applied sciences

MDPI

Article

Explainable Artificial Intelligence to Support Work Safety in Forestry: Insights from Two Large Datasets, Open Challenges, and Future Work

Ferdinand Hoenigsberger [1,†], Anna Saranti [1,†], Anahid Jalali [1,2], Karl Stampfer [1] and Andreas Holzinger [1,*]

[1] Human-Centered AI Lab, Institute of Forest Engineering, Department of Forest and Soil Sciences, University of Natural Resources and Life Sciences, 1190 Vienna, Austria; ferdinand.hoenigsberger@boku.ac.at (F.H.); anna.saranti@boku.ac.at (A.S.); anahid.jalali@boku.ac.at (A.J.); karl.stampfer@boku.ac.at (K.S.)
[2] Austrian Institute of Technology, 1210 Vienna, Austria
* Correspondence: andreas.holzinger@boku.ac.at
† These authors contributed equally to this work.

Abstract: Forestry work, which is considered one of the most demanding and dangerous professions in the world, is claiming more and more lives. In a country as small as Austria, more than 50 forestry workers are killed in accidents every year, and the number is increasing rapidly. This serves as a catalyst for us to implement more stringent measures for workplace safety in order to achieve the sustainability objective of SDG 3, which focuses on health and well-being. This study contributes to the analysis of occupational accidents and focuses on two large real-world datasets from both the Austrian Federal Forests (ÖBf) and the Austrian Workers' Compensation Board (AUVA). Decision trees, random forests, and fully connected neural networks are used for the analysis. By exploring different interpretation methods, this study sheds light on the decision-making processes ranging from basic association to causal inference and emphasizes the importance of causal inference in providing actionable insights for accident prevention. This paper contributes to the topic of explainable AI, specifically in its application to occupational safety in forestry. As a result, it introduces novel aspects to decision support systems in this application domain.

Keywords: explainable AI; occupational accidents; forestry; work safety

Citation: Hoenigsberger, F.; Saranti, A.; Jalali, A.; Stampfer, K.; Holzinger, A. Explainable Artificial Intelligence to Support Work Safety in Forestry: Insights from Two Large Datasets, Open Challenges, and Future Work. *Appl. Sci.* **2024**, *14*, 3911. https://doi.org/10.3390/app14093911

Academic Editor: Chilukuri K. Mohan

Received: 29 February 2024
Revised: 26 April 2024
Accepted: 1 May 2024
Published: 3 May 2024

1. Introduction

Forest operations, encompassing various activities related to harvesting, logging, and managing woodland areas, particularly motor-manual felling, represent a domain fraught with inherent risks and challenges, causing a lot of yearly deaths [1–3].

This field has become increasingly perilous, especially under the duress of escalating extreme weather conditions, which compound the inherent dangers of working in such unpredictable environments. Moreover, the intensifying pressure to meet logging demands further exacerbates these risks, leading to a worrying trend in the frequency and severity of forest accidents. These developments necessitate innovative approaches to enhance safety and operational efficiency. In this context, artificial intelligence (AI) emerges as a promising avenue for helping to mitigate the risks associated with forest operations. Particularly, explainable AI (XAI) offers a huge transformative potential in this sector. By its nature, XAI provides transparent and understandable AI decisions and predictions, which is crucial in high-stakes environments like forestry operations.

This paper effectively illustrates the effectiveness of integrating various analytical methods to examine and make sense of a detailed and multifaceted dataset, aiming to improve occupational safety within the forestry industry with a focus on explainable AI (XAI). It interlaces classical descriptive statistics, advanced machine learning algorithms, and probabilistic models to deliver an exhaustive analysis that reveals the complex patterns of occupational accidents. The research methodically progresses from establishing a foundation using descriptive statistics to applying machine learning for more profound insights and utilizing probabilistic models to navigate through data constraints, with each technique significantly contributing to the comprehensive understanding and forecasting of accidents. This thorough examination not only pinpoints the primary risk factors but also proposes practical steps for reducing these hazards, thereby aiding in the enhancement of safer workplace conditions in the forestry sector. Moreover, the manuscript emphasizes the importance of cross-disciplinary approaches in exploiting data for occupational safety and underscores the crucial role of informed, data-driven decision-making in promoting a more secure and well-informed environment for forestry work.

The research presented in this paper extends the preliminary findings of a conference paper [4] and explores new insights based on Actionable Explainable AI [5], thereby exploring novel dimensions and unearthing deeper insights into forest operation safety.

However, it is important to acknowledge that despite the advancements presented in this paper, numerous open research areas remain unaddressed. The complexity of forest operations and the multifaceted nature of the risks involved mean that this paper can only scratch the surface of potential AI applications in this domain. Future research is required to fully realize the potential of AI, particularly XAI, in enhancing safety and operational efficiency in forest operations. This paper aims to provide a foundation for such future explorations, offering insights and directions that could shape subsequent research efforts in this vital field.

2. Background and Related Work

The meticulous documentation and subsequent analysis of work accidents are crucial for gaining insights into the factors contributing to occupational hazards and legal compliance. Reports of incidents in hazardous work environments can be analyzed to identify common hazards, which will help prevent future accidents. While some studies do this, most focus on causes involving physical risks. Hinze et al. (2021) [6] proposed an alternative approach and shed light on the causes of accidents in forestry from the perspective of employee failure and fatigue. It is interesting to note that 70% of incidents are attributed, at least in part, to worker negligence. Secondly, 78% of the causes attributable to worker error show signs of fatigue. This suggests that a significant number of forestry accidents are due to worker fatigue.

A wide array of studies have been dedicated to examining different aspects of accident records. Tsioras et al. (2014) [7], for example, analyzed timber harvesting accidents in Austria. This line of research extends to other countries, with notable studies in Poland [8], Slovakia [9,10], Italy [11], and Brazil [12]. Additionally, cross-national comparative research has been conducted [13]. Notably, a considerable segment of this research is centered on understanding accidents involving chainsaw use or motor-manual tree felling [14–16].

The adoption of machine learning (ML) in accident prediction is chiefly attributed to two key factors: its explanatory capacity, which aids in identifying primary causes of accidents [17], and its predictive ability, enhancing accident prevention efforts [18]. This methodology is increasingly used in various high-risk professions to analyze occupational accidents. For instance, ML has been applied to study serious accidents in the chemical industry [19] and the steel industry [20]. In the field of electrical engineering, Oyedele et al. (2021) [21] explored the use of deep learning and boosted tree algorithms to evaluate their effectiveness in safety risk management. Further examples of ML's application in accident prediction can be found in studies such as [22–25].

In the following sections, we introduce two datasets, followed by a series of experiments to uncover the explanatory variables influencing forest accidents. We conclude by outlining future research directions.

3. Dataset 1: Austrian Federal Forests

3.1. Description

Data from the 2021 statistics of accidents during forestry work in Austria, as reported by ÖBf (2021), indicated that out of 1189 recorded accidents, 21 tragically resulted in fatalities, in only one year! This evidence firmly positions forestry work among the most hazardous professions across all production sectors.

The first dataset we analyzed as part of this study contained the records of occupational accidents reported by the Austrian Federal Forests from the years 2005 to 2021. The dataset incorporated information on fatal and non-fatal incidents. In non-fatal cases, individuals may be unable to work for a specified period or on sick leave, measured in hours or days. The columns deemed most descriptive, based on domain knowledge and insights from previous models [7,19] primarily included the time of day (in minutes), the age of the individual, and the day of the week (with one corresponding to Monday) when the accident occurred. Additionally, the dataset distinguished between workers and employees, as severe accidents are more likely among workers. Details on the accident's cause, the sector of employment, and the injured body part were also provided, each encoded appropriately. While potential causal relationships existed between these variables, this research focused solely on exploring their linear and non-linear correlations.

The dataset underwent exploration and analysis utilizing Python libraries, primarily `pandas` [26,27]. Preprocessing tasks were predominantly carried out with `scikit-learn` [28], while visualization was performed using `matplotlib` [29] and `Bokeh` [30].

3.2. Preprocessing

Between 1 October 2005 and 21 December 2021, a total of 2481 registered accidents were recorded. Among these, only nine resulted in fatalities, which was not sufficient to provide significant insights into the factors influencing these occurrences. However, classifying between non-fatal and fatal accidents remains crucial and could benefit from additional data. A dedicated preprocessing phase involved eliminating unfilled entries and rows with invalid content. In the final analysis, a set of 7 input features was used: "age at accident", "worker or employee", "day of the week", "time in the day", "injured body part", "accident cause", and "working sector". After removing invalid entries (such as empty strings or out-of-range values like working sector 51, where valid values range from 1 to 35), 1965 rows containing valid data remained.

Before examining predictive modeling, histograms were constructed to visually depict the distributions of the input features. It is important to note that the domain cardinality of each input feature significantly influences its relevance. Features, whether continuous or categorical with numerous categories, tend to provide more descriptive information compared to those with fewer categories [31,32]. To address this, certain input features like "injured body part", "working sector", and "accident cause" underwent value grouping. For example, accidents impacting the head and neck area were grouped together.

Prior to employing the features in predictive models, we conducted assessments to evaluate both linear and non-linear correlations between all pairs of features. Linear correlations were quantified using the Pearson coefficient [33], while non-linear correlations were measured using mutual information (MI) [34]. Among these, the highest linear correlation was observed between the working sector and the cause of the accident (0.26), followed by the correlation between the working sector and worker or employee (−0.26), and then between the cause and body part (−0.13). Correspondingly, the MI values for these pairs were 0.25, 0.19, and 0.06, respectively. It is noteworthy that all other pairs exhibited MI values below 0.1 [4]. These findings are consistent with both domain knowledge and previous research. Interestingly, no significant correlations were detected among any pairs, thereby eliminating the necessity to streamline the model by removing redundant input features or applying dimensionality reduction techniques before training [33,35,36].

3.3. Predictive Models

3.3.1. Regression Task

Decision Tree

In the context of non-fatal accidents, the primary predictive objective is to anticipate the duration of recovery for the injured individual, which varies between workers and employees. This distinction arises from the distinct nature of the recovery process for each group, influenced by the nature of their respective tasks. As depicted in Figure 1, the distributions of recovery duration for workers and employees exhibit noticeable differences, aligning with expectations due to the higher incidence of severe accidents in manual labor settings. The used approach involved developing a unified model capable of discerning both shared patterns and unique characteristics between these two groups. However, it is important to remain cognizant of the potential need to construct separate, more distinct models tailored to each group's specific attributes in future iterations.

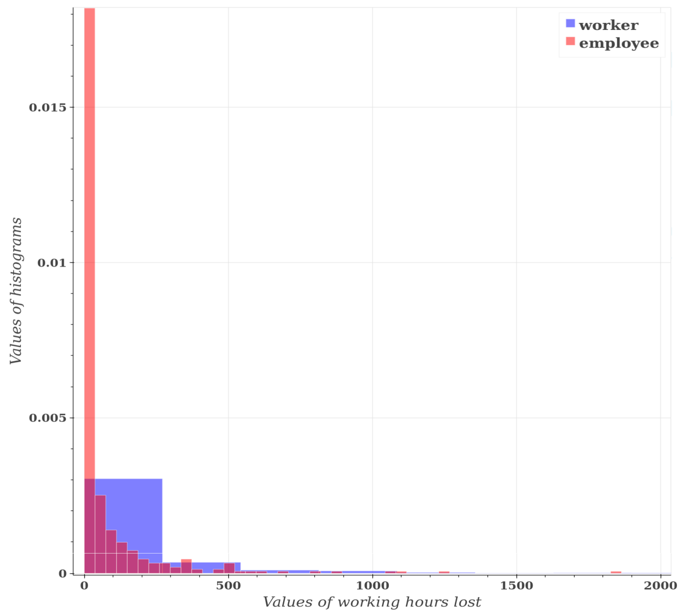


Figure 1. Working hours lost for workers (blue-collar workers) and employees (white-collar workers), reprinted with permission from [4], Springer 2022.

This problem constitutes a regression task [37], and employing a decision tree (DT) [33] is a straightforward and inherently interpretable approach for such a task. Given that the numeric or continuous features did not require scaling, attention shifted to non-ordinal input features such as injured body parts, cause of the accident, and the working sector. Some features already possessed integer mappings, but in most instances, these mappings lacked an ordinal relationship, so it was necessary to prevent the predictive model from assuming such a relationship in its decision-making or generalization. Moreover, as mentioned in the preceding Section 3.2, a grouping of conceptually similar values was applied to these features, consolidating them into single categories. To encode these categories, one-hot encoding was employed [36], where each category is represented by a vector with only one entry orthogonal to the other. However, this approach introduces challenges; the resulting feature values matrix tends to be sparse and large due to the presence of a column for each category.

The analysis of the decision tree (see Figure 2) revealed that younger employees and workers were anticipated to require a relatively short recovery period, depending upon the time of the accident. Additionally, distinctions emerged between accidents occurring during the start of the week and those happening later in the week.

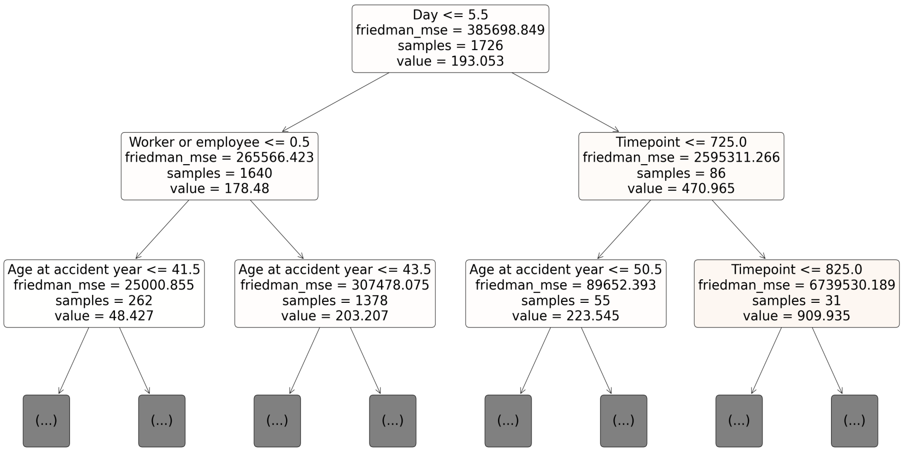


Figure 2. The acquired rules of the decision tree, which was trained using 4 features; reprinted with permission from [4], Springer 2022.

The dataset underwent shuffling and division into two sets: 80% designated as the training set and 20% for the test set. A grid search was conducted on the training set to determine the optimal parameters for the decision tree. These parameters were:

- Maximum depth: five
- Maximum number of features to consider: "auto" (suggesting that the best split was determined by considering all available input features)
- Minimum number of samples required for a split: two

After evaluation on the test set utilizing four features—specifically, the individual's age at the time of the accident, the day of the week, the time of day, and the classification as a worker or an employee—the decision tree showed a Mean Squared Error (MSE) of 151,222.99 and a Relative Root Mean Squared Error (RRMSE) of 1.15 [4]. The significance of these input features is consistent with previous research findings [7,19], as illustrated in Figure 3.

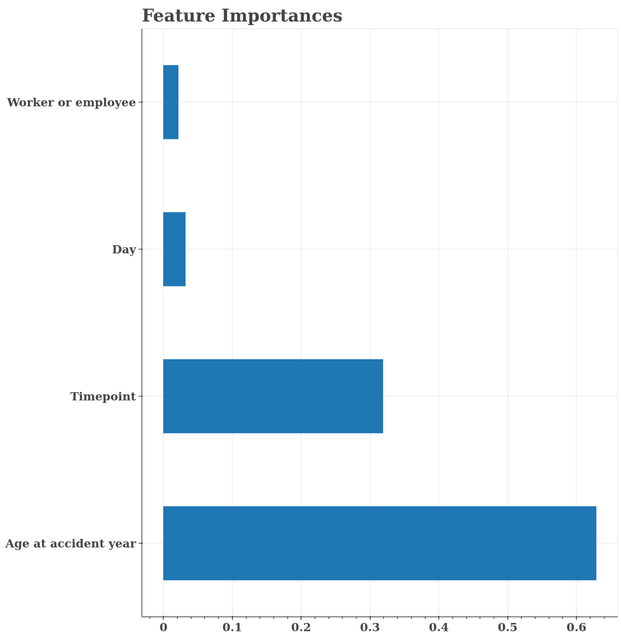


Figure 3. The importance of features in the decision tree trained with 4 features; reprinted with permission from [4], Springer 2022.

Training with all seven input features indeed affected the performance metrics, resulting in an MSE of 164,610.79 and an RRMSE of 1.83. Additionally, the learned rules of the decision tree became more detailed, leading to a change in the order of importance of features.

Random Forest

The random forest was trained following a similar approach to the decision tree regarding preprocessing, dataset splitting, and grid search for the best parameters. With 200 estimators, the model achieved acceptable prediction performance. For the four features, the MSE was found to be 121,154.73, with an RRMSE of 1.84. Similarly, for all seven features, the MSE decreased to 114,203.96, with a corresponding RRMSE of 1.71 [4]. The feature importance values derived from random forests are typically more robust than those from individual decision trees [33]. However, unlike decision trees, random forests do not provide explicit rules for interpretation. Despite this lack of interpretability, random forests generally outperform individual decision trees regarding predictive performance.

Fully Connected Neural Network

A fully connected neural network [37] was implemented to address the regression problem. Utilizing four layers with 50 neurons each, the neural network achieved an MSE of 93,215.32 and an RRMSE of 35.39 with the four basic features. When all seven features were used, the MSE increased to 155,141.41, with a corresponding RRMSE of 37.97 [4]. The optimizer employed was "RMSprop," and the activation function utilized was the Rectified Linear Unit (ReLU). Figure 4 depicts the prediction results for the four features.

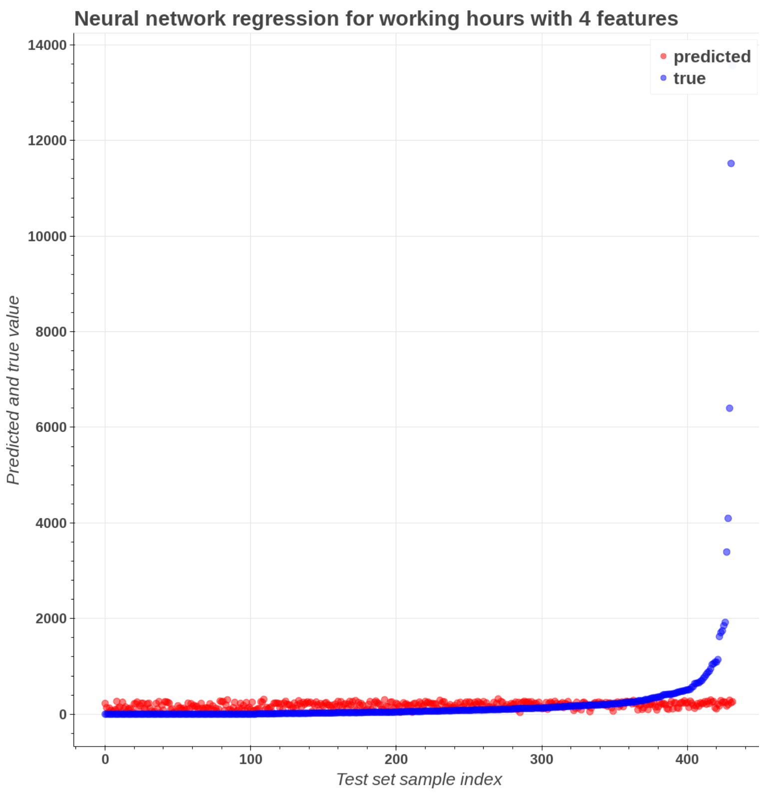


Figure 4. The outcomes of the neural network model on the test set, utilizing 4 input features; reprinted with permission from [4]. Springer 2022.

One method for computing the importance of features is through the SHAP (Shapley values) explainable AI method [38–40]. Figure 5 illustrates the feature importance values obtained using SHAP values for four features (and refer to Table 1).

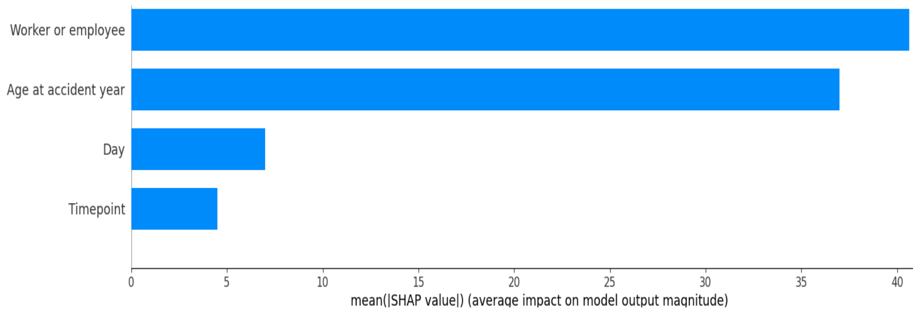


Figure 5. The importance of features in the neural network trained with 4 features; reprinted with permission from [4]. Springer 2022.

Table 1. Comparing the performance of different models with 4 (a) and 7 (b) features; reprinted with permission from [4], Springer 2022.

	Decision Tree	Random Forest	Neural Network
MSE (a)	151,222.99	121,154.73	93,215.32
RRMSE (a)	1.15	1.84	35.39
MSE (b)	164,610.79	114,203.96	155,141.41
RRMSE (b)	1.83	1.71	37.97

3.3.2. Classification

Dealing with the challenge of classifying fatal and non-fatal cases was complex due to a significant class imbalance, where only nine fatal cases were present among 1965 data samples. This imbalance often leads to classifiers achieving a high accuracy but yielding very low, almost negligible mutual information (MI), as discussed in [34]. Therefore, for classification purposes, metrics such as the confusion matrix and mutual information (MI) [33] were utilized. To address the imbalance issue, oversampling was performed on the minority class. Leveraging the Synthetic Minority Oversampling Technique (SMOTE) [41] with the assistance of the imbalanced-learn Python package [42] facilitated the creation of a balanced dataset. Employing four as the number of neighbors to generate synthetic samples yielded the most promising performance outcomes.

Decision Tree

Following a method similar to that outlined in the regression section (refer to Section 3.3.1), the decision tree parameters were determined through a grid search. For the balanced dataset after oversampling, using four features, we observed a mutual information (MI) of 0.60, achieving an accuracy rate of 0.98. The confusion matrix showed the following values: true negatives: 463, true positives: 470, false positives: 17, and false negatives: 2 [4]. Notably, the feature “worker or employee” emerged as the most significant in terms of importance despite being binary. This suggested its substantial discriminatory power, even compared to other categorical or continuous features. It is worth mentioning that among the raw data, five out of nine fatal accidents occurred among employees. Textual representations of the decision tree’s rules, similar to those derived for the regression task in Section 3.3.1, can be presented.

Transitioning to the dataset featuring seven input features, the mutual information (MI) increased to 0.67, accompanied by an accuracy rate of 0.99. Figure 6 visually depicts the feature importance values.

Random Forest

For the random forest model with 100 estimators and four input features, the mutual information (MI) was 0.63, with an accuracy of 0.98. With seven input features, the MI increased to 0.69, and the accuracy reached 1.0 [4]. Moreover, the confusion matrix showed no false positives or negatives, indicating a “perfect” classification outcome. However, it is essential to approach these results with caution as they may suggest overfitting, necessitating further examination.

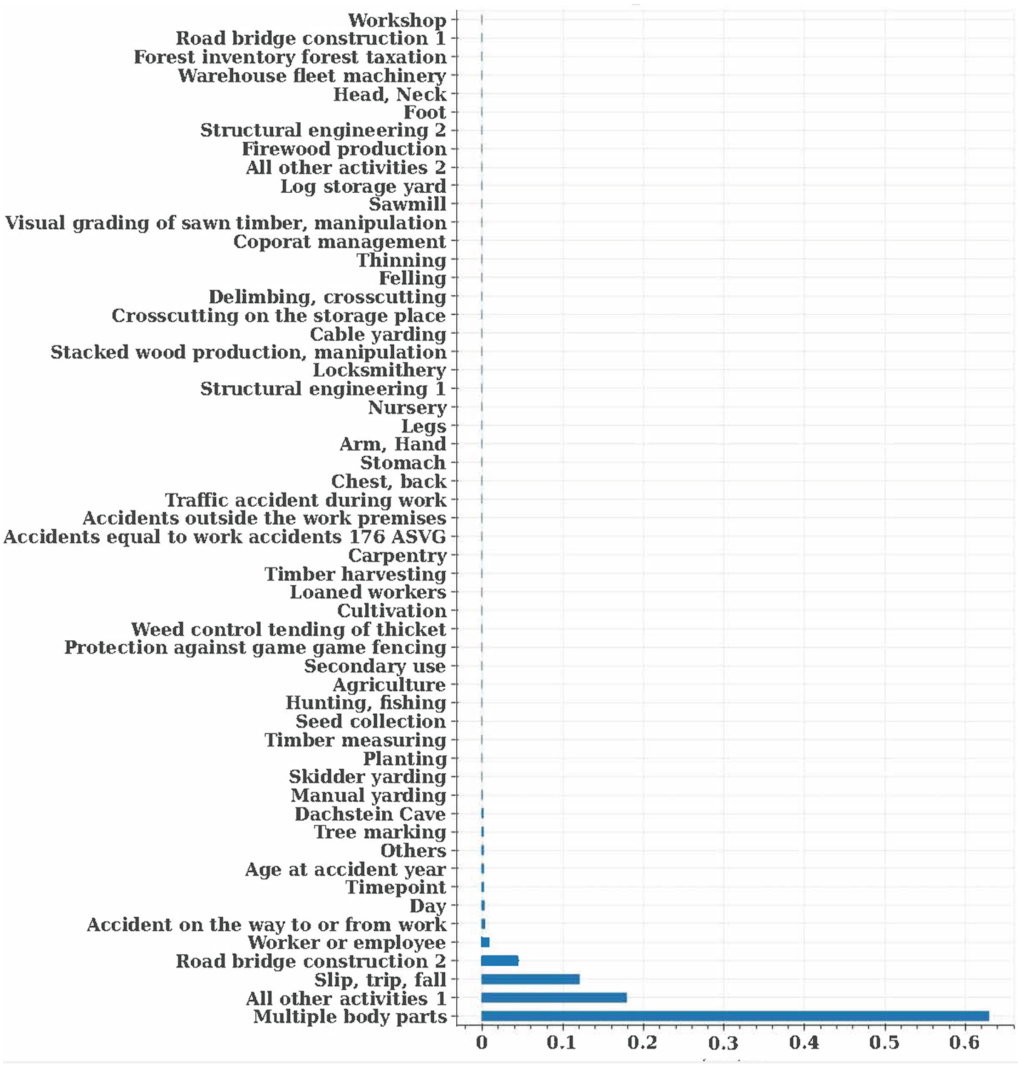


Figure 6. The feature importances in the trained decision tree; on the x axis the feature importance, on the y axis the features; reprinted with permission from [4], Springer 2022.

Fully Connected Neural Network

In the classification task (refer also to Table 2), the neural network architecture matched the one utilized for the regression, with a notable alteration: the use of a sigmoid non-linearity in the output layer. When considering the dataset with four features, the mutual information (MI) was measured at 0.64, alongside an accuracy rate of 0.98. Upon expanding to seven features, the MI improved to 0.67, while the accuracy further improved to 0.99 [4].

Table 2. Comparison of the classification models with 4 (a) and 7 (b) features [4].

	Decision Tree	Random Forest	Neural Network
Accuracy (a)	0.98	0.98	0.98
Mutual information (a)	0.60	0.63	0.64
Accuracy (b)	0.99	1.0	0.99
Mutual information (b)	0.67	0.69	0.67

3.4. Summary and Next Step

The study on dataset 1 investigated occupational accidents within the Austrian Federal Forest, which oversees approximately 10% of the Austrian state area [43]. Decision trees, random forests, and fully connected neural networks were utilized to predict the length of absence resulting from accidents and to categorize incidents as fatal or non-fatal. The decision tree analysis revealed that younger employees and workers typically experienced shorter recovery periods, contingent upon the time of the accident. Furthermore, distinctions were evident between accidents occurring during the initial days of the week and those occurring later in the week.

The findings underscored the significance of age, particularly concerning the prediction of recovery duration for injured individuals. These results align with prior studies [9,44], emphasizing the importance of paying attention to the health and safety of elderly workers and employees, particularly in light of demographic shifts. Furthermore, our analysis confirmed the anticipated disparity between workers and employees, indicating that manual labor was associated with a higher frequency of severe accidents. Moreover, time emerged as a crucial factor in our models, with peaks in occupational accidents observed between 10 and 12 a.m., and a smaller peak occurring between 2 and 4 p.m. This finding echoes previous research attributing variations in accident probabilities throughout the day to factors such as fatigue and dehydration [7,12].

In the subsequent phase of this study, expanding the database holds the potential to enhance and extend the methodology. This expansion could involve the inclusion of additional factors such as company size, level of training, or professional backgrounds of individuals involved. Additionally, conducting comparative analyses between this dataset and data from other institutions across Europe could unveil hidden patterns or correlations. A long-term objective of this research is to develop a causal model [45–47]. It is evident that categorical variables exhibited relationships with each other; a notable example was the "body part" feature, where certain values encompassed accidents affecting multiple body parts, while others focused on single parts (particularly before grouping). Investigating the existence, structure, and parameters of relationships among variables such as the working sector, cause of the accident, and affected body parts will be the focus of future investigations, necessitating higher data quality requirements.

4. Dataset 2 Austrian Workers' Compensation Board (AUVA)

4.1. Descriptive Statistics for the Austrian Forestry Accident Statistics (AUVA)

Before diving into the creation of ML models, some first data explorations were performed after a data preprocessing stage. Since not all accidents deal with forest workers, it is important to separate them and see where they differ in their characteristics, particularly in the severity of the diagnosis. The five more common items involved in accidents among forest workers involve mostly branches and tree trunks, slippery floors, and stony, rocky areas.

The dataset comprises 14,611 observations, each providing unique individual information regarding accidents, including the date and time of the incident, age, and sick leave duration. "AgeId" in this dataset represents an encoded feature, where observations between ages $x0$ to $x9$ are encoded as AgeID $x5$, where x belongs to the set $[1, 2, 3, 4, 5, 6, 7]$. For instance, observations between 20 and 29 are encoded as age 25, indicating that they are in their twenties. Similar encoding principles apply to other age groups. With this encoding, the dataset presents an average age of 45, with the 25th and 75th quartiles at 35 and 55 years old, respectively, and a median age of 45. The average sick leave duration is 14 days, with 75% of the data showing 11 days or fewer and a maximum leave of 539 days.

Some highlights from the dataset can be extracted by inspecting each factor individually. Looking at which day of the week most of the accidents occurred, it was visible that Tuesday (2569) had the maximum number and Sunday the least (210).

The time when most accidents happened was analyzed in previous studies [7,12] and the results are depicted in Figure 7. Most accidents occurred between 10:00 and 11:00, with another local maximum around 15:00–16:00.

The relationship between the month of the accident and their number was also quite reasonable; January saw the most accidents, with 1618, and June, the least, with 875, as depicted in Figure 8.

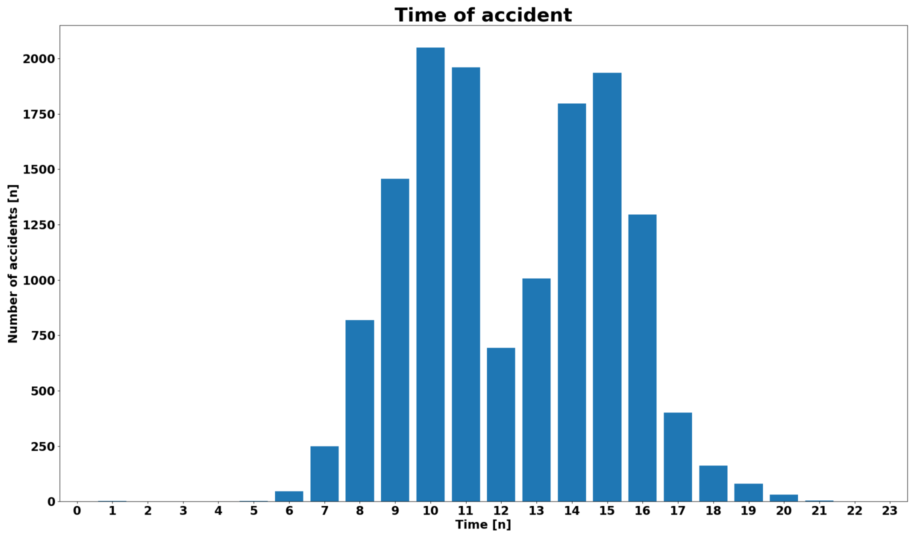


Figure 7. Histogram of the times when accidents occurred.

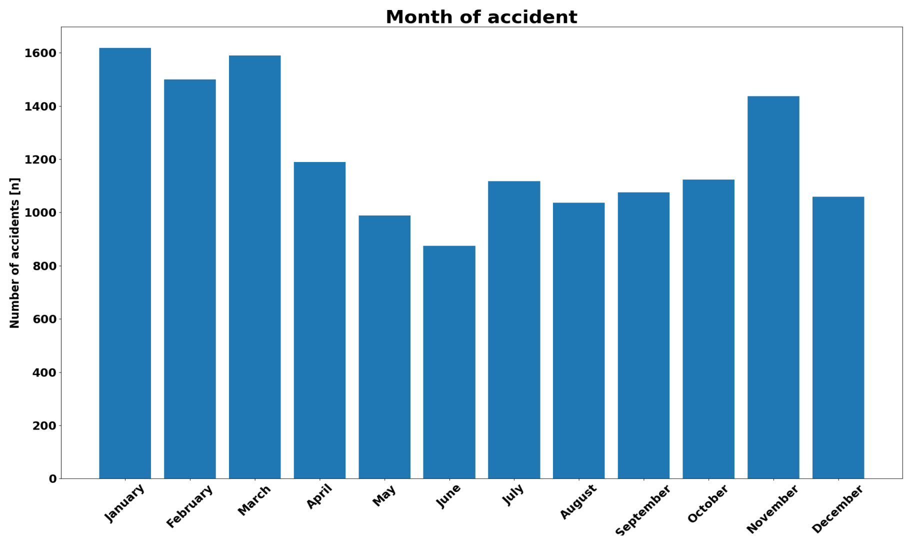


Figure 8. Histogram of the months when accidents occurred.

Figure 9 depicts the distribution of sick leave days among various age groups. The samples were categorized as follows: (1) under twenty, (2) individuals in their twenties, (3) individuals in their thirties, (4) individuals in their forties, (5) individuals in their fifties, (6) individuals in their sixties, and (7) individuals in their seventies. A notable observation is that the longest sick leave duration values were associated with the fifties age group.

Upon further investigation into the relationship between age ID and sick leave duration, it became apparent that most accidents were observed among individuals in their fifties. This analysis is visualized in Figure 10. However, instances of severe injuries requiring a three-month recovery period (90 days of sick leave) were comparable among individuals aged twenty to the end of their fifties. Similarly, shorter sick leaves (less than three days) exhibited similar patterns across these age groups. Notably, age IDs 53, 45, and 55 each featured four sick leave cases lasting longer than a year, while age ID 25 recorded only three such cases.

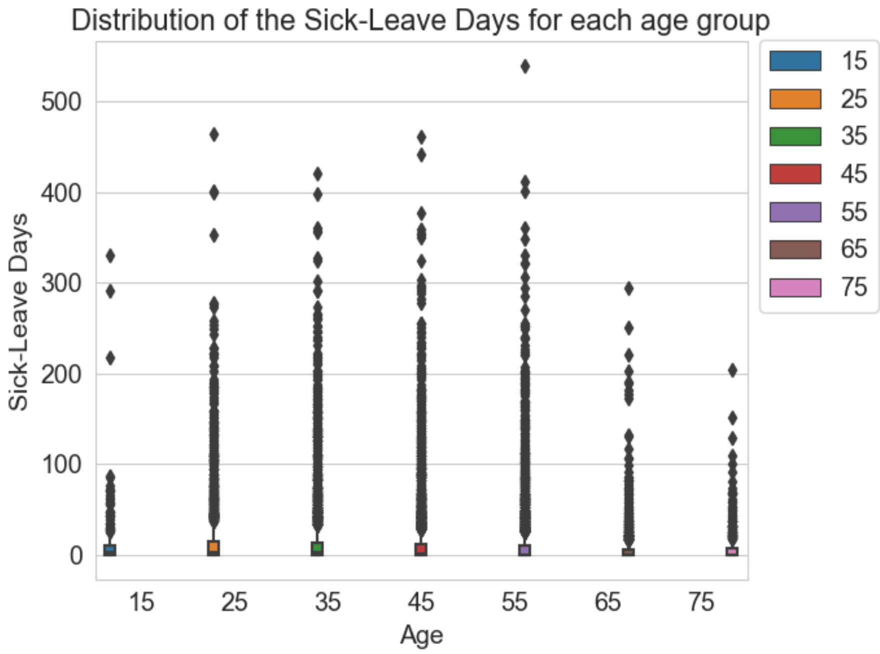


Figure 9. Visualization of the distribution of sick leave days across age groups.

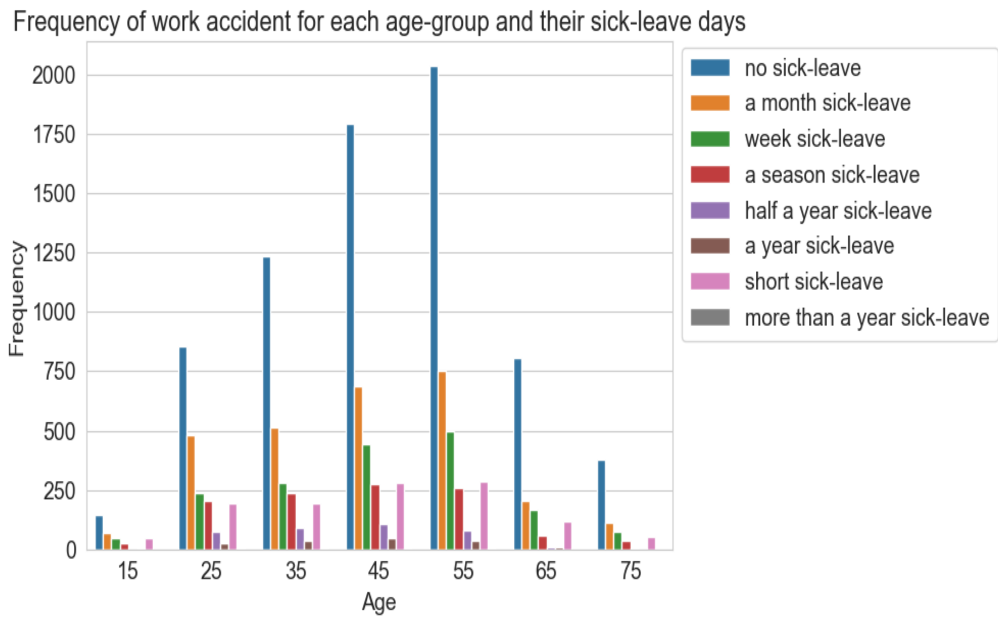


Figure 10. Visualization of the frequency of work accidents categorized by sick leave days across different age groups.

To further examine the correlation among these features, we assumed linearity between each two features and measured the correlations using the Pearson method. This revealed no significant interaction between the features, except for a slight positive correlation between the recorded accident month and sick leave days (in German: "Krankenstandstage") and a slight negative correlation between the duration of sick leave and age ID. We further calculated the Spearman correlation coefficients with the assumption of monotonic relationships. Similar to the Pearson correlations, we did not observe any significant relationship between the features (see Figure 11). It is important to note that observations with no sick leave days were excluded from this correlation analysis.

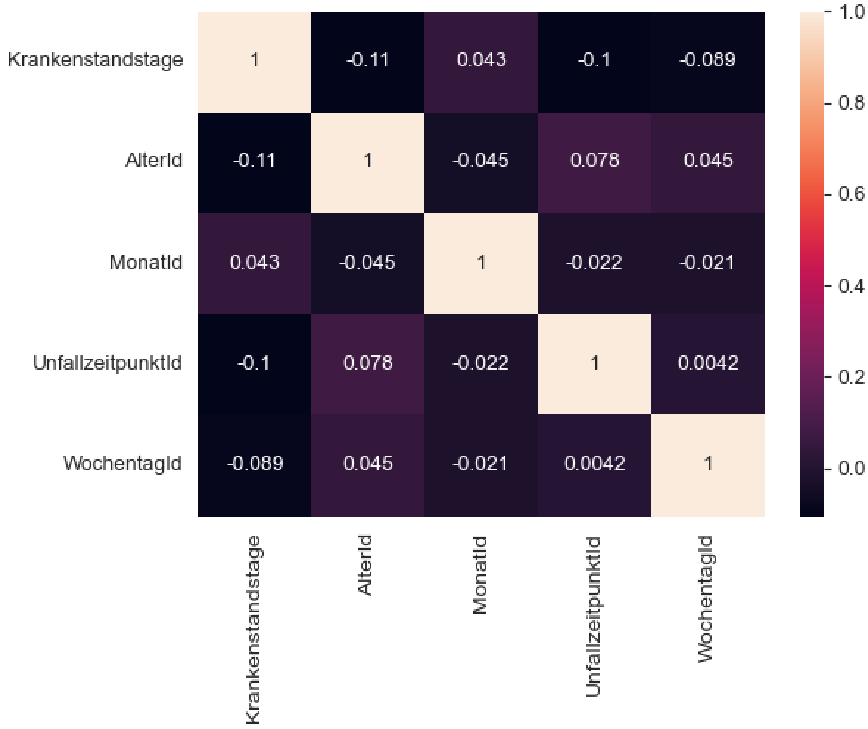


Figure 11. Illustration of the Spearman correlation results for the selected features.

We further calculated the mutual information (MI) between all the discrete features and the continuous target variable sick leave days. Figure 12 depicts the feature and estimated MI values for sick leave days vs. each input variable. MI is a non-negative value that quantifies the dependency between two random variables. It equals zero only when the variables are independent, while higher values indicate a stronger dependency between them. We observed that the sick leave days were highly dependent on the weekdays.

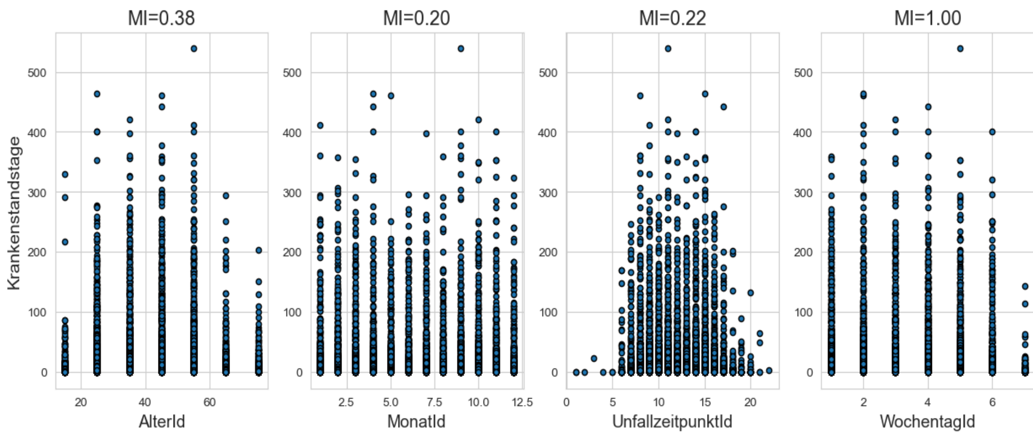


Figure 12. Illustration of the estimated mutual information for the selected features.

4.2. Machine Learning (ML) Models

Machine Learning (ML) models such as decision trees (DTs), random forests (RFs), and neural networks (NNs) can be used to make predictions about the fatality of the accident (classification) and the number of days a patient spends being hospitalized. Those predictions are based on the learning of a non-linear relationship between the factors that are gathered to describe the phenomenon and the variable that is being predicted. If the prediction reaches a sufficiently large performance value, then the internal decision-making process of such models can provide some insights into the data that the models were trained upon and can uncover associations between multiple influencing factors that act at the same time and not evaluate the effect of each of the factors alone (as seen in Section 4.1).

The prediction of the number of sick leave days can be turned into a classification task when a threshold >0 days denotes an accident resulting in downtime, and anything equal to 0 hospitalization days is not considered so effective. Human domain knowledge was incorporated to choose the first six factors that were believed to be decisive for predicting the severity of accidents: time of accident, month, age, weekday, type of injury, and reason for injury. Although ML algorithms perform amazingly well today, human domain knowledge can sometimes be very important for certain tasks, and the involvement of a human expert can bring advantages [48]. Month and time had to be cyclically encoded with the use of `sine` and `cosine` functions since conditions in December were considered to be similar to the ones in January; similarly, 00:00 midnight should be encoded as 01:00. Furthermore, the type of injury and reason for injury contained textual information, which in this first stage was encoded with the simple word2vec algorithm [49]. The resulting continuous three-dimensional embeddings had prevalence over the other input features regarding decision-making importance because of the inductive bias, which is an inherent property of several ML solutions, including decision trees (DTs) [50]. The data were augmented with a Variational Autoencoder (VAE) [51] before they were used for training. The resulting augmented dataset was imbalanced (52,007 for 0 and 12,286 for >0 sick leave days), correctly reflecting the class imbalance in the original dataset; therefore, the Adaptive Synthetic Sampling approach (ADASYN) [52] was used for minority class sampling.

The performance of a decision tree (DT) trained on the aforementioned factors and hyperparameter search is depicted in Figure 13. The number of true negatives (TN) was 10,722, that of true positives (TP) was 10,301, that of false positives (FP) was 801, and that of false negatives (FN) was 945 [33]. The classification had an overall balanced accuracy score of 0.923, and the mutual information (MI) was 0.422 [34].

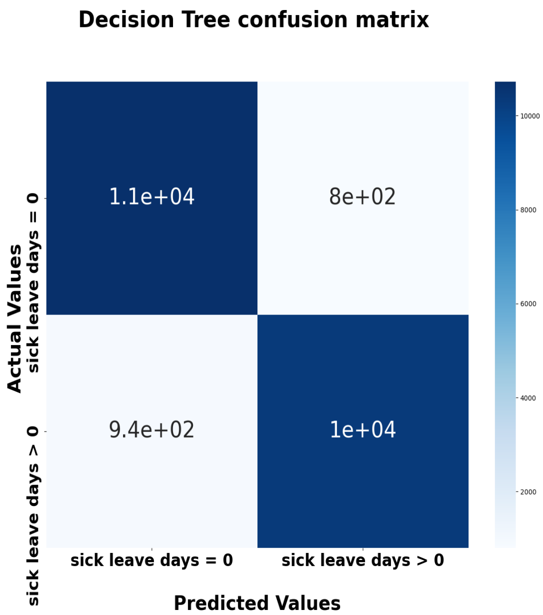


Figure 13. Confusion matrix of the decision tree.

The rules by which the DT decided whether a particular accident case should be predicted as having 0 or >0 hospitalization days are depicted in the Figure 14. For better readability, only the two first levels of the DT are presented. It is observable that the type of injury was selected as the most relevant feature for the classification, which is questionable with respect to expert domain knowledge. This phenomenon can be explained through the aforementioned inductive bias [50], and together with the fact that the decision tree contains the necessary encoding of the input features in its rules, the DT is overall a suboptimal solution for modeling forest accidents.

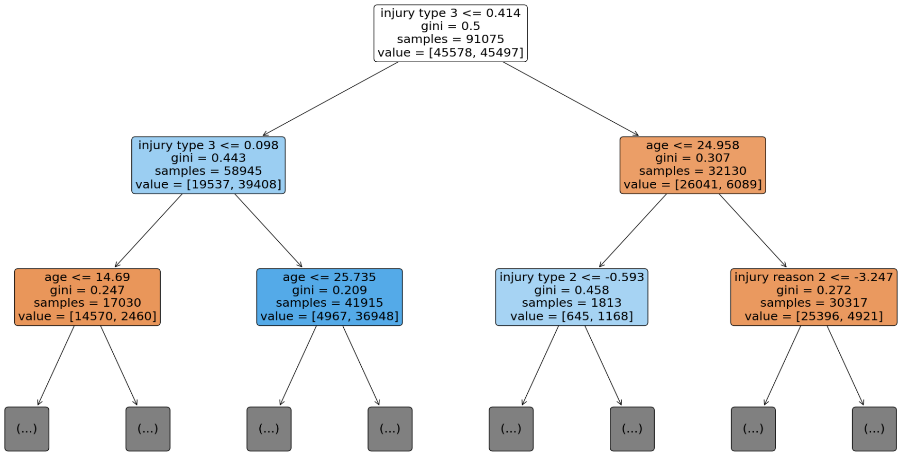


Figure 14. Decision tree rules (first 2 levels).

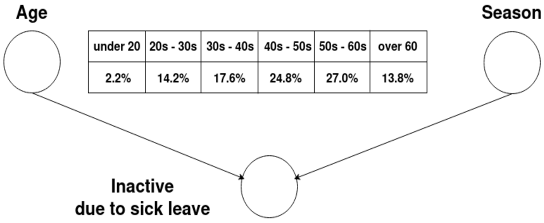
4.3. Probabilistic Graphical Models (PGMs)

Probabilistic Graphical Models (PGMs) represent human beliefs about causal dependencies and their strength between a set of random variables (RVs) that describe a phenomenon [46]. A PGM's modeled causal dependencies are depicted as edges in a Directed Acyclic Graph (DAG) where the nodes are visible or hidden RVs; together with the numerical description of each RV, the PGM represents the joint distribution over all RVs. This model does not learn associations between the input factors and the predicted variable (although it makes predictions) but goes beyond that and learns causal dependencies both from the human domain expertise and the available data. Furthermore, it can be used for answering probabilistic queries in cases where evidential information is present and as a generative model as well.

Figure 15 depicts a PGM trained on the AUVA dataset containing only two of the twenty-six possible features. The age RV has six possible outcomes: "under 20", "20s–30s", "30s–40s", "40s–50s", "50s–60s", "over 60". The season RV is deduced from monthly information, and the inactivity due to sick leave has two possible outcomes: "minor" (meaning $\leq$3 sick leave days because of the accident) and "severe" (>3 sick leave days). The graphical structure of the PGM expresses the prior beliefs of the human domain experts about the causal relationships and dependencies between the RVs. It was assumed that season and age were independent of each other and that the inactivity of the forest worker due to hospitalization was directly caused by those two.

The training was done with Maximum Likelihood Estimation (MLE) [46], meaning that the parameters of the local distributions in Figure 15 were influenced only by the available data and not from prior knowledge of the human domain expert (a soft and uniform prior). After training was accomplished, the model could be used to provide readable explanations directly from the CPTs. For example, the highlighted row in Figure 15 is the answer to an informative statement:

If the worker is young (under 20) and the season is winter, then the probability of a severe accident, and thereby more than 3 sick leave days, is 35.8%. The evaluation of long hospitalization times for older people (over 60) can be computed through a marginalization over the season variable; it was 37.9%.



under 20	20s - 30s	30s - 40s	40s - 50s	50s - 60s	over 60
2.2%	14.2%	17.6%	24.8%	27.0%	13.8%

spring	summer	autumn	winter
25.7%	20.7%	24.8%	28.5%

	minor	severe
under 20, spring	54.1%	45.9%
under 20, summer	43.2%	56.8%
under 20, autumn	51.2%	48.8%
under 20, winter	64.2%	35.8%
20s - 30s, spring	48.3%	51.7%
...	...	...
over 60, autumn	62.8%	37.2%
over 60, winter	60.6%	39.4%

Figure 15. Naive Bayes model with two parent RVs (age and season) and one predicted RV (inactivity due to hospitalization days). Each row of each Conditional Probability Table (CPT) – also called Conditional Probability Distribution (CPD) – contains the probabilities (posterior) for each of the outcomes conditioned on one of the combinations of the values of the parent RVs. RVs that have no parents do not need conditioning.

5. Discussion

This work elucidated the efficacy of integrating diverse analytical approaches to harness the potential of a comprehensive dataset enriched with an array of recorded factors, aiming to bolster occupational safety within the forestry sector. It delineated a methodological triad comprising classical descriptive statistics, traditional machine learning techniques, and advanced probabilistic models, each contributing uniquely to the analysis and interpretation of the data.

In the initial phase of the descriptive statistical examination, patterns emerged that aligned with findings from previous research [7,12], highlighting two peaks in the frequency of work-related accidents within the daily work schedule: notably between 10:00 and 11:00 and between 15:00 and 16:00. Additionally, the analysis revealed that forestry workers predominantly encountered accidents involving interactions with branches, tree trunks, encounters with slippery soil surfaces, and navigating through areas laden with stones and rocks and gravel. A discernible seasonal trend in the incidence of accidents was also observed, with a higher occurrence rate during the winter months spanning November to March, albeit with a notable dip in December, attributed to the holiday season's reduction in work activities.

Within the scope of this study, probabilistic modeling techniques were utilized to validate or refute expert opinions on the causality of accidents. The applied explanatory model, for instance, revealed that the likelihood of a young forestry worker (under 20 years of age) experiencing a severe occupational accident (entailing more than 3 days of sick leave) during the winter season was estimated to be 35.8%.

Machine learning methodologies were applied to enhance the interpretability of the dataset. Specifically, a decision tree model was employed, integrating six variables from the dataset—including the time of the accident, month, age of the worker, day of the week, nature of the injury, and cause of the injury—to ascertain the likelihood of an accident leading to downtime. This model attained a commendable balanced accuracy score of 0.923, accompanied by a mutual information (MI) score of 0.422.

This paper effectively illustrates that multiple methodologies can be deployed to distill insights from a dataset, offering valuable perspectives on practical queries. The choice of the approach is contingent upon the nature of the question at hand, with different techniques potentially offering the most significant insights.

6. Conclusions and Future Work

In conclusion, this manuscript effectively demonstrated the power of combining diverse analytical methodologies to analyze and interpret a rich and complex dataset for enhancing occupational safety in the forestry sector. By weaving together classical descriptive statistics, machine learning techniques, and probabilistic models, the study provided a comprehensive analysis that unveiled the intricate dynamics of work-related accidents. From laying the groundwork with descriptive statistics to harnessing machine learning for deeper insights and employing probabilistic models to overcome data limitations, each approach played a crucial role in understanding and predicting accident occurrences. This holistic analysis not only identified key risk factors but also offered actionable recommendations for mitigating these risks, thereby contributing to the advancement of safer work environments in forestry. The paper underscored the significance of interdisciplinary strategies in leveraging data for occupational safety, highlighting the essential role of data-driven decision-making in fostering a safer and more informed forestry work setting.

With modern AI models, it is possible to shed light on possible causal relationships of forest accidents, which can lead to faster prevention and avoidance. The gradual takeover of these tasks by AI agents, and later possibly by robots, could also help; however, of course, accidents with and by robots cannot be ruled out. In any case, and this is very important, all the information collected from human accidents can be encoded in the prior, which can then be used to predict robot accidents in the forest. Hopefully, this work will contribute to an enhanced comprehension of the underlying causes of accidents in the perilous realm of forestry work, thereby aiding in the prevention of fatalities in the future.

A key area of future work includes the exploration of causal machine learning. Causal machine learning enables AI models to generate precise predictions by considering causal relationships rather than solely relying on correlations [53]. Causal approaches enhance the robustness, explainability, and fairness of models [54]. This must go hand in hand with the development of a comprehensive knowledge graph, which holds significant potential for advancing our understanding of the causal relationships underlying forest accidents. The incorporation of the human-in-the-loop approach [55] promises to enhance the relevance and applicability of AI solutions in this context. In future research, we aim to integrate counterfactual explanations [56] for analyzing forestry accidents. Such counterfactual explanations, describing alternative scenarios that could have prevented an accident, offer a powerful tool for understanding the complex dynamics of forestry work environments. By providing insights into "what-if" scenarios, such as how altering certain conditions might have averted an accident, this approach can enhance predictive models and lead to more effective preventative strategies of high practical use in accident prevention. Ultimately, these explanations will not only deepen the current understanding of accident causation but also assist in developing targeted interventions for improving safety measures.

Another relevant future work is on time- and energy-efficient modeling with Prior-Data Fitted Networks (PFNs). New ML architectures like Transformers [57,58] have proven their success not only in language but also in tabular data processing tasks [59]. Being able to solve all aforementioned classification and regression tasks just in milliseconds after a one-time (pre-)training procedure provides enormous benefits as far as energy consumption and near-real-time processing are concerned. Furthermore, the (pre-)training procedure can incorporate human expert domain knowledge in the form of a Bayesian Neural Network (BNN) [60,61] and/or a Structural Causal Model (SCM) [62,63], thereby being an excellent practical example of the human-in-the-loop principle [64] due to the fact that humans are able to bring in conceptual understanding and contextual knowledge [65]. Nevertheless, the explainability aspect has to be addressed and fulfill the expectations of forest operation experts to the same–if not a greater degree–as the PGMs mentioned in Section 4.3.

A very important future work would be the integration of knowledge graphs. Basically, knowledge graphs (KGs) [66] can encapsulate vast quantities of structured and unstructured data related to forestry operations, including geographical, environmental, and operational parameters. Researchers can enrich the analytical process by integrating these graphs with existing AI models such as TabPFN. This integration would allow one to leverage relational information and contextual knowledge, enhancing the accuracy and relevance of predictive models. Combined with AI, knowledge graphs could identify previously unnoticed patterns and relationships in the data, leading to a more comprehensive understanding of the factors contributing to accidents. The work of Shi et al. (2024) [67] recently used ontologies for analyzing accidents in the construction industry where they used a domain word discovery algorithm to build a construction safety ontology and combined this ontology with a Text Convolutional Neural Network (TextCNN). Ontologies and knowledge graphs, while both crucial in the organization and representation of knowledge, have distinct differences and applications. An ontology is a formal and structured representation of concepts and their relationships within a domain, primarily focused on establishing a common understanding that can be shared among people and computational entities. Knowledge graphs, conversely, are expansive networks of interconnected entities and their attributes, often encompassing a variety of data sources, including both structured and unstructured data. They are less rigid in structure compared to ontologies and excel in scalability, flexibility, and integration with AI and machine learning technologies. This makes knowledge graphs particularly effective for handling large, complex datasets and for applications that require sophisticated data analytics and real-time information processing, surpassing the capabilities of traditional ontologies in practical, dynamic environments [66].

Another possible future work is the synergy of counterfactual explanations and knowledge graphs. This combination could provide a multi-faceted view of accident scenarios, encompassing both the specific details of individual cases and the broader operational context. Employing knowledge graphs to inform and generate counterfactual scenarios could result in more realistic and applicable safety recommendations. This integrated approach would be particularly beneficial in tailoring safety protocols to the unique conditions of specific forestry operations, thereby enhancing the overall effectiveness of occupational safety strategies.

Author Contributions: Conceptualization, F.H., K.S. and A.H.; methodology, F.H., A.S. and A.J.; software, A.S. and A.J.; validation, K.S.; formal analysis, A.S. and A.J.; investigation, F.H., A.S., A.H. and A.J.; resources, A.H.; data curation, A.S.; writing—original draft preparation, F.H., A.S., A.J. and A.H.; writing—review and editing, F.H., A.S., A.J. and A.H.; visualization, A.S.; supervision, K.S.; funding acquisition, A.H. All authors have read and agreed to the published version of the manuscript.

Funding: Parts of this work have been funded by the Austrian Science Fund (FWF), Project: P-32554 explainable Artificial Intelligence.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is unavailable due to privacy restrictions.

Conflicts of Interest: The authors declare that they have no conflicts of interest. This work does not raise any ethical issues.

Abbreviations

ADASYN	Adaptive Synthetic Sampling approach
AI	Artificial intelligence
AUVA	Allgemeine Unfallversicherungsanstalt
BNN	Bayesian Neural Network
CPD	Conditional Probability Distribution
CPT	Conditional Probability Table
DAG	Directed Acyclic Graph
DT	Decision Tree
FCNN	Fully connected neural networks
FN	False negative
FP	False positive
KG	Knowledge graph
LRP	Layer-wise Relevance Propagation
ML	Machine learning
ÖBf	Österreichische Bundesforste AG
PFN	Prior-Data Fitted Networks
PGM	Probabilistic Graphical Model
RF	Random forest
RV	Random variable
SCM	Structural Causal Model
TN	True negative
TP	True positive
TabPFN	Tabular Prior-Data Fitted Network
VAE	Variational Autoencoder
xAI	Explainable AI

References

1. Slappendel, C.; Laird, I.; Kawachi, I.; Marshall, S.; Cryer, C. Factors affecting work-related injury among forestry workers: A review. *J. Saf. Res.* **1993**, *24*, 19–32. [CrossRef]
2. Rickards, J. The human factor in forest operations: Engineering for health and safety. *For. Chron.* **2008**, *84*, 539–542. [CrossRef]
3. Landekić, M.; Martinić, I.; Mijoč, D.; Bakarić, M.; Šporčić, M. Injury patterns among forestry workers in Croatia. *Forests* **2021**, *12*, 1356. [CrossRef]
4. Hoenigsberger, F.; Saranti, A.; Angerschmid, A.; Retzlaff, C.O.; Gollob, C.; Witzmann, S.; Nothdurft, A.; Kieseberg, P.; Holzinger, A.; Stampfer, K. Machine Learning and Knowledge Extraction to Support Work Safety for Smart Forest Operations. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*; Springer Lecture Notes in Computer Science LNCS 13480; Springer: Cham, Switzerland, 2022; pp. 362–375. [CrossRef]
5. Saranti, A.; Hudec, M.; Minarikova, E.; Takac, Z.; Großschedl, U.; Koch, C.; Pfeifer, B.; Angerschmid, A.; Holzinger, A. Actionable explainable AI (AxAI): A practical example with aggregation functions for adaptive classification and textual explanations for interpretable Machine Learning. *Mach. Learn. Knowl. Extr.* **2022**, *4*, 924–953. [CrossRef]
6. Hinze, A.; König, J.L.; Bowen, J. Worker-fatigue contributing to workplace incidents in New Zealand Forestry. *J. Saf. Res.* **2021**, *79*, 304–320. [CrossRef] [PubMed]
7. Tsioras, P.A.; Rottensteiner, C.; Stampfer, K. Wood harvesting accidents in the Austrian state forest enterprise 2000–2009. *Saf. Sci.* **2014**, *62*, 400–408. [CrossRef]
8. Grzywiński, W.; Skonieczna, J.; Jelonek, T.; Tomczak, A. The influence of the privatization process on accident rates in the forestry sector in Poland. *Int. J. Environ. Res. Public Health* **2020**, *17*, 3055. [CrossRef] [PubMed]
9. Jankovskỳ, M.; Allman, M.; Allmanová, Z. What are the occupational risks in forestry? Results of a long-term study in Slovakia. *Int. J. Environ. Res. Public Health* **2019**, *16*, 4931. [CrossRef]
10. Allman, M.; Jankovskỳ, M.; Allmanová, Z.; Ferenčík, M.; Messingerová, V.; Vlčková, M.; Stoilov, S. Work accidents during cable yarding operations in Central Europe 2006–2014. *For. Syst.* **2017**, *26*, 13. [CrossRef]
11. Laschi, A.; Marchi, E.; Foderi, C.; Neri, F. Identifying causes, dynamics and consequences of work accidents in forest operations in an alpine context. *Saf. Sci.* **2016**, *89*, 28–35. [CrossRef]
12. Lima, K.S.; Meira Castro, A.C.; Torres Costa, J.; Baptista, J.S. Occupational accidents in native and planted forests in Brazil: 2007–2018. *Work* **2022**, *71*, 719–728. [CrossRef]
13. Akay, A.O.; Akgül, M.; Esin, A.I.; Demir, M.; Şentürk, N.; Öztürk, T. Evaluation of occupational accidents in forestry in Europe and Turkey by k-means clustering analysis. *Turk. J. Agric. For.* **2021**, *45*, 495–509. [CrossRef]
14. López-Toro, A.A.; Pardo-Ferreira, M.C.; Martínez-Rojas, M.; Carrillo-Castrillo, J.A.; Rubio-Romero, J.C. Analysis of occupational accidents during the chainsaws use in Andalucía. *Saf. Sci.* **2021**, *143*, 105436. [CrossRef]

15. Montorselli, N.B.; Lombardini, C.; Magagnotti, N.; Marchi, E.; Neri, F.; Picchi, G.; Spinelli, R. Relating safety, productivity and company type for motor-manual logging operations in the Italian Alps. *Accid. Anal. Prev.* **2010**, *42*, 2013–2017. [CrossRef]
16. Bentley, T.A.; Parker, R.J.; Ashby, L. Understanding felling safety in the New Zealand forest industry. *Appl. Ergon.* **2005**, *36*, 165–175. [CrossRef] [PubMed]
17. Martin, J.E.; Rivas, T.; Matías, J.; Taboada, J.; Argüelles, A. A Bayesian network analysis of workplace accidents caused by falls from a height. *Saf. Sci.* **2009**, *47*, 206–214. [CrossRef]
18. Matías, J.M.; Rivas, T.; Martín, J.; Taboada, J. A machine learning methodology for the analysis of workplace accidents. *Int. J. Comput. Math.* **2008**, *85*, 559–578. [CrossRef]
19. Tamascelli, N.; Solini, R.; Paltrinieri, N.; Cozzani, V. Learning from major accidents: A machine learning approach. *Comput. Chem. Eng.* **2022**, *162*, 107786. [CrossRef]
20. Sarkar, S.; Pramanik, A.; Maiti, J.; Reniers, G. Predicting and analyzing injury severity: A machine learning-based approach using class-imbalanced proactive and reactive data. *Saf. Sci.* **2020**, *125*, 104616. [CrossRef]
21. Oyedele, A.; Ajayi, A.; Oyedele, L.O.; Delgado, J.M.D.; Akanbi, L.; Akinade, O.; Owolabi, H.; Bilal, M. Deep learning and Boosted trees for injuries prediction in power infrastructure projects. *Appl. Soft Comput.* **2021**, *110*, 107587. [CrossRef]
22. Sarkar, S.; Vinay, S.; Raj, R.; Maiti, J.; Mitra, P. Application of optimized machine learning techniques for prediction of occupational accidents. *Comput. Oper. Res.* **2019**, *106*, 210–224. [CrossRef]
23. Mohanta, B.K.; Jena, D.; Mohapatra, N.; Ramasubbareddy, S.; Rawal, B.S. Machine learning based accident prediction in secure iot enable transportation system. *J. Intell. Fuzzy Syst.* **2022**, *42*, 713–725. [CrossRef]
24. Venkat, A.; KP, G.V.; Thomas, I.S. Machine Learning Based Analysis for Road Accident Prediction. *Int. J. Emerg. Technol. Innov. Eng. IJETIE* **2020**, *6*, *2*. Available online: https://ssrn.com/abstract=3550462 (accessed on 28 February 2024).
25. Galatioto, F.; Catalano, M.; Shaikh, N.; McCormick, E.; Johnston, R. Advanced accident prediction models and impacts assessment. *IET Intell. Transp. Syst.* **2018**, *12*, 1131–1141. [CrossRef]
26. Pandas Development Team April, 10, 2024 v.2.2.2. In *Zenodo: EU Open Research Repository*; 2024. Available online: https://zenodo.org/records/10957263 (accessed on 28 February 2024). [CrossRef]
27. McKinney, W. Data Structures for Statistical Computing in Python. In Proceedings of the 9th Python in Science Conference, Austin, TX, USA, 28 June–3 July 2010; van der Walt, S., Millman, J., Eds.; 2010; pp. 56–61. Available online: http://conference.scipy.org.s3-website-us-east-1.amazonaws.com/proceedings/scipy2010/mckinney.html (accessed on 28 February 2024). [CrossRef]
28. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; et al. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.
29. Hunter, J.D. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
30. Bokeh Development Team. *Bokeh: Python Library for Interactive Visualization*; 2018. Available online: https://docs.bokeh.org (accessed on 28 February 2024).
31. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*; Morgan Kaufmann: San Francisco, CA, USA, 2011.
32. Kalousis, A.; Prados, J.; Hilario, M. Stability of feature selection algorithms: A study on high-dimensional spaces. *Knowl. Inf. Syst.* **2007**, *12*, 95–116. [CrossRef]
33. Géron, A. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2022.
34. MacKay, D.J. *Information Theory, Inference and Learning Algorithms*; Cambridge University Press: Cambridge, UK, 2003.
35. Van Der Maaten, L.; Postma, E.O.; van den Herik, H.J. Dimensionality reduction: A comparative review. *J. Mach. Learn. Res.* **2009**, *10*, 13.
36. Zheng, A.; Casari, A. *Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists*; O'Reilly Media, Inc.: Sebastopol, CA, USA, 2018.
37. Bishop, C.M.; Nasrabadi, N.M. *Pattern Recognition and Machine Learning*; Springer: New York, NY, USA, 2006.
38. Shapley, L.S. A value for n-person games. *Class. Game Theory* **1997**, *69*.
39. Štrumbelj, E.; Kononenko, I. Explaining prediction models and individual predictions with feature contributions. *Knowl. Inf. Syst.* **2014**, *41*, 647–665. [CrossRef]
40. Staniak, M.; Biecek, P. Explanations of model predictions with live and breakDown packages. *arXiv* **2018**, arXiv:1804.01955.
41. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic minority over-sampling technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
42. Lemaître, G.; Nogueira, F.; Aridas, C.K. Imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning. *J. Mach. Learn. Res.* **2017**, *18*, 1–5.
43. Österreichische Bundesforste. Insights. 2020. Available online: https://www.bundesforste.at/unternehmen/ueber-uns/zahlen-fakten.html (accessed on 28 February 2024).
44. Lilley, R.; Feyer, A.M.; Kirk, P.; Gander, P. A survey of forest workers in New Zealand: Do hours of work, rest, and recovery play a role in accidents and injury? *J. Saf. Res.* **2002**, *33*, 53–71. [CrossRef] [PubMed]
45. Barber, D. *Bayesian Reasoning and Machine Learning*; Cambridge University Press: Cambridge, UK, 2012.
46. Koller, D.; Friedman, N. *Probabilistic Graphical Models: Principles and Techniques*; MIT Press: Cambridge, MA, USA, 2009.

47. Saranti, A.; Taraghi, B.; Ebner, M.; Holzinger, A. Insights into learning competence through probabilistic graphical models. In Proceedings of the International Cross-Domain Conference for Machine Learning and Knowledge Extraction, Canterbury, UK, 26–29 August 2019; Springer: New York, NY, USA, 2019; pp. 250–271.
48. Holzinger, A.; Saranti, A.; Angerschmid, A.; Retzlaff, C.O.; Gronauer, A.; Pejakovic, V.; Medel, F.; Krexner, T.; Gollob, C.; Stampfer, K. Digital Transformation in Smart Farm and Forest Operations needs Human-Centered AI: Challenges and Future Directions. *Sensors* **2022**, *22*, 3043. [CrossRef]
49. Mikolov, T.; Chen, K.; Corrado, G.; Dean, J. Efficient estimation of word representations in vector space. *arXiv* **2013**, arXiv:1301.3781.
50. Witten, I.H.; Frank, E.; Hall, M.A. *Data Mining: Practical Machine Learning Tools and Techniques*, 3rd ed.; Morgan Kaufmann Series in Data Management Systems; Morgan Kaufmann: Amsterdam, The Netherlands, 2011.
51. Kingma, D.P.; Welling, M. Auto-encoding variational bayes. *arXiv* **2013**, arXiv:1312.6114.
52. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, China, 1–6 June 2008; IEEE: Piscataway, NJ, USA, 2008; pp. 1322–1328.
53. Ness, R.O. *Causal AI*; Manning: Shelter Island, NY, USA, 2024.
54. Angerschmid, A.; Zhou, J.; Theuermann, K.; Chen, F.; Holzinger, A. Fairness and Explanation in AI-Informed Decision Making. *Mach. Learn. Knowl. Extr.* **2022**, *4*, 556–579. [CrossRef]
55. Retzlaff, C.O.; Das, S.; Wayllace, C.; Mousavi, P.; Afshari, M.; Yang, T.; Saranti, A.; Angerschmid, A.; Taylor, M.E.; Holzinger, A. Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities. *J. Artif. Intell. Res. (JAIR)* **2024**, *79*, 349–415. [CrossRef]
56. Del Ser, J.; Barredo-Arrieta, A.; Díaz-Rodríguez, N.; Herrera, F.; Saranti, A.; Holzinger, A. On Generating Trustworthy Counterfactual Explanations. *Inf. Sci.* **2024**, *655*, 119898. [CrossRef]
57. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
58. Bai, Y.; Chen, F.; Wang, H.; Xiong, C.; Mei, S. Transformers as statisticians: Provable in-context learning with in-context algorithm selection. *Adv. Neural Inf. Process. Syst. (NeurPIS 2023)* **2023**, *36*, 1–87.
59. Hollmann, N.; Müller, S.; Eggensperger, K.; Hutter, F. Tabpfn: A transformer that solves small tabular classification problems in a second. *arXiv* **2022**, arXiv:2207.01848.
60. Hinton, G.E.; Srivastava, N.; Krizhevsky, A.; Sutskever, I.; Salakhutdinov, R.R. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv* **2012**, arXiv:1207.0580.
61. Blundell, C.; Cornebise, J.; Kavukcuoglu, K.; Wierstra, D. Weight uncertainty in neural network. In Proceedings of the International Conference on Machine Learning, PMLR, Lille, France, 6–11 July 2015; pp. 1613–1622.
62. Peters, J.; Janzing, D.; Schölkopf, B. *Elements of Causal Inference: Foundations and Learning Algorithms*; The MIT Press: Cambridge, MA, USA, 2017.
63. Pearl, J. *Causal Inference in Statistics: An Overview*; UCLA: Los Angeles, CA, USA, 2009.
64. Holzinger, A.; Saranti, A.; Hauschild, A.C.; Beinecke, J.; Heider, D.; Roettger, R.; Mueller, H.; Baumbach, J.; Pfeifer, B. Human-in-the-Loop Integration with Domain-Knowledge Graphs for Explainable Federated Deep Learning. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*; Lecture Notes in Computer Science (LNCS) Volume 14065; Springer: Cham, Switzerland, 2023; pp. 45–64. [CrossRef]
65. Holzinger, A.; Saranti, A.; Angerschmid, A.; Finzel, B.; Schmid, U.; Mueller, H. Toward human-level concept learning: Pattern benchmarking for AI algorithms. *Patterns* **2023**, *4*, 1–21. [CrossRef]
66. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Philip, S.Y. A survey on knowledge graphs: Representation, acquisition, and applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 494–514. [CrossRef]
67. Shi, D.; Li, Z.; Zurada, J.; Manikas, A.; Guan, J.; Weichbroth, P. Ontology-based text convolution neural network (TextCNN) for prediction of construction accidents. *Knowl. Inf. Syst.* **2024**, *66*, 2651–2681. [CrossRef]