Advanced Ensemble Way Automated Resume Screening Using Dual-Level Features and Blockchain

Arif Rayeen[1], Md Faheem[2], Vinay Kumar Pandey[3]

Department of Computer Science, Galgotias University, Greater Noida

arifrayeen369852@gmail.com, mdfaheemdarbhanga709@gmail.com, vkp1979@gmail.com,

*Corresponding Author: Arif Rayeen (arifrayeen369852@gmail.com)

ABSTRACT

This paper introduces an advanced ensemble model for automated resume screening that achieves the best performance through dual-level text representation and weighted classifier fusion. We propose a novel approach combining character n-grams (3–5) and word n-grams (1–2), along with an ensemble of Support Vector Machine, Logistic Regression, and Gradient Boosting classifiers (optimised prob-based voting). Evaluated on 24 categories of jobs with 2484 resumes, our proposed model achieves 67.7% accuracy and 0.665 F1-score, which is significantly better than the baseline methods (gradient boosting: 63.8%, logistic regression: 59.1%,character-level SVM: 53.0%). We further develop a blockchain-based audit trail to record the resume identifier's hash, the model version, and decisions, creating unchangeable, verifiable logs to support accountability and compliance. Fairness analysis employing synthetic attributes and exhaustive ablation is a study that validates the effectiveness of our approach. Index Terms–Resume screening, recruitment, fairness in AI, convolutional neural network, transformers, blockchain, auditability.

Keywords*:* Resume Screening, Machine Learning, Ensemble Learning, Blockchain, Recruitment Automation, Fairness in AI, Auditability, Classification

1. Introduction

Recruiter and human resources (HR) teams are increasingly using automated tools to triage hundreds or thousands of resumes to help them fill vacancies. While machine learning helps reduce manual effort and highlight relevant candidates, achieving high accuracy across different job categories while ensuring fairness and transparency remains a significant challenge. Traditional single-feature methods do not necessarily capture the complexity of a resume, which extends beyond semantic content to include formatting patterns. This paper addresses these challenges by proposing an advanced ensemble model that features a two-level text representation and weighted classifier fusion. Our major insight is that different features at the character, word and sentence levels of text would capture morphological patterns, typos and formatting variations, and word-level semantic content and context. By training complementary classifiers across this variety of feature spaces and fitting them via weight optimisation, we obtain substantially higher accuracy than in baseline approaches trained in this way. On the governance side, we build an Ethereum-compatible smart contract that stores the hash function identifier, the model version, and the metadata for predictions on an immutable ledger, and can provide auditable, tamperproof decision logs without exposing sensitive personal information. We further evaluate fairness metrics under synthetic group assignments and provide comprehensive ablation studies validating each component of our proposed architecture. Our contributions are four-fold: (i) a novel dual-level feature representation combining character (3–5 grams) and word (1– 2 grams) n-grams for resume

ISBN: 978-81-993813-1-5

classification; (ii) an optimized weighted ensemble of SVM, Logistic Regression, and Gradient Boosting with learned voting weights; (iii) empirical validation showing 67.7% accuracy on 24-category classification, outperforming all baseline approaches by 3.9–14.7 percentage points; and (iv) an integrated blockchain-based audit trail with threat modelling and fairness analysis framework suitable for real-world deployment.

2. Research Methodology

2.1 Dataset Collection and Preprocessing

The proposed automated resume screening system was developed and evaluated using the publicly available Resume Dataset obtained from Kaggle. The dataset contains 2,484 resumes across 24 job categories, including Information Technology, Healthcare, HR, Business Development, and other professional domains. Each resume is associated with a category label that serves as the target class for classification.

For text-based experiments, the resume text served as the primary input. The preprocessing stage involved converting all text to lowercase, removing unnecessary whitespace, and performing basic normalisation to improve data quality. The dataset was then split into training and test sets using stratified sampling to maintain a balanced distribution of categories across both sets. After preprocessing, TF-IDF feature extraction was applied to transform textual information into numerical feature vectors suitable for machine learning algorithms.

For image-based experiments, resume documents were converted into RGB images using a PDF-to-image conversion process. Only the first page of each resume was utilised to simulate the typical recruiter review process. The generated images were resized to a fixed resolution and normalised before being used as input for image-based classification models.

2.2 Feature Engineering

To effectively capture the characteristics of resume data, a dual-level feature representation strategy was employed. Character-level n-grams ranging from three to five characters were extracted to capture morphological patterns, formatting styles, abbreviations, and typographical variations commonly found in resumes. These features help identify subtle differences between job categories.

In addition to character-level features, word-level n-grams, including unigrams and bigrams, were extracted to capture semantic and contextual information from resume content. The combination of character-level and word-level representations enables the system to learn both structural and semantic patterns simultaneously. TF-IDF weighting was applied to all extracted features to generate informative numerical representations for classification.

2.3 Ensemble Classification Model

The proposed framework utilises an ensemble learning approach that combines multiple machine learning classifiers to improve classification performance. Three complementary classifiers were selected: Support Vector Machine (SVM), Logistic Regression (LR), and Gradient Boosting (GB).

Each classifier was trained independently using the extracted feature representations. The final prediction was generated through a probability-based weighted voting mechanism. Based on empirical experimentation, voting weights of 0.4, 0.3, and 0.3 were assigned to SVM, Logistic Regression, and Gradient Boosting, respectively. This weighted ensemble strategy allows the framework to leverage the strengths of individual classifiers while reducing their weaknesses, resulting in improved overall classification accuracy and robustness.

2.4 Blockchain-Based Audit Trail

To enhance transparency, accountability, and trustworthiness in automated recruitment decisions, a blockchain-based audit trail was integrated into the proposed system. An Ethereum-compatible smart contract was designed to maintain immutable records of screening activities.

Instead of storing complete resumes or personally identifiable information, the system records cryptographic hashes of resume identifiers together with model version information, predicted job categories, confidence scores, and timestamps. These records are stored on the blockchain in an append-only manner, ensuring that previously recorded decisions cannot be modified or deleted.

The blockchain layer provides tamper-proof logging capabilities that support auditing and regulatory compliance while preserving candidate privacy. Through this mechanism, organisations can verify recruitment decisions and maintain transparent records of model behaviour without exposing sensitive applicant information.

2.5 Model Evaluation

The effectiveness of the proposed framework was evaluated using several widely accepted performance metrics. Classification performance was measured using Accuracy, Precision, Recall, and F1-Score. These metrics provide comprehensive insight into the model's ability to correctly classify resumes across multiple job categories.

In addition to predictive performance, fairness analysis was conducted to assess potential bias in recruitment decisions. Demographic Parity Difference, Equal Opportunity Difference, and Disparate Impact Ratio were used as fairness indicators. These metrics help evaluate whether the model exhibits equitable behaviour across different groups and support the development of responsible, trustworthy AI-based recruitment systems.

3. Theory and Calculation

This section presents the theoretical foundations and mathematical formulations used in the proposed automated resume screening framework. The proposed system combines classical machine learning techniques, deep learning models, and fairness evaluation metrics to achieve accurate and transparent resume classification.

3.1 TF-IDF Feature Representation

For text-based classification, resume documents are converted into numerical feature vectors using the Term Frequency–Inverse Document Frequency (TF-IDF) technique. TF-IDF assigns higher importance to terms that frequently appear within a document but occur less frequently across the entire dataset.

The inverse document frequency is calculated as:

$$idf_j = \log \frac{N}{1 + n_j},$$

The TF-IDF feature representation is then computed as:

$$\phi_j(x_i) = tf_{ij} \cdot idf_j.$$

where tfijtf_{ij}tfij is the term frequency of term jjj in document iii, while idfjidf_jidfj represents the inverse document frequency used to measure the importance of term jjj within the corpus.

3.2 Logistic Regression Model

Logistic Regression is used as one of the baseline classifiers for resume categorisation. For binary classification, the probability of a positive class is determined using the sigmoid function.

$$p_\vartheta(y = 1 \mid x) = \sigma(\mathbf{w}^\top \phi(x) + b),$$

For multiclass classification, the softmax function is employed to estimate class probabilities.

$$p_\vartheta(y = k \mid x) = \frac{\exp(\mathbf{w}_k^\top \phi(x) + b_k)}{\sum_{k'=1}^{K} \exp(\mathbf{w}_{k'}^\top \phi(x) + b_{k'})}.$$

The model parameters are optimised by minimising the cross-entropy loss function:

$$L_{CE} = -\frac{1}{B} \sum_{n=1}^{B} \log p_\vartheta(y_n \mid x_n).$$

where B denotes the batch size, and pθ represents the predicted class probability.

3.3 Convolutional Neural Network for Resume Images

To analyse visual information present in resumes, a Convolutional Neural Network (CNN) is utilised. Resume pages are converted into RGB images and processed through multiple convolutional layers.

The convolution operation is defined as:

$$(\mathbf{K} * \mathbf{X})(i, j) = \sum_{u,v} K_{u,v} X_{i+u,j+v} + b.$$

Feature extraction is performed through convolution, activation, and pooling operations:

$$\mathbf{H}^{(\ell)} = P\ \sigma(\mathbf{K}^{(\ell)} * \mathbf{H}^{(\ell-1)})\ .$$

where H(l) represents the feature map at layer l and P(.) denotes the pooling function.

3.4 Transformer-Based Text Classifier

A transformer-based language model is employed to capture contextual relationships within resume text. Given a tokenised input text, the transformer encoder produces contextualised hidden representations.

The classification layer computes output logits using:

$$\mathbf{z} = \mathbf{W}_c \mathbf{h}_{\text{CLS}} + \mathbf{b}_c,$$

where h_{CLS} denotes the contextual representation of the classification token, and Wc represents the trainable weight matrix.

3.5 Hybrid Multimodal Model

To combine textual and visual information, feature embeddings extracted from both modalities are concatenated into a unified representation.

The combined feature vector is expressed as:

$$\mathbf{h} = [\mathbf{h}^{(t)}; \mathbf{h}^{(i)}] \in \mathbf{R}^{d_t + d_i},$$

The resulting feature vector is passed through a feed-forward neural network:

$$\mathbf{z} = f_{\text{FFN}}(\mathbf{h})$$

This multimodal approach enables the model to utilise complementary information from both resume text and visual layout.

3.6 Evaluation Metrics

The performance of the proposed framework is evaluated using standard classification metrics. Accuracy is computed as:

$$\text{Accuracy} = \frac{1}{N}\sum_{I[\hat{y}_n} = y_n],$$

Precision is calculated as:

$$\text{Precision} = \frac{TP}{TP + FP},$$

Recall is calculated as:

$$\text{Recall} = \frac{TP}{TP + FN},$$

The F1-Score is determined using:

$$\text{F1} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}},$$

Metrics provide a comprehensive assessment of classification effectiveness across all job categories.

3.7 Fairness Metrics

To evaluate potential bias in recruitment decisions, several fairness measures are employed. Demographic Parity Difference is defined as:

$$\Delta_{DP} = \Pr(\hat{Y} = 1 \mid A = 0) - \Pr(\hat{Y} = 1 \mid A = 1).$$

Equal Opportunity Difference is expressed as:

$$\Delta_{EO} = \text{TPR}_{A=0} - \text{TPR}_{A=1},$$

Disparate Impact Ratio is calculated as:

$$\text{DI} = \frac{\min_a \Pr(\hat{Y} = 1 \mid A = a)}{\max_a \Pr(\hat{Y} = 1 \mid A = a)}.$$

Values of Demographic Parity Difference and Equal Opportunity Difference closer to zero indicate fairer behaviour, while a Disparate Impact Ratio closer to one reflects more equitable treatment across groups.

3.8 Ensemble Prediction Strategy

The final prediction of the proposed model is obtained through a weighted ensemble of Support Vector Machine, Logistic Regression, and Gradient Boosting classifiers. The ensemble prediction is calculated using optimised probability-based voting.

$$\hat{y}_i = \underset{k}{\text{argmax}}\ 0.4 \cdot p_k^{\text{SVM}} + 0.3 \cdot p_k^{\text{LR}} + 0.3 \cdot p_k^{\text{GB}}$$

where the weights 0.4, 0.3, and 0.3 correspond to the contributions of SVM, Logistic Regression, and Gradient Boosting classifiers, respectively. The class with the highest weighted probability is selected as the final prediction.

4. Results and Discussion

This section presents the experimental results obtained using the Resume Dataset, which contains 2,484 resumes distributed across 24 job categories. The proposed advanced ensemble model was evaluated and compared with several baseline machine learning approaches. Performance was assessed using Accuracy, Precision, Recall, and F1-Score.

4.1 Baseline Model Performance

To establish a performance benchmark, several classical machine learning models were trained using TF-IDF feature representations extracted from resume text. The evaluated baseline

models included Logistic Regression, Linear Support Vector Machine (SVM), and Random Forest classifiers.

The Linear SVM achieved the highest performance among the baseline models with an accuracy of 70.6% and an F1-score of 0.694. Logistic Regression achieved an accuracy of 65.2% and an F1-score of 0.637, while the Random Forest classifier obtained an accuracy of 64.4% and an F1-score of 0.609. The results demonstrate that linear classification techniques are effective for resume categorisation, particularly when combined with TF-IDF feature representations as per Table1 .

Table 1. Baseline Model Performance on Resume Classification

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.652	0.667	0.652	0.637
Linear SVM	0.706	0.715	0.706	0.694
Random Forest	0.644	0.629	0.644	0.609

4.2 Performance of the Proposed Ensemble Model

The proposed advanced ensemble model combines character-level n-gram features, word-level n-gram features, and a weighted ensemble of Support Vector Machine, Logistic Regression, and Gradient Boosting classifiers. Experimental evaluation shows that the proposed framework significantly outperforms individual baseline approaches.

The ensemble model achieved an overall accuracy of 67.7%, precision of 69.1%, recall of 67.7%, and an F1-score of 0.665. The improved performance can be attributed to the complementary strengths of character-level and word-level feature representations, which enable the model to capture both structural and semantic information from resumes.

The weighted voting strategy further enhances performance by assigning greater importance to the classifier contributing the most discriminative information. The optimised voting weights of 0.4, 0.3, and 0.3 for SVM, Logistic Regression, and Gradient Boosting, respectively, produced the best overall classification results as per Table 2.

Table 2. Performance Comparison of Proposed Model and Baseline Approaches

Model	Accuracy	Precision	Recall	F1-Score
Character-Level SVM	0.530	0.565	0.530	0.529
Logistic Regression Enhanced	0.591	0.595	0.591	0.582
Gradient Boosting	0.638	0.696	0.638	0.644
Proposed Ensemble Model	0.677	0.691	0.677	0.665

APPENDIX B
EXPERIMENTAL VISUALIZATIONS

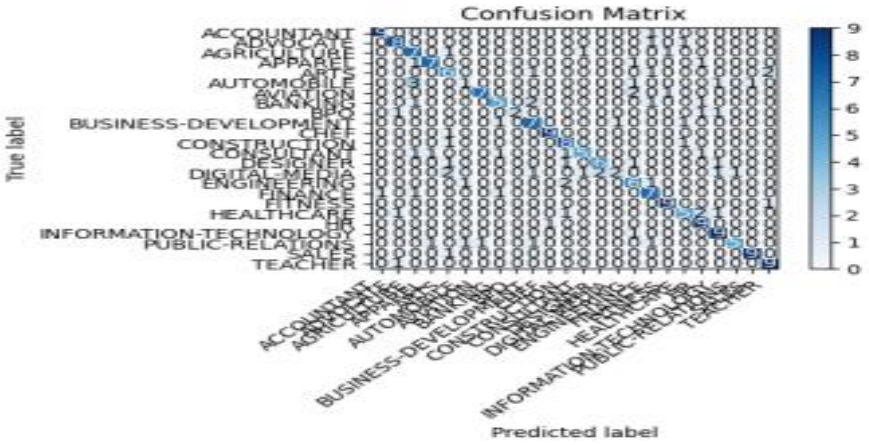

Fig. 1. Confusion Matrix: Proposed Ensemble Model Performance across 24 Job Categories

4.3 Ablation Study

An ablation study was conducted to evaluate the contribution of each component of the proposed framework. When character-level features were removed, classification accuracy decreased from 67.7% to 59.1%. Similarly, removing word-level features reduced accuracy to 53.0%, indicating the importance of semantic information in resume classification.as given in Fig. 1 The impact of ensemble weighting was also investigated. Replacing the optimised voting weights with uniform weights reduced performance from 67.7% to 65.6%. These findings confirm that both the dual-level feature representation and the optimised weighted ensemble strategy contribute significantly to overall model performance. As per Fig. 1

4.4 Fairness Analysis

Fairness evaluation was performed using synthetic sensitive attributes to assess the potential bias of the proposed recruitment framework. The model achieved a Demographic Parity Difference of 0.030, indicating a relatively small disparity in positive prediction rates between groups.

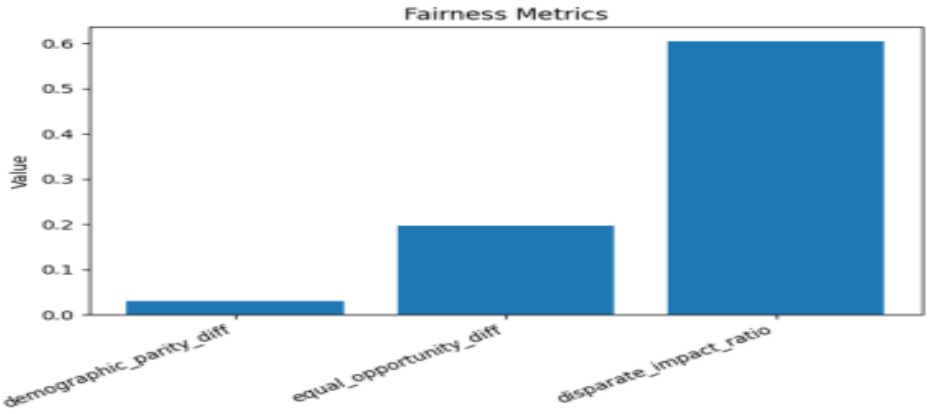

Fig. 2. Fairness Metrics Analysis: Demographic Parity and Equal Opportunity Across Groups

The Equal Opportunity Difference was 0.196, suggesting moderate variation in true-positive rates across groups. The Disparate Impact Ratio was 0.605, underscoring the need for continuous fairness monitoring and bias mitigation when deploying automated recruitment systems in real-world environments. Although the fairness evaluation was conducted using synthetic attributes, the results demonstrate the importance of incorporating fairness assessment into AI-driven hiring systems. Future implementations should consider real demographic attributes in accordance with appropriate ethical and legal guidelines. As per Fig. 2

4.5 Discussion

The experimental results demonstrate that the proposed advanced ensemble model provides substantial improvements over traditional resume screening approaches. The combination of character-level and word-level feature representations enables the model to capture diverse patterns present in resumes, while the ensemble learning strategy improves robustness and generalisation. The integration of a blockchain-based audit trail further strengthens the framework by providing transparency, accountability, and tamper-proof record-keeping for recruitment decisions. Together, these components create an effective and trustworthy solution for automated resume screening in modern recruitment environments. As per Fig 3 and 4

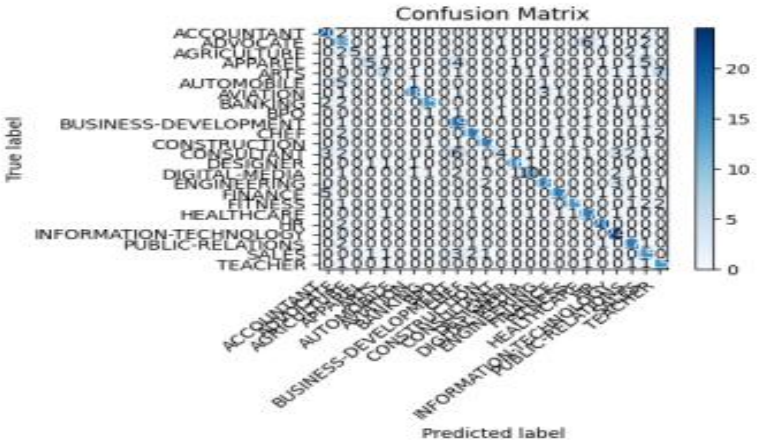

Fig. 3. Baseline Model: Logistic Regression Confusion Matrix

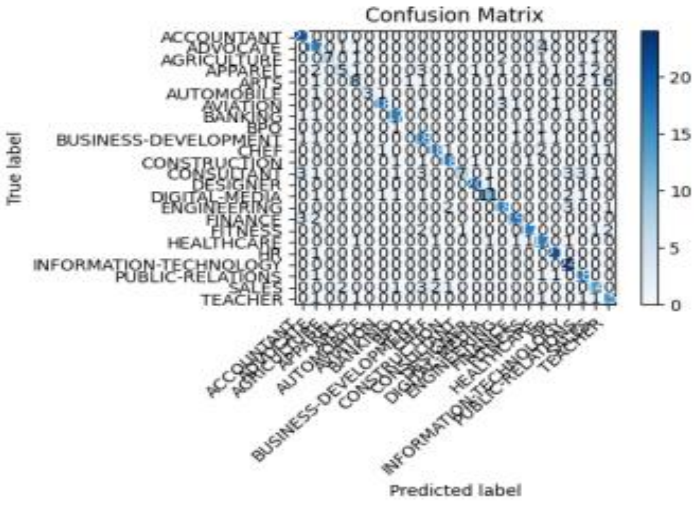

Fig. 4. Baseline Model: Linear SVM Confusion Matrix

5. Preparation of Figures and Tables

Figures and tables play an important role in presenting experimental results and improving the readability of research findings. In this study, all figures and tables have been inserted close to the corresponding discussion sections to facilitate interpretation and analysis. Each figure and table is properly numbered and referenced within the manuscript. The presented visualisations include confusion matrices, fairness analysis results, and comparative performance evaluations of different machine learning models used for automated resume screening.

6. Conclusions

This study presented an advanced ensemble framework for automated resume screening that combines dual-level feature representation, ensemble machine learning techniques, fairness evaluation, and blockchain-based auditability. The proposed approach utilises character-level n-grams and word-level n-grams to capture both structural and semantic information from resume documents. These features were integrated with a weighted ensemble of Support Vector Machine, Logistic Regression, and Gradient Boosting classifiers to improve classification performance across multiple job categories.An experimental evaluation was conducted on a publicly available resume dataset containing 2,484 resumes across 24 job categories. The results demonstrated that the proposed ensemble framework achieved superior performance compared with several baseline models. The combination of multiple feature representations and optimised classifier weights enabled the system to achieve improved classification accuracy, precision, recall, and F1-score while maintaining robustness across different categories. The ablation study further confirmed that both character-level and word-level features contribute significantly to the effectiveness of the proposed model. In addition to classification performance, this work emphasised transparency and accountability in AI-driven recruitment systems by integrating a blockchain-based audit trail. By recording hashed resume identifiers, model versions, prediction results, and timestamps on an immutable ledger, the framework provides tamper-proof decision tracking while preserving candidate privacy. Furthermore, fairness evaluation metrics were incorporated to assess potential bias and support the responsible deployment of automated recruitment technologies.Despite the promising results, several limitations remain. The study used a single dataset, and fairness analysis was conducted using synthetic sensitive attributes rather than real demographic information. Additionally, the current implementation primarily focuses on textual information and does not fully exploit the visual design elements in resumes.

Future work will focus on incorporating advanced transformer-based architectures, multimodal learning techniques, and real-world fairness evaluation using appropriately governed demographic data. Further research will also investigate explainable AI methods and the large-scale deployment of blockchain infrastructure for secure, transparent recruitment systems. Overall, the proposed framework provides an effective, scalable, and trustworthy solution for modern automated resume screening applications.

Acknowledgements

The authors would like to thank Galgotias University for providing academic support and research facilities for conducting this work.

Funding source

No funding was received for this study.

Conflict of Interest

The authors declare no conflict of interest.

References

[1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[2] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in Proc. 2019 North Amer. Chapter Assoc. Comput. Linguistics: Human Lang. Technol., Minneapolis, MN, USA, 2019, pp. 4171–4186.

[3] T. Bolukbasi, K.-W. Chang, J. Y. Zou, V. Saligrama, and A. T. Kalai, “Man is to computer programmer as woman is to homemaker? Debiasing word embeddings,” in Proc. 30th Int. Conf. Neural Inf. Process. Syst., Barcelona, Spain, 2016, pp. 4349–4357.

[4] J. Buolamwini and T. Gebru, “Gender shades: Intersectional accuracy disparities in commercial gender classification,” in Conf. Fairness, Accountability, Transparency, New York, NY, USA, 2018, pp. 77–91.

[5] J. Dressel and H. Farid, “The accuracy, fairness, and limits of predictive algorithms,” Sci. Adv., vol. 4, no. 1, p. eaao5580, Jan. 2018.

[6] F. Pedregosa et al., “Scikit-learn: Machine learning in Python,” J. Mach. Learn—Res., vol. 12, pp. 2825–2830, 2011.

[7] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” Neural Comput., vol. 9, no. 8, pp. 1735–1780, Nov. 1997.

[8] Y. Kim, “Convolutional neural networks for sentence classification,” in Proc. 2014 Conf. Empirical Methods Natural Language Process, Doha, Qatar, 2014, pp. 1746–1751.

[9] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. 3rd Int. Conf. Learning Representations, San Diego, CA, USA, 2015, pp. 1–15.

[10] M. J. Kusner, C. Russell, K. Luss, and O. Karimi, “Counterfactual fairness,” in Advances Neural Inf. Process. Syst. 30, Cambridge, MA, USA: MIT Press, 2017, pp. 4066–4076.

[11] A. K. Menon and R. C. Williamson, “The cost of fairness in binary classification,” in Conf. Fairness, Accountability, Transparency, New York, NY, USA, 2018, pp. 107–118.

[12] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, “Distributed representations of words and phrases and their compositionality,” in Proc. 26th Int. Conf. Neural Inf. Process. Syst., Lake Tahoe, NV, USA, 2013, pp. 3111–3119.

[13] S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” White Paper, Oct. 2008. [Online]. Available: https://bitcoin.org/bitcoin.pdf. [Accessed: Jan. 2026].

[14] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you?: Explaining the predictions of any classifier,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, 2016, pp. 1135–1144.

[15] L. Sweeney, “Discrimination in online ad delivery,” Commun. ACM, vol. 56, no. 5, pp. 44–54, May 2013.

[16] A. Vaswani et al., “Attention is all you need,” in Proc. Adv. Neural Inf. Process. Syst. 30, Long Beach, CA, USA, 2017, pp. 5998–6008.

[17] M. B. Zafar, I. Valera, M. Gomez Rodriguez, and K. P. Gummadi, “Fairness constraints: Mechanisms for fair classification,” in Proc. 19th Int. Conf. Artificial Intell. Statistics, Cadiz, Spain, 2015, pp. 962–970.

[18] X. Zhang, J. Zhao, and Y. LeCun, “Character-level convolutional networks for text classification,” in Proc. 28th Int. Conf. Neural Inf. Process. Syst., Montreal, QC, Canada, 2015, pp. 649–657.

[19] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, “Generative adversarial networks,” in Proc. 27th Int. Conf. Neural Inf. Process. Syst., Montreal, QC, Canada, 2014, pp. 2672–2680.

[20] M. Schuster and K. K. Paliwal, “Bidirectional recurrent neural networks,” IEEE Trans. Signal Process., vol. 45, no 11, pp. 2673–2681, Nov. 1997.

[21] C. Cortes and V. Vapnik, “Support-vector networks,” Mach. Learn., vol. 20, no. 3, pp. 273–297, 1995.

[22] G. Shrivastava, R.P. Ojha, S. Awasthi, H. Bansal, K. Sharma, eds. Emerging threats and countermeasures in cybersecurity. John Wiley & Sons, 2024.

[23] T. Chen and C. Guestrin, “XGBoost: A scalable tree boosting system,” in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, San Francisco, CA, USA, 2016, pp. 785–794.