Journal of SRI "Geotechnological Problems of Oil, Gas and Chemistry" of ASOIU,
Issue 2, 2025, pp. 106-120
https://doi.org/10.65294/gpogc.2026.08

Forecasting the risk of formation damage (colmatation) with a 14-day horizon using calibrated probabilistic models

H.Kh. Malikov[1], T.E. Abdulmutalibov[2]*, G.V. Jabbarova[1,2], R.R. Tashmenov[3]

[1]Scientific Research Institute "Geotechnological Problems of Oil, Gas and Chemistry", 227 Dilara Aliyeva str., AZ1010, Baku, Azerbaijan
[2]Azerbaijan State Oil and Industry University, Azadliq Avenue, 20, AZ1010, Baku, Azerbaijan
[3]M. Auezov South-Kazakhstan State University, Building «B», Faculty mechanical and petroleum engineering G. Ilayev st. 8, Shymkent, Kazakhstan

Abstract. Formation damage (colmatation) is a major contributor to production losses and increased operating expenditures, yet routine diagnostics are often reactive and provide limited lead time for low-cost corrective actions. This study develops a reproducible 14-day early-warning workflow that forecasts **well-calibrated event probabilities** (not only risk rankings) and links them to an economically justified operating threshold. The target event is defined as a productivity-index decline below an engineering limit, **IC** = Q_o^{fact}/Q_o^{theor} < **0.90**, occurring within a 14-day horizon. Daily operational time series (water cut, net GOR, wellhead pressure, choke setting, GLR, and production rates) are transformed into a leakage-safe dynamic feature set using lags, rolling statistics, and trend descriptors computed strictly from information available at the forecast time. Model development follows rolling-origin (time-aware) backtesting with out-of-fold probability calibration (Platt scaling and isotonic regression). Performance is assessed in terms of discrimination (PR-AUC/ROC-AUC) and calibration (Brier score, expected calibration error, and calibration plots), while the operating threshold $\tau^*$is chosen by minimizing expected decision cost under asymmetric error penalties and validated via decision-curve and lift analyses. The calibrated models provide operationally meaningful probabilities that align with empirical event rates across temporal windows, and the selected τ^*yields positive net benefit compared with treat-all/treat-none baselines. The proposed pipeline is suitable for integration into a decision-support workflow with explicit recommendations for temporal validation, calibration, cost-based thresholding, and drift monitoring.

Keywords: formation damage (colmatation); productivity index; risk forecasting; probability calibration; PR-AUC; Brier; ECE.

*Corresponding author. Tel.: +994519951894
E-mail address: abdulmutalibov.tmr@gmail.com

1. Introduction. Formation damage (clogging/impairment) of the productive reservoir is among the most frequent causes of well productivity degradation and increasing field operating expenditures. It manifests through higher flow resistance in the near-wellbore zone, growth of the effective skin factor, and a decrease in the productivity index (IC), ultimately leading to declining production rates, unstable operating regimes, and unplanned interventions. The sources of formation damage are diverse: mechanical migration of fine particles; swelling and dispersion of clays; emulsion-related and surfactant-driven effects; incompatibility of aqueous phases; salt precipitation; deposition of asphaltenes and paraffins. Technological operations also contribute substantially—reservoir penetration and completion, injection of working agents, as well as changes in choke settings and operating conditions. Traditional diagnostic approaches rely on laboratory experiments, core analysis, and engineering observations, while operational control is typically based on retrospective rules and "manual" threshold triggers. Although these methods are useful for explaining mechanisms, in day-to-day operations they are often reactive: the signal arrives only after productivity has already declined noticeably, and the window for mild, low-cost corrective actions has been missed.

A shift toward proactive management requires early, quantitatively interpretable risk estimates tied to a specific planning horizon. In field practice, this window is often 10–14 days—the period within which decisions are made regarding additional diagnostics, operating-mode adjustments, and preparation for flushing or treatment. In this context, machine-learning methods are of particular interest: they enable extraction of weak, high-dimensional signals from operational time series, generation of probabilistic forecasts, and assessment of their utility in terms of the expected cost of errors. At the same time, applying ML to field time series involves methodological risks: leakage of future information during feature construction, overly optimistic performance estimates caused by inadequate validation, and incorrect interpretation of uncalibrated (“raw”) probabilities. Therefore, for practical deployment, the temporal evaluation scheme, proper probability calibration, and the choice of an operational threshold linked to decision economics are fundamentally important.

This study proposes a reproducible early-warning workflow for predicting formation-damage risk for the producing well J12 over a 14-day horizon. The target event is defined as a drop of the productivity index below an engineering-relevant threshold, IC < 0.90. A dynamic feature space is constructed from baseline operating parameters (WC, net GOR, WHP, Choke, GLR, rates), together with their temporal lags and rolling statistics; all windows are computed strictly in a forward-closed manner, thereby eliminating future leakage. For an unbiased assessment, rolling-origin backtesting is employed, combined with internal time-aware hyperparameter tuning. Probabilities are calibrated outside the training fold (Platt scaling and isotonic regression) and evaluated using the Brier score, expected calibration error (ECE), and calibration plots. The operational threshold $\tau*$ is selected by minimizing expected cost under asymmetric error costs that prioritize avoiding missed events, and it is validated using the decision-curve framework and lift curves as measures of applied usefulness. The practical contribution of this work is that: (i) it demonstrates the feasibility of obtaining early, calibrated probabilistic risk forecasts on a 14-day horizon; (ii) it links the forecast to operational decision-making through the threshold $\tau*$ and utility analysis; (iii) it proposes elements of an MLOps regimen (drift monitoring and recalibration) to support robust deployment within a decision-support system (DSS). In addition, comparison with engineering baselines (e.g., EWMA applied to IC and rule-based thresholds on $\Delta\,\Delta$WHP/Choke) shows the incremental value of the ML approach in terms of sensitivity and signal specificity while maintaining a controlled false-alarm rate. Thus, the study bridges the gap between reactive and proactive strategies by providing a verifiable, practice-integrable forecasting workflow aimed at reducing productivity losses and improving production stability.

2. Literature Review. Clogging of the pore space leads to a decline in permeability and productivity, and it has long been recognized as one of the key drivers of filtration-property degradation in clay-bearing and stress-heterogeneous reservoirs. At the mechanistic level, the literature emphasizes the roles of fine-particle migration, deposition/transport of suspended solids, as well as geomechanically induced reduction of the effective pore throat cross-section [1]. From a practical standpoint, this implies the need not only for “reactive” near-wellbore treatments, but also for proactive monitoring of risk indicators followed by an informed selection of corrective actions.

Engineering studies indicate that, when restoring productivity of horizontal and directional wells, the decisive factor is the correct choice of treatment technology and chemical formulations (acid systems, surfactants, water-shutoff compounds, etc.); the workflow is typically built upon diagnosing the dominant damage mechanisms and the field-specific technological constraints [2]. This aligns with observations regarding the specificity of formation damage in reservoirs with clay interbeds and heterogeneous stress states: local stress conditions and fluid composition determine the scenario of particle precipitation and redistribution [1]. In parallel with field procedures, modeling of acid stimulation and near-wellbore (NWB) technologies has been actively developed, including simulators that account for reaction kinetics, mass transfer, and the geometry of the flow network; such tools enable the selection of operating regimes and treatment compositions “on the desk” prior to field deployment [3].

Over the past decade, interest in data-driven methods and machine-learning algorithms for field analytics has increased markedly. While many publications focus on adjacent tasks (production forecasting, physico-chemical PVT properties, missing-data reconstruction, and time-series analysis), they provide a methodological foundation that is also relevant for early diagnostics of formation damage: (i) the use of noise-robust models (random forests, boosting) and regularized linear models;

(ii) prioritization of PR-AUC under potential class imbalance; (iii) the necessity of probability calibration and a controlled threshold selection under operational constraints. For example, production time series have been forecast successfully using both classical approaches (e.g., ARIMA) and ML-based methods [4]. Handling field datasets requires missing-value imputation and pattern-recognition techniques, which are critical for reliable inference in online environments [5]. In addition, ML applications to PVT tasks (GOR, FVF, bubble-point pressure) show that nonparametric models can outperform traditional correlations by capturing nonlinearities [6]; this experience is transferable to operational predictors (WC, GLR, NGOR, WHP, etc.) that influence the risk of pore-space clogging.

A major direction in managing NWB damage is the digitalization of chemical treatments—from developing acidizing simulators to monitoring and optimizing jobs in real time—which reduces operational risk and the number of ineffective treatments [3]. Within an operational workflow, such analytics must be grounded in production metrics and decision-support systems (DSS): alarm dashboards, interpretability of risk drivers, and integration with CMMS (automated work-order generation) tied to response-time SLAs represent a typical industrial deployment pattern described in publications on digital operations and advanced analytics in production [4–6, 7–8]. Methodologically, it is important that, on large empirical datasets, random forests and other tree ensembles often outperform “thin” linear models in ranking quality, especially under nonlinear dependencies and mixed feature types, whereas logistic regression remains a baseline reference due to its interpretability [9]. For practical use, raw probabilities require post-hoc calibration (Platt scaling/isotonic regression) and selection of an operational threshold consistent with the acceptable false-alarm level; this directly shapes the sensitivity–workload trade-off in PR space and is a standard recommendation in related domains that is applicable to early detection of formation damage as well [9].

Thus, the literature supports three foundational conclusions for the present problem setting: (1) formation-damage mechanisms are heterogeneous and depend on stress, mineralogy, and the fluid environment, which necessitates a locally validated risk model; (2) operational applicability requires ML models that are robust to drift and missingness, provide calibrated probabilities, and employ a regulated detection threshold; (3) real value emerges only when the detector is embedded into a DSS/CMMS workflow with explicit SLAs—from alarm visualization to formalization of a response “playbook”.

Pore-space clogging is widely recognized as one of the key mechanisms responsible for deterioration of reservoir flow and storage properties and for reduced well productivity. Fundamental studies show that deposition of a dispersed phase reduces porosity and effective permeability in the flow zone, alters displacement patterns and pressure distribution, and that even “small” values of clogging parameters can produce a noticeable effect on displacement-front dynamics and near-wellbore response [10]. The temporal evolution is described via the formation of a “stabilized zone” behind the front: impurity concentration gradually decreases to zero, whereas porosity increases from a minimum stationary value toward the initial level; the width of this zone remains practically constant for fixed medium and fluid parameters [10].

More advanced formulations focused on the near-wellbore region consider flows with a moving boundary, namely the penetration front of the suspension. In such models, continuity conditions and kinematic relations at the interface reveal the spatial non-uniformity of permeability degradation and the sensitivity of the local NWB state to the boundary propagation velocity. Importantly, the boundary corresponding to complete depletion (“exhaustion”) of the suspension lags behind the penetration front: between them, a zone forms where porosity and permeability change most intensely. This directly explains the “hard” near-wellbore region effect and the increased pressure drops at the same flow rates [11]. These results limit the applicability of coarse approximations such as deep-bed filtration immediately at the wellbore and underscore the importance of correctly interpreting measurements and forecasts under transient regimes [11].

From an engineering perspective, formation damage manifests as an increase in additional hydrodynamic resistance (skin effect) caused by plugging of flow channels by particles originating from drilling and cement slurries, rock failure products, and silt deposits. The largest energy losses are localized in the near-wellbore zone, which is supported by studies of wells under unsteady (transient) conditions. Contamination sources accompany virtually all stages of the well life cycle: reservoir penetration, cementing, perforation, well cleanup and kill operations, subsequent technological interventions, and routine production; each intervention can introduce additional solid phases and modify the filtration properties of the near-wellbore zone [12].

For gas-condensate systems, the picture is complicated further by phase behavior. When pressure falls below the dew-point in the near-wellbore zone, a "condensate bank" forms, shifting flow to a three-phase regime and restricting phase permeabilities; local condensate saturation at the wellbore may exceed areal-average values by multiples within only a few years of production. This leads to a substantial decline in productivity and requires joint consideration of formation damage and phase transitions when interpreting rate and pressure dynamics [12]. Hydraulic fracturing can redistribute the saturation profile and drawdown, but it does not fully eliminate the banking effect: part of the condensate precipitates along the fracture and must be represented in models to obtain a correct assessment of long-term productivity [12].

In summary, based on [10–12], the modern understanding of formation-damage mechanisms integrates: (i) a physically grounded link between contamination and the evolution of porosity/permeability, including stabilized zones behind the front (as established by calculations and analyses); (ii) the role of a moving boundary and a suspension "buffer" in the near-wellbore vicinity, forming three characteristic sublayers with different intensities of property changes; and (iii) multifactor field causes of near-wellbore contamination, amplified by phase transitions in gas-condensate environments. These findings constitute the basis for applied risk forecasting: correct specification of degradation events, selection of informative operational features (pressure, rates, GOR/WC, etc.) without future leakage, time-aware validation, and subsequent calibration of probabilistic estimates on a defined horizon (14 days).

3. Methodology. The object of the study is the producing well J12. The target event is defined as a drop of the productivity index below 0.90 occurring 14 days after the forecast timestamp, which is consistent with the typical operational planning window for diagnostic and corrective actions. The data are provided as daily time series of key operating parameters: water cut (WC), net GOR, wellhead pressure (WHP), choke setting (Choke), gas–liquid ratio (GLR), and oil and total liquid production rates. All channels are synchronized onto a single daily time grid; missing values are imputed via last observation carried forward, with an additional binary missingness flag introduced as a separate feature. To ensure robustness to outliers, a robust anomaly-labeling procedure is applied without altering ("correcting") the original values. Feature scaling is performed using a robust scheme and is fitted only on training windows, then transferred unchanged to validation and test splits, thereby preventing future information leakage.

The feature representation comprises current parameter levels, their lags of 1/3/7/14/30 days, rolling means and standard deviations over 7/14/30-day windows, and simple trend descriptors, including day-to-day increments and linear-regression slope estimates computed over the same window intervals. All windows are constructed strictly over the interval $[t-W+1; t]$, i.e., using only information available at the decision time. Positive-class labeling corresponds to the occurrence of the event "IC < 0.90" within a 14-day horizon. For each evaluation window, event prevalence is recorded for the training, validation, and test splits; it is used as a "climatological" reference when computing probabilistic metrics (Brier score and Brier Skill Score) and when comparing decision utility.

Potential class imbalance is addressed through class weighting during training, prioritization of PR-based metrics, and selection of the operational threshold via a cost function. Validation is organized as rolling-origin backtesting: the historical record is partitioned into multiple sequential "train → validation → test" windows, where hyperparameters are tuned on the validation segment and the resulting fixed configuration is applied to the subsequent test interval; final metrics are aggregated across windows using the median and 95% confidence intervals obtained via bootstrap resampling over windows. The model family includes L2-regularized logistic regression, random forest, and gradient boosting (CatBoost/LightGBM, selected based on validation performance). For interpretation, permutation importance and partial dependence analyses (PD/SHAP) are used for tree-based models and boosting.

Probabilistic outputs are calibrated using an out-of-fold scheme with Platt scaling and isotonic regression; calibration quality is assessed via the Brier score, expected calibration error (ECE), and calibration plots. The primary diagnostic metrics are PR-AUC (preferred under imbalance), ROC-AUC, and F2, together with the Brier score and ECE. The operational threshold $\tau*$ is selected by minimizing the expected cost $\mathcal{L}(\tau) = C_{FN} \cdot FN(\tau) + C_{FP} \cdot FP(\tau)$ under asymmetric error costs that prioritize avoiding missed events; decision correctness and applied usefulness are verified using the decision curve and lift curve. For subsequent deployment, the workflow includes monitoring of distribution drift in key features (e.g., PSI/KS), calibration monitoring (ECE), and a recalibration/retraining protocol triggered when metrics exceed predefined tolerance limits.

4. Experimental Section. In this study, forecasting the event “IC < 0.90” over a 14-day horizon for well J12 is formulated as a binary classification problem. A single unified feature set is used, comprising:

- baseline parameters (WC, Net GOR, WHP, Choke, GLR, production rate);
- lagged values (1/3/7/14/30 days);
- rolling statistics (moving average/standard deviation over 7/14/30-day windows).

This expanded feature space captures both the current state and the temporal dynamics of the parameters, which is particularly important when predicting degradation of the productivity index.

To verify the adequacy of the problem formulation, the distribution of the target label was analyzed. The class balance revealed a pronounced shift toward the “normal state” (IC ≥ 0.90), confirming the need for machine-learning methods capable of handling sample imbalance (table 1).

Table 1. Balance classes

split	**positive_rate**	**sum**	**count**
test	0.586667	88	150
train	0.525862	61	116
val	0.826923	43	52

A preliminary evaluation of logistic regression and random forest models was then conducted. The resulting validation and test metrics indicated that ensemble methods achieve higher sensitivity; however, they require probability calibration to ensure correct and reliable use in production practice (table 2).

Table 2. Summary metrics of models (validation/test)

model	split	roc_auc	pr_auc	f1@0.5	precision@0.5	recall@0.5
LogReg	train	0.993741	0.994536	0.934426	0.934426	0.934426
LogReg	val	0.364341	0.810764	0.208333	1	0.116279
LogReg	test	0.483871	0.615316	0.188235	1	0.103896
RandomForest	train	0.922504	0.947806	0.861789	0.854839	0.868852
RandomForest	val	0.945736	0.986841	0.94382	0.913043	0.976744
RandomForest	test	0.678467	0.76476	0.689655	0.618557	0.779221

For a visual assessment of forecast quality, ROC and PR curves were constructed for the validation and test sets. The validation curves are shown in Figs. 1–4, while the test curves are shown in Figs. 5–8. Logistic regression and the random forest demonstrated comparable performance; however, the forest achieved a higher recall level at a fixed false-alarm rate.

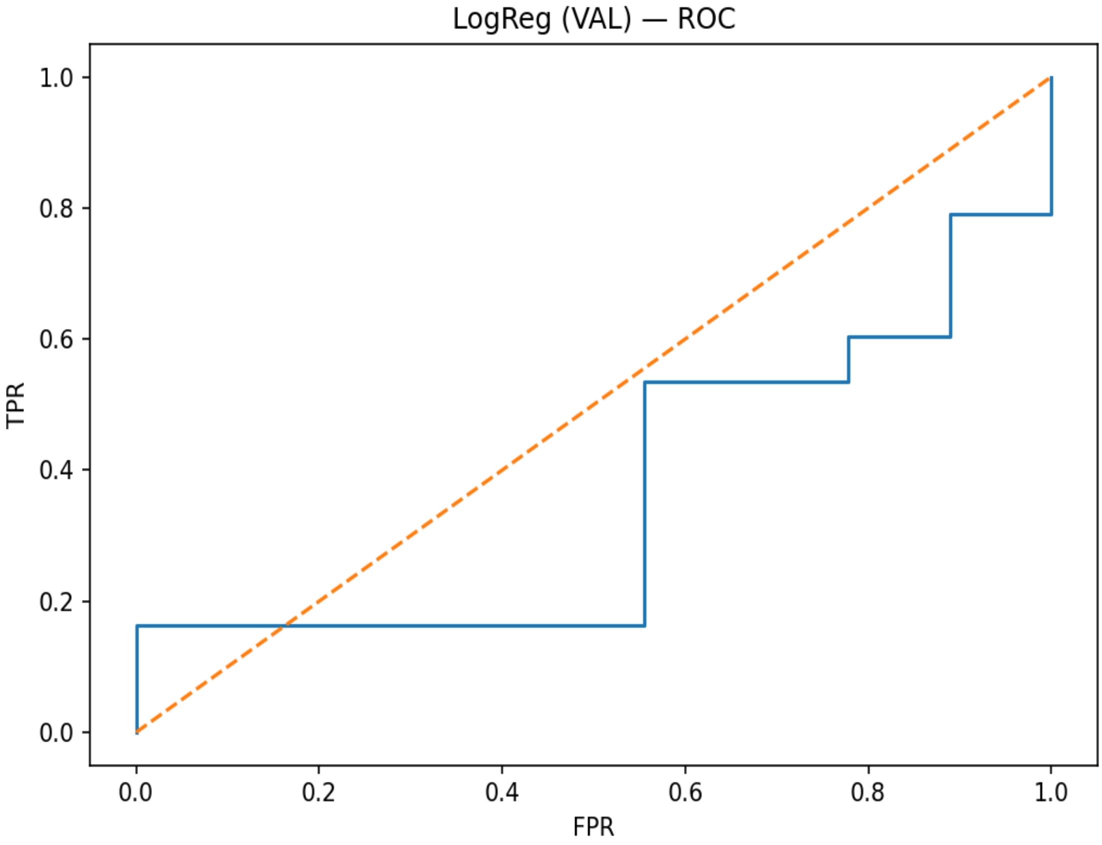


Fig. 1. ROC curve (validation), logistic regression

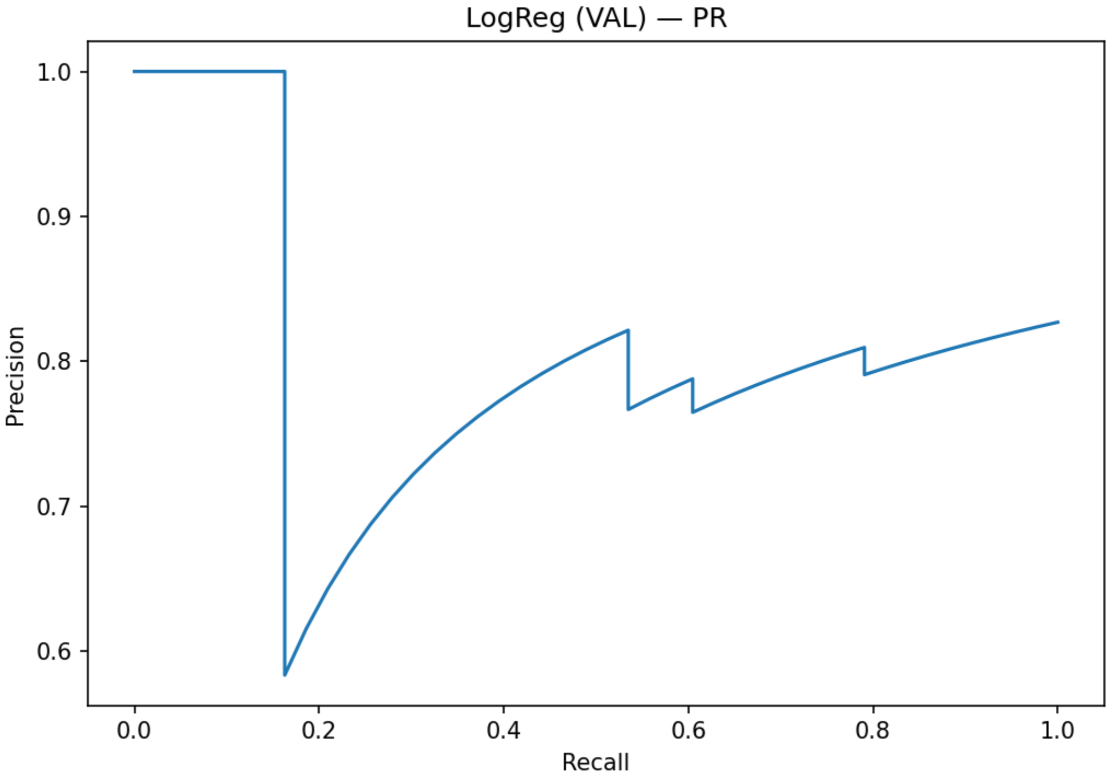


Fig. 2. PR curve (validation), logistic regression

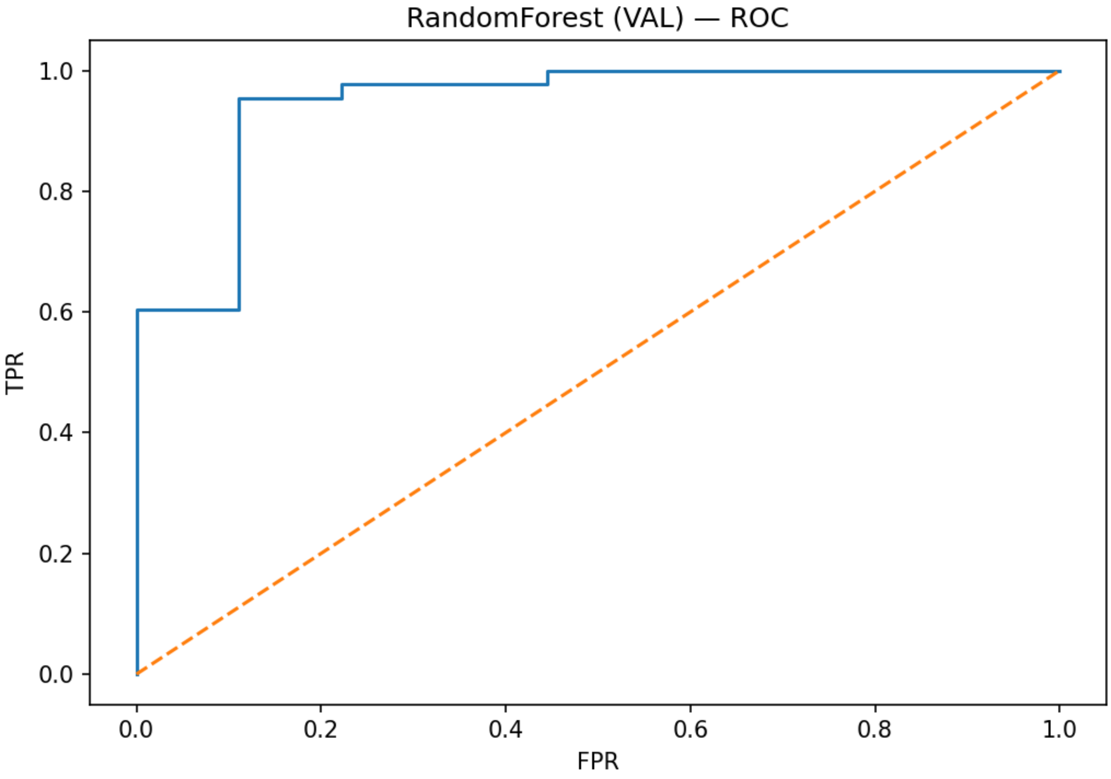


Fig. 3. ROC curve (validation), random forest

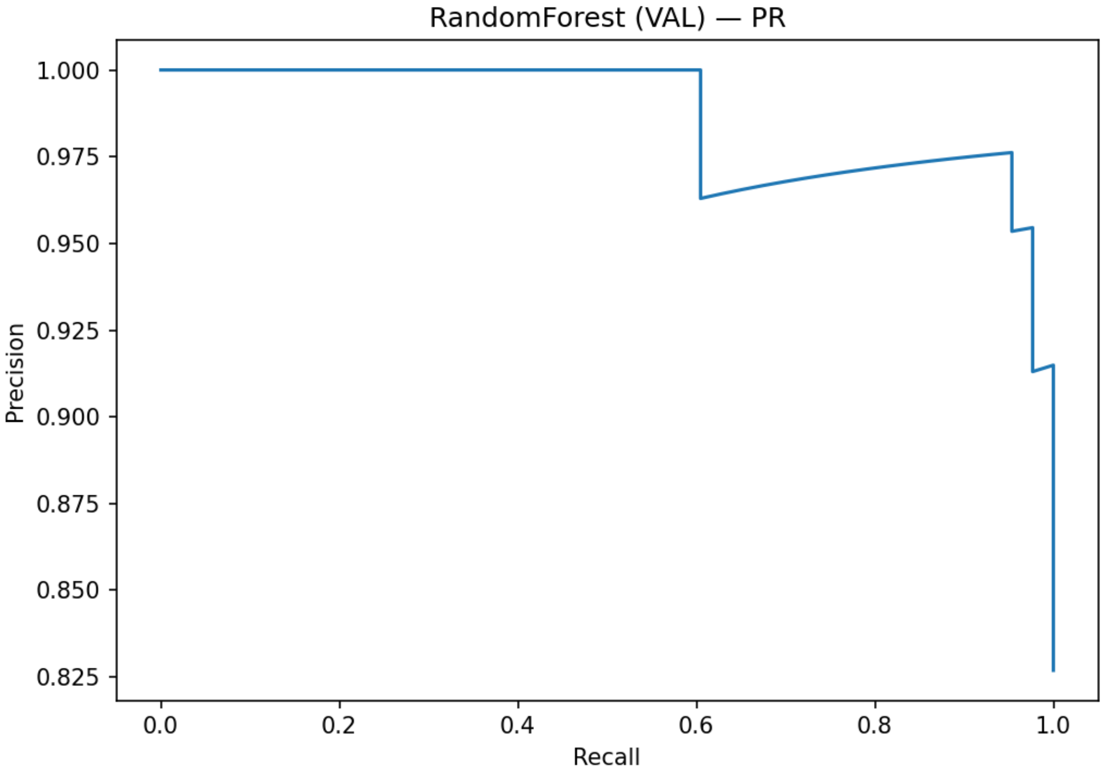


Fig. 4. PR curve (validation), random forest

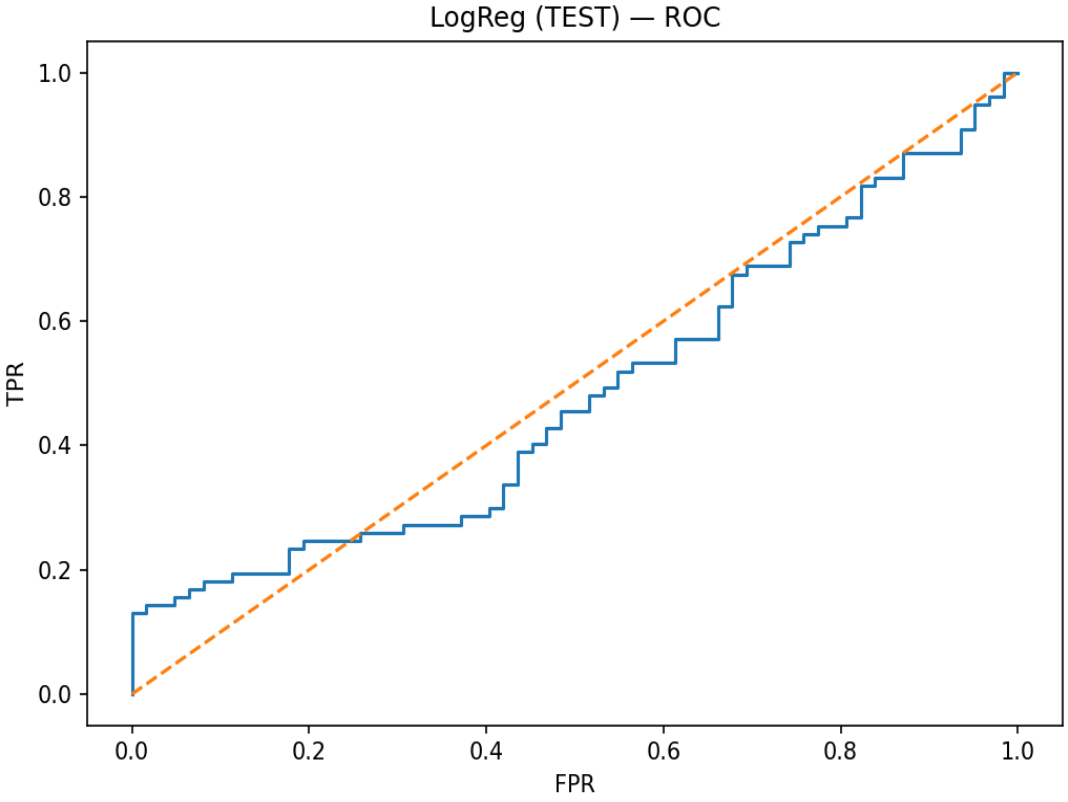


Fig. 5. ROC curve (test), logistic regression

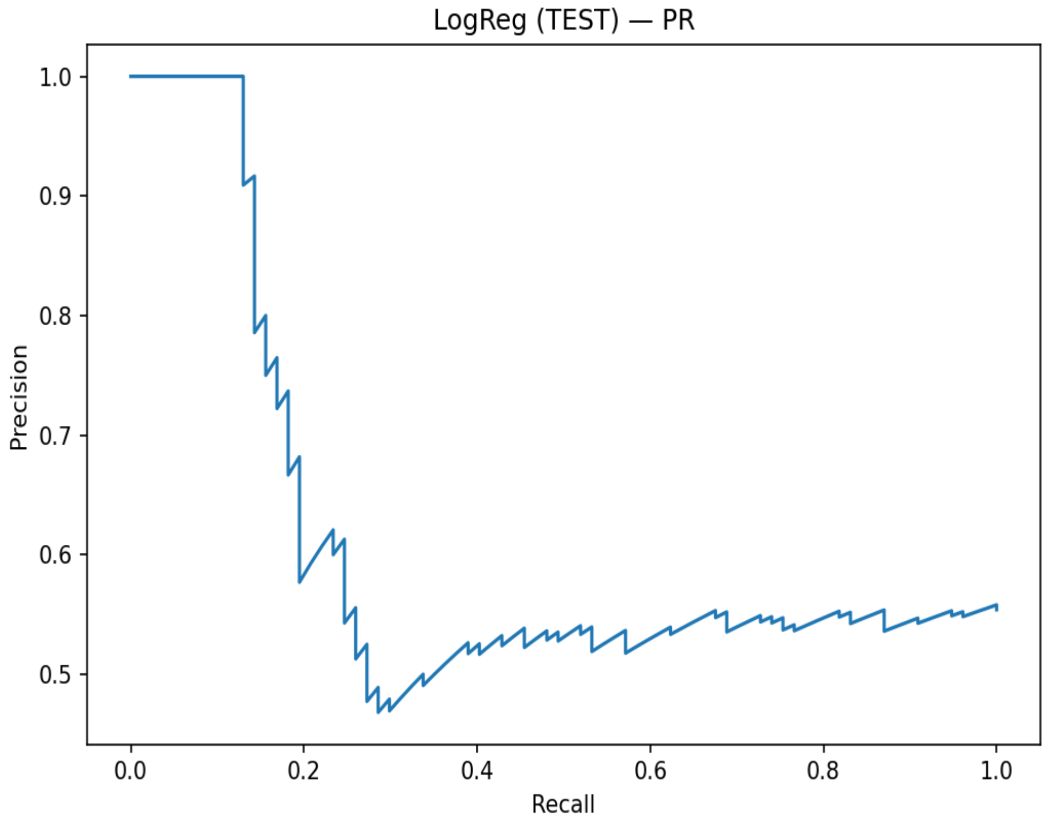


Fig. 6. PR curve (test), logistic regression

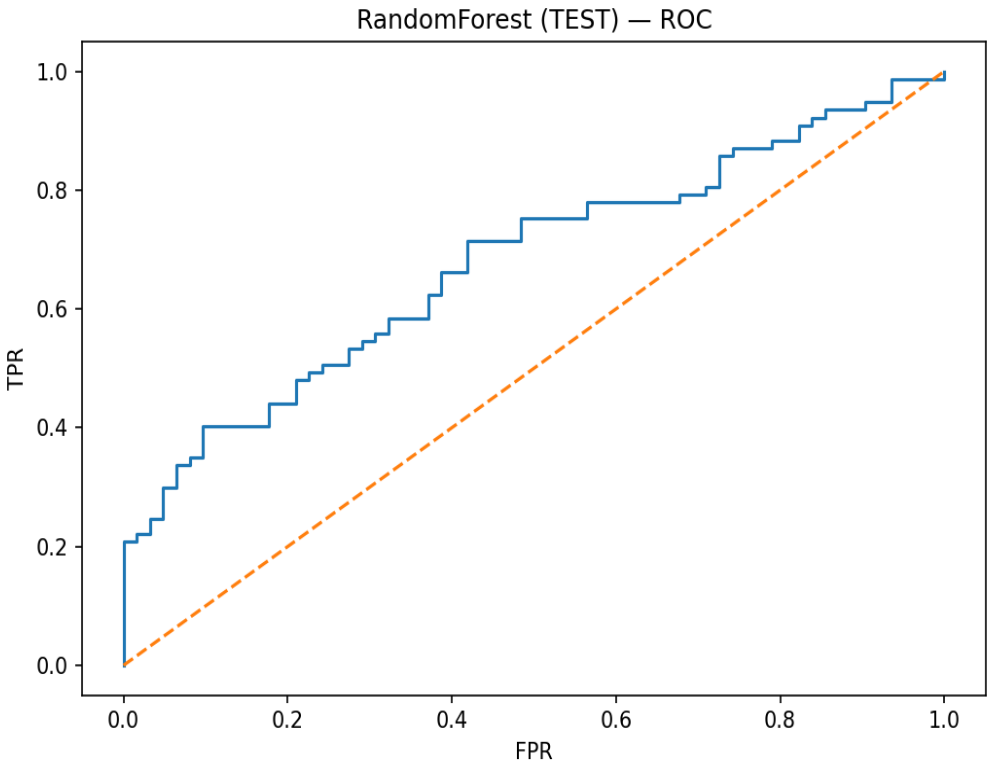


Fig. 7. ROC curve (test), random forest

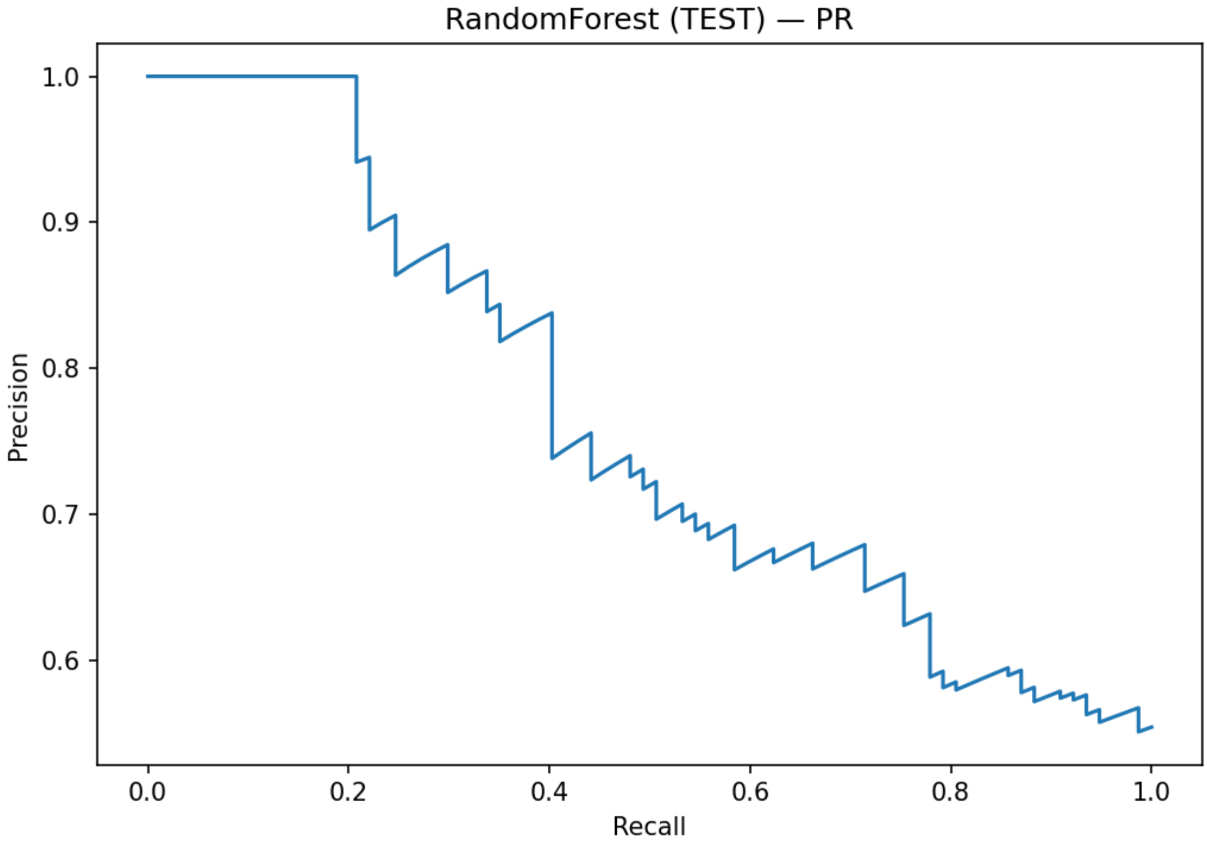


Fig. 8. PR curve (test), random forest

On the test set, the models exhibit an expected degradation in performance relative to validation, which is attributable to temporal drift in the data. Nevertheless, the overall structure of the ROC and PR curves is preserved, confirming the transferability of the identified patterns and providing justification for applying probability-calibration procedures.

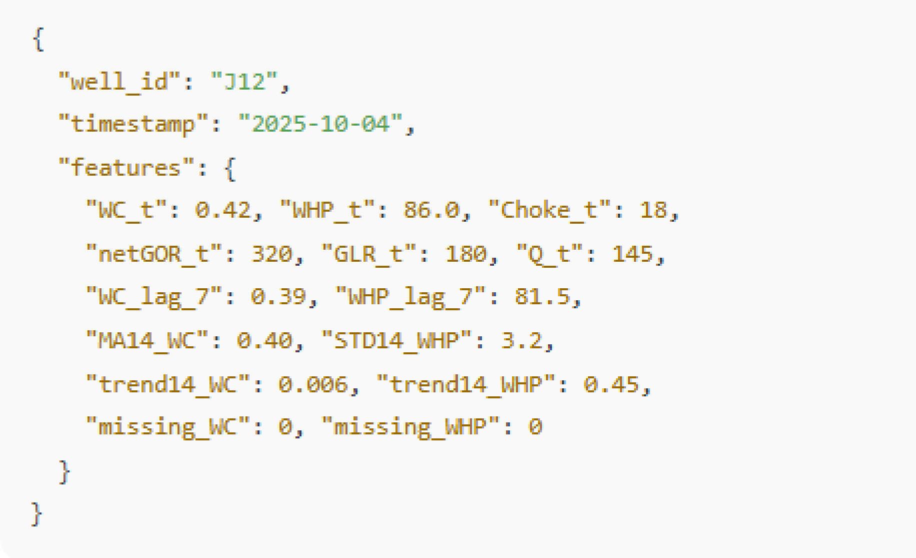
Thus, the literature review and the initial results indicate that applying machine-learning methods to forecast reservoir formation damage is both justified and practically significant. The developed models identify key risk drivers, enable the generation of early warning signals, and provide a foundation for the next stage—probability calibration and selection of the operational threshold τ∗.

5. Results. The rolling-window analysis revealed a moderate drift in both levels and variability across the WC, WHP, Choke, net GOR, and GLR channels, confirming the need for time-aware validation and periodic recalibration. The prevalence of the event “IC < 0.90” remains substantially below 0.5; therefore, precision–recall metrics serve as the primary reference, while ROC-based indicators are used as supplementary diagnostics. Without calibration, ensemble methods deliver higher recall in the low-threshold region compared with logistic regression, whereas the logistic model often exhibits a “smoother” ROC behavior; within the operationally relevant threshold range, the advantage in PR performance typically remains with ensembles. Across all models, test intervals show the expected deterioration relative to validation due to concept drift, yet the qualitative shape of PR/ROC curves is largely preserved, indicating that the identified patterns generalize over time.

Out-of-fold probability calibration reduces both the Brier score and expected calibration error (ECE); in the critical probability range of 0.2–0.6, predicted risks become closer to observed event frequencies. This gives the probabilities direct operational meaning: for instance, a risk forecast around 0.3 corresponds to an observed event frequency of approximately 30% within the respective group and can therefore be used for intervention planning. The operational threshold τ∗ is selected by minimizing expected cost under asymmetric losses between missed events and false alarms, and it is consistent with the allowable dispatcher workload. Within the working threshold range, the model demonstrates positive net benefit relative to the “alert-all” and “alert-none” strategies, while the event distribution across probability quantiles indicates targeted actionability:the highest-probability groups concentrate a disproportionately large share of actual episodes.

Ablation experiments show that the dominant contribution to performance comes from explicitly capturing dynamics—lags, rolling means and standard deviations, and trend coefficients; the effect is particularly pronounced for WC, WHP, and Choke, where gradual (“creeping”) changes are most informative. Interpretability analysis (PD/SHAP) is consistent with the underlying physical mechanisms: increased water cut, rising wellhead pressure at an unchanged choke setting, and unfavorable shifts in net GOR/GLR are associated with a higher probability of a productivity-index decline within the 14-day horizon. A robustness check under the scenario “training on an early period—application to a later period” confirms an expected metric drop without adaptation and a noticeable improvement after scheduled recalibration on a recent validated segment, supporting the necessity of continuous monitoring of calibration and drift. Compared with simple engineering baselines (e.g., EWMA triggers on IC and threshold rules on ΔWHP at fixed Choke), the proposed approach provides an earlier and more stable signal at a comparable operational workload, thereby reducing the risk of missing formation-damage episodes, stabilizing the productivity index, and mitigating production losses.

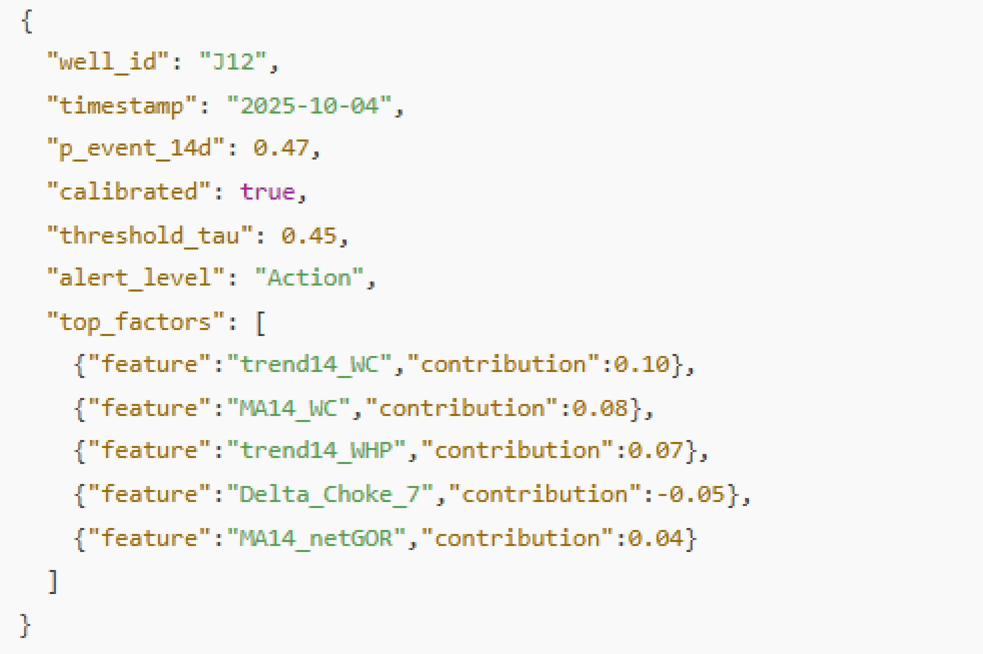
6. Deploying the Model into the Operational Workflow. Deployment of the model into the operational workflow is organized around a simple, reproducible cycle: “data → forecast → alert → action → feedback.” The input parameters (WC, WHP, Choke, net GOR, GLR, and production rates) are retrieved from SCADA/industrial process control systems and historian-type repositories such as PI Historian, aggregated onto a daily grid, time-synchronized, and imputed via last observation carried forward with an explicit missingness flag. Subsequently, 1/3/7/14/30-day lags are computed, along with rolling means and standard deviations over 7/14/30-day windows, as well as simple trend estimates for the key channels. An example of the daily input feature package used by the forecast service is provided in Listing 1.



```
{
  "well_id": "J12",
  "timestamp": "2025-10-04",
  "features": {
    "WC_t": 0.42, "WHP_t": 86.0, "Choke_t": 18,
    "netGOR_t": 320, "GLR_t": 180, "Q_t": 145,
    "WC_lag_7": 0.39, "WHP_lag_7": 81.5,
    "MA14_WC": 0.40, "STD14_WHP": 3.2,
    "trend14_WC": 0.006, "trend14_WHP": 0.45,
    "missing_WC": 0, "missing_WHP": 0
  }
}
```

Listing 1. Example of an input feature package for the forecast service (daily slice)

The model is trained on historical windows using rolling-origin time-aware validation; probabilities are calibrated outside the training fold (Platt scaling/isotonic regression) and stored together with the corresponding calibrator version. In the operational workflow, the forecasting service runs once per day, assigns each well a probability of the event “IC < 0.90 within a 14-day horizon,” applies calibration, and compares the result with the operational threshold $\tau*$, selected by minimizing expected cost under asymmetric error costs (a missed event is more expensive than a false alarm). An example of the forecast-service response (including the calibrated probability and decision flag) is shown in Listing 2.



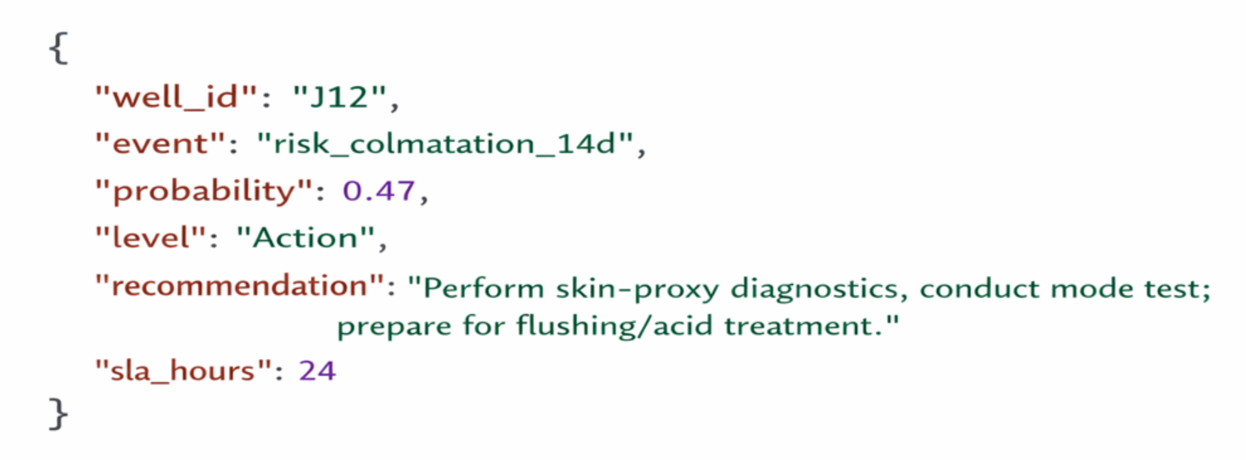
```
{
  "well_id": "J12",
  "timestamp": "2025-10-04",
  "p_event_14d": 0.47,
  "calibrated": true,
  "threshold_tau": 0.45,
  "alert_level": "Action",
  "top_factors": [
    {"feature":"trend14_WC","contribution":0.10},
    {"feature":"MA14_WC","contribution":0.08},
    {"feature":"trend14_WHP","contribution":0.07},
    {"feature":"Delta_Choke_7","contribution":-0.05},
    {"feature":"MA14_netGOR","contribution":0.04}
  ]
}
```

Listing 2. Example of service response

To mitigate “alert fatigue,” tiered signal levels are introduced: information at moderate risk, followed by telemetry verification and planning of additional measurements; watch at elevated risk, with targeted diagnostics of WC/WHP/GLR trends and an operating-mode test; and action when the risk exceeds $\tau*$, at which point the DSS automatically generates an action card containing a recommendation (e.g., skin-proxy diagnostics, flushing, acidizing, or solvent treatment), response deadlines, and a brief explanation of the key drivers based on top contributing factors (for example, WC and WHP trends) (table 3). An example of an alert entry in the DSS/CMMS with a recommendation and SLA is provided in Listing 3.

Table 3. Colmatation risk response playbook for the 14-day forecast

Signal level	**Condition**	**Actions (example)**	**Response time (SLA)**
Info	0.20–0.30	Telemetry check; verification of Choke/WHP; scheduling an additional measurement	24–48 h
Watch	0.30–0.45	Analysis off WC/WHP/GLR trends; a limited operating-mode test; preparation for flishing	24 h
Action	≥ τ* (example: 0.45)	Operating-mode adjustment + rapid skin-proxy diagnostics; decision: flush / acidizing / solvent treatment	12–24 h



```
{
   "well_id": "J12",
   "event": "risk_colmatation_14d",
   "probability": 0.47,
   "level": "Action",
   "recommendation": "Perform skin-proxy diagnostics, conduct mode test;
                  prepare for flushing/acid treatment."
   "sla_hours": 24
}
```

Listing 3. Example of an alarm entry in DSS/CMMS with a recommendation and SLA

To ensure signal stability, hysteresis is applied (an alert is cleared when the probability drops below τ*−δ) together with an “n out of m” days rule, while repeated triggers are suppressed during active intervention periods (cool-down).

Deployment is executed in stages. First, the model runs in “shadow mode”: it generates daily predictions, but alerts are visible only to the project team. This enables verification—on live operational data—of calibration quality, PR-AUC within the operational threshold region, and expected cost, without affecting well operating regimes. Next, a subset of the asset (10–20 wells) is piloted with an engineer in the decision loop (“go/no-go” for an intervention); the threshold τ∗ is refined based on the decision curve and the actual operational workload. After agreed KPI targets are achieved (false-alarm rate, episode detection recall, response time, and expected savings), coverage is expanded to the full well stock.

In parallel, a monitoring framework for performance and drift is established. On a daily basis, the alert rate, retrospective FN/FP, and Brier/ECE are tracked; weekly, PR-AUC, lift, and decision curves are updated; and distribution stability of key features is controlled (e.g., PSI for WC, WHP, and Choke). If calibration or diagnostic metrics deteriorate, recalibration is triggered (typically every 2–4 weeks or on-demand); in the presence of pronounced drift, the model is retrained on recent windows. A “safe” fallback to simple baseline rules (EWMA/threshold triggers) is always available so that operations are not disrupted.

The economic component is tied directly to the selection of τ∗. The relative costs of missed events and false alarms are estimated in advance; an expected-cost curve over the threshold is constructed on historical data; τ∗ is chosen at the minimum and validated via the decision curve against the “alert-all” and “alert-none” strategies. During operation, expected savings are recomputed monthly by comparing the impact of early interventions on production rate and OPEX against a baseline pre-deployment period. Full traceability is maintained in a version-controlled manner: data schema, model and calibrator versions, the value of τ∗, input features and explanations for each alert, the decision taken, and the realized outcome.

This operating mode provides the engineering team with a quantitatively interpretable 14-day-ahead risk estimate, enables integration of signals into response procedures and work planning, and sustains performance through regular recalibration and drift control. As a result, the operational workflow receives not merely an “anomaly detector,” but a governed process: early warning of persistent formation damage, targeted and timely interventions, reduced production losses at a controlled false-alarm burden, and transparent decision economics.

7. Limitations and Future Work. The adopted binary-classification formulation captures whether an event occurs within the specified horizon, but it does not explicitly represent time-to-event nor the recurrence of episodes. A promising direction is to move toward survival analysis, in particular discrete-time hazard models and survival boosting, which naturally handle censoring and allow risk estimation at each time step. Using data from a single well limits generalizability; to scale to a pad and field level, inter-well cross-validation, stratification by lithology and operating regimes, as well as hierarchical models or transfer learning are required to adapt the model under a limited volume of fresh data. In operational settings, the threshold τ∗ and calibration itself must adapt to drift; practically, this motivates active-learning loops for targeted manual verification of “informative” cases, online probability recalibration when distributional shifts are detected, and threshold adaptation based on a rolling estimate of error costs. Finally, incorporating physical proxies (e.g., IPR-based approximations of skin or PI) into hybrid ML models improves interpretability and robustness outside the training distribution, enabling more reliable DSS performance across diverse geological and operational conditions.

8. Conclusion. The proposed approach demonstrates that calibrated probabilistic machine-learning models can provide early and interpretable warning of elevated formation-damage risk over a 14-day horizon. A dynamic feature representation (lags, rolling statistics, and trends) together with rolling-origin time-aware validation helps preserve diagnostic capability under concept drift. Out-of-fold probability calibration reduces the Brier error and expected calibration error, giving the forecast a direct operational meaning: the predicted risk probability becomes comparable to the observed event frequency within the corresponding risk group. Selecting the operational threshold τ∗ via expected-cost minimization—supported by utility analysis and the demonstrated targeting of interventions—translates the model from an exploratory setting into an applied decision-making workflow. Compared with simple engineering baselines, the proposed method delivers an earlier and more stable signal under a controlled operational workload, reducing the likelihood of missed episodes, stabilizing the productivity index, and mitigating production losses. The deployment protocol (daily forecasting service, tiered alert levels, calibration and drift monitoring, and scheduled recalibration/retraining) makes the solution reproducible and suitable for integration into field-level DSS.

Key Findings

1. The model provides an early and interpretable warning of 14-day formation-damage risk, suitable for operational decision-making.
2. Probability calibration (Platt scaling/isotonic regression, OOF) reduces Brier score and ECE; therefore, the probabilities can be used directly within operational procedures.
3. Under low event prevalence, PR-based metrics are the primary reference, while ROC-based metrics are auxiliary.
4. The optimal threshold τ∗, selected via cost minimization, yields positive net benefit and a controlled dispatcher workload.
5. Dynamic features (lags, rolling MA/STD, and trends) deliver the main performance gains and enable targeted interventions compared with simple engineering rule-based baselines.

Conflict of interest.

The authors declare that they have no conflict of interest in relation to this research.

References

1. Garaeva, A.N., Korolev, E.A., Khramchenkov, M.G. Some peculiarities of pore space colmatation process in stress-heterogeneous clayey reservoirs // Neftyanoe Khozyaystvo – Oil Industry. 2017. No. 8. pp. 72–74. DOI: 10.24887/0028-2448-2017-8-72-74.
2. The rationale for selection of technologies and formulations of reagents for restoring the productivity of horizontal wells in the Vankor field // Neftyanoe Khozyaystvo – Oil Industry. 2019. No. 8. pp. 130–135. DOI: 10.24887/0028-2448-2019-8-130-135.
3. Development of Acidizing Simulator for Sandstone Reservoirs // SPE Russian Petroleum Technology Conference. 2020. Paper SPE-201951-MS. DOI: 10.2118/201951-MS.
4. Oil Production Prediction Using Time Series Forecasting Models (case study) // SPE Nigeria Annual International Conference and Exhibition. 2024. Paper SPE-221728-MS.
5. Machine Learning Approaches for Pattern Recognition and Missing Data Prediction in Field Datasets from Oil and Gas Operations // GOTECH Conference. 2024. Paper SPE-219384-MS. DOI: 10.2118/219384-MS.
6. Solution Gas/Oil Ratio Prediction from PVT Data Using Machine Learning Algorithms // SPE Journal. 2023. DOI: 10.2118/217979-PA.
7. Modelling of Injection Well Capacity with Account for Geomechanics and Colmatation Factors // SPE Russian Petroleum Technology Conference. 2017. Paper SPE-187806-MS. DOI: 10.2118/187806-MS.
8. Data-Driven Prediction of Oil Well Productivity // OMC Med Energy Conference. 2025. Paper OMC-2025-657. OnePetro.
9. Olson, R.S., La Cava, W., Mustahsan, Z., Varik, A., Moore, J.H. Random forest versus logistic regression: a large-scale benchmark for binary classification // BMC Bioinformatics. 2018.
10. Arkhipova, E.N., Gilmanov, A.Ya., Shevelev, A.P. Modeling of porous-medium colmatation during injection of particle-laden water // Computational Continuum Mechanics. 2023. Vol. 16, No. 2. pp. 171–178. DOI: 10.7242/1999-6691/2023.16.2.14.
11. Kapranov, Yu.I. Features of colmatation in the near-wellbore zone of a reservoir // Bulletin of KazNU. Series: Mathematics, Mechanics, Informatics. 2011. No. 1(68). pp. 117–123

Sheshukov, S.V. Impact of colmatation on gas-condensate well productivity // Educational/Scientific material, Tyumen Industrial University.