processes

MDPI

Article

Coal Mine Accidents in China and the United States, 2011–2025: Harmonized Spatio-Temporal Patterns, Key Risk Factors, and Interpretable Machine Learning Severity Prediction from Public Accident Data

Shiwu Zhang [1] and Zeyu Feng [2,*]

[1] School of Mining Engineering, Shanxi Open University, Taiyuan 030024, China; zzzzwww1234@126.com
[2] College of Safety and Emergency Management Engineering, Taiyuan University of Technology, Taiyuan 030024, China
* Correspondence: feng18636976878@163.com

Abstract

The two largest coal producers, China and the United States, operate markedly different accident-reporting regimes, yet comparable cross-country evidence is scarce. We harmonized 47,986 U.S. Mine Safety and Health Administration (MSHA) Part 50 records for 2011–2025, documenting reportable injuries, occupational illnesses and accidents, with Chinese official statistics and a published 532-report structured dataset. Because the archives report different objects, and China reports fatalities per million tonnes while U.S. reporting is event-based, we placed both on a common production-normalized basis, tested correlates of fatal outcomes with chi-square statistics and Cramér's V and benchmarked four classifiers under explicit class-imbalance controls with SHAP attribution and leakage-controlled narrative text. China's rate fell from 0.564 (2011) to 0.045 (2025) while the U.S. rate stayed flat (0.020 to 0.017), narrowing the gap from 28-fold to under 3-fold. U.S. fatalities concentrated in powered haulage (31.2%) and machinery (25.9%); Chinese deaths were dominated by gas explosions and outbursts (20.8 and 9.6 per event). Scrubbed narratives raised cross-validated ROC–AUC from 0.892 to 0.966 and held-out PR–AUC from 0.049 to 0.252; a sensitivity analysis confirmed the composition findings under the Chinese dataset's severity bias. The framework supports screening-oriented risk triage of newly reported events, not automated adjudication.

Keywords: coal mine safety; accident severity prediction; interpretable machine learning; SHAP; accident report text mining; China–United States comparison

Check for updates

Academic Editor: Davide Papurello

Received: 1 September 2026
Revised: 21 September 2026
Accepted: 23 September 2026
Published: 28 September 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Coal still supplies a substantial share of primary energy in both China and the United States, and underground extraction remains exposed to roof failure, gas and dust explosions, outbursts, water inrush and powered-haulage hazards. The two countries anchor the global picture of coal mine safety: China produces roughly half of the world's coal, and the United States operates the longest-running mandatory accident-reporting system, the Part 50 program administered by the Mine Safety and Health Administration (MSHA) [1]. Both systems have improved markedly. China's official fatality rate per million tonnes of coal (FRPMT) fell from 0.564 in 2011 to a historic low of 0.044 in 2021 [2–5], and annual U.S. coal mine fatal events declined from 20 in 2011 to single digits in several recent years [1,6]. Yet

the two trajectories diverge in instructive ways: China's rate rebounded to 0.054 in 2022 and 0.094 in 2023 amid supply security-driven production pressure [7,8], before falling to 0.059 in 2024 and 0.045 in 2025 [5,7], while the U.S. trend flattened at a low level dominated by single-victim events [6].

A large body of work has examined accident statistics within a single country. For China, studies have characterized temporal and regional patterns of gas explosions [9–12], extraordinarily severe accidents [13,14], fires and explosions [15], roof accidents [16], water inrush [17], and human factor profiles of fatal events [18–20]. For the United States, retrospective analyses of MSHA data cover long-run disaster history [21] and panel regressions of injury rates [22]. Two limitations persist. First, cross-country comparisons are usually restricted to aggregate death tolls and fatality rates [23,24], because national archives differ in granularity, definitions and accessibility; event-level, harmonized comparisons remain rare. Second, the growing machine learning literature on accident severity [25–28] is dominated by single-country road traffic applications, and mining applications rarely address extreme class imbalance or open the black box, limiting regulatory usefulness.

Accident investigation reports also contain free-text narratives that encode causal circumstances absent from structured fields. Text mining of such reports has matured in construction [29] and in Indian coal mines [30], and keyword- or BERT-based extraction has been applied to Chinese coal mine reports [31], but narrative information is seldom fused with structured variables for severity prediction, and outcome leakage through narrative wording is rarely controlled.

To address these gaps, this study harmonizes event-level U.S. MSHA Part 50 coal records for 2011–2025 with Chinese official annual statistics and a published 532-report structured accident dataset and pursues three research questions. RQ1: how do the two countries differ in temporal trends, accident-type composition, and severity structure? RQ2: which mine, worker and situational factors are associated with fatal outcomes, and with what relative weight? RQ3: can machine learning models predict fatality risk accurately and interpretably, and do report narratives add incremental predictive signal? The contributions are threefold: (i) a transparent, auditable cross-country harmonization of public accident data with explicit scope definitions; (ii) a dual-track design that pairs descriptive inference (chi-square tests with Cramér's V effect sizes) with interpretable machine learning under extreme class imbalance (189 fatal events among 47,986 records), reported with precision–recall metrics appropriate to a 0.39% prevalence; and (iii) a leakage-controlled text-mining increment that quantifies what narratives add beyond structured fields. Section 2 reviews related work; Section 3 details data and methods; Section 4 reports results; Section 5 discusses mechanisms, implications and limitations; Section 6 concludes.

2. Literature Review

2.1. Data-Driven Accident Analysis and Severity Prediction

Statistical characterization of mine accidents has a long tradition in both countries. Chinese studies have documented the dominance of roof and gas accidents [9,23,32], the role of unsafe behavior in gas explosions [18,20], and decade-scale declines driven by consolidation of small mines and regulatory reform [8,14]. U.S.-focused work includes the CDC's disaster history 1900–2006 [21] and a 36-year generalized-estimating-equation panel of MSHA data linking mine type, tenure and injury characteristics to accident rates [22]. Predictive modeling of injury severity has expanded rapidly in road safety, where random forest, XGBoost and hybrid logit–ML designs are benchmarked with SHAP-based interpretation [25–27]; mining-specific applications include roof-accident risk assessment with ensemble models [16] and risk coupling analysis of gas explosions [33]. A recent structured dataset of 532 Chinese mine accident reports (2010–2025) demonstrated the

feasibility of full-element coding from official sources [24]. These studies establish that ensemble learners handle the nonlinear, interactive structure of accident data better than single regressions, but most report accuracy on balanced or moderately imbalanced targets, which does not transfer to fatality prediction where the positive class is below 1%.

2.2. Cross-Country Comparison of Mine Accidents

Direct cross-country analyses are few. Wang et al. [23] compared accident types across major coal producing countries using aggregated statistics and found that Chinese accidents concentrate in roof and gas categories while U.S. fatalities concentrate in falls, haulage and machinery. Fatality rate comparisons against production show China converging toward, but still above, U.S. and Australian levels [23,24]. Comparative work on regulatory regimes describes the U.S. inspectorate model with mandatory 15-min electronic reporting through the Mine Data Retrieval System [1] and China's evolution toward a national inspection, local regulation and enterprise responsibility system, strengthened by the 2024 Coal Mine Safety Production Regulation [34] and the 'eight hard measures' campaign against major accidents [35]. What is missing is an event-level comparison built on a unified scope, in which the same severity definitions and the same time window are applied to both archives, with composition differences expressed as effect sizes rather than narrative description.

2.3. Text Mining of Accident Investigation Reports

Investigation narratives have been mined with latent Dirichlet allocation (LDA), keyword extraction and transformer embeddings. In construction, LDA plus association-rule mining over 1223 Chinese investigation reports recovered 24 causal topics spanning personnel, equipment, management and environment [29]. For Indian coal mines, LDA with rapid automatic keyword extraction reproduced inspector-assigned accident classes and exposed the root-cause structure [30]. KeyBERT-based coupling analysis has mapped disaster-causing factor networks for six Chinese coal mine disaster types [31]. Pipeline-safety and aviation studies similarly show that narrative co-occurrence structure carries causal information that coded fields miss [36]. Two methodological issues remain under-addressed: narratives are written after the outcome is known, so uncorrected text features can leak the label into predictive models; and topic structures are rarely linked back to supervised severity models to quantify incremental value. Both issues are handled explicitly here.

2.4. Research Gap and Positioning

In contrast to single-country event-level studies [9,16,22], this work applies one scope definition and one analytical pipeline to two national archives. In contrast to aggregate cross-country comparisons [23], it operates at event level on the U.S. side and links Chinese composition statistics through an explicit harmonization table. In contrast to black box severity models [25,27], every predictive claim is paired with SHAP-based attribution and imbalance-appropriate metrics, and the text-mining increment is measured after scrubbing outcome-revealing tokens, a control absent from prior report-mining work [29–31].

3. Data and Methods

3.1. Data Sources and Scope

U.S. data were taken from the MSHA Open Government Data portal, Part 50 'Accidents' file (arlweb.msha.gov/OpenGovernmentData/DataSets/Accidents.zip; accessed 1 September 2026; file version dated 28 August 2026), which covers operator-reported accidents, injuries and occupational illnesses with quarterly updates and de-identified records [1]. All records flagged as coal (COAL_METAL_IND = 'C') in calendar years 2011–2025 were retained, yielding 47,986 event-level records, each uniquely identified by a

document number. A record is coded fatal when the injury degree field equals '01' (fatality); 189 records (0.39%) met this definition. Chinese data were assembled from three official or peer-reviewed public sources: (i) the National Bureau of Statistics Statistical Communiqués for 2011–2025, which publish the national coal mine FRPMT and, for most years, the death toll [3–5,7]; (ii) the National Mine Safety Administration (NMSA) and Ministry of Emergency Management releases on annual accident counts and policy campaigns [2,35]; and (iii) the tabulated composition of a peer-reviewed 532-report structured dataset of Chinese mine accidents (2010–2025) constructed from NMSA and provincial emergency-management publications [24]. Sources, fields, coverage and access dates are listed in Table 1. No personally identifiable information is present in any source: MSHA records are de-identified before release, and the Chinese sources are aggregate statistics or coded reports without named individuals. All retrieval used public download endpoints without bypassing access controls, and the processing pipeline is fully scripted for reproducibility (Section 3.7). Because only publicly available, de-identified secondary data were used and no human participants were recruited, the national measures for ethical review of life science and medical research involving human participants do not apply to this study [37].

Table 1. Data sources, coverage and access dates.

Data Stream	Source	Coverage	Level	Access Date
U.S. accident events	MSHA Open Government Data, Part 50 Accidents file	2011–2025, coal, n = 47,986	Event-level, de-identified	1 September 2026
U.S. context statistics	MSHA Mine Safety and Health At a Glance [6]	FY 2019–2025	Aggregate	1 September 2026
U.S. production series	EIA Quarterly Coal Report, Table 1 (compiled from MSHA Form 7000-2 operator returns) [38]	2011–2025	Aggregate, official	1 September 2026
China annual series	NBS Statistical Communiqués 2011–2025 [3–5,7]	National FRPMT; deaths	Aggregate, official	1 September 2026
China policy context	NMSA/MEM releases [2,35]	2019–2025	Official releases	1 September 2026
China composition	Published 532-report structured dataset [24]	2010–2025	Coded reports, published tables	1 September 2026

Two scope decisions require explicit justification. The window 2011–2025 is the longest period over which both archives are simultaneously usable: continuous event-level U.S. coverage in the form used here begins in 2011, the Chinese official FRPMT series covers the same years, and the published Chinese structured dataset spans from January 2010 to June 2025. Extending the window further back would break common coverage on the U.S. side, whereas shortening it would exclude the 2021–2023 rebound that is central to RQ1. The two archives also differ greatly in size (47,986 U.S. records against 532 Chinese reports) because they count different objects rather than different populations: a U.S. record documents one reportable injury, illness or accident, whereas the Chinese dataset is a curated collection of investigated accident reports, so the two counts are not comparable as sample sizes and no statistical power is claimed for the Chinese side. U.S. coal production was taken from the EIA Quarterly Coal Report, whose production series is compiled from the same MSHA Form 7000-2 returns that underpin the accident file [38], so the numerator and the denominator share a single reporting basis.

3.2. Cross-Country Harmonization

The two archives measure different objects: MSHA records every reportable injury, illness and accident, whereas Chinese official series report fatal accidents and deaths.

Harmonization therefore proceeded at three explicitly declared levels. Level 1 (rates): rather than comparing Chinese FRPMT with U.S. fatal-event counts, which are not equivalent, we recomputed the U.S. series on China's basis, expressing U.S. fatalities as deaths per million tonnes of coal from EIA production converted to metric tonnes (1 short ton = 0.9071847 t) [38]; both countries are therefore placed on one production-normalized scale before any comparison is drawn. A useful internal check follows from the per-individual structure of U.S. reporting: the 189 records flagged fatal correspond exactly, year by year, to MSHA's published annual coal mine fatality totals for 2011–2025 [1,6], so on the U.S. side 'fatal event' and 'death' coincide numerically in this window, and we reserve 'event' for any record in the file, fatal or not. Level 2 (composition): the MSHA accident classification shares of events and of fatal events were compared against the 532-report dataset's type shares [24], with the caveat that the latter describes investigated accidents (Section 4.7). Level 3 (event-level modeling): it was performed on the U.S. archive only, because Chinese event-level public records with comparable covariate depth are not available; this asymmetry is stated as a scope limitation rather than smoothed over. MSHA classifications were mapped to the Chinese ten-category taxonomy (roof, gas, transport, water, fire, electromechanical, blasting, outburst, rock burst, other) for qualitative comparison; the mapping table is provided in Supplementary Materials.

3.3. Descriptive and Inferential Analysis

Annual event and fatal-event counts, month-of-year and shift-period distributions, state-level concentration and classification shares were computed for the U.S. archive; FRPMT and death toll series were tabulated for China. Associations between candidate factors (mine subunit, accident classification, accident type, occupation, activity, injury source, underground location, mining method, state, shift period) and fatal outcome were tested with Pearson's chi-square; because significance is trivially attainable with n = 47,986, Cramér's V is reported as the effect size alongside *p*-values, and interpretation follows V rather than p. Two-sided $\alpha = 0.05$.

3.4. Severity Prediction Models

The prediction target was the fatal versus non-fatal outcome of U.S. coal mine events. Predictors were restricted to attributes knowable at or immediately after occurrence, before outcome ascertainment included calendar year, quarter, month and shift period; mine subunit, underground location and mining method; accident classification and type; occupation, activity and injury source; state; and total, mine-level and job-specific experience of the most junior involved worker, with missingness indicators (values median-imputed and clipped at 50 years). Post-outcome fields (nature of injury, injured body part, days lost) were excluded to prevent leakage. Categorical levels with low frequency were collapsed into an 'other' level and one-hot encoded, yielding 126 features. Four classifiers were benchmarked: logistic regression (class-weighted, standardized features), random forest (class-weighted, 300 trees, depth 14 [39]), gradient boosting (250 stages, sample-weighted) and XGBoost (500 rounds, depth 6, scale_pos_weight equal to the negative-to-positive ratio [40]), plus an XGBoost variant trained on SMOTE-resampled data (sampling ratio 0.3 [41]). Model selection used five-fold stratified cross-validation on an 80% train split; the held-out 20% test set was evaluated once per model. Given 0.39% prevalence, primary metrics are ROC–AUC and area under the precision–recall curve (PR–AUC); recall, precision, F1 and the Brier score are reported at the Youden threshold estimated from out-of-fold training predictions. Confusion matrices are reported in full so that false-positive burden is visible.

The imbalance is extreme and is handled in four explicit ways rather than by resampling alone. First, because the positive class is 0.39% of records (189 of 47,986), overall

accuracy is uninformative, as a null model already scores 99.6%; we therefore treat ROC–AUC and PR–AUC as primary metrics and always read PR–AUC against its 0.004 prevalence baseline, so that a value such as 0.049 is understood as roughly twelve times the no-skill level rather than as a failure. Second, imbalance is addressed inside every training fold: logistic regression and random forest use balanced class weights, gradient boosting and XGBoost use sample weights, and the XGBoost variant sets scale_pos_weight to the negative-to-positive ratio, so no weighting is informed by the validation fold. Third, SMOTE is applied only to the training partition of each fold and never to validation or test data, so synthetic minority samples cannot leak into evaluation. Fourth, the decision threshold is not tuned on the test set: it is the Youden point estimated from out-of-fold training predictions, and confusion matrices at that threshold are reported in full, keeping the false-positive burden visible. The consequence of the imbalance is reported rather than hidden: at any operationally useful recall, precision remains between 0.02 and 0.03, because a false-positive rate of only 2% already generates more false alarms than there are true fatal events. This is why deployment is framed as enrichment for inspection triage rather than as automated classification.

3.5. Interpretability

Feature attribution used SHAP (Shapley additive explanations) computed natively from the XGBoost booster's pred_contribs path on the held-out test set [42]. Global importance is summarized as mean absolute SHAP values, and direction-of-effect is read from beeswarm plots against normalized feature values. Attributions are interpreted as associations within a predictive model, not as causal effects.

3.6. Narrative Text Mining

Free-text narratives accompany 47,700 records (99.4%). Because narratives are written after the outcome is known, a scrubbing step replaced outcome-revealing tokens (e.g., 'fatal', 'died', 'deceased', 'killed', 'pronounced') with a neutral placeholder before any feature extraction; this step is the central leakage control and is reported as part of the pipeline. Scrubbed texts were vectorized with TF–IDF (unigrams and bigrams, minimum document frequency 15, 4000 features) and reduced to 40 latent components by truncated SVD (15.9% cumulative explained variance). These text components were appended to the structured feature matrix to measure the incremental value of narratives for severity prediction with XGBoost under the same cross-validation protocol. Separately, LDA with ten topics (batch learning, seed 42 [43]) was fitted on scrubbed narratives for thematic analysis, and mean topic weights were contrasted between fatal and non-fatal events.

3.7. Reproducibility

Analyses ran on Python 3.12 (Python Software Foundation, Beaverton, OR, USA) with pandas 2.3.2 (NumFOCUS, Austin, TX, USA), scikit-learn 1.7.2 (Inria, Paris, France) [44], XGBoost 3.4.1 (Distributed (Deep) Machine Learning Community; open source, https://xgboost.ai), imbalanced-learn 0.14.2 (Université de Lorraine, Nancy, France) and SciPy 1.16.2 (NumFOCUS, Austin, TX, USA); all random seeds were fixed at 42. The sensitivity analysis of the Chinese composition data (Section 4.7) runs in the same pipeline and uses only the counts and mean death tolls published in the source dataset [24]. The processing scripts, the event-level modeling table and per-figure source data accompany this paper as Supplementary Materials (File S1) and in the data availability statement.

4. Results

4.1. Temporal Trends (RQ1)

China's FRPMT declined from 0.564 in 2011 to 0.044 in 2021, a 92.2% reduction, then rebounded to 0.054 in 2022 and 0.094 in 2023 before falling to 0.059 in 2024 and 0.045 in 2025 (Figure 1a). Reported coal mine deaths followed the same arc: 1973 in 2011, 588 in 2015, 316 in 2019, a low of 178 in 2021, and a rebound to 245 in 2022 and 443 in 2023 [2–5,7]. In the U.S. archive, reportable coal mine events fell from 6089 in 2011 to 1994 in 2025 (−67.3%), and fatal events from 20 per year in 2011–2013 to 5 in 2020 and 8 in 2025 (Figure 1b). Both countries therefore improved, but along different geometries: China's curve is steep with a policy-visible rebound during the 2021–2023 supply-security period, whereas the U.S. curve is a long, flattening tail of single-victim events.

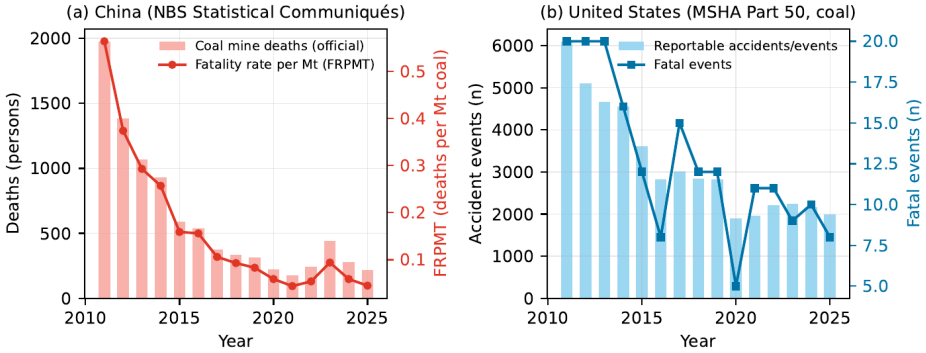


Figure 1. Temporal trends, 2011–2025. (**a**) China: coal mine deaths (bars) and fatality rate per million tonnes of coal (line), from NBS Statistical Communiqués; 2024–2025 deaths derived from FRPMT × raw-coal output. (**b**) United States: reportable coal mine accident events (bars) and fatal events (line), event-level MSHA Part 50 records.

Placing the two countries on a single production-normalized scale changes the picture in an important way. U.S. coal production fell from 993.9 to 479.4 million tonnes across the window, so although U.S. fatalities declined from 20 to 8, U.S. fatalities per million tonnes of coal were essentially flat, moving within a narrow 0.010–0.022 band and showing no downward trend: 0.020 in 2011, 0.022 at its highest in 2013, 0.010 at its lowest in 2020, and 0.017 in 2025. The 2020 minimum reflects depressed output in that year rather than a step change in safety, since the fatality count was also the lowest on record (5). China's rate, by contrast, fell by an order of magnitude over the same period. The two rates therefore converged sharply even though only one of them genuinely improved: the ratio of the Chinese to the U.S. rate narrowed from 28.0 in 2011 to 2.1 in 2021 and 2.7 in 2025. China's absolute death toll also fell, from 1973 in 2011 to 218 in 2025, on a raw-coal output that grew from roughly 3.5 to 4.8 billion tonnes (Table 2). Read with Figure 1b, this indicates that the U.S. record over 2011–2025 is one of declining exposure rather than declining risk per unit of production, a reading that a comparison of FRPMT against fatal-event counts would otherwise have concealed.

Table 2. Harmonized annual safety series. China: NBS Statistical Communiqués (2024–2025 deaths derived from FRPMT × raw-coal output). United States: fatalities from MSHA Part 50 records, coal production from EIA converted to metric tonnes (1 short ton = 0.9071847 t) [38]; the last column is the ratio of the two, and Section 3.2 explains why U.S. fatal events and U.S. deaths coincide.

Year	China FRPMT	China Deaths	U.S. Events (n)	U.S. Deaths (n)	U.S. Coal Output (Mt)	U.S. FRPMT
2011	0.564	1973	6089	20	993.9	0.02
2012	0.374	1384	5105	20	922.1	0.022

Table 2. *Cont.*

Year	China FRPMT	China Deaths	U.S. Events (n)	U.S. Deaths (n)	U.S. Coal Output (Mt)	U.S. FRPMT
2013	0.293	1067	4670	20	893.4	0.022
2014	0.257	931	4547	16	907.2	0.018
2015	0.159	588	3608	12	813.7	0.015
2016	0.156	538	2830	8	660.8	0.012
2017	0.106	375	3022	15	702.7	0.021
2018	0.093	333	2836	12	686.0	0.017
2019	0.083	316	2820	12	640.8	0.019
2021	0.044	178	1954	11	523.8	0.021
2022	0.054	245	2210	11	539.0	0.02
2023	0.094	443	2244	9	524.3	0.017
2024	0.059	281	2163	10	465.0	0.022
2025	0.045	218	1994	8	479.4	0.017

4.2. Composition, Severity Structure and Geography (RQ1)

Composition differs sharply. In the U.S. archive, routine injury events are dominated by handling of materials (26.7% of events), falls of roof or back (15.9%) and slips or falls of persons (14.7%), but fatal events concentrate in powered haulage (31.2% of 189 fatal events), machinery (25.9%) and falls of face, rib or highwall (14.3%) (Figure 2a). In the Chinese 532-report dataset, roof accidents are most frequent (19.9%) followed by transport (14.4%), yet severity concentrates in gas explosions (mean 20.84 deaths per event), fires (15.45), coal-dust explosions (10.33), water inrush (9.96) and outbursts (9.55) [24] (Figure 2b). The structural contrast is consistent: U.S. fatalities arise chiefly from energy-exchange events involving mobile equipment, typically one victim at a time, whereas Chinese mortality is driven by low-frequency, high-consequence underground hazards. Geographically, U.S. events concentrate in West Virginia (35.1% of events; 38.1% of fatal events), Kentucky (16.4%; 17.5%) and Pennsylvania (12.3%; 10.6%) (Figure 3a); within-year seasonality is mild, with a late-summer peak (Figure 3b). Chinese accidents cluster in the central and western coal provinces, with Shanxi, Gansu and Inner Mongolia heading the 532-report dataset [24].

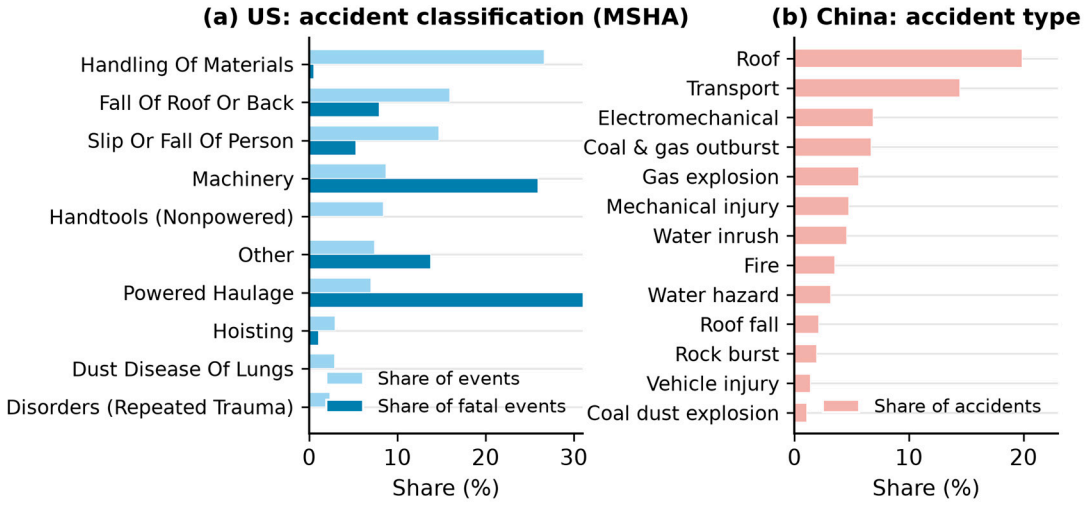


Figure 2. Accident composition. (**a**) U.S. accident classifications: shares of all events and of fatal events (MSHA, n = 47,986 events; 189 fatal). (**b**) Chinese accident types by share, from the published 532-report structured dataset [24].

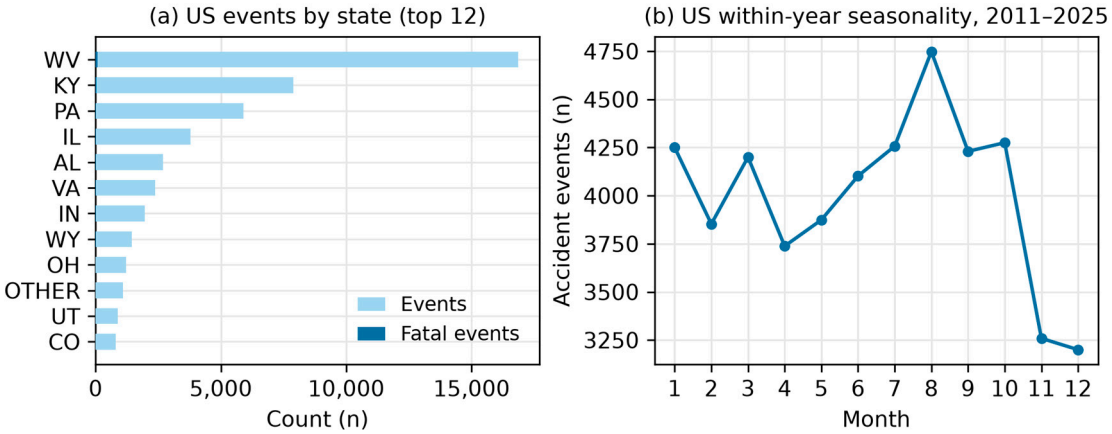


Figure 3. Spatial and seasonal distribution of U.S. coal mine events, 2011–2025. (**a**) Events and fatal events by state (top 12). (**b**) Month-of-year distribution of events.

4.3. Correlates of Fatal Outcomes (RQ2)

Chi-square tests with effect sizes (Table 3) show that accident classification is the strongest situational discriminator of fatal outcome ($\chi^2 = 529.7$, $p = 1.46 \times 10^{-106}$, Cramér's V = 0.105), followed by activity (V = 0.077), accident type (V = 0.068) and injury source (V = 0.067). Occupation carried a smaller but clear signal (V = 0.048). Mine subunit (underground vs. surface operations; V = 0.032), underground mining method (V = 0.020) and shift period (V = 0.019) were statistically significant but weak, and state was not associated with fatality share once event mix was accounted for ($p = 0.98$). The pattern indicates that what the worker was doing and what energy source was involved matter far more than where or when the event occurred.

Table 3. Chi-square tests of association with fatal outcome (n = 47,986; 189 fatal).

Factor	χ^2	dof	p	Cramér's V
classification	529.7	11	1.46×10^{-106}	0.105
activity	282.4	12	2.31×10^{-53}	0.077
accident type	224.7	12	2.51×10^{-41}	0.068
injury source	217.5	12	7.94×10^{-40}	0.067
occupation	112.3	15	5.91×10^{-17}	0.048
subunit	49.7	8	4.72×10^{-8}	0.032
mining method (UG)	19.4	7	7.07×10^{-3}	0.02
shift period	16.7	4	2.21×10^{-3}	0.019
state	7.4	17	9.77×10^{-1}	0.012

4.4. Severity Prediction Performance (RQ3)

Cross-validated performance on structured features (Table 4) ranged from ROC–AUC 0.878 ± 0.013 (gradient boosting) to 0.919 ± 0.011 (random forest); XGBoost achieved the best cross-validated PR–AUC among non-resampled models (0.064 ± 0.015) and SMOTE resampling raised PR–AUC to 0.087 ± 0.037 at some cost to ROC–AUC stability. On the held-out test set (9598 events, 38 fatal), ROC curves separated the models clearly (Figure 4a), and the precision–recall view (Figure 4b) quantified the difficulty of the 0.4% prevalence regime: at the Youden threshold the random forest recovered all 38 fatal events at precision 0.005, gradient boosting achieved the best precision–recall balance (recall 0.71, precision 0.026, F1 0.050), and XGBoost with SMOTE attained recall 0.87 with precision 0.021 (Figure 5). These numbers are reported deliberately: in this prevalence regime the models are screening instruments that enrich fatal-event risk by roughly an order of magnitude

over base rate, not standalone classifiers, and any deployment must budget for the implied inspection load.

Table 4. Model performance: five-fold stratified cross-validation (mean ± SD) and held-out test at the Youden threshold.

Model	CV ROC–AUC	CV PR–AUC	Test ROC–AUC	Test PR–AUC	Recall	Precision	F1
Logistic regression	0.902 ± 0.021	0.047 ± 0.014	0.878	0.026	0.947	0.009	0.018
Random forest	0.919 ± 0.011	0.062 ± 0.020	0.929	0.039	1	0.005	0.009
Gradient boosting	0.878 ± 0.013	0.037 ± 0.010	0.892	0.038	0.711	0.026	0.05
XGBoost	0.892 ± 0.019	0.064 ± 0.015	0.905	0.049	0.921	0.012	0.023
XGBoost + SMOTE	0.913 ± 0.021	0.087 ± 0.037	0.916	0.044	0.868	0.021	0.041
XGBoost + text	0.966 ± 0.008	0.219 ± 0.053	0.966	0.252	–	–	–

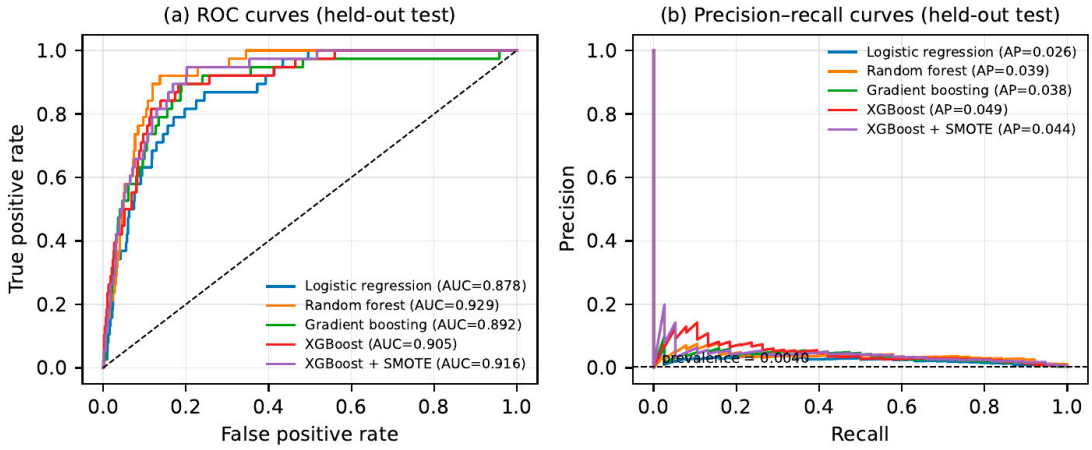


Figure 4. Held-out test-set discrimination (9598 events, 38 fatal). (**a**) ROC curves. (**b**) Precision–recall curves; the dashed line marks outcome prevalence (0.004).

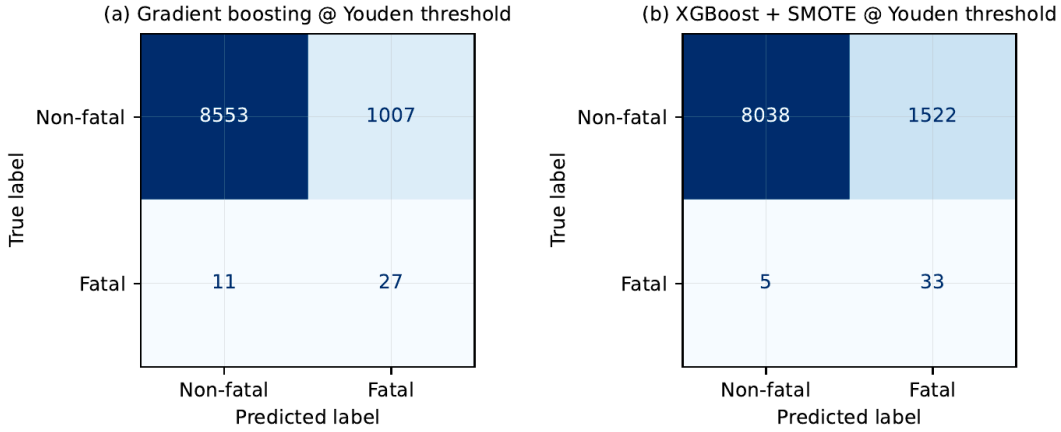


Figure 5. Confusion matrices at the Youden threshold estimated from out-of-fold training predictions. (**a**) Gradient boosting. (**b**) XGBoost with SMOTE.

4.5. SHAP Attribution (RQ3)

Global SHAP importance for the ten highest-ranked features (structured XGBoost) is listed in Table 5. SHAP attribution of the structured XGBoost model (Figure 6) ranked the material-handling classification first (mean |SHAP| = 1.20): its high SHAP values were associated with low feature values, i.e., events outside material handling carried elevated fatality risk, consistent with the descriptive finding that material handling dominates minor-injury volume but only 0.5% of fatal events. The three experience measures

followed (0.88–0.93), with shorter total, mine-level and job-specific experience pushing predictions toward fatal, and missing experience records forming a distinct risk profile. Calendar year and month captured the declining baseline risk and seasonal residual. Among hazard classes, falls of roof or back and slip/fall events carried positive fatal attribution, continuous mining as the underground method added risk, and the roof-bolter occupation, the trade most exposed to the roof-fall energy source, ranked within the top fifteen. The attribution structure is therefore coherent with domain knowledge rather than an artifact of feature noise.

Table 5. SHAP global importance, top 10 features (structured XGBoost, held-out test set).

Rank	Feature	Mean |SHAP|
1	Class: handling of materials	1.204
2	Total experience (yr)	0.925
3	Job experience (yr)	0.922
4	Mine-site experience (yr)	0.879
5	Calendar year	0.47
6	Month of year	0.464
7	Activity: other	0.397
8	Class: fall of roof or back	0.344
9	Experience not reported	0.323
10	Class: slip or fall of person	0.32

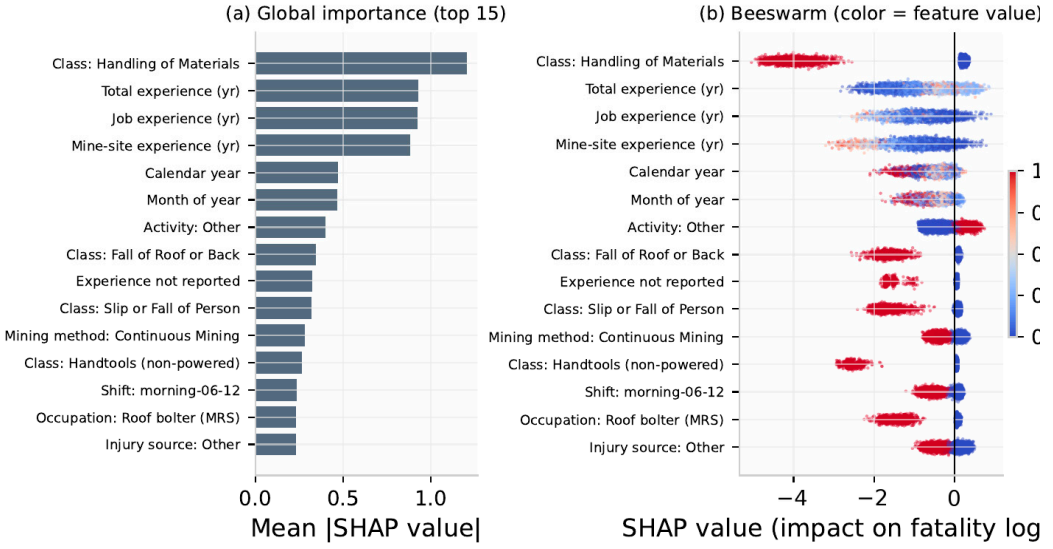


Figure 6. SHAP attribution of the structured-feature XGBoost model on the held-out test set. (**a**) Mean absolute SHAP values, top 15 features. (**b**) Beeswarm plot; point color encodes the normalized feature value.

4.6. Narrative Text: Themes and Predictive Increment (RQ3)

Adding scrubbed narrative components raised cross-validated ROC–AUC from 0.892 ± 0.019 (structured-only XGBoost) to 0.966 ± 0.008 and cross-validated PR–AUC from 0.064 to 0.219; on the held-out test set PR–AUC increased from 0.049 to 0.252, roughly 64 times the 0.004 prevalence baseline (Table 4). Because outcome-revealing tokens were scrubbed before vectorization, this increment reflects situational and causal wording rather than label leakage. LDA themes separated cleanly along severity lines (Figure 7); as in the caption, every pair of topic weights quoted below gives the fatal-narrative value first. Topics loading on mobile equipment and haulage (0.297 in fatal narratives versus 0.129 in non-fatal) and on shaft/elevator service (0.135 versus 0.046) were weighted more heavily in fatal narratives, whereas topics loading on hand laceration (0.139 versus 0.199),

musculoskeletal pain (0.025 versus 0.175) and occupational disease (0.022 versus 0.057) were weighted more heavily in non-fatal narratives; the roof-and-rib topic sat between the two groups (0.157 versus 0.140), mirroring its mixed severity profile in Table 2. The text stream thus supplies both an interpretable thematic layer and a large, leakage-controlled predictive increment.

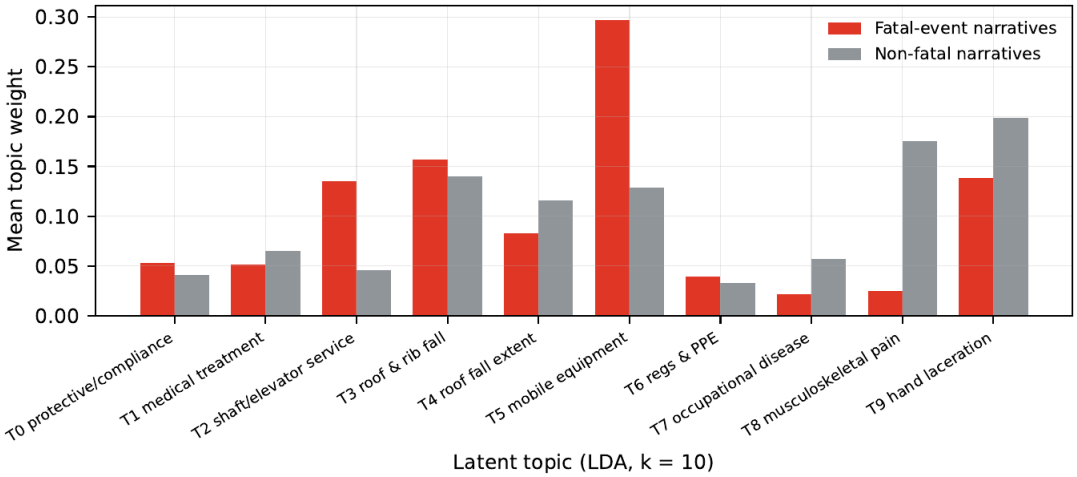


Figure 7. LDA topic structure of scrubbed accident narratives (k = 10): mean topic weights in fatal-event narratives (dark bars) versus non-fatal narratives (light bars). Wherever a pair of topic weights is quoted in the text, the fatal-narrative value is given first.

4.7. Sensitivity of the Chinese Composition Data to Severity Bias (RQ1)

The Chinese structured dataset is a collection of investigated accidents and therefore over-represents severe events, which limits what its composition can say about the full accident spectrum; we address this directly rather than by assertion, in three steps. First, we quantify the coverage gap. Summing the type-level counts in the published table weighted by their mean death tolls implies roughly 2266 deaths, against 9095 officially reported Chinese coal mine deaths for 2011–2025 [2–5,7], so the dataset documents about a quarter of recorded fatalities; it is a severity-enriched subsample rather than a census. Second, we test whether the dataset is nevertheless confined to catastrophic events, and it is not: 65.6% of its entries are general-grade accidents, 17.1% are major, 16.0% are comparatively major and only 1.3% are extraordinarily severe, and the mean death toll is 4.6 per accident with a median of 1 [24]. The dataset is therefore biased toward investigated cases without being a disasters-only collection, which bounds the direction and the rough size of the distortion. Third, and most importantly, we separate the two kinds of statistic that this paper draws from the dataset. Frequency shares are sensitive to severity enrichment, but conditional statistics such as deaths per event within an accident type are not, because they are computed inside a category rather than across the sampling frame. Table 6 shows the consequence: the ranking by event frequency (roof, transport, electromechanical) is nearly the reverse of the ranking by implied deaths (gas explosion, coal and gas outburst, fire, water inrush), and the divergence survives an explicit worst-case reweighting in which each type is weighted by the reciprocal of its mean death toll, which removes severity enrichment by construction. Under that reweighting, the frequency ranking shifts further toward low-consequence types, while the high-consequence mean death tolls that carry the paper's severity claim are unchanged, because they are conditional on type. One reporting detail of the source dataset should be noted: its type table assigns counts to 13 named categories, and the percentages printed there use a different denominator, so we work from the counts and mean death tolls themselves throughout.

Table 6. Sensitivity of the Chinese composition data to severity enrichment. Counts and mean death tolls are as published [24] for the 13 accident types tabulated there, and implied deaths is their product; the published percentages in that table use a different denominator, so counts are used here. The final column reweights each type by the reciprocal of its mean death toll to emulate a severity-neutral sampling frame.

Accident Type	Events (n)	Mean Deaths/Event	Implied Deaths (n)	Rank by Events	Rank by Implied Deaths	Severity-Neutral Share (%)
Roof	113	2.01	227	1	5	21.8
Transport	82	1.20	98	2	7	26.5
Electromechanical	39	1.00	39	3	10	15.1
Coal and gas outburst	38	9.55	363	4	2	1.5
Gas explosion	32	20.84	667	5	1	0.6
Mechanical injury	27	0.41	11	6	13	25.5
Water inrush	26	9.96	259	7	4	1.0
Fire	20	15.45	309	8	3	0.5
Water hazard	18	6.22	112	9	6	1.1
Roof fall	12	1.00	12	10	12	4.7
Rock burst	11	7.09	78	11	8	0.6
Vehicle injury	8	3.62	29	12	11	0.9
Coal dust explosion	6	10.33	62	13	9	0.2
Total	432	–	2266	–	–	100.0

The Chinese claims are restricted accordingly: the composition statistics describe the severity structure of investigated accidents, and we make no claim that the event-frequency shares in Figure 2b or Table 6 approximate the full population of Chinese coal mine accidents. The rate comparison in Section 4.1 and Figure 8 is unaffected, being built on the official national series rather than on the structured dataset, and the harmonized conclusion that Chinese mortality is concentrated in high-consequence underground hazards is if anything strengthened by the correction, since over-sampling of severe accidents would imply an even lower true frequency for gas explosions and outbursts while leaving their conditional lethality intact.

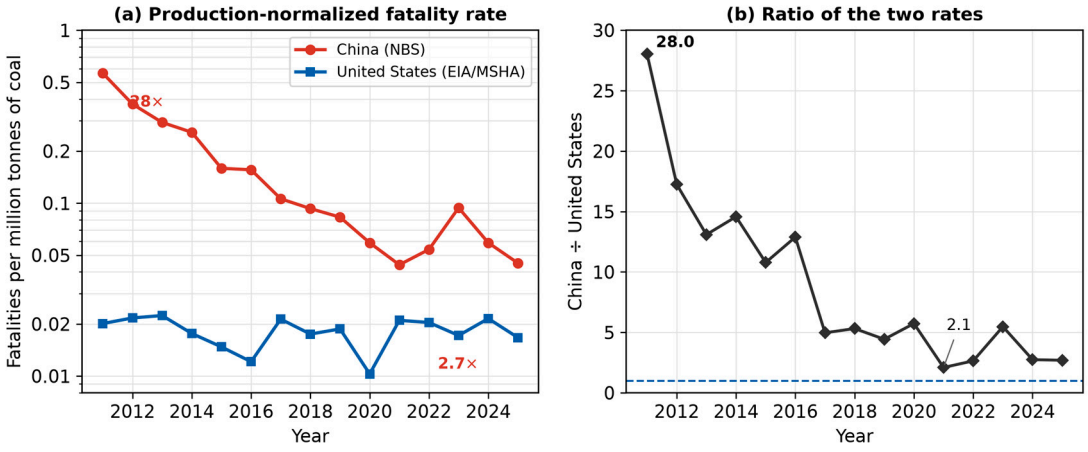


Figure 8. Production-normalized comparison, 2011–2025. (**a**) Coal mine fatalities per million tonnes of coal; China (circles) from the official NBS series, and United States (squares) computed as annual fatalities divided by EIA coal production, so both are on one scale. Note the logarithmic axis. (**b**) Ratio of the Chinese to the U.S. rate; the dashed line marks parity. Values are listed in Table 2.

5. Discussion

The harmonized comparison answers RQ1 with a structural asymmetry that aggregate fatality rates alone conceal. China's risk profile remains dominated by catastrophic under-

ground energy releases, so progress depends on suppressing low-frequency, high-death toll events; the 2022–2023 rebound coincided with production-pressure conditions that official commentary linked to over-capacity operation and concealed hazards [8,35], and the subsequent decline followed the 2024 regulatory tightening [34,35]. The U.S. profile is the opposite: a dispersed tail of single-victim equipment events in which powered haulage and machinery account for over half of fatal events, a pattern stable across the study window and consistent with long-run analyses of MSHA data [21–23]. The production-normalized series adds a second and less comfortable reading. Measured per million tonnes of coal, U.S. fatal risk was flat across 2011–2025 while Chinese risk fell by an order of magnitude, so the convergence from 28-fold to below 3-fold reflects Chinese improvement rather than U.S. progress; the absolute decline in U.S. fatalities is real, but it tracks a 52% contraction in coal production, and per unit of coal mined the underlying hazard has not changed materially. The distinction matters for policy transfer, because a regulator reading only incident counts may mistake reduced exposure for reduced risk. For RQ2, effect sizes order the levers clearly: the immediate work activity and the involved energy source outweigh geography and shift timing, and worker experience operates as a continuous gradient rather than a threshold effect, echoing human-factor analyses of Chinese gas explosions [18,20] and U.S. panel findings on mine-tenured workers [22]. For RQ3, class-imbalanced ensemble models are viable fatality-risk screeners, and the SHAP structure recovered domain-plausible drivers without being tuned toward them; the large narrative increment shows that the circumstances inspectors choose to record carry most of the discriminating signal, which argues for preserving and standardizing narrative fields in reporting systems [24,30].

Three implications follow. For regulators, fatality-risk screening can prioritize inspection queues: a model with test recall 0.71–0.87 at precision 0.02–0.03 concentrates fatal-event likelihood roughly 10-fold over base rate, which is operationally meaningful when inspection capacity is rationed, provided the false-positive load is accepted explicitly. For training programs, the experience gradient and the roof-bolter and continuous-mining signals point to task-specific mentoring for new and transferred workers rather than generic safety courses. For technical prevention, the U.S. pattern argues for proximity-detection and collision-avoidance retrofits on mobile equipment, while the Chinese pattern continues to argue for gas drainage, ventilation reliability and outburst prediction in deep or gassy seams [9,10,16].

The prospective value of the framework needs to be stated precisely, because it is built on historical accident records and cannot forecast accidents that have not happened. What it offers is not forecasting but triage. When a new reportable event is entered into a system such as Part 50, its structured fields and its scrubbed narrative can be scored immediately, and events whose score falls in the top decile of predicted fatality risk can be routed to inspection, follow-up or learning-review queues ahead of the routine backlog. The retrospective archive supplies the mapping from event characteristics to outcome that makes this prioritization possible; the prospective step is the deployment of that fixed mapping on incoming events, which is why the design is described as screening-oriented risk triage and not as early warning of future accidents. Two conditions bound the mechanism: the model enriches lethal-event risk roughly ten-fold over base rate rather than separating fatal from non-fatal events, so what it allocates is scarce inspection capacity rather than a binary decision; and the text increment is measured only after outcome-revealing tokens have been scrubbed, which removes the most obvious channel by which a post hoc narrative might otherwise appear predictive. The abstract and the conclusions have been revised to describe screening-oriented risk triage in place of the earlier wording, which overstated the mechanism.

Several limitations bound these claims. First, the comparison is asymmetric by necessity: Chinese event-level public records with covariate depth comparable to MSHA do not exist, so modeling was performed on the U.S. archive and the Chinese side contributes rates and composition; findings about predictors should not be projected onto Chinese mines. Second, MSHA captures operator reports under Part 50, so under-reporting of minor events is possible, which would bias prevalence but is less likely to distort fatal-event covariates; because the EIA production denominators are drawn from the same MSHA Form 7000-2 returns, the numerator and denominator share any reporting bias in that system even though the two are internally consistent. Third, the 532-report Chinese dataset over-represents investigated, and therefore more severe, accidents [24]; Section 4.7 quantifies the resulting coverage gap and shows that the paper's conditional severity statistics are insensitive to it, but frequency shares for China should still be read as describing investigated accidents only. Fourth, with 189 fatal events, PR–AUC estimates carry wide uncertainty (cross-validated SD up to 0.04), and model rankings should be read as indicative. Fifth, narrative scrubbing is conservative but cannot remove all indirect outcome cues, so the text increment is an upper bound of the deployable value. Finally, all attributions are associational; no causal claim is made about any single factor.

6. Conclusions

Using harmonized public accident data for 2011–2025, this study shows that, measured per million tonnes of coal, Chinese coal mine fatality risk fell from 0.564 in 2011 to 0.045 in 2025 with a rebound peaking at 0.094 in 2023, while U.S. risk remained essentially flat (0.020 to 0.017) even as reportable accidents declined by two-thirds and fatalities fell from 20 to 8; the ratio between the two rates therefore narrowed from 28-fold to below 3-fold, and the U.S. decline in absolute fatalities reflects contracting production rather than a lower hazard per unit mined (RQ1). Fatal risk in the U.S. archive concentrates in powered haulage (31.2% of fatal events) and machinery (25.9%), and correlates most strongly with what the worker was doing and which energy source was involved (Cramér's V up to 0.105), with worker experience acting as a graded protective factor (RQ2). Interpretable ensemble models predict fatal outcomes at cross-validated ROC–AUC up to 0.919 on structured data and 0.966 with leakage-scrubbed narratives, and SHAP attribution recovers domain-consistent drivers, supporting screening-oriented risk triage of newly reported events rather than automated judgment (RQ3). A sensitivity analysis of the investigated-accident bias in the Chinese dataset confirms that the conditional severity results are robust to it, while restricting what its frequency shares can claim (RQ1). The harmonization protocol, leakage controls and reporting template transfer directly to other jurisdictions with public accident archives, and the contrast between the two national risk profiles, catastrophic-hazard suppression in China versus mobile-equipment control in the United States, gives each regulator a mirror in which its own priorities become visible.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/pr14193115/s1, S1: Classification mapping (MSHA → Chinese ten-category taxonomy); S2: Full SHAP importance ranking (top 25); S3: LDA topics: full keyword lists and severity contrast; S4: Month-of-year distribution of U.S. coal mine events, 2011–2025; S5: Robustness and pipeline notes; S6: Sensitivity of the Chinese composition data to severity enrichment (Table S1). File S1. Source_Data_Workbook.

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

Funding: This research was funded by the Fundamental Research Program of Shanxi Province, China (Young Scientific Research Project, free-exploration category), grant number 202203021212198.

Data Availability Statement: The U.S. accident data are publicly available from the MSHA Open Government Data portal (https://arlweb.msha.gov/OpenGovernmentData/DataSets/Accidents.zip; accessed on 1 September 2026). Chinese annual statistics are publicly available from the National Bureau of Statistics Statistical Communiqués (https://www.stats.gov.cn) and NMSA releases (https://www.chinamine-safety.gov.cn). Chinese composition statistics were taken from the published 532-report dataset description [24]. The processed event-level modeling table, per-figure source data and analysis scripts are provided in the accompanying source-data workbook and Supplementary Materials.

Acknowledgments: During the preparation of this manuscript, the authors used an AI-based writing assistant for language polishing and consistency checks. The authors have reviewed and edited the output and take full responsibility for the content of this publication.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript: MSHA, Mine Safety and Health Administration; FRPMT, fatality rate per million tonnes of coal; SHAP, Shapley additive explanations; LDA, latent Dirichlet allocation; TF–IDF, term frequency–inverse document frequency; SVD, singular value decomposition; ROC–AUC, area under the receiver operating characteristic curve; PR–AUC, area under the precision–recall curve; SMOTE, synthetic minority over-sampling technique; NBS, National Bureau of Statistics of China; NMSA, National Mine Safety Administration.

References

1. Mine Safety and Health Administration. *Open Government Data: Accidents Data Set (30 CFR Part 50), File Dated 28 August 2026*; U.S. Department of Labor: Washington, DC, USA, 2026. Available online: https://arlweb.msha.gov/OpenGovernmentData/DataSets/Accidents.zip (accessed on 1 September 2026).
2. China Energy News. Coal mine Fatality Rate per Million Tonnes Hits Record Low (2019 National Coal Mine Safety Work Conference). *People's Daily Online*. 13 January 2020. Available online: http://paper.people.com.cn/zgnyb/html/2020-01/13/content_1966990.htm (accessed on 1 September 2026). (In Chinese)
3. National Bureau of Statistics of China. *Statistical Communiqué of the People's Republic of China on the 2019 National Economic and Social Development*; NBS: Beijing, China, 2020. Available online: https://www.stats.gov.cn/english/PressRelease/202002/t20200228_1728917.html (accessed on 1 September 2026). (In Chinese)
4. National Bureau of Statistics of China. *Statistical Communiqué of the People's Republic of China on the 2020 National Economic and Social Development*; NBS: Beijing, China, 2021. Available online: https://www.stats.gov.cn/english/PressRelease/202102/t20210228_1814177.html (accessed on 1 September 2026). (In Chinese)
5. National Bureau of Statistics of China. *Statistical Communiqué of the People's Republic of China on the 2024 National Economic and Social Development*; NBS: Beijing, China, 2025. Available online: https://www.stats.gov.cn/sj/zxfb/202502/t20250228_1958817.html (accessed on 1 September 2026). (In Chinese)
6. Mine Safety and Health Administration. *Mine Safety and Health at a Glance: Fiscal Year*; U.S. Department of Labor: Washington, DC, USA, 2026. Available online: https://www.msha.gov/mine-safety-and-health-glance-fiscal-year (accessed on 1 September 2026).
7. National Bureau of Statistics of China. *Statistical Communiqué of the People's Republic of China on the 2025 National Economic and Social Development*; NBS: Beijing, China, 2026. Available online: https://www.stats.gov.cn/english/PressRelease/202602/t20260228_1962661.html (accessed on 1 September 2026). (In Chinese)
8. Wang, Y.; Fu, G.; Lyu, Q.; Wu, Y.; Jia, Q.; Yang, X.; Li, X. Reform and development of coal mine safety in China: An analysis from government supervision, technical equipment, and miner education. *Resour. Policy* **2022**, *77*, 102777. [CrossRef]
9. Chen, H.; Qi, H.; Long, R.; Zhang, M. Research on 10-year tendency of China coal mine accidents and the characteristics of human factors. *Saf. Sci.* **2012**, *50*, 745–750. [CrossRef]
10. Wang, L.; Cheng, Y.P.; Liu, H.Y. An analysis of fatal gas accidents in Chinese coal mines. *Saf. Sci.* **2014**, *62*, 107–113. [CrossRef]
11. Zhang, P.; Li, F.; Zhu, H. Statistical analysis and prevention countermeasures of coal mine accidents in China from 2008 to 2020. *Min. Saf. Environ. Prot.* **2022**, *49*, 128–134. (In Chinese)

12. Zhao, Y.; Yan, Y.; Liu, K.; Zhao, X.; Li, H.; Cao, J.; Zhang, S.; Ma, K. Analysis of coal mine safety accident features in China, 2017–2022. *Geohazard Mech.* **2024**, *2*, 108–120. [CrossRef]
13. Zhang, J.; Cliff, D.; Xu, K.; You, G. Focusing on the patterns and characteristics of extraordinarily severe gas explosion accidents in Chinese coal mines. *Process Saf. Environ. Prot.* **2018**, *117*, 390–398. [CrossRef]
14. Zhang, J.J.; Xu, K.L.; Reniers, G.; You, G. Statistical analysis of the characteristics of extraordinarily severe coal mine accidents (ESCMAs) in China from 1950 to 2018. *Process Saf. Environ. Prot.* **2020**, *133*, 332–340. [CrossRef]
15. Zhu, Y.F.; Wang, D.M.; Shao, Z.L.; Xu, C.H.; Zhu, X.L.; Qi, X.Y.; Liu, F. A statistical analysis of coalmine fires and explosions in China. *Process Saf. Environ. Prot.* **2019**, *121*, 357–366. [CrossRef]
16. Guan, Z.-Y.; Xie, J.-L.; Wu, S.-K.; Liang, C. Research on predicting the risk level of coal mine roof accident based on machine learning. *Sci. Rep.* **2025**, *15*, 24028. [CrossRef] [PubMed]
17. Sun, W.; Li, W.; Wu, Q.; Zhao, S.; Liu, W.; Wang, C. Unraveling the spatio-temporal pattern of mine water accidents in China. *Mine Water Environ.* **2025**, *44*, 226–238. [CrossRef]
18. Chen, H.; Qi, H.; Feng, Q. Characteristics of direct causes and human factors in major gas explosion accidents in Chinese coal mines: Case study spanning the years 1980–2010. *J. Loss Prev. Process Ind.* **2013**, *26*, 38–44. [CrossRef]
19. Guo, Y.; Li, J.; Cliff, D. Exploring the relationship between miners' physiological signals and safety behavior in four emergency scenarios in coal mines: A virtual reality study. *Int. J. Ind. Ergon.* **2025**, *109*, 103801. [CrossRef]
20. Yin, W.; Fu, G.; Yang, C.; Jiang, Z.; Zhu, K.; Gao, Y. Fatal gas explosion accidents on Chinese coal mines and the characteristics of unsafe behaviors: 2000–2014. *Saf. Sci.* **2017**, *92*, 173–179. [CrossRef]
21. Centers for Disease Control and Prevention. Underground coal mining disasters and fatalities—United States, 1900–2006. *MMWR Surveill. Summ.* **2009**, *58*, 105–109. Available online: https://stacks.cdc.gov/view/cdc/183316 (accessed on 1 September 2026). [PubMed]
22. National Institute for Occupational Safety and Health. *Accident Analysis of Mining Industry in the United States: A Retrospective Study of MSHA Accident Databases, 1983–2018*; CDC Stacks: Washington, DC, USA, 2022. Available online: https://stacks.cdc.gov/view/cdc/230171 (accessed on 1 September 2026).
23. Wang, F.; Yang, Y.; Wu, L.; Meng, H.; Li, P.; Pang, L. Analysis on accident types of coal mine in global major coal producing countries. *Smart Cities Adv. Technol.* **2023**, *1*, 1–9. Available online: https://www.extrica.com/article/22419 (accessed on 1 September 2026). [CrossRef]
24. Wan, M.; Li, H.; Wang, H.; Gong, H.; Hou, J. China's 15-year mine accident report dataset (2010–2025): Construction and analysis. *Data* **2025**, *10*, 202. [CrossRef]
25. Al Hasanat, H.A.; Hassanat, A.B.; Alharasees, O.; Tarawneh, A.S.; Altarawneh, G.A.; Alhasanat, L.J.; Begovic, M. Interpretable machine learning for imbalanced pedestrian injury severity prediction in urban Jordan. *J. King Saud Univ.-Eng. Sci.* **2025**, *37*, 39. [CrossRef]
26. Wang, C.; Serre, T. A hybrid approach to investigating factors associated with crash injury severity: Integrating interpretable machine learning with logit model. *Appl. Sci.* **2025**, *15*, 10417. [CrossRef]
27. Wang, Y.; Zhang, J.; Hao, X.; Li, H.; Zhang, B.; Wei, Y.; Li, B. Analysis of the characteristics and causal factors of coal mine transportation accidents in China. In *Proceedings of the 7th International Conference on Electrical Engineering and Information Technologies for Rail Transportation*; Springer: Singapore, 2026. [CrossRef]
28. Xie, X.; Fu, G.; Xue, Y.; Zhao, Z.; Chen, P.; Lu, B.; Jiang, S. Risk prediction and factors risk analysis based on IFOA-GRNN and Apriori algorithms: Application of artificial intelligence in accident prevention. *Process Saf. Environ. Prot.* **2019**, *122*, 169–184. [CrossRef]
29. Li, Z.; Zhang, Y. High-altitude fall accidents in construction: A text mining analysis of causal factors and COVID-19 impact. *Modelling* **2025**, *6*, 124. [CrossRef]
30. Agarwal, S.; Chugh, Y.P.; Singh, A.; Sakinala, V.; Mukherjee, A.; Prasad, B.; Dagli, C.; Zou, Y. Application of natural language processing and machine learning for analyzing mining accident reports and automating the process of root cause analysis. *Int. J. Coal Sci. Technol.* **2025**, *12*, 91. [CrossRef]
31. Ouyang, Z.; Xu, Q.; Zhang, T. Coupling analysis of disaster-causing factors in coal mines and dual prevention mechanism based on the KeyBERT model and accident-causation theory model. *Front. Earth Sci.* **2025**, *13*, 1586785. [CrossRef]
32. Jing, G.; Liu, M. Statistics and regularity analysis of coal mine gas accidents in China, 2015–2019. *J. Saf. Environ.* **2022**, *22*, 1680–1686. (In Chinese) [CrossRef]
33. Zhang, J.; Xu, K.; You, G.; Wang, B.; Zhao, L. Causation analysis of risk coupling of gas explosion accident in Chinese underground coal mines. *Risk Anal.* **2019**, *39*, 1634–1646. [CrossRef] [PubMed]
34. State Council of the People's Republic of China. *Regulation on Coal Mine Safety Production (Decree No. 774, Effective 1 May 2024)*; State Council: Beijing, China, 2024. (In Chinese)

35. National Mine Safety Administration. *Eight Firm Measures for Preventing and Curbing Extraordinarily Serious Production Safety Accidents in the Mining Sector*; NMSA: Beijing, China, 2024. Available online: https://www.chinamine-safety.gov.cn (accessed on 1 September 2026). (In Chinese)
36. Pipeline and Hazardous Materials Safety Administration. *Natural Language Processing and Text Mining of Pipeline Incident Reports: Final Project Report*; U.S. Department of Transportation: Washington, DC, USA. Available online: https://primis.phmsa.dot.gov/rd/FileGet/16431/PHMSA_project_Final_report_v.0.pdf (accessed on 1 September 2026).
37. National Health Commission of China; Ministry of Education; Ministry of Science and Technology; National Natural Science Foundation of China. *Measures for Ethical Review of Life Science and Medical Research Involving Humans*; National Health Commission of the People's Republic of China: Beijing, China, 2023. (In Chinese)
38. U.S. Energy Information Administration. *Quarterly Coal Report (U.S. Coal Production, Table 1)*; U.S. Department of Energy: Washington, DC, USA, 2026. Available online: https://www.eia.gov/coal/production/quarterly/ (accessed on 1 September 2026).
39. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [CrossRef]
40. Chen, T.; Guestrin, C. XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 785–794. [CrossRef]
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. Lundberg, S.M.; Lee, S.-I. A unified approach to interpreting model predictions. In *Advances in Neural Information Processing Systems 30 (NeurIPS 2017)*; Curran Associates: Red Hook, NY, USA, 2017; pp. 4765–4774.
43. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
44. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V. Scikit-learn: Machine learning in Python. *J. Mach. Learn. Res.* **2011**, *12*, 2825–2830.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.