International Scientific Conference 20 UNITECH 26

INTERNATIONAL SCIENTIFIC CONFERENCE
19-21 November 2026
GABROVO

FIELD-AWARE OPTIMISATION AND SUCCESS-RATE BOUNDS FOR DOCUMENT OCR

Plamen Nakov*

Technical University of Sofia, Bulgaria
**Corresponding author: p.nakov@tu-sofia.bg*

Abstract
Improving an individual OCR field does not uniquely determine the probability that every required field is correct. This paper develops field-aware performance budgets for a machine-readable zone and a document number, using reported baseline accuracies of 98.60%, 81.10% and 81.10% joint correctness as analytical inputs. Standard intersection bounds show that increasing document-number accuracy to 95% permits joint correctness between 93.6% and 95%, but not above 95%. At fixed machine-readable-zone accuracy, 96.4% document-number accuracy is sufficient for a 95% joint target regardless of error dependence in the point-probability model. A recovery–regression model shows that, without newly introduced errors, 79.43% of the cases with a correct machine-readable zone and an incorrect number must be repaired. A 2% regression rate increases this requirement to 88.70%. The analysis is extended to multiple required fields and to the distinction between full-population correctness and selective automatic processing. A component-wise validation protocol translates these conditions into testable localisation and recognition upgrades. No improved OCR configuration or new empirical dataset is claimed.

Keywords: document OCR; joint correctness; region of interest; error dependence; optimisation.

INTRODUCTION

A document-processing application may require several fields to be exactly correct before a record is usable. Improving the average score of one field is then not sufficient to predict the final success rate. What matters is the overlap between successful cases: an upgrade helps joint correctness only when it repairs a missing requirement without invalidating another one.

A technical tests reports PaddleOCR accuracies of 98.60% for a machine-readable zone (MRZ), 81.10% for the document number and 81.10% for their joint correctness. It identifies localisation as a possible limitation and anticipates substantial gains after improving the number field. Those observations motivate a quantitative question: what field-level improvements would be sufficient for a joint target of at least 95%?

This paper answers that question through probability bounds, a recovery–regression model and a proposed component-wise validation protocol. The reported aggregates serve as point inputs to transparent scenarios, not as a second independent experiment. The bounds themselves are standard probability results; the contribution is their application to OCR upgrade requirements, including error-dependence sensitivity, preservation of correct cases and extension to additional mandatory fields.

The analysis does not implement a new detector, fine-tune a recogniser or establish a deployment success rate. It distinguishes three quantities that are often conflated: field accuracy, joint correctness over the full population, and correctness among records selected for automatic processing. Keeping them separate turns a projected improvement into a testable engineering requirement.

JOINT-CORRECTNESS MODEL

Events and admissible dependence

Let M denote exact MRZ transcription and D exact document-number transcription

This is an open access article licensed under Creative Commons Attribution 4.0 International
Print ISSN: 1313-230X, Online ISSN: 3033-1404

on the same population or declared finite sample. Write $p_M = \mathrm{P}(M)$, $p_D = \mathrm{P}(D)$ and $q = \mathrm{P}(M \cap D)$. The matching convention and eligibility criteria must remain identical across configurations. Neither a plausible format nor a successful check-digit calculation defines ground-truth correctness.

The four mutually exclusive outcomes have probability masses q, $p_M - q$, $p_D - q$ and $1 - p_M - p_D + q$. Their non-negativity gives the standard intersection bounds:

$$\max(0, p_M + p_D - 1) \le q \le \min(p_M, p_D) \quad (1)$$

The upper bound follows because the intersection is contained in each event. The lower bound follows because their union cannot exceed one. Both endpoints are attainable: suitable dependence can place every success of the less accurate field inside the other success set, or can minimise the overlap while preserving both marginals. Thus the interval describes genuinely different admissible populations, not uncertainty caused by numerical approximation.

Only under an additional independence assumption is joint correctness the product $p_M\, p_D$. Shared blur, glare, perspective and crop quality provide reasons to investigate dependence rather than assume it away. A shared recogniser or preprocessing stage can also make the two errors related. No independence assumption is required for (1).

At the reported baseline, (1) gives $0.797 \le q \le 0.811$. Independence would give 0.799646. The reported joint rate lies at the upper endpoint at the displayed precision. Since the percentages are rounded and per-image outcomes are unavailable, this numerical equality is not evidence of exact nesting in the original sample. Subsequent scenarios use the stated probabilities as analytical point values only.

A counterexample to an optimistic projection

Suppose an upgrade raises document-number accuracy to 95% and leaves MRZ accuracy at 98.6%. Then (1) restricts joint correctness to 93.6–95.0%. Table 1 constructs two possible populations with those same marginal accuracies. Their different joint results arise entirely from the overlap of errors.

Table 1. Analytical outcome masses, %, for p_M = 98.6% and p_D = 95.0%; not observed counts.

Outcome	Nested success	Minimum overlap
Both correct	95.00	93.60
Only MRZ correct	3.60	5.00
Only number correct	0.00	1.40
Neither correct	1.40	0.00

In the nested-success case, every correct number accompanies a correct MRZ. In the minimum-overlap case, the 1.4% MRZ failures occur where the number is correct. Both satisfy the field targets, but the joint rates differ by 1.4 percentage points. Independence gives 93.67%, close to the lower bound in this particular scenario. None of these calculations predicts which dependence a trained system will exhibit.

The immediate implication is exact: number accuracy of 95% cannot establish joint correctness above 95% under the same two-field requirement. Better localisation may still be valuable; the projection must simply be evaluated against a valid joint metric instead of being inferred from an isolated field score.

PERFORMANCE TARGETS AND BUDGETS

Necessary and dependence-robust targets

For a target $q \ge \tau$, a necessary condition is that both marginal accuracies reach τ. When $p_M \ge \tau$, a sufficient condition based only on the marginals follows from the lower bound:

$$p_D \ge 1 + \tau - p_M \quad (2)$$

For $\tau = 0.950$ and $p_M = 0.986$, the sufficient number target is 0.964. Accuracy between 95.0% and 96.4% may still achieve the joint target under favourable dependence; it cannot ensure it from the marginals alone. When independence is explicitly assumed, the corresponding threshold is:

$$p_D \ge \frac{\tau}{p_M} \quad (3)$$

Equation (3) gives approximately 96.35%, rather than 96.40%. The difference is due to the added assumption, not to a measured improvement. Figure 1 displays the complete admissible interval and the independent scenario as the number accuracy varies.

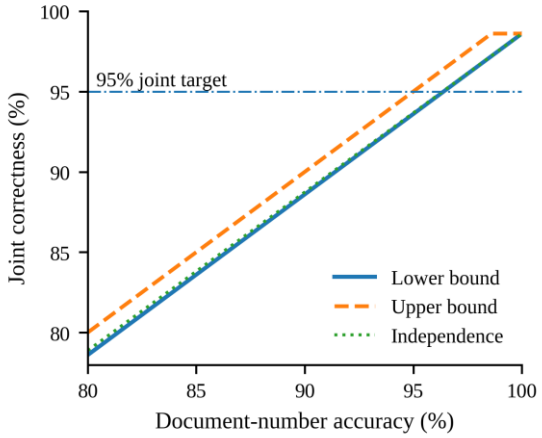


Fig. 1. *Joint-correctness bounds at p_M = 98.6%. Independence is an explicitly conditional scenario; the curves are analytical, not measured upgrade results*

Table 2 gives sensitivity to the accuracy of the supposedly easier MRZ field. Raising that accuracy reduces the number-field target one-for-one under (2). Conversely, an MRZ regression consumes an equal part of the permissible number-error budget. If $p_M < \tau$, no improvement to the number alone can attain the target.

Table 2. Number-accuracy thresholds, %, for a 95% joint target. All rows are analytical scenarios.

MRZ accuracy	Sufficient, any dependence	Under independence
98.00	97.00	96.94
98.60	96.40	96.35
99.00	96.00	95.96
99.50	95.50	95.48

All thresholds above refer to "at least 95%". To support "above 95%", the relevant threshold must be exceeded strictly. These are logical guarantees for exact point probabilities, not confidence guarantees for unseen documents. A finite-sample accuracy estimate and a rounded value are not substitutes for an exact population parameter.

Extension to several required fields

For K mandatory fields with correct-transcription events F_j, let $p_j = P(F_j)$ and $q_K = P(\cap F_j)$. Applying the union bound to the field failures gives:

$$\max(0, \sum_{j=1}^{K} p_j - K + 1) \le q_K$$

$$q_K \le \min_j p_j \qquad (4)$$

A convenient dependence-robust design rule allocates a total error budget across fields:

$$\sum_{j=1}^{K}(1 - p_j) \le 1 - \tau \qquad (5)$$

For three mandatory fields with $p_M = 0.986$ and an additional field accuracy of 0.980, the number field must reach at least 0.984 for this sufficient 95% joint guarantee. This is an illustrative scenario, not a measured name- or address-field accuracy. With four equally accurate required fields, assigning at most 1.25% error to each satisfies the same total budget. The rule prevents a two-field result from being silently extended to an entire document.

The rule is conservative when failures overlap strongly. Its value is that it needs no favourable dependence assumption. Where paired data establish a less conservative joint model, direct joint evaluation may justify different field targets. The choice should be made explicitly rather than by treating the minimum field accuracy as the achieved document accuracy.

RECOVERY AND REGRESSION ANALYSIS

An upgrade confined to one field

Consider an upgrade to number processing that leaves the MRZ correctness of every image unchanged. This per-image invariance is stronger than merely retaining the same aggregate MRZ accuracy. Let q_0 be baseline joint correctness. The probability mass eligible to improve joint success through a number-only repair is:

$$b = p_M - q_0 \qquad (6)$$

Define r as the fraction of this eligible group whose number becomes correct. Define λ as the fraction of initially jointly correct cases whose number becomes incorrect after the upgrade. These are conditional, retrospective evaluation rates, not quantities observable without ground truth during deployment. Partitioning the MRZ-correct cases gives:

$$q_{\text{new}} = (1 - \lambda)q_0 + rb \quad (7)$$

The expression does not assume that an improved crop is always beneficial. The same change can reveal a previously missing digit in one image and introduce neighbouring text in another. Repairs in the MRZ-incorrect group do not increase q while MRZ correctness remains fixed, even though they may increase marginal number accuracy.

At the baseline, $b = 0.175$. With no regressions and complete recovery of this group, the largest joint success allowed by this number-only model is 98.6%. This is a theoretical ceiling. It is not a prediction that every eligible error is recoverable by a particular detector or training procedure.

Required repairs and tolerated regressions

For $b > 0$, the minimum recovery fraction required by a target follows directly from (7):

$$r_{\text{min}} = \frac{\tau - (1 - \lambda)q_0}{b} \quad (8)$$

A value above one makes the target infeasible for the stated regression rate and frozen MRZ outcomes. A non-positive value means the target is already met without recovery. When $b = 0$, there is no MRZ-correct, number-incorrect group to repair, so division by b is not permitted. Table 3 evaluates the feasible baseline scenarios.

Table 3. Conditional recovery required for $q_{new} \geq 95\%$, using $p_M = 98.6\%$ and $q_0 = 81.1\%$.

Regression λ, %	Required r, %	Maximum q_new, %
0	79.43	98.60
1	84.06	97.79
2	88.70	96.98
4	97.97	95.36
5	102.60*	94.55

*Infeasible: recovery cannot exceed 100%. Maximum values use r = 1.

Without regressions, 79.43% of the eligible 17.50% group must be repaired; this is not 79.43% of all documents. At a 2% regression rate among the originally joint-correct cases, the required recovery rises to 88.70%. Conditional rates must therefore always be accompanied by their denominators. Reporting a count of repaired errors alone conceals the effect on retained correct records.

Even complete recovery is insufficient when regressions are too frequent. Provided $q_0 > 0$, the largest feasible regression rate is:

$$\lambda_{\text{max}} = \frac{p_M - \tau}{q_0} \quad (9)$$

For the stated baseline and target, the limit is approximately 4.44%. A 5% regression rate restricts the maximum joint result to 94.545%, despite perfect recovery of the eligible failures. This exposes a useful acceptance criterion for a localisation upgrade: preservation of successful cases is part of the target, not an optional secondary check.

If a shared preprocessing change alters MRZ outcomes, (7) is no longer the correct reduced model. The general identity is new joint accuracy equals baseline joint accuracy plus the probability of joint failure becoming success, minus the probability of joint success becoming failure. All relevant transitions then need to be measured on paired images, including changes in both fields.

FIELD-AWARE UPGRADE AND VALIDATION

Localisation and recognition as separate tests

The source associates some number errors with incomplete or inaccurate localisation. A testable upgrade sequence therefore begins with field identification and controlled cropping, followed by recognition and format checking. MRZ and document-number regions need different rules. ICAO Doc 9303 supplies the relevant MRZ conventions [2], but a compliant structure is a consistency check, not proof that the transcribed text is correct.

DB provides a separately trainable localisation component [3]. A crop-margin change should be evaluated through complete-character retention and final field matching, not only region overlap. A larger margin can retain a missing character while also introducing neighbouring text; its benefit cannot be assumed. Manually annotated crops can help isolate the recogniser's residual errors, but represent supplied localisation rather than an automatic operating mode.

Recognition alternatives must be compared using compatible alphabets, crop sizes and evaluation contracts. PP-OCRv3 documents an SVTR-based recogniser [4], and CPPD denotes Context Perception Parallel Decoder for Scene Text Recognition [5]. Both are candidates for controlled experiments, not upgrades tested here. The PaddleOCR 2.x line-orientation module is a 0°/180° classifier [6]; correcting a whole document rotated by 90° is a different operation.

Paired records and experimental controls

A proposed ablation compares a frozen baseline with annotated-crop recognition, localisation-only changes, recogniser-only changes and their combination. Every configuration should use the same held-out images, reference strings and normalisation policy. A versioned manifest should preserve weights, dictionaries, thresholds, preprocessing, random seeds where relevant and the exact set of evaluated images.

Training and validation must be separated from the final test. Repeated captures of the same document or holder should be grouped before splitting. The paired output log should retain the two field indicators before and after the change, crop coordinates, raw text, normalised text and timing. This permits direct measurement of r, λ and joint transitions instead of reconstructing them from rounded marginal percentages.

Error strata should be defined before inspecting upgrade outcomes: document type, field, image quality and crop completeness are possible categories. The test should retain empty outputs, missed fields and timeouts. Inference should use the independent grouping unit actually supported by the data, rather than assuming that every repeated photograph is an independent trial. No such statistical result is claimed from the present aggregate inputs.

Automatic processing is a separate policy

Selective classification distinguishes predictive quality from the fraction of inputs accepted automatically [7]. Let A be the event that a record is accepted without review and let $C = M \cap D$ denote joint correctness, with E its complement. Coverage a and, for $a > 0$, accepted-record error ε_{A} are:

$$a = \mathrm{P}(A), \quad \varepsilon_{\mathrm{A}} = \frac{\mathrm{P}(E \cap A)}{\mathrm{P}(A)} \tag{10}$$

The fraction of all inputs that are both automatically accepted and correct is then $a(1 - \varepsilon_{\mathrm{A}})$, which cannot exceed q. A system may achieve very low accepted-record error by reviewing many inputs, without improving full-population OCR correctness. Therefore $q = 95\%$ alone does not establish that only 5% of documents require review; that conclusion would require the policy to identify every error correctly.

Confidence scores also require validation against the relevant correctness event. Neural-network calibration research shows why confidence and empirical correctness need not coincide [8]. A recognition score for one crop is not automatically a calibrated probability that both required fields are correct. The acceptance threshold should be selected on validation data, then coverage and

accepted-record errors should be reported on the held-out test with uncertainty.

From point targets to statistical evidence

The deterministic target can guide a future acceptance test, but it must not be applied to uncertain estimates as though they were population parameters. Suppose valid simultaneous lower confidence limits L_M and L_D are obtained for the two marginal accuracies. Then $\max(0, L_M + L_D - 1)$ is a conservative lower limit for joint correctness on the event that both marginal limits hold. Marginal limits constructed separately need simultaneous coverage; their individual confidence levels cannot simply be inherited by the joint claim.

Direct estimation of joint correctness from paired outcomes may be less conservative and should be reported alongside the marginal analysis. The inference method must respect repeated images and any document-level grouping. A planned acceptance criterion should specify the target, confidence level, evaluation unit and treatment of missing outputs before testing. The rounded source percentages do not support those confidence calculations; no such limits are estimated here.

CONCLUSION

Joint OCR success is an intersection probability, not the smaller marginal accuracy by definition. At fixed MRZ accuracy of 98.6%, a number-field accuracy of 95% allows joint success between 93.6% and 95.0%. A number target of 96.4% is sufficient for at least 95% joint success without assuming favourable error dependence in the point-probability model.

The recovery–regression analysis adds an operational requirement: a useful field-specific upgrade must repair eligible failures while preserving correct records. At the stated baseline, the 95% target requires 79.43% eligible recovery without regressions, increasing to 88.70% at 2% regression. Multi-field error budgets and selective-processing metrics prevent these results from being overstated as whole-document or automatic-acceptance guarantees. The next step is the proposed paired component evaluation, not the presentation of these analytical thresholds as measured OCR gains.

Data and preparation statement

All derivations, scenarios and conclusions were developed by the author. ChatGPT (OpenAI) was used only for structuring, language editing and reformulating the presentation of derivations the author had already completed; it did not originate any equation, scenario or numerical result. After editing, the author re-derived every equation from the four-cell outcome partition, and checked the standard intersection bounds against [Fréchet 1951 / Nelsen 2006]. An author-written Python script, provided as supplementary material, recomputes all values in Tables 1–3, Fig. 1 and the text, including the infeasible and boundary cases ($r > 1$, $b = 0$, λ_max). The author takes full responsibility for the content.

REFERENCES

[1] Kravets A.G., Semenochkin D.O., Markov A.K. Development of a New Experimental Method for Evaluating OCR Tools for the Task of Digital Document Classification. Proceedings of Voronezh State University. Series: Systems Analysis and Information Technologies 2024;(3):114–126. doi:10.17308/sait/1995-5499/2024/3/114-126.

[2] ICAO. Doc 9303: Machine Readable Travel Documents. Parts 3–5, 8th ed., 2021. https://www.icao.int/publications/doc-series/doc-9303 (accessed 4 October 2026).

[3] Liao M., Wan Z., Yao C., Chen K., Bai X. Real-Time Scene Text Detection with Differentiable Binarization. Proceedings of the AAAI Conference on Artificial Intelligence 2020;34(07):11474–11481. doi:10.1609/aaai.v34i07.6812.

[4] Li C. et al. PP-OCRv3: More Attempts for the Improvement of Ultra Lightweight OCR System. arXiv:2206.03001, 2022. doi:10.48550/arXiv.2206.03001.

[5] PaddleOCR. CPPD: Context Perception Parallel Decoder for Scene Text Recognition. Version 2.10.0 documentation. https://www.paddleocr.ai/v2.10.0/en/algorithm/text_recognition/algorithm_rec_cppd.html (accessed 4 October 2026).

[6] PaddleOCR. Text Direction Classification. Version 2.10.0 documentation. https://www.paddleocr.ai/v2.10.0/en/ppocr/model_train/angle_class.html (accessed 4 October 2026).

[7] Geifman Y., El-Yaniv R. Selective Classification for Deep Neural Networks. Advances in Neural Information Processing Systems 2017;30. arXiv:1705.08500.

[8] Guo C., Pleiss G., Sun Y., Weinberger K.Q. On Calibration of Modern Neural Networks. Proceedings of ICML, PMLR 2017;70:1321–1330.