A Hybrid AI Framework for CV Screening with CNN-Based Layout Classification and Open-Source LLMs

by Johan Setiawan

Submission date: 03-Feb-2026 04:18PM (UTC+0700)
Submission ID: 2870104408
File name: yout_Classification_and_Open-Source_LLMs_-_After_Revision_2.pdf (770.08K)
Word count: 3879
Character count: 23933

A Hybrid AI Framework for CV Screening with CNN-Based Layout Classification and Open-Source LLMs

1st Emilio Yanvrent
Department of Information Systems
Universitas Multimedia Nusantara
Tangerang, Indonesia
emilio.yanvrent@student.umn.ac.id

2nd Johan Setiawan
Department of Information Systems
Universitas Multimedia Nusantara
Tangerang, Indonesia
johan@umn.ac.id

***Abstract*— The growing influx of digital Curriculum Vitae (CV) has created significant challenges for recruiters in efficiently identifying qualified candidates. This study proposed a dual-stage hybrid artificial intelligence (AI) framework that integrated Convolutional Neural Network (CNN)-based visual classification with semantic analysis using an open-source Large Language Model (LLM) to streamline the recruitment workflow. In the first stage, the CNN model classifies CV with 98% accuracy and evaluates their compatibility with Applicant Tracking Systems (ATS), achieving 92% classification accuracy. In the second stage, a fine-tuned LLM extracts and evaluates candidate information, generates concise summaries, and identifies missing key competencies. Unlike proprietary API-based solutions, the proposed system leverages open-source models to ensure transparency, enable local deployment, and support ethical handling of personally identifiable information (PII). Experimental results show an F1-score above 0.86 for skill extraction and summary evaluation. The framework outputs structured data in JSON format for seamless integration with recruitment platforms. Despite limitations related to dataset availability and evaluation diversity, the proposed approach demonstrates strong potential for scalable, interpretable, and customizable CV analysis in modern human resource systems. Future work would explore job-matching capabilities, multilingual support, bias detection, and compliance with global data protection regulations. This research underscores the feasibility of integrating vision and language models to develop ethical and intelligent recruitment workflows.**

***Keywords*— Applicant Tracking System, Convolutional Neural Network, CV Screening, Large Language Model, Recruitment Automation**

I. Introduction

Recruiting the right talent is a strategic imperative for any organization [1]. Employees drive innovation, productivity, and growth while shaping company culture and long-term competitiveness [2]. Modern hiring is increasingly complex due to online platforms and remote work, which have dramatically increased applicant volume and extended recruitment timelines [3], [4].

Traditional recruitment methods, particularly manual CV screening, are no longer sufficient in high-volume environments [5]. These processes are time-consuming, inconsistent, and prone to unconscious bias, often causing qualified candidates to be overlooked [6]. To address these challenges, many organizations have adopted artificial intelligence (AI) tools to automate and streamline recruitment. However, most current AI implementations rely on simplistic keyword matching or proprietary large language models (LLMs) that lack transparency, interpretability, and adaptability to specific organizational needs [7].

This study proposed a hybrid AI framework that integrates layout-based structural analysis of CV with semantic processing using open-source LLMs to deliver a scalable, accurate, and explainable recruitment solution [8]. The framework operated in two stages: first, a convolutional neural network (CNN) processed CVs in various formats including PDFs, images, and scanned documents to classify visual layout elements such as headers and sections and assess compatibility with Applicant Tracking Systems (ATS), thereby standardizing formatting and filtering out low-quality inputs. Second, a fine-tuned open-source LLM extracts key candidate attributes, including skills, experience, and education, evaluates their relevance to specific job roles, generates concise summaries, and detects missing competencies.

This dual-layer approach offers several advantages: improved efficiency through automation, reduced bias via consistent evaluation, and enhanced transparency using open-source models [9]. The framework is modular and customizable, enabling integration with existing ATS platforms and adaptation to enterprise-specific requirements. By supporting diverse CV formats and delivering detailed candidate assessments, the proposed system is particularly suited to high-demand sectors such as technology, healthcare, and customer service. Furthermore, by digitizing and automating core recruitment tasks, it supports remote workflows, allowing recruiters and hiring managers to manage candidate pipelines effectively from any location [10].

This paper details the design, development, and evaluation of the proposed system and examines its potential to make candidate screening faster, fairer, and more interpretable through responsible, open-source AI technologies. The following section reviews existing research on AI-driven recruitment, highlighting current capabilities and persistent limitations.

II. Literature Review

The integration of Artificial Intelligence (AI) in recruitment has notably improved efficiency and accuracy. AI-driven tools reduce time-to-hire by up to 45% through automated CV screening and scheduling [11], with adoption rising to 55% by 2020 among U.S. firms using predictive analytics [12]. Surveys further confirm this trend: by 2022, 92% of HR leaders planned to expand AI use, with 73% already applying it to recruitment. By 2024, 64% of U.S.

organizations reported talent acquisition as the main AI application area, ahead of learning and development (43%) and performance management (25%) [13].

Broader industry surveys reinforce AI's value in streamlining hiring. In 2017, 67% of 8,815 HR professionals in a Statista survey believed AI helps reduce recruitment time [14]. Similar studies show widespread adoption, with 99% of Fortune 500 companies implementing AI-based recruitment systems [15]. Collectively, these findings highlight AI's growing role in accelerating hiring timelines while improving precision, scalability, and consistency.

III. Methodology

Our approach adheres to the Knowledge Discovery in Databases (KDD) paradigm, as depicted in Fig. 1, integrating document preprocessing, visual feature extraction, and convolutional neural network (CNN)-based modeling to enhance the automated classification of CV. This process supports intelligent recruitment systems by streamlining the identification of ATS-compliant documents and optimizing candidate screening.

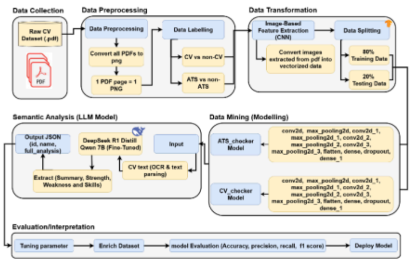

Fig. 1. Knowledge Discovery in Databases (KDD) process

A. Dataset

A structured dataset of CV was obtained from internal company records through a private recruitment partner. These authentic CV, submitted by job applicants, were selected to represent diverse layouts and formatting styles. Due to the proprietary nature of the data, the original documents cannot be publicly released. Each PDF was converted into per-page PNG images to preserve key visual elements such as layout hierarchy, typography, and symbols. Although the dataset is not publicly accessible, all experiments used anonymized data and followed a standardized preprocessing pipeline to ensure consistency. Future replications may be used synthetically generated CV to simulate similar structural diversity.

B. Data Labeling

After preprocessing, each converted image was manually annotated with two binary labels to support supervised learning. The first label identifies whether the document is a CV or a non-relevant document, ensuring irrelevant files are excluded from the recruitment pipeline. The second label assesses the CV compatibility with Applicant Tracking Systems (ATS), based on features such as clear headings, minimal graphics, and machine-readable layouts.

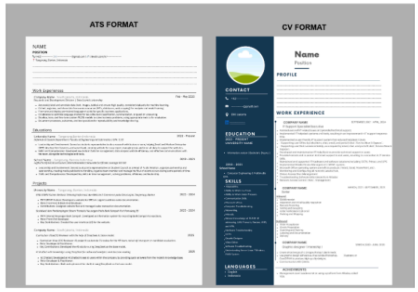

Fig. 2. Examples of CV layouts

To illustrate this distinction, Fig. 2 show examples of ATS-compliant and non-compliant CV, respectively. This dual-labeling approach provides a reliable ground truth for training models to evaluate both content relevance and formatting quality

C. Feature Extraction

Feature extractions were performed using Convolutional Neural Networks (CNNs), which transformed the visual content of each image into high-dimensional feature vectors[16]. These vectors encapsulated structural and stylistic cues, enabling the model to assess document classification and formatting compliance. The core operation in CNNs for feature extraction was the two-dimensional convolution [17].

For multi-channel inputs, such as RGB images, the convolution operation extends across all channels. To ensure robust evaluation, the dataset was split using stratified sampling into 80% training and 20% testing subsets, preserving class distributions across both sets

D. CNN-Based Modelling

To support automated document classification, a unified Convolutional Neural Network (CNN) architecture was employed for both CV detection and ATS compliance assessment:

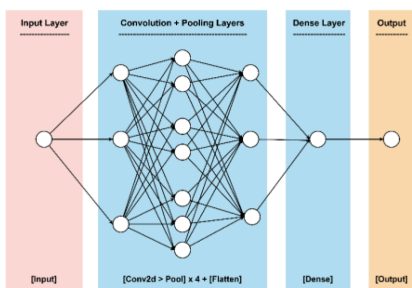

Fig. 3. CNN architecture for detection

As illustrated in Fig. 3, the architecture consisted of sequential layers: an input layer for preprocessed PDF images, four convolutional layers with max-pooling to extract typographic and structural features [18]. and a flatten layer to convert feature maps into a vector. This vector was passed to fully connected dense layers, with dropout applied to reduce overfitting, before reaching a final SoftMax layer for binary classification [19].

The same CNN design was trained separately for two tasks: (1) distinguishing CV from non-CV, and (2) evaluating ATS-compliant versus non-compliant CV. Training was performed with labeled data using backpropagation and gradient descent, while hyperparameters (learning rate, dropout rate, batch size, filter count) were tuned to optimize performance and prevent overfitting.

This shared modeling approach streamlines the pipeline, ensuring consistent architecture while enabling accurate classification across both tasks [20].

E. Semantic Analytics (LLM Model)

To complement CNN-based structural classification, this study integrated a semantic analysis module using the open-source LLM DeepSeek R1 Distill Qwen-7B. It was fine-tuned on anonymized CV data to extract attributes such as strengths, weaknesses, and key skills [21]. The text was then processed with a recruiter-style prompt that organized extracted information into four structured sections: summary, strengths, weaknesses, and skills. This approach ensured clarity, consistency, and a professional tone in the outputs.

F. Evaluation and Interpretation

Both classifiers were evaluated using standard binary metrics: accuracy, precision, recall, and F1-score. Accuracy measured overall correctness [22]. Precision reflected reliability in detecting positives [23]. Recall captured sensitivity to true cases [24], and the F1-score balanced both, making it valuable where misclassification costs differed [25].

Insights from the evaluation informed two feedback mechanisms. First, low-performing samples were flagged for re-labeling or augmentation to enhance data quality [26]. Second, models were refined through hyperparameter tuning and architectural adjustments to reduce recurring errors [27]. This iterative process improved performance and demonstrated the system's reliability in filtering non-CV and non-ATS-compliant documents, significantly streamlining recruitment workflows.

IV. RESULTS AND DISCUSSION

This section presents the outcomes of our hybrid AI framework, emphasizing the performance of the Convolutional Neural Network (CNN)-based classifiers for CV and ATS compliance detection, as well as the deployment of the system for practical recruitment applications.

A. CV Classification Model Performance

Classification Report (Binary: 1 = CV, 0 = Not CV):

	precision	recall	f1-score	support
0	0.98	0.98	0.98	10834
1	0.98	0.98	0.98	10834
accuracy			0.98	21668
macro avg	0.98	0.98	0.98	21668
weighted avg	0.98	0.98	0.98	21668

Fig. 4. Classification report of CV detection model

The performance of the CV classification model was illustrated in Fig. 4, which presented the classification report with key evaluation metrics. The model achieved an accuracy of 98%, along with precision, recall, and F1-scores of 0.98 for both CV (label 1) and non-CV (label 0). The balanced class support (10,834 samples each) and high macro and weighted averages demonstrated the model's strong capability in handling both positive and negative classifications with minimal bias.

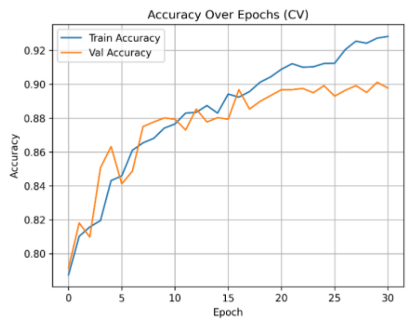

Fig. 5. Training and validation accuracy trends of CV

Figure 5 visualized the training and validation accuracy trends over 30 epochs. The training accuracy rapidly increased and stabilized above 0.97, indicating effective learning of the training data. The validation accuracy followed a similar upward trend, peaking at around 0.96 and maintaining stability with minor fluctuations, which suggested good generalization and minimal overfitting. Together, these figures confirmed the model's robustness, consistent learning behavior, and reliability in distinguishing CV documents from irrelevant content in diverse formats.

B. ATS Compliance Classification Model Performance

Classification Report (Binary: 1 = ATS, 0 = Not ATS):

	precision	recall	f1-score	support
0	0.92	0.92	0.92	3002
1	0.92	0.92	0.92	3002
accuracy			0.92	6004
macro avg	0.92	0.92	0.92	6004
weighted avg	0.92	0.92	0.92	6004

Fig. 6. Performance metrics of the ATS detection model

Figure 6 presented the performance metrics of the ATS compliance classification model, showing balanced results across both classes (ATS-compliant and non-compliant) with precision, recall, and F1-score of 0.92, and an overall accuracy of 92%.

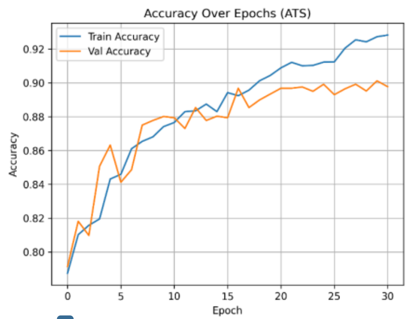

Fig. 7. Training and validation accuracy trends of ATS

Figure 7 depicted the training and validation accuracy trends over 30 epochs, which indicated steady improvement in accuracy with the training set, while the validation accuracy plateaued around 0.89. This suggested good generalization without significant overfitting.

C. CNN Architecture

The CNN architecture applied in both the CV classification and ATS compliance task was identical and was illustrated in Fig. 3. It consisted of four convolutional layers interspersed with max pooling, followed by a flattening operation, dense layers, dropout, and a final SoftMax output. This design was trained separately for each task using task-specific labeled datasets, but the underlying structure remained unchanged. By maintaining unified architecture, the framework ensures consistency in feature extraction and computational efficiency, while allowing independent optimization for CV detection and ATS compliance evaluation.

D. System Development

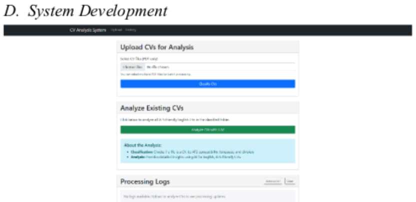

Fig. 8. User interface of deployed system

Figure 8 highlighted the system deployment interface, which provided a user-friendly platform for uploading and analyzing CV. The interface included functionalities for uploading new documents, reviewing existing CV, and tracking processing status through logs, thereby supporting seamless interaction between the user and the AI-powered analysis system.

E. Large Language Model (LLM)

The semantic analysis module was evaluated using the fine-tuned DeepSeek R1 Distill Qwen-7B model with a structured prompt to extract candidate information into four sections: summary, strengths, weaknesses, and skills. This ensured clarity, consistency, and professional tone in the outputs.

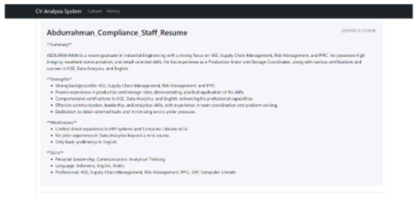

Fig. 9. Example output of LLM-based semantic analysis

Figure 9 illustrated an example analysis of a CV, where the model correctly identified strengths (e.g., HSE, supply chain, risk management), weaknesses (limited ERP experience), and categorized skills (personal, language, professional). Although the model occasionally generated conversational phrases (e.g., "Alright, let's dive into this CV analysis"), these did not affect accuracy and mainly reflected challenges in prompt control.

TABLE I. LLM BENCHMARK

Extracted Information	Precision	Recall	F1-Score
Summary	0.90	0.86	0.88
Strengths	0.85	0.80	0.82
Weakness	0.77	0.73	0.75
Skills	0.88	0.85	0.86

Table I shows a benchmarking study with five anonymized CV compared model outputs against manually annotated ground truth using precision, recall, and F1-score. Results are summarized in Table 1, showing a powerful performance across categories, with highest accuracy for skill extraction (F1 = 0.86) and summary generation (F1 = 0.88). Slightly lower performance was observed for weaknesses (F1 = 0.75), due to the implicit nature of such information. Overall, the model demonstrated reliable semantic extraction while highlighting areas for improvement in prompt engineering and fine-tuning.

F. *Result Comparison*

TABLE II. PREVIOUS STUDIES

Study and Method	Dataset Types	Accuracy	Limitations
NLP + SVM with TF-IDF [28]	Text Only	96.6%	Ignores Layout
CNN-GRU + BERT [29]	Text + Embedding	93.5%	Limited generalizability
CV2Vec [30]	Embedding-based	92%	Proprietary pipeline
LLM for ATS [31]	Text + LLM	85-90%	No visual Features
Proposed framework	Layout + Semantic	98% (CV) & 92% (ATS)	Private dataset

Table II compared our proposed framework with related studies. While text-only and embedding-based models achieved strong accuracy, they neglected layout features or depended on proprietary pipelines. Our dual-stage CNN + LLM framework provided competitive accuracy while integrating both structural and semantic analysis, offering a more holistic and interpretable solution.

G. *Ethical Considerations and Bias Mitigation*

Ethical integrity is critical in AI-driven recruitment because of its impact on career outcomes. This study addresses bias, fairness, transparency, and data protection through a multi-layered ethical framework.

The pipeline applies to two CNN models: one trained on 10,834 samples to identify CV, and another on 3,002 samples for ATS compliance. Only documents passing both stages are analyzed by the fine-tuned LLM, which extracts structured insights (strengths, weaknesses, and skills) using a prompt-guided approach. These outputs, formatted in JSON, enable human oversight and support interpretability.

To safeguard privacy, all data was processed locally with minimal handling of personally identifiable information. Although not fully aligned with GDPR or Indonesia's PDP Law, the system follows privacy-aware practices suitable for deployment. To mitigate bias, the system incorporates: (1) fairness audits using SHAP and AIF360, (2) balanced synthetic datasets to reduce gender and age imbalance, (3) evaluation on diverse demographic samples, and (4) automated alerts during retraining to flag bias. These measures strengthened accountability and ensured equitable outcomes.

H. *Implication for Practices*

The hybrid AI framework streamlines recruitment by combining CNN-based classification with LLM-driven semantic analysis. CNNs filter out non-CV or poorly formatted documents, while the LLM extracts structured insights (summary, strengths, weaknesses, and skills) from ATS-compliant CV.

Outputs in JSON format integrate easily with recruitment systems, supporting human decision-making without replacing it. Its modular design and local data storage ensure adaptability and privacy, making the framework suitable even in environments with limited cloud access. Although not intended for automated ranking, it provides a scalable and transparent tool that reduces recruiter workload in high-volume fields such as technology, administration, and customer service.

V. CONCLUSION

This study proposed a hybrid AI framework integrating CNN-based document classification with fine-tuned open-source LLM semantic analysis for CV screening. The system achieved 98% accuracy in CV identification and 92% in ATS compliance, confirming its reliability for early-stage candidate filtering.

To ensure transparency and fairness, LLM analysis was applied only to ATS-compliant CV under a predefined prompt, producing structured summaries of candidate strengths, weaknesses, and skills to support rather than replace human decision-making. Ethical principles guided design choices, including explainable outputs, local data storage, and minimal use of personal data. While not fully compliant with GDPR or Indonesia's PDP Law, the system follows privacy-aware practices suitable for localized deployment.

Limitations include the LLM's difficulty in handling nuanced, job-specific, or cultural context, and the absence of automated ranking or job matching. Future work might explore fairness audits, multilingual processing, fairness-aware learning, and job description analysis to improve candidate-role alignment.

Overall, the framework offers a scalable, interpretable, and ethically conscious approach to AI-assisted CV screening, balancing efficiency with fairness and human oversight.

ACKNOWLEDGMENT

This study was supported by the Institution of Research and Community Services at Multimedia Nusantara University. We also thank our colleagues from the Big Data Laboratory, Information Systems Department, for their valuable insights and contributions.

REFERENCES

[1] F. M. M. Opada, M. B. H. Ibrahim, A. Irawan, M. A. Akbar, and A. Rasyid, "Talent acquisition strategies: A comprehensive examination of recruitment policies for organizational success," *Advances in Human Resource Management Research*, vol. 2, no. 2, May 2024, doi: 10.60079/ahrmr.v2i2.185.

[2] M. Kaliannan, D. Darmalinggam, M. Dorasamy, and M. Abraham, "Inclusive talent development as a key talent management approach: A systematic literature review," *Human Resource Management Review*, vol. 33, no. 1, p. 100926, Mar. 2023, doi: 10.1016/j.hrmr.2022.100926.

[3] S. M. Aparna and S. Kumar, "Changing landscapes in recruitment: A critical analysis of past, present and future," *Personnel Review*, May 2025, doi: 10.1108/PR-01-2024-0098.

[4] M. D. Solanki and P. Gujarati, "The digital revolution in recruitment: Unraveling the impact and challenges of e-recruitment," *Educational Administration Theory and Practices*, Jun. 2024, doi: 10.53555/kuey.v30i6(S).5362.

[5] L. Sivabalan, R. Yazdanifard, and N. H. Ismail, "How to transform the traditional way of recruitment into online system," *International Business Research*, vol. 7, no. 3, Feb. 2014, doi: 10.5539/ibr.v7n3p178.

[6] E. Albaroudi, T. Mansouri, and A. Alameer, "A comprehensive review of AI techniques for addressing algorithmic bias in job hiring," *AI*, vol. 5, no. 1, pp. 383–404, Feb. 2024, doi: 10.3390/ai5010019.

[7] Z. Atf and et al., "The challenge of uncertainty quantification of large language models in medicine," Apr. 2025. [Online]. Available: http://arxiv.org/abs/2504.05278

[8] A. Ujlayan, S. Bhattacharya, and Sonakshi, "A machine learning-based AI framework to optimize the recruitment screening process," *International Journal of Global Business and Competitiveness*, vol. 18, no. S1, pp. 38–53, Dec. 2023, doi: 10.1007/s42943-023-00086-y.

[9] M. Soori, F. K. G. Jough, R. Dastres, and B. Arezoo, "AI-based decision support systems in Industry 4.0: A review," *Journal of*

Economy and Technology, Aug. 2024, doi: 10.1016/j.ject.2024.08.005.
[10] J. Setiawan and R. G. Alamsari, "Prediction of work from home post COVID-19 using classification model," in *Proc. 7th Int. Conf. Informatics Comput. (ICIC)*, Dec. 2022, pp. 1–6. doi: 10.1109/ICIC56845.2022.10006898.
[11] S. A. Shanubhog, "Revolutionizing recruitment: AI-powered talent acquisition in the Microsoft ecosystem," *International Journal of Advanced Research in Engineering and Technology*, vol. 16, no. 2, pp. 1–14, Mar. 2025, doi: 10.34218/IJARET_16_02_001.
[12] A. Tursunbayeva, V. Fernandez, E. Gallardo-Gallardo, and L. Moschera, "Artificial intelligence and digital data in recruitment: Exploring business and engineering candidates' perceptions of organizational attractiveness," *European Management Journal*, Mar. 2025, doi: 10.1016/j.emj.2025.03.002.
[13] G. Koman, P. Boršoš, and M. Kubina, "The possibilities of using artificial intelligence as a key technology in the current employee recruitment process," *Adm Sci*, vol. 14, no. 7, Jul. 2024, doi: 10.3390/admsci14070157.
[14] Abdurrahim, "Artificial intelligence: Optimizing the recruitment process," Nov. 2024. [Online]. Available: https://plus62.isha.or.id/index.php/cebong/article/view/284
[15] R. Meshram, "The role of artificial intelligence (AI) in recruitment and selection of employees in the organisation," *Russian Law Journal*, vol. 11, no. 9s, Apr. 2023, doi: 10.52783/rlj.v11i9s.1624.
[16] L. Alzubaidi and others, "Review of deep learning: concepts, CNN architectures, challenges, applications, future directions," *J Big Data*, vol. 8, no. 1, p. 53, Mar. 2021, doi: 10.1186/s40537-021-00444-8.
[17] A. Mjahad, A. Polo-Aguado, L. Llorens-Serrano, and A. Rosado-Muñoz, "Optimizing image feature extraction with convolutional neural networks for chicken meat detection applications," *Applied Sciences*, vol. 15, no. 2, p. 733, Jan. 2025, doi: 10.3390/app15020733.
[18] S. A. Sanjaya, D. A. Kristiyanti, I. Irmawati, F. I. Hadinata, and C. N. Karaeng, "A comparative analysis of building hidden layer, activation function, and optimizer on neural network sentiment analysis," *International Journal on Informatics Visualization*, vol. 9, no. 3, p. 1202, May 2025, doi: 10.62527/joiv.9.3.2906.
[19] L. Zhao and Z. Zhang, "An improved pooling method for convolutional neural networks," *Sci Rep*, vol. 14, no. 1, p. 1589, Jan. 2024, doi: 10.1038/s41598-024-51258-6.
[20] L. Liao, H. Li, W. Shang, and L. Ma, "An empirical study of the impact of hyperparameter tuning and model optimization on the performance properties of deep neural networks," *ACM Transactions on Software Engineering and Methodology*, vol. 31, no. 3, pp. 1–40, Jul. 2022, doi: 10.1145/3506695.
[21] I. H. Sarker, "Data science and analytics: An overview from data-driven smart computing, decision-making and applications perspective," *SN Comput Sci*, vol. 2, no. 5, p. 377, Sep. 2021, doi: 10.1007/s42979-021-00765-8.
[22] P. Sisodia, "Optical character recognition development using Python," *Journal of Informatics Electrical and Electronics Engineering*, vol. 4, no. 3, pp. 1–13, 2023, doi: 10.54060/jieee.2023.75.
[23] F. Saeik and others, "Task offloading in edge and cloud computing: A survey on mathematical, artificial intelligence and control theory solutions," *Computer Networks*, vol. 195, Aug. 2021, doi: 10.1016/j.comnet.2021.108177.
[24] "Accuracy vs. precision vs. recall in machine learning," Jun. 2025. [Online]. Available: https://encord.com/blog/classification-metrics-accuracy-precision-recall/
[25] P. Christen, D. J. Hand, and N. Kirielle, "A review of the F-measure: Its history, properties, criticism, and alternatives," *ACM Comput Surv*, vol. 56, no. 3, pp. 1–24, Mar. 2024, doi: 10.1145/3606367.
[26] K. Maharana, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Global Transitions Proceedings*, vol. 3, no. 1, pp. 91–99, Jun. 2022, doi: 10.1016/j.gltp.2022.04.020.
[27] T. Agrawal, *Hyperparameter Optimization in Machine Learning*. Berkeley, CA: Apress, 2021. doi: 10.1007/978-1-4842-6579-6.
[28] I. Ali, N. Mughal, Z. H. Khan, J. Ahmed, and G. Mujtaba, "CV classification system using natural language processing and machine learning techniques," *Mehran University Research Journal of Engineering and Technology*, vol. 41, no. 1, pp. 65–79, Jan. 2022, doi: 10.22581/muet1982.2201.07.
[29] A. Qostal, A. Moumen, and Y. Lakhrissi, "CV classification using neural network approaches combined with BERT and Gensim: CV of Moroccan engineering students," *Data (Basel)*, vol. 9, no. 6, p. 74, May 2024, doi: 10.3390/data9060074.
[30] R. V. K. Bevara and others, "CV2Vec: Transforming applicant tracking systems with intelligent CV embeddings for precise candidate matching," *Electronics (Basel)*, vol. 14, no. 4, p. 794, Feb. 2025, doi: 10.3390/electronics14040794.
[31] S. Anand and A. K. Giri, "Optimizing CV design for ATS compatibility: A large language model approach," *International Journal on Smart & Sustainable Intelligent Computing*, vol. 1, no. 2, pp. 49–57, Oct. 2024, doi: 10.63503/j.ijssic.2024.33.

A Hybrid AI Framework for CV Screening with CNN-Based Layout Classification and Open-Source LLMs

ORIGINALITY REPORT

10%	5%	8%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Proceedings of IEMTRONICS 2025", Springer Science and Business Media LLC, 2026 Publication	1%
2	Thangaprakash Sengodan, Sanjay Misra, M Murugappan. "Advances in Electrical and Computer Technologies", CRC Press, 2025 Publication	1%
3	1library.net Internet Source	1%
4	Sanjaya Samuel Ady, Suyoto. "The transformation from Pandemic to Endemic of Covid-19: Spatio-temporal Analysis of Citizen Mobility in Asia Countries", 2022 13th International Conference on Information and Communication Technology Convergence (ICTC), 2022 Publication	1%
5	Arvind Dagur, Karan Singh, Pawan Singh Mehra, Dhirendra Kumar Shukla. "Artificial Intelligence, Blockchain, Computing and Security", CRC Press, 2023 Publication	1%
6	Submitted to University of Lincoln Student Paper	1%
7	Vankamamidi S. Naresh, Ayyappa D. "PPDNN-CRP: privacy-preserving deep neural network processing for credit risk prediction in cloud:	1%

	a homomorphic encryption-based approach", Journal of Cloud Computing, 2024 Publication	
8	"Recent Trends in AI Enabled Technologies", Springer Science and Business Media LLC, 2026 Publication	<1 %
9	Submitted to Loyola University, Chicago Student Paper	<1 %
10	jurnal.polibatam.ac.id Internet Source	<1 %
11	Fabeyo, Stephen. "Ensuring ADA Compliance in AI Hiring: Review, Analysis, and Disability Discrimination Framework.", Middle Georgia State University Publication	<1 %
12	eudoxuspress.com Internet Source	<1 %
13	www.mdpi.com Internet Source	<1 %
14	"2024 Real-Time Intelligent Systems", Springer Science and Business Media LLC, 2025 Publication	<1 %
15	dspace.univ-guelma.dz Internet Source	<1 %
16	wrap.warwick.ac.uk Internet Source	<1 %
17	www.aaem.pl Internet Source	<1 %
18	www.frontiersin.org Internet Source	<1 %
19	www.joig.net Internet Source	<1 %

20	Abdulrazak H. Almaliki, Afaq Khattak. "Short- and long-term tidal level forecasting: A novel hybrid TCN+LSTM framework", Journal of Sea Research, 2025 Publication	<1 %

Exclude quotes	On	Exclude matches	< 7 words
Exclude bibliography	On		