RecruitGuard: A Review of Deep Learning Techniques for Fraudulent Job Posting Detection

Mohammed Faiz[1], Prof. A. D. Wakhare[2]
[1]*MTech Student, Department of CSE, Deogiri Institute of Engineering and Management Studies, Chh. Sambhajinagar, Maharashtra, India*
[2]*Assistant Professor, Department of CSE, Deogiri Institute of Engineering and Management Studies, Chh. Sambhajinagar, Maharashtra, India*

Abstract*—The rapid growth of online recruitment platforms has increased the risk of fake job postings and recruitment fraud. Fraudulent job advertisements can mislead job seekers, compromise personal information, and damage organizational credibility. Traditional detection methods are often inefficient due to the large volume and dynamic nature of online job data. This review paper analyzes existing fake job detection techniques based on Machine Learning, Deep Learning, Natural Language Processing, and hybrid models. It highlights the advantages and limitations of approaches such as LSTM, Bi-LSTM, and ensemble learning. The study also identifies key research gaps and emphasizes the need for accurate, explainable, and real-time fake job detection systems.

***Keywords*—Fake Jobs Detection, Online Recruitment Fraud, LSTM, NLP, Machine Learning, Deep Learning, Bi-LSTM, Fraud Detection, Real-Time Verification**

I. INTRODUCTION

The rapid advancement of internet technologies and online recruitment platforms has significantly transformed the modern hiring process. Organizations increasingly rely on digital job portals such as LinkedIn, Indeed, Glassdoor, and Naukri to advertise vacancies and recruit candidates efficiently. While these platforms provide convenience, accessibility, and faster communication, they have also created opportunities for cybercriminals and fraudulent recruiters to exploit job seekers through fake job advertisements and recruitment scams. As online job applications continue to grow worldwide, fake job postings have emerged as a serious cybersecurity and social issue.

Fake job advertisements are deceptive recruitment postings designed to mislead candidates for malicious purposes such as financial fraud, identity theft, phishing attacks, and unauthorized collection of personal information. These fraudulent postings often imitate genuine organizations and use professional language to appear legitimate. Job seekers, especially students and unemployed individuals, are highly vulnerable to such scams because they frequently share sensitive information including resumes, educational records, contact details, and banking information during the application process. In many cases, victims are asked to pay registration fees, training charges, or security deposits, resulting in financial losses and emotional distress.

Traditional methods for detecting fake job postings mainly rely on manual verification, rule-based filtering, and blacklist mechanisms. However, these approaches are no longer sufficient due to the enormous volume of online recruitment data and the constantly evolving strategies used by fraudsters. Consequently, researchers have focused on developing automated fake job detection systems using Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) techniques.

Machine Learning algorithms such as Decision Trees, Naive Bayes, Support Vector Machines, and Random Forest classifiers analyze textual and metadata features to classify job postings as real or fake. More advanced Deep Learning models, including Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Bi-directional

LSTM (Bi-LSTM), provide improved contextual understanding and sequential learning capabilities. Additionally, NLP techniques help extract semantic, linguistic, and behavioral patterns from job descriptions for more accurate fraud detection.

Despite significant advancements, existing fake job detection systems still face challenges such as dataset imbalance, lack of explainability, high computational complexity, and limited real-time detection capability. Therefore, there is a growing need for intelligent, scalable, and explainable fake job detection frameworks capable of improving recruitment security and protecting job seekers from online fraud.

II. LITERATURE REVIEW

Fake job detection has gained significant attention with the increasing growth of online recruitment platforms and digital hiring systems. Researchers have proposed various Machine Learning (ML), Deep Learning (DL), and Natural Language Processing (NLP) techniques to identify fraudulent job advertisements effectively. Ramya et al. (2025) proposed a fake job prediction framework using multiple Machine Learning algorithms for classifying job advertisements as real or fake. Their approach improved prediction accuracy but lacked deep contextual understanding [1]. Kumar and Gupta (2025) developed a Machine Learning-based fake job post detection system using textual and metadata analysis. Although the model achieved good classification performance, it showed limitations in handling complex linguistic patterns [2]. Shankar and Patel (2025) introduced Artificial Neural Networks (ANN) combined with ML techniques for fraudulent job detection. Their model enhanced detection capability but required higher computational resources [3].

Reddy and Thomas (2025) proposed an LSTM-based framework for detecting fake job postings using sequential text analysis. The model effectively captured contextual dependencies but suffered from increased training complexity [4]. Roy and Mazumdar (2025) applied Deep Convolutional Neural Networks (CNN) for fake job detection, achieving improved feature extraction and classification performance; however, the approach required large datasets for effective training [5]. Gupta and Dave (2025) used TF-IDF feature extraction combined with ML classifiers for fake job posting detection. Their system improved interpretability but lacked semantic contextual learning [6].

Rahman and Basu (2024) implemented deep learning techniques for fraudulent online job advertisement detection. Their approach achieved higher accuracy compared to traditional ML models but involved increased computational cost [7]. Kulkarni and Ahmed (2024) developed a hybrid ML-NLP pipeline integrating textual preprocessing, feature engineering, and classification models. The system improved robustness but required extensive preprocessing [8]. Zhang and Lin (2024) proposed a transformer-based architecture for fake job detection, demonstrating superior contextual understanding and semantic analysis, although the model demanded high computational power [9].

Shah and Mehta (2024) focused on real-time fake job filtering systems using Machine Learning techniques. Their framework improved real-time applicability but faced challenges in maintaining prediction accuracy under high data volume [10]. Nashaat et al. (2023) incorporated Explainable Artificial Intelligence (XAI) using SHAP analysis for fake job detection. Their approach improved transparency and interpretability but increased processing overhead [11]. Verma and Goyal (2023) compared different ML models for job fraud identification and concluded that ensemble techniques provide better classification accuracy than standalone classifiers [12].

Yadav and Arora (2023) proposed ensemble learning methods for fake job detection, combining multiple classifiers to improve robustness and prediction performance [13]. Prajapati and Patel (2023) emphasized NLP-based feature engineering techniques for scam job detection, highlighting the importance of semantic and linguistic analysis [14]. Johnson and Smith (2022) introduced multimodal learning approaches combining textual and metadata analysis for recruitment scam detection, improving overall detection capability [15]. Sultana et al. (2022) applied LSTM networks for job scam detection and demonstrated improved sequential learning and contextual representation [16]. Patel and Shah (2022) implemented Gradient Boosting algorithms for fake job advertisement classification, achieving high

prediction accuracy with optimized feature selection [17]. Jain and Srivastava (2021) explored NLP-based fake employment posting detection using textual preprocessing and semantic analysis [18]. Mishra and Singh (2021) studied various deep learning architectures for job scam detection and reported improved performance using hybrid neural models [19]. Sharma and Verma (2021) applied Logistic Regression techniques for predicting job posting authenticity, providing computational efficiency but limited contextual understanding [20].

III. RESEARCH GAPS

Based on the analysis of existing fake job detection techniques, several significant research gaps have been identified. Most current systems primarily focus on offline analysis and are not optimized for real-time fake job verification, limiting their practical applicability in dynamic online recruitment platforms. Although Machine Learning and Deep Learning models have achieved promising accuracy, many of these approaches behave as black-box systems and fail to provide clear explanations for their predictions, thereby reducing transparency and user trust.

Another major challenge is dataset imbalance, where genuine job postings significantly outnumber fake advertisements. This imbalance often leads to biased models that perform poorly in identifying fraudulent postings. Additionally, many existing studies rely heavily on textual analysis while ignoring important external verification factors such as company authenticity, recruiter credibility, and contact information validation. Cross-platform generalization also remains a major issue, as models trained on one recruitment platform often fail to maintain similar performance on other job portals due to differences in formatting, writing style, and data structure. Furthermore, Deep Learning models such as LSTM and Transformer-based architectures require high computational resources, making them difficult to deploy in real-time environments. These limitations highlight the need for an intelligent, explainable, scalable, and real-time fake job detection framework for improved fraud detection performance.

IV. PROBLEM STATEMENT

The increasing number of fake job postings across online recruitment platforms poses significant risks to job seekers and organizations. Existing fake job detection systems suffer from limitations such as poor contextual understanding, lack of explainability, dataset imbalance, high computational complexity, and limited real-time detection capability. Therefore, there is a need for an intelligent and efficient fake job detection framework capable of accurately identifying fraudulent recruitment advertisements.

V. PROPOSED METHODOLOGY

To overcome the limitations of existing fake job detection systems, a hybrid intelligent framework is proposed that integrates Natural Language Processing, Deep Learning, and heuristic verification mechanisms.

A. Hybrid Detection Framework

The proposed system combines textual analysis, contextual learning, and metadata verification for improved fake job detection accuracy.

B. Data Collection

Job posting data is collected from online recruitment platforms and benchmark datasets. The dataset contains both real and fake job advertisements.

C. Data Preprocessing

The collected data undergoes preprocessing steps including tokenization, stop-word removal, lowercasing, lemmatization, and noise removal.

D. Feature Extraction

Important textual and metadata features are extracted including TF-IDF vectors, word embeddings, salary patterns, company profile analysis, contact information consistency, and URL verification.

E. Deep Learning Module

The processed textual data is passed through an LSTM/Bi-LSTM model for sequential contextual learning.

F. SMOTE Balancing

SMOTE is applied to handle class imbalance and improve minority class prediction.

G. Classification

The extracted features are classified using hybrid classifiers integrating LSTM, Random Forest, and Logistic Regression.

H. Explainability Layer

The system highlights suspicious indicators such as unrealistic salary offers, missing company information, suspicious email domains, and excessive promotional language.

I. Output Generation

The system predicts whether the job posting is a Real Job or Fake Job.

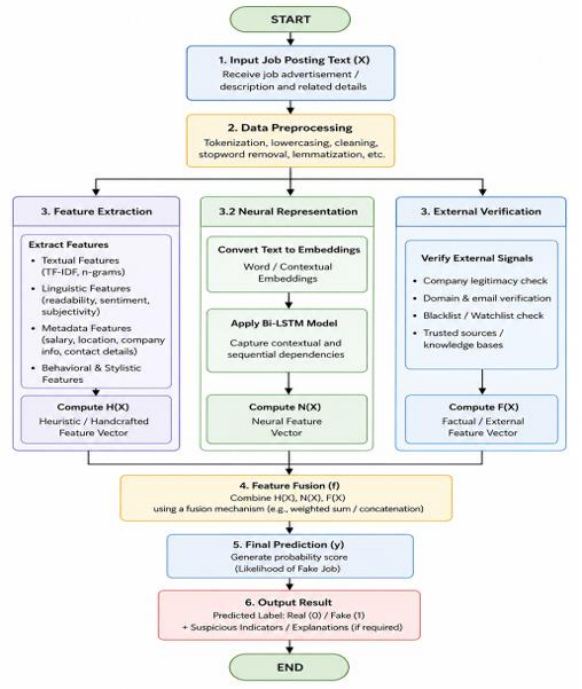

Figure 1: Proposed Hybrid Deep Learning Framework for Fake Job Detection Using NLP and Bi-LSTM

VI. MATHEMATICAL MODEL

Fake job detection can be represented as a binary classification problem. Let X = {x1, x2, x3, ..., xn} represent the input job posting text sequence. The overall prediction function is represented as:

$$y = f(H(X), N(X), M(X)) \quad (I)$$

Where H(X) = heuristic and metadata features, N(X) = neural contextual representation, M(X) = metadata verification component, and y ∈ {0, 1} where 0 = Real Job and 1 = Fake Job. The hybrid prediction model is expressed as:

$$y = \sigma(WhH(X) + WnN(X) + WmM(X) + b) \quad (II)$$

Where Wh, Wn, Wm are learnable weight matrices, b is bias, and σ is the sigmoid activation function. The neural contextual representation is obtained using Bi-LSTM:

$$N(X) = BiLSTM(Embedding(X)) \quad (III)$$

VII. ALGORITHM

Algorithm: Hybrid Fake Job Detection Using LSTM and NLP

Input: Job posting text X

Output: Predicted label y ∈ {0, 1}

Steps:

1. Receive input job advertisement text.
2. Perform preprocessing and tokenization.
3. Extract textual and metadata features.
4. Apply TF-IDF and word embeddings.
5. Balance the dataset using SMOTE.
6. Pass embeddings through Bi-LSTM model.
7. Combine heuristic, metadata, and neural features.
8. Perform classification using hybrid classifier.
9. Generate prediction result.
10. Display suspicious indicators and classification output.

Fig. 1: Proposed Hybrid Deep Learning Framework for Fake Job Detection Using NLP and Bi-LSTM

VIII. CONCLUSION

This paper reviewed existing fake job detection techniques based on Machine Learning, Deep Learning, NLP, and hybrid approaches. The study shows that advanced models such as LSTM and Bi-LSTM provide improved contextual understanding and better detection accuracy compared to traditional methods. Hybrid frameworks integrating linguistic analysis, metadata verification, and neural models further enhance system reliability and robustness.

Despite these advancements, challenges such as dataset imbalance, lack of explainability, high computational complexity, and limited real-time applicability still exist. Future research should focus on developing lightweight, explainable, and scalable fake job detection systems capable of performing accurate real-time analysis across multiple recruitment platforms.

REFERENCES

[1] S. Ramya, J. Sravani, K. Nandini, D. Mahesh, and Y. Nikhil, "Fake Job Prediction Using Machine Learning Algorithms," Int. J. Eng. Res. Sci. Tech., vol. 21, no. 2, pp. 2000–2005, 2025.

[2] R. Kumar and P. Gupta, "Fake Job Post Detection Using Machine Learning," Int. J. Eng. Res., vol. 10, no. 1, pp. 54–62, 2025.

[3] A. Shankar and D. Patel, "Detection of Fraudulent Job Ads Using ANN and ML Techniques," Int. J. Res. Publ. Rev., vol. 11, no. 3, pp. 112–119, 2025.

[4] P. K. Reddy and S. Thomas, "LSTM-Based Approach for Detecting Fake Job Posting," Int. J. Futu. Mod. Res., vol. 4, no. 2, pp. 45–52, 2025.

[5] M. Roy and A. Mazumdar, "Deep Convolutional Neural Networks for Fake Job Detection," Foundry J., vol. 13, no. 1, pp. 88–96, 2025.

[6] S. Gupta and L. Dave, "Fake Job Posting Detection Using TF-IDF and ML Models," IRJMETS, vol. 6, no. 4, pp. 221–229, 2025.

[7] H. Rahman and S. Basu, "Fraudulent Online Job Advertisement Detection Using Deep Learning," IJIRSET, vol. 12, no. 10, pp. 1405–1415, 2024.

[8] D. Kulkarni and F. Ahmed, "A Hybrid ML–NLP Pipeline for Fake Job Ad Detection," Int. J. Intell. Comput., vol. 9, no. 2, pp. 55–64, 2024.

[9] Y. Zhang and H. Lin, "Transformer-Based Fraudulent Job Posting Detection," IEEE Access, vol. 12, pp. 22001–22012, 2024.

[10] P. Shah and K. Mehta, "Real-Time Fake Job Filtering System Using ML," Int. J. Comput. Appl., vol. 182, no. 31, pp. 33–40, 2024.

[11] A. Nashaat, H. El-Bakry, and C. Alken, "Explainable AI for Fake Job Detection Using SHAP," Int. J. Adv. Comput. Sci., vol. 13, no. 4, pp. 120–129, 2023.

[12] R. Verma and J. Goyal, "Machine Learning Models for Job Fraud Identification," Int. J. Inf. Technol., vol. 15, no. 1, pp. 78–86, 2023.

[13] S. Yadav and N. Arora, "Ensemble Learning for Fake Job Detection," Int. J. Comput. Sci. Mobile Comput., vol. 12, no. 9, pp. 44–53, 2023.

[14] H. Prajapati and D. Patel, "NLP-Based Feature Engineering for Scam Job Detection," Int. J. Data Sci., vol. 7, no. 3, pp. 149–158, 2023.

[15] T. Johnson and M. Smith, "Detecting Online Recruitment Scams Using Multimodal Learning," IEEE Trans. Inf. Forensics Security, vol. 17, pp. 889–899, 2022.

[16] R. Sultana, M. Rahman, and M. Haque, "Job Scam Detection Using LSTM Networks," in Proc. ICCIT, 2022, pp. 560–566.

[17] N. Patel and R. Shah, "Fake Job Advertisement Detection Using Gradient Boosting," Int. J. Comput. Eng. Res., vol. 10, no. 6, pp. 101–110, 2022.

[18] A. Jain and P. Srivastava, "Fake Employment Posting Detection Using NLP," Procedia Comput. Sci., vol. 193, pp. 105–112, 2021.

[19] S. Mishra and K. Singh, "Job Scam Detection Using Deep Learning Models," Int. J. Artif. Intell. Tools, vol. 30, no. 5, pp. 215–226, 2021.

[20] K. Sharma and S. Verma, "Application of Logistic Regression in Predicting Job Posting Authenticity," Int. J. Eng. Res. Technol., vol. 10, no. 3, pp. 58–65, 2021.

[21] R. Choudhary and N. Gupta, "Classification of Fraudulent Job Postings Using Text Mining Techniques," Int. J. Emerg. Technol. Eng. Res., vol. 9, no. 5, pp. 101–108, 2021.

[22] A. Prajapati and H. Mehta, "Hybrid ML Model for Job Scam Detection," Int. J. Innov. Res. Comput. Commun. Eng., vol. 9, no. 7, pp. 67–75, 2021.

[23] P. Ravula and S. Kanagala, "Predicting Fake Job Postings Using Supervised Learning Models," Int. J. Sci. Res. Comput. Sci., vol. 7, no. 4, pp. 23–30, 2020.

[24] H. Gupta and R. Bansal, "Analysis of Machine Learning Techniques for Fake Content Detection," J. Inf. Security Appl., vol. 52, pp. 102–115, 2020.

[25] A. Kumar, R. Sharma, and S. Jain, "Detection of Fake Job Postings Using Machine Learning Algorithms," Int. J. Comput. Appl., vol. 180, no. 43, pp. 1–7, 2019.

[26] V. Yadav and A. Arora, "Identifying Fake Online Job Offers Using Machine Learning," Int. J. Comput. Sci., vol. 14, no. 2, pp. 155–162, 2019.

[27] P. Jain and R. Srivastava, "Detection of Fraudulent Job Postings Using ML," Int. J. Adv.

Comput. Commun. Eng., vol. 8, no. 11, pp. 300–306, 2018.

[28] M. Singh and D. Gupta, "Text Analytics-Based Fake Job Offer Identification," Int. J. Data Min. Knowl. Manage., vol. 13, no. 1, pp. 14–22, 2018.

[29] D. Srivastava and M. Goyal, "Early Detection of Online Recruitment Fraud Using NLP," Int. J. Web Eng., vol. 6, no. 2, pp. 45–55, 2017.

[30] A. Verma and R. Mehta, "Identifying Job Fraud Using Supervised Machine Learning," Proc. Int. Conf. Comput. Intell., pp. 89–96, 2017.