Trusted AI Challenge
Stage-1: Summer 2024

Submitted By

Nishat Ara Nipa
Graduate Research Assistant
Old Dominion University
Email: nishataranipa@gmail.com

Dr. Sachin Shetty
Professor
Old DominionUniversity
Email: sshetty@odu.edu

Introduction:

The system under analysis integrates an Unmanned Ground Vehicle (UGV) for mine-clearing operations with an Unmanned Aerial Vehicle (UAV) equipped with a multi-spectral video collection system. The UAV captures video data for each segment of the network, which is then analyzed by an AI model to predict whether the area contains mines or is clear. However, the AI's accuracy is variable and influenced by different metadata conditions. The Command and Control Center uses these predictions to determine the UGV's path. If the UGV encounters a mine, it takes 1 hour to clear the link, whereas if no mine is detected, it clears the link in 20 minutes. A human review of the video data takes 30 minutes per link, compared to just 1 minute for the AI model.

The core of this research revolves around critical questions regarding the integration of AI in mine-clearing operations. Specifically, we seek to determine whether the system should rely on AI predictions, if a human operator should assess the imagery instead, and to what extent humans and AI should share responsibilities within this system. A pivotal consideration is whether the system's architecture and operation should be designed to influence trust in AI decision-making. To address these questions, we have been provided with data capturing AI and Human Subject Matter Expert (SME) abilities to detect mines, collected during four test events at two different locations.

The expected outcomes of this research include characterizing the provided performance map to develop a nuanced understanding of how AI performance varies with different metadata conditions. Additionally, we aim to establish relationships between the model's performance and the overall system performance, which will inform decisions on system design and operational strategies.

Research Approach:

To address this problem, our core focus was to characterize the provided performance map under varying metadata conditions. The key objectives were to thoroughly analyze and compare human and AI performance, and to establish relationships between these performances across different terrain and environmental conditions. The following tasks were accomplished to draw comparison and identify pattern to use in the next stage.

- Data cleaning and conversion to numeric values were done to ensure accuracy in subsequent analyses.
- Key metrics, such as mean accuracy, standard deviation, mean difference, and performance ratio for AI and human operators across different terrains, were calculated. We averaged the condition values for each terrain type. The plots generated offer a visual and quantitative means to compare AI and human performance, identify strengths and weaknesses, and make informed decisions on system design and operational strategies based on terrain conditions.

- Next we compared performance metrics with factors like temperature, wind speed, and visibility to identify which conditions are more challenging for AI or humans and make recommendations on how to optimize system performance based on environmental factors
- We calculated a correlation matrix to examine the relationships between AI performance across different links and the various environmental conditions to find consistency in AI performance and the impact of environmental conditions.
- Finally we performed a performance threshold Analysis to identify specific environmental condition envelop that allow AI to perform optimally. This help in understanding conditions under which the AI systems are most reliable.

Key Findings from Stage-1:

- The accuracy comparison plots in Figure 1(a), 1(c) and 1(d) shows that the AI model generally performs well across most terrains, with mean accuracy exceeding 0.8 in grassy, rocky, and sandy terrains. However, the wooded terrain shows a noticeable drop in mean accuracy for AI, which is significantly lower than human performance. We also measured the variability in performance by measuring the standard deviation of accuracy as shown in Figure 1(b). AI performance is relatively consistent in grassy and rocky terrains (low standard deviation), but variability increases in sandy, swampy, and particularly wooded terrains. The higher standard deviation in wooded and swampy terrains highlights that AI performance is less reliable in these conditions, reinforcing the need for human oversight in such environments. We can also visualize the mean accuracy by terrain types in the heatmap illustrated in Figure 2.

- For a deeper understanding of how environmental conditions impact AI and human peroformance, we compared performance metrics with factors like temperature, wind speed and visibility. The findings are as below:
    - ➢ AI performance fluctuates more significantly than human performance as temperature changes. Notably, AI accuracy decreases sharply at lower temperatures (Night A) but peaks during Day B when the temperature is highest. Human performance remains relatively stable across varying temperatures, with only slight deviations, suggesting that AI is more sensitive to temperature changes than humans [Figure 3(a)]
    - ➢ A similar pattern to temperature is observed, where AI accuracy declines sharply during periods of high wind speed (Night A) and improves when wind speed is moderate (Day B).Human performance shows a slight increase during high wind speeds, indicating that AI is more negatively affected by increased wind speed compared to human operators [Figure 3(b)]
    - ➢ AI performance drops drastically during periods of low visibility (Night A) and improves significantly when visibility is higher (Day B).Human performance, on the other hand, remains relatively stable, albeit with some minor fluctuations. This suggests that AI models are more reliant on clear visibility conditions for optimal

performance, whereas human performance is less dependent on visibility changes. [Figure 3(c)]

- We further calculated a correlation matrix to examine the relationships between AI performance across different links and the various environmental conditions. It can be visualized in [Figure 4] by using a heatmap where the color intensity indicates the strength and direction of correlations. As we can observe, most links demonstrate high positive correlation with each other, especially between Link_3 to Link_9, indicating consistent performance behavior
    - ➢ The column "currentTemperature" shows varied correlations with the links. Positive correlation, especially with Link_8 (0.84) and Link_2 (0.47), suggesting performance on these links increases with temperature.
    - ➢ The "currentWindSpeed" column has a strong positive correlation with Link_7 (0.92) but a negative correlation with Link_1 (-0.91). This suggests that wind speed affects performance differently across links.
    - ➢ The "currentVisibility" column negatively correlates with most links except Link_9 (0.64) and Link_1 (1), indicating visibility generally negatively impacts performance, but Link_9 serves as an exception.

- Finally we identified the specific environmental conditions that allow AI to perform optimally. The boxplot provided in [Figure 5] illustrates the range of environmental conditions during different times of the day when AI performance exceeds a predefined threshold (0.90). This analysis can be used to make informed decisions about deploying AI systems in different environments, ensuring that they operate under conditions where they are most likely to succeed.

Future Research Direction:

Building upon the insights gained from our analysis of how AI performance varies with different metadata conditions and the comparative assessment between AI and human performance, future research will focus on two primary directions:

1. Incorporation of Metadata-Driven Insights into AI Models: The knowledge acquired from understanding the specific metadata conditions under which AI performance is deemed acceptable will be integrated into the AI models. This integration aims to refine the models' decision-making processes by tailoring them to operate optimally under known favorable conditions and to identify scenarios where human intervention is necessary.
   Insights from the correlation matrix can drive improvements in the underlying AI algorithms. For example, if visibility negatively affects certain links, computer vision models can be enhanced to handle low-visibility conditions better.

2. Optimizing Operational Environments: Taking insights from how different environmental conditions (e.g., temperature, wind speed, visibility) impact AI performance, systems can be optimized to function under varying conditions. For instance, algorithms can be

retrained or fine-tuned to account for specific variables that significantly affect their performance. Real-time monitoring can be implemented to dynamically adjust AI system parameters in response to changing environmental factors, ensuring consistent performance.

3. Enhancement of Model Reliability through Uncertainty Estimation: To improve the reliability and interpretability of AI models, advanced uncertainty estimation techniques such as Bayesian Inference and Deep Ensembles can be employed. These methods will allow for more robust predictions by quantifying the confidence levels of the AI models, thereby reducing the risk of errors in critical decision-making processes. This will not only enhance model performance but also increase trust in AI systems operating in complex environments.

These research directions will contribute to the development of more accurate, reliable, and interpretable AI systems capable of operating effectively under diverse conditions.

Appendix

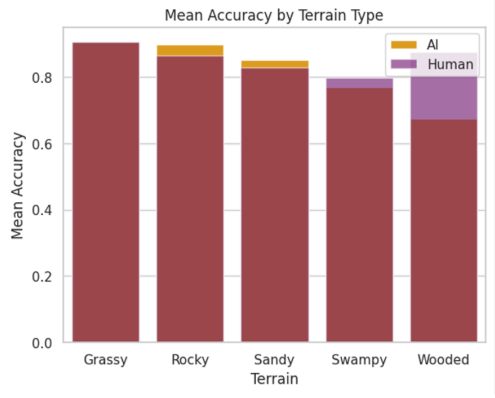

Figure 1(a). Mean Accuracy by Terrain Type

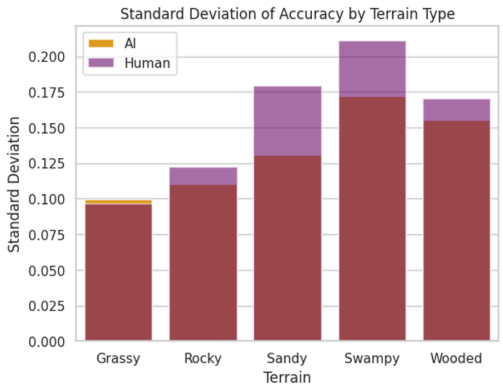

Figure 1(b). Standard Deviation of Accuracy by Terrain Type

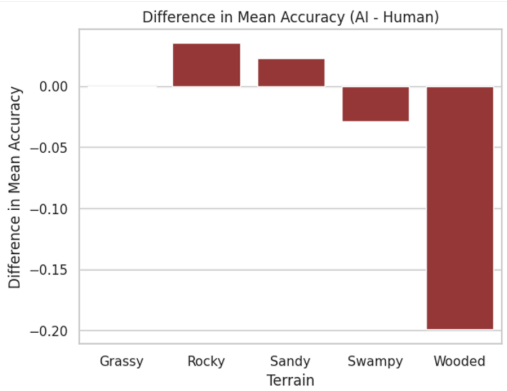

Figure 1(c). Difference in Mean Accuracy

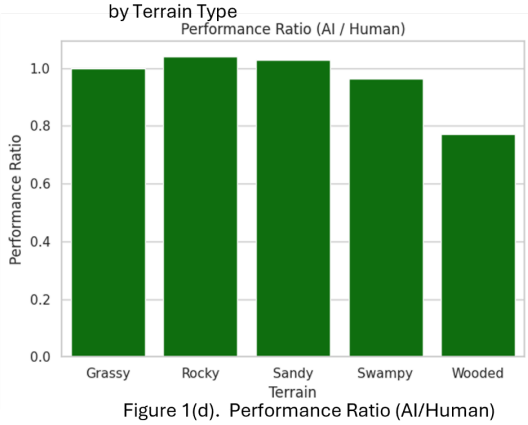

Figure 1(d). Performance Ratio (AI/Human)

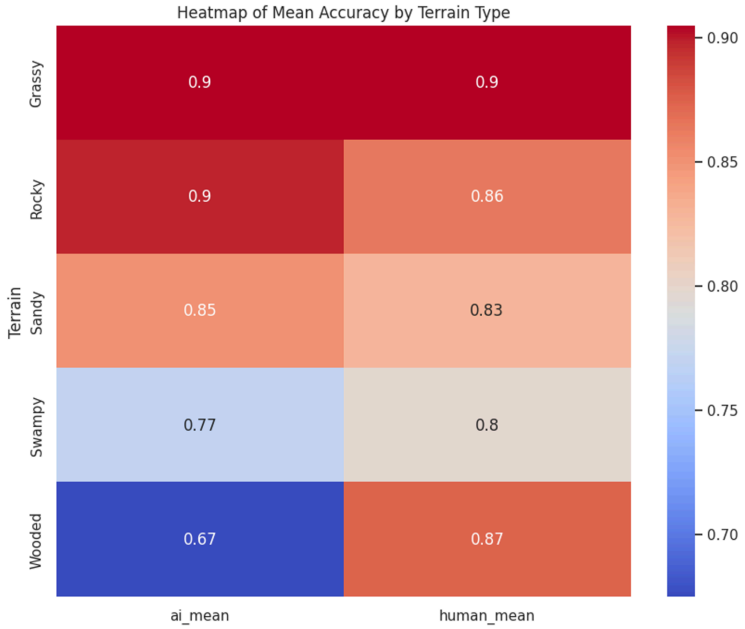

Figure 2. Heatmap of Mean Accuracy by Terrain Type

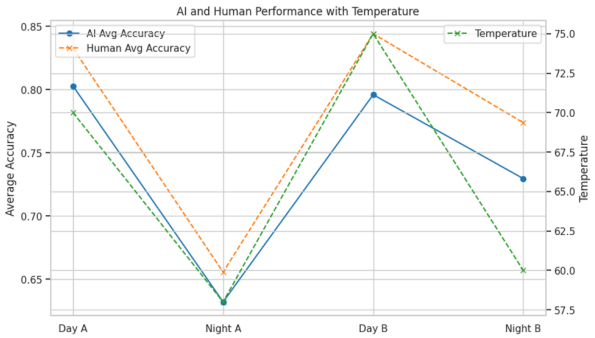

Figure 3(a). AI and Human performance with temperature

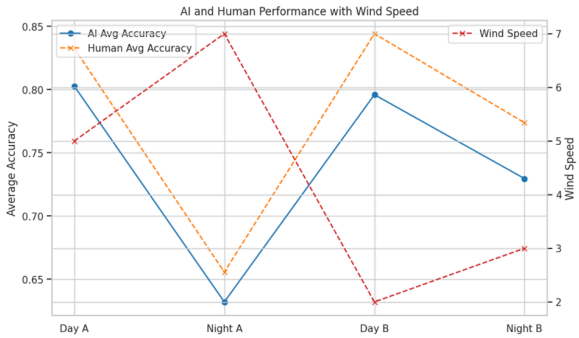

Figure 3(b). AI and Human performance with Wind Speed

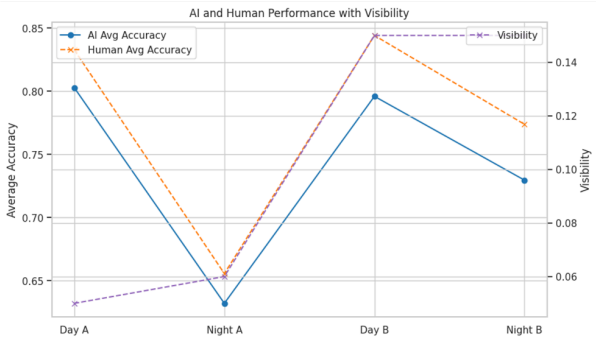

Figure 3(c). AI and Human performance with Visibility

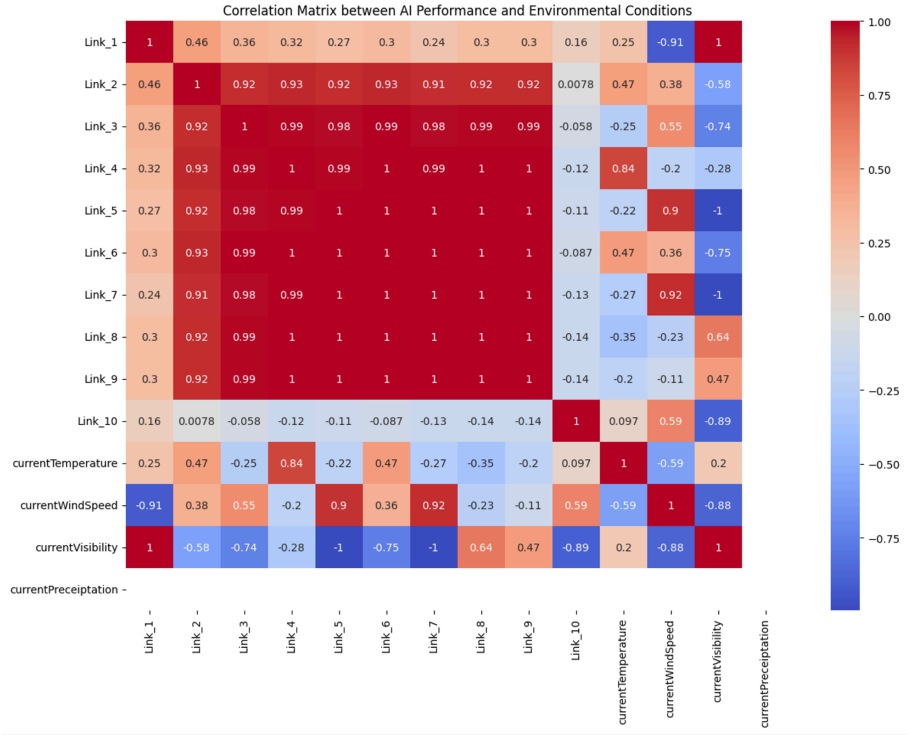

Figure 4. Correlation Matrix Between AI Performance and Environmental Conditions

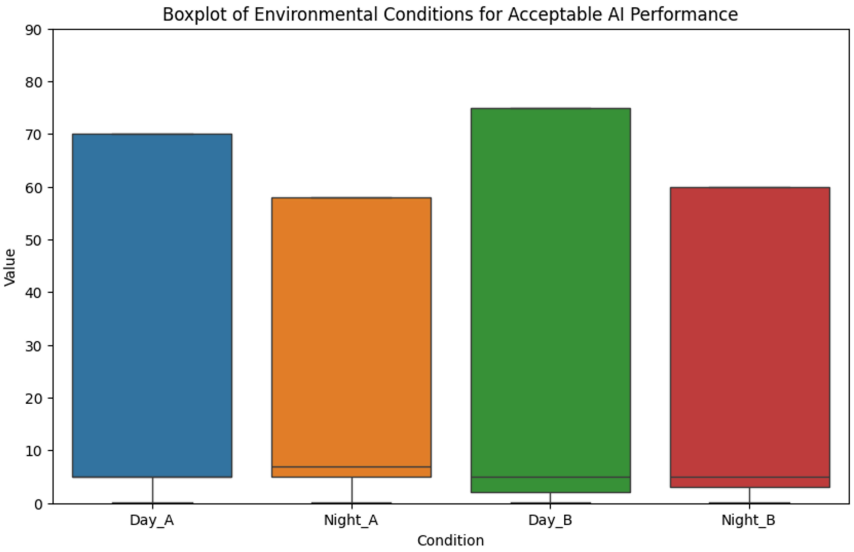

Figure 5. Boxplot for Environmental Conditions for Acceptable AI Performance