Machine Learning and Knowledge Extraction to Support Work Safety for Smart Forest Operations

Ferdinand Hoenigsberger[1(✉)], Anna Saranti[1], Alessa Angerschmid[1], Carl Orge Retzlaff[1], Christoph Gollob[1], Sarah Witzmann[1], Arne Nothdurft[1], Peter Kieseberg[2], Andreas Holzinger[1], and Karl Stampfer[1]

[1] University of Natural Resources and Life Sciences, Vienna, Austria
ferdinand.hoenigsberger@boku.ac.at
[2] University of Applied Sciences St.Poelten, St.Poelten, Austria

Abstract. Forestry work is one of the most difficult and dangerous professions in all production areas worldwide - therefore, any kind of occupational safety and any contribution to increasing occupational safety plays a major role, in line with addressing sustainability goal SDG 3 (good health and well-being). Detailed records of occupational accidents and the analysis of these data play an important role in understanding the interacting factors that lead to occupational accidents and, if possible, adjusting them for the future. However, the application of machine learning and knowledge extraction in this domain is still in its infancy, so this contribution is also intended to serve as a starting point and test bed for the future application of artificial intelligence in occupational safety and health, particularly in forestry. In this context, this study evaluates the accident data of Österreichische Bundesforste AG (ÖBf), Austria's largest forestry company, for the years 2005–2021. Overall, there are 2481 registered accidents, 9 of which were fatal. For the task of forecasting the absence hours due to an accident as well as the classification of fatal or non-fatal cases, decision trees, random forests and fully-connected neuronal networks were used.

Keywords: Occupational accident · Artificial Intelligence · Explainable AI · Forestry · Machine learning · Explainability · Human-in-the-Loop

1 Introduction

Even though mechanization significantly reduced the number of accidents and fatalities in forestry [26], they remain on a high level. The 2021 accident statistics for forestry work in Austria [2] recorded 1189 accidents, 21 of which were fatal. Forestry work is thereby still one of the most dangerous and difficult occupations in all fields of production and also appears in the list of heavy physical labor (§ 1 Abs 1 Z 4 SchwerarbeiterVO, BGBl. II 104/2006 as amended by BGBl. II 413/2019). New approaches are needed to further reduce the number of accidents.

Published by Springer Nature Switzerland AG 2022
A. Holzinger et al. (Eds.): CD-MAKE 2022, LNCS 13480, pp. 362–375, 2022.
https://doi.org/10.1007/978-3-031-14463-9_23

Extraordinarily high hopes are therefore being placed in digital transformation to support the sustainable development goals [13], and its applications to make a major step toward improving ergonomics and occupational safety addressing the sustainability goal SDG 3 [37]. Detailed records of work accidents and the analysis of these data play an important role (and are therefore also legally required) in order to understand the interacting factors that lead to occupational accidents and, if possible, adjust them for the future. Various studies have looked at and analyzed different aspects of such accident records. Tsioras et al. (2014) [34] for example evaluated accidents during timber harvesting in Austria. Such studies were also completed for other countries (e.g. Poland [10], Slovakia [3,15], Italy [17], or Brazil [18]) as well as comparisons between different countries [1]. One focus of research centers on accidents during the work with chainsaws respectively during motor-manual tree felling [5,19,24]. All these approaches and contributions reach their limits where a multitude of factors and influences interact. Therefore, this contribution compares ML methods with "conventional" statistical methods. Today, machine learning and knowledge extraction permeate practically all application domains and especially in the domain of forestry it has great future potential [12]. The use of ML owes it especially to the application in the field of accident prediction to two factors: on the one hand due to its explanatory capacity and with that the ability to gain insights into major causes of accidents [21], and secondly with its predictive capacity [22], enabling better accident prevention. Other professions, mainly those with high accident risk or high accident severity, also start to exploit the capabilities of machine learning for the analysis of occupational accidents. For example, the analysis of serious occupational accidents in the chemical industry [33], or in the steel industry [28]. For electrical engineering, Oyedele et al. (2021) [25] used deep learning and boosted trees algorithms to test the possibilities and effectiveness for the use in the area of safety risk management. Refer to [8,23,29,35] for other examples of the application of ML in the field of accident prediction.

In general, however, the application of machine learning and knowledge extraction in this domain is still in its infancy. Therefore, this contribution is intended to serve as an important starting point and test bed not only for the application of machine learning in occupational safety and health, especially in forestry, but also for new developments, especially in emerging subfields of artificial intelligence, including explainable AI and counterfactual explanations ("what if ... questions ") - especially with respect to a future trustworthy AI [11]. In this context, this study evaluates accident data from Österreichische Bundesforste AG (ÖBf), Austria's largest forestry company, for the years 2005–2021 using "conventional" statistical methods as well as machine learning algorithms to analyze and contrast the context of occupational accidents and lay the foundation for further work.

2 Dataset

2.1 Description

The dataset consists of tabular data containing information about occupational accidents. In principle, an accident can be fatal or non-fatal; in the latter case, the injured person will not be able to work for a particular number of hours or days. The information that is most descriptive from domain knowledge and previous models [33,34] is mainly contained in the columns that specify the time on the day on which the accident happened (given in minutes), the age of the person, and the day in the week (starting with 1 = Monday). Furthermore, there is a differentiation between workers and employees in that manner, since severe accidents rather happen to workers. The cause of the accident, working sector, and body part are also provided with an appropriate encoding. It is apparent that there are associations between the columns and potential causal relationships, but for the scope of this research work, only their linear and non-linear correlations were explored.

The dataset was explored and analyzed with the use of Python libraries `pandas`: https://pandas.pydata.org/, preprocessed mainly with https://scikit-learn.org and visualized with `matplotlib`: https://matplotlib.org/ and `Bokeh`: https://bokeh.org/:.

2.2 Preprocessing

Overall, there are 2481 registered accidents from 01 October 2005 until 21 December 2021. Of those, only 9 were fatal, which does not provide enough information on the influencing and decisive factors of those accidents. Nevertheless, a classification between non-fatal and fatal accidents is of great importance and can be supported with the gathering of more relevant data. A dedicated pre-processing stage consisted of the removal of non-filled entries and invalid content rows. In total, there were 7 input feature columns that were used for prediction: "worker or employee", "age at accident", "day of the week", "time in the day", "working sector", "accident cause" and "injured body part". After the removal of invalid entries (empty string, invalid value as f.e. working sector 51 where the valid ones are only between 1 and 35) there were 1965 rows with valid data left.

Before using the data as input for the predictive models, histograms that visualized their contents were computed. It is important to note that the cardinality of the domain of the input feature is playing a substantial role in its importance. Continuous features and categorical features with many categories have a bias towards being more descriptive than ones with less categories [14]. In the case of the input features "injured body part", "working sector" and "accident cause", grouping of values was used. For example, all accidents that affected the head and neck area were grouped together.

Both linear and non-linear correlations between all pairs of features were computed before they were used as input to the predictive models. The linear correlations were measured with the Pearson coefficient [9] and the non-linear with Mutual Information (MI) [20]. The highest linear correlation was found between working sector and cause of the accident (0.26), the second-highest between working sector and worker or employee (−0.26) and the third between cause and body part (−0.13). All other pairs had values < 0.1. The MI is correspondingly 0.25, 0.19 and 0.06 for the aforementioned pairs, and anything else is < 0.1. Those values are supported by domain knowledge and by previous research work. No substantial correlations were found in any of the pairs, so it was not necessary to remove redundant input features to trim the model, or even apply some dimensionality reduction before the training [9,36].

3 Predictive Models

3.1 Regression Task

Decision Tree. The main predictive task is - in case of a non-fatal accident - to predict the number of hours the injured person will not be able to work, because of the need to recover. There is a differentiation between workers and employees in that manner; the hours that employees need for recovery are quite different from the ones of the workers. Figure 1 shows the difference between the two distributions of values; this was to be expected, since manual work tends to cause more and more serious accidents. We created one model that picks up commonalities and differences between those two groups, but we are well aware that in the future, if necessary, we could create two different detailed models.

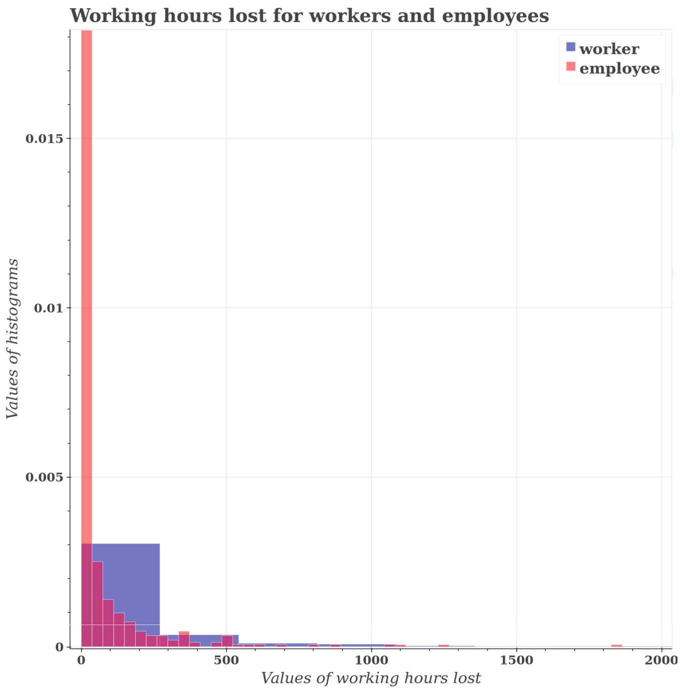

Fig. 1. Difference in the distributions of recovery working hours after an accident in workers and employees. The figure is cropped on the right side, to zoom on the vast majority of the values that are < 2000.

This is a regression problem [6] and the use of a Decision Tree (DT) [9] is a straightforward, explainable-per-design method for such a task. It is not necessary to scale the numeric or continuous features; nevertheless, as far as non-ordinal input features are concerned - such as the working sector, the cause of the accident and the injured body part - there is already a mapping to integer numbers. For example, in the body part feature, the ears are encoded with the number 102, the eyes with $113, 123, 133$ to denote right, left or both eyes, the face with 104, the teeth with 105. As one can observe, those body parts are not in an ordinal relationship with each other, that means a predictive model should be prevented to assume so or base its decision and generalization on this. Furthermore, as mentioned in the previous Sect. 2.2, on those features, grouping of conceptually similar values was applied and composed one category. The one-hot encoding was used [36], where each category was encoded by a vector containing only one 1 while the rest of the entries are 0 and being orthonormal to the rest. This still contains issues in this case, since there is a column for each category, the feature values matrix is quite sparse and large.

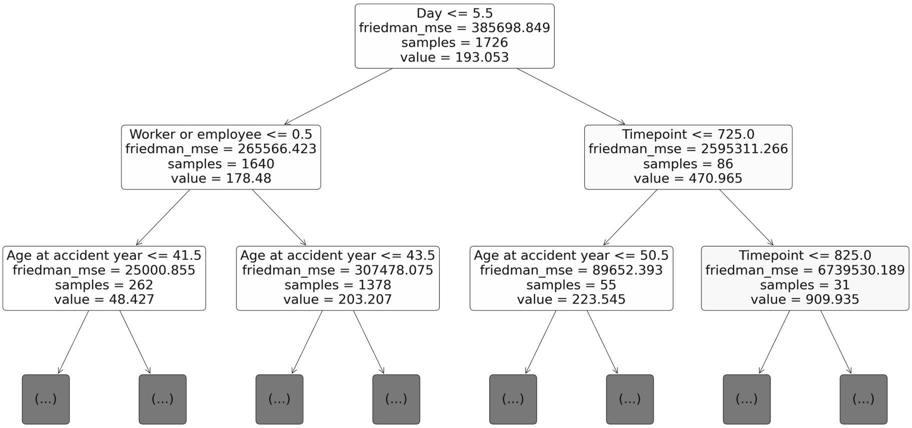

Fig. 2. The learned rules of the decision tree trained with 4 features.

The dataset was shuffled and split in two parts, the training set contains 80% of the original data, whereas 20% are used for the test set. There was a grid search applied to find out the best parameters for the decision tree on the training set, which were maximum depth of tree: 5, maximum number of features to consider: "auto" (denoting that the number of features considered for the best split equals the number of all input features), and the minimum number of sample necessary for split: 2. The performance of the decision tree on the test set with the use of 4 features, namely the age of the person at the time of the accident, the day in the week, the point of time in the day and the information if the person is a worker or an employee was a Mean Squared Error (MSE) of 151222.99 and a Relative Root Mean Squared Error (RRMSE) of 1.15.

The learned rules of the decision tree trained with 4 features are depicted in Fig. 2. The decision tree rules can also be extracted and compared to previous research insights. According to the decision tree it can be shown that a rather small number of recovery hours is expected for younger employee and workers in dependence to the time of the accident; furthermore differences can be observed between the first days of the week and the rest of the week.

```
|--- Day <= 5.50
|   |--- Worker or employee <= 0.50
|   |   |--- Age at accident year <= 41.50
|   |   |   |--- Timepoint <= 1275.00
|   |   |   |   |--- Age at accident year <= 20.50
|   |   |   |   |   |--- value: [58.00]
|   |   |   |   |--- Age at accident year >  20.50
|   |   |   |   |   |--- value: [17.67]
|   |   |   |--- Timepoint >  1275.00
|   |   |   |   |--- Timepoint <= 1320.00
|   |   |   |   |   |--- value: [504.00]
|   |   |   |   |--- Timepoint >  1320.00
|   |   |   |   |   |--- value: [80.00]
|   |   |--- Age at accident year >  41.50
|   |   |   |--- Age at accident year <= 45.50
|   |   |   |   |--- Timepoint <= 525.00
|   |   |   |   |   |--- value: [484.80]
|   |   |   |   |--- Timepoint >  525.00
|   |   |   |   |   |--- value: [61.33]
|   |   |   |--- Age at accident year >  45.50
|   |   |   |   |--- Age at accident year <= 54.50
|   |   |   |   |   |--- value: [34.57]
|   |   |   |   |--- Age at accident year >  54.50
|   |   |   |   |   |--- value: [70.95]
...
|--- Day >  5.50
...
```

The input features importances are in accordance to previous research work [33,34] and are depicted in Fig. 3.

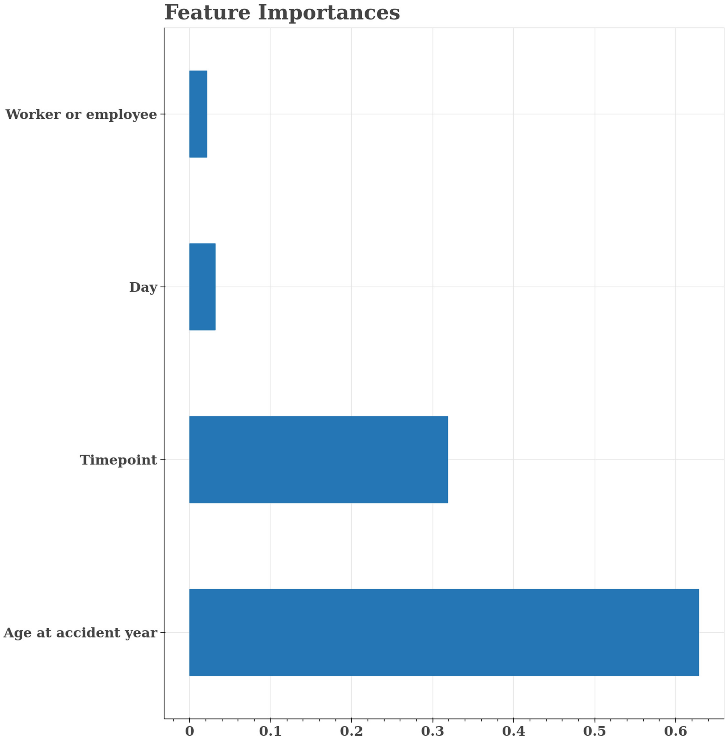

Fig. 3. The feature importances of the decision tree that is trained with 4 features.

Training with all 7 input features does have an impact on the MSE 164610.79 and RRMSE 1.83. The learned rules of the decision tree are more detailed, and the order of feature importances changes.

Random Forest. A random forest was also trained in a way similar to the decision tree as far as pre-processing, dataset split and grid search of the best parameters is concerned. 200 estimators were found, enough to achieve an acceptable prediction performance with MSE for 4 features: 121154.73 and RRMSE 1.84, whereas for 7 features: 114203.96 and 1.71 correspondingly.

The feature importances are more robust than the ones from the decision trees [9], but there is no tree to depict the rules. Random forests are not considered interpretable, but have generally better performance than decision trees.

Fully-Connected Neural Network. A fully-connected neural network [6] was implemented to also tackle the problem of regression. Four layers with 50 neurons each are enough to reach an MSE of 93215.32 and an RRMSE 35.39 with the 4 basic features, whereas 155141.41 and 37.97 with all 7 correspondingly.

The optimizer that was used was "RMSprop" and the activation unit was the Rectified Linear Unit (ReLU). The results of the prediction for 4 features can be seen in Fig. 4.

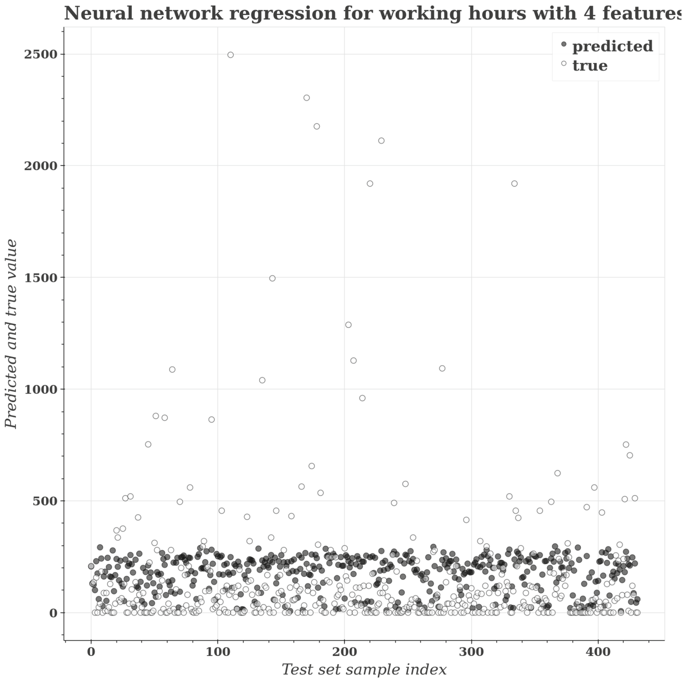

Fig. 4. The results of the neural network on the test set for 4 input features.

One way to compute the importance of features is the SHAP (Shapley) values Explainable AI method [30–32] as depicted in Fig. 5 for 4 features.

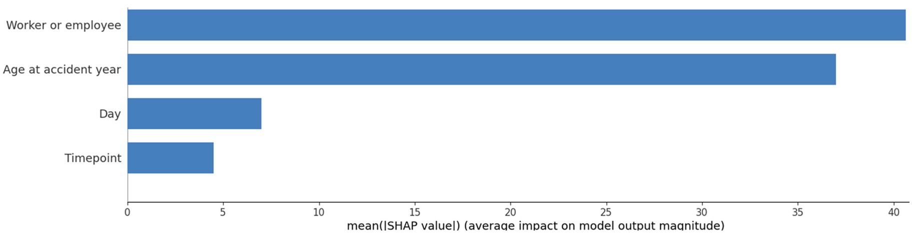

Fig. 5. The feature importances of the neural network that is trained with 4 features.

Table 1. Comparison of different regression models using 4 features for the prediction of lost working hours due to accidents

	Decision tree	Random forest	Neural network
MSE	151222.99	121154.73	93215.32
RRMSE	1.15	1.84	35.39

Table 2. Comparison of different regression models using all features for the prediction of lost working hours due to accidents.

	Decision tree	Random forest	Neural network
MSE	164610.79	114203.96	155141.41
RRMSE	1.83	1.71	37.97

3.2 Classification

The classification of fatal and non-fatal cases is a task that is characterized by its great imbalance - 9 fatal cases in 1965 data samples led to classifiers with high accuracy but very low, near zero Mutual Information (MI) - which was expected, as explained in [20]. Therefore, the metrics that were used for the classification were the confusion matrix and Mutual Information (MI) [9]. Another way to counteract the imbalance is to perform oversampling on the class with the smaller number of samples. The Synthetic Minority Oversampling TEchnique (SMOTE) [7] was used with the help of the `imbalanced-learn` Python package https://imbalanced-learn.org to create a balanced dataset. The number of neighbors that were used to construct synthetic samples was 4, this has provided the most profitable performance results.

Decision Tree. The decision tree parameters were defined by a grid search similar to the one described in the regression Sect. 3.1. The Mutual Information (MI) is 0.60, and accuracy 0.98 for the balanced dataset after oversampling with 4 features. The confusion matrix is described by the true negatives: 463, true positives: 470, false positives: 17 and false negatives: 2. As far as the feature importances is concerned, the most important feature is found to be the "worker or employee" feature; since this is a binary feature and does not carry such discriminating information in comparison to the other categorical with more categories or the continuous ones, it must be of great significance. From the raw data 5 out of 9 fatal accidents occurred among employees. The rules can be extracted as text, equal to the regression task 3.1.

For 7 input features the MI is 0.67 and the accuracy is 0.99. The feature importances are depicted in the Fig. 6.

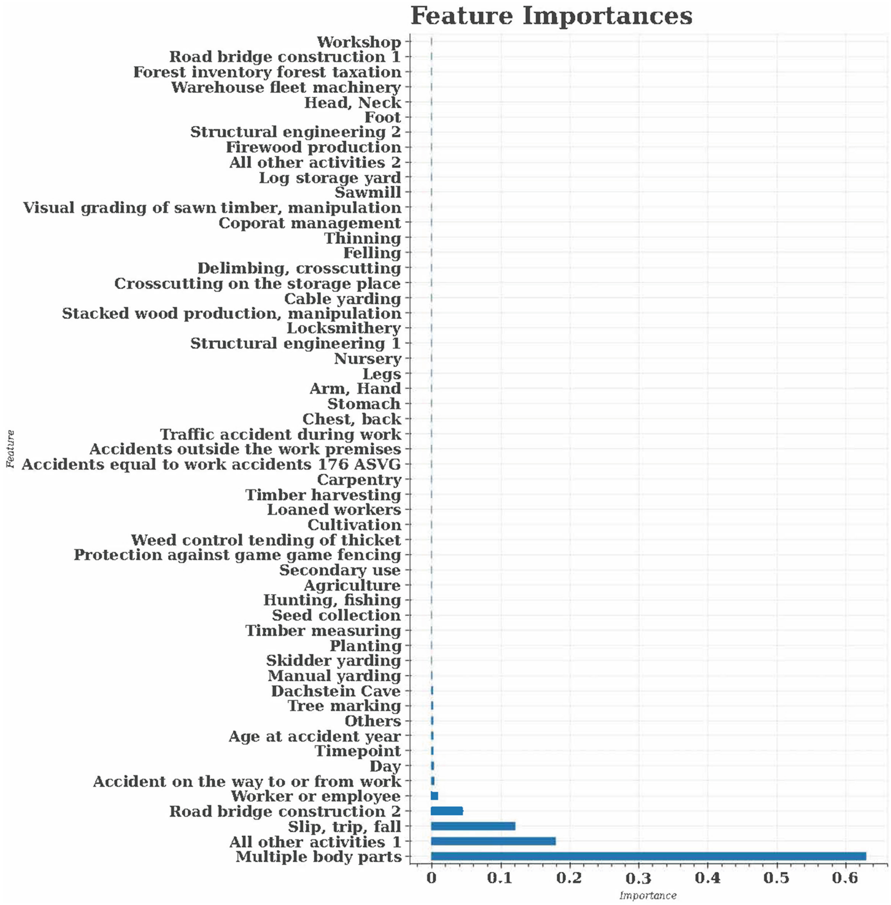

Fig. 6. The feature importances of the decision tree that is trained with all features.

Random Forest. MI with random forest, 100 estimators and 4 input features 0.63, whereas accuracy reaches 0.98. 7 input features reach MI 0.69, accuracy 1.0 with a "perfect" confusion matrix (no false positives or negatives), although this needs to be further examined with care since it might be an over-fitting indication.

Fully-Connected Neural Network. The architecture of the neural network is similar to the one used in the regression task 3.1. The fundamental difference lies in the use of a sigmoid non-linearity in the output layer. MI with 4 features: 0.64, with accuracy 0.98 and with 7 features: MI 0.67 and accuracy 0.99.

Table 3. Comparison of different classification models with 4 features for fatal and non-fatal accidents.

	Decision tree	Random forest	Neural network
Accuracy	0.98	0.98	0.98
Mutual Information	0.60	0.63	0.64

Table 4. Comparison of different classification models with all features for fatal and non-fatal accidents.

	Decision tree	Random forest	Neural network
Accuracy	0.99	1.0	0.99
Mutual Information	0.67	0.69	0.67

4 Conclusion and Future Research Questions

The current study analyzes the occupational accidents of the ÖBf AG, which manages about 10% of the Austrian state area [38]. For the task of forecasting the absence hours due to an accident as well as the classification of fatal or non-fatal cases, a decision tree, a random forest, and a fully-connected neuronal network were used. According to the decision tree it can be shown that a rather small number of recovery hours is expected for younger employee and workers in dependence to the time of the accident; furthermore differences can be observed between the first days of the week and the rest of the week.

Our results show further, that age is a decisive factor especially when predicting the number of hours the injured person will not be able to work. Our results are consistent with other studies [15] and once again show that attention must be paid to the health of elder workers and employees. In the light of the demographic change this statement is especially relevant. The found difference between workers and employees confirms the expectation, that manual work tends to cause more and more serious accidents. The time is also an important factor in our models. Peaks in occupational accidents occur between 10 and 12 am. A second smaller peak occurs between 2 and 4 pm. Our study is here in line with other studies of occupational accidents, which highlighted fatigue and dehydration as possible explanations [18,34] for different accident probabilities during the day.

In the future, by expanding the database, the methodology could be further refined and extended to include factors such as the size of the company, the level of training or the professional background of the people involved. Furthermore, it would be insightful to compare this data to data from other institutions throughout Europe and search for hidden patterns or correlations. A long-term goal of this research is a causal model [4,16,27]. As seen, the values of the categorical variables do have some relationship with each other; one of the most prominent examples is the "body part" feature, where one of the values encompasses accidents in several parts of the body whereas the rest concentrate on only one (especially before grouping). The existence, structure and parameters of relationships between working sector, cause of the accident and the affected body part are subject of future investigations and are expected to have higher data quality requirements in general.

Acknowledgements. The authors declare that there are no conflict of interests. This work does not raise any ethical issues. Parts of this work have been funded by the Austrian Science Fund (FWF), Project: P-32554, explainable AI.

The authors would like to thank ÖBf AG, and especially Mr. Stefan Trzesniowski, for the kind provision of accident data.

References

1. Akay, A.O.: Evaluation of occupational accidents in forestry in Europe and Turkey by k-means clustering analysis. Turkish J. Agric. Forest. **45**(4), 495–509 (2021). https://doi.org/10.3906/tar-2010-55
2. Allgemeine Unfallversicherungsanstalt (AUVA): Unfallstatistik 2021 forstwirtschaftliche arbeiten. AUVA Schwerpunktauswertungen (2022)
3. Allman, M., et al.: Work accidents during cable yarding operations in central Europe 2006–2014. Forest Syst. **26**(1), 13 (2017). https://doi.org/10.5424/fs/2017261-10365
4. Barber, D.: Bayesian Reasoning and Machine Learning. Cambridge University Press, Cambridge (2012)
5. Bentley, T.A., Parker, R.J., Ashby, L.: Understanding felling safety in the New Zealand forest industry. Appl. Egonom. **36**(2), 165–175 (2005). https://doi.org/10.1016/j.apergo.2004.10.009
6. Bishop, C.M., Nasrabadi, N.M.: Pattern recognition and machine learning. Springer (2006)
7. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: synthetic minority over-sampling technique. J. Artif. Intell. Res. **16**, 321–357 (2002)
8. Galatioto, F., Catalano, M., Shaikh, N., McCormick, E., Johnston, R.: Advanced accident prediction models and impacts assessment. IET Intell. Transp. Syst. **12**(9), 1131–1141 (2018)
9. Géron, A.: Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems. " O'Reilly Media, Inc." (2019)
10. 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 **17**(9), 3055 (2020). https://doi.org/10.3390/ijerph17093055
11. Holzinger, A.: The next frontier: AI we can really trust. In: Kamp, M. (ed.) Proceedings of the ECML PKDD 2021, CCIS 1524, pp. 1–14. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-93736-2-33
12. Holzinger, A., et al.: Digital transformation in smart farm and forest operations needs human-centered AI: challenges and future directions. Sensors **22**(8), 3043 (2022). https://doi.org/10.3390/s22083043
13. Holzinger, A., Weippl, E., Tjoa, A.M., Kieseberg, P.: Digital transformation for sustainable development goals (SDGS) - a security, safety and privacy perspective on AI. In: Springer Lecture Notes in Computer Science, LNCS 12844, pp. 1–20. Springer (2021). https://doi.org/10.1007/978-3-030-84060-0

14. Ian, H.W., Eibe, F.: Data mining: Practical Machine Learning Tools And Techniques (2005)
15. 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 **16**(24), 4931 (2019). https://doi.org/10.3390/ijerph16244931
16. Koller, D., Friedman, N.: Probabilistic Graphical Models: Principles And Techniques. MIT Press, London (2009)
17. Laschi, A., Marchi, E., Foderi, C., Neri, F.: Identifying causes, dynamics and consequences of work accidents in forest operations in an alpine context. Safety Sci. **89**, 28–35 (2016). https://doi.org/10.1016/j.ssci.2016.05.017
18. 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 **71**(3), 719–728 (2022). https://doi.org/10.3233/WOR-210543
19. 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. **143**, 105436 (2021). https://doi.org/10.1016/j.ssci.2021.105436
20. MacKay, D.J., Mac Kay, D.J., et al.: Information Theory, Inference and Learning Algorithms. Cambridge University Press , Cambridge (2003)
21. 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.Saf. Sci. **47**(2), 206–214 (2009)
22. Matías, J.M., Rivas, T., Martín, J., Taboada, J.: A machine learning methodology for the analysis of workplace accidents. Int. J. Comput. Math. **85**(3–4), 559–578 (2008)
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. **42**(2), 713–725 (2022)
24. Montorselli, N.B., et al.: Relating safety, productivity and company type for motor-manual logging operations in the Italian Alps. Acc. Anal. Preven. **42**(6), 2013–2017 (2010). https://doi.org/10.1016/j.aap.2010.06.011
25. Oyedele, A., et al.: Deep learning and boosted trees for injuries prediction in power infrastructure projects. Appl. Soft Comput. **110**, 107587 (2021). https://doi.org/10.1016/j.asoc.2021.107587
26. Rickards, J.: The human factor in forest operations: engineering for health and safety. For. Chron. **84**(4), 539–542, (2008). https://doi.org/10.5558/tfc84539-4
27. Saranti, A., Taraghi, B., Ebner, M., Holzinger, A.: Insights into learning competence through probabilistic graphical models. In: Holzinger, A., Kieseberg, P., Tjoa, A.M., Weippl, E. (eds.) CD-MAKE 2019. LNCS, vol. 11713, pp. 250–271. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-29726-8_16
28. 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. Safety science **125**, 104616 (2020). https://doi.org/10.1016/j.ssci.2020.104616
29. Sarkar, S., Vinay, S., Raj, R., Maiti, J., Mitra, P.: Application of optimized machine learning techniques for prediction of occupational accidents. Comput. Operat. Res. **106**, 210–224 (2019)
30. Shapley, L.S.: A value for n-person games. Classics Game Theory **69** (1997)
31. Staniak, M., Biecek, P.: Explanations of model predictions with live and breakdown packages. arXiv preprint arXiv:1804.01955 (2018)

32. Štrumbelj, E., Kononenko, I.: Explaining prediction models and individual predictions with feature contributions. Knowl. Inf. Syst. **41**(3), 647–665 (2013). https://doi.org/10.1007/s10115-013-0679-x
33. Tamascelli, N., Solini, R., Paltrinieri, N., Cozzani, V.: Learning from major accidents: A machine learning approach. Comput. Chem. Eng. **162**, 107786 (2022)
34. Tsioras, P.A., Rottensteiner, C., Stampfer, K.: Wood harvesting accidents in the Austrian state forest enterprise 2000–2009. Safety Sci. **62**, 400–408 (2014)
35. Venkat, A., KP, G.V., Thomas, I.S., et al.: Machine learning based analysis for road accident prediction. IJETIE **6**(2) (2020)
36. Zheng, A., Casari, A.: Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists. " O'Reilly Media, Inc." (2018)
37. Zink, K.J.: Designing sustainable work systems: the need for a systems approach. Appl. Ergonom. **45**(1), 126–132 (2014). https://doi.org/10.1016/j.apergo.2013.03.023
38. Österreichische Bundesforste: Insights (2020),https://www.bundesforste.at/fileadmin/bundesforste/Zahlen___Fakten/2020/Oesterreichische_Bundesforste_Folder_Einblicke_EN_200630_Screen.pdf