Iowa Initiative for Artificial Intelligence

Final Report

Project title:	Enhancing Agricultural Safety: Predictive Analysis of Injury and Fatality Factors
Principal Investigator:	Ibrahim Demir
Prepared by:	Omer Mermer, Ibrahim Demir (UIHILab) and Yanan Liu (IIAI)
Other investigators:	Omer Mermer, Charles Jennissen
Date:	07/31/2025

Were specific aims fulfilled:	Yes
Readiness for extramural proposal?	Yes
If yes … Planned submission date	Submitted October 2024
Funding agency	NIH- NIOSH R21
Grant mechanism	Research grant
If no … Why not? What went wrong?	NA

Brief Summary of Accomplished Results:

Research Report:

Aims (provided by PI):

Agricultural work carries inherent injury risks due to hazardous equipment, physically demanding conditions, and exposure to environmental hazards. The primary aim of this project is to improve agricultural safety by building a predictive framework that can assess the severity of injuries and identify actionable preventive strategies. To achieve this, we leverage advanced machine learning (ML) models and Explainable Artificial Intelligence (XAI) tools. These approaches enable both predictive accuracy and transparent insights into the factors most associated with severe injuries in agricultural settings. The project is structured around the following specific aims:

Aim 1: Data Collection and Dataset Preparation. Develop a comprehensive dataset of agricultural injury cases by sourcing and curating incident-level data from AgInjuryNews.com, a publicly available platform aggregating media-reported injuries and fatalities across the United States.

Aim 2: Machine Learning (ML) Model Development and Comparison. Train and evaluate a variety of ML models, including Random Forest, AdaBoost, Gradient Boosting, and XGBoost, to classify injury severity. Performance will be measured using accuracy, precision, recall, and AUC-ROC to identify the most effective model(s).

Aim 3: Explainable AI (XAI) for Model Interpretability. Apply SHAP (SHapley Additive exPlanations) to interpret model outputs and identify the most influential factors in injury severity prediction. This interpretability enables stakeholders to make informed decisions based on model results. By enabling stakeholders to understand which factors contribute most to injury risk, this aim builds trust and supports data-driven decision-making, allowing model insights to be effectively translated into safety measures.

Data:

The dataset used in this study was obtained from AgInjuryNews.com, an open-access surveillance platform that aggregates news media reports of agricultural injuries and fatalities occurring across the

United States and Canada. Covering a time span from 2015 to 2024, AgInjuryNews collects incident-level data that includes a wide range of variables such as geographic location, demographic information, accident characteristics, and the presence or absence of safety equipment at the time of the incident. These records are identified through automated keyword filtering and digital media monitoring, after which expert volunteers manually review and code the information into a structured format.

To prepare the data for analysis, the raw dataset underwent extensive preprocessing, including removal duplicates, outliers and records with missing values. The resulting dataset comprised 2,472 valid records of agricultural injury incidents. Of these, 791 cases (32%) were classified as non-fatal, and 1,681 cases (68%) were identified as fatal. The binary target variable in this study, injury severity, was coded as 0 for fatal incidents (in which at least one person died) and 1 for non-fatal incidents (which included cases involving minor injuries or property damage only).

The modeling framework employed in this study utilized a comprehensive set of features that were grouped into four major categories: temporal attributes, personal characteristics, accident attributes, and roadway/environmental factors. These explanatory variables are summarized in Table 1 and include both categorical and continuous types, carefully selected based on their relevance to injury risk and severity. Temporal attributes included the season of occurrence (spring, summer, autumn, or winter), whether the incident occurred on a weekday or weekend, the time of day (morning, afternoon, evening, or night), and the calendar month. These variables allowed for the exploration of patterns related to seasonal work, time-of-day effects, and potential influences of agricultural cycles.

Personal characteristics comprise the victim's gender, age, alcohol consumption, and the use of safety equipment such as seatbelts, helmets, and other personal protective equipment (PPE). Additional binary indicators like drowning and the presence of rollover protective structures (ROPS) were also included. These factors provide insight into the role of human behavior and safety compliance in injury outcomes. Accident attributes captured information about the injury agent involved in the incident—categorized into types such as machinery, livestock, ATVs, or environmental hazards—as well as the total number of victims, the individual's role (e.g., operator, bystander, passenger), and whether the incident was intentional. Geographic state was included to account for potential regional variations in practices or regulations.

Environmental and roadway features were also incorporated, including whether the incident occurred in a confined space, whether grain was involved, and whether the location was part of an agri-tourism activity. These variables reflect the physical context of the incident and its potential contribution to severity. The outcome variable, injury severity, was binarized for supervised machine learning classification.

Table 1. Details of all variables in the dataset

Variables	**Variable Category**	**Features**	**Type**	**Description/Labeling**
Independent Variables	Temporal Attributes	Season	Categorical	1: Spring (Mar-May); 2: Summer (Jun-Aug); 3: Autumn (Sep-Nov); 4: Winter (Dec-Feb)
		Day	Binomial	0: Weekday (Monday to Friday); 1: Weekend (Saturday and Sunday)
		Time	Categorical	1: Morning (6am – 12pm); 2: Afternoon (12pm – 6pm); 3: Evening (6pm – 12am); 4: Night (12am – 6am)

		Month	Categorical	[1,12]
	Personal Features	Gender	Binomial	0: Male; 1: Female
		Age	Continuous	[0, 98]
		Alcohol	Binomial	0: No; 1: Yes
		Seatbelt	Binomial	0: No; 1: Yes
		Helmet	Binomial	0: No; 1: Yes
		Drowning	Binomial	0: No; 1: Yes
		ROPS	Binomial	0: No; 1: Yes
		Other PPE	Binomial	0: No; 1: Yes
	Accident	Injury Agent	Categorical	1: ATV/Off-Road Vehicle; 2: Building; 3: Environment; 4: Fall; 5: Fishing/Forestry; 6: Livestock; 7: Machinery; 8: Pesticides/Plants; 9: Vehicle
		Number of Victims	Continuous	[1, 17]
		Location	Categorical	1: Roadways; 2: Agricultural Env.; 3: Forestry; 4: Fishing
		Role	Categorical	0: Bystander; 1: Certified First Responder; 2: Operator; 3: Other; 4: Passenger; 5: Worker/Farmer/Fisher
		Intentional	Binomial	0: No; 1: Yes
		State	Categorical	[0, 49]
	Roadway	Confined Space	Binomial	0: No; 1: Yes
		Grain Involved	Binomial	0: No; 1: Yes
		Agri-tourism	Binomial	0: No; 1: Yes
Dependent Variable		Injury Severity	Binomial	0: Fatal; 1: Non-Fatal

Table 2 provides the frequency and percentage distribution of injury severity categories (non-fatal and fatal) across different years from 2016 to 2024, highlighting a consistent pattern in the predominance of fatal injuries and a gradual decline in overall incidents over time. Out of a total of 2,472 incidents, 791 (32%) were classified as non-fatal injuries and 1,681 (68%) as fatal injuries. Fatal injuries represented the majority each year, with the highest recorded in 2018—260 fatal cases, accounting for 69% of that year's total. Starting in 2019, both fatal and non-fatal injuries began to decrease, with a notable increase in the proportion of non-fatal injuries in 2020, reaching 34% despite a drop in overall incidents. This upward shift in non-fatal cases continued modestly through 2023, where non-fatal injuries accounted for 36% of that year's incidents, the highest proportion across the dataset. By 2024, there is a sharp drop in total reported cases, with only 12 non-fatal (32%) and 25 fatal (68%) injuries recorded, suggesting either an actual decline in incidents or partial data collection for the year. Overall, the data indicates a long-term trend toward reduced injury frequency and a slight improvement in non-fatal outcomes in recent years.

Table 2. Frequency and percentage distribution of injury severity categories over the years.

Year	Injury Severity Category	Frequency	Percent (%)
2016	Non-Fatal	109	32%
	Fatal	234	68%
2017	Non-Fatal	111	34%
	Fatal	220	66%
2018	Non-Fatal	117	31%
	Fatal	260	69%
2019	Non-Fatal	100	29%
	Fatal	242	71%
2020	Non-Fatal	106	34%
	Fatal	204	66%
2021	Non-Fatal	75	31%
	Fatal	170	69%
2022	Non-Fatal	79	30%
	Fatal	181	70%
2023	Non-Fatal	82	36%
	Fatal	145	64%
2024	Non-Fatal	12	32%
	Fatal	25	68%

Collectively, the dataset represents a rich and multidimensional resource for modeling injury severity in agriculture. The diversity of variables allows for the application of advanced machine learning models that not only provide accurate predictions but also support explainability through methods like SHAP. The combination of structured incident-level data and contextual factors offers a unique opportunity to generate actionable insights for policymakers, safety professionals, and researchers in agricultural health and safety.

AI/ML Approach:

Machine learning techniques, particularly ensemble methods, have gained significant traction in predictive modeling due to their ability to enhance accuracy, handle complex datasets, and reduce model variance and bias. This section outlines several widely used ensemble algorithms—ranging from boosting-based approaches to bagging methods, highlighting their mechanisms and advantages in classification task relevant to predictive analytics.

Adaptive Boosting (AdaBoost):

AdaBoost is one of the earliest and most influential boosting ensemble methods. It operates on the principle that a strong classifier can be created by sequentially combining multiple weak classifiers, each refining the accuracy of the overall prediction. The algorithm adjusts the weights of training samples at every iteration, giving more focus to misclassified instances. This leads to an iterative refinement process that continues until a certain classification error threshold is reached, helping to reduce bias in complex prediction scenarios. Unlike some ensemble techniques, AdaBoost is relatively robust to both overfitting and underfitting in various classification tasks. It improves model performance by dynamically updating the sample distribution and selecting only the most effective classifiers, discarding weaker ones. This approach results in a cumulative increase in predictive accuracy and has proven useful in areas such as accident severity prediction.

Extreme Gradient Boosting (XGBoost):

XGBoost is an advanced ensemble learning algorithm built on decision trees and designed for efficiency and flexibility in supervised learning tasks. It implements a refined form of gradient boosting, where multiple weak learners are combined iteratively into a single, powerful model. Each iteration focuses on correcting the errors of the previous learner, using gradient-based optimization guided by an objective function. The algorithm includes regularization to control overfitting and improve generalization, and it supports feature importance ranking, which is useful in high-dimensional datasets. XGBoost also employs random sampling to reduce variance and uses both first- and second-order derivatives of the loss function for precise gradient calculations. It supports parallel and distributed computing, enabling faster training times and scalability to large datasets. XGBoost has been applied effectively in various domains, including injury severity and accident analysis.

Light gradient boosting machine (LightGBM):

LightGBM is a highly efficient implementation of gradient boosting decision trees designed to overcome common issues such as high computational cost and long training times. It is suitable for both classification and regression and introduces innovative techniques like Exclusive Feature Bundling (EFB) and Gradient-based One-Side Sampling (GOSS) to improve scalability and performance. LightGBM uses a histogram-based learning algorithm and a leaf-wise tree growth strategy with depth constraints, which boosts accuracy while reducing the risk of overfitting. It also natively supports categorical features, eliminating the need for one-hot encoding. LightGBM's design makes it ideal for large datasets and complex prediction tasks, including real-world applications like road traffic injury analysis.

Histogram gradient boosting (HistGBRT):

HistGBRT is a variant of gradient boosting that focuses on improving training speed and memory efficiency. It follows the standard boosting framework by iteratively training weak learners to correct the mistakes of their predecessors. A key distinction of HistGBRT is its use of histogram-based techniques, where continuous features are discretized into bins, significantly accelerating computation. This method reduces training time and memory usage by storing values in histograms instead of processing continuous data directly. HistGBRT offers a practical balance of accuracy and efficiency, especially for large-scale datasets.

Random Forest (RF):

Random Forest is a popular ensemble learning technique that aggregates the output of multiple decision trees to improve predictive performance. It is effective for both regression and classification, especially in high-dimensional data environments. By building each tree using a random subset of features and data samples, Random Forest minimizes overfitting and ensures robustness to noise and outliers. The algorithm also provides useful metrics like feature importance, helping to interpret the model's decisions. Its scalability and versatility have made it a valuable tool in numerous domains, including crash severity prediction.

Gradient Boosting (GB):

Gradient Boosting is an ensemble method that builds models sequentially, where each new model corrects the residual errors of the previous one. It uses gradient descent to minimize a specified loss function, making it highly effective in improving predictive accuracy. The approach allows models to focus on the most challenging data points and incorporates regularization techniques—such as limiting tree depth and using a learning rate—to reduce overfitting and enhance model generalization. Gradient Boosting has demonstrated strong performance on complex, non-linear datasets, often outperforming traditional models like linear regression and standalone decision trees in both classification and regression tasks.

Experimental methods, validation approach:

Figure 1 shows a comprehensive workflow diagram for ML agricultural injury severity prediction. Initial step involves obtaining and preparing data from AgInjuryNews. After applying the preprocessing, organized data are split into training and test dataset, where the chosen ML models are trained using the training dataset, and test dataset are employed to observe the behavior of the trained ML models. Since our dataset contains many more samples than features, we were careful not to reduce the number of features too much to avoid losing important information. This strategy helps us balance capturing a wide range of influencing factors and maintaining the robustness of our model. The min-max method (MinMaxScaler) was applied to standardize the features to eliminate potential model bias due to difference in scales and units. The min-max method transforms the data into a range 0 and 1 and ensures that all features contribute equally to the model by bringing them to a common scale, which helps improve the performance of machine learning algorithms.

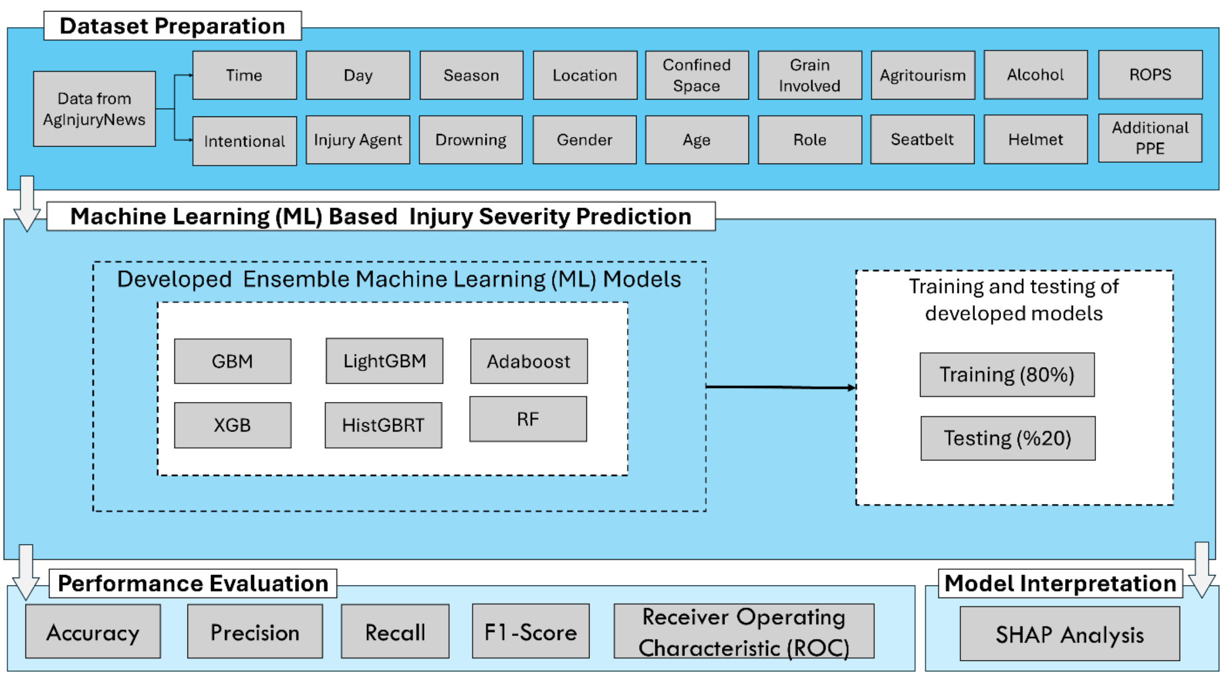


Figure 1. Workflow diagram for all scenarios in agricultural injury severity prediction.

Model Evaluation:

When comparing classification models, various performance metrics derived from a confusion matrix are commonly used. Table 3 presents the confusion matrix for a binary classifier. This matrix, structured as a contingency table, illustrates how observations are distributed across actual and predicted classes. In classification problems, the confusion matrix comprises four possible outcomes: true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), as shown in Table 3. It provides a clear representation of both correct and incorrect classifications under a specified target, allowing for the calculation of accuracy and other performance metrics.

Table 3. Confusion Matrix for Binary Classification

Actual Class	Predicted Class	
	Positive	**Negative**
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

In this study, the binary confusion matrix for each machine learning model was utilized to compute various quantitative performance metrics. The metrics used for evaluating model performance are outlined and explained below. Accuracy, expressed by Equation 1, represents the proportion of correctly classified samples out of the total number of samples. Precision (equation 2) assesses the alignment between data labels and the positive labels predicted by the classifier. Recall, or sensitivity (equation 3), indicates how effectively the classifier identifies positive labels. Finally, the F1-score (equation 4) is the harmonic mean of recall and precision, providing a balanced measure of the that considers both false positives and false negatives in a classification model.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad \text{Eq. 1}$$

$$Precision = \frac{TP}{TP + FP} \quad \text{Eq. 2}$$

$$Recall = \frac{TP}{TP + FN} \quad \text{Eq. 3}$$

$$F1 - Score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad \text{Eq. 4}$$

Explainability Analysis with SHAP:

As machine learning models become increasingly complex and accurate, the need for interpretability is more critical than ever, especially in safety-critical areas like traffic accident prediction. Among the various explainable AI (XAI) techniques, SHAP (SHapley Additive exPlanations) has become a leading method for interpreting model predictions. Based on principles from cooperative game theory, SHAP assigns each input feature a contribution value toward a model's output using Shapley values. It offers a consistent, model-agnostic approach that can explain predictions from any machine learning model, including complex ensemble and deep learning systems.

One of SHAP's key strengths lies in its ability to quantify both the size and direction of a feature's impact on an individual prediction. This dual capability supports both global and local interpretability. Globally, SHAP generates summary plots that rank features by their overall importance across the entire dataset, helping analysts identify the most influential variables. Locally, SHAP provides visualizations—such as force plots and decision plots—that show how specific feature values contributed to an individual prediction relative to a baseline, offering insights into why a particular outcome was predicted.

SHAP is especially valuable when applied to high-performing ensemble models like XGBoost, LightGBM, and AdaBoost, which often sacrifice transparency for predictive power. By uncovering non-linear interactions, revealing hidden dependencies, and highlighting threshold effects, SHAP enhances the interpretability of these models. This makes it particularly effective for analyzing agricultural injury severity, where the relationships among driver behavior, vehicle characteristics, environmental conditions, and road factors are often complex.

In this study, SHAP is used to assess feature importance and interaction effects within ensemble models designed to predict agricultural injury severity. It provides both broad policy-level insights—such as identifying key safety factors—and detailed, case-level explanations to support operational decision-making. This layered interpretability bridges the gap between prediction accuracy and practical application, offering valuable insights to first responders, transportation planners, and policymakers. By enabling a deeper understanding at both population-wide and individual crash levels, SHAP enhances the diagnostic and prescriptive utility of machine learning models.

Results:

Figure 2 presents the annual distribution of injury severity categories—fatal and non-fatal—between 2016 and 2024. The data reveal that fatal injuries consistently outnumber non-fatal ones across all observed years. The frequency of incidents peaked in 2018, with fatal cases reaching their highest recorded level during that year. Starting in 2019, there is a noticeable decline in both fatal and non-fatal injuries. Notably, in 2020, the proportion of non-fatal injuries increased to 33% of that year's total incidents, despite an overall decrease in reported cases. This trend of declining injury frequency continues through 2023 and into 2024. While this downward trend may suggest improvements in safety practices or interventions, it may also be partially attributed to reporting lags or incomplete data for more recent periods.

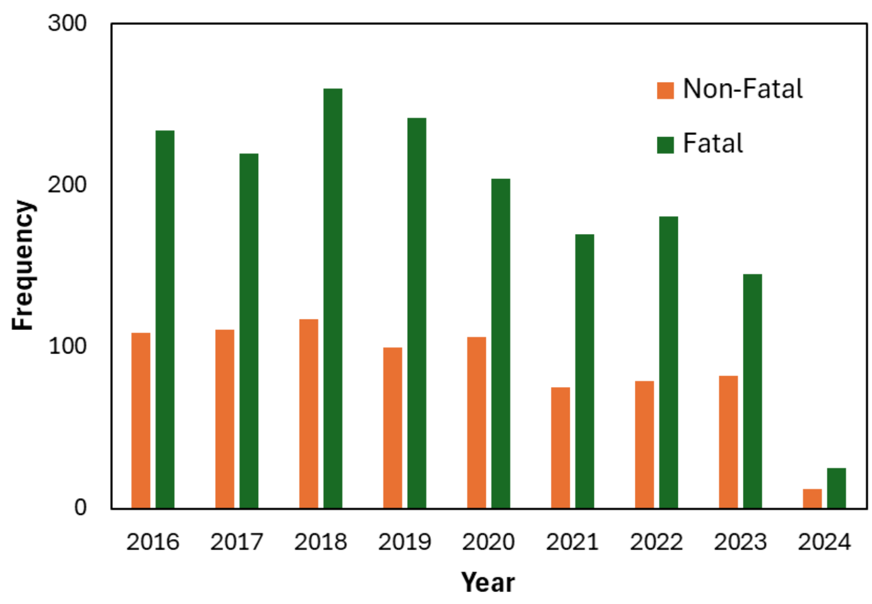


Figure 2: Year-wise distribution of agricultural injuries by severity from 2016 to 2024.

Table 4 presents the confusion matrices for all machine learning (ML) methods tested in this study. Across the models, there is a consistent trend of high correct classification rates for fatal cases compared to non-fatal cases. For example, Random Forest (RF) achieved 1,574 correct fatal predictions versus 154 correct non-fatal predictions, while Gradient Boosting (GB) and XGBoost (XGB) displayed similar patterns. This imbalance suggests that the models were more adept at identifying fatal injuries, likely due to the higher representation of fatal cases in the dataset (68% of total). While LightGBM and HistGBRT exhibited slightly higher correct non-fatal classifications than other models, these gains came at the expense of reduced fatal classification accuracy, hinting at a trade-off between sensitivity to the minority class and overall accuracy.

Table 4. Confusion Matrix for All ML Methods

Model	Actual Class	Predicted Class	
		Non-Fatal	Fatal
GB	Non-Fatal	213	565
	Fatal	137	1536
XGB	Non-Fatal	183	595
	Fatal	120	1553
RF	Non-Fatal	154	624
	Fatal	99	1574
Adaboost	Non-Fatal	216	562
	Fatal	152	1521
LightGBM	Non-Fatal	265	513
	Fatal	268	1405
HistGBRT	Non-Fatal	261	517
	Fatal	253	1420

Table 5 summarizes the performance metrics for each ML method, highlighting their comparative strengths. Most models achieved an accuracy of approximately 0.71, with F1-scores clustering around 0.81, indicating balanced performance between precision and recall. Random Forest and XGBoost demonstrated the highest recall values (0.94 and 0.93, respectively), reflecting their strong ability to detect non-fatal cases. Conversely, LightGBM recorded the lowest overall accuracy (0.68) and F1-score (0.78), suggesting challenges in achieving balanced predictions across classes. Despite these differences, the close metric values across most models underscore the robustness of the ensemble approaches in predicting injury severity within this dataset.

Table 5. Performance Metrics for all Linear ML Models

Model	Accuracy	Precision	Recall	F1-Score
GB	0.71	0.73	0.92	0.81
XGB	0.71	0.72	0.93	0.81
RF	0.71	0.72	0.94	0.81
Adaboost	0.71	0.73	0.91	0.81
LightGBM	0.68	0.73	0.84	0.78
HistGBRT	0.69	0.73	0.85	0.79

Figure 3 shows the Receiver Operating Characteristic (ROC) curves for all evaluated ML models, offering a visual comparison of their classification performance. The curves for Gradient Boosting, XGBoost, Random Forest, and AdaBoost exhibit a high area under the curve (AUC), reflecting strong discriminative ability between fatal and non-fatal cases. While LightGBM and HistGBRT showed slightly lower AUC scores, their curves still indicated acceptable performance above random chance. The overall proximity of the curves suggests that all tested models effectively leveraged the dataset's feature space to differentiate between severity classes, with marginal differences in predictive capacity.

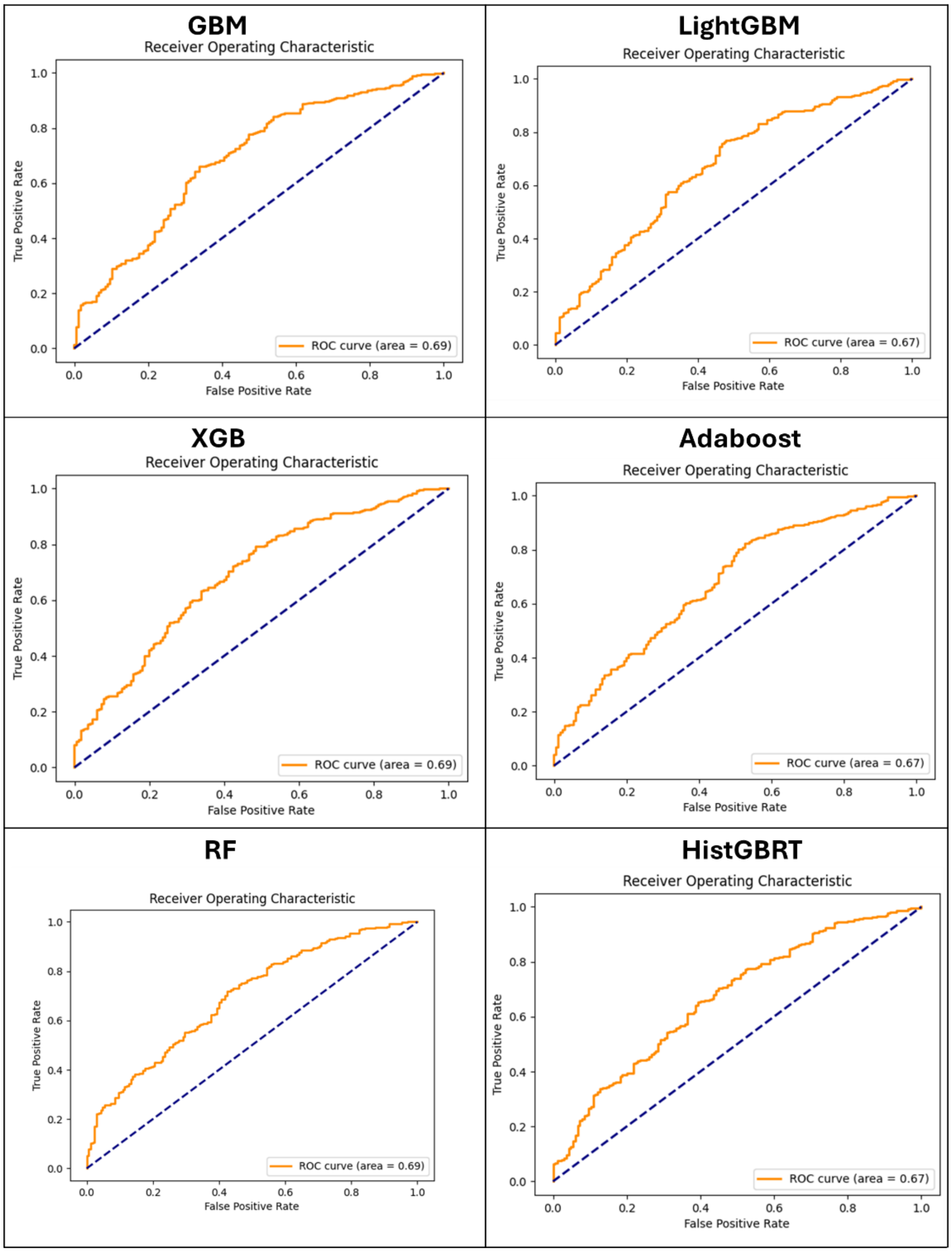
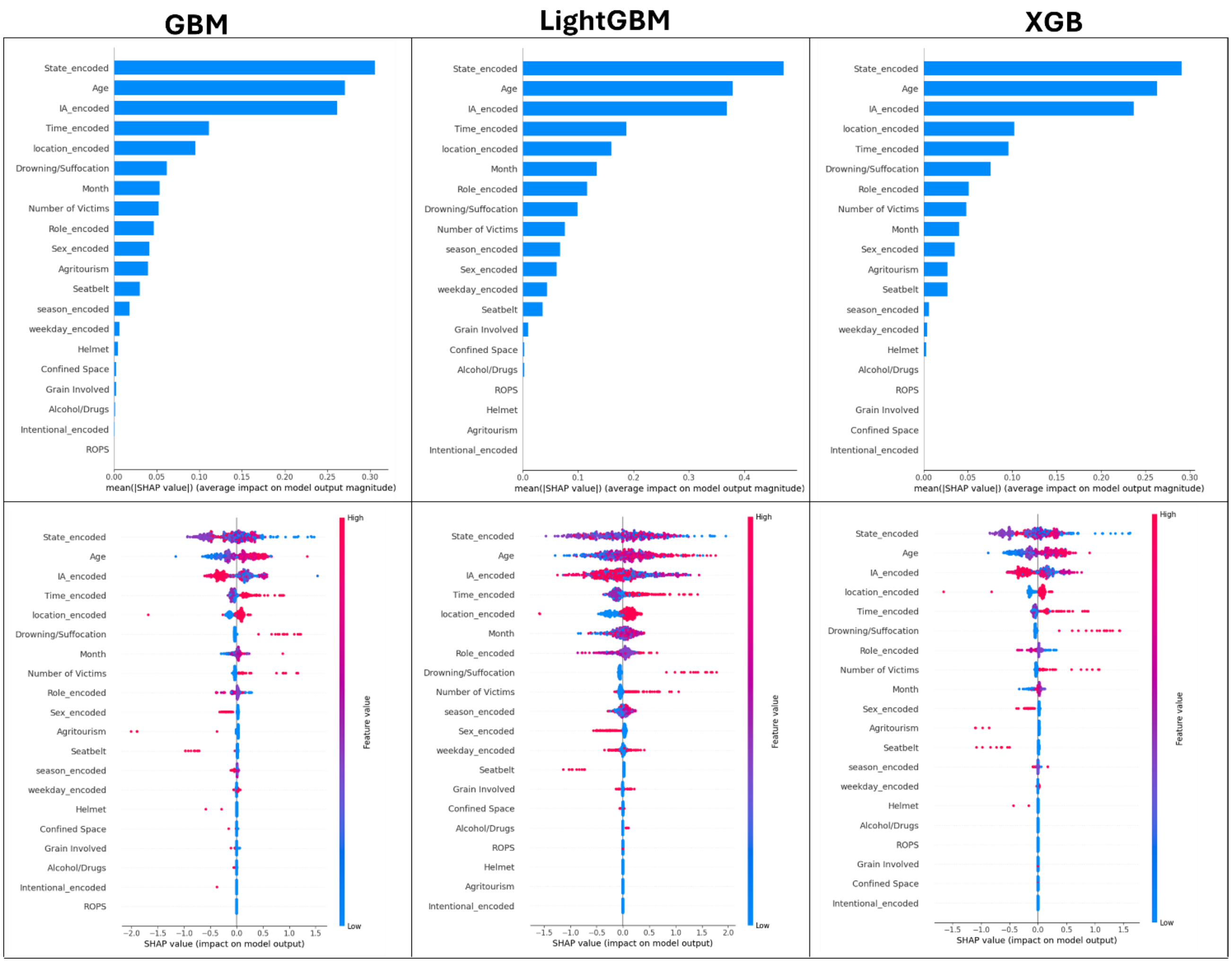


Figure 3: The ROC curves generated from ML models for injury severity classification based on data we used.

Figure 4 provides the global SHAP interpretation results, with the left panel ranking features by their mean absolute SHAP values and the right panel offering a summary plot of their impact. Key features such as the presence of rollover protective structures (ROPS), helmet use, age, and injury agent type emerged as the most influential in determining injury severity. Positive SHAP values (shown in red) indicate a higher likelihood of predicting a non-fatal outcome, while negative values (blue) are associated with fatal outcomes. The summary plot reveals clear threshold and interaction effects—for instance, higher age values generally increased fatality risk, while safety equipment presence strongly shifted predictions toward non-fatal outcomes. These insights not only validate domain knowledge but also offer practical targets for safety interventions in agricultural settings.



Cont.

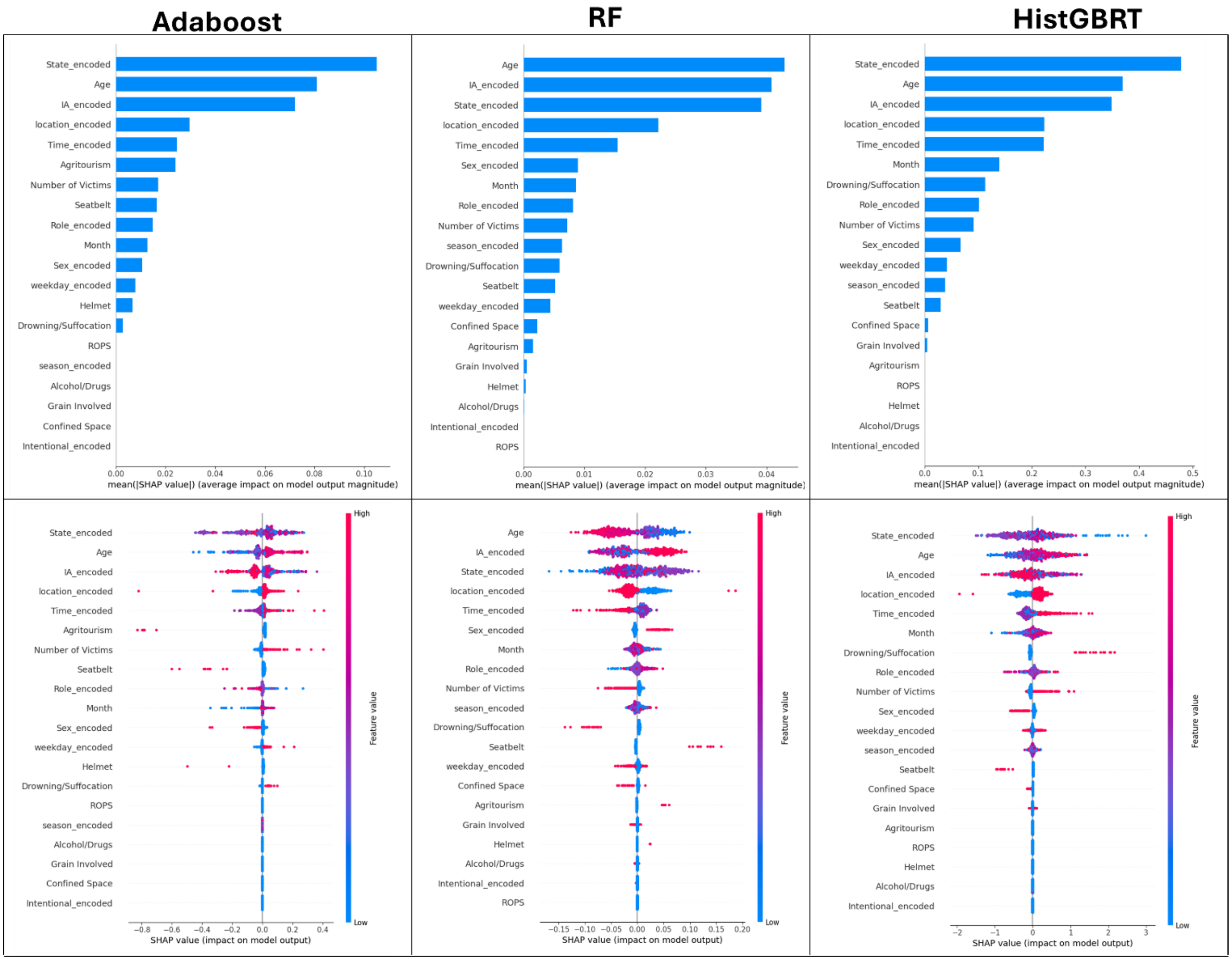


Figure 4: Global interpretation utilizing SHAP based on features importance (upper panel) and summary plot (lower panel).

Figure 5 presents the SHAP local interpretability plots for individual predictions generated by the different ML models used in this study. Each plot highlights the most influential features for a specific prediction, showing both the magnitude and direction of their impact. Features in red push the model's output toward predicting a non-fatal outcome, while those in blue push it toward a fatal outcome. Across models, the *state* variable consistently emerges as a dominant factor, often exerting a strong negative influence on predicted survival likelihood. Other recurring influential features include location type, role, age, and injury agent, though their impact varies in magnitude and sign between models and individual cases. These local explanations reveal how the same feature can have opposite effects depending on the contextual combination of other variables, illustrating the complex interactions captured by ensemble methods. This level of interpretability is valuable for case-specific analysis, enabling stakeholders to understand the reasoning behind individual severity classifications and potentially tailor interventions to specific high-risk scenarios.

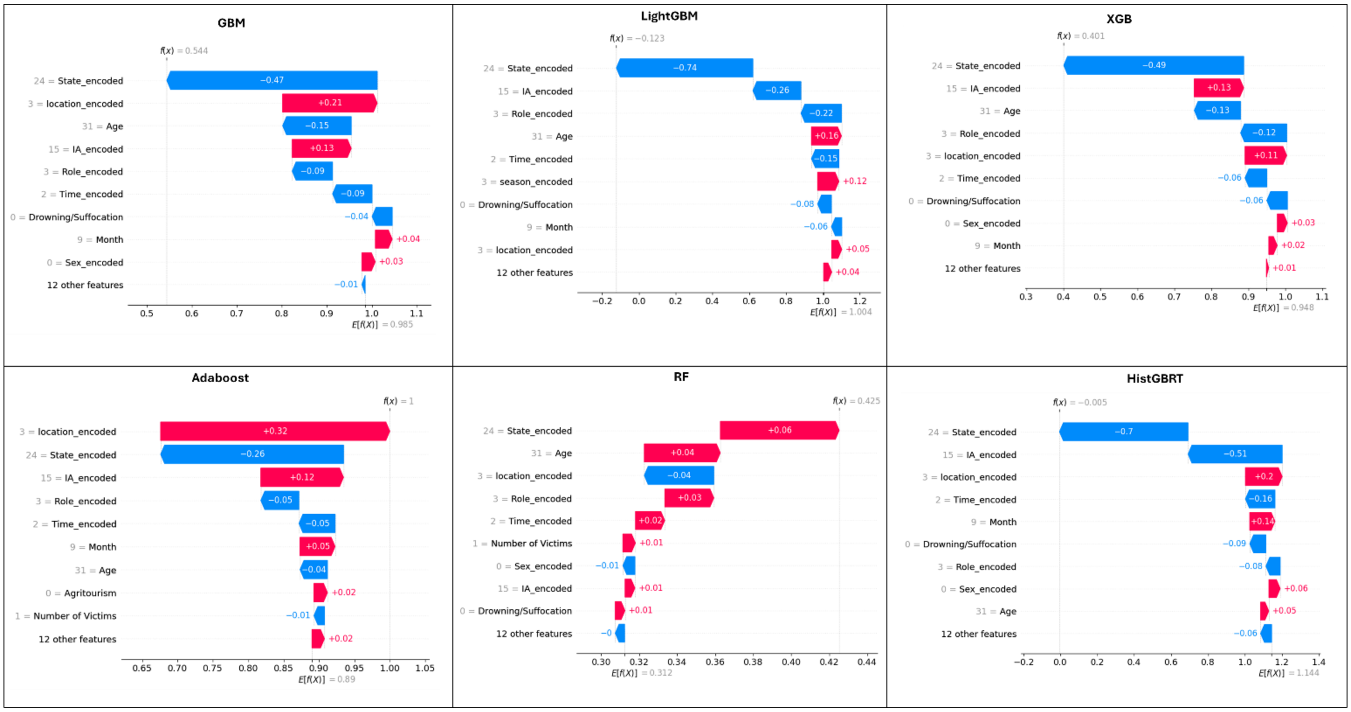


Figure 5: SHAP local interpretability plots for different ML models we use.

Ideas/Aims for Future Extramural Project:

In a future extramural project, the ultimate goal is to is to enhance agricultural safety by developing a predictive framework that not only assesses injury severity based on diverse risk factors but also provides actionable insights into preventive measures. By employing advanced deep learning (DL) techniques and Explainable AI (XAI), this project will deliver a comprehensive analysis of the key risk factors contributing to injuries in agricultural settings. These insights will guide safety protocols and targeted interventions, ultimately working towards a safer agricultural work environment. This approach is particularly vital in an industry where data availability and interpretability can greatly influence the adoption of safety measures. The results from this IIAI pilot project can be extended in developing the above-mentioned approaches. The following specific aims shall be addressed: 1) comparative Study in Agricultural Safety Using Advanced DL Models; 2) Integration of Explainable AI for Interpretability and Stakeholder Engagement; 3) Practical Application for Data-Driven Safety Interventions in Agriculture.

Publications Resulting from Project:

O. Mermer, Y. Liu, I. Demir, “Predicting agricultural injury severity using ensemble machine learning: global versus local explainability using SHAP”, International Journal of Injury Control and Safety Promotion, 2025 (in preparation)