WGU C951

Task 3

MACHINE LEARNING PROJECT PROPOSAL

Cole Linke

Student ID #011917160

8/30/2025

A. Project Overview

My project proposition is to implement an image recognition system for the Minnesota Department of Natural Resources (DNR). The machine learning (ML) model will analyze crowd-sourced photos, submitted by park visitors, to monitor the condition of park infrastructure, detect the presence of flora and fauna, and track overall environmental health trends. Photos will fall into two categories. The first includes images from designated monitoring stations, which provide consistent data on specific points of interest over time. The second includes images taken freely throughout the park and submitted with their geolocation data.

A.1. Organizational Need

The DNR relies on staff and rangers to manually inspect parks, trails, and rivers for signs of infrastructure damage, wildlife activity, and environmental changes. This time-consuming process is limited by staff availability. By implementing a crowd-sourced image recognition system, the DNR could collect data and allocate staff more efficiently and effectively.

A.2. Context and Background

The DNR is dedicated to protecting and managing land, water, wildlife, and natural habitat (Minnesota Department of Natural Resources, n.d.). Though the Minnesota Department of Natural Resources is made up of many divisions, this proposal focuses on the Parks and Trails division. This division oversees 64 state parks, comprised of over 200,000 acres of land, 1500 miles of developed trails, and 3000 public water accesses (Minnesota Department of Natural Resources, n.d.). The manual approach to maintaining these lands, while possible, limits the DNR's efficacy and scope of monitoring. This proposal will free up much-needed staffing to proactively address issues as they develop and aid in environmental data collection.

Examples:

- A photo submitted by a hiker of a trail deep in the woods reveals the spread of an invasive plant species. The resource management team has not yet reviewed this

site, but because of the submission, the DNR can send a crew to remove the invasive species before it becomes a larger problem.

- Photos taken at a designated monitoring station show a gradual decline in vegetation over the course of a year. The image recognition system flags the trend and notifies the DNR, who tests the soil for possible contaminants. Subtle environmental changes like this can be hard to notice with the human eye, especially when monitoring is infrequent or done by different people.
- Multiple photos submitted in a single month reveal the presence of an endangered species in an area where it was previously undocumented. The DNR is notified and begins conservation efforts to protect the newly identified population.

A.3. Outside Works Review

1. "The iNaturalist Species Classification and Detection Dataset" by Grant Van Horn

   This study presents a large-scale dataset designed to train machine learning models for species classification. The dataset, from the iNaturalist platform, includes more than 800,000 photos representing over 5,000 species. iNaturalist is a free platform where users share observations of plants, animals, and other living things by uploading photos. The model is trained by citizen scientists who validate and verify the specimens to ensure quality and train the model. Researchers from the study used convolutional neural networks (CNNs) to test the dataset, demonstrating its effectiveness for training models to classify species. (Van Horn, et al., 2018)

2. "Identifying Animal Species in Camera Trap Images Using Deep Learning and Citizen Science" by Marco Willi

   This study evaluates the effectiveness of CNNs for differentiating between camera trap images containing different animal species and empty images (images with no animals) using data from the Zooniverse platform. Zooniverse is an online platform that utilizes citizen scientists as volunteers to help researchers with mass classification projects. The models achieved 91-98%

accuracy for detecting empty images and 88-92% accuracy for identifying specific species. The research also showed that using transfer-learning (copying weights learned on a base model to a target model) improved accuracy on smaller datasets compared to training from scratch. In addition, filtering out low-confidence predictions also increased accuracy. (Willi, et al., 2019)

3. "Evaluating the Use of Automated Plant Identification Tools in Biodiversity Monitoring" by Manuel R. Popp

   This study conducted a comprehensive field survey, evaluating publicly available automated plant identification tools, at 151 sites in Switzerland. The research found that 85% of species were accurately identified if multiple images were supplied, and over 90% of species were accurately identified at least once. They also found that using a narrower geographic scope reached higher identification accuracy. (Popp, Zimmermann, & Brun, 2025)

These three studies demonstrate how machine learning, citizen science, and automated tools can be combined to improve environmental monitoring. The first study shows how machine learning models can be trained with large datasets to return accurate results for identifying plants and animals. The second study shows that using transfer-learning is more efficient than building models from scratch, especially when working with smaller datasets. The third study additionally highlights that region-specific models can show even greater accuracy. For the Minnesota Department of Natural Resources, these findings mean less time spent on manual inspections and more time on conservation efforts.

A.4. Solution Summary

My proposed solution is to use a machine learning image recognition system trained on large, tagged datasets of plants, animals, and environmental features from Minnesota. The model will use convolutional neural networks to classify crowd-sourced photos that are submitted by park visitors. Photos taken at designated monitoring stations will allow the DNR to track specific locations over time, while freely submitted photos with geolocation data will further expand monitoring coverage. This system will provide the

DNR with ecological data and reduce the time spent on manual inspections, allowing staff to focus more on solving problems.

A.5. Machine Learning Benefits

One benefit of my proposed machine learning system is that by relying on crowd-sourced photos, the DNR can more broadly and consistently monitor its parks and trails. Minnesota state parks receive over 11 million visitors annually, which, if utilized correctly, could be overseen by more than staff alone (Minnesota Department of Natural Resources, n.d.). By using these submissions, the system shifts staff time from finding issues to solving them. It also helps identify trends by creating a long-term digital environmental record.

B. Machine Learning Project Design

B.1. Scope

In Scope:

- Develop a machine learning image recognition model using convolutional neural networks
- Train the model on tagged datasets of Minnesota flora, fauna, and environmental features like trail conditions, signs of overgrowth, water levels, and litter
- Building and installation of monitoring stations (small mounts where visitors can place phones to take photos of fixed areas with consistent points of view)
- Create a system for accepting, storing, and processing crowd-sourced photos from visitors
- Generate reports of environmental trends based on classified images

Out of Scope:

- Creation of a public-facing mobile application that identifies species or provides feedback to users
- Real-time monitoring or emergency response

B.2. Goals, Objectives, and Deliverables

Goals

- Improve the DNR's ability to monitor parks and trails by reducing reliance on manual inspections
- Improve the DNR's ability to track and predict environmental changes over time
- Support conservation efforts by providing early detection of invasive species
- Support infrastructure maintenance by providing early detection of wear and tear and overgrowth of paths
- Create a system that takes advantage of the significant amount of visitors by using crowd-sourced photos

Objectives

- Achieve at least 85% classification accuracy for flora and fauna species with the model
- Achieve at least 85% identification accuracy for overgrowth, trail damage, excess litter, and water levels
- Collect and process a sufficient number of tagged images to train the model in the pilot year
    - 50,000 – 100,000 photos or .5 – 1% of the annual visitor population
    - Tag using citizen science platforms like Zooniverse

Deliverables

- A trained CNN model for classifying flora, fauna, and environmental features

- A cloud-based system for the collection, storage, and archiving of crowd-sourced photos
- Install monitoring station mounts at designated locations
- An interactive dashboard for the DNR to visualize and interpret processed data
- Documentation and training for staff on system use and maintenance

B.3. Standard Methodology

Business Understanding:

The goal is to improve the Minnesota Department of Natural Resources' ability to monitor parks and trails. This is achieved by utilizing crowd-sourced photos from park visitors and training a machine learning model to detect species and environmental conditions. Achieving this goal will shift the energy spent on manual inspection to addressing the problems that are found.

Data Understanding:

The data for this project will come entirely from crowd-sourced photos submitted by park visitors in two variations. Monitoring station submissions will provide photos of the same location over time. This creates a consistent record of environmental change in a specific, predetermined spot. Free submissions, with geolocation data, will give a broader coverage of trails, rivers, and areas that are not regularly monitored. Using crowd-sourced photos will collect data during different seasons and in a wide range of locations.

Data Preparation:

The collected photos will need to be organized before training the model. To create usable training data, photos will be tagged through the citizen science platform Zooniverse, where volunteers help classify and label species or environmental conditions in the images. Once tagged, the photos will be integrated into a structured database. The dataset will include separate sections for flora, fauna, and environmental features that the model can learn from.

Modeling:

The prepared and tagged dataset will be used to train a convolutional neural network to classify flora, fauna, and environmental features in park photos. The model will be trained by introducing the data in subsets. This technique will allow other subsets of newly introduced data to measure accuracy, test validation, and prevent overfitting. Once the model is trained, it will be able to analyze new crowd-sourced photos automatically.

Evaluation:

The performance of the trained model will be evaluated using a separate subset of data that was not used during training. The model will be tested on classification accuracy for flora, fauna, and environmental features, as well as identifying empty images. Low-confidence predictions will be filtered out to reduce false positives, and any misclassifications will be reviewed and retested to refine the model.

Deployment:

Once the model has been trained and evaluated, it will be deployed to automatically process new crowd-sourced photos submitted from park visitors. The system will use a cloud-based database to store, organize, and track images and their classifications. The DNR staff will be provided with an interactive dashboard that will display processed data, potential issues, and trends. Maintenance will include verifying that new data is collected and classified correctly.

B.4. Projected Timeline

01/01/2026 – 01/09/2026:	Project proposal and approval of goals, objectives, and deliverables
01/12/2026 – 01/23/2026:	Build, install, and test physical monitoring stations in parks
01/26/2026 – 02/06/2026:	Set up cloud system for storing data
02/09/2026 – 02/13/2026:	Create a project on Zooniverse and recruit volunteers

02/16/2026 – 03/06/2026: Build data pipeline and connect to Zooniverse

03/09/2026 – 03/12/2027: Collect and tag 50,000 – 100,000 photos

03/15/2027 – 04/16/2027: Train and test CNN model

04/19/2027 – 05/07/2027: Design and build an interactive dashboard for staff

05/10/2027 – 06/04/2027: Final integration and testing of model, database, and dashboard

06/07/2027 – 06/18/2027: Staff training on use of new system

06/21/2027 – 07/02/2027: Project finalization and closeout

Sprint Schedule

Sprint	Start	End	Tasks
1	01/01/2026	01/09/2026	Business Understanding: • Draft and approve project proposal • Confirm goals, objectives, and deliverables
2	01/12/2026	02/13/2026	Data Understanding: • Build, install, and test monitoring stations • Set up cloud system for storing data • Create project on Zooniverse and recruit volunteers

3	02/16/2026	03/12/2027	Data Preparation: • Build data pipeline and connect to Zooniverse • Collect and tag 50,000-100,000 photos from visitors • Organize and structure datasets for training
4	03/15/2027	04/16/2027	Modeling & Evaluation: • Train and test CNN model with tagged datasets
5	04/19/2027	07/02/2027	Deployment: • Design and build interactive dashboard for staff • Final integration and testing of model, database, and dashboard • Staff training • Project finalization

B.5. Resources and Costs

Resource	Description	Cost

Development Staff	Two full-time developers (salary) 2 x $80,000	$160,000
Additional Staff	Field staff, QA testers, additional developers, and field scientists (contract-based) $40,000	$40,000
Monitoring stations	Building and installation (two at each state park) 2 x 64 x $100	$12,800
Cloud services	Cloud storage (monthly cost for one year) 12 x $500	$6,000
Zooniverse Platform	Project created on Zooniverse (free project creation & volunteer-driven)	$0
Data	Photos collected from visitors (free & crowd-sourced)	$0
Basic Server	Model training (two servers) 2 x $5,000	$10,000
DNR Staff Training	Train staff on system use taught by project staff (two weeks & included in staff salary)	$0
	Total	$228,800

B.6. Evaluation Criteria

Objective	Success Criteria

Classification Accuracy (flora & fauna)	Model correctly classifies flora and fauna with at least 85% accuracy
Identification Accuracy (overgrowth, damage, litter, water levels)	Model correctly identifies environmental conditions like trail overgrowth or damage, water levels, and excess litter with at least 85% accuracy
Data Collection	Collection and tagging of 50,000 – 100,000 photos in the six-month data preparation period
Trend Detection	Model correctly identifies verifiable environmental trends (The model flags projected trends, which are then monitored over time to confirm accuracy)
Overall Effectiveness	Compare projects identified and completed before and after implementation

C. Machine Learning Solution Design

C.1. Hypothesis

If a convolutional neural network is trained on a large, tagged dataset of Minnesota flora, fauna, and environmental features, then it will be able to classify crowd-sourced park photos with at least 85% accuracy. This hypothesis will be tested by validating the model with a separate, unseen subset of data and comparing its predictions against photos tagged by experts.

C.2. Selected Algorithm

The algorithm I will implement is a convolutional neural network, which is a supervised learning algorithm. In supervised learning, the model is taught by introducing labeled data, and it must determine a method for arriving at the correct conclusions. Once trained, the CNN can automatically classify new, unseen images and make predictions based on learned patterns. (Wakefield, n.d.)

C.2.a. Algorithm Justification

I chose a CNN because it excels at pattern recognition in images, which supports the goal of my proposal. Its ability to process large amounts of data and automatically classify new photos accurately makes it ideal for identifying species and environmental conditions.

C.2.a.i. Algorithm Advantage

One advantage of supervised learning models is that they can achieve high accuracy in image classification when trained on reliable, labeled data. It is also straightforward to evaluate accuracy because the method uses known outcomes to test the model. (Coursera Staff, 2024)

C.2.a.ii. Algorithm Limitation

One limitation of supervised learning models is that they can overfit the training data. Overfitting occurs when the model becomes too tailored to its training examples and fails to predict or classify new, unseen data. (Belcic & Stryker, n.d.)

C.3. Tools and Environment

The primary programming language will be Python, as it is considered one of the best for machine learning because of its ease of use, libraries, and framework support. TensorFlow, an open-source framework used for image classification, will be used to build and train the convolutional neural network model (Geeks for Geeks, 2025). A physical GPU server running Linux will be used to train the model, while cloud storage will be used to collect and organize photos. The initial training data will be tagged on Zooniverse, a citizen science platform, with the help of volunteers.

C.4. Performance Measurement

The model's performance will be measured by its accuracy in classifying Minnesota flora, fauna, and environmental conditions. The dataset will be divided into training, validation, and testing subsets to establish a baseline accuracy and prevent overfitting. The main metric will be classification accuracy, with a minimum success rate of at least 85%.

D. Description of Data Sets

D.1. Data Source

The data will come entirely from crowd-sourced photos submitted by park visitors. Volunteers from the Zooniverse platform will classify and label the submitted photos to create a complete dataset for training the machine learning model.

D.2. Data Collection Method

Although all photos will be collected from park visitors, there will be two variants of submissions. The first type, monitoring station submissions, will provide consistent data on specific points throughout the parks. The second type, free submissions, allows visitors to submit photos taken anywhere throughout the park, as long as they include geolocation data. Both types of photos will be uploaded to a cloud-based system for storage.

D.2.a.i. Data Collection Method Advantage

This method allows the DNR to gather a large and diverse collection of photos for its training dataset. Even if only a small percentage of annual visitors submit photos, the total could reach tens to hundreds of thousands, which is far more than the staff alone could collect. The variety in perspective and conditions from amateur photos will also help with the model's effectiveness.

D.2.a.ii. Data Collection Method Limitation

One limitation of crowd-sourced photos is that they will vary in quality. This will require additional effort to tag or discard unusable photos before training. Another limitation is that the stream of photos might not be consistent through different seasons, potentially resulting in disproportionate or biased data.

D.3. Quality and Completeness of Data

Before training, collected photos will be processed through the Zoonivers platform. Volunteers will label each image to identify flora, fauna, and environmental conditions. Because the proposal uses a public platform to process data, all photos that reveal personally identifiable information will be removed before processing. Unusable, duplicate, or photos missing required data, such as location or monitoring station number,

will also be excluded from the dataset. Outliers, such as species not from the target region or unexpected environmental conditions, will be manually reviewed to maintain quality. Finally, any photos submitted outside the collection timeline will not be accepted.

D.4. Precautions for Sensitive Data

Because the DNR will manage the project, it will follow government standards for handling and securing data. Volunteers tasked with tagging photos on Zooniverse are required to register for an account, which adds an additional layer of accountability. All data will be backed up regularly in secure cloud storage and on the physical training servers. Access to the complete dataset will be limited to authorized staff and necessary contract workers. The collected data itself is not highly sensitive because all personally identifiable information is removed, leaving only photos of plants, animals, and environmental conditions on public property.

References

Belcic, I., & Stryker, C. (n.d.). *What Is Supervised Learning*. Retrieved from IBM: https://www.ibm.com/think/topics/supervised-learning

Coursera Staff. (2024, October 4). *Supervised vs. Unsupervised Learning: Pros, Cons, and When to Choose*. Retrieved from Coursera: https://www.coursera.org/articles/supervised-vs-unsupervised-learning

Geeks for Geeks. (2025, August 6). *10 Best Language for Machine Learning*. Retrieved from Geeks for Geeks: https://www.geeksforgeeks.org/machine-learning/best-language-for-machine-learning/

Minnesota Department of Natural Resources. (n.d.). *About the DNR*. Retrieved from Minnesota Department of Natural Resources: https://www.dnr.state.mn.us/aboutdnr/index.html

Minnesota Department of Natural Resources. (n.d.). *Minnesota State Parks and Trails*. Retrieved from Minnesota Department of Natural Resources: https://www.dnr.state.mn.us/parks_trails/index.html

Minnesota Department of Natural Resources. (n.d.). *State Parks*. Retrieved from Minnesota Department of Natural Resources: https://www.dnr.state.mn.us/faq/mnfacts/state_parks.html

Popp, M., Zimmermann, N., & Brun, P. (2025). Evaluating the use of automated plant identification tools in biodiversity monitoring—a case study in Switzerland. *Ecological Informatics*, 103316.

Van Horn, G., Mac Aodha, O., Song, Y., Cui, Y., Sun, C., Shepard, A., . . . Belongie, S. (2018). The

iNaturalist Species Classification and Detection Dataset. *Proceedings of the IEEE conference on computer vision and pattern recognition*, (pp. 8769-8778).

Wakefield, K. (n.d.). *A Guide to the Types of Machine Learning Algorithms*. Retrieved from SAS: Data and AI Solutions: https://www.sas.com/en_gb/insights/articles/analytics/machine-learning-algorithms.html

Willi, M., Pitman, R., Cardoso, A., Locke, C., Swanson, C., Boyer, A., . . . Fortson, L. (2019). Identifying animal species in camera trap images using deep learning and citizen science. *Methods in Ecology and Evolution*, 80-91.