Evaluating Impact of AI on Cognitive Load of Technicians During Diagnosis Tasks in Maintenance

Hyunjong Shin, Vittaldas Prabhu

HAL Id: hal-02177865
https://inria.hal.science/hal-02177865

Submitted on 9 Jul 2019

Evaluating Impact of AI on Cognitive Load of Technicians during Diagnosis Tasks in Maintenance

Hyunjong Shin[1] and Vittaldas V. Prabhu[1]

[1] Pennsylvania State University, University Park PA 16802, USA
hvs5026@psu.edu

Abstract. Even today, many maintenance activities are still done manually because maintenance is one of the most difficult areas to be automated in manufacturing. Many technicians spend their time on non-technical activities such as retrieving instructions from manuals. If AI (Artificial Intelligence) can alleviate some of these tasks, the time to diagnosis and repair can be shortened. However there are limited works about the effects of using AI during maintenance activities on a technician's cognitive load. Therefore, as an initiative, we conducted a pilot experiment with 10 participants to analyze the effects of the AI-based support system on diagnosis tasks in the manufacturing. In the experiment, participants were divided into two groups: the group used an AI-based support system and the other group used a Fault Tree (FT) based support system; two groups' mean task completion time and task load of participants using NASA Task Load were measured. According to the experiment results, the group which used the AI-based support system to diagnose the model completed task 53% lesser time than the group which used the FT-based support system. In addition, participants who used the AI-based support system reported relatively lower task loads compared to participants who used the FT-based support system. This experiment results imply that maintenance time and a variability can be reduced if an AI-based support system supports maintenance technicians

Keywords: Smart Maintenance, Artificial Intelligence, Machine Learning, Naïve Bayesian, Diagnosis, Task Load, Cognitive Load, HAII

1 Introduction

Although maintenance activities are very critical in the manufacturing industry, only few maintenance activities are fully automated yet because it is one of the last areas to be automated in the manufacturing [1, 2]. Recent study also reports that over 30% of total workforce contributes to maintenance activities [3]. Maintenance activities are often composed of technical activities and non-technical activities. Retrieving instructions or information from manuals, for instance, take up about 45% of maintenance technicians' time [4]. Therefore, if a technology such as an AI can alleviate some of technicians' task by supporting their activities, the diagnosis and repair time will be shortened. However, AI must be cautiously implemented to the maintenance process

because there was a case that an AI was meant to improve operators' performance but it, instead, acted as a barrier and created even more challenges [5, 6].

Since 1996, as AI has become more popular, the number of annually published AI papers has soared in the field of computer science; the annual investment in AI startups by venture capitals has increased six fold since 2000 [7]; more and more people are paying attention to the potential benefits of AI.
In the field of the manufacturing numerous AI related papers can be found. In the manufacturing, AI is often used to detect product quality problems [8]. For example, Nguyen et al. and Yang et al used an AI to detect defective wafers in the semiconductor industry [9, 10]. Similarly, Liu & Jin used an AI to detect defective tail lights in the automobile industry [11]. Outside of detecting product quality problems, research has also investigated different applications of AI. Huang et al. used AI to diagnose vehicle fault. Hong et al. used AI to detect faults in the semiconductor manufacturing equipment [12]. Similarly Zhang et al used AI to identify degradation machines and tools [13].

The usage of an AI is also studied in the field of the human factors. For example, Overmeyer et al, studied the cognitive load of the operator who commands autonomous vehicles through an AI agent [14]. Similarly Strayer et al. studied the cognitive load of drivers who used an intelligent personal assistant [5].

Therefore, to explore this issue, we conducted a controlled pilot experiment to investigate the effect of AI-based support system on diagnosis task in the maintenance process.

The rest of this paper is structured as follows. In Section 2, we explain the experiment that we conducted to evaluate the effect of the AI on the diagnosis task. Next, in Section 3, we present the results of the experiment. Lastly, in Section 4, we state discussion and conclusions of this experiment.

2 Experimental Design and Setup

A proximity sensor is widely used to detect the presence of an object in many automated machines. However, proximity sensors frequently fail in CNC (Computer Numerical Control) machines. In addition, even though a technician identifies that the cause of a machine failure is related to the proximity sensor, the maintenance activity is not as simple as replacing a proximity sensor. The technician must check conditions of all components such as cable, power, I/O board, and sensor itself in order to repair the machine. Therefore, in this experiment, the model operated by a proximity sensor is chosen to evaluate the effect of AI on diagnosis tasks in maintenance

2.1 Experimental Task

Proximity Sensor Model. Every component in the proximity sensor model represents some component in a real industry machine as shown Table 1. The sensor in the experiment model detects whether the door in front of the sensor is closed or not. When the sensor detects the door, it shuts down the power to turn off the light. On the other hand, when there is no object, and every component is in working order, the light bulb is illuminated (See Fig. 1).

In the experiment, 4 components of the proximity sensor model were purposely in bad condition: battery, switch, light bulb and signal cable to light bulb. Then the participants were divided into two groups. The first group, also known as the FT group, were asked to diagnose problems and fix the model according to a fault tree-based support system. The participants in the second group, also known as the AI group, were asked to diagnose problems and fix the model according to an AI-based support system.
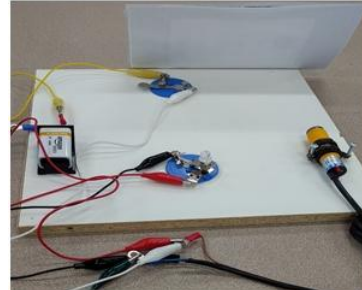
(a) When the Door is Closed – Light Off
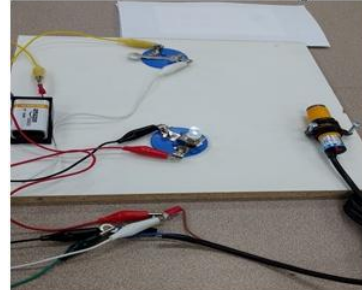
(b) When the Door is Opened – Light ON

Fig. 1. Proximity Sensor Model

Table 1. Proximity Sensor Model Setting

Real Machine Setting	Model Setting	Malfunction
Sensor	Sensor	No
Power	Battery	Yes
Power Switch	Switch	Yes
Sensor Connection	Signal Cable to Light Bulb	Yes
Cables	Cables	No
I/O Board	Light bulb	Yes

Support Systems. Two support systems were provided to support participants' diagnosis tasks. The FT, which is a common practice to repair the machines in many small and medium enterprises, based support system helped participants diagnose the locations of problems by deductive failure analysis method. On the other hand, the AI-based support system helped participants diagnose the locations of problems based on the pre calculated probability using the Naïve Bayesians classifier method. The Naïve Bayesians classifier method is used in this experiment because the method is known to require less input, work great in practice even if NB assumptions doesn't hold, and good for showing casual relationship [1]. (See Fig. 2).
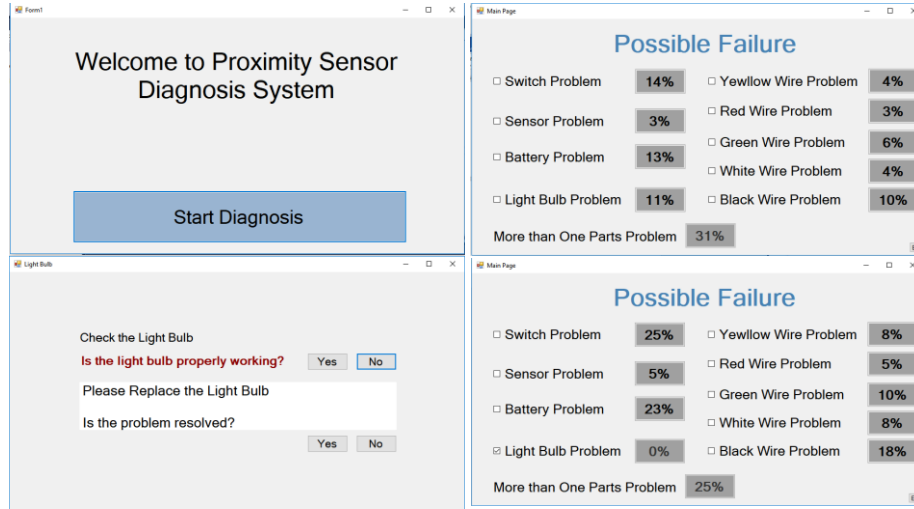
Fig. 2. AI-based Support System Interface

2.2 Participants

Five subjects were participated in each group. The total participants for this experiment were 10. The average age of participants in the FT and the AI group was 29.2 and 29.6 respectively. The youngest participant was 25 years old and the oldest was 32 years old. Of 10 participants, 80% of them were male. In each group, equal number of female participants was assigned to minimize gender effects. Twenty percent of the participants did not major in either engineering or science. All other participants' majors were either engineering or science.

2.3 Hypotheses

The following hypotheses were tested by using above experimental design and setup

- H1: The task completion time of the group which uses the AI-based support system will be shorter
- H2: The cognitive load of the group which uses the AI-based support system will be lower

2.4 Experiment Procedures

Experiment participants are going to be divided into two groups depending on their assigned group and participated in the experiment as stated in Table 2.

Table 2. Experiment Procedures

Step	Procedure	Description
1	Subject Arrival	The subject will be introduced to the testing facility, locations of exits and restrooms will be provided
2	Eligibility Verification	Subject eligibility will be checked prior to continuation. The requirements include: 18 years of age minimum and English speaking.
3	Consent	A summary of the study will be given to participants. Participants will be allowed to ask questions about the study. Verbal consent will be obtained prior to continuation to the following steps
4	Demographic Questionnaire	The subject will be asked to complete a short demographic questionnaire. Data collected with include: age, gender and major
5	Training Session	The subject will attend a training session. Methods of using a multimeter and support system will be introduced using PowerPoint slides. The subject can ask any question during the training session
6	Break	The participant will take a break. The participant may go to the restroom and drink water during this time
7	Experiment	The subject will diagnose the problem of a simple circuit and correct the circuit accordingly with A.I. support or without A.I. support depending on the group that the participant is assigned
8	Work Load Questionnaire	A subjective workload questionnaire (NASA TLX) will be administered after the task, which includes six rating scales in total to measure workload along six different dimensions (mental demand, physical demand, temporal demand, effort, frustration, and performance).
9	Debrief	To conclude, subjects will be thanked and provided with monetary compensation. Any concerns or questions will be addressed.

3   Experiment Results

Task Completion Time. Task completion time is the time that a participant takes to diagnose components and fix them accordingly. It is comprised of diagnosis time, such as using a diagnosis support system and a multimeter, and time to replace or fix components. By measuring the task completion time, the effect of the AI-based support system on diagnosis time can be identified.

The mean task completion time for the FT group was 372.4 seconds. The standard deviation of this group was 72.2. The mean task completion time of the AI group, on the other hand, was 176.4 seconds and its standard deviation was 21.1. The coefficient of variation for FT and AI group was 22% and 12% respectively. Based on the level of the coefficient of variation, the AI group had less variation in the task completion time. The mean task completion time

Fig. 3. Mean Task Completion Time Difference

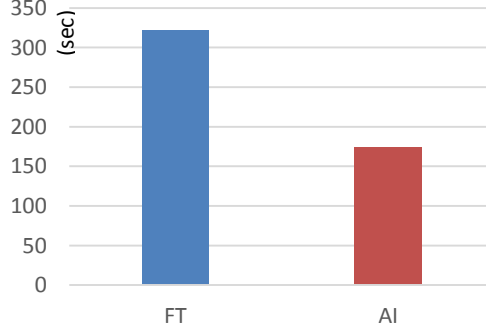difference between the two groups was 196 seconds (see Fig. 3). A two-sample t-test was used to test the difference between two groups. The calculated t-value was 5.83 and p-value was 0.004. Therefore at α equals to 0.05, we conclude that there was a mean task completion time difference between two groups.

NASA Task Load. The NASA Task Load index (TLX) is a subjective assessment tool that rates perceived workload of participants in order to assess a system. The TLX is divided into six subscales or categories: mental demand, physical demand, temporal demand, performance, effort and frustration. By measuring TLX, the effect of AI-based support system on operators' cognitive load and workload can be identified.
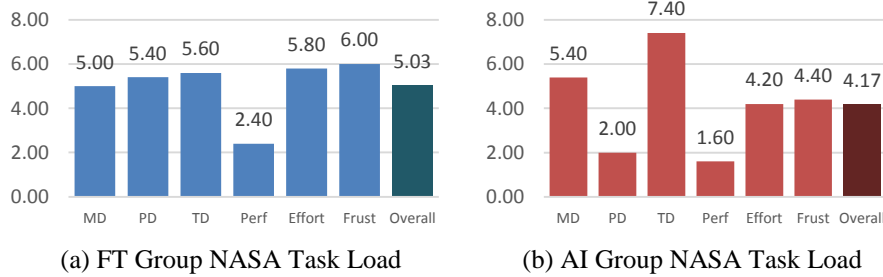

(a) FT Group NASA Task Load

(b) AI Group NASA Task Load

Fig. 4. The NASA TASK Load of Two Groups: FT Group and AI Group

The average overall task load of the FT group was 5.03. For the FT group, the frustration load turned out to be the highest load among six sub-scales. The other loads were around 5.00 or above except the performance load. The average performance load for this group was 2.40. Furthermore, in average, the mental load was less than the physical load as shown in Fig. 4. On the other hand, the average overall task load for AI group was about 4.2. Most of the loads' levels were similar to the overall task load level. However, the temporal demand load was 1.5 times more than the overall task load. The second highest load was the mental load which was 5.40. Among six task loads, the performance load was the lowest.
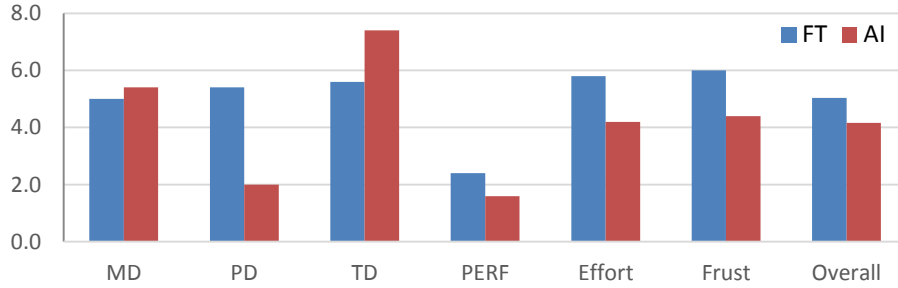
Fig. 5. NASA Task Load Differences

A two-sample t-test was used to identify the significance of task load differences between the two groups. The two sample t-test revealed that none of task loads' differences were statistically significant at alpha 0.05. Although visually there were some differences between the two groups, the differences were not large enough to have statistical meaning.

Performance Accuracy. The performance accuracy(PA) was defined as the number of parts replaced divided by the number of malfunctioning parts. If PA is greater than one, it implies that the participant replaced unnecessary parts while they were diagnosisng the model. Of 10 participants, none of them replaced unnecessary parts.

4 Discussion and Conclusions

The experiment that we conducted to investigate effects of the AI-based support system on maintenance reveals several interesting points.

First of all, the experiment result shows that the AI-based support system not only can reduce the diagnosis time but also can reduce the variation of the diagnosis time compared to the FT-based support system. This is possible that the AI-based support system allows participants to diagnose less numbers of parts compared to the FT-based support system if and only if the reliability of AI-based support system is high.

Secondly, AI-based support system must be carefully implemented to the maintenance process because the experiment result shows that the mean mental load of the AI group is higher than the mean mental of the FT group although the difference was not verified by the two-sample t-test

In sum, the experiment showed that the AI-based support system can reduce the diagnosis time and increase the mental load of technicians. However, the above points must be carefully interpreted since these results are based on our preliminary experiment in which only 10 subjects participated. Since a prerequisite of a two-sample t-test is a normality and the normality could not be assumed with 10 participants, the result of the two-sample t-test has to be interpreted cautiously. In addition, the power test requires at least 8 participants for each group. Therefore, there is a possibility that participants in this experiment do not truly represent the population. This pilot exper-

iment was conducted as part of an exploratory study. In the future study, several additional factors that might influence the cognitive load of a technician during the maintenance task will be included and investigated.

References

1. Langer, R., Li, J., Biller, S., Chang, Q., Huang, N., Xiao, G.: Simulation study of a bottleneck-based dispatching policy for a maintenance workforce. Int. J. Prod. Res. 48, 1745–1763 (2010)
2. Sheikhalishahi, M., Pintelon, L., Azadeh, A.: Human factors in maintenance: a review. J. Qual. Maint. Eng. 22, 218–237 (2016)
3. Kothamasu, R., Huang, S.H., Verduin, W.H.: System health monitoring and prognostics - A review of current paradigms and practices. Handb. Maint. Manag. Eng. 337–362 (2009)
4. Neumann, U., Majoros, A.: Cognitive , Performance , and Systems Issues for Augmented Reality Applications in Manufacturing and Maintenance. IEEE Virtual Real. (VR '98). 4–11 (1998)
5. Wiener, E.L.: Human factors of advanced technology (glass cockpit) transport aircraft. (Nasa-Cr-177528). 222 (1989)
6. Wiener, E.L., Curry, R.E.: Flight-deck automation: promises and problems. Ergonomics. 23, 995–1011 (1980)
7. Shoham, Y., Perrault, R., Brynjolfsson, E., Clark, J., LeGassick, C.: Artificial Intelligence Index, November 2017. 101 (2017)
8. Ademujimi, T.T., Brundage, M.P., Prabhu, V. V.: A Review of Current Machine Learning Techniques Used in Manufacturing Diagnosis. IFIP Adv. Inf. Commun. Technol. 513, 407–415 (2017)
9. Nguyen, D.T., Duong, Q.B., Zamai, E., Shahzad, M.K.: Fault diagnosis for the complex manufacturing system. Proc. Inst. Mech. Eng. Part O J. Risk Reliab. 230, 178–194 (2015)
10. Yang, L., Lee, J.: Bayesian Belief Network-based approach for diagnostics and prognostics of semiconductor manufacturing systems. Robot. Comput. Integr. Manuf. 28, 66–74 (2012)
11. Liu, Y., Jin, S.: Application of Bayesian networks for diagnostics in the assembly process by considering small measurement data sets. Int. J. Adv. Manuf. Technol. 65, 1229–1237 (2013)
12. Hong, S.J., Lim, W.Y., Cheong, T., May, G.S.: Fault detection and classification in plasma etch equipment for semiconductor manufacturing e-diagnostics. IEEE Trans. Semicond. Manuf. 25, 83–93 (2012)
13. Zhang, Z., Wang, Y., Wang, K.: Fault diagnosis and prognosis using wavelet packet decomposition, Fourier transform and artificial neural network. J. Intell. Manuf. 24, 1213–1227 (2013)
14. Overmeyer, L., Podszus, F., Dohrmann, L.: Multimodal speech and gesture control of AGVs, including EEG-based measurements of cognitive workload. CIRP Ann. - Manuf. Technol. 65, 425–428 (2016)