International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

AI-based Multimodal Resume Ranking Web Application for Large Scale Job Recruitment

Dr. Abdul Rasool MD[1], Mr. Mohammad Zaid Khan[2], Mr. Mudassir Yameen[3], Mr. Ibrahim Mohammed [4], Mr. Mohammed Ibrahim [5]

[1]Associate Professor, Dept. of CSE-AIML, Lords Institute of Engineering and Technology

[2,3,4,5]B.E. Students, Dept. of CSE-AIML, Lords Institute of Engineering and Technology

ABSTRACT

The exponential surge in online job applications has placed immense strain on Human Resources (HR) departments worldwide, rendering traditional resume screening methods inefficient and error-prone. This paper presents a novel AI-based multimodal resume ranking web application designed to automate and enhance large-scale recruitment processes. The proposed system integrates a suite of state-of-the-art deep learning models: YOLOv9 fine-tuned on a custom-annotated dataset for resume segment detection, EasyOCR for robust multilingual text recognition, a fine-tuned multilingual BERT (mBERT) model for structured text classification, and GLiNER for zero-shot named entity recognition supplemented by regular expressions. A hybrid matching pipeline combining dense vector cosine similarity — leveraging the gte-large-en-v1.5 embedding model — with BM25 keyword relevance scoring is employed to rank candidates against job descriptions with high precision.

A custom dataset of 2,751 diverse resume templates was assembled and annotated using Roboflow for YOLOv9 training. Among evaluated object detection models including DETR and Detectron2, YOLOv9 achieved the highest mean average precision (mAP) of 0.84. The text classification model attained an F1-score of 0.9652 and accuracy of 0.9621. The web application supports batch upload of up to 200 resumes, configurable weighting of resume attributes, and real-time ranked output for HR professionals. Experimental results demonstrate that the multimodal architecture achieves strong end-to-end performance, though challenges remain in OCR accuracy under diverse fonts and zero-shot NER confidence. The proposed system significantly reduces manual screening effort and improves candidate-job matching accuracy, establishing a practical AI-driven solution for modern recruitment workflows.

Keywords *— Resume ranking, Deep learning, YOLOv9, Named Entity Recognition, Optical Character Recognition, mBERT, Cosine similarity, BM25, Large-scale recruitment, NLP, Web application, Multimodal AI.*

I. INTRODUCTION

The rapid proliferation of digital recruitment platforms and remote work has fundamentally transformed the hiring landscape, resulting in an unprecedented surge in job applications. Human Resources (HR) departments are increasingly challenged to efficiently process and evaluate large volumes of resumes while ensuring fairness, consistency, and accuracy in candidate selection. Traditional manual screening — wherein recruiters individually review resumes — is no longer viable at scale [1][2].

Manual screening methods exhibit several inherent limitations. Recruiters can evaluate only a limited number of resumes per day, introducing delays in the hiring process. Additionally, cognitive biases, including affinity and confirmation bias, contribute to inconsistent assessments, potentially leading to the rejection of qualified candidates or the selection of unsuitable applicants. The heterogeneity of resume formats — such as PDF, DOCX, and scanned documents — further complicates standardized information extraction and comparison due to variations in layout, structure, and language [3].

Automated resume processing systems provide a scalable and efficient alternative. Early approaches based on keyword matching and rule-based parsing lacked semantic understanding and adaptability. However, recent advances in deep learning — particularly transformer-based language models and convolutional architectures — enable sophisticated analysis, including document layout detection, contextual text classification, and semantic similarity-based matching [4].

This paper proposes an AI-based multimodal resume ranking web application that integrates state-of-the-art techniques into a unified pipeline. The system utilizes YOLOv9 for visual segmentation, EasyOCR for text extraction, mBERT for classification, and GLiNER for named entity recognition, alongside a hybrid ranking mechanism combining cosine similarity and BM25 scoring.

This paper makes three primary contributions: (1) a specialized, geographically diverse dataset of Indian traffic signs with comprehensive annotation and augmentation; (2) adaptation of the YOLOv9 architecture for the specific characteristics of Indian traffic environments; and (3) integration of the recognition system with smart traffic management and ADAS platforms. Extensive experiments demonstrate that the proposed YOLOv9-based approach achieves state-of-the-art performance in detecting and recognizing Indian traffic signs across complex urban settings.

II. LITERATURE SURVEY

Automated resume processing has attracted sustained research attention across computer vision, natural language processing, and information retrieval communities. Early resume parsing systems relied primarily on rule-based extraction using hand-crafted patterns and regular expressions. While computationally inexpensive, such approaches proved brittle in practice — failing to generalize across the diverse structural conventions adopted in resumes from different industries, geographies, and time periods.

The introduction of Optical Character Recognition (OCR) and Natural Language Processing (NLP) substantially extended the capabilities of automated resume analysis systems. Palshikar et al. [4] developed RINX, a comprehensive system for information and knowledge extraction from resumes using NLP techniques, demonstrating the effectiveness of structured extraction over raw text parsing. Komanduri et al. [5] utilized the Tesseract OCR engine combined with adaptive binarization and contour detection for text identification across diverse document formats, achieving competitive recognition accuracy. Subramanian et al. [6] further combined YOLO-based character localization with Super Resolution preprocessing, achieving high word accuracy on document image datasets.

Machine learning approaches progressively replaced rule-based methods for resume classification. Alamelu et al. [7] developed an NLP-based system to compare extracted resume data with job descriptions, significantly reducing screening time for HR professionals. Roy et al. [8] applied n-gram text classification with machine learning models, achieving 78.53% accuracy in automated resume recommendation. Kinge et al. [9] evaluated multiple machine learning configurations on resume screening tasks, reporting accuracy figures of 78% and 98% for different classification targets. Nisha B. et al. [10] constructed an Automated Resume Parsing and Ranking System (ARRS) using NLP techniques for essential information extraction and candidate ranking.

Deep learning methods subsequently achieved substantial performance gains across multiple resume processing sub-tasks. Mhatre et al. [17] reported up to 99.48% accuracy in resume screening and ranking using convolutional neural networks (CNN) and long short-term memory (LSTM) architectures, demonstrating the representational power of deep sequence models for resume text. Jayakumar et al. [14] applied Random Forest classifiers to the hiring recommendation problem, achieving 92.9% accuracy. Mohanty et al. [15] developed Resumate — an NLP-based resume parsing prototype using XGBoost — attaining 96% training accuracy on skill classification tasks.

Transformer-based language models have most recently emerged as the dominant paradigm for resume NLP tasks. Mukherjee [13] demonstrated the effectiveness of DistilBERT and XLM for resume ranking and shortlisting, outperforming conventional machine learning baselines on multilingual datasets. Mehboob et al. [19] created an automated CV shortlisting tool achieving 99% job-matching accuracy. Alderham and Jaha enhanced candidate-career matching by combining machine learning and NLP, achieving up to 92% accuracy across diverse job categories. Chandak et al. [18] integrated a resume parser with a job recommendation engine using skill extraction and semantic job matching.

The application of object detection models to resume layout analysis represents a more recent development. Tanberk et al. utilized OCR combined with NLP models including DistilBERT and YOLOv8 for resume matching and ranking, demonstrating that visual layout detection prior to text extraction substantially improves downstream processing accuracy. Gunaseelan et al. [16] employed machine learning techniques including XGBoost for automatic segment extraction from resumes, confirming the value of structured visual parsing. Despite these advances, existing systems typically address only isolated components of the resume processing pipeline. The present work distinguishes itself through end-to-end multimodal integration combining YOLOv9 visual detection, OCR, transformer-based classification, zero-shot NER, and hybrid semantic matching within a unified production web application.

III. PROPOSED WORK

This section presents the proposed methodology for the AI-based multimodal resume ranking system. The proposed work is organized into four primary components: system architecture and dataset construction, model implementation and training, evaluation and testing, and web application integration.

A. System Architecture and Dataset Construction

The proposed resume ranking system employs a multi-tiered architecture integrating deep learning models with classical information retrieval techniques. The system processes resumes through three principal pipelines: resume information extraction, job description processing, and hybrid candidate matching.

1) Resume Information Extraction:

HR professionals upload resumes in PDF, DOCX, or image formats. Uploaded files are converted to JPG format for processing by the YOLOv9 object detection model, which detects and segments text regions within the resume layout. Text content is subsequently extracted from detected regions using EasyOCR. A fine-tuned mBERT model then classifies extracted text segments into 12 predefined

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

categories: awards, certificates, contact/name/title, education, interests, languages, paragraphs, professional experiences, projects, skills, soft skills, and summary. The GLiNER zero-shot NER model further extracts structured entities — including locations, persons, job titles, dates, and organizations — from classified segments, supplemented by regular expressions for standardized fields such as phone numbers, email addresses, and URLs. All processed information is stored in a structured database for subsequent matching .

2) Job Description Processing:

Recruitment professionals define job descriptions directly within the web application interface. The system supports specification of job title, company, location, job type, required skills, and detailed job description text. Users configure attribute weights for Skills, Experience, Education, Miscellaneous content, and Keywords, reflecting their specific hiring priorities. Job description fields are combined into a single textual representation for downstream embedding.

3) Hybrid Matching and Ranking:

The system employs a hybrid matching pipeline combining dense vector cosine similarity with BM25 keyword relevance scoring. Cosine similarity is computed between the job description embedding and each resume category embedding using the gte-large-en-v1.5 model. BM25 matching focuses on location and language keywords extracted by the NER component. An overall matching score S is computed by combining user-weighted cosine similarity scores with the keyword matching score, producing a ranked candidate list ordered by relevance to the job description .

B. Dataset Construction and Preprocessing

A comprehensive dataset of 2,751 diverse resume templates was assembled from multiple online sources including royalty-free resume websites, professional networking platforms, open-source resume builders, and academic institution career centres. This collection encompasses diverse styles, formats, content structures, and languages. Dataset annotation for object detection was conducted using Roboflow, with a single class label — 'segment' — denoting all text regions within each resume image.

The dataset was partitioned into training (75%, comprising 3,220 images), validation (19%, comprising 819 images), and test (6%, comprising 265 images) subsets following established practice. Preprocessing standardized all images to 640×640 pixels with automatic orientation adjustment. Augmentation strategies applied through Roboflow included grayscale conversion (25% of images), saturation adjustment (±25°), blur (up to 1 pixel), and noise injection (up to 5.01% of pixels), generating two augmented variants per original image to enhance model generalization across varied resume designs and print qualities.

C. Model Implementation and Training

Three object detection architectures were evaluated for resume segment detection: YOLOv9, DETR, and Detectron2. Training parameters for each model followed guidance from their respective research publications. YOLOv9's GELAN architecture with Programmable Gradient Information (PGI) achieved the highest mAP and fastest inference time, confirming its suitability as the primary detection model for the system . The fine-tuned mBERT model for text classification was sourced from the Hugging Face model repository and fine-tuned on a labeled resume text dataset. GLiNER was deployed in zero-shot mode with dynamically defined entity labels based on classified segment categories, without additional fine-tuning on resume-specific data.

For the hybrid matching component, the gte-large-en-v1.5 embedding model was selected following evaluation on the MTEB Benchmark for its superior performance on long-text sequences and optimal balance between embedding quality and model size. BM25 keyword matching was applied specifically to location and language fields extracted by the NER component. The overall matching score S is computed as: $S = Ws \cdot Cs + We \cdot Ce + Wed \cdot Ced + Wm \cdot Cm + Wk \cdot K$, where Cs, Ce, Ced, Cm are cosine similarity scores for skills, experience, education, and miscellaneous categories respectively, K is the BM25 keyword score, and Ws, We, Wed, Wm, Wk are user-defined weights.

D. Web Application Development

The web application provides an intuitive interface for HR professionals to upload up to 200 resumes simultaneously in PDF, DOCX, or image formats. Users define job descriptions including job title, company details, location, job type, required skills, and attribute importance weights. The application displays ranked resume results with detailed per-candidate score breakdowns across skill, experience, education, and keyword dimensions, enabling HR professionals to transparently review how each resume aligns with the defined job criteria. The interface supports real-time model output visualization and adjustable matching parameters for iterative candidate shortlisting.

IV. RESULTS AND DISCUSSION

A. Dataset Analysis

The final annotated dataset comprises 2,694 original resume images, extending to 4,304 images after augmentation. The dataset features thorough segment-level annotations totaling 19,111 bounding box annotations, with an average of approximately 7.1 annotations per image. Annotations are distributed across training (10,898), validation (6,238), and test (1,975) subsets, ensuring a comprehensive and balanced evaluation framework. The distribution of object count per image is broadly concentrated in the 4–9 segment range, reflecting the typical structural complexity of professional resumes

International Journal of Information Technology & Computer Engineering

ISSN 2347–3657

Volume 14, Issue 2, 2026

with multiple distinct text sections. The diversity of resume layouts, fonts, and languages in the dataset presents a realistic and challenging benchmark for segment detection model training.

B. Object Detection Model Comparison

Table I presents a comparative evaluation of the three object detection models — YOLOv9, DETR, and Detectron2 — trained and evaluated on the custom resume dataset. Models were assessed using Average Precision (AP) at IoU thresholds of 0.50:0.95, 0.50, and 0.75, and Average Recall (AR) at IoU thresholds of 0.50:0.95 for up to 100 detections. Inference times were measured on an L4 GPU using Google Colab.

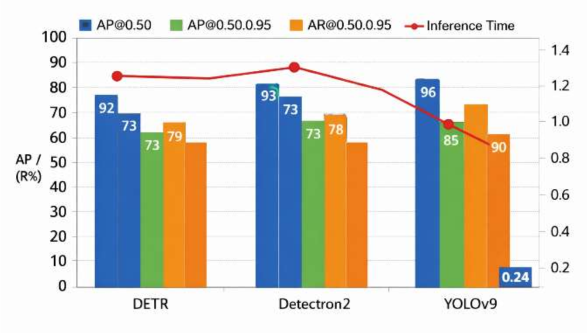

FIG. 1. Resume Segment Detection Model Performance Comparison

The results confirm that YOLOv9 achieves the highest detection precision at both relaxed and standard IoU thresholds, with AP of 0.965 at IoU=0.50 and 0.850 at IoU=0.50:0.95, outperforming both DETR and Detectron2. Critically, YOLOv9's inference time of 0.24 seconds per image is approximately 4× faster than DETR (1.05s) and 5× faster than Detectron2 (1.17s), making it highly suitable for real-time processing of large resume batches. These results are consistent with the broader literature demonstrating YOLOv9's advantages in document layout analysis tasks .

C. Text Classification and NER Results

Table II presents the performance metrics for the fine-tuned mBERT text classification model evaluated on an unseen resume dataset. The model demonstrates high accuracy and reliability across all predefined classification categories.

TABLE I. Text Classification Model Performance Metrics

Metric	Value
Loss	0.0369
F1 Score	0.9652
ROC AUC	0.9808
Accuracy	0.9621

The classification model achieves an F1 score of 0.9652 and accuracy of 0.9621, indicating high reliability in categorizing resume text segments into classes including education, professional experience, skills, and contact information. The strong ROC AUC of 0.9808 confirms robust discriminative performance across all 12 classification categories. The NER component implemented using GLiNER in zero-shot mode produces averagely good entity extraction results, though with lower confidence scores compared to fine-tuned NER models — a characteristic attributable to the zero-shot deployment without domain-specific training data. OCR accuracy using EasyOCR was generally high across standard fonts and layouts, with identified challenges including character-level misinterpretations (e.g., 'O' vs '0', 'l' vs '1') and spacing errors in stylized resume headers.

D. Hybrid Matching Results

Table III illustrates example hybrid matching scores for a representative recruitment scenario, demonstrating the system's flexible weighted scoring across resume attributes. Resumes are ranked in descending order of overall matching score S, computed from user-configured attribute weights for Skills (0.2), Experience (0.4), Education (0.1), Miscellaneous (0.2), and Keywords (0.1).

TABLE II. Hybrid Matching Scores for an Example Scenario

ID	S	E	Ed	M	K	Final
4	0.14	0.296	0.07	0.146	0	0.65
5	0.12	0.298	0.047	0.146	0	0.61
3	0.118	0.277	0.00	0.141	0.05	0.59
6	0.148	0.273	0.062	0.108	0	0.59

The results demonstrate that the hybrid matching pipeline effectively differentiates candidates based on multi-attribute relevance to the job description. The experience dimension — assigned the highest weight of 0.4 — contributes most significantly to final scores, reflecting the typical priority of HR professionals in candidate evaluation. The configurable weighting system enables the application to adapt to diverse recruitment contexts with minimal reconfiguration, providing a flexible and practical tool for HR workflows.

V. CONCLUSION

This paper has presented a comprehensive AI-based multimodal resume ranking web application for large-scale job recruitment. The proposed system integrates four principal deep learning components — YOLOv9 for visual segment detection, EasyOCR for text recognition, fine-tuned mBERT for text classification, and GLiNER for zero-shot named entity recognition — within a unified end-to-end pipeline. A hybrid matching mechanism combining dense vector cosine similarity with BM25 keyword relevance produces ranked candidate lists aligned with configurable job description criteria and user-defined attribute weights.

Experimental results confirm that YOLOv9 achieves the highest object detection performance among evaluated models, with an mAP of 0.85 at IoU=0.50:0.95 and an inference time of 0.24 seconds per resume —

approximately 4–5× faster than transformer-based alternatives. The mBERT text classification model achieves an F1 score of 0.9652 and accuracy of 0.9621 across 12 resume section categories. The hybrid matching framework with configurable attribute weighting provides a practical and flexible solution for aligning job descriptions with candidate profiles across diverse recruitment contexts.

Key challenges identified include OCR character misinterpretation under diverse font styles and layouts, lower confidence scores in zero-shot NER outputs, and limitations in capturing full semantic meaning with dense vector similarity alone. Future work will focus on developing advanced post-processing correction mechanisms for OCR output, constructing a domain-specific NER training dataset for resume entities, incorporating hybrid sparse-dense vector retrieval techniques for improved semantic matching, and expanding the system to support additional languages and document formats. Regular dataset updates and model fine-tuning will further enhance system robustness and accuracy across the evolving diversity of real-world resume formats. The proposed system represents a significant step towards fully automated, AI-driven recruitment workflows that benefit both HR professionals and job seekers.

REFERENCES

[1] E. Baykal, "Digital Era and New Methods for Employee Recruitment," in Digital Transformation in the Age of Digitalisation, IGI Global, 2020. doi: 10.4018/978-1-7998-1125-1.ch018.

[2] S. Solanki and Dr. Gujarati, "The Digital Revolution in Recruitment: Unraveling the Impact and Challenges of E-Recruitment," Educational Administration Theory and Practices, vol. 30, Jun. 2024, doi: 10.53555/kuey.v30i6(S).5362.

[3] S. D. Rozario, S. Venkatraman, and A. Abbas, "Challenges in recruitment and selection process: An empirical study," Challenges, vol. 10, no. 2, p. 35, 2019.

[4] G. K. Palshikar et al., "RINX: A system for information and knowledge extraction from resumes," Data and Knowledge Engineering, vol. 147, p. 102202, 2023.

[5] S. Komanduri, Y. M. Roopa, and M. M. Bala, "Novel approach for image text recognition and translation," in Proc. 3rd Int. Conf. Computing Methodologies and Communication (ICCMC), 2019, pp. 596–599.

[6] S. Subramanian et al., "TEYSuR-Text Extraction with YOLO and Super Resolution," in Proc. Int. Conf. Advancement in Technology (ICONAT), 2022, pp. 1–7.

[7] M. Alamelu et al., "Resume Validation and Filtration using Natural Language Processing," in Proc. 10th Int. Conf. Internet of Everything, Microwave Engineering, Communication and Networks (IEMECON), 2021, pp. 1–5.

[8] P. K. Roy, S. S. Chowdhary, and R. Bhatia, "A Machine Learning approach for automation of Resume Recommendation system," Procedia Computer Science, vol. 167, pp. 2318–2327, 2020.

[9] B. Kinge et al., "Resume Screening using Machine Learning and NLP: A proposed system," Int. J. Scientific Research in Computer Science, Engineering and Information Technology, vol. 8, no. 2, pp. 253–258, 2022.

[10] B. Nisha et al., "HR Tech Analyst: Automated Resume Parsing and Ranking System through Natural Language Processing," in Proc. 2nd Int. Conf. Automation, Computing and Renewable Systems (ICACRS), 2023, pp. 1681–1686.

[11] S. Amin et al., "Web application for screening resume," in Proc. Int. Conf. Nascent Technologies in Engineering (ICNTE), 2019, pp. 1–7.

[12] H. Artajaya et al., "Job Recommendation System based on Resume using Natural Language Processing and Distance-based Algorithm," in Proc. IEEE Int. Conf. Artificial Intelligence and Mechatronics Systems (AIMS), 2024, pp. 1–6.

[13] A. Mukherjee, "Resume Ranking and Shortlisting with DistilBERT and XLM," in Proc. IEEE Int. Conf. for Women in Innovation, Technology & Entrepreneurship (ICWITE), 2024, pp. 301–304.

[14] N. Jayakumar et al., "On-Demand Job-Based Recruitment for Organisations Using Artificial Intelligence," in Proc. Int. Conf. Networking and Communications (ICNWC), 2023, pp. 1–6.

[15] S. Mohanty et al., "Resumate: A prototype to enhance recruitment process with NLP based resume parsing," in Proc. 4th Int. Conf. Intelligent Engineering and Management (ICIEM), 2023, pp. 1–6.

[16] B. Gunaseelan, S. Mandal, and V. Rajagopalan, "Automatic extraction of segments from resumes using machine learning," in Proc. IEEE 17th India Council Int. Conf. (INDICON), 2020, pp. 1–6.

[17] S. Mhatre et al., "Resume Screening and Ranking using Convolutional Neural Network," in Proc. Int. Conf. Sustainable Computing and Smart Systems (ICSCSS), 2023, pp. 412–419.

[18] A. V. Chandak et al., "Resume Parser and Job Recommendation System using Machine Learning," in Proc. Int. Conf. Emerging Systems and Intelligent Computing (ESIC), 2024, pp. 157–162.

[19] M. Mehboob et al., "Evaluating automatic CV shortlisting tool for job recruitment based on machine learning techniques," in Proc. Mohammad Ali Jinnah Univ. Int. Conf. Computing (MAJICC), 2022, pp. 1–4.

[20] T. G. Sougandh, N. S. Reddy, and M. Belwal, "Automated Resume Parsing: A Natural Language Processing Approach," in Proc. 7th Int. Conf. Computation System and Information Technology for Sustainable Solutions (CSITSS), 2023, pp. 1–6.