



International Journal of Artificial Intelligence and Machine Learning

Publisher's Home Page: <https://www.svedbergopen.com/>



Research Paper

Open Access

Beyond OCR: Quantifying The Field-Association Gap in Automated Electrical Nameplate Extraction

Amitkumar Subhashchandra Vishwakarma¹, Dr. Ganesh Mhatre², Samadhan Patil³, Dhananjay Labde⁴

¹ Students of Masters of Power System, Yadavrao Tasgaonkar Institute of Engineering & Technology, Karjat

² Assistant Professor at Department of Mechatronics Engineering, New Horizon Institute of Technology & Management

³ Head of Department at Department of Electrical Engineering, Yadavrao Tasgaonkar Institute of Engineering & Technology, Karjat

⁴ Assistant Professor at Department of Electrical Engineering, Yadavrao Tasgaonkar Institute of Engineering & Technology, Karjat

Corresponding Author(s) Amitkumar Subhashchandra Vishwakarma¹ Email Id : akvish.4198@gmail.com

Abstract: This paper evaluates, on a public 286-image real-world dataset, the same extraction architecture implemented in the authors' nameplate web application: local optical character recognition followed by deterministic rule-based field association and engineering validation, with no call to a hosted API. The application originally used a hosted multimodal model (Gemini) for extraction; that call has since been replaced, in the application's own source, by an in-browser OCR-plus-rules pipeline, so the design evaluated here is the design the application now runs. Two OCR engines are scored with an engineering-aware protocol, field exact match, all-critical-fields-correct rate, hallucination and order-of-magnitude error rates, and validation-rule flag recall and precision, adapted from a companion review of deep-learning nameplate extraction. Reference values for the dataset were produced by a large vision-language model (Qwen3.6-35B) rather than by human annotators, so every number below is a measurement of agreement with that machine-generated reference, not of ground-truth accuracy, and is worded that way throughout rather than only in a caveat. RapidOCR-plus-rules achieves 20.5% field-level agreement with the reference labels and 67.0% precision (bootstrap 95% CI 17.5-23.9%), against 8.8% and 50.3% for Tesseract-plus-rules; both are far below the all-critical-fields-correct rate a deployment would require (1.0% and 0.3% respectively). A diagnostic breakdown splits the loss between reading and association: 52.5% of reference critical-field values can be found in the raw RapidOCR transcript, but only 20.5% are correctly assigned to their field by the deterministic parser, so 61% of the values the engine does read are lost at the association step. The 51 reference records that fail the engineering-validation rules (17.8% of the dataset) were adjudicated against their OCR transcripts: 76% are consistent with a genuine reference-label error, 10% are false positives on machines the validator does not model (synchronous, servo, geared and DC motors), and 14% could not be resolved from text alone. Every point at which this evaluation fell short of its original protocol is disclosed explicitly, most importantly that GPU-dependent recognizers and a human-verified gold-label subset could not be produced in the environment used for this analysis, and that no public transformer-nameplate dataset with field labels exists, so a motor-nameplate dataset was used to validate the pipeline design rather than the application's exact field list.

Keywords: Nameplate information extraction, optical character recognition, rule-based field extraction, engineering-aware evaluation, on-device inference, no external API.

I. INTRODUCTION

Reading a nameplate correctly is a precondition for almost every downstream use of the data it carries [1]-[3]: asset registers, maintenance scheduling, procurement and compliance checks all depend on values such as rated power, voltage, current and insulation class being transcribed exactly, not merely close enough for a human to recognize. A companion review of deep-learning approaches to nameplate information extraction argued that conventional OCR benchmarks [4], character error rate [5] and word accuracy [6], do not capture this requirement, because a single-digit or unit error (11 kV read as 1.1 kV, or Dyn11 read as Dyn1 [7]) can be catastrophic for an engineering use case while leaving the character error rate almost unchanged. That review proposed an engineering-aware evaluation protocol built around field-level exact match, an all-critical-fields-correct (ACFC) rate, hallucination and order-of-magnitude error rates, and a validation layer that checks extracted values against basic physical constraints.

This paper applies that protocol to the pipeline implemented in the authors' own nameplate web application, run end to end on a real-world image set. No result in this paper was obtained from synthetic

or hand-constructed test images: every number was computed offline from the archived output files of one complete run of the evaluation notebook over the full dataset.

Two design decisions shape everything that follows and are stated up front. First, the application, and every system evaluated in this paper, runs with no external API call as the application originally sent each image to a hosted multimodal model, and that call has been replaced in the application's source by an in-browser pipeline, described in Section II, so the results in this paper describe the design the application now runs, not the earlier one. Second, the dataset's reference values were generated by a large vision-language model (Qwen3.6-35B) [8], not transcribed by a person, so every comparison to them is reported as agreement with an LLM-generated reference, never as accuracy [9]; Section VI returns to the practical consequences of this.

II. THE PIPELINE UNDER TEST

A. Application architecture

The nameplate application follows the same four-step design before and after the change described below: a photograph is passed through an extraction step that returns a value and a confidence for each of fifteen fields (extraction.js), the returned values are normalized to engineering quantities (validation.js, ported from nameplate/normalize.py), the normalized record is checked against a set of physical-consistency rules (validation.js, ported from nameplate/validate.py), and each field is shown to the user marked Accepted or Needs review with the specific reason, alongside a JSON export in the format the evaluation toolkit reads. What changed is only the extraction step. The application originally sent the image to a hosted multimodal model (Gemini [10]) with a fixed JSON schema and a temperature-zero prompt. That call, and the API key it required, have been removed from the application's source; extraction now runs Tesseract.js [11], an in-browser WebAssembly [12] build of the Tesseract OCR engine [13], followed by a deterministic rule-based field-association module (ocrRules.js) ported line-for-line from the same rule logic used in the evaluation toolkit's Python parser (nameplate/rules.py), so that the application and the offline experiments run the same association logic. No image, and no extracted text, leaves the browser at any point.

This is the OCR-plus-rules design evaluated throughout this paper. RapidOCR-plus-rules and Tesseract-plus-rules, in Sections IV and V, follow the same architecture the application runs with OCR text lines, then label-keyword and value-pattern matching over neighboring lines, then normalization and engineering validation (Fig. 1). They differ from the application in three respects. The OCR engine differs as RapidOCR [14], [15] is a stronger, deep-learning-based engine not yet available as a WebAssembly build suitable for in-browser use, while Tesseract is the engine the application runs in the browser. The field set differs: the motor dataset's fourteen fields replace the application's fifteen transformer fields, so the evaluated parser applies the same label-proximity rules to motor labels such as HP, kW, RPM, VOLTS and AMPS. And the evaluated parser looks for a value only to the right of its label on the same row, whereas ocrRules.js also checks the line directly below a label for several fields. The metrics below therefore measure the application's association design on real photographs, not a line-for-line copy of its transformer-specific code.

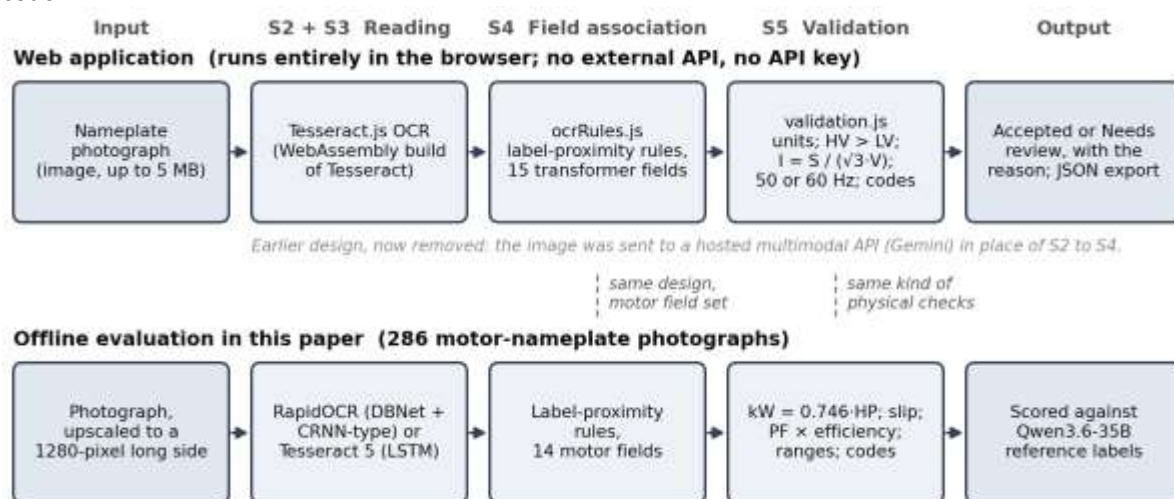


Fig. 1. The pipeline under test: the web application after the change described in Section II-A (top) and the offline evaluation reported in this paper (bottom). S2 to S5 are the pipeline stages of the companion review.

B. Why the hosted call was removed

The extraction step was changed for a project requirement stated independently of this evaluation: no part of the pipeline may depend on an external, hosted API. Removing it has three further consequences relevant to this paper. It removes the per-request cost and rate limits a hosted call would impose during a large evaluation run. It removes the reproducibility risk of an externally served model changing behavior between runs [16], which the application's earlier design had already flagged by pinning a model identifier. Most importantly for what follows, it means the application's extraction step now follows the same deterministic design evaluated in Sections IV and V, so the measurements reported here bear directly on the deployed pipeline rather than on a different kind of system.

C. Dataset

No public dataset of power-transformer nameplate photographs with field-level labels could be located [1], a gap already noted in the companion review. The empirical run in this paper therefore uses sergiudanstano/motor-nameplate on Hugging Face [8], 286 web-sourced photographs of electric-motor nameplates. This substitution is explicit and load-bearing for how the results should be read: it validates the pipeline design, extraction followed by normalization, association and engineering validation, and it validates the OCR-plus-rules association logic that the application runs client-side, but it does not validate the application's fifteen transformer-specific fields (rated power, HV/LV voltage and current, vector group, cooling code and so on), which have not yet been run against labeled transformer photographs. Fourteen fields are scored for the motor dataset: rated horsepower, rated kW, rated speed, rated voltage, rated current, frequency, phase count, service factor, insulation class, frame, enclosure, duty cycle, serial number and model or catalogue number [17], [18]; the first nine are treated as critical fields for the ACFC and validation-flag metrics, following the definition used in the companion review. Reference values for all 286 images were produced by Qwen3.6-35B [8] reading each image directly, with no human review; this is disclosed by the dataset and is the reason every metric in Section IV is reported as agreement rather than accuracy.

III. EVALUATION METHOD

Field exact match (EM) is the fraction of critical fields, printed on the plate according to the reference and successfully normalized, for which the system's normalized value equals the reference's; precision is the same numerator over the fields the system actually returned a value for. ACFC is the fraction of plates on which every critical field is simultaneously correct and no field is hallucinated (a value returned for a field the reference says is blank), the threshold a record would need to clear to enter an asset register unedited. Hallucination rate and order-of-magnitude (OOM, a roughly 10 times or greater numeric error typical of a misread decimal point or unit) follow the companion review's definitions. Field CER is the character error rate on the raw returned string.

The engineering-validation layer applies four checks, the motor-domain equivalents of the transformer checks the application's own validation.js runs (unit parsing, HV greater than LV, and rated current equal to S divided by root3 times V [7]): kW equal to 0.746 times HP within a 0.72-0.77 ratio band when both are present; slip, rated speed must sit at or below a synchronous speed of 120f/p for a plausible pole count p [19], within a tolerance widened for sub-1 kW motors; implied power factor times efficiency, equal to P divided by root3 times V times I, required to fall in 0.25-0.99 for at least one plausible voltage-current pairing on multi-voltage plates; and range and grammar checks (50 or 60 Hz, 1 or 3 phases, service factor 1.0-1.5 [17], insulation class in the set A, B, E, F, H [20]). Flag recall and precision measure how often these checks correctly catch a wrong field, using the reference labels as ground truth for whether a field is wrong. Accepted coverage and accuracy report the share of fields that pass validation and their accuracy among the passed set. ECE (expected calibration error) [21] uses each engine's own line-confidence scores. All confidence intervals are bootstrap 95% intervals over images, 1000 resamples, fixed seed [22].

Because the reference labels are themselves machine-generated, Section V goes one step further and adjudicates the subset of reference records that the engineering-validation layer itself flags as physically inconsistent. Adjudication in this paper was performed by inspecting the full RapidOCR and Tesseract transcripts of the corresponding image alongside the flagged reference values, not by viewing the image, which the analysis environment could not obtain, as explained in Section VI. Every adjudication verdict below is disclosed as text-based, and cases the OCR transcript does not resolve are reported as ambiguous rather than guessed.

IV. RESULTS

A. Overall agreement with the reference labels

Table I and Table II report the full metric set for both OCR-plus-rules configurations over all 286 images.

TABLE I FIELD-LEVEL AGREEMENT WITH THE QWEN3.6-35B REFERENCE LABELS (NOT GROUND-TRUTH ACCURACY)

System	n	Field EM %	95% CI	Precision %	ACFC %	95% CI
RapidOCR+rules	286	20.5	17.5-23.9	67.0	1.0	0.0-2.4
Tesseract+rules	286	8.8	6.8-11.1	50.3	0.3	0.0-1.0

TABLE II ERROR, VALIDATION-FLAG, CALIBRATION AND LATENCY METRICS (SAME RUNS AS TABLE I)

System	Halluc. %	OOM %	CER	Flag recall %	Flag prec. %	Cov. %	Acc.@cov %	ECE	Lat. mean/p95 s
RapidOCR+rules	1.3	9.1	0.776	37.7	63.9	80.5	74.5	0.086	2.6 / 5.4
Tesseract+rules	2.5	18.2	0.901	29.7	77.0	80.8	56.8	0.202	1.6 / 2.9

RapidOCR-plus-rules more than doubles Tesseract-plus-rules' field-level agreement with the reference labels (20.5% against 8.8%) and roughly halves its OOM error rate, consistent with RapidOCR's deep-learning detector and recognizer [23], [24] handling the oblique angles, glare and low contrast common in web photographs [25]-[27] better than Tesseract's classical pipeline [13]. Neither configuration comes close to being usable unedited: an ACFC of 1.0% means that on 283 of 286 plates at least one critical field is wrong or missing after validation. Accepted-coverage accuracy, 74.5% for RapidOCR-plus-rules, shows that the validation layer, when it does accept a field, is a meaningfully better predictor of correctness than raw output, 67.0% precision; but the checks accept 80.5% of the returned fields, and about one accepted field in four still disagrees with the reference, an error the checks cannot see. Most disagreements are omissions rather than misreadings: RapidOCR-plus-rules returns no value at all for 70.0% of the reference critical values, a value with wrong characters for 7.0% and an order-of-magnitude error for 2.5%. Fig. 2 shows these comparisons graphically; Section V examines where the omissions arise.

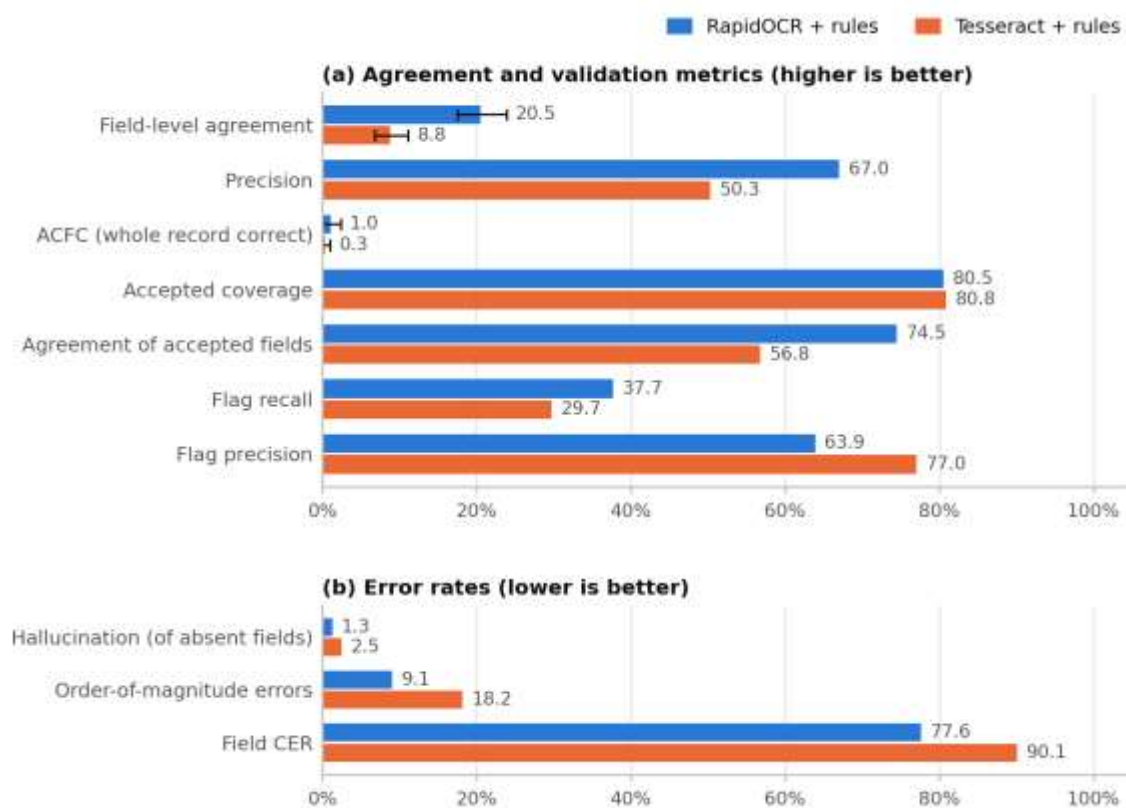


Fig. 2. Agreement, validation and error metrics of both OCR-plus-rules configurations against the Qwen3.6-35B reference labels (the values of Tables I and II). Whiskers show bootstrap 95% confidence intervals.

B. Per-field breakdown

Table III and Fig. 3 break field-level agreement down by field. The pattern is consistent across both engines: short, low-variability tokens (duty cycle, enclosure code) are recovered far more reliably than numeric ratings that require correct label-to-value association (voltage, current, service factor), the same weakness the application's field-association module inherits by design, since it is a label-proximity parser rather than a learned mapper [28]-[30].

TABLE III FIELD-LEVEL EXACT MATCH, RANKED BY RAPIDOCR+RULES PERFORMANCE

Field	n on plate	RapidOCR+rules %	Tesseract+rules %
duty	173	47.4	24.9
enclosure	171	35.7	15.8
rpm	227	31.3	9.3
hz	230	28.3	9.6
hp	158	27.2	12.7
frame	146	23.3	6.8
volts	236	19.1	6.8
kw	151	18.5	9.9
amps	226	17.7	7.5
service_factor	127	14.2	5.5
phase	242	14.0	10.7
model/cat no.	236	13.6	1.3
insul_class	220	12.7	7.3
serial_no	164	9.8	3.7

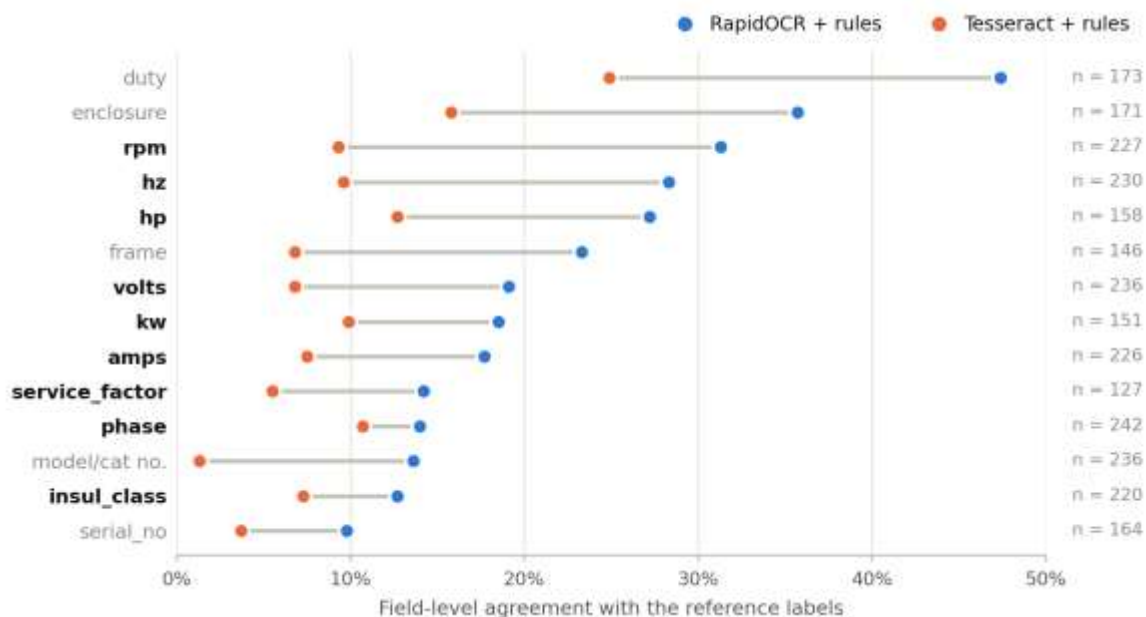


Fig. 3. Field-level agreement with the reference labels for each of the fourteen fields (the values of Table III). Critical fields are shown in bold; n is the number of plates on which the reference lists the field.

V. WHERE THE ERRORS COME FROM

A. Recognition versus association

A field that is wrong in the final output can fail at two different pipeline stages: the OCR engine may never recognize the text at all, a detection or recognition failure, stages S2-S3 of the taxonomy used in the companion review, or the text may be recognized correctly but assigned to the wrong field or dropped by the rule parser, an association failure, stage S4, which in the application is ocrRules.js. Distinguishing the two matters because they call for entirely different fixes: better image preprocessing and a stronger recognizer for the first, a learned or more context-aware field mapper for the second.

For every critical field with a parseable reference value, it was checked whether that value can be found anywhere in the raw RapidOCR transcript, independently of whether the rule parser assigned it correctly. A numeric value counts as found only when every number in it appears as a number token in the transcript (all of 208, 230 and 460 for a reference of 208-230/460), and an insulation-class letter only when it follows a class or insulation keyword, since a bare letter such as F occurs in almost any transcript. Of 1,814 reference critical-field values checked, 953 (52.5%) are found in the raw transcript, but only 372 (20.5%) are correctly extracted by the rule parser: 584 values, 32.2% of all and 61% of those found, are read but lost at the association step, while 858 (47.3%) are neither found nor extracted, either because the engine did not read them or because the reference value itself is wrong. Table IV and Fig. 4 give the field-level split. The association loss is largest for phase (65.3% found against 14.0% extracted) and frequency (69.1% against 28.3%), although for short tokens such as a phase count, presence anywhere in the transcript is a lenient test; it remains large for rated voltage (57.2% against 19.1%), where every rated value must be present. Insulation class is the opposite case: its letter follows a class keyword in only 17.5% of transcripts, so for that field most of the loss happens before association. For Tesseract, only 29.8% of the reference values are found in the transcript, which accounts for most of its lower agreement.

TABLE IV TEXT IS READ MORE OFTEN THAN IT IS CORRECTLY ASSIGNED TO A FIELD (RAPIDOCR, CRITICAL FIELDS)

Field	n	Present in raw OCR text %	Correctly extracted %
hz	230	69.1	28.3
phase	242	65.3	14.0
rpm	227	62.6	31.3
hp	158	60.1	27.2
volts	236	57.2	19.1
kw	151	51.0	18.5
amps	226	44.2	17.7
service_factor	127	38.6	14.2
insul_class	217	17.5	12.9

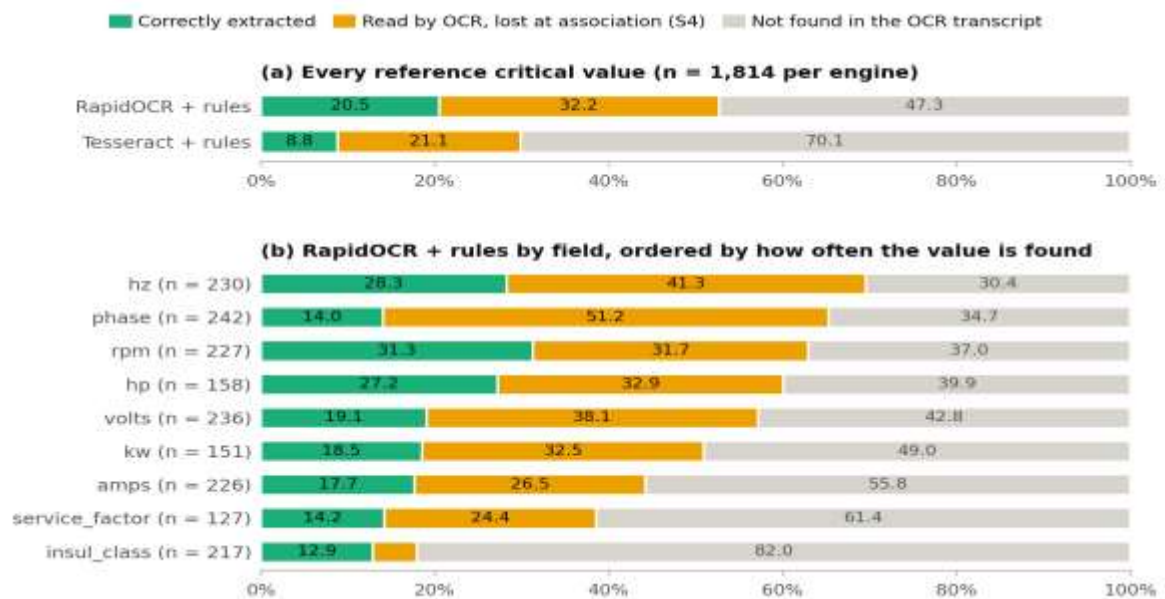


Fig. 4. Where reference critical values are lost. (a) Share of all reference critical values that are correctly extracted, found in the raw OCR transcript but lost at the association step, or not found in the transcript. (b) The same split for each critical field, RapidOCR plus rules (the values of Table IV).

This is the single clearest empirical finding of this evaluation: on this dataset, the observed upper bound from raw RapidOCR transcript availability, that is, the share of reference values findable somewhere in the transcript regardless of correct field assignment, is 52.5%, while the deterministic rule parser realizes only about two fifths of that figure, 20.5%. This is not a theoretical ceiling on OCR performance, only what was observed to already be present in this particular engine's output; a different OCR engine could recover a different share. What it does show directly is that most of the text the engine reads, 61% of it, is not attached

to the right field, suggesting that field association is a major bottleneck for this pipeline and a promising target for further investigation, alongside, not necessarily ahead of, further recognition improvements. Fig. 5 shows the mechanism on a single plate, on which the OCR engine reads every rated value and the parser still recovers only three of them. Because the application's ocrRules.js uses the same label-proximity design, and additionally checks the line directly below a label, the finding is expected to carry over to it, but this has not yet been measured on transformer plates. Confirming which improvement actually yields the larger gain would require running the S4 ablation itself, a learned or LLM-based field mapper over the same OCR text, which Section VI-B explains could not be executed in this environment. Fig. 6 summarizes the failure categories met in this analysis, each with a real example.

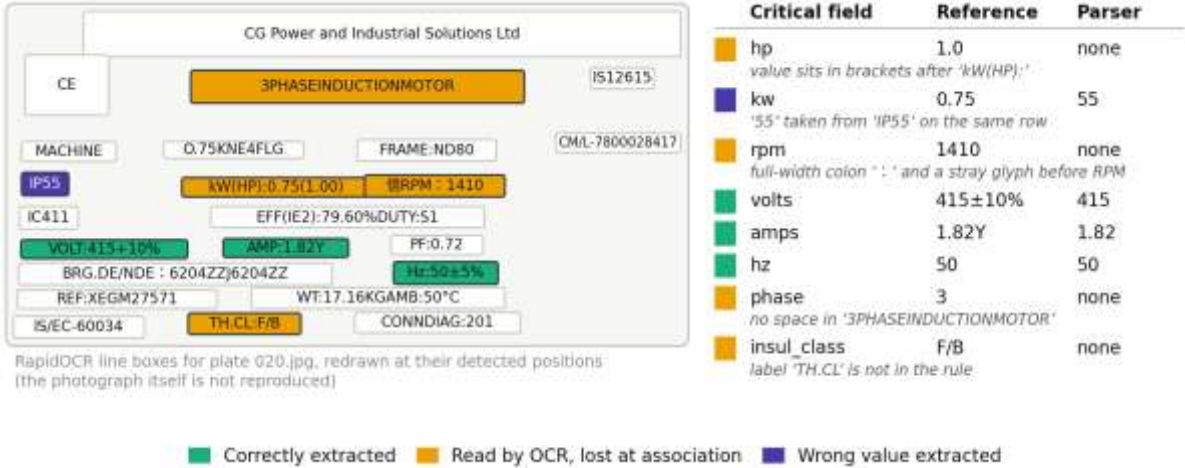


Fig. 5. Worked example, plate 020.jpg: the OCR engine reads every rated value, but the label-proximity rules recover only voltage, current and frequency. The reason for each loss is given beside the field.

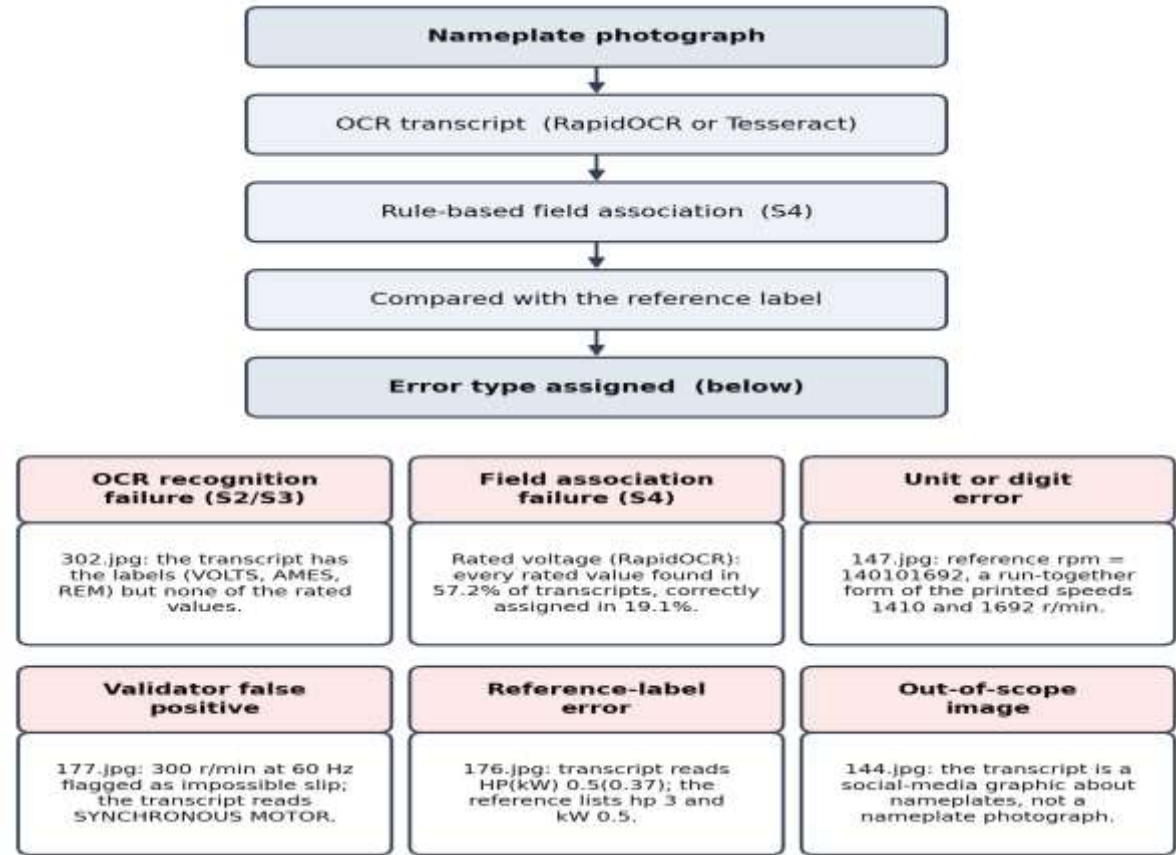


Fig. 6. Six failure categories observed during this analysis, each with a real example from the run: a recognition failure, an association failure, a digit error, a validator false positive, a reference-label error and an out-of-scope image.

B. Adjudicating the engineering-validation flags

Fifty-one of the 286 reference records (17.8%) fail at least one of the four engineering-validation checks in Section III, that is, the reference labels themselves, not the system output, are internally inconsistent according to basic physics. Each of the 51 was adjudicated by reading the full RapidOCR and Tesseract transcripts of the image against the flagged values (image inspection was not possible in this environment, see Section VI) and classified into one of three outcomes, recorded with a one-line reason per record in the adjudication file:

- Reference-label error (39 of 51, 76%).** The transcript contradicts the flagged value or shows it to be an artifact, or the record breaks a hard physical limit for the machine type the transcript identifies. Examples include an image whose transcript reads "HP(kW)0.5(0.37)" against a reference of hp 3 and kW 0.5, where the reference appears to have shifted the printed 0.5 into the wrong field and invented a new horsepower value; a reference rpm of 140101692 against a transcript showing the two printed speeds, 1410 and 1692 r/min; an amps value of 6600, identical to the plate's printed voltage of 6600 V; a reference of 90 kW on a plate whose rating table reads 30 kW in every row; values combined from different rating rows, such as the horsepower of a 60 Hz NEMA row with the kilowatts of an IEC row; and malformed entries, such as a voltage field holding "240V/50HZ". Records whose implied power factor times efficiency exceeds 1, or whose rated speed exceeds synchronous speed, on plates the transcript identifies as ordinary AC motors, are counted here as well, since no correct plate can produce them.
- Validator false positive (5 of 51, 10%).** The flagged values agree with the transcript, but the plate belongs to a machine or rating the checks do not model: a synchronous motor, whose transcript explicitly reads "SYNCHRONOUS MOTOR" and whose 300 r/min at 60 Hz is exactly synchronous for a 24-pole machine, which the slip check rejects because it expects an induction motor's slip; a synchronous servomotor whose printed speed is a maximum (nmax 2000 r/min); a two-speed geared motor whose printed speeds, 43 and 176 r/min, are gearbox output speeds; a small 25.2 V DC motor; and a legitimate service factor of 1.60 on a small single-phase motor.
- Ambiguous (7 of 51, 14%).** The transcript is too fragmentary to decide, because of heavy glare, an extreme oblique angle or near-total OCR failure, or it agrees with a record that is still physically implausible, so that a misprint on the plate cannot be told apart from a misreading shared by the OCR engine and the labeling model.

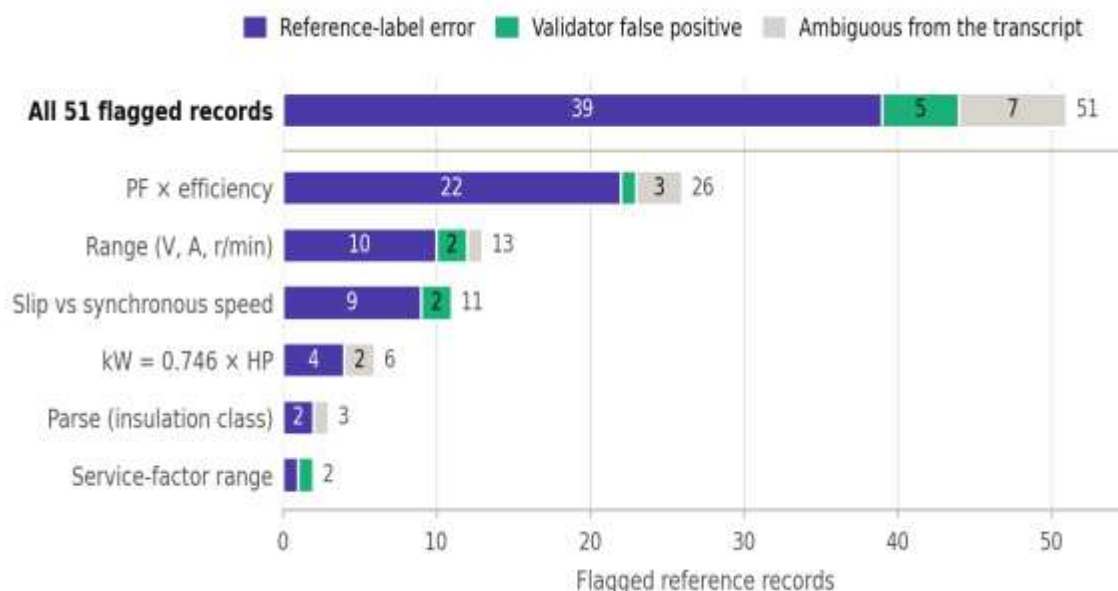


Fig. 7. Adjudication outcomes for the 51 flagged reference records, overall and by the check that raised the flag; a record that fails several checks is counted once under each.

Two practical conclusions follow. First, three quarters of the flags correspond to genuine defects in the machine-generated reference labels, which is itself evidence that any accuracy number computed against this reference, uncorrected, understates true system performance to some degree: a system that reads a plate correctly will be marked wrong whenever the reference itself is wrong. Second, the validator false positives, 5 of 51, all come from machines or ratings the checks do not model: none of the checks currently conditions on machine type, so they misfire predictably on synchronous and servo motors, geared outputs,

small DC motors and unusual but legitimate ratings. Fig. 7 breaks the outcomes down by the check that raised the flag: the power-factor check raises half of all flags, 26 of 51, and 22 of them are genuine label errors, while the false positives come from the slip, range and service-factor checks, and once from the power-factor check. A revised rule set that first classifies device type, or simply relaxes the affected checks for synchronous, servo, geared and DC machines, should remove these false positives without materially changing flag recall on ordinary induction motors, the large majority of the dataset.

As a by-product of this adjudication, the dataset itself contains a handful of images arguably out of scope for a motor-nameplate label: three transcripts read as informational graphics about how to read a nameplate rather than as nameplates, one image is a frequency-inverter rating label, one a hydraulic motor and one a blank template with placeholder text. At least two plates also appear twice: 056.jpg and 291.jpg show the same plate but carry mutually inconsistent reference labels, and 220.jpg and 253.jpg show another. At under 3% of the dataset these do not change the headline numbers, but they are worth excluding before any hand-labeling effort, and are reported here so that a future annotator does not have to rediscover them.

VI. DISCUSSION AND LIMITATIONS

A. What these numbers do and do not show

Every percentage in Sections IV and V is agreement with, or an adjudication against, an LLM-generated reference, never a measurement of ground-truth accuracy. This distinction is not a formality: Section V-B shows that at least 76% of the specific records where the reference itself is flagged as inconsistent are indeed wrong, so the reference is demonstrably imperfect, and a system that happens to make correlated errors with Qwen3.6-35B, for instance by reading the same ambiguous glyph the same way, would be scored as more accurate than it is, while a system that reads a genuinely mislabeled plate correctly would be scored as less accurate than it is. Absent human-verified gold labels, the field EM and ACFC numbers in Table I should be read as an upper-bound-adjacent agreement figure, not as the accuracy figure a deployment decision should be based on.

B. Why no learned field-association or vision-language result is reported

The companion review's central S4 ablation question, whether a learned or LLM-based field mapper closes the association gap identified in Section V-A, requires either a locally-run vision-language model, such as Qwen2.5-VL [31], or a transformer-based OCR recognizer, such as TrOCR [32], both of which need a GPU to run in reasonable time and, in the Colab notebook built for this study (Section VI-D), a download of model weights and the dataset itself. The most recent run available going into this paper (run_info.json, device "cpu", gpu null) confirms that run executed without a GPU runtime, so its TrOCR and Qwen2.5-VL-3B fields carried no new information beyond the OCR-plus-rules baselines already reported here.

C. Why no human-verified gold-label subset is reported

The evaluation protocol calls for a human-labeled sample, 60 to 100 images, as the basis for any accuracy claim, collected through the notebook's hand-labeling form. As of this analysis, that sample had not been produced, and the 51-row adjudication file intended to record a second check on the flagged records had an empty verdict column throughout. Section V-B's adjudication is offered as a partial, disclosed substitute for that step: it establishes that the reference has real errors and estimates their rate on the flagged subset specifically, but it is not a substitute for a full 60 to 100 image hand-labeled sample scored across all fields, which remains necessary before any accuracy number can be reported for this system.

D. Constraints of the analysis environment used for this paper

This paper's results were computed offline, in an analysis environment separate from the Colab runtime that produced the original run. That environment has two CPU cores, no GPU and no network access to huggingface.co, which its egress policy blocks; that host serves both the dataset images and the TrOCR and Qwen2.5-VL-3B model weights. Consequently, within this environment it was not possible to download the dataset's images to run the GPU-dependent recognizers directly, to view the images and hand-label a human gold-label sample, or to perform the flagged-record adjudication in Section V-B by inspecting images rather than OCR transcripts. All of the numbers reported in Sections IV and V are instead computed from artifacts already produced by an earlier, real run of the Colab notebook over the full 286-image dataset, the raw OCR transcripts, the reference labels and the rule-based predictions, none of which required a further download. The headline metrics were recomputed with the notebook's own scoring code and reproduce its summary files exactly. No number in this paper is estimated, simulated or computed on synthetic data; every one is derived from that real run's output files. The Colab notebook itself, current as of that run, accompanies this paper for reference; it does not need to be run again to reproduce the numbers reported here.

No external hosted API of any kind was used to produce any result in this paper, or is used by the application, the evaluation toolkit or the Colab notebook it accompanies. The OCR-plus-rules systems evaluated here, and the application described in Section II, run entirely with local, open-source software, RapidOCR and Tesseract for OCR, and, for the GPU-dependent comparison the Colab notebook was built for but could not be executed here, locally-downloaded open-weight models (TrOCR, Qwen2.5-VL), never a paid or rate-limited service.

VII. CONCLUSION

The application's extraction pipeline design, local OCR followed by deterministic rule-based field association and engineering validation, recovers 20.5% of critical nameplate fields correctly with RapidOCR and 8.8% with Tesseract, when compared against a real-world dataset's LLM-generated reference labels, and produces a fully correct, non-hallucinated record on at most 1% of plates, well short of what unedited use in an asset register would require. The loss is shared between reading and association: 52.5% of reference values can be found in the raw RapidOCR transcript, but only 20.5% are correctly attached to their field, so the parser loses 61% of the values the engine does read. This suggests that field association, the application's ocrRules.js stage, is a major bottleneck and a promising target for further investigation, while the 47.3% of values never found in the transcript show that better reading remains a target of comparable size, although part of that share reflects errors in the reference itself. A text-based adjudication of the 51 reference records that fail basic engineering-consistency checks finds that three quarters, 76%, are genuine reference-label defects, while 10% expose a real, fixable gap in the validation rules themselves, their unstated assumption of a fixed-speed induction motor at line frequency. The GPU-dependent comparison against a learned field-association model, a human-verified accuracy figure, and a run against the application's own transformer-nameplate fields on labeled transformer photographs remain the open items needed to complete this evaluation, and are named here as such rather than filled in with results this analysis was not in a position to produce honestly.

Reproducibility

Dataset: sergiudanstani/motor-nameplate (Hugging Face, MIT licence), 286 images, reference labels by Qwen3.6-35B [8]. OCR engines: RapidOCR 1.4.4 [14] (ONNX Runtime [33], DBNet detector [23], CRNN-type recognizer [24]), Tesseract 5 [13] (pytesseract 0.3.13), both at 1280-pixel long side. Bootstrap confidence intervals: 1000 resamples, seed 42. Source artifacts for every number in this paper (ocr_rapidocr.jsonl, ocr_tesseract.jsonl, pred_qwen_dataset_labels.jsonl, pred_rapidocr_rules.jsonl, pred_tesseract_rules.jsonl, summary_vs_silver.csv, per_field_vs_silver.csv, silver_flagged_for_adjudication.csv, run_info.json) are the outputs of the notebook's own run over the full dataset and are provided alongside this manuscript, together with the adjudication file holding the verdict and reason for each flagged record (silver_flagged_adjudicated.csv) and the Colab notebook itself, for reference.

REFERENCES

- [1] C. Alvarez-Murillo, A. Ospina-Bedoya, J. W. Branch-Bedoya, G. D. Zapata Madrigal, and R. García Sierra, "Computer vision-based information extraction from electrical assets: A systematic literature review," *Sensors*, vol. 26, no. 16, Art. no. 5009, 2026, doi: 10.3390/s26165009.
- [2] M. I. Mossad, M. Azab, and A. Abu-Siada, "A novel evolutionary technique to estimate induction machine parameters from name plate data," in *Proc. 22nd Int. Conf. Electrical Machines (ICEM)*, Lausanne, Switzerland, 2016, pp. 66-71, doi: 10.1109/ICELMACH.2016.7732507.
- [3] Y. Wang, Y. Du, Z. Pang, J. Qi, and H. Ge, "The text detection and recognition method for electrical nameplates based on deep learning," in *Electronic Engineering and Informatics: Proc. 5th Int. Conf. Electronic Engineering and Informatics (EEI 2023)*, 2024, pp. 589-598, doi: 10.3233/ATDE240123.
- [4] D. Karatzas et al., "ICDAR 2015 competition on robust reading," in *Proc. 13th Int. Conf. Document Analysis and Recognition (ICDAR)*, 2015, pp. 1156-1160, doi: 10.1109/ICDAR.2015.7333942.
- [5] V. I. Levenshtein, "Binary codes capable of correcting deletions, insertions, and reversals," *Soviet Physics Doklady*, vol. 10, no. 8, pp. 707-710, 1966.
- [6] J. Baek, G. Kim, J. Lee, S. Park, D. Han, S. Yun, S. J. Oh, and H. Lee, "What is wrong with scene text recognition model comparisons? Dataset and model analysis," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2019, pp. 4715-4723, doi: 10.1109/ICCV.2019.00481.
- [7] Power Transformers - Part 1: General, IEC Standard 60076-1, Ed. 3.0, 2011.
- [8] sergiudanstani, "Motor Nameplate," Hugging Face dataset, 2026. [Online]. Available: <https://huggingface.co/datasets/sergiudanstani/motor-nameplate>

- [9] Y. Liu, Z. Li, M. Huang, B. Yang, W. Yu, C. Li, X.-C. Yin, C.-L. Liu, L. Jin, and X. Bai, "OCRBench: On the hidden mystery of OCR in large multimodal models," *Sci. China Inf. Sci.*, vol. 67, no. 12, Art. no. 220102, 2024, doi: 10.1007/s11432-024-4235-6.
- [10] Gemini Team, Google, "Gemini: A family of highly capable multimodal models," arXiv:2312.11805, 2023.
- [11] Project Naptha, "Tesseract.js: Pure Javascript OCR for more than 100 languages," 2016. [Online]. Available: <https://github.com/naptha/tesseract.js>
- [12] A. Haas, A. Rossberg, D. L. Schuff, B. L. Titzer, M. Holman, D. Gohman, L. Wagner, A. Zakai, and J. F. Bastien, "Bringing the web up to speed with WebAssembly," in *Proc. 38th ACM SIGPLAN Conf. Programming Language Design and Implementation (PLDI)*, Barcelona, Spain, 2017, pp. 185-200, doi: 10.1145/3062341.3062363.
- [13] R. Smith, "An Overview of the Tesseract OCR Engine," in *Proc. ICDAR*, 2007, pp. 629-633.
- [14] RapidAI Team, "Rapid OCR: OCR Toolbox," 2021. [Online]. Available: <https://github.com/RapidAI/RapidOCR>
- [15] Y. Du, C. Li, R. Guo, X. Yin, W. Liu, J. Zhou, Y. Bai, Z. Yu, Y. Yang, Q. Dang, and H. Wang, "PP-OCR: A practical ultra lightweight OCR system," arXiv:2009.09941, 2020.
- [16] L. Chen, M. Zaharia, and J. Zou, "How is ChatGPT's behavior changing over time?," *Harvard Data Science Review*, vol. 6, no. 2, 2024, doi: 10.1162/99608f92.5317da47.
- [17] Motors and Generators, ANSI/NEMA Standard MG 1-2021, 2021.
- [18] Rotating Electrical Machines - Part 1: Rating and Performance, IEC Standard 60034-1, Ed. 15.0, 2026.
- [19] S. J. Chapman, *Electric Machinery Fundamentals*, 5th ed. New York, NY, USA: McGraw-Hill, 2012.
- [20] Electrical Insulation - Thermal Evaluation and Designation, IEC Standard 60085, Ed. 4.0, 2007.
- [21] C. Guo, G. Pleiss, Y. Sun, and K. Q. Weinberger, "On calibration of modern neural networks," in *Proc. 34th Int. Conf. Machine Learning (ICML)*, PMLR, vol. 70, 2017, pp. 1321-1330.
- [22] B. Efron and R. J. Tibshirani, *An Introduction to the Bootstrap*. New York, NY, USA: Chapman & Hall, 1993.
- [23] M. Liao, Z. Wan, C. Yao, K. Chen, and X. Bai, "Real-time scene text detection with differentiable binarization," in *Proc. AAAI*, vol. 34, no. 7, pp. 11474-11481, 2020, doi: 10.1609/aaai.v34i07.6812.
- [24] B. Shi, X. Bai, and C. Yao, "An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 11, pp. 2298-2304, 2017, doi: 10.1109/TPAMI.2016.2646371.
- [25] X. Zhang, W. Bai, and H. Cui, "Optical character recognition of power equipment nameplate for energy systems based on recurrent neural network," *Front. Energy Res.*, vol. 9, Art. no. 834283, 2022, doi: 10.3389/fenrg.2021.834283.
- [26] Y. Zhao, Y. Xiong, and W. Li, "Text detection of power nameplate based on improved DBNet," in *Proc. China Automation Congress (CAC)*, 2023, pp. 4788-4793, doi: 10.1109/CAC59555.2023.10451112.
- [27] H. Wu, Y. Liu, Z. Jin, and Y. Zhou, "Exposure image correction of electrical equipment nameplate based on the LMPEC algorithm," *PLoS ONE*, vol. 19, no. 6, Art. no. e0300792, 2024, doi: 10.1371/journal.pone.0300792.
- [28] Y. Xu, M. Li, L. Cui, S. Huang, F. Wei, and M. Zhou, "LayoutLM: Pre-training of text and layout for document image understanding," in *Proc. 26th ACM SIGKDD Int. Conf. Knowledge Discovery & Data Mining*, 2020, pp. 1192-1200, doi: 10.1145/3394486.3403172.
- [29] Y. Huang, T. Lv, L. Cui, Y. Lu, and F. Wei, "LayoutLMv3: Pre-training for document AI with unified text and image masking," in *Proc. 30th ACM Int. Conf. Multimedia*, 2022, pp. 4083-4091, doi: 10.1145/3503161.3548112.
- [30] G. Kim, T. Hong, M. Yim, J. Nam, J. Park, J. Yim, W. Hwang, S. Yun, D. Han, and S. Park, "OCR-free document understanding transformer," in *Proc. European Conf. Computer Vision (ECCV)*, 2022, pp. 498-517, doi: 10.1007/978-3-031-19815-1_29.
- [31] S. Bai, K. Chen, X. Liu, J. Wang, W. Ge et al., "Qwen2.5-VL Technical Report," arXiv:2502.13923, 2025.
- [32] M. Li, T. Lv, J. Chen, L. Cui, Y. Lu, D. Florencio, C. Zhang, Z. Li, and F. Wei, "TrOCR: Transformer-based optical character recognition with pre-trained models," in *Proc. AAAI*, vol. 37, no. 11, pp. 13094-13102, 2023, doi: 10.1609/aaai.v37i11.26538.
- [33] ONNX Runtime developers, "ONNX Runtime," 2018. [Online]. Available: <https://onnxruntime.ai>