Construction Robotics (2024) 8:16
https://doi.org/10.1007/s41693-024-00134-w

ORIGINAL PAPER

Autonomous AI shaft excavator: a case study on AI fairness for sustainability and green technology

Yoshiyasu Takefuji[1]

Received: 14 March 2024 / Accepted: 16 July 2024

Abstract
In Japan, the number of skilled shaft excavator engineers is decreasing. To complete the Linear Central Shinkansen line of 286 km between Tokyo and Nagoya, an AI-equipped shaft excavator was prototyped to absorb the tacit knowledge of highly skilled engineers. The AI can predict penetration resistance and optimally controlling the excavator. This not only reduces drilling time, enhancing sustainability, but also cuts CO_2 emissions by half. Datasets are built based on standard penetration tests and an ensemble–ensemble method with 16 determinants is used, achieving a prediction accuracy of 0.9. This paper presents a case study that AI capabilities are there to fill the gap, extend the skills or meet the shortage in the labor market. Trust of AI in fairness is addressed by calculating fairness as a benchmark with a variety of fairness metrics from all disciplines. From an information management perspective, this paper explores methods for managing the tacit knowledge of highly skilled, diminishing workers in civil engineering to enhance the sustainability of services and products. Tacit knowledge can drive innovation to boost sustainability.

Keywords AI shaft excavator · Standard penetration test (SPT) · Harsh shaft drilling environment

Abbreviations

lightgbm	Light gradient boosted machine
xgboost	Extreme gradient boosting
mlxtend	Machine learning extensions

1 Introduction

The construction sector is a major contributor to global CO_2 emissions, accounting for 23% of the total. Highly skilled engineers with heavy construction machines such as excavators can play a significant role in reducing these CO_2 emissions. In this paper, a highly skilled engineer is defined as one who possesses drilling skills at least at least twice as fast as those of an average engineer. One way to harness the expertise of these engineers is to use AI-equipped shaft excavators. These machines can learn from the tacit knowledge of highly skilled engineers, which can then be used to improve efficiency and reduce CO_2 emissions. AI-equipped shaft excavators can be programmed to identify the most efficient drilling patterns for different types of soil. This can help to reduce the amount of time and energy required to drill, which in turn can lead to significant reductions in CO_2 emissions. Autonomous AI shaft excavators can help construction engineers reduce dangerous working environments and promote green technology. The scope of this paper focuses on the transfer of tacit knowledge from highly skilled engineers to AI, and the enhancement of trust in AI through fairness. This paper demonstrates that tacit knowledge can drive innovation to boost sustainability by reducing 50% CO_2 emission. Tacit knowledge from highly skilled engineers has been harnessed to generate the current datasets. It is essential that all datasets are scrutinized from the perspective of fairness metrics. Similarly, all AI projects should be evaluated using fairness metrics for potential comparison. This paper serves as a pioneering project, introducing the first application of fairness metrics in construction robotics.

According to NIST (National Institute of Standards and Technology) special publication (NIST 1996), the Japanese construction industry was found to be large, solid, and progressive, with state spending on infrastructure and construction in general accounting for twice the GDP of the United States. In a country prone to earthquakes, typhoons, and other natural disasters, Japan led the world in the modernity

✉ Yoshiyasu Takefuji
takefuji@keio.jp

[1] Faculty of Data Science, Musashino University, 3-3-3 Ariake Koto-ku, Tokyo 135-8181, Japan

Published online: 25 July 2024

and quality of its built facilities and in the size and quality of its physics laboratories as of 1996 (NIST 1996). However, the number of skilled engineers is rapidly decreasing due to difficulties in passing on skills (MLIT.GO.JP. 2024). Skills are not manualized in Japan. Skilled Japanese engineers tend not to explain detailed techniques to other engineers. Therefore, generally, all types of shaft excavator engineers are in shortage in Japan (MLIT.GO.JP. 2024).

Knowledge comes in three forms: explicit, implicit, and tacit. Explicit knowledge can be shared, and when applied practically, it becomes implicit knowledge. Tacit knowledge, on the other hand, is gained from experience and is hard to share. This paper discusses a method to capture the tacit knowledge of highly skilled engineers in a dataset, effectively representing their expertise. We utilized fairness metrics to assess the fairness of our datasets.

This paper showcases a case study demonstrating how AI capabilities can bridge the gap, enhance skills, or address the scarcity in the labor market. While the article does discuss the potential of AI to perform tasks traditionally done by highly skilled engineers, it is important to clarify that the ultimate goal of AI is not to replace humans. Instead, AI can be used as a tool to augment human abilities, allowing us to work more efficiently and effectively. AI can handle repetitive tasks, analyze large amounts of data quickly, and even learn from its experiences. However, it lacks the ability to understand context, make ethical judgments, and bring a truly innovative and creative approach to problem-solving—these are uniquely human traits. Therefore, the vision should not be about replacing engineers with AI, but about creating a collaborative environment where AI and human expertise complement each other. This approach not only leverages the strengths of both but also addresses ethical concerns. In this light, the focus should be on developing AI systems that are transparent, explainable, and designed with human oversight in mind. This way, we can ensure that AI is used ethically and responsibly, and that it serves to enhance our capabilities rather than replace them. In response to the dwindling number of skilled workers, CEOs of construction corporations initiated this proposed AI project to bolster their human workforces.

In order to complete the Linear Central Shinkansen line with insufficient manpower, it was necessary to accelerate the prototype of a shaft excavator or shaft boring machine equipped with artificial intelligence (AI). The Chuo Shinkansen or Central Shinkansen is a Japanese maglev line under construction between Tokyo and Nagoya, running at a maximum speed of 505 km/h. About 90% of the 286 km between Tokyo and Nagoya will be tunnels while 10% will be above ground. The project requires more than 50 vertical shafts, each 40 m in diameter and 100 m deep to gain access to the tunnel horizon, provide ventilation and emergency access to the completed tunnels. No one knows the underground conditions. However, the AI shaft excavator can handle gravel, bedrock, sand and clay soils.

The autonomous prototype AI shaft excavator we developed is actually in use at construction sites such as the Central Shinkansen Line and others in Japan. The AI shaft excavator prototype is a fully automated machine with skills learned from the highly skilled operators. The skill of the engineer is to accurately predict penetration resistance and to optimally control the shaft excavator based on its penetration resistance. Because of the harsh shaft drilling environment, the conventional IoT (Internet of Things) sensors cannot be used. Shaft drilling bits are hot, high pressure, and humid so that the conventional IoT sensors cannot be used to measure the penetration resistance.

This paper presents a case study of human management in which highly skilled engineers possessing unshared tacit knowledge are captured by the proposed method. This method can significantly improve sustainability by reducing drilling times by a half. In other words, the skills of highly engineers, which are difficult to pass on, will be successfully absorbed by the multi-sensory datasets using the proposed method. A precise prediction of soil resistance can contribute to a reduction in both drilling time and CO_2 emissions. In additions, this paper addresses the trust of AI by calculating fairness as a benchmark, which will contribute to sustainability in green infrastructure and technology.

This paper details how the AI shaft excavator can predict penetration resistance to optimally control and thus reduce drilling time. The shorter the excavation time, the more sustainable the construction. This paper focuses on shaft excavators or shaft drilling machines. A digitized shaft excavator already existed, but a new AI was added to the digitized shaft excavator.

A comprehensive literature review was conducted on AI shaft excavator. In 1997, Shimizu et al. proposed the world's first unmanned deep foundation construction method in shaft which means covering from excavation to concrete placing (Shimizu, et al. 1997). Operators must remotely control the machines, not automated. However, it did not elaborate on how the operator's skills in the shaft could be learned by machine learning with the dataset.

In 1998, Bradley et al. studied the development, control and operation of an autonomous robotic excavator (Bradley and Seward 1998). They developed artificial intelligence-based control system utilizing motion control strategy for movement of the excavator bucket through ground. However, the developed excavator is not for deep foundation construction in shaft but solely for excavation.

Hara et al. developed small-diameter vertical shafts constructed in the shallow space of steep mountainous areas (Hara et al. 2019). Four small-diameter vertical shafts of 2.5 or 3.0 m as the foundation for high-voltage transmission line towers in mountainous areas are traditionally constructed manually in Japan. The developed unmanned excavation of a vertical shaft can be achieved by the combination of three components: remote-controlled excavation, excavated soil conveyance, and assembling and inserting the retaining wall at the top of the vertical shaft. Therefore, the shaft excavator must be controlled by engineers.

Lee et al. developed unmanned excavator vehicle system for performing dangerous construction work (Lee et al. 2019). However, the developed excavator needs human engineers to control it.

Zhiqiang et al. studied key technologies of drilling process with raise boring method (Liu and Meng 2015). They concluded that the mechanized method can significantly reduce the number of workers, increase the safety of shaft construction, reduce worker injuries, and improve operational efficiency compared to the conventional drill-and-blast method. The development and application of raise boring technology are promising. However, they have never applied their technologies to the real-world application.

Kim, K. et al. investigated and simulated modeling, and velocity-field control of autonomous excavator with main control valve (Kim et al. 2019). However, their technologies have not yet been applied in practice.

Liu W. et al. conducted a literature review on control systems and control strategies for excavators (Liu, et al. 2022). Mayer presented drilling into the autonomous future of the industry (Mayer 2021). Oybek Maripjon UgliEraliev et al. conducted a literature review on sensing, perception, decision, planning and action of autonomous excavators (UgliEraliev et al. 2022).

HeshanFernando et al. investigated autonomous excavators using proprioceptive force sensing and machine learning and achieved classification accuracy (90%) using basic signal features and simple classifiers (HeshanFernando, et al. 2020). However, they have not mentioned how to improve the speed of penetration in shaft drilling.

The goal of the proposed AI shaft excavator is to learn the unshared tacit skill from highly skilled engineers with datasets and to reduce the drilling time with optimal control. Veteran operators in Japan can drill at least twice as fast as non-veteran operators. The proposed excavators can reduce CO_2 emissions from construction sites by up to 50%. The digging performance of highly skilled workers surpasses that of average workers by more than double. Consequently, the proposed excavators have the potential to reduce CO_2 emissions at construction sites by up to 50%.

To the best of our knowledge, there are no papers on unshared tacit knowledge management of highly skilled engineers. In other words, the dataset created can represent the unshared and unexplained tacit knowledge of highly skilled engineers.

Drilling efficiency is largely determined by the engineer's experience, but the underlying operational mechanisms are often overlooked (Shen, et al. 2022). The national renewable energy laboratory (NREL) is working with geothermal operators to introduce lean manufacturing principles to the geothermal drilling industry (Gov and Efficiency xxxx). Time and experience have shown that this can lead to significant savings. The practical experience gained from existing research has shown that the oil and gas construction industry can improve its sustainability by incorporating innovative strategies (Waqar et al. 2023).

The total CO_2 emission of the construction sector was 5.7 billion tons which made up 23% of the emissions of global economic activity in 2009 (Sizirici et al. 2021). This indicates the contribution of this paper is significant.

Fairness has been raised in machine learning issues in a variety of areas (Chen et al. 2023; Drukker et al. 2023; Kaminwar et al. 2023). Computing fairness metrics in machine learning is beneficial for several reasons. It aids in the identification and reduction of biases in model predictions, which can arise from imbalanced datasets or biased feature selection. This ensures that models can generalize better to diverse populations, leading to more accurate and reliable predictions across different groups. Fairness in machine learning is also crucial for ethical decision-making, particularly in sensitive domains such as healthcare, finance, and criminal justice where biased decisions can have significant real-world consequences. In many jurisdictions, laws and regulations require algorithmic decisions to be fair, and computing fairness metrics can help demonstrate compliance with these regulations. Lastly, fair models can enhance trust among end-users. When users understand that a model makes decisions fairly, they are more likely to accept and use the technology. However, it is important to remember that while fairness metrics can guide us towards more equitable models, they are just one piece of the puzzle. The context in which the model is used, as well as continual monitoring and adjustment of the model, are also necessary. In general, it is crucial to evaluate datasets utilized for machine learning from a fairness standpoint. Our investigation reveals a lack of research on fairness metrics computing within the realm of construction robotics. This paper aims to bridge that gap by showcasing the application of fairness computing in this field.

A literature review was conducted on fairness on datasets and algorithms. Chen et al. addressed current AI systems in healthcare can be unfair and lead to disparities in care (Chen et al. 2023). Their perspective discussed algorithmic biases that arise in data acquisition, genetic variation, and intra-observer labeling variability, and emerging technologies for mitigating biases such as disentanglement, federated learning, and model explainability (Chen et al. 2023).

Drukker et al. found that medical imaging AI can introduce bias at multiple steps in the development and deployment process (Drukker et al. 2023). A multi-institutional team identified 29 sources of potential bias and developed recommendations for best practices to mitigate bias (Drukker et al. 2023). Kaminwar et al. argued that the democratization of AI requires the democratization of the verification process of ML systems, in order to ensure fairness and mitigate bias (Kaminwar et al. 2023).

Nazer et al. found that the adoption of AI algorithms in healthcare is rapidly increasing, but there is a risk of bias and disparities in the development and implementation of these algorithms (Nazer et al. 2023). It is important to understand the sources of bias and to develop strategies to mitigate them, in order to ensure fairness in healthcare (Nazer et al. 2023).

Kim et al. developed a deep learning model to predict financial losses due to accidents at apartment construction sites (Kim et al. 2022). The model was trained on insurance claim payout data and can be used to prevent and reduce the risk of financial loss. However, the study also found that the AI model can be unfair, as it can unilaterally provide results of predictions without explaining the basis and process for drawing the results. This could negatively affect the reliability and fairness of future models (Kim et al. 2022). The findings of the review highlight the absence of fairness computing in construction robotics, underscoring the significance of this paper's contribution in introducing fairness metrics.

The proposed method can be applied to general tacit knowledge management if datasets can be created with highly skilled experts or engineers. This paper suggests that by eliminating hazardous working environments for humans and reducing drilling time with AI, we can enhance sustainability and promote green technologies with trust of AI from a fairness perspective. Trust of AI with fairness can be improved by calculating and showing values of fairness metrics with datasets.

The proficiency of shaft excavator engineers can be gauged by the accuracy of their predictions regarding soil penetration resistance, as measured by the standard penetration test (SPT). Greater accuracy in these predictions leads to reduced drilling time. The proprietary dataset is composed of 24,377 instances and 16 determinants with N-value named as SPT which is a full list of variables. The standard penetration test (SPT) procedure is notably time intensive. It involves manually measuring 'N' values per 300 mm, a process repeated 24,377 times for a total soil tube extraction of 400 m by boring machines. Details will be explained in Sect. 2.

Fairness metrics in machine learning offer numerous benefits. They help identify and reduce biases, ensuring models generalize better across diverse groups. This leads to more accurate predictions and ethical decision-making, particularly in sensitive areas such as healthcare and finance. Fair models can also enhance user trust and demonstrate regulatory compliance. Despite their importance, there is a noticeable gap in applying fairness metrics within construction robotics. This paper aims to address this by highlighting the significance of fairness in machine learning in construction robotics and the need for continual monitoring and adjustment of models and comparison in the future with other projects for researchers.

This paper makes three significant contributions: (1) it demonstrates the superior performance of the ensemble–ensemble regressor over the random forest regressor and extra trees regressor, (2) it discusses the creation of datasets by highly skilled engineers for the acquisition of tacit knowledge using *N*-values or soil penetration resistance for controlling shaft excavators, and (3) it introduces the application of fairness metrics in machine learning within the field of construction robotics to validate the fairness of created datasets.

1.1 Shaft excavator machine learning

Supervised machine learning is based on a skilled engineered dataset: $y = f(X)$, consisting of a target variable y or y penetration resistance and a set of input variables X.

The penetration resistance (*N*-value) of target y is given by $y = f(X) = f(x_1, x_2, \ldots, x_n)$ where x_1, x_2, and x_n are influence determinants and the function f() can be trained as the relationship function between the influence determinants of X and the target y.

The effectiveness of shaft excavator engineers is determined by their ability to accurately predict soil penetration resistance, as quantified by the standard penetration test (SPT). Enhanced prediction accuracy results in decreased drilling time. The proprietary dataset, comprising 24,377 instances and 16 determinants, includes an '*N*' value referred to as SPT, which represents a comprehensive list of variables. The SPT procedure, which is notably time-consuming, involves the manual measurement of '*N*' values. This process is repeated 24,377 times, resulting in a total soil tube extraction of 400 m by boring machines with multiple times.

Despite the challenges posed by the lack of public access to the dataset for independent validation, this paper offers a detailed methodology to address this issue. It encompasses the algorithms, Python programs, parameters, and statistical methods utilized to capture SPT N-values. These values are essential for engineers to predict soil resistance, a critical factor in the control of shaft excavators. More precise predictions not only reduce drilling time but also decrease CO_2 emissions. Essentially, this paper proposes a novel method for generating datasets that enable AI shaft excavators to optimally predict soil resistance. Consequently, those with access to similar data can replicate the study, thereby validating its findings.

This paper examines and compares three ensemble methods: (1) random forest regressor, (2) extra trees regressor, and (3) stacking ensemble method. The stacking ensemble method is an ensemble–ensemble method which can ensemble multiple ensemble methods: lightgbm, xgboost, extra trees regressor, and random forest regressor.

Tenfold cross-validation is used for evaluating three methods for comparison. The proprietary dataset is composed of 24,377 instances and 16 multi-sensory input variables. Existing studies do not provide a comparison between random forest algorithms and ensemble–ensemble machine learning algorithms using real-world datasets in construction robotics. The advantages of ensemble–ensemble algorithms over existing ones become evident in this context by showing the improvement in the prediction accuracy.

To the best of our knowledge, there are no papers in the construction machine field on ensemble–ensemble methods demonstrated. Therefore, a new contribution of this paper is to present an ensemble–ensemble method with source code disclosure and to validate and compare the results of conventional ensemble methods with tenfold cross-validation. The effectiveness of the ensemble–ensemble method over the conventional ensemble methods is demonstrated. The AI shaft excavator proposed in this study is indeed utilized in a variety of fields and locations, distinguishing it from existing research that is largely based on simulations rather than real-world applications. In other words, the proposed paper will demonstrate the effectiveness of the computer application in industry and business.

1.2 Fairness based on computational ethics

The trust of AI systems must be addressed in all existing AI systems, but in many original papers (NIST 1996; Shimizu, et al. 1997; Bradley and Seward 1998; Hara et al. 2019; Lee et al. 2019; Liu and Meng 2015; Kim et al. 2019; Liu, et al. 2022; Mayer 2021; UgliEraliev et al. 2022; HeshanFernando et al. 2020; Shen, et al. 2022; Gov and Efficiency xxxx; Waqar et al. 2023; Sizirici et al. 2021) in civil engineering, only the quality of accuracy with cross-validation is usually reported. This paper addresses fairness based on computational ethics by calculating fairness metrics such as disparate impact, average odds error, average odds difference, equal opportunity difference, between group generalized entropy error, conditional demographic disparity, Theil index, consistency score, KL divergence, statistical parity difference, and class imbalance, respectively. Scores of fairness metrics from all disciplines can show a variety of biases to ensure trust of AI. Fairness metrics on origin area and description are summarized in Table 1. The shaded rows indicate that fairness metrics can only be applied to classification tasks. The proposed method is used for solving regression problems.

The details of fairness metrics, including "disparate impact", average odds error (AOE), conditional demographic disparity (CDD), and Theil index, are thoroughly discussed. According to Merriam-Webster legal, "disparate impact" is a legal theory that allows challenges to practices

Table 1 Fairness metrics on original area and description

Fairness metric	Original area	Description
Disparate impact	Law	Measures the difference in the rates at which a protected group is denied a loan, job, or other outcome, compared to a non-protected group
Average odds error	Statistics	Measures the difference in the average odds of receiving a favorable outcome for two groups
Average odds difference	Statistics	Measures the difference in the average odds of receiving a favorable outcome for two groups, after adjusting for covariates
Equal opportunity difference	Statistics	Measures the difference in the false-positive rates for two groups
Between group generalized entropy error	Economics	Measures the difference in the expected value of the entropy of the predicted probabilities for two groups
Conditional demographic disparity	Statistics	Measures the difference in the distributions of predicted probabilities for two groups, conditional on a set of covariates
Theil index	Economics	Measures the overall inequality in a distribution
Consistency score	Machine learning	Measures the consistency of a model's predictions with a set of fairness constraints
KL divergence	Information theory	Measures the distance between two probability distributions
Statistical parity difference	Statistics	Measures the difference in the proportions of two groups that receive a favorable outcome
Class imbalance	Machine learning	Measures the imbalance in the distribution of classes in a dataset

or policies that appear to be nondiscriminatory but have a disproportionately adverse effect on members of a protected class, such as race, color, religion, sex, or disability. Disparate impact can be used to identify unlawful discrimination in employment, housing, education, and other areas. To prove disparate impact, plaintiffs must show that the practice or policy has an unjustified negative impact on the protected group compared with non-members of the group.

Average odds error (AOE) is a measure of fairness in machine learning models. It is a relaxed version of equality of odds, which means that it does not require the model to have the same error rate for all protected groups. Instead, AOE measures the difference in error rates between the protected and unprivileged groups.

Conditional demographic disparity (CDD) is a measure of fairness in machine learning models. It is a relaxed version of demographic parity, which means that it does not require the model to have the same proportion of positive outcomes for all protected groups. Instead, CDD measures the difference in the proportion of positive outcomes between the protected and unprivileged groups, conditional on a set of covariates.

The Theil index is a measure of inequality that is used to quantify the distribution of income or wealth in a population. It is named after the Dutch economist Henri Theil, who developed it in the 1960s.

In machine learning, consistency score is a measure of how similar the labels are for similar instances. It is a metric of fairness that is used to assess whether a machine learning model is making consistent predictions across similar instances, regardless of their protected class.

Since there is no paper on fairness in civil engineering, we can only show the result as a benchmark.

2 Methods

Eight datasets (each at every 50 m depth) were created with highly skilled engineers. Eight datasets were converted to a single dataset. Each dataset is with *N*-value and 16 multi-sensory variables such as head lubrication pressure, air pressure, frame tilt angle X, hydraulic pump pressure P2, frame tilt angle Y, spindle speed, spindle torque, water rate flow, hydraulic pump pressure P1, oscillator pressure, engine speed, bit load, water pressure, engine water temperature, drilling speed, and battery voltage. The *N*-value represents penetration resistance which plays a key role in reducing the drilling time. The more accurate the prediction of penetration resistance, the shorter the drilling time. The N-value will be detailed.

The purpose of the standard penetration test (SPT) is to determine the relative density and consistency of the subsoils, and to obtain disturbed soil samples for field visual identification of soil samples. SPT is from 0 to 100 in the dataset.

The SPT procedure is described in ISO 22476-3. It produces an *N*-value, which represents the number of blows of a standardized sampler driven into the soil a standardized distance. The sampler is 51 mm O.D. (outside diameter) and it is driven into the soil with a 63.5 kg weight having a free fall of 760 mm. The first 150 mm of soil is neglected. The next 300 mm of soil constitutes the test. The number of blows for that 300 mm becomes the *N*-value. *N*-value represents standard penetration resistance. Therefore, creating a dataset is a time-consuming task. It took 3 months to create proprietary multi-sensory eight datasets with highly skilled engineers. The eight datasets were converted to a single dataset. The SPT procedure is the most time-consuming, with N-values measured manually 24,377 times out of a total of 400 m.

We have compared three ensemble methods: (1) random forest regressor, (2) extra trees regressor, and (3) mlxtend stacking regressor (ensemble–ensemble method) with four ensemble methods: lightgbm, xgboost, extra trees regressor, and random forest regressor.

A random forest is a meta-estimator that fits a number of classifying decision trees on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.

An extra trees is also a meta-estimator that fits a number of randomized decision trees. Random forest chooses the optimum split, while extra trees chooses it randomly.

Stacked generalization consists in stacking the output of individual estimator and use a regressor to compute the final prediction. Stacking allows to use the strength of each individual estimator using their output as input of a final estimator.

The mlxtend stacking allows us to use the ensemble–ensemble method. In other words, the mlxtend stacking allows us to ensemble three regressors including lightgbm, xgboost, extra trees with meta-regressor such as random forest regressor. There are four possible cases choosing a meta-regressor among four ensemble methods: lightgbm, xgboost, extra trees regressor, and random forest regressor. The random forest regressor performed best as a meta-regressor. APPENDIX shows the ensemble–ensemble method. In APPENDIX, 'SPT' represents the N-value of penetration resistance.

3 Results

3.1 Accuracy analysis with cross-validation

The results of the tenfold cross-validation showed a significant difference in the prediction accuracy of the three methods such as random forest regression, extra trees regression, and stacking regression (ensemble–ensemble) with four ensemble methods such as lightgbm, xgboost, extra trees with meta-regressor such as random forest regressor. Table 2 summarizes the average prediction accuracy results from the tenfold cross-validation of random forest regression, extra tree regression, and ensemble–ensemble regression. Values in Table 2 represent the average prediction accuracy of tenfold cross-validation. Table 2 presents two columns: 'Method' and 'Cross-Validation Average Prediction Accuracy'. It compares three methods: the random forest regressor, extra trees regressor, and ensemble–ensemble regressor. The 'Cross-Validation Average Prediction Accuracy' column includes an average value and ten additional values in brackets, which represent the results of a tenfold cross-validation. As indicated in Table 2, the ensemble–ensemble regressor consistently outperforms the random forest and extra trees regressors.

The proprietary dataset is composed of 24,377 instances and 16 determinants with N-value named as SPT. It took more than 3 months with highly skilled engineers to create the proprietary dataset.

The result of the proposed regression is as shown:

Accuracy is 0.9047, MSE: 58.5980, RMSE: 7.6549, MAE: 4.0938, R^2: 0.9047.

3.2 Fairness computing for trust of AI

Due to the regression nature of this problem, only three fairness metrics were examined using our datasets**.** In other words, a greater number of fairness metrics can be applied to classification problems. The results of our study will serve as a benchmark for comparisons with other datasets. The Theil index is a measure of economic inequality and, similar to other inequality metrics, a lower Theil index value indicates a more equal distribution. Therefore, a lower Theil index is generally considered better.

A consistency score closer to 1 is generally considered better. This score indicates a high level of consistency in the data or model, which is a desirable attribute in many contexts, particularly in machine learning and data analysis. A score of 1 typically signifies perfect consistency.

The statistical parity difference is a measure used in fairness metrics, particularly in machine learning models. A value close to zero indicates that the outcomes of the model are fair across different groups. Therefore, a statistical parity difference value closer to zero is generally considered better as it signifies a more fair and unbiased model. The findings suggest that our dataset exhibits consistent fairness. It is important to note, however, that these results have not been compared with those from existing studies:

theil_index: 0.1182.

consistency_score: 0.9047.

statistical parity difference: 0.0

4 Discussion

Many fairness metrics can be used in classification problems between two groups to be compared on fairness. However, in regression, three fairness metrics such as Theil index, consistency score, and statistical difference are calculated. Theil index of 0.1182 in regression suggests that there is a very weak positive relationship between the independent variable and the dependent variable. The consistency score with (1 − mse/(y_test)): 0.9047 in regression indicates a very good fit between the independent and dependent variables. Statistical parity difference is a measure of fairness in a model relative to a protected attribute. A value of 0 indicates that the model is perfectly fair, while values between −0.1 and 0.1 are generally considered to be reasonably fair. In the context of regression analysis, statistical parity difference can be used to evaluate the fairness of a regression model with respect to a protected attribute. A statistical parity difference of 0.0 would indicate that the model is perfectly fair with respect to the protected attribute, meaning that there is no difference in the predicted outcomes for different groups defined by the protected attribute.

This paper shows how to build the dataset for two ensemble methods such as random forest and extra trees and the ensemble–ensemble method using stacking regressor with three regressors including lightgbm, xgboost, extra trees with meta-regressor including random forest regressor. The mlxtend stacking is an ensemble–ensemble machine learning method. The random forest regressor performed best as a meta-regressor.

Table 2 Cross-validation average prediction accuracy

Method	Cross-validation average prediction accuracy
Random forest regressor	0.5986: [0.60847767 0.60346399 0.59708295 0.60437557 0.57247037 0.61394713 0.59389243 0.60619872 0.58295351 0.60391978]
Extra trees regressor	0.592: [0.59799453 0.59845032 0.58842297 0.60118505 0.57064722 0.60665451 0.59434822 0.5952598 0.58067457 0.58659982]
Ensemble–ensemble regressor	0.908: [0.9121502 0.90925775 0.90681729 0.9073827 0.91054546 0.91035381 0.91701385 0.91333652 0.90334382 0.89268235]

Table 2 shows that the stacking method or ensemble–ensemble method achieved high prediction accuracy exceeding 0.9 with tenfold cross-validation. For the random forest regressor and extra tree regressor, the prediction accuracy is about 0.6 with tenfold cross-validation. Stacking regressor or ensemble–ensemble can take advantage of strengthening individual ensemble methods automatically for improving the prediction accuracy.

The proposed stacking method or ensemble–ensemble method can be applied to improving deep learning. The better the prediction accuracy, the shorter the drilling time is. The shorter the drilling time, the more sustainable the shaft excavator will be.

The prototype autonomous AI shaft excavator we developed is actually in use at construction sites such as the Central Shinkansen Line and others in Japan. Remember that skilled operators and engineers are needed to create high-quality datasets for the tacit knowledge transfer. The larger the dataset, the better the prediction accuracy is.

The proposed stacking method or ensemble–ensemble method achieved the prediction accuracy over 0.9 in tenfold cross-validation. The proposed prototype AI shaft excavators have been used in actual construction fields and sites in Japan. This paper presented a case study of human management in which highly skilled engineers are replaced by AI. In other words, the skills of super-engineers, which are difficult to pass on, were successfully absorbed by the multi-sensory dataset with 16 multi-sensory determinants. This paper presented a specific real-world problem of how to replace operators or engineers with tacit knowledge for green technology and sustainability. This problem is significant because the construction sector is a major contributor to global CO_2 emissions. The proposed method uses AI to capture the tacit knowledge of highly skilled operators and engineers. This knowledge can then be used to train AI-powered machines that can perform the same tasks as the human operators. The proposed method can be applied to general tacit knowledge management as long as datasets can be created.

The results of the fairness metrics examination on our datasets (theil_index: 0.1182, consistency_score: 0.9047, statistical parity difference: 0.0) affirm their consistent fairness.

5 Conclusion

The proposed stacking method or ensemble–ensemble method achieved high prediction accuracy exceeding 0.9 with tenfold cross-validation. This method can be applied to improving deep learning and reducing drilling time, which will make shaft excavators more sustainable. The proposed method can also be applied to general tacit knowledge management as long as datasets can be created. This paper demonstrated computational fairness with datasets. This study illustrates how tacit knowledge can fuel innovation to enhance sustainability.

APPENDIX

APPENDIX: Ensemble–ensemble source code: crossst_fair.py

```
# -*- coding: utf-8 -*-
import sys,codecs.
import numpy as np.
np.random.seed(0).
import pandas as pd.
import sys.
import warnings.
warnings.filterwarnings("ignore").
import logging.
logging.getLogger().setLevel(logging.ERROR).
filename = all2. 'csv'.
trees = 493.
dig = pd.read_csv(filename).
dig.fillna(0,inplace = True).
crossv = 10.
print('filename:',filename).
print("data instances and parameters:",dig.shape).
#print("cross-validation:",crossv).
print("trees:",trees).
X = dig.drop(['Layer', 'SPT'],axis = 1).
y = dig['SPT'].
print('max SPT',max(y)).
print('min SPT',min(y)).
from sklearn.ensemble import ExtraTreesRegressor.
from sklearn.ensemble import RandomForestRegressor.
from mlxtend.classifier import StackingClassifier.
from mlxtend.regressor import StackingRegressor.
import xgboost as xgb.
from sklearn.model_selection import cross_val_score.
from sklearn.metrics import *
from sklearn.model_selection import ShuffleSplit.
from sklearn.model_selection import train_test_split.
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score.
from aif360.sklearn.metrics import disparate_impact_ratio,average_odds_error,generalized_fpr,average_odds_difference,kl_divergence,between_group_generalized_entropy_error.
```

```
from aif360.sklearn.metrics import generalized_fnr,difference,statistical_parity_difference,equal_opportunity_difference,conditional_demographic_disparity,theil_index,consistency_score,class_imbalance.
from aif360.metrics import BinaryLabelDatasetMetric.
from aif360.metrics import ClassificationMetric.
from aif360.datasets import BinaryLabelDataset.
from aif360.sklearn.detectors import bias_scan.
from aif360.detectors.mdss.ScoringFunctions.ScoringFunction import ScoringFunction.
import lightgbm as lgb.
#cv = ShuffleSplit(n_splits = crossv, test_size = 0.3, random_state = 54).
lg = lgb.LGBMRegressor(num_leaves = 31,learning_rate = 0.48,n_estimators = 182).
xg = xgb.XGBRegressor(n_estimators = trees).
ext = ExtraTreesRegressor(n_estimators = trees, max_depth = None,min_samples_split = 2, random_state = 0).
rf = RandomForestRegressor(n_estimators = trees).
sclf = StackingRegressor(regressors = [lg,ext,xg],meta_regressor = rf).
X_train,X_test,y_train,y_test = train_test_split(X,y, test_size = 0.2, random_state = 54).
#scores = cross_val_score(sclf, X, y, cv = cv).
#print(scores,scores.mean(),round(scores.std(),6)).
sclf.fit(X_train,y_train).
y_pred = sclf.predict(X_test).
print(sclf.score(X_test,y_test)).
mse = mean_squared_error(y_test, y_pred).
rmse = mean_squared_error(y_test, y_pred, squared = False).
mae = mean_absolute_error(y_test, y_pred).
r2 = r2_score(y_test, y_pred).
print(‘MSE:’, mse).
print(‘RMSE:’, rmse).
print(‘MAE:’, mae).
print(‘R^2:’, r2).
##print(‘disparate impact:’,disparate_impact_ratio(y_test, y_pred)).
#print(‘average odds error:’,average_odds_error(y_test, y_pred)).
#print(‘average_odds_difference:’,average_odds_difference(y_test,y_pred)).
#print(‘equal_opportunity_difference:’,equal_opportunity_difference(y_test,y_pred)).
#print(‘between_group_generalized_entropy_error:’,between_group_generalized_entropy_error(y_test,y_pred)).
#print(‘conditional_demographic_disparity:’,conditional_demographic_disparity(y_test,y_pred)).
#print(‘KL_divergence:’,kl_divergence(y_train + y_test)).
#print(‘class_imbalance:’,class_imbalance(y_train + y_test)).
y_pred = pd.DataFrame(y_pred)[0].
print(y_pred.shape,y_test.shape,len(y_test)).
def theil_index(y_test, y_pred):
# Calculate the ratio of predicted to actual values.
ratio = y_pred / y_test.
# Calculate the natural logarithm of the ratio.
log_ratio = np.log(ratio).
# Calculate the Theil index.
theil_index = np.sum(ratio * log_ratio)/len(y_test).
return theil_index.
print(“The Theil index is:”, theil_index(y_test,y_pred)).
print(‘theil_index:’,theil_index(y_test,y_pred)).
print(‘consistency_score:’,1—(mse / np.var(y_test))).
#print(‘consistency_score:’,consistency_score(X_test,y_pred)).
print(‘statistical parity difference:’,statistical_parity_difference(y_train + y_test)).
```

Acknowledgements We would like to thank Mr. Haruhiko Ito of Toatone boring for his work on the AI shaft excavator project.

Author contributions YT completed this research and wrote this article.

Funding This research has no fund.

Data availability The proprietary dataset used in this paper is not available.

Declarations

Conflict of interest The author has no conflict of interest.

References

Bradley DA, Seward DW (1998) The development, control and operation of an autonomous robotic excavator. J Intell Rob Syst 21:73–97. https://doi.org/10.1023/A:1007932011161

Chen RJ, Wang JJ, Williamson DFK, Chen TY, Lipkova J, Lu MY, Sahai S, Mahmood F (2023) Algorithmic fairness in artificial intelligence for medicine and healthcare. Nature Biomed Eng 7(6):719–742. https://doi.org/10.1038/s41551-023-01056-8

Drukker K, Chen W, Gichoya J, Gruszauskas N, Kalpathy-Cramer J, Koyejo S, Myers K, Sá RC, Sahiner B, Whitney H, Zhang Z, Giger M (2023) Toward fairness in artificial intelligence for medical image analysis: identification and mitigation of potential biases in the roadmap from data collection to model deployment. J Med Imag 10(6):061104. https://doi.org/10.1117/1.JMI.10.6.061104

Hara T et al (2019) Small-diameter vertical shafts constructed in the shallow space of steep mountainous areas. Underground Space. https://doi.org/10.1016/j.undsp.2018.10.005

HeshanFernando, et al (2020) What lies beneath: material classification for autonomous excavators using proprioceptive force sensing and machine learning. Autom Constr. https://doi.org/10.1016/j.autcon.2020.103374

Kaminwar SR, Goschenhofer J, Thomas J, Thon I, Bischl B (2023) Structured verification of machine learning models in industrial settings. Big Data 11(3):181–198. https://doi.org/10.1089/big.2021.0112

Kim K et al (2019) Modeling and velocity-field control of autonomous excavator with main control valve. Automatica. https://doi.org/10.1016/j.automatica.2019.02.041

Kim JM, Bae J, Park H, Yum SG (2022) Predicting financial losses due to apartment construction accidents utilizing deep learning techniques. Sci Rep 12(1):5365. https://doi.org/10.1038/s41598-022-09453-w

Lee J, Kim B, Sun D, Han C, Ahn Y (2019) Development of unmanned excavator vehicle system for performing dangerous construction work. Sensors 19(22):4853. https://doi.org/10.3390/s19224853

Liu Z, Meng Y (2015) Key technologies of drilling process with raise boring method. J Rock Mechan Geotech Eng 7(4):385–394. https://doi.org/10.1016/j.jrmge.2014.12.006

Liu W. et al. (2022). Review on control systems and control strategies for excavator. J. Phys.: Conf. Ser. 2301 012023. https://doi.org/10.1088/1742-6596/2301/1/012023

Maksim Mayer. (2021). Drilling into the autonomous future of the industry. https://coringmagazine.com/article/autonomous-future/

MLIT.GO.JP. Securing Leaders for the Development of Infrastructures, Improving On-site Productivity, and Adopting New Technologies. P.123 Accessed on June 12, 2024. https://www.mlit.go.jp/common/001157859.pdf

Nazer LH, Zatarah R, Waldrip S, Ke JXC, Moukheiber M, Khanna AK, Hicklen RS, Moukheiber L, Moukheiber D, Ma H, Mathur P (2023) Bias in artificial intelligence algorithms and recommendations for mitigation. PLOS Digital Health 2(6):e0000278. https://doi.org/10.1371/journal.pdig.0000278

NIST. (1996). *Innovation in the Japanese construction.* https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication898.pdf

NREL.Gov. Geothermal Drilling Efficiency. https://www.nrel.gov/geothermal/drilling-efficiency.html. Accessed 23 July 2024

Shen Q et al (2022) Efficiency evaluation of a percussive drill rig using rate-energy ratio based on rock drilling tests. J Petrol Sci Eng. https://doi.org/10.1016/j.petrol.2022.110873

Shimizu, M. et al. (1997). Development of unmanned deep foundation construction method in shaft. Proceedings of the 14th ISARC, Pittsburgh, USA. https://doi.org/10.22260/ISARC1997/0029

Sizirici B, Fseha Y, Cho CS, Yildiz I, Byon YJ (2021) A review of carbon footprint reduction in construction industry, from design to operation. Mater (basel, Switzerland) 14(20):6094. https://doi.org/10.3390/ma14206094

UgliEraliev OM et al (2022) Sensing, perception, decision, planning and action of autonomous excavators. Autom Constr. https://doi.org/10.1016/j.autcon.2022.104428

Waqar A, Othman I, Shafiq N, Mansoor MS (2023) Applications of AI in oil and gas projects towards sustainable development: a systematic literature review artificial intelligence review. Adv Online Public. https://doi.org/10.1007/s10462-023-10467-7