending on their specialty areas, they were divided into generalist and specialist groups, as shown in Table 1. The generalist group consisted of 34 engineers and technicians with a basic understanding of machinery and bearing fault symptoms. However, they had limited experience in the bearing fault diagnosis area because they were mainly engaged in the O&M of wind farms and typically performed various maintenance tasks of wind turbine equipment. The specialist group consisted of 20 engineers with more than three years of experience in bearing fault diagnosis. They were either inspection engineers or engineering consultants, and their routine duties included bearing fault diagnosis.

4.2. Procedure and inspection task

The experiment procedure of this study is shown in Fig. 2**.** For the experiment, half of the participants in each generalist and specialist group were randomly selected and assigned to teams 1 and 2. In the instruction step, all participants were first informed about their right to withdraw from the experiment at any time without any consequences. Then, they were explained about the procedure and the inspection task of classifying endoscopic images into normal and abnormal. In addition, the inspectors were familiarized with the two different user interfaces for human-only setting (Fig. 3-a) and AI-assistance setting (Fig. 3-b) before starting the experiment.

The participants were asked to perform the inspection task over two rounds throughout the procedure. For the inspection task, a total of 34 problems were developed based on 34 images that were randomly selected from the test dataset. The problems were evenly split into two different problem sets—set A and B—and used in the two rounds of inspection tasks across the different teams. This experimental design was for canceling out the selection bias among the problem sets and participant teams. The ratio of normal to abnormal in set A and B was 9 to 8 and 8 to 9, respectively.

In the first round, the human-only setting was applied so that participants performed the inspection task without AI-assistance (Fig. 3-a). In the second round, otherwise, an AI-assisted setting

Table 1
Participant demography.

Demography		Generalist (n = 34)	Specialist (n = 20)
Gender − Number of males (% of males)		34 (100%)	19 (95%)
Average work experience in related areas		6.38 years	8.15 years
Major	Mechanical engineering	33	17
	Electrical, electronic engineering	14	11
	Others	7	6
Specialty area	Inspection and failure analysis		18
	Bearing specialist		2
	Operation and maintenance	31	
	Gearbox design and analysis	3	
Nationality	Korea	25	5
	US		7
	Japan	5	
	UK		4
	India		3
	Others	4	1

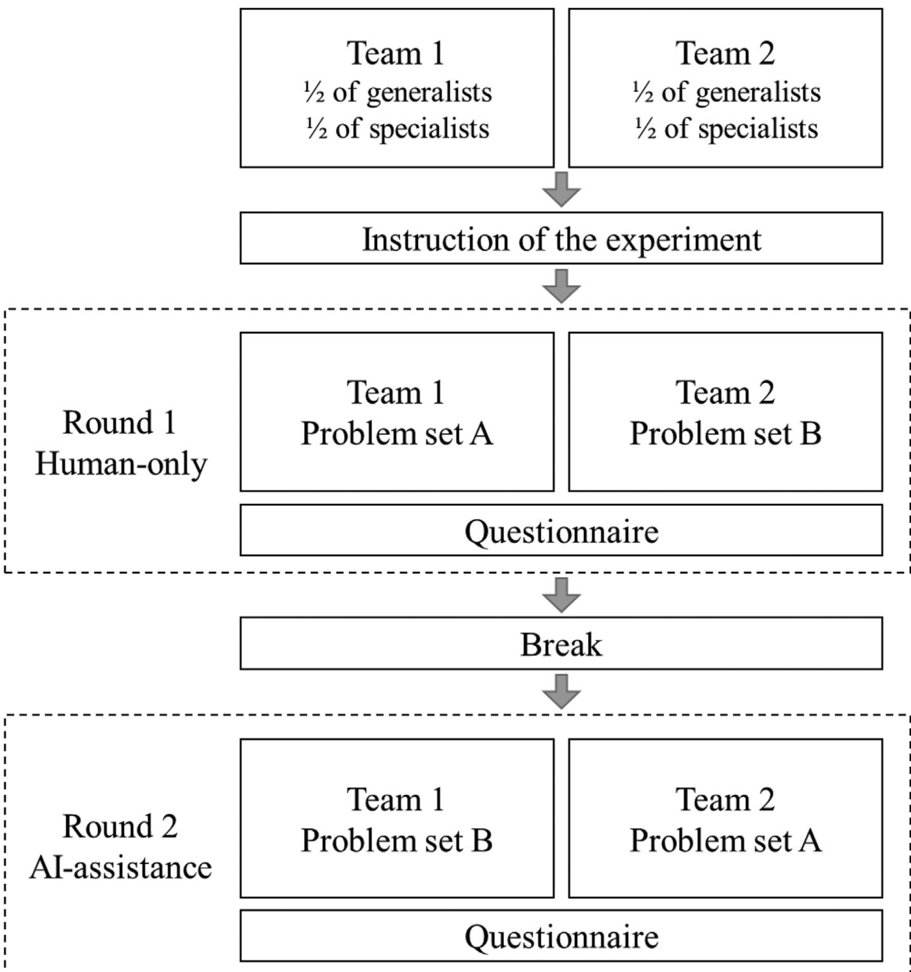
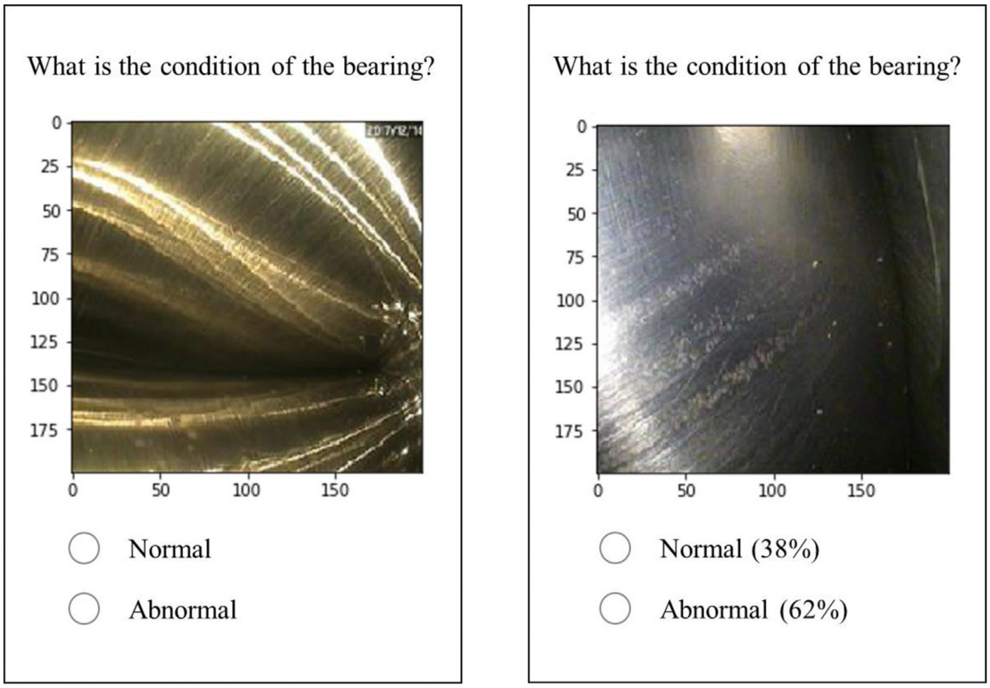


Fig. 2. Experimental procedure.

was applied, so the participants were provided with the probability of each class estimated by the AI model (Fig. 3-b). In this round, the participant was also informed about the AI model's performance, so that they could establish their trust level on the provided probability. Between the two rounds of inspection tasks, a break time was provided to prevent side effects due to fatigue or stress.

At the end of each round, the participant was asked to respond to a questionnaire designed to assess the cognitive load of the task. Particularly, at the end of the second round, surveys for assessing participants' intention to reuse and usefulness of the AI-assistance were conducted. In this survey, the participants could voluntarily leave their own feedback regarding AI-assistance as well as the overall experiment.



(a) Round 1: Human-only (b) Round 2: AI-assistance

Fig. 3. User interface for the inspection task.

4.3. Performance metrics

The participants' performances on the inspection task were assessed using three metrics: specificity, sensitivity, and time efficiency (Table 2). Specificity and sensitivity are the most commonly used statistical measures for diagnostic systems [41]. Specificity measures the proportion of actual normality (negatives) that are correctly identified as normal. It is also called the true negative rate. High specificity means that actual normality is rarely mistaken as abnormal, so false positives are few. Sensitivity measures the proportion of actual abnormality (positives) that is correctly identified as abnormal. It is also called the true positive rate or recall. High sensitivity means that actual abnormality is rarely overlooked, and hence, false negatives are few. The terms "positive" and "negative" refer to the presence or absence of an abnormality condition, which is the condition of interest. Apart from the two metrics for diagnostic systems, time efficiency of the inspector was also analyzed because time is an essential constraint for industrial tasks [16] because it is directly related to the cost. Time efficiency was defined as the average number of images that were analyzed per minute.

4.4. Perception metrics

The participants' perceptions of the inspection task and AI-assistance were assessed in terms of cognitive load, intention to reuse, and usefulness. Cognitive load refers to the effort used in the working memory to perform the task [42]. If decision support through AI-assistance reduces the user's effort to complete the fault diagnosis task, the cognitive load metric is expected to be reduced. Intention to reuse is the user's behavioral intention to use the system again in a similar situation in the future [43]. Usefulness can be defined as the degree to which a person believes that using the system would enhance his or her job performance [44]. A good fit between task and technology positively influences the perceived intention to reuse and usefulness. All perceptions were measured

Table 2
Performance metrics.

Metric	Equation
Specificity	$\frac{\text{True Negative}}{\text{True Negative} + \text{False Positive}}$
Sensitivity	$\frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$
Time efficiency	$\frac{\text{Number of images in the task}}{\text{Time to complete the task (minutes)}}$

using questionnaires adapted from Refs. [44–46] by considering the context of this study. Participants responded to the questionnaires after each round without knowing the scores of the inspection task. All items in the questionnaires were assessed using a 5-point Likert scale (strongly disagree = 1; strongly agree = 5), and example items are provided in Table 3 with their reliabilities (Cronbach's alpha).

5. Analysis results

To investigate the effects of an AI-assisted approach on the inspector's performance and perception, a descriptive analysis was first conducted on all the measurements of the metrics. Then, a two-way mixed ANOVA was conducted to investigate the impact of the two main factors, i.e., the existence of AI-assistance and task proficiency of human inspectors, on the three performance measurements and cognitive load. When significant effects were found, post-hoc analysis was subsequently conducted using Bonferroni's multiple comparison tests to investigate the two factors' effects in detail. For measurement of intention to reuse and usefulness, which were measured once at the end of the second round, a *t*-test was performed to investigate the significant differences between the two inspector groups. Missing values were excluded from the analysis.

Table 3
Perception metrics.

Metric	Number of items	Example	Cronbach's alpha
Cognitive load	4	This task required much mental and perceptual activity to complete	0.75–0.76
Intention to reuse	3	If I had a software that provides AI-assistance, I will perform bearing inspection by using it	0.77
Usefulness	4	Using AI-assistance would enhance the effectiveness of my job	0.95

5.1. Effects on performance

5.1.1. Specificity

The results of the analysis regarding the inspectors' specificity are summarized in Table 4. In both unassisted and assisted settings, the specialist group showed a higher specificity than the generalist group. The specificity for both groups was improved by AI-assistance; in particular, this improvement was greater in the generalist group than the specialist group (Table 4–a). Generalists' specificity was improved by 24.6% (from M = 0.65 to M = 0.81), whereas the specialists' specificity was improved by 4.7% (from M = 0.85 to M = 0.89). Two-way mixed ANOVA (Table 4–b) yielded a significant effect both for the level of task proficiency ($F(1,52) = 12.94$, $p < 0.001$) and for the existence of AI-assistance ($F(1,52) = 9.30$, $p < 0.01$); however, the interaction between the two factors was not significant ($F(1,52) = 3.40$, $p < 0.01$). In the results of post-hoc analysis (Table 4–c), specificity improvement of the generalist group due to AI-assistance was found to be significant (difference = 0.15, $p < 0.01$). Under the human-only setting, a significant difference was observed in the specificity between the two groups (difference = 0.20, $p < 0.01$), but this difference was not significant when AI-assistance was provided (difference = 0.09, $p = 0.55$). As expected, assisted specialists significantly outperformed the unassisted generalists (difference = 0.24, $p < 0.001$).

5.1.2. Sensitivity

The analysis results regarding inspectors' sensitivity are summarized in Table 5. In sensitivity measurements, the differences between the two groups were very small, and the improvement in sensitivity because of the AI-assistance was also very slight (Table 5–a). ANOVA yielded no significant main effects both for the level of task proficiency and the existence of AI-assistance (Table 5–b). There was also no significant interaction between the two factors.

5.1.3. Time efficiency

The results of analysis regarding inspectors' time efficiency are summarized in Table 6. In the absence of AI-assistance, the specialist group showed higher time efficiency than the generalist group, but this difference was reversed by AI-assistance provision (Table 6–a); the generalist group's time efficiency was improved by 25.3% (from M = 5.88 to M = 7.38), whereas the generalist group's efficiency was improved by 6.4% (from M = 6.28 to M = 6.68). In time efficiency, a significant main effect was only found in the existence of AI-assistance factor ($F(1, 47) = 5.52$, $p < 0.05$, Table 6–b). In the following post-hoc analysis, no significant difference was found.

5.2. Effects on perception

5.2.1. Cognitive load

The results of the analysis of the inspectors' cognitive load are summarized in Table 7. The descriptive analysis shows that the overall cognitive load of specialists was lower than that of generalists (Table 7–a). This was not surprising because specialists are more familiar with the task and use less effort than generalists. Interestingly, however, changes in cognitive load because of AI-assistance provision occurred in opposite directions in the two groups; the generalist group's cognitive load was reduced by 7.0% (from M = 5.88 to M = 7.38), whereas the specialists' cognitive load was improved by 2.4% (from 3.29 to 3.37). In ANOVA, a significant main effect was found only in the level of task proficiency ($F(1,46) = 4.34$, $p < 0.01$, Table 7–b). In the post-hoc analysis, all the differences were not statistically significant, but there was a

Table 4
Analysis results of specificity.

(a) Descriptive analysis

Existence of AI-assistance \ Level of task proficiency	Generalist M (SD)	Specialist M (SD)
Human-only	0.65 (0.24)	0.85 (0.13)
AI-assistance	0.81 (0.16)	0.89 (0.12)

(b) ANOVA

Effect	F
Level of task proficiency	12.94***
Existence of AI-assistance	9.30**
Level of task proficiency × Existence of AI-assistance	3.28

(c) Post-hoc analysis

Pairwise comparison	Difference
Generalist: human-only – AI-assistance	0.15**
Specialist: human-only – AI-assistance	0.04
Human-only: generalist – specialist	0.20**
AI-assistance: generalist – specialist	0.09
Generalist, human-only – specialist, AI-assistance	0.24***
Generalist, AI-assistance – specialist, human-only	0.04

Note. **$p < 0.01$, ***$p < 0.001$ (2-tailed).

Table 5
Analysis results of sensitivity.

(a) Descriptive analysis

Existence of AI-assistance \ Level of task proficiency	Generalist M (SD)	Specialist M (SD)
Human-only	0.93 (0.13)	0.93 (0.11)
AI-assistance	0.95 (0.89)	0.96 (0.67)

(b) ANOVA

Effect	F
Level of task proficiency	0.05
Existence of AI-assistance	1.55
Level of task proficiency × Existence of AI-assistance	0.19

Table 6
Analysis results of time efficiency.

(a) Descriptive analysis

Existence of AI-assistance \ Level of task proficiency	Generalist M (SD)	Specialist M (SD)
Human-only	5.88 (2.36)	6.28 (2.87)
AI-assistance	7.37 (3.09)	6.68 (3.16)

(b) ANOVA

Effect	F
Level of task proficiency	0.04
Existence of AI-assistance	5.52*
Level of task proficiency × Existence of AI-assistance	1.82

(c) Post-hoc analysis

Pairwise comparison	Difference
Generalist: human-only – AI-assistance	1.49
Specialist: human-only – AI-assistance	0.40
Human-only: generalist – specialist	0.39
AI-assistance: generalist – specialist	-0.69
Generalist, human-only – specialist, AI-assistance	0.80
Generalist, AI-assistance – specialist, human-only	-1.09

Note. *p < 0.05 (2-tailed).

marginal difference between generalists and specialists under the human-only setting (difference = −0.68, p = 0.08).

5.2.2. Intention to reuse and usefulness

Table 8 presents the results of the analysis regarding inspectors' intention to reuse AI-assistance and its' usefulness. Both groups showed positive perceptions regarding both metrics. In terms of intention to reuse, a positive willingness was found in the generalist group ($M = 3.95$, $SD = 0.89$) and the specialist group ($M = 3.93$, $SD = 0.64$). In terms of usefulness, the specialist group showed a slightly higher response ($M = 3.97$, $SD = 0.46$) than the generalist group ($M = 3.79$, $SD = 0.89$). The differences between the two groups regarding the metrics were not significant.

6. Discussion

The effects of AI-assistance on human inspectors were assessed for performing bearing fault diagnostics. The results show the potential for a positive influence, and a few important related topics are discussed in this section.

6.1. Advantages of AI-assistance

Our experiment was limited to the task of bearing diagnostics of wind turbines, and caution is required for interpreting the implications to the general image-based diagnostics. With this in mind, we addressed the advantages that might be widely applicable to AI-assisted diagnostics.

6.1.1. Overall effects of AI-assistance

Regarding the performance metrics, AI-assistance improved all three metrics for both generalists and specialists. While only some of the performance improvements were statistically significant, it is still promising that the effects were positive for specialists and even for sensitivity. The specialists were technical experts with years of experience in bearing fault diagnostics. For sensitivity, AI alone performed worse than the generalists, but it was able to enhance the human's performance for both generalists and specialists.

Regarding the perception metrics, the effects of AI-assistance on cognitive load were small (7.0% reduction for generalists and 2.4% increase for specialists) and statistically insignificant. Therefore, the experimental results indicate that AI-assistance can be adopted without a significant downside on cognitive load. In addition, the

Table 7
Analysis results of cognitive load.

(a) Descriptive analysis

Existence of AI-assistance \ Level of task proficiency	Generalist M (SD)	Specialist M (SD)
Human-only	3.97 (0.90)	3.29 (0.85)
AI-assistance	3.69 (1.00)	3.37 (0.81)

(b) ANOVA

Effect	F
Level of task proficiency	4.34*
Existence of AI-assistance	0.67
Level of task proficiency × Existence of AI-assistance	2.19

(c) Post-hoc analysis

Pairwise comparison	Difference
Generalist: human-only – AI-assistance	-0.28
Specialist: human-only – AI-assistance	0.08
Human-only: generalist – specialist	-0.68
AI-assistance: generalist – specialist	-0.32
Generalist, human-only – specialist, AI-assistance	-0.60
Generalist, AI-assistance – specialist, human-only	-0.40

Note. *$p < 0.05$ (2-tailed).

Table 8
Analysis results of intention to reuse and usefulness.

Existence of AI-assistance	Level of task proficiency		t
	Generalist M (SD)	Specialist M (SD)	
Intention to reuse	3.81 (0.89)	3.97 (0.46)	0.71
Usefulness	3.95 (0.88)	3.93 (0.64)	−0.11

metrics for intention to reuse and usefulness were positive for both groups (3.79–3.97 out of 5). According to the technology acceptance model [46,47], intention to reuse and usefulness are known as the major factors driving an individual's willingness to accept a new technology tool in an organizational setting, and it is encouraging that all of the three metrics were neutral or positive.

6.1.2. *Impacts on generalists*

The influence of AI-assistance on specialists was limited, but its impact on generalists was relatively large. Consequently, the difference in performance between the generalist and specialist groups became statistically insignificant for all three performance metrics under AI-assistance. In particular, generalists even outperformed specialists in terms of time efficiency. Similar results have been reported in other disciplines. For instance, a similar improvement was observed in the field of clinical radiology where the performance of the generalists was improved with Computer-aided Detection [21].

If AI-assistance can enable generalists to perform as well as specialists, the implication can be significant. Typically, it takes years for a generalist to become a specialist, and any task that requires a specialist is constrained by the cost and availability of the specialist. With AI-assistance, the constraints can be removed or mitigated without sacrificing the quality of diagnostics, and skilled specialists can be released from routine diagnostic works and the resources can be invested in new areas [48]. As an example, a specialist's time might be better utilized by working on image labeling (creating ground truth labels for AI model training) or by supporting a team of generalists. Similar results were obtained for tasks in other industries, including legal services [49] and medical diagnostics [50].

6.1.3. *Time efficiency*

High sensitivity (i.e. recall) and low specificity are common characteristics of fault diagnostics because the penalty for a false negative is far larger than the penalty for a false positive [51]. Because all inspectors are concerned about missing a fault, even generalists show a high sensitivity performance, as shown in Table 6-(a). In the absence of AI-assistance, the generalists achieved as high a sensitivity performance as that of the specialists (0.93 vs. 0.93) at the cost of low specificity performance (0.65 vs. 0.85). The AI model's detection threshold, however, was chosen neutrally and therefore the AI model achieved a sensitivity of 0.89 and a specificity of 0.91. The high specificity performance compared to the human inspectors is an important factor to be considered, and its influence on time efficiency might be important.

When AI-assistance was provided, the time efficiency was improved for both generalists and specialists where the improvement was larger for generalists than for specialists (25.3% vs. 6.4%). The improvement in time efficiency indicates that the inspectors were able to make faster decisions with AI-assistance without sacrificing sensitivity and specificity. Therefore, a reasonable conjecture is that AI-assistance was helpful in making quick decisions for obvious cases while it was not harmful for the borderline cases. The effect might have been even stronger for obviously negative cases because of the AI model's superior specificity performance. This explanation was also supported by a participant's feedback. The participant mentioned that "The AI assistance helped me weed out obviously normal and abnormal cases quickly". Therefore, time saving was more likely from obvious cases than from subtle borderline cases and probably more from relieving inspectors from the worry of obtaining false negatives.

6.2. Toward frequent and periodic predictive maintenance

Many fields are actively adopting image-based predictive maintenance. For instance, structure monitoring based on drone camera inspection is another growing field. Similar to renewable energy and medical fields, the goal is to perform periodic inspections such that a possible problem in buildings, bridges, construction sites, or telecommunication sites can be detected early. In all such fields, the recent revolution in image classification through deep learning algorithms is leading the trend of pursuing frequent and periodic predictive maintenance. A few issues related to the role of AI-assistance in predictive maintenance are discussed below.

One of the basic problems that is encountered with predictive maintenance is the increased workload. Inspection needs to become a periodic and routine process, where image data collection is the first part of work and reading and diagnosis of the images is the second part of the work. In the early phase of adopting predictive maintenance, the inspectors often need to use image devices to collect the image data manually. While this part cannot benefit from AI-assistance, it can be automated in some of the fields if the manufacturers start to include built-in image-based monitoring systems as equipment. The reading and diagnosis part, however, is unlikely to be fully automated in the near future because AI-only is limited in the sense that it can recognize only the faults that it has learned well from the training data. When there are new types of problems that can occur, humans are required to be kept in the loop and detect possibly disastrous cases in their early stages.

As long as the human's role cannot be eliminated, AI-assistance's utility might only be in the form on time efficiency. As in our experiment, even a direct and simple integration of AI-assistance might be sufficient to improve the human inspector's time efficiency. Additionally, more sophisticated approaches might helpful. For instance, a careful trade-off between the AI model's specificity and sensitivity might result in an additional improvement depending on the characteristics of human decisions. In our case, human-only was biased toward a higher sensitivity at the cost of lower specificity. Another possible approach is to train multiple AI models with different goals. A version of AI models might be designed and trained with the goal of confidently detecting the obvious cases while simply providing 'unknown' as the output for the difficult cases. Another version of the AI model might be designed and trained with the goal of confidently identifying only a particular fault type that commonly occurs. If high quality datasets are available, many other versions can be learned, resulting in a good chance of additional improvement. In this case, however, the effects on perception should be carefully studied for choosing the ideal subset of AI models. In addition, some of the extreme versions might be used in AI-only mode where their utility is for automatically filtering out the trivial cases.

Apart from time efficiency, another important role of AI-assistance is supporting generalists such that they can perform as well as specialists. This issue has been already addressed, but it is worthwhile to point out its explicit effect on predictive maintenance. As visual inspection with imaging devices is being widely adopted as a routine maintenance procedure, the demand for skilled inspectors is sharply increasing. However, years are required for a generalist to become a specialist, and the availability of qualified inspectors is a bottleneck for quickly adopting predictive maintenance. This was one of the reasons why predictive maintenance was only slowly adopted in the past decades. However, with the recent deep learning solutions that can provide close-to-human performance on image classification tasks, the problem can be greatly alleviated through AI-assistance. If a minimum number of specialists is available and the collected dataset results in a high-performance AI model for a given diagnostics task, there is a good chance that AI-assistance can be used for accelerated deployment of predictive maintenance.

6.3. Limitations of this work

6.3.1. Cognitive load

There are a few things that can be addressed regarding cognitive load. For real image-based diagnostics, the inspectors are required to process a large number of images over many hours. Because mental fatigue can lead to frequent mistakes and degraded efficiency, it is important to reduce the cognitive load or at least prevent it from increasing. In our experiment, each of set A and set B contained only 17 images. For a real work environment with a much larger number of images, it is unclear if the relative influence of AI-assistance on cognitive load will be positive or negative. It might become positive if the mental fatigue of repeated image screening is reduced. In contrast, it might become negative because of information overload [22] or the initial lack of trust on AI-assistance [52]. A further study is required to address this issue.

6.3.2. Human performance, AI performance, and their relative relationship

In our experiment, the AI model's performance was comparable to a human's performance. For a given diagnostic task, however, the situation can be different and there are a few important factors to be considered. The first that needs to be considered is human performance. If the diagnostic task is too difficult even for specialists, there might be no incentive to pursue predictive maintenance. Note that AI models of today are typically trained with human generated labels, and therefore, AI models are unlikely to perform well, either. The second factor that needs to be considered is AI performance. If the diagnostics task is simple and sufficiently easy that an AI model can diagnose all types of faults with almost perfect accuracy, then using AI-only would be sufficient for predictive maintenance. For such a simple and easy task, however, the chances are that the situation is already serious and obvious that even simple rule-based or other hand-crafted algorithms might perform as well as any sophisticated AI models. In this case, automation of diagnostics can be easy without using AI models or AI-assistance.

When a diagnostic task is workable enough for specialists and yet difficult enough for generalists, AI-assistance can be a viable solution. The usefulness of AI-assistance, however, is not an easy one to determine. Human performance depends on many elements such as training and development of domain knowledge. AI performance also depends on many elements such as image quality and dataset enhancement. Furthermore, each task can have many details and can be complex, so that it is difficult to clearly define. When all these factors are considered together, a careful study might be necessary on a task-by-task basis to draw a reliable conclusion regarding the usefulness of AI-assistance. In this regard, our study can only serve as a case study, and the results should be carefully interpreted when applying them to other tasks. When ultimate cost-saving through predictive maintenance is very large as in the renewable energy sector, repeating a study like ours can be a very prudent and timely topic to work on.

7. Conclusion

Predictive maintenance is a crucial component for the successful operation of renewable energy systems, and adopting the latest AI technology is expected to be inevitable for scaling predictive maintenance to sufficiently frequent and periodic inspections. In this work, we studied a representative case of bearing fault detection where endoscope images are collected for the inspection. We considered the limitations of human-only and AI-only approaches and evaluated the benefits of AI-assistance. The results show that all factors, specificity, sensitivity, and time efficiency, can be improved with AI-assistance, for both the generalist group and specialist group. In fact, AI-assistance was helpful for inspectors even when the performance of AI-only was worse than human performance. The perception effects were also investigated, and positive results were obtained. Possible future works include experiments under a regular operation mode and extended experiments in the renewable energy fields outside of wind farm operation.

Author statement

Won Shin: Conceptualization, Methodology, Software, Validation, Investigation, Resources, Writing – original draft, Writing – review & editing, Project administration, Jeongyun Han: Methodology, Validation, Formal analysis, Writing – original draft, Writing – review & editing, Visualization, Wonjong Rhee: Conceptualization, Methodology, Resources, Writing – original draft, Writing – review & editing, Supervision

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work was supported by the National Research Foundation of Korea, South Korea; (NRF) grant funded by the Korea government (MSIT) (Grant No. NRF-2017R1E1A1A03070560 and Grant No. NRF-2020R1A2C2007139).

References

[1] Irena. The power to change: Solar and wind cost reduction potential to 2025. 2016.
[2] Bosch J, Staffell I, Hawkes AD. Global levelised cost of electricity from offshore wind. Energy 2019;189:116357.
[3] ReliaPlant. Predictive Maintenance Explained, https://www.reliableplant.com/Read/12495/preventive-predictive-maintenance.
[4] Life Cycle Engineering. Preventive and Predictive Maintenance, https://www.lce.com/pdfs/The-PMPdM-Program-124.pdf.
[5] Addepalli S, Roy R, Axinte D, Mehnen J. In-situ' inspection technologies: trends in degradation assessment and associated technologies. Procedia CIRP 2017;59:35–40.
[6] GE Energy Power Conversion. External Report on maintenance Schedule Checklist for GEPC - RM Commissioning & services generators for wind turbine application. 2017.
[7] Hur S, Recalde-Camacho S, Leithead W. Detection and compensation of anomalous conditions in a wind turbine. Energy 2018;124:74–86.
[8] Zhou D, Zhang H, Weng S. A novel prognostic model of performance degradation trend for power machinery maintenance. Energy 2014;78:740–6.
[9] Akram M, Li G, Jin Y, Chen X, Zhu C, Zhao X, Khaliq A, Faheem M, Ahmad A. CNN based automatic detection of photovoltaic cell defects in electroluminescence images. Energy 2019;189:116319.
[10] Ren S, He K, Girshick R, Sun J, Faster R-CNN. Towards real-time object detection with region proposal networks. IEEE Trans Pattern Anal Mach Intell 2017;39(6):1137–49.
[11] Cha YJ, Choi W, Suh G, Mahmoudkhani S, Büyüköztürk O. Autonomous structural visual inspection using region-based deep learning for detecting multiple damage types. Comput Civ Infrastruct Eng 2018;33(9):731–47.
[12] Bughin J, Hazan E. Artificial intelligence – the next digital frontier. 2017.
[13] Brynjolfsson E, Mitchell T. What can machine learning do? Workforce implications: profound change is coming, but roles for humans remain. Science 2017;358:6370. American Association for the Advancement of Science. 2017;1530–1534.
[14] ISO. ISO 15243:2004(en): rolling bearings – damage and failures – terms, characteristics and causes. 2004.
[15] James WH, Daugherty PR. Collaborative intelligence: humans and AI are joining forces. Harv Bus Rev 2018;2018(July-August).
[16] Megaw ED. Factors affecting visual inspection accuracy. Appl Ergon 1979;10(1):27–32.
[17] Reiner BI, Krupinski E. The insidious problem of fatigue in medical imaging practice. J Digit Imag 2012;25(1):3–6.
[18] Mostefa Khelil C, Amrouche B, Benyoucef A, Kara K, Chouder A. New intelligent fault diagnosis (IFD) approach for grid-connected photovoltaic systems. Energy 2020;211:118591.
[19] Brkovic A, Gajic D, Gligorijevic J, Savic-Gajic I, Georgieva O, Gennaro S. Early fault detection and diagnosis in bearings for more efficient operation of rotating machinery. Energy 2017;136:63–71.
[20] Bode G, Thul S, Baranski M, Müller D. Real-world application of machine-learning-based fault detection trained with experimental data. Energy 2020;198:117323.
[21] Fisichella VA, Jäderling F, Horvath S, Stotzer PO, Kilander A, Båth M, Hellström M. Computer-aided detection (CAD) as a second reader using perspective filet view at CT colonography: effect on performance of inexperienced readers. Clin Radiol 2009;64(10):972–82.
[22] Mălăescu I, Sutton SG. The effects of decision aid structural restrictiveness on cognitive load, perceived usefulness, and reuse intentions. Int J Account Inf Syst 2015;(17):16–36.
[23] Dale L, Ronald L. Task-technology fit and individual performance. MIS Q 1995;19(2):213–36.
[24] Smith-Bindman R, Chu P, Miglioretti D, Quale C, Rosenberg R, Cutter G, Geller B, Bacchetti P, Sickles EA, Kerlikowske K. Physician predictors of mammographic accuracy. J Natl Cancer Inst 2005;97(5):358–67.
[25] Gulshan V, Peng L, Coram M, Stumpe MC, Wu D, Narayanaswamy A, Venugopalan S, Widner K, Madams T, Cuadros J, Kim R, Raman R, Nelson P, Mega JL, Webster DR. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. J Am Med Assoc 2016;316(22):2402.
[26] Wang M, Tong W, Liu S. fault detection for power line based on convolution neural network. In: Proceedings of the 2017 international conference on deep learning technologies - ICDLT '17; 2017. p. 95–101.
[27] Doi K. Computer-aided diagnosis in medical imaging: historical review, current status and future potential. Comput Med Imag Graph 2007;31(4–5):198–211.
[28] Doi K. Current status and future potential of computer-aided diagnosis in medical imaging. Br J Radiol 2005;78(1):s3–19.
[29] van Ginneken B, Schaefer-Prokop CM, Prokop M. Computer-aided diagnosis: how to move from the laboratory to the clinic. Radiology 2011;261(3):719–32.
[30] Brown MS, McNitt-Gray F, Pais R, Shah K, Qing P, Da I, Aberle R, Goldin G. CAD in clinical trials: current role and architectural requirements. Comput Med Imag Graph 2007;31(4–5):332–7.
[31] Cooper JG, West RM, Clamp SE, Hassan TS. Does computer-aided clinical decision support improve the management of acute abdominal pain? A systematic review. Emerg Med J 2011;28(7):553–7.
[32] Freer TW, Ulissey MJ. Screening mammography with computer-aided detection: prospective study of 12,860 patients in a community breast center. Radiology 2001;220(3):781–6.
[33] Birdwell RL, Bandodkar P, Ikeda DM. Abbreviations: BI-RADS breast imaging reporting and data system CAD computer-aided detection PPV positive predictive value. Radiology 2005;236:451–7.
[34] Helvie MA, Hadjiiski L, Makariou E, Chan HP, Petrick N, Sahiner B, Lo SC, Freedman M, Adler D, Bailey J, Blane C, Hoff D, Hunt K, Joynt L, Klein K, Paramagul C, Patterson SK, Roubidoux MA. Sensitivity of noncommercial computer-aided detection system for mammographic breast cancer detection: pilot clinical trial. Radiology 2004;231(1):208–14.
[35] Ko JM, Nicholas MJ, Mendel JB, Slanetz PJ. Prospective assessment of computer-aided detection in interpretation of screening mammography. Am J Roentgenol 2006;187(6):1483–91.
[36] Halligan S, Altman D, Mallett S, Taylor S, Burling D, Roddie M, Honeyfield L, McQuillan J, Amin H, Dehmeshki J. Computed tomographic colonography: assessment of radiologist performance with and without computer-aided detection. Gastroenterology 2006;131(6):1690–9.
[37] Mani A, Napel S, Paik DS, Jeffrey Jr RB, Yee J, Olcott EW, Prokesch R, Davila M, Schraedley-Desmond P, Beaulieu CF. Computed tomography colonography: feasibility of computer-aided polyp detection in a first reader paradigm. J Comput Assist Tomogr 2004;28(3):318–26.
[38] Alberdi E, Povyakalo A, Strigini L, Ayton P. Effects of incorrect computer-aided detection (CAD) output on human decision-making in mammography. Computer Assisted Radiology and Surgery 1004; 11(8):909-918.
[39] Eom Sb. The development of decision support systems research: a bibliometrical approach. Edwin Mellen Press; 2007.
[40] Shibl R, Lawley M, Debuse J. Factors influencing decision support system acceptance. Decis Support Syst 2013;54(2):953–61.
[41] Gonçalves VM, Delamaro ME, Nunes FDLDS. A systematic review on the evaluation and characteristics of computer-aided diagnosis systems. Rev Bras Eng Biomédica 2014;30(4):355–83.
[42] Sweller J. Cognitive load during problem solving: effects on learning. Cognit Sci 1988;12(2):257–85.
[43] Venkatesh V. Determinants of perceived ease of use: integrating control, intrinsic motivation, and emotion into the technology acceptance model. Inf

Syst Res 2000;11(4):342–65.
[44] Venkatesh V, Smith RH, Morris MG, Davis GB, Davis FD, Walton SM. User acceptance of information technology: toward A unified view. User Accept IT MIS Q 2003;27(3):425–78.
[45] Hart SG, Staveland LE. Development of NASA-TLX (task load index): results of empirical and theoretical research. Adv Psychol 1998;52(C):139–83.
[46] Davis FD. Perceived usefulness, perceived ease of use, and user acceptance of information technology. MIS Q 1989;13(3):319.
[47] Ortega Egea JM, Román González MV. Explaining physicians' acceptance of EHCR systems: an extension of TAM with trust and risk factors. Comput Hum Behav 2011;27(1):319–32.
[48] Benke K, Benke G, Benke K, Benke G. Artificial intelligence and big data in public health. Int J Environ Res Publ Health 2018;15(12):2796.
[49] Reid M. A call to arms: why and how lawyers and law schools should embrace artificial intelligence, vol. 50. Univ Toledo Law Rev; 2018.
[50] Jha S, Topol EJ. Adapting to artificial intelligence: radiologists and pathologists as information specialists. J Am Med Assoc 2016;316(22):2353.
[51] Knop K, Olejarz E, Ulewicz R. Evaluating and improving the Effectiveness of visual inspection of Products from the Automotive industry. Cham: Springer; 2019. p. 231–43.
[52] Jianlong ZS, Chen F. Effects of uncertainty and cognitive load on user trust in predictive decision making. Int Fed Inf Process 2017;10516(October):23–39.