



# TIME SERIES FORECASTING OF OCCUPATIONAL INJURY RISK

GIUSEPPE COCOMAZZI

THESIS SUBMITTED IN PARTIAL FULFILLMENT  
OF THE REQUIREMENTS FOR THE DEGREE OF  
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY  
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES  
OF TILBURG UNIVERSITY

**STUDENT NUMBER**

547014

**COMMITTEE**

dr. Stijn Rotman  
dr. Travis Wiltshire

**LOCATION**

Tilburg University  
School of Humanities and Digital Sciences  
Department of Cognitive Science &  
Artificial Intelligence  
Tilburg, The Netherlands

**DATE**

December 30, 2025

**WORD COUNT**

4782

**ACKNOWLEDGMENTS**

# TIME SERIES FORECASTING OF OCCUPATIONAL INJURY RISK

GIUSEPPE COCOMAZZI

## Abstract

### 1 DATA SOURCE, ETHICS, CODE, AND TECHNOLOGY STATEMENT

The SIR OSHA dataset is publicly available at [Severe Injury Dashboard | Occupational Safety and Health Administration](#). The data is anonymised. Work on this thesis did not involve collecting data from human participants or animals. The original owner of the data and code used in this thesis retains ownership of the data and code during and after the completion of this thesis.

All the figures belong to the author. In terms of writing, a generative language model, OpenAI's ChatGPT 5.1, was used to improve the author's original content, for paraphrasing, spell checking and grammar, and help debug LaTex syntax. The same model assisted in writing and documenting the code, included the function to fetch data from the U.S. Bureau of Statistics. No other typesetting tools or services were used.

### 2 PROBLEM STATEMENT & RESEARCH GOAL

#### 2.1 *Context*

Occupational safety and health (OSH) organizations across the world help foster the well-being of workers by preventing workplace accidents, promoting safe practices, and enforcing regulatory standards. Employers are legally responsible for ensuring a healthy working environment, as employees may be exposed to a wide range of physical, biological, and psychosocial hazards ([Howe et al., 2024](#); [Vassiley, Barratt, Dayaram, & Burgess, 2025](#); [Vitrano & Micheli, 2024](#)).

According to the International Labour Organization (ILO), an estimated 364 million workers globally suffer non-fatal occupational accidents each

year, with substantial geographical disparities in incidence rates. Injured workers often face long-term health complications, reduced employability, and financial hardship.

Beyond its societal implications, workplace safety has a considerable economic dimension. It is widely recognized that improving workplace safety can reduce economic losses at both firm and national levels. ILO estimates suggest that almost an annual 4% loss in global Gross Domestic Product is accountable to occupational accidents and diseases ([Organization, 2020](#)).

Traditional prevention measures to reduce occupational injury risk include compliance with safety standards and regulatory frameworks; inspections and audits; environmental controls and maintenance schedules; safety education programs, and post-incident analysis. Despite the efforts, manual reporting increasingly struggles to keep up with both rapidly evolving working conditions and the increasing volume of historical safety data. Post-hoc interventions are no longer sufficient for effective prevention.

For these reasons, real-time responsiveness has become crucial to further mitigate workplace injury risks. The integration of the Internet of Things technologies into the workplace allows for a continuous streaming of data coming from sensors, wearables, and smart equipment, supporting constant monitoring of safety conditions and triggering timely preventive actions ([Michaels & Wagner, 2025](#)).

From a scientific point of view, this complex data flow may pose a challenge to traditional time series methods such as Error, Trend, Seasonal models (ETS), autoregressive integrated moving average (ARIMA) and its seasonal extension (SARIMA), which have long served as standard baselines for forecasting tasks in the public-health domain. These models provide statistically grounded forecasts that often perform well in settings with clear seasonal structure and relatively smooth temporal dynamics ([Spiliotis, 2023](#)). On the other hand, Machine learning (ML) methods can leverage rich and dynamical data sources to capture complex, non-linear patterns, thereby marking a shift from a reactive to a proactive safety management strategy ([Kontopoulou, Panagopoulos, Kakkos, & Matsopoulos, 2023](#); [Masini, Medeiros, & Mendes, 2023](#)). Because the temporal factor will be the main leading axis in our analysis, we will first explore the data to extract components such as trend and seasonality to guide our model selection. Additionally, we will explore the geographical distribution as well to test whether encoding states at different levels can enhance the interpretability of the models and improve the forecasts both at the local and the global level (respectively, state and national level).

While most studies have focused on classification tasks, the area of time series forecasting for occupational injury prevention remains still

underexplored, despite its potential to anticipate future risk patterns and support more targeted interventions.

## 2.2 Research strategy

As outlined in subsection 2.1, knowing *when* an intervention may be required is crucial for increasing the likelihood of alleviating workers' exposure to severe injuries. If meaningful patterns in historical data can be reliably captured by forecasting algorithms, stakeholders can make more informed decisions about how to prioritize preventive resources, and about which types of data need continuous monitoring. As different temporal granularities can impact the accuracy of the generated forecasts (Rostami-Tabar, Goltsova, & Wang, 2023), the first question we aim to explore is to what extent different levels of temporal resolution can help smooth out an intermitting time series at the original lead time. Moreover, this question gains a further dimension when the data is organised according to a grouping factor, such as having time series across multiple states (Qu, Timmermann, & Zhu, 2024). The model choice will take into account the explainability of the models.

**RQ1:** *How does the forecasting performance of a set of classical statistical models (Holt-Winter method, ARIMA, and SARIMA) and ML models (divided into two main families: linear and tree-based ensemble models) at two temporal aggregations (monthly vs. weekly) in one-step-ahead forecasting of severe injuries risk across a twelve-month horizon, as evaluated by RMSE and MAE?*

This question addresses to what extent the temporal resolution affects the model performance when data are aggregated at a national level, yielding a single time series. Weekly aggregation increases the number of observations but it may introduce noise and variability, while monthly aggregation may produce coarser predictions. Understanding this trade-off is essential for designing reliable predictive pipelines (Barhrhouj, Ananou, & Ouladsine, 2025).

Given the absence of a comparable supervised regression setting in prior work on occupational injury forecasting, we will train different families of regressors based on the exploratory initial step and on the main findings from related datasets. The choice for linear models with penalty terms (Ridge, Lasso, ElasticNet) allows us to assess how regularization can affect feature selection in a strongly seasonal data structure across a balanced panel (Han, Zhou, Sun, & Li, 2025). Among tree-based ensemble models, CatBoost is deemed particularly suited for our panel data, as it can handle categorical features (Hall & Rasheed, 2025). XGBoost and LightGBM

appears to consistently be the most accurate models in recent forecasting benchmarks and machine-learning competitions ([Makridakis, Spiliotis, & Assimakopoulos, 2022](#)). The sample size and a low-dimensional feature matrix allowed us to compare a wide variety of models with negligible computational cost.

**RQ2:** *How does the predictive accuracy change when estimated as a Global Forecasting Model (across all states jointly) rather than as a Local Forecasting Model (state-specific), for both monthly and weekly temporal resolutions?*

With this question we evaluate whether pooling information across states enhances the predictive performance relative to modeling each state independently. A global model may exploit shared temporal knowledge and increase the available sample size, while a local model may better capture state-specific patterns. This question is closely related to the literature on panel data discussed in [3.3](#).

*SQ2.1: How does the inclusion of an exogenous categorical variable, namely the industry code, affect the forecasts of the ML models?*

Some industry sectors may explain better a given amount of the variability in the distribution of hospitalizations, as severe injuries occur more often in e.g. construction sites. To answer this question, we will explore feature importances to quantify how much each feature contributes to the predictive power of the model.

**RQ3:** *How sensitive are the forecasting results when cross-validated on different training and test splits?*

Forecasting performance estimates can be highly sensitive to the validation design, particularly in time-series settings where temporal dependence violates the assumptions of standard cross-validation. To assess the robustness of the results, we conduct a sensitivity analysis based on a rolling-origin cross-validation scheme.

### 2.3 Societal Relevance

OSH agencies must strategically allocate their resources to maximize injury reduction. Severe injuries, in particular, can pose an even higher burden for both employees and the society at large. Forecasting the relative risk of these life-altering events in the area where they will most likely occur can

provide actionable insights to guide OSH agencies on how to better target their interventions.

Recent evidence underscores the potential of such data-driven targeting strategies. As [Johnson, Levine, and Toffel \(2023\)](#) suggest, a machine learning approach to prioritize inspections could have averted twice as many injuries compared to the current OSHA prevention programs. Building on this premise, the present study focuses on forecasting relative risk trajectories across U.S. states to support more informed and proactive resource deployment.

#### 2.4 *Scientific Relevance*

A debated topic in the forecasting literature concerns to what extent global models – trained on pooled cross-sectional time-series – can outperform local models, which are trained independently for each unit (e.g., state, firm, or region). We systematically test whether a global pooling strategy improves forecasting accuracy compared to a local strategy when applied to a real-world panel dataset. Multiple ML regressors will be trained and compared under both frameworks, providing empirical evidence on the benefits and limitations, if any, of pooling information across states. Additionally, by framing the problem as a supervised regression task, we incorporate lagged features, rolling statistics, and exogenous variables, in line with the best practice in time series forecasting with ML regressors ([Bojer, 2022](#); [Makridakis et al., 2022](#)).

To the best of our knowledge, this is the first study to comparatively assess global versus local forecasting strategies for occupational injury risk prediction. More specifically, although time series forecasting is well established in adjacent health-related domains such as epidemiology and biostatistics, the regression-based framing of our study extend methodological practices common in these scientific fields to occupational safety research ([Reich et al., 2019](#)).

### 3 RELATED WORK

#### 3.1 *Occupation Injury Risk Prediction*

The field of occupation injury risk prediction using ML models has been largely dominated by classification tasks. In their comprehensive survey, [Vivian, Bauder, and Khoshgoftaar \(2025\)](#) identified three main research directions: (i) identifying workplace injury risk factors, (ii) predicting return-to-work rates, and (iii) analyzing sociodemographic features across different worker populations. Most studies rely on business-level datasets

or focus on specific injury types such as traumatic brain injuries or foot and ankle injuries.

The ML models most frequently used in this field are selected for their ability to capture patterns in heterogeneous sets of categorical variables. While no single model clearly outperforms the others, Random Forest, Support Vector Machines, Decision Trees, and Linear Regression have achieved the best results across evaluation metrics such as accuracy, sensitivity, and F-1 score, underscoring the importance of tailoring model selection to data characteristics. However, regression-based approaches that model the temporal evolution of occupational risk remain virtually absent from the literature. NOTE: should I find some references in a related field?

The work most closely aligned with our research is [Cerqua, Giannantoni, Letta, and Pinto \(2024\)](#). The authors develop a supervised regression framing for the time-series forecasting of workplace fatalities in Italy, focusing on spatial heterogeneity and ex-ante allocation of regulatory interventions. The methodology explicitly addresses how to apply machine learning pipelines to panel data, implementing a rolling forecasting origin approach for hyperparameter tuning, and evaluating model performance on a one-year hold-out test set. They selected a mix of models: linear, tree-based, and the Long Short-Term Memory neural network, ultimately identifying Partial Least Squares (PLS) as the best performing model across different sets of predictors.

### 3.2 *Studies on OSHA's SIR dataset*

A notable study examining the temporal aspects of occupational injury data is [Gomes, Parasram, Collins, and Socias-Morales \(2023\)](#). The authors analysed the Severe Injury Report (SIR) dataset provided by the Occupational Safety and Health Administration (OSHA), using time series decomposition models to uncover seasonal effects and general trends. Moreover, they evaluated industry and injury-type breakdowns, revealing that manufacturing and construction are the most affected industries. Importantly, they identified systematic temporal patterns in the occupational injuries, such as the predominance of injuries on weekdays compared to weekends, and a higher injury incidence in the summer months. While their analysis provided valuable descriptive insights into the temporal features of the SIR data, it did not extend to predictive or regression-based modeling. We will use these findings to inform model selection

On the classification side, [Khairuddin et al. \(2022\)](#) trained five machine learning models on the same SIR dataset to predict hospitalization and amputation outcomes, finding that Random Forest achieved the highest performance on metrics such as accuracy and F1-score. Building on this

work, Hasan Khalleefah Hassan and Khalifa (2025) evaluated a different set of machine learning models, and reported that the AdaBoost Classifier showed the best performance on accuracy and F1-score. As both studies aimed to categorize past events, their results and model selection can not serve as a benchmark for our work, but only as a point of reference. In fact, a regression-based forecasting approach applied to SIR data is still absent from the current literature. NOTE: maybe I should make clear that the experimental setup and the model selection for this project is based on the theoretical findings in the section below, much more than on the empirical ones on this specific dataset.

### 3.3 Theoretical Framework

When observations are collected over time from multiple entities—such as states, firms, or individuals—the resulting structure is referred to as *time-series cross-sectional data*, or more simply, *panel data*. Panel data can be *balanced*, when all variables are observed over the same time span for every unit, or *unbalanced*, when at least one unit (i.e., grouping factor) has missing observations for at least one time period.

A further, core distinction concerns how the model accounts for unit-specific, time-invariant components (in this study, the U.S. states included in the SIR dataset). In a *fixed-effects* model, each unit is assigned a fixed intercept that captures all characteristics that do not vary over time. In this framework, unobserved heterogeneity across units is not explicitly modeled; in other words, the effects of time-invariant variables—such as structural differences between states—are not estimated. Conversely, a *mixed-effects* (or *random-effects*) model treats unit-specific effects as random variables drawn from a common normal distribution. This allows the model to estimate their variance, thereby capturing how the time-invariant characteristics of one unit relate to those of others. Both approaches can be integrated within a *Hierarchical Linear Modeling* (HLM) framework, where fixed effects represent the average impact of predictors across the entire population, and random effects account for unit-specific deviations from this overall trend (Baltagi, 2013; Chen, 2021).

In our study, we aggregate monthly relative risks across all units to estimate the overall risk of hospitalization and amputation, while still allowing for state-level deviations by computing relative risks relative to each state's population rather than the national total. Although hierarchical modeling provides a more flexible framework, explicitly modeling between-state variation falls beyond the scope of the present research.

The question whether to implement a global as opposed to a local forecasting model has been extensively addressed by Montero-Manso and

[Hyndman \(2021\)](#). In their study, they found that global methods show good results not only in groups of similar time series, but even in heterogeneous groups. When a large number of time series is available, global models may generalize better with fewer parameters. Importantly, global models can afford much larger memory in terms of lags, being able to capture long-memory patterns that local models could learn only if manually engineered. We empirically test these findings by training different models, both linear and nonlinear ones, locally – fitting a single function to each cross-sectional time-series – and globally – by pooling all state-level series into a single learning problem.

Such a global model strategy raises the question of how panel data should be pooled. Panels can be combined on both the temporal as the time-invariant dimension, or along each axis separately. Appropriate handling of the panel structure during the temporal split of the data into a train and a test set can help prevent a wide range of potential data leakage ([Cerqua, Letta, & Pinto, 2024](#)).

Finally, hierarchical forecasting methods are designed to produce forecasts at multiple levels of aggregation. In our case, we could forecast total hospitalizations at a national and a state level, and with enough data at a county level. Hierarchical approaches reconcile forecasts so that the sum of the disaggregated predictions equal the aggregate forecast. However, we excluded these methods because the target variables are relative risks, which are not additive across hierarchical levels ([Hyndman & Athanasopoulos, 2018](#)).

### 3.4 Literature Gaps & Contribution

From the above review we can conclude that the SIR dataset has been used to gain descriptive insights into the temporal dimension of the injury report as well as to classify injury severity based on multiple categorical features. However, much remains unanswered when it comes to how severe injuries evolve over time and across states from a forecasting perspective. A truly forecasting framing is thus worthy of being developed, as it could improve risk monitoring and enable decision makers on a proactive management of prevention resources. Moreover, we contribute to the discussion in the forecasting community about global versus local modelling of panel data by providing an empirical comparison of both approaches in a novel domain, using a unified feature set and consistent evaluation metrics.

#### 4 METHODS

As discussed in subsection 3.3, a crucial methodological consideration for panel time-series data is whether to fit a single function to all series or to repeatedly fit a function to each series. In panel data, grouping factors are referred to as *cross-sectional units*. In this study, each state is observed over the same time span and shares the same features as the remaining units. Therefore, in the SIR dataset a single time series at a state level is a cross-sectional unit.

In a Global Forecasting Model cross-sectional units are assumed to be generated from the same data-generating process. In a Local Forecasting Model each time series is independent (Hewamalage, Bergmeir, & Bandara, 2022). In the following subsections, we illustrate how we constructed the dataset in a way suitable to both time-series forecasting frameworks.

Another key distinction concerns the forecasting strategy. Traditionally, most researchers compared statistical methods against a one-step ahead forecast horizon, where the forecaster predicts the next step in a time series. Moreover, classical models such as ARIMA and Vector Autoregression (VAR) are able to generate recursive multi-step forecasts natively, thereby lowering the need to target the multi-step task with dedicated forecasting strategies. With the advent of machine learning methods for time series forecasting, researchers must design explicit forecasting strategies, as those models lack the built-in capacity to roll forward recursive predictions. Under the recursive strategy, a model generates a one-step ahead forecast which is fed back to the historical window, and used to recompute all the features, before the model can generate a new one-step ahead forecast. This process is repeated until the final time point in the horizon is reached.

In their seminal review, Taieb, Bontempi, Atiya, and Sorjamaa (2011) formalized five strategies commonly used for multi-step forecasting: (i) recursive, (ii) direct, (iii) direct-recursive hybrids, (iv) multi-input multi-output, and (v) direct multi-output. The most relevant strategy for answering our research questions is the direct strategy. Because our models are built under different assumptions - linear, tree-based, and artificial networks - the direct approach allows for comparability by deploying the same pipeline under the same evaluation paradigm.

In fact, under the direct approach, a separate model is independently trained to predict a single output for each horizon. For instance, forecasting four horizons (e.g.,  $h = 1, 3, 6, 12$ ) requires fitting four separate models, one per horizon. Unlike the recursive strategy, the direct approach does not use forecasts as inputs in the historical window. As a consequence, errors do not compound at each next step, overcoming the risk of bias accumulation over longer horizons.

As a further motivation, the data complexity of the SIR data set is estimated to be sufficiently low to run separate models at a reasonable computation time. One of the drawbacks of the direct approach is the computational burden, since a separate model has to be fitted for every forecast horizon. experimental setup is missing: models configuration; train-test split; possibly, cv strategy,

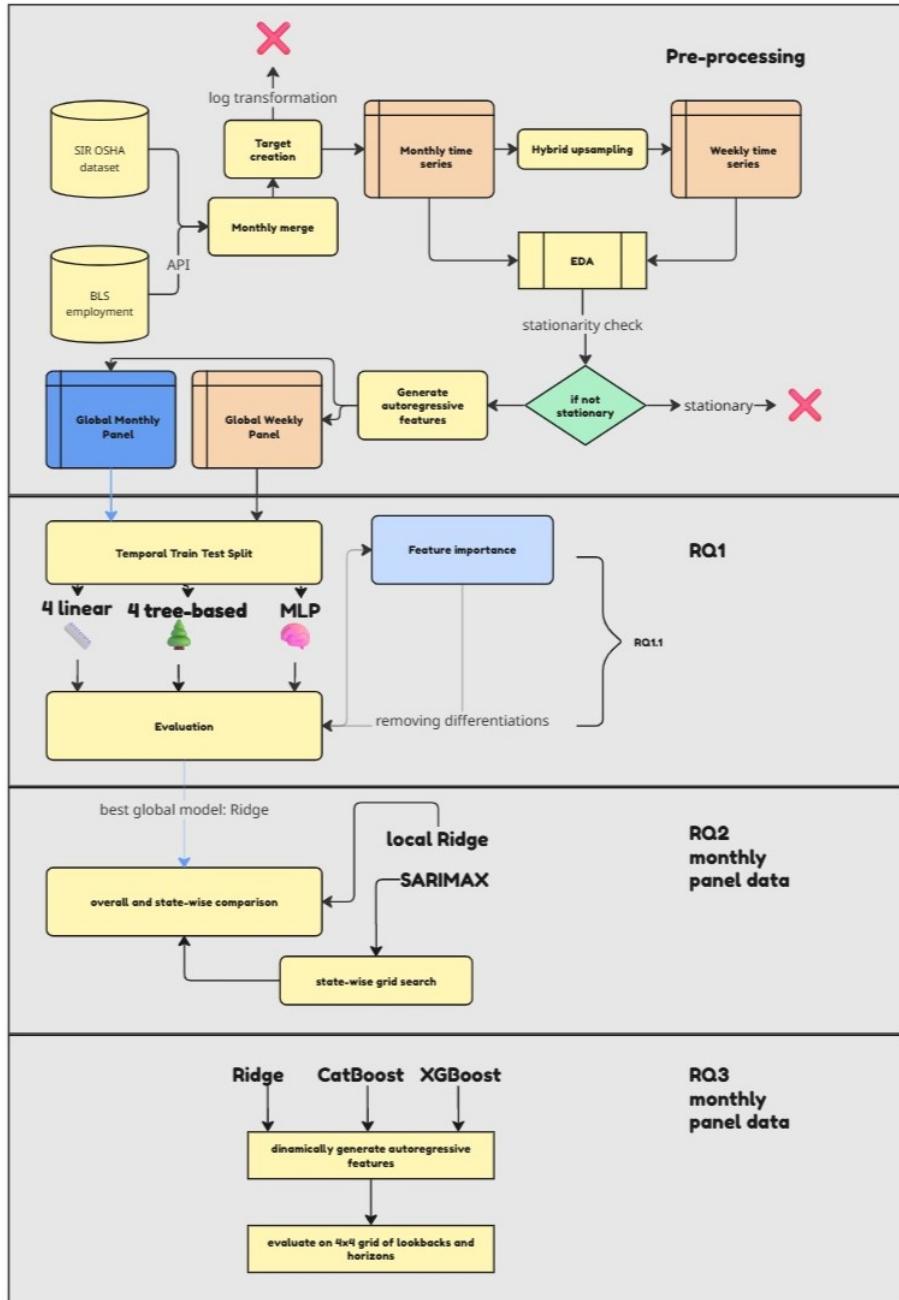


Figure 1: Workflow

#### 4.1 Dataset Description

The dataset is publicly available through the OSHA Severe Injury Dashboard<sup>1</sup>. OSHA collects reports about severe injuries occurred within U.S. states under its jurisdiction. These reports form the Severe Injury Reports (SIR) dataset. There are 34 states with a federal plan under OSHA jurisdiction, although the SIR dataset also contains reports from non-federal-plan states when the injured employees are federal workers. In addition, seven states have state plans covering only state and local government workers, while private-sector workers are covered by federal plans. To ensure a fair comparison between states, we included only states operating under a federal OSHA jurisdiction, together with the seven states mentioned above, as all reports from those states are collected from the private sector.

The dataset contains 98,801 entries, spanning from January 1, 2015 to February 28, 2025. Each entry represents a report submitted by an employer. A single report may mention multiple severe injuries (e.g. one hospitalization and one amputation), as employers are required to specify "*the number of employees who suffered a fatality, in-patient hospitalization, amputation, or loss of an eye*"<sup>2</sup>. This means that the feature used to construct the target series (Hospitalization) is a count variable, and that we can not aggregate the data at the report level.

To validate the target feature, we cross-tabulated Hospitalization and Amputation counts in Table 1 and we performed a qualitative analysis of some edge cases. In particular, we studied the *Final Narrative* feature for a sample of accidents where the employer reported both an amputation and a hospitalization. Even in these cases, the record involves only one employee; whereas accidents with multiple hospitalizations do refer to multiple employees. Even if this reporting inconsistency does not significantly impact the hospitalization size, we have to explicitly define our outcome variable as *prevalence of reported inpatient hospitalization*.

Table 1: Cross-tabulation of hospitalization and amputation indicators

	<b>Amputation</b>	<b>Hospitalized</b>						
		0	1	2	3	4	5	6
0		41	70,152	550	11	2	2	2
1		18,448	7,154	6	0	1	0	0
2		6	5	6	0	0	0	0

<sup>1</sup> <https://www.osha.gov/severe-injury-reports>. The dataset was retrieved on August 28, 2025.

<sup>2</sup> <https://www.osha.gov/laws-regulations/standardnumber/1904/1904.39>

The dataset provides 26 attributes, ranging from geographical information about the city, the address, and the state where the accident occurred, to details about the nature of the injury, the body part affected, the employer, and the business sector. Only variables required to construct state-level monthly injury counts and relative risks were retained.

To make our model more transparent for decision makers, we derived two new target variables from the two attributes Hospitalization and Amputation. Since these two attributes represent only absolute numbers, they may vary depending on the worker population. More populated states will probably report a larger number of severe injuries than states with a small workforce. Even within the same state, workforce trends over time can enormously impact the number of accidents. As an example, the COVID period saw a dramatic shrinking of the workforce, possibly leading to a number of severe injuries much lower than in any other period. Moreover, as we aim to evaluate both a global and a local forecasting strategy, the range of the outcome variable should be on the same scale. For this reason, we normalize the raw count of hospitalization with the workforce for each state and for each month included in the dataset.

To achieve interstate comparability, data from the Bureau of Labor Statistics (BLS) were fetched about seasonally adjusted, monthly employment at a state level. Employment is estimated on nonfarm, payroll jobs by the Current Employment Statistics (CES) survey. The covered time period spans from January 1, 2015 to December 31, 2024. Three states, namely American Samoa, Guam, and Northern Mariana Islands, were excluded from the panel since they are not covered by the BLS. After inspecting the testing set, one more U.S. territory, Virgin Islands, was excluded because of few data points. The ratio of both target features over the monthly employment was separately computed for each state. Therefore, the target to be predicted is not the absolute count of injuries per state, but the relative risk of getting either hospitalized or amputated per 100,000 state employees.

---

#### 4.2 Data Cleaning and Preparation

To evaluate for forecastability, we computed the percentage of days with no reports over the total temporal range for each state (mean value across all states=0,62). Given the data sparsity at the daily level, we aggregated the data to a lower resolution. The main preparation step involved aggregating both hospitalization and amputation risks at the monthly level to ensure a consistent merge with the monthly state employment statistics from the Bureau of Labor Statistics (BLS). After merging the two datasets, some missing values originated from a mismatch between the SIR time period

(ending in February 2025) and the BLS time period (ending in December 2024). These entries were removed from the panel.

For the weekly forecasting task, the state employment data were upsampled from monthly to weekly frequency using a hybrid strategy. Missing weekly values were interpolated with the exception of the period from March to September 2020. During this interval, the COVID-19 pandemic caused an abrupt drop in employment that the interpolation method was not able to capture properly. For this particular period, each initial monthly value was forward-filled until the next observed monthly value. Upsampling monthly employment data to weekly frequency assumes relatively smooth workforce trends within each month.

#### 4.3 Exploratory Data Analysis

A skewed data distribution of the target feature may lead to poor generalization, as the time series can be sensitive to extreme values. Figure 2 shows the counts of hospitalization and amputation before adjusting for the workforce size. The hospitalization distribution is slightly positively skewed.

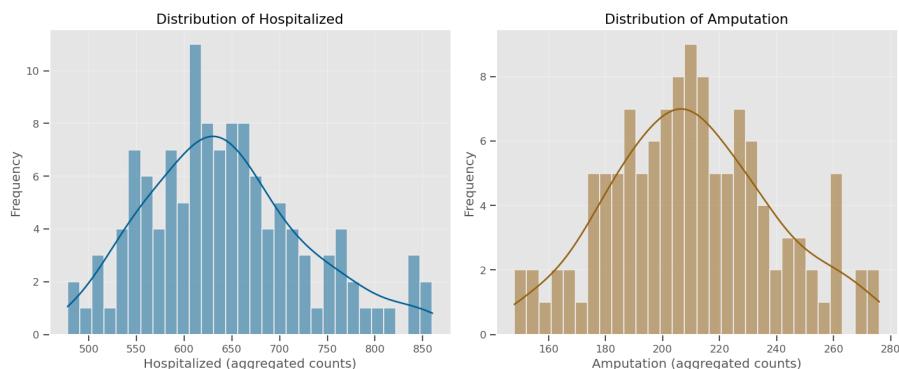


Figure 2: Distributions of hospitalizations and amputations (raw counts)

After adjusting for the workforce, the hospitalization risk distribution is shown in Figure 3 and its normality was assessed with a Shapiro-Wilk test, which rejected the null hypothesis of normality ( $W = 0.96$ ,  $p < 0.001$ ).

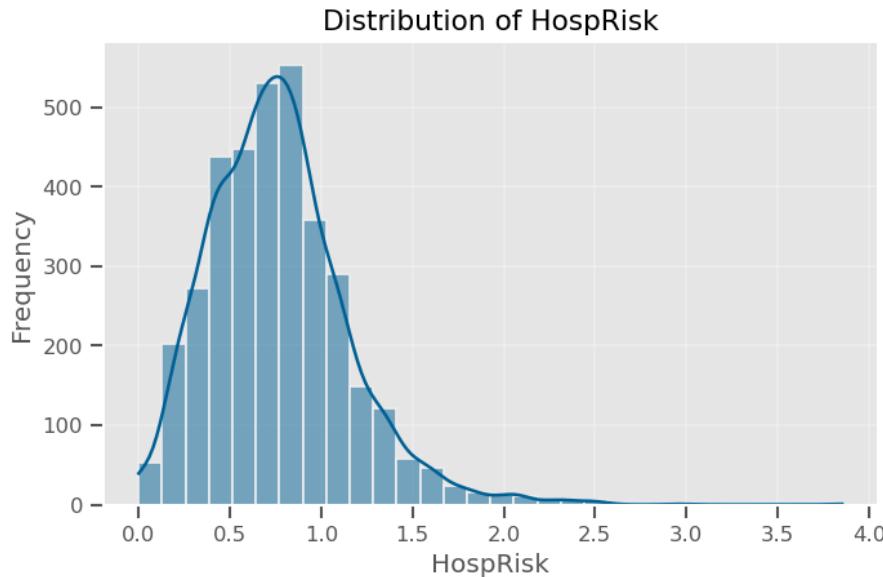


Figure 3: Distribution of hospitalizations adjusted for state workforce

A common approach to normalize a skewed distribution is to apply a log transformation to the target feature. However, because our target is already normalized by workforce size and because the target naturally included zero values (days with no report at all), we do not apply further transformations to our target (cit. ealing with Logs and Zeros in Regression Models).

To assess time serie components, we first visually inspect the mean monthly time series for hospitalization risk across all states in the panel. Figure 4 shows seasonal fluctuations, with regular upwards trends in the summer and a general decline after the COVID period.

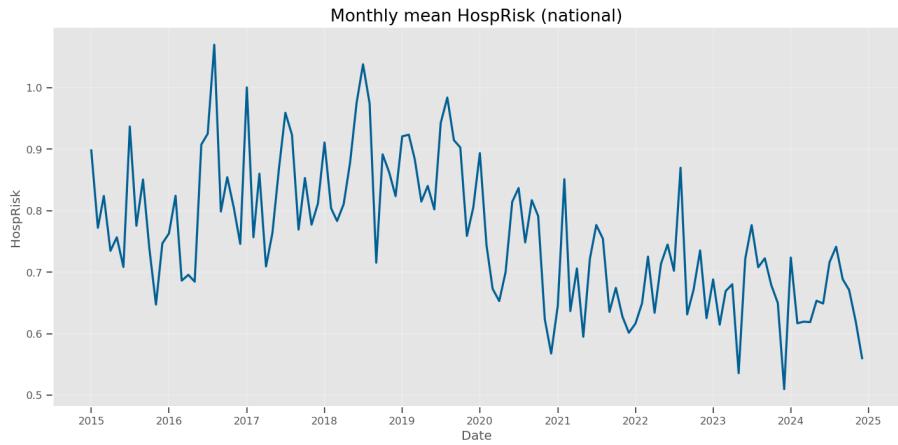


Figure 4: Monthly mean hospitalization risk

A visual inspection further confirms the seasonal components of the time series at a state level, with higher volatility when states with a lower workforce are interested by occasional spikes, as it is clear from Figure ??, which shows three states selected according to their relative workforce size.

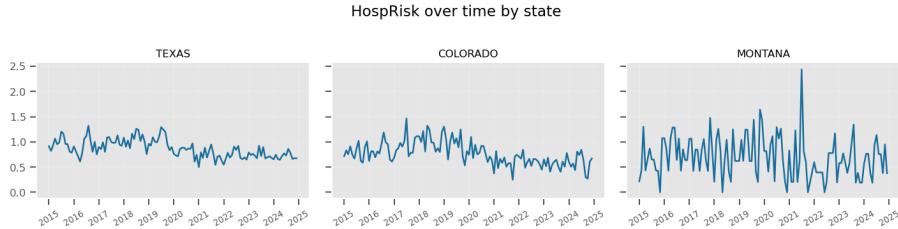


Figure 5: Monthly hospitalization risk for Texas, Colorado, and Montana

#### 4.4 Feature Engineering

In this crucial step, each time series was provided with exogenous variables. Although it is not strictly necessary for a ML model to be explicitly informed about data patterns, exogenous variables can still benefit the model's capabilities to predict temporal values. This is a viable option when the data complexity is low, the data are likely to be properly fitted by a linear model, and seasonal patterns can be detected from a visual inspection of the time series. The exogenous variables are divided into four categories:

Table 2: Overview of target variables, autoregressive features, and exogenous features used in the monthly forecasting models.

Category	Variables
Target variable	Hospitalization: sum of hospitalizations for each state at a monthly frequency.
Calendar and seasonal features	Year, Month, Quarter, Week of the Year.
Lagged features	lag1, lag2, lag3, lag6, lag12.
Rolling features	3-month rolling mean: roll3, 6-month rolling mean: roll6, 12-month rolling mean: roll12, 3-, 6-, 12-month exponential weighted mean: ewm3, ewm6, ewm12.
Exogenous variables	State (categorical identifier, one-hot encoded in the models), top k NAICS 2-digit industry: NAICSmix

## 5 RESULTS

## 6 DISCUSSION

### 6.1 Answering Research Questions

### 6.2 Limitations and further research

## 7 CONCLUSION

## REFERENCES

- Baltagi, B. H. (2013). Chapter 18 - Panel Data Forecasting. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (Vol. 2, pp. 995–1024). Elsevier. Retrieved from <https://www.sciencedirect.com/science/article/pii/B978044462731500018X> (ISSN: 1574-0706) doi: <https://doi.org/10.1016/B978-0-444-62731-5.00018-X>
- Barhrhouj, A., Ananou, B., & Ouladsine, M. (2025). Assessing the Impact of Temporal Data Aggregation on the Reliability of Predictive Machine Learning Models. In V. Julian et al. (Eds.), *Intelligent Data Engineering and Automated Learning – IDEAL 2024* (Vol. 15346, pp. 481–492). Cham: Springer Nature Switzerland. Retrieved 2025-12-22, from [https://link.springer.com/10.1007/978-3-031-77731-8\\_43](https://link.springer.com/10.1007/978-3-031-77731-8_43) (Series Title:

- Lecture Notes in Computer Science) doi: 10.1007/978-3-031-77731-8  
–43
- Bojer, C. S. (2022, October). Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities. *International Journal of Forecasting*, 38(4), 1555–1561. Retrieved 2025-11-17, from <https://linkinghub.elsevier.com/retrieve/pii/S0169207021001771> doi: 10.1016/j.ijforecast.2021.11.003
- Cerqua, A., Giannantoni, C., Letta, M., & Pinto, G. (2024, December). '*Dead Man Working': A Place-based Approach to Workplace Fatalities* [SSRN Scholarly Paper]. Rochester, NY: Social Science Research Network. Retrieved 2025-09-08, from <https://papers.ssrn.com/abstract=5040905> doi: 10.2139/ssrn.5040905
- Cerqua, A., Letta, M., & Pinto, G. (2024). *On the (Mis)Use of Machine Learning with Panel Data*. SSRN. Retrieved 2025-11-07, from <https://www.ssrn.com/abstract=5014594> doi: 10.2139/ssrn.5014594
- Chen, J. M. (2021, February). An Introduction to Machine Learning for Panel Data. *International Advances in Economic Research*, 27(1), 1–16. Retrieved 2025-11-07, from <https://doi.org/10.1007/s11294-021-09815-6> doi: 10.1007/s11294-021-09815-6
- Gomes, H., Parasram, V., Collins, J., & Socias-Morales, C. (2023, September). Time series, seasonality and trend evaluation of 7FIX ME!!!!years (2015–2021) of OSHA severe injury data. *Journal of Safety Research*, 86, 30–38. Retrieved 2025-09-14, from <https://www.sciencedirect.com/science/article/pii/S0022437523000798> doi: 10.1016/j.jsr.2023.06.005
- Hall, T., & Rasheed, K. (2025, January). A Survey of Machine Learning Methods for Time Series Prediction. *Applied Sciences*, 15(11), 5957. Retrieved 2025-11-17, from <https://www.mdpi.com/2076-3417/15/11/5957> (Publisher: Multidisciplinary Digital Publishing Institute) doi: 10.3390/app15115957
- Han, C., Zhou, Y., Sun, J., & Li, Z. (2025, November). An optimized ridge regression for forecasting time series with a fixed period. *Pattern Recognition Letters*, 197, 274–281. Retrieved 2025-12-22, from <https://www.sciencedirect.com/science/article/pii/S0167865525002971> doi: 10.1016/j.patrec.2025.08.017
- Hasan Khalleefah Hassan, M., & Khalifa, W. M. S. (2025). Work Place Safety: Machine Learning Techniques for Assessing Workplace Incident Severity. *IEEE Access*, 13, 34211–34226. Retrieved 2025-09-10, from <https://ieeexplore.ieee.org/abstract/document/10891569> doi: 10.1109/ACCESS.2025.3543136
- Hewamalage, H., Bergmeir, C., & Bandara, K. (2022, April). Global models

- for time series forecasting: A Simulation study. *Pattern Recognition*, 124, 108441. Retrieved 2025-11-07, from <https://www.sciencedirect.com/science/article/pii/S0031320321006178> doi: 10.1016/j.patcog.2021.108441
- Howe, A. S., Tan, J., Yuen, B., Saini, H., Saade-Cleves, N., Obeidat, D., ... Nowrouzi-Kia, B. (2024, August). Physical and Psychosocial Correlates of Occupational Physical Injury in the Global Construction Industry: A Scoping Review. *Environmental Health Insights*, 18, 11786302241270371. Retrieved 2025-12-24, from <https://pmc.ncbi.nlm.nih.gov/articles/PMC11345736/> doi: 10.1177/11786302241270371
- Hyndman, R. J., & Athanasopoulos, G. (2018). Forecasting: Principles and Practice. Retrieved 2025-11-29, from <https://research.monash.edu/en/publications/forecasting-principles-and-practice-2/> (Publisher: OTexts)
- Johnson, M. S., Levine, D. I., & Toffel, M. W. (2023, October). Improving Regulatory Effectiveness through Better Targeting: Evidence from OSHA. *American Economic Journal: Applied Economics*, 15(4), 30–67. Retrieved 2025-11-06, from <https://pubs.aeaweb.org/doi/10.1257/app.20200659> doi: 10.1257/app.20200659
- Khairuddin, M. Z. F., Lu Hui, P., Hasikin, K., Abd Razak, N. A., Lai, K. W., Mohd Saudi, A. S., & Ibrahim, S. S. (2022, October). Occupational Injury Risk Mitigation: Machine Learning Approach and Feature Optimization for Smart Workplace Surveillance. *International Journal of Environmental Research and Public Health*, 19(21), 13962. Retrieved 2025-09-08, from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9653932/> doi: 10.3390/ijerph192113962
- Kontopoulou, V. I., Panagopoulos, A. D., Kakkos, I., & Matsopoulos, G. K. (2023, August). A Review of ARIMA vs. Machine Learning Approaches for Time Series Forecasting in Data Driven Networks. *Future Internet*, 15(8), 255. Retrieved 2025-11-29, from <https://www.mdpi.com/1999-5903/15/8/255> (Publisher: Multidisciplinary Digital Publishing Institute) doi: 10.3390/fi15080255
- Makridakis, S., Spiliotis, E., & Assimakopoulos, V. (2022, October). M5 accuracy competition: Results, findings, and conclusions. *International Journal of Forecasting*, 38(4), 1346–1364. Retrieved 2025-11-29, from <https://www.sciencedirect.com/science/article/pii/S0169207021001874> doi: 10.1016/j.ijforecast.2021.11.013
- Masini, R. P., Medeiros, M. C., & Mendes, E. F. (2023, February). Machine learning advances for time series forecasting. *Journal of Economic Surveys*, 37(1), 76–111. Retrieved 2025-11-07, from <https://onlinelibrary.wiley.com/doi/10.1111/joes.12429> doi:

- 10.1111/joes.12429
- Michaels, D., & Wagner, G. R. (2025, April). OSHA Injury Data: An Opportunity for Improving Work Injury Prevention. *American Journal of Public Health*, 115(4), 588–595. Retrieved 2025-09-04, from <https://ajph.aphapublications.org/doi/full/10.2105/AJPH.2024.307934> (Publisher: American Public Health Association) doi: 10.2105/AJPH.2024.307934
- Montero-Manso, P., & Hyndman, R. J. (2021, March). *Principles and Algorithms for Forecasting Groups of Time Series: Locality and Globality*. arXiv. Retrieved 2025-11-05, from <http://arxiv.org/abs/2008.00444> (arXiv:2008.00444 [cs]) doi: 10.48550/arXiv.2008.00444
- Organization, I. L. (2020). *Safety + health for all : an ILO Flagship Programme : key facts and figures (2016-2020)* - International Labour Organization. Retrieved 2025-11-29, from [https://labordoc.ilo.org/discovery/fulldisplay/alma995108293102676/41IL0\\_INST:41IL0\\_V2](https://labordoc.ilo.org/discovery/fulldisplay/alma995108293102676/41IL0_INST:41IL0_V2)
- Qu, R., Timmermann, A., & Zhu, Y. (2024, July). Comparing forecasting performance with panel data. *International Journal of Forecasting*, 40(3), 918–941. Retrieved 2025-11-08, from <https://www.sciencedirect.com/science/article/pii/S0169207023000766> doi: 10.1016/j.ijforecast.2023.08.001
- Reich, N. G., Brooks, L. C., Fox, S. J., Kandula, S., McGowan, C. J., Moore, E., ... Shaman, J. (2019, February). A collaborative multiyear, multimodel assessment of seasonal influenza forecasting in the United States. *Proceedings of the National Academy of Sciences of the United States of America*, 116(8), 3146–3154. Retrieved 2025-11-29, from <https://pmc.ncbi.nlm.nih.gov/articles/PMC6386665/> doi: 10.1073/pnas.1812594116
- Rostami-Tabar, B., Goltsos, T. E., & Wang, S. (2023, February). Forecasting for lead-time period by temporal aggregation: Whether to combine and how. *Computers in Industry*, 145, 103803. Retrieved 2025-12-20, from <https://www.sciencedirect.com/science/article/pii/S0166361522001993> doi: 10.1016/j.compind.2022.103803
- Spiliotis, E. (2023). Time Series Forecasting with Statistical, Machine Learning, and Deep Learning Methods: Past, Present, and Future. In M. Hamoudia, S. Makridakis, & E. Spiliotis (Eds.), *Forecasting with Artificial Intelligence: Theory and Applications* (pp. 49–75). Cham: Springer Nature Switzerland. Retrieved 2025-12-24, from [https://doi.org/10.1007/978-3-031-35879-1\\_3](https://doi.org/10.1007/978-3-031-35879-1_3) doi: 10.1007/978-3-031-35879-1\_3
- Taieb, S. B., Bontempi, G., Atiya, A., & Sorjamaa, A. (2011, August). *A review and comparison of strategies for multi-step ahead time series forecasting based on the NN5 forecasting competition*. arXiv. Retrieved 2025-11-13, from <http://arxiv.org/abs/1108.3259> (arXiv:1108.3259 [stat]) doi:

- 10.48550/arXiv.1108.3259
- Vassiley, A., Barratt, T., Dayaram, K., & Burgess, J. (2025, April). Psychosocial workplace hazards and industrial relations: An introduction. *Journal of Industrial Relations*, 67(2), 177–201. Retrieved 2025-12-24, from <https://doi.org/10.1177/00221856251326664> (Publisher: SAGE Publications Ltd) doi: 10.1177/00221856251326664
- Vitrano, G., & Micheli, G. J. L. (2024, May). Effectiveness of Occupational Safety and Health interventions: a long way to go. *Frontiers in Public Health*, 12, 1292692. Retrieved 2025-12-24, from <https://pmc.ncbi.nlm.nih.gov/articles/PMC11111875/> doi: 10.3389/fpubh.2024.1292692
- Vivian, G. A., Bauder, R. A., & Khoshgoftaar, T. M. (2025, July). A comprehensive survey on machine learning for workplace injury analysis: risk prediction, return to work strategies, and demographic insights. *Journal of Big Data*, 12(1), 167. Retrieved 2025-09-01, from <https://doi.org/10.1186/s40537-025-01229-z> doi: 10.1186/s40537-025-01229-z

#### APPENDIX A

(page 21).

Table 3: Hyperparameter configurations for all models

<b>Model</b>	<b>Hyperparameters</b>
Ridge Regression	$\alpha = 1.0$ , random_state = 0
Lasso Regression	$\alpha = 0.001$ , max_iter = 10,000, random_state = 0
Elastic Net	$\alpha = 0.001$ , l1_ratio = 0.5, max_iter = 10,000, random_state = 0
PLS Regression	n_components = 10
XGBoost	n_estimators = 300, learning_rate = 0.05, max_depth = 3, subsample = 1.0, colsample_bytree = 1.0, reg_lambda = 1.0, objective = reg:squarederror, random_state = 0, n_jobs = -1
LightGBM	n_estimators = 300, learning_rate = 0.05, max_depth = 3, num_leaves = 7, subsample = 1.0, colsample_bytree = 1.0, objective = regression, reg_lambda = 0.0, random_state = 0, n_jobs = -1, verbose = -1
CatBoost	iterations = 300, learning_rate = 0.05, depth = 3, l2_leaf_reg = 3.0, loss_function = RMSE, random_state = 0, verbose = False