Seventh International Conference in Information and Communication Technologies, Lusaka, Zambia 15th to 16th October 2025

A Modular AI-Driven Framework for Automating HR Case Processing: OCR, NLP, Sentiment Analysis and Imbalanced Learning

Chitalu Ilunga
School of Computing, Technology, & Applied Science
Department of Computer Science
ZCAS *University*
Lusaka, Zambia
chitaluc3@gmail.com

Aaron Zimba
School of Computing, Technology, & Applied Science
Department of Computer Science
ZCAS University
Lusaka, Zambia
aaron.zimba@zcasu.edu.zm

***Abstract*—This study investigates the feasibility and effectiveness of integrating Optical Character Recognition (OCR), Regular Expressions (RegEx), Aspect-Based Sentiment Analysis (ABSA), and supervised learning to automate structured and semi-structured Human Resource (HR) case processing in African public-sector contexts. The proposed modular pipeline comprises OpenCV-based pre-processing for noise reduction, skew correction, and ROI detection; tuned Tesseract 4.1 OCR; RegEx-driven attribute extraction; ABSA for narrative sentiment; and Balanced Bagging Random Forest classification with SMOTE-ENN applied to enhance minority-class representation and improve sensitivity to rare but decision-critical HR outcomes. Evaluated on anonymised promotion and transfer cases from Zambia's Teaching Service Commission, the system achieved 95% OCR accuracy, F2-Score of 0.97, weighted F1 = 0.95, PR AUC = 0.98, and minority-class F1 = 0.36, demonstrating improved detection of low-frequency, high-priority HR cases. End-to-end processing reduced manual timelines from ~3 days to <9.4 s per two-page case batch, scaling to 1,312 records in 3.3 s on commodity hardware, enabling timely decision support in resource-constrained environments. The framework's domain-specific integration and ethical alignment provide a scalable, adaptable solution for HR digitisation in policy-bound, resource-limited sectors, supporting compliance with governance and transparency requirements.**

Keywords—Imbalanced Learning, Document Digitisation, AI-Driven Decision Support Systems, Human Resource Automation, Aspect-Based Sentiment Analysis, Optical Character Recognition (OCR).

I. INTRODUCTION

This study addresses the automation of Human Resource (HR) case processing in governmental and bureaucratic contexts reliant on paper-based workflows. The developed framework converts multi-case forms into structured records, enabling precise digitisation rather than raw text capture. Integrating Optical Character Recognition (OCR) using Tesseract, Natural Language Processing (NLP) with Regular Expressions (RegEx), lexicon-based Aspect-Based Sentiment Analysis (ABSA), and Artificial Intelligence (AI) classification, the pipeline mitigates bottlenecks of inefficiency and inaccuracy while delivering scalable decision support. Structuring data at the point of digitisation ensures consistent input for the machine learning model and extensibility for wider organisational applications. Artificial Intelligence (AI) has emerged as a transformative force across industries, recognised for its ability to automate routine tasks, support data-driven decision-making, and enhance operational efficiency [14]. Within Human Resource Management (HRM), this capacity makes AI particularly valuable for document-intensive processes such as case evaluation, recruitment, and performance management [16]. Yet while adoption is advancing globally, its application to public-sector HR case processing in Africa remains largely untapped, leaving critical inefficiencies in paper-based workflows unaddressed. This gap underscores the contribution of the present work, which demonstrates how AI can be contextualised to augment bureaucratic HR processes in low-resource, paper-reliant environments.

Physical documents are circulated across offices and hierarchical approval stages before decisions are reached. This manual routing introduces substantial delays, increases susceptibility to transcription errors, and limits the capacity for timely, data-driven decision support. Security and compliance remain critical, particularly when handling sensitive HR data. To address this, anonymisation is applied in line with legislation on data privacy, with ethical safeguards grounded in AI-specific HR frameworks [15], complemented by broader considerations of technology adoption and AI impact. Ethical concerns are considered as well through transparency in algorithmic decision-making to prevent bias and preserve fairness [16]. Despite the efficiency gains achievable with AI and OCR, human oversight remains essential in complex or ambiguous cases where contextual judgment is indispensable [13].

HR case processing spans promotions, transfers, retirements, and other personnel functions where timeliness and accuracy are vital to both employees' careers and institutional integrity. Delays or errors not only stall progression but also expose institutions to disputes or litigation, with employees experiencing the consequences directly. Yet in government contexts, paper-based methods often cause deterioration, misplacement, duplication, and backlogs that compound inefficiency. This study focused on promotion and transfer cases for testing, but the framework generalises through RegEx tuning, allowing adaptation to other HR case types such as retirements and disciplinary records. Crucially, AI augments rather than replaces human oversight, delivering efficiency gains while preserving contextual judgement.

To guide this investigation, the following research questions (RQs) are posed:

RQ1. Can OCR, RegEx, ABSA, and supervised learning be effectively integrated into a modular pipeline for automating HR case processing?

RQ2. How accurately and efficiently can the proposed system process HR cases, particularly with respect to minority-class detection in imbalanced data?

RQ3. To what extent does the framework reduce processing time and scale effectively compared to manual or spreadsheet-based methods?

RQ4. Can the framework meet ethical, fairness, and compliance requirements necessary for deployment in public-sector HR systems?

RQ5. How adaptable is the framework to other HR case types and future advances in NLP/ML methods?

The contributions of this work are threefold: (1) academically, it advances research through a novel integration of OCR, a custom lexicon-based ABSA module, and imbalanced-learning strategies into a unified pipeline, validated with a Random Forest model enhanced by Balanced Bagging and SMOTEENN to achieve both high recall and improved minority-class sensitivity; (2) practically, it delivers a modular and deployable framework that reduces case processing time from days to seconds without reliance on high-end resources, while addressing challenges of backlog and document deterioration; and (3) in ICT4D, it offers a scalable, resource-conscious model for digitisation and decision support in African public-sector institutions where paper-based records remain the norm.

II. RELATED WORKS

• *AI in Human Resource Management*

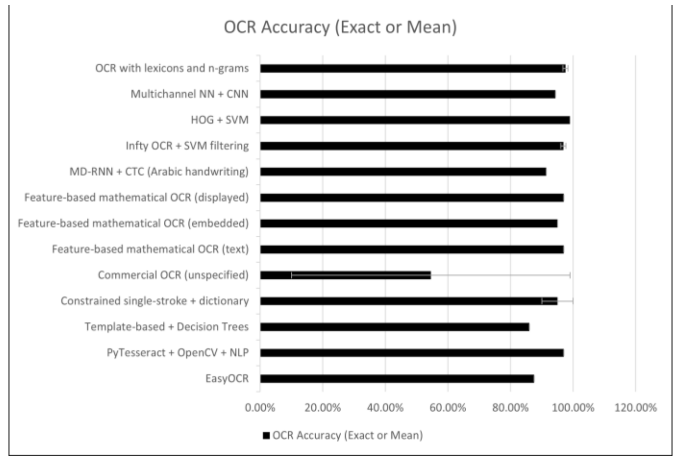
Artificial Intelligence (AI) is increasingly applied to Human Resource Management (HRM) for recruitment, performance evaluation, and employee engagement. These applications highlight AI's ability to automate labour-intensive processes and improve decision support [16], [14]. However, document-level HR case processing in public-sector environments remains underexplored, especially in Africa where reliance on paper-based systems persists. This gap motivates the present work, which situates itself at the intersection of HR automation and AI-driven document digitisation.

• *Optical Character Recognition*

OCR remains a cornerstone of digitisation. Prior studies demonstrate that recognition accuracy is shaped by preprocessing methods, document quality, and language support. A common formulation of OCR accuracy is given by:

$$\text{Accuracy} = \frac{\text{Correctly Recognised Characters}}{\text{Total Characters in Ground Truth}} \times 100\%$$

Open-source tools such as Tesseract have achieved results comparable to proprietary systems when coupled with techniques like noise reduction, skew correction, and ROI detection [4], [7]. These findings justify the adoption of Tesseract 4.1 in this work, tuned for HR-specific documents. Figure 1 provides a comparative view of OCR model accuracies, while Table 1 lists exact performance values and error ranges for reference.



[1]. OCR accuracy comparison with error margins

• OCR MODEL ACCURACY COMPARISON

OCR Model / Method	Accuracy (midpoint ± error)
EasyOCR	87.5%
PyTesseract + OpenCV + NLP	97%
Template-based + Decision Trees	86%
Constrained input + dictionary aid	95% ± 5%
Commercial OCR (general vs. math content)	99% (general), 10% (math)
Feature-based mathematical OCR	96% ± 1%
MD-RNN + CTC (Arabic handwriting)	91.4%
Infty OCR + SVM filtering	96.9% ± 0.8%
HOG + SVM	99%
Multichannel NN + CNN	94.4%
OCR with lexicons and n-grams	97.5% ± 0.8%

• *Aspect-Based Sentiment Analysis*

Sentiment analysis has traditionally been applied to domains like consumer reviews and social media. Aspect-Based Sentiment Analysis (ABSA) in particular, provides attribute-linked opinions, enabling transparency and interpretability [5], [9]. A common formulation of ABSA defines sentiment toward an aspect a in a document d as:

$$\text{Sentiment Score}(d, a) = \sum_{w \in d_a} \text{lexicon}(w)$$

where d_a are the words in document d that are linked to aspect a, and lexicon(w) returns sentiment polarity value of word w.

Yet HR-focused ABSA remains underexplored, leaving a gap in narrative-driven contexts such as promotion or grievance cases. This study adapts a lexicon-based ABSA approach to personnel documents, extending interpretability into HR decision support.

- *Ethical and Fairness Considerations*

AI deployment in HR must account for fairness, compliance, and transparency. Prior studies emphasise that algorithmic decision-making risks perpetuating bias unless interpretability and safeguards are built in [15], [16]. In public-sector contexts, these risks are heightened by sensitive data and high-stakes decisions. To mitigate them, the proposed framework incorporates anonymisation aligned with data privacy legislation, as well as ethical safeguards grounded in responsible AI for HR [15].

- *Imbalanced Learning for Machine Learning Models*

Imbalanced datasets are a prevalent challenge in machine learning, where classifiers often prioritise the majority class at the expense of the minority class of greatest interest [1], [18]. Two main strategies have been developed: cost-sensitive learning and sampling-based methods, with the latter receiving extensive attention [1]. Sampling approaches include undersampling, oversampling, and hybrids. The Synthetic Minority Oversampling Technique (SMOTE) has gained prominence for generating synthetic minority samples, proving effective across domains such as fraud detection and medical diagnosis [18]. However, SMOTE can introduce class overlap and noise; hybrid methods such as SMOTE-ENN mitigate these issues by combining oversampling with Edited Nearest Neighbours to simultaneously balance classes and filter noisy data, resulting in more stable and generalisable models [18].

In parallel, ensemble methods provide robust alternatives for learning from skewed distributions. Random Forests, for example, leverage multiple bootstrap trees with random feature selection to deliver resilient classification performance [1], [12]. Balanced Bagging integrates oversampling and undersampling within the bootstrapping process, offering competitive performance with reduced computational overhead [12]. More advanced refinements, such as unevenly balanced bagging, further improve prediction stability on minority classes [3]. Collectively, these approaches mark a progression from basic resampling to hybrid and ensemble strategies for effectively handling class imbalance.

- *Human-AI Collaboration*

The literature consistently stresses that human oversight remains indispensable for ambiguous or sensitive cases. While AI systems accelerate routine case processing, human judgment ensures contextual sensitivity and accountability [13]. This aligns with a hybrid decision-support paradigm in which automation augments, rather than replaces, HR officers.

- *ICT4D and Responsible AI in African Contexts*

Governments in Sub-Saharan Africa score among the lowest in AI readiness, with few national strategies in place, and wholesale adoption of Global North models risks suboptimal outcomes [11]. Instead, tailored frameworks and responsible AI approaches; transparent, ethical, and aligned with local contexts are critical to avoiding entrenched divides [11]. Financial and institutional constraints also shape outcomes, with weak infrastructure, limited capacity, and fiscal trade-offs challenging sustainable investment [17]. Frimpong stresses that sequencing, capacity building, and ethical governance must guide deployment through a “digital social compact” that prioritises equity [17]. In parallel, AI deployments across Africa highlight diverse applications, notably in agriculture where AI enhances productivity without displacing labour. Transferring this principle to empowered public offices suggests that targeted AI, designed to augment rather than replace, could drive positive and responsible digital transformation in the HR sector.

III. METHODOLOGY

- *Research Design*

This study follows a Design Science Research Methodology (DSRM) [2] to iteratively develop and validate a modular AI framework for HR case processing. The problem domain concerns delays and inefficiencies in manual, paper-based workflows; thus, the proposed solution integrates OCR, RegEx-driven extraction, aspect-based sentiment analysis, and AI-based classification [10]. We employ a lexicon-based ABSA module to maintain auditability and feature-level transparency, while recognising that transformer-based ABSA can be substituted when resources permit [8], [9].

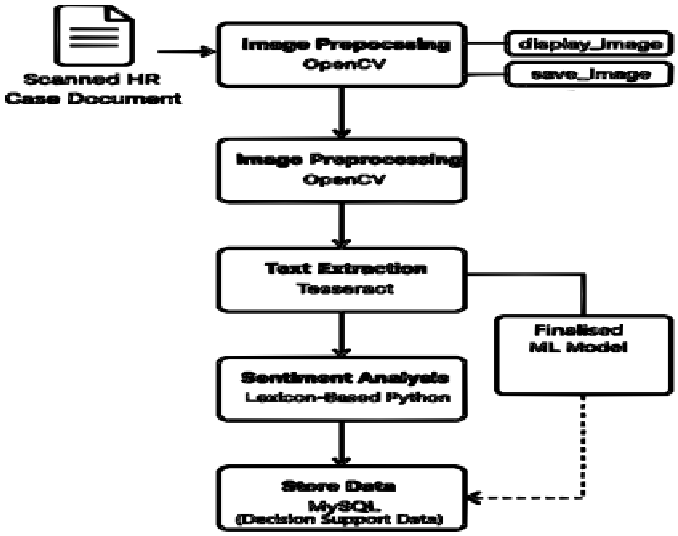
- *Data Sources and Pre-Processing*

Two primary data sources were utilised:

- Institution-provided cases from the Teaching Service Commission of Zambia, comprising scanned promotion and transfer documents with associated decision outcomes.
- Public datasets (Kaggle, GitHub) on employee promotion and performance [19], [20], used for baseline modelling.

Data underwent image-level pre-processing (grayscale conversion, noise removal, skew correction, ROI detection) and tabular cleaning (anonymisation, standardisation, gap resolution). This ensured high OCR accuracy and reliable downstream model training.

- *System Architecture*



[2]. Framework Architecture

The framework (Fig. 2) is modular:

- Image Processing: OpenCV pipeline for enhancement and ROI detection.
- OCR: Tesseract (via PyTesseract) for character extraction.
- Text Analytics: RegEx for attribute identification, lexicon-based ABSA for sentiment signals, adopted for interpretability in HR contexts; transformer-based ABSA remains a drop-in alternative [8], [9].
- Database Layer: MySQL for structured storage.
- AI Classification: Comparative training of Logistic Regression, Random Forest, XGBoost, and ensemble models. Final model selected was Balanced Bagging Random Forest with SMOTE-ENN, chosen for sensitivity to minority classes in imbalanced data [18], [12].

- *Evaluation Strategy*

System performance was evaluated across:

- OCR accuracy, using character-level match rates;
- Model performance, using F1/F2, PR AUC, confusion matrices, with emphasis on minority-class F1;
- Efficiency, measured as runtime per case batch on commodity hardware.

- *Ethical Considerations*

Given the sensitive nature of HR data, the framework incorporated anonymisation, compliance with Zambia's Data Protection Act (2021), and human-in-the-loop oversight to avoid automation-driven bias [15], [13]. The system is designed to augment rather than replace HR professionals, preserving human judgement for ambiguous cases.

To support transparency and compliance (RQ5), the pipeline logs predictions with confidence scores, preserves feature-level traces for ABSA/RegEx extractions, and enforces anonymisation, role-based access control, and audit logging. These operational safeguards align with regional responsible-AI guidance [11] and ethical HR decision-making principles [15], while preserving human oversight for ambiguous or contested cases [13].

IV. Experiments/ Setup

- *Development Approach*

The system was developed using an evolutionary prototyping approach, enabling modular construction and iterative refinement of each component. This method aligned with the Design Science Research methodology, ensuring continuous evaluation and scalability of the solution.

- *Framework Modules*

The pipeline was implemented in Python with modular components:

- Image Pre-processing: OpenCV for grayscale conversion, noise removal, skew correction, and border cropping, optimising scanned HR documents for OCR.
- Region of Interest Detection: Thresholding and contour-based bounding boxes to isolate textual blocks.
- Optical Character Recognition (OCR): Tesseract 4.1 via PyTesseract for extracting raw text.
- Text Analytics: Regular Expressions for structured field extraction; lexicon-based Aspect-Based Sentiment Analysis (ABSA) for attribute-linked opinion detection.
- Database Integration: MySQL for structured storage of extracted fields and sentiment values.
- AI Classification: Comparative testing of models (Logistic Regression, Random Forest, XGBoost) with Balanced Bagging Random Forest + SMOTE-ENN selected for minority-class sensitivity.

- *Experimental Protocol*

- Datasets: Two sources were used — (i) institution-provided HR promotion/transfer cases (anonymised), and (ii) public datasets sourced from Kaggle and Github for baseline training.
- Evaluation Metrics: OCR performance was assessed via character-level accuracy. Classification models were evaluated with F1, F2, Precision-Recall AUC, and confusion matrices, with emphasis on minority-class F1.
- Implementation Environment: All modules were developed in Python with OpenCV, scikit-learn, imbalanced-learn, and MySQL integration, executed on commodity hardware to reflect resource-constrained public-sector environments.

V. Results and Discussion

- *System Performance*

Our results align with prior findings that AI improves HR efficiency [16], OCR digitisation reduces backlogs [10], and ABSA supports decision-making [4], [6], [8]. Unlike isolated component studies, this project integrates these into one deployable pipeline.

The HR case processing framework was deployed and tested over three months on a modest laptop (Intel i7, 16GB RAM). To illustrate this workflow, Figure 2 provides an overview of the system pipeline, from OCR to classification. Average runtimes remained efficient: preprocessing at ~1–2s per image, complete OCR/ABSA/DB insertion at ~5.4s per two-page case batch (6-8 cases), and batch processing (1312 records) in ~3.3s. This contrasts sharply with the manual workflow (~3 days per case). OCR accuracy was strong, with 95% recognition for single-column text. To formalise this, the accuracy measure is defined in Equation 1. Regex-based extraction achieved ~80–85% accuracy, with errors mainly from overlapping patterns.

- *Evaluation of AI Models*

The classification pipeline addressed class imbalance using SMOTE-ENN and Balanced Bagging. Table II presents the comparative results across models and imbalance treatments. Performance of the Random Forest (Balanced Bagging + SMOTE-ENN) was highest, reaching F2 = 0.97, PR-AUC = 0.98, Weighted F1 = 0.95. Minority-class detection improved markedly: precision 0.31 and recall 0.44 versus near-zero baselines.

- *Model Evaluation Metrics*

To evaluate model performance on imbalanced data, we use metrics beyond raw accuracy, which can be misleading as majority-class predictions dominate the score. Instead, recall-focused and balance-sensitive metrics are emphasised. The F2-score is prioritised, since it weights recall more than precision, better reflecting the model's ability to capture minority-class cases. Precision–Recall AUC further illustrates precision–recall trade-offs and is more reliable than ROC AUC under imbalance.

The metric evaluation formula are listed below:

- $\text{Precision} = \frac{TP}{TP+FP}$
- $\text{Recall} = \frac{TP}{TP+FN}$
- $F_{\text{1-Score}} = \frac{2\cdot\text{Precision}\cdot\text{Recall}}{\text{Precision}+\text{Recall}}$
- $F_{\text{2-Score}} = \frac{5\cdot\text{Precision}\cdot\text{Recall}}{4\cdot\text{Precision}+\text{Recall}}$
- $\text{PR AUC} = \int_0^1 \text{Precision}(r)\ dr$

where TP, TN, FP, FN denote True Positives, True Negatives, False Positives, and False Negatives respectively. $PR\ AUC$ is expressed as the area under the precision-recall curve as recall r varies.

- *Achieved Metrics*

A comparison table of the achieved metrics of varied configured models is shown below:

- AI MODEL METRIC COMPARISONS

Metric	Baseline LogReg	LogReg	LogReg + SMOTE	Random Forest (Base)	RF + SMOTEENN	RF + Balanced Bagging	RF + Balanced Bagging + SMOTEENN
F2-Score	–	0.81	0.84	0.98	0.86	0.78	**0.97**
PR AUC	–	0.99	0.99	0.98	0.98	0.99	**0.98**
Accuracy	0.93	0.78	0.81	0.95	0.83	0.75	**0.95**
Minority Precision	0.78	0.11	0.13	0.17	0.09	0.11	**0.31**
Minority Recall	0.25	0.78	0.78	0.11	0.44	0.89	**0.44**
Minority F1	0.38	0.19	0.22	0.13	0.15	0.19	**0.36**
Majority Precision	0.94	0.99	0.99	0.97	0.98	0.99	**0.98**
Majority Recall	0.99	0.78	0.81	0.98	0.84	0.74	**0.96**
Majority F1	0.96	0.87	0.89	0.97	0.90	0.85	**0.97**

- *Key Strengths*

The framework demonstrated:

- Accuracy: OCR at 95%, regex >80%, AI model with reliable weighted F1.
- Efficiency: Days-long manual processing reduced to seconds.
- Lightweight design: Runs on ordinary hardware.
- Flexibility: Modular Python pipeline allowing upgrades (e.g., deep-learning OCR or transformer ABSA).

- *Efficiency Gains*

To evaluate efficiency (RQ4), manual and automated case processing times across case formats were compared. Manual handling requires ~3 days per case under current workflows. The proposed OCR + ABSA pipeline processes a two-page HR document containing 6–8 cases in 5.4 s, while structured spreadsheet registers (1,312 entries) complete in 3.3 s on commodity hardware. These represent ~48,000× (manual→OCR/ABSA), ~78,000× (manual→records), and 1.6× (OCR/ABSA→records) speed-ups, underscoring throughput gains sufficient for backlog elimination at departmental scale.

- *Comparison with Related Work*

Our results align with prior findings that AI improves HR efficiency [14], [16], OCR digitisation reduces backlogs [10], and ABSA supports decision-making [8],[5],[4],[6]. Unlike isolated component studies, this project integrates these into one deployable pipeline.

- *Unique Contributions*

Based on the evaluation, the contributions of this work are threefold:

- Academic contribution: Demonstrated a novel integration of OCR, lexicon-based ABSA, and imbalanced-learning strategies into a unified pipeline,

 ISBN: 978-9982-95-500-3

validated with a Random Forest classifier enhanced by Balanced Bagging and SMOTE-ENN.

- Practical contribution: Achieved end-to-end case processing time reductions from multiple days to seconds (see section V-D), while maintaining high OCR accuracy and improving sensitivity to minority classes. This addresses challenges of backlog and deterioration without reliance on high-end computational resources.
- ICT4D contribution: Provided a scalable, resource-conscious digitisation and decision-support model tailored for African public-sector institutions where paper-based workflows remain dominant.

VI. CONCLUSION

This work set out to evaluate whether the integration of OCR, lexicon-based ABSA, and AI could accelerate HR case processing while retaining decision support accuracy. The results show that the system reliably digitises HR case records through automated OCR text extraction, structured by regex, and enriched by sentiment analysis. This reduced processing time from days to seconds, while minimising manual errors, thereby confirming the feasibility of research question RQ1. In addressing RQ2, the study demonstrates that machine learning models tuned for imbalance—via SMOTE, SMOTEENN, and balanced bagging—achieve improved minority-class sensitivity compared to baselines. In particular, Random Forest with SMOTEENN achieved strong recall, while Balanced Bagging stabilised performance across classes, with F-2 score and PR AUC offering more faithful assessments than accuracy. Finally, RQ3 was confirmed through the contextualisation of the framework within ICT4D. The modular design and resource-conscious implementation illustrate that advanced NLP and ML can be adapted for African public-sector environments reliant on paper records. This contribution bridges a recognised digital divide, validating the academic and practical value of the system.

Limitations remain in predefined extraction patterns and narrow case coverage, but these also point to future work in expanding training data, developing adaptive templates, and integrating ML-driven decision support with ABSA interpretability.

In sum, this study offers three contributions: academically, the unification of OCR, ABSA, and imbalanced learning into a deployable pipeline; practically, a demonstrable reduction in HR processing time with increased reliability; and in ICT4D, a scalable model for digitisation in low-resource governmental settings.

ACKNOWLEDGMENT

Thanks are made to the Teaching Service Commission of Zambia, for their data provisioning, utilised for the purposes of conducting this research.

REFERENCES

C. Chen, A. Liaw, and L. Breiman, "Using Random Forest to Learn Imbalanced Data," Dept. Stat., Univ. California, Berkeley; Biometrics Res., Merck Res. Labs, 2004. [Online]. Available: https://www.stat.berkeley.edu/sites/default/files/tech-reports/666.pdf

C. Lawrence, T. Tuunanen, and M. D. Myers, "Extending Design Science Research Methodology for a Multicultural World," in Human Benefit through the Diffusion of Information Systems Design Science Research, J. Pries-Heje, J. R. Venable, D. Bunker, N. L. Russo, and J. I. DeGross, Eds., Perth, Australia: Springer, IFIP Adv. Inf. Commun. Technol., vol. 318, pp. 108–121, 2010. doi: 10.1007/978-3-642-12113-5_9.

G. Liang and A. G. Cohn, "An Effective Approach for Imbalanced Classification: Unevenly Balanced Bagging," in Proc. 27th AAAI Conf. Artif. Intell. (AAAI'13), Washington, DC, USA, Jul. 2013, pp. 1625–1631.

M. Pontiki, D. Galanis, J. Pavlopoulos, H. Papageorgiou, I. Androutsopoulos, and S. Manandhar, "SemEval-2014 Task 4: Aspect Based Sentiment Analysis," in Proc. 8th Int. Workshop Semantic Eval. (SemEval 2014), pp. 27–35.

W. Che, Y. Zhao, H. Guo, Z. Su, and T. Liu, "Sentence compression for Aspect Based Sentiment Analysis," IEEE/ACM Trans. Audio, Speech, Lang. Process., vol. 23, no. 12, pp. 2111–2124, Dec. 2015. doi: 10.1109/TASLP.2015.2443982.

M. Pontiki, D. Galanis, H. Papageorgiou, I. Androutsopoulos, S. Manandhar, M. Al-Smadi, M. Al-Ayyoub, Y. Zhao, B. Qin, O. De Clercq, V. Hoste, M. Apidianaki, X. Tannier, N. Loukachevitch, E. Kotelnikov, N. Bel, S. M. Jiménez-Zafra, and G. Eryiğit, "SemEval-2016 Task 5: Aspect based sentiment analysis," in Proc. 10th Int. Workshop Semantic Eval. (SemEval-2016), pp. 19–30, 2016.

M. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, and L. Zettlemoyer, "Deep contextualized word representations," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol. (NAACL-HLT 2018), vol. 1, pp. 2227–223.

M. Hoang, O. A. Bihorac, and J. Rouces, "Aspect based sentiment analysis using BERT," in Proc. 22nd Nordic Conf. Comput. Linguist. (NoDaLiDa 2019), pp. 187–196. Linköping, Sweden: Linköping Univ. Electronic Press, 2019. [Online]. Available: https://www.aclweb.org/anthology/W19-5317

J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of deep bidirectional transformers for language understanding," in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguist.: Hum. Lang. Technol. (NAACL-HLT 2019), vol. 1, pp. 4171–4186.

J. Memon, M. Sami, R. A. Khan, and M. Uddin, "Handwritten Optical Character Recognition (OCR): A Comprehensive Systematic Literature Review (SLR)," IEEE Access, vol. 8, pp. 142642–142668, 2020. doi: 10.1109/ACCESS.2020.3012542.

A. Gwagwa, E. Kraemer-Mbula, N. Rizk, et al., "Responsible Artificial Intelligence in Sub-Saharan Africa: Landscape and Approach," Int. Res. Centre Artif. Intell. (IRCAI), Ljubljana, Slovenia, Mar. 2021. [Online]. Available: https://ircai.org/wp-content/uploads/2021/03/AI4D_Report_Responsible_AI_in_SSA.pdf

A. Newaz, S. Hassan, and F. S. Haq, "An Empirical Analysis of the Efficacy of Different Sampling Techniques for Imbalanced Classification," Dept. Elect. Electron. Eng., Islamic Univ. of Technology, Gazipur, Bangladesh, 2021.

Z. Xu, "Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities," Appl. Artif. Intell., Advance online publication, 2021. doi: 10.1080/08839514.2021.2013652.

P. Li, A. Bastone, T. A. Mohamad, and F. Schiavone, "How does artificial intelligence impact human resources performance? Evidence from a healthcare institution in the United Arab Emirates," J. Innov. Knowl., vol. 8, no. 2, Art. 100340, 2023. doi: 10.1016/j.jik.2023.100340.

W. Rodgers, J. M. Murray, A. Stefanidis, W. Y. Degbey, and S. Y. Tarba, "An artificial intelligence algorithmic approach to ethical decision-making in human resource management processes," Hum. Resour. Manage. Rev., vol. 33, no. 1, Art. 100925, 2023. doi: 10.1016/j.hrmr.2022.100925.

N. Nawaz, H. Arunachalam, B. K. Pathi, and V. Gajenderan, "The adoption of artificial intelligence in human resources management practices," Int. J. Inf. Manage. Data Insights, vol. 4, no. 1, Art. 100208, 2024. doi: 10.1016/j.jjimei.2023.100208.

V. Frimpong, "Artificial Intelligence Investment in Resource-Constrained African Economies: Financial, Strategic, and Ethical Trade-Offs with Broader Implications," World, vol. 6, no. 2, p. 70, May 2025. doi: 10.3390/world6020070.

G. Husain, D. Nasef, R. Jose, J. Mayer, M. Bekbolatova, T. Devine, and M. Toma, "SMOTE vs. SMOTEENN: A Study on the Performance of Resampling Algorithms for Addressing Class Imbalance in Regression Models," Algorithms, vol. 18, no. 1, p. 37, Jan. 2025. doi: 10.3390/a18010037.

R. Tulluri, Employee Promotion Prediction Dataset. [Online]. Available: https://github.com/rajtulluri/EmployeePromotionPrediction/blob/master/employeePromotion.csv [Accessed: Aug. 11, 2025].

T. M. Takkimsncn, HR Classification for Promotion. [Online]. Available: https://www.kaggle.com/code/takkimsncn/hr-classification-for-promotion/input