



## Detecting Indonesian Online Gambling Promotions in Digital Images: An OCR-LSTM Pipeline

*Ariefin Nur Hidayat<sup>1</sup>, Endang Wahyu Pamungkas<sup>2</sup>*

<sup>1</sup> Universitas Muhammadiyah Surakarta, Sukoharjo, Indonesia

**Abstract:** The rapid spread of online gambling promotions via digital images on social media renders manual identification processes highly inefficient, necessitating an automated detection system. This study aims to implement a system to identify online gambling promotions in digital images using a combination of Optical Character Recognition (OCR) and Long Short-Term Memory (LSTM) methods. The research utilizes a dataset of 2,200 images, evenly balanced between gambling and non-gambling categories. The proposed system involves several key stages: image preprocessing using the Contrast Limited Adaptive Histogram Equalization (CLAHE) method to improve contrast, text extraction via PaddleOCR, text and vocabulary correction using FastText, and finally, contextual classification using the LSTM architecture. The testing results on 330 test data demonstrated excellent and stable model performance in recognizing text patterns. The model successfully achieved an impressive accuracy rate of 94%, with a precision of 0.96, a recall of 0.92, and an F1-score of 0.94. In conclusion, the combination of OCR and LSTM technologies is proven to be highly effective for automatically detecting online gambling promotional content.

**Keywords:** Online Gambling, Optical Character Recognition, Long Short-Term Memory, Promotion Detection, Text Extraction

### Article History:

Received: 12 May 2026

Accepted: 30 June 2026

Published: 30 June 2026

**Corresponding Author:** Ariefin Nur Hidayat, Email: [l200220031@student.ums.ac.id](mailto:l200220031@student.ums.ac.id)

DOI: 10.65917/aisa.v2i1.68

## 1 Introduction

The rapid advancement of digital technology has significantly impacted modern society, particularly in the accelerated and widespread dissemination of information through social media. However, this progress brings new challenges, notably the misuse of digital platforms to distribute illegal content, including online gambling advertisements. Gambling is broadly defined as an intentional act of betting in which individuals risk valuable assets on an uncertain event, fully aware of the associated risks [1]. Easy access to online gambling through digital devices such as smartphones often triggers addictive behavior, as initial wins create an illusion of greater future profits and lead players to continuously increase their bets [2]. Consequently, the number of online gambling participants continues to rise.

The spread of online gambling currently relies heavily on image-based promotional content on social media because of its visual appeal and shareability. These promotional strategies, which often combine persuasive language with bonus offers, have become increasingly massive, and the lack of firm enforcement action against online gambling advertisers has exacerbated the situation [3]. The resulting volume of content makes manual moderation highly inefficient, motivating the need for an automated system capable of detecting digital images that contain online gambling promotions.

Indonesian online gambling promotions in digital images tend to follow several recurring patterns. Textually, they often combine informal Indonesian gambling-related slang (e.g., terms referring to winning, depositing, or withdrawing) with distinctive site or brand names, frequently formed by combining a word with a number, a naming convention common to gambling platforms seeking to present a coherent brand identity. Promotional phrasing typically emphasizes bonuses, ease of registration, and guaranteed or high-probability winning, often paired with sports, entertainment, or celebrity-style imagery to increase appeal. Visually, these images commonly use bold, high-contrast typography overlaid on photographic backgrounds, designed to be eye-catching within a fast-scrolling social media feed. Because automated content moderation on social media platforms typically scans caption text or metadata rather than the pixels within an image, embedding promotional text directly inside the image itself is also an effective way to evade simple keyword-based filters. This pattern is precisely what motivates an OCR-based approach: by extracting the text embedded in the image itself, the proposed system



can apply text-based classification to content that would otherwise bypass caption-only moderation.

One effective approach to address this issue is Optical Character Recognition (OCR) technology, which enables computers to extract text from images by recognizing writing patterns and structures [4]. However, text extraction alone is insufficient to determine whether the extracted content constitutes a gambling promotion, since the same words (e.g., sports terms, numbers, or generic promotional language) may also appear in legitimate advertisements. Long Short-Term Memory (LSTM), a recurrent neural network architecture capable of capturing sequential and contextual dependencies in text [5], is therefore used to analyze and classify the context of the extracted text. This research implements a combined OCR and LSTM pipeline to detect online gambling promotions in digital images.

The main contributions of this study are as follows:

1. Construction of a manually curated, balanced dataset of 2,200 Indonesian-language digital images for gambling-promotion detection. This dataset was newly and independently collected from scratch by the authors and is not derived from any pre-existing published dataset, thereby presenting a novel benchmark resource specifically tailored for Indonesian online gambling content moderation.
2. Development of an OCR and text-normalization pipeline (grayscale conversion, CLAHE-based contrast enhancement, PaddleOCR extraction, and FastText-based correction) tailored to noisy, promotional Indonesian-language images.
3. Evaluation of an LSTM-based classifier on OCR-extracted Indonesian text for binary gambling-promotion detection.
4. An error analysis identifying the principal sources of misclassification, including distorted text, noisy backgrounds, and ambiguous gambling slang.

## **2 Related Work (if applicable)**

Previous studies have demonstrated the efficacy of LSTM in natural language processing tasks. Amriza et al. [6] utilized LSTM for emotion detection in social media text, achieving an accuracy of 92.33%, outperforming standard RNN and GRU models.

Similarly, Pipin and Kurniawan [7] applied LSTM for sentiment analysis regarding government policy using Twitter data, achieving a maximum accuracy of 80.42%.

While these studies demonstrate that LSTM is effective for Indonesian-language text classification, both operated on clean, natively digital text collected directly from social media platforms. Neither study addressed the additional noise introduced when text must first be extracted from an image via OCR, such as misrecognized characters, broken or merged words, and inconsistent spacing, nor did they address domain-specific adversarial wording such as intentionally misspelled or coded gambling slang designed to evade keyword-based filters. This study extends the LSTM paradigm to this more challenging setting, in which the input text itself is imperfect and the vocabulary is domain-specific.

On the OCR side, Al amin and Aprilino [8] combined YOLO and Tesseract OCR for automatic vehicle license-plate recognition, a task involving structured, fixed-format alphanumeric text rather than the free-form, stylistically varied text found in promotional images. Shi et al. [9] proposed an end-to-end trainable convolutional-recurrent network (CRNN) for scene-text recognition, a foundational approach to sequence-based text recognition that underlies many modern OCR engines, including the PaddleOCR engine adopted in this study. Compared with these prior applications, the present task is distinguished by its reliance on promotional Indonesian-language text embedded in visually noisy, stylistically diverse images.

### **3 Methods**

This research was conducted following a structured machine learning pipeline, chosen for its systematic approach to handling image-based data and sequential text analysis. The proposed methodology comprises five main phases: data collection, image preprocessing and OCR extraction, text normalization and data splitting, LSTM architecture training, and testing and evaluation.

#### **3.1 Data Collection**

The dataset comprises 2,200 digital images collected from X and various public web sources between November 2025 and March 2026. Images were obtained through manual browsing and targeted keyword searches, followed by the individual evaluation and



downloading of each promotional image to guarantee strict alignment with the inclusion criteria.

The dataset is evenly balanced between two classes: 1,100 images containing online gambling promotions and 1,100 images of non-gambling promotional content. To avoid a narrow or biased representation of the negative class, the non-gambling images were drawn from five broad promotional categories: food and beverage advertisements, retail product advertisements, service advertisements (e.g., travel, finance, or education), event or activity announcements, and general brand or corporate advertisements. This diversification was intended to ensure that the model learns gambling-specific visual and textual cues rather than cues that merely distinguish gambling images from a single type of non-gambling content.

Each image was manually labeled as gambling or non-gambling. This approach was deliberately adopted to guarantee a completely uniform and unified evaluation standard across the entire dataset. An image was labeled as a gambling promotion if it contained explicit visual or textual indicators of gambling activities, including gambling site logos, domain URLs, or specific promotional catchphrases offering bonuses and high winning probabilities.

### ***3.2 Image Preprocessing and OCR Extraction***

To ensure optimal text extraction, each image undergoes a preprocessing pipeline before OCR is applied. First, images are converted to grayscale to simplify subsequent processing and reduce computational load [10]. Contrast Limited Adaptive Histogram Equalization (CLAHE) is then applied to improve local contrast without excessively amplifying noise [11]. By equalizing the local histogram distribution [12], CLAHE sharpens text edges while mitigating the over-enhancement of background noise that can occur with global histogram equalization [13]. Following enhancement, PaddleOCR is used to extract raw text from the image.

### ***3.3 Text Normalization and Data Splitting***

A FastText module is used to correct OCR-related typographical errors in the extracted text. The corrected text then undergoes case folding, tokenization, and normalization of

gambling-related slang into a consistent vocabulary form guided by a lookup dictionary that was manually compiled by extracting the most frequent gambling-related slang words and non-standard variations encountered during the initial data collection phase. The dataset is then divided using stratified random sampling into 70% training (1,540 images), 15% validation (330 images), and 15% testing (330 images). The validation set is used to monitor whether the model is learning generalizable patterns during training [14], while the held-out test set provides an estimate of performance on unseen data and helps control for overfitting [15], [16].

### 3.4 LSTM Architecture and Training

The classification model uses a sequential LSTM architecture, illustrated in Figure 1, motivated by the architectural comparison discussed in Section 2.

- Embedding Layer: maps a 10,000-word vocabulary into 32-dimensional dense vectors.
- Spatial Dropout: applied at a rate of 0.3 to reduce overfitting.
- LSTM Layer: 64 units, with a standard dropout of 0.2 and a recurrent dropout of 0.1, to capture long-term contextual dependencies in the (potentially noisy) extracted text.
- Dense Layers: a fully connected layer of 32 units, followed by a final dense layer with a sigmoid activation function for binary classification.

The model is optimized using the Adam optimizer, which adapts the learning rate for each parameter and accelerates convergence [17]; maintaining an appropriate learning rate is important to avoid slow training or convergence to a poor local minimum [18]. Binary cross-entropy is used as the loss function to measure the discrepancy between predictions and ground-truth labels [19]. Training is conducted for 10 epochs with a batch size of 32.



Figure 1. LSTM Architecture used for classifying OCR-extracted text.



### 3.5 Text Normalization and Data Splitting

After training, the model is evaluated on the held-out 15% test set (330 images) using a confusion matrix that compares predictions against ground-truth labels. From the resulting True Positive, True Negative, False Positive, and False Negative counts, the performance metrics summarized in Table 1 are derived. Evaluation in this study is based on a single stratified train-validation-test split rather than k-fold cross-validation; this choice and its implications are discussed as a limitation in Section 4.5.

Metric	Formula	Description
Accuracy	$(TP + TN) / (TP + TN + FP + FN)$	Proportion of correctly predicted instances out of all data; a high value indicates good overall predictive capability.
Precision	$TP / (TP + FP)$	Proportion of positive predictions that are actually correct; a high value means the model rarely misclassifies negative data as positive.
Recall	$TP / (TP + FN)$	Ability of the model to detect all actual positive data; a high value means the model captures most gambling promotions.
F1-Score	$2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall})$	Harmonic mean of precision and recall, providing a balanced evaluation [27].

Table 1. Evaluation metrics, formulas, and descriptions.

## 4 Results and Discussion

The preprocessing pipeline improved the legibility of text within the images, allowing PaddleOCR to extract character sequences with higher fidelity. Training also showed a stable convergence pattern, as detailed below.

### 4.1 Preprocessing and OCR Analysis

Grayscale conversion followed by CLAHE noticeably improved image quality by clarifying text characters, as illustrated in Figure 2. This step is particularly important because online gambling promotions often use visually complex backgrounds, such as photographic imagery and decorative typography, that can interfere with standard OCR. By sharpening text edges and balancing contrast locally, PaddleOCR was able to extract promotional keywords with higher fidelity, including from comparatively low-quality digital images.



Figure 2. Example of a digital image containing an online gambling promotion.

### 4.2 Training Performance and Learning Curves

Training over 10 epochs showed stable convergence without strong indications of overfitting, as shown in Figure 3. Both training and validation accuracy increased steadily, reaching approximately 94% by the final epoch, while training and validation loss decreased consistently and remained close to one another throughout training. This pattern is consistent with the Adam optimizer guiding the model toward a stable solution while learning generalizable patterns from the 1,540 training samples, although, as noted above, this was assessed on a single train-validation split rather than across multiple cross-validation folds.

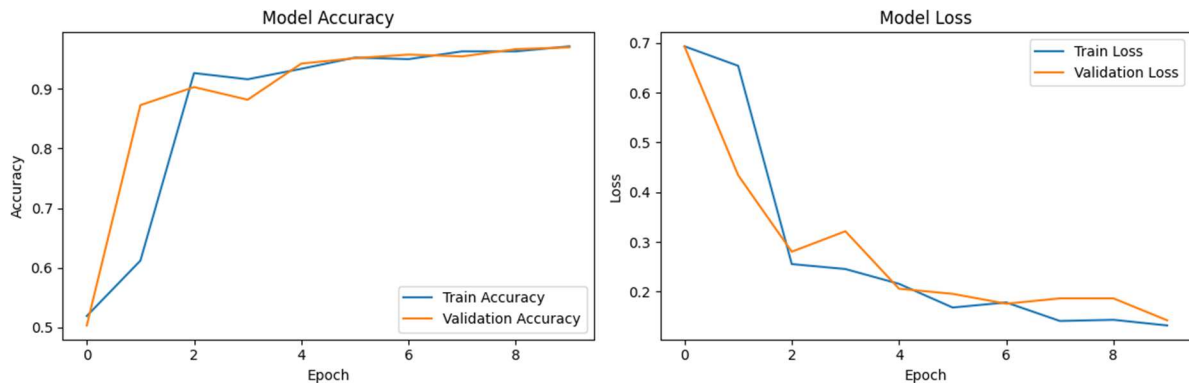


Figure 3. Training and validation accuracy and loss curves over 10 epochs.

### 4.3 Classification Performance

Evaluation on the 330 held-out test images is summarized in Table 2. The system achieved a macro-averaged F1-score of 0.94, with consistent performance across both classes: an F1-score of 0.94 for the non-gambling class (recall 0.96, precision 0.92) and 0.94 for the gambling class (recall 0.92, precision 0.96). The relatively high precision for the gambling class (0.96) indicates that when the model flags an image as a gambling promotion, it is rarely wrong, while the recall of 0.92 indicates that a modest fraction of actual gambling promotions were missed; this trade-off is examined further in the error analysis below.

Label	Recall	Precision	F1-Score	Accuracy
Non-Gambling	0.96	0.92	0.94	<b>94%</b>
Gambling	0.92	0.96	0.94	
<b>Total (macro avg.)</b>	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>	

Table 2. Classification report.

### 4.4 Error Analysis

The confusion matrix in Figure 4 provides a detailed view of the model's predictive behavior. The model correctly identified 151 gambling images (true positives) and 159 non-gambling images (true negatives). It produced 6 false positives, in which non-gambling content was incorrectly classified as gambling, and 14 false negatives, in which actual gambling images

went undetected. Manual inspection of the misclassified images suggested that these errors were primarily associated with heavily distorted or stylized text, extreme background noise, or slang terminology that was ambiguous even on manual review. The comparatively higher number of false negatives relative to false positives suggests that the model is somewhat conservative in flagging content as gambling-related, which may be preferable from a precision standpoint but implies that some real promotions could currently evade detection.



Figure 4. Confusion matrix on the held-out test set.

#### 4.5 Limitations

While the proposed pipeline achieved strong performance on the held-out test set, several limitations should be acknowledged. First, evaluation relied on a single stratified train-validation-test split rather than k-fold cross-validation, so the reported metrics may not fully reflect performance variability across different data partitions. Second, the LSTM classifier was not directly benchmarked against alternative approaches such as OCR+SVM, OCR+CNN, or transformer-based classifiers (e.g., IndoBERT), which would clarify the specific contribution of the LSTM component. Third, all images were collected from X and various public web sources within the specific timeframe of November 2025 to March 2026. so the extent to which the model generalizes to gambling promotions from other platforms, later time periods, or newer evasion tactics remains untested. These limitations directly motivate the future-work directions outlined in Section 5.



## 5 Conclusion

This study presented an OCR-LSTM pipeline for detecting Indonesian-language online gambling promotions in digital images, together with a manually curated, balanced dataset of 2,200 images spanning gambling and diverse non-gambling promotional categories. The pipeline combines grayscale conversion and CLAHE-based contrast enhancement, PaddleOCR-based text extraction, FastText-based correction and slang normalization, and an LSTM classifier. On a held-out test set of 330 images, the system achieved an accuracy of 94% and a macro-averaged F1-score of 0.94, suggesting that the combination of OCR and LSTM is a viable approach for this task under the conditions evaluated. Future work should: (1) validate the pipeline on images drawn from additional platforms, time periods, and visual styles; (2) benchmark the LSTM classifier against alternative approaches such as OCR+SVM, OCR+CNN, OCR+IndoBERT, and multimodal vision-language models; (3) assess robustness using k-fold cross-validation; and (4) expand the dataset to include more examples of heavily distorted or adversarially obscured text.

## Acknowledgments

The author wishes to thank Endang Wahyu Pamungkas, S.Kom., M.Kom., Ph.D. for supervision and guidance throughout this research.

## Funding Information

The authors declare no funding was received for this study.

## Conflict of Interest Statement

The authors declare no conflicts of interest.

## Ethical Approval

This study did not involve human or animal subjects.

## Data Availability

Data is available upon request from the corresponding author.

## References

- [1] A. Nasyiah, "Game Multi-Platform Untuk Adab Dan Akhlak Anak Muslim Menggunakan Metode Game Development Live Cycle (Gdlc)," *Jurnal Informatika Dan Rekayasa Perangkat Lunak*, vol. 3, no. 3, pp. 254-265, 2022.
- [2] A. S. Sitanggang, R. Sabta, and H. Fani Yuli, "Perkembangan Judi Online dan Dampaknya Terhadap Masyarakat: Tinjauan Multidisipliner," *Triwikama: Jurnal Ilmu Sosial*, vol. 1, no. 5, pp. 1-7, 2023.
- [3] P. P. Rastika Paramartha, S. Laksamani Dewi, and I. P. Gede Saputra, "Sanksi Pidana terhadap Para Pemasang dan Promosi Iklan Bermuatan Konten Judi Online," *Jurnal Preferensi Hukum*, vol. 2, no. 1, pp. 156-160, 2021.
- [4] Z. R. Mair and M. A. Rahmanda, "Perbandingan Versi Terbaik YOLO Dalam Mendeteksi Jarak Spasi Antar Baris Tulisan Tangan," *Jurnal Sains, Nalar, Dan Aplikasi Teknologi Informasi*, vol. 4, no. 2, pp. 103-110, 2025.
- [5] A. Fauzul, A. Kustiwa, M. Aminul Akbar, and A. Pinandito, "Studi Perbandingan pada Metode CNN-LSTM dan LSTM dalam Mendeteksi Emosi pada Data Teks Berbahasa Indonesia pada Media Sosial Twitter," *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, vol. 1, no. 1, pp. 2548-2964, 2017.
- [6] R. N. S. Amriza, D. Supriyadi, K. Banyumas, and J. Tengah, "Komparasi Metode Machine Learning dan Deep Learning untuk Deteksi Emosi pada Text di Sosial Media," *Jurnal JUPITER*, vol. 13, no. 2, pp. 130-139, 2021.
- [7] S. J. Pipin and H. Kurniawan, "Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM," *Jurnal SIFO Mikroskil*, vol. 23, no. 2, pp. 197-208, 2022.
- [8] I. H. Al amin and A. Aprilino, "Implementasi Algoritma Yolo Dan Tesseract Ocr Pada Sistem Deteksi Plat Nomor Otomatis," *Jurnal Teknoinfo*, vol. 16, no. 1, p. 54, 2022.
- [9] 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 Transactions on Pattern Analysis and Machine Intelligence*, 2015.
- [10] I. Christian and G. P. Kusuma, "Improving OCR Performance on Low-Quality Image Using Pre-processing and Post-processing Methods," *International Journal of Engineering Trends and Technology*, vol. 71, no. 6, pp. 396-405, 2023.
- [11] Z. Al-ameen, G. Sulong, A. Rehman, A. Al-dhelaan, and T. Saba, "An innovative technique for contrast enhancement of computed tomography images using adaptive histogram equalization," *EURASIP Journal on Advances in Signal Processing*, pp. 1-12, 2015.
- [12] N. Sapkota, "Real Time Digital Night Vision Using Nonlinear Contrast Enhancement," *UNLV Theses, Dissertations, Professional Papers, and Capstones*, 2013.



- [13] M. Opoku, B. A. Weyori, A. F. Adekoya, and K. Adu, "CLAHE-CapsNet: Efficient retina optical coherence tomography classification using capsule networks with contrast limited adaptive histogram equalization," pp. 1-22, 2023.
- [14] H. Imaduddin, F. Y. A, and Y. S. Nugroho, "Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach," vol. 14, no. 8, pp. 113-117, 2023.
- [15] Y. K. Bintang and H. Imaduddin, "Pengembangan model deep learning untuk deteksi retinopati diabetik menggunakan metode transfer learning," vol. 9, no. 3, pp. 1442-1455, 2024.
- [16] V. Singh et al., "Impact of train/test sample regimen on performance estimate stability of machine learning in cardiovascular imaging," Scientific Reports, pp. 1-8, 2021.
- [17] H. Iiduka, "Appropriate Learning Rates of Adaptive Learning Rate Optimization Algorithms for Training Deep Neural," IEEE Transactions on Cybernetics, pp. 1-22, 2021.
- [18] M. Reyad and A. M. Sarhan, "A modified Adam algorithm for deep neural network optimization," Neural Computing and Applications, vol. 35, no. 23, pp. 17095-17112, 2023.
- [19] C. Guo and X. Chen, "Multi-Stage Attentive Network for Motion Deblurring via Binary Cross-Entropy Loss," Entropy, pp. 1-14, 2022.
- [20] D. R. Alghifari, M. Edi, and L. Firmansyah, "Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia," Jurnal Manajemen Informatika (JAMIKA), vol. 12, pp. 89-99, 2022.