Improving Online Job Authenticity Detection using Deep Learning and Focal Loss

1st K. Akhila
Department of Information Technology
BVRIT Hyderabad College of Engineering for Women
Hyderabad, India
kolipakaakhila15@gmail.com

2nd Mukhtar Ahmad Sofi
Department of Computer Science
BVRIT Hyderabad College of Engineering for Women
Hyderabad, India
sofimukthar@bvrithyderabad.edu.in

3rd Ch. Likitha Reddy
Department of Information Technology
BVRIT Hyderabad College of Engineering for Women
Hyderabad, India
likithachinthamredddy@gmail.com

4th A. Sarayu
Department of Information Technology
BVRIT Hyderabad College of Engineering for Women Hyderabad, India
arramshettysarayu@gmail.com

5th M. Sannitya
Department of Information Technology
BVRIT Hyderabad College of Engineering for Women Hyderabad India
mssannitya@gmail.com

2024 International Conference on Data Science and Network Security (ICDSNS) | 979-8-3503-7311-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/ICDSNS62112.2024.10691323

***Abstract*—The rise of fake job postings in online job marketplaces presents a significant challenge for job seekers, necessitating robust detection methods. Existing methods often struggle to adapt to evolving fraudulent tactics, leading to suboptimal accuracy. This project introduces a novel Deep Learning approaches which are Recurrent Neural Network (RNN) variants leveraging Bi-LSTM and Bi-GRU architectures to enhance fake job detection and used a loss function for removing the imbalanced class distribution which is called "Focal Loss". Using the EMSCAD dataset, we trained and evaluated the performance of the proposed model against state-of-the-art methods, aiming for broader accessibility through integration into a user-friendly webserver for real-time job posting analysis which will help jobseekers to identify whether the job posting is Genuine/Fake. In this project we got accuracy of value 98.6% with precision of 85%, recall of 84% and f1-score of 85%.**

Keywords—fake job, deep learning, prediction, LSTM, GRU, focal loss.

I. INTRODUCTION

The contemporary job marketplace has indeed undergone a remarkable transformation charged by the rapid evolution of social communication and technological advancements[6]. There's been a surge in platforms and websites designed to link people looking for work with companies hiring, giving job seekers more options than ever before in different fields and industries. The digital era has revolutionized the way individuals search for and apply to jobs, making the process more efficient and accessible than ever before. However, amidst this progress, a serious problem has emerged: the pervasive rise of fake job postings. These deceptive listings pose a significant threat to job seekers, as they can ruin their efforts and subject them to potential scams [1]. The perpetrators behind fake job postings often exploit the vulnerability of individuals seeking employment, using enticing job descriptions and promises of high salaries to lure unsuspecting candidates into their traps[7]. Once engaged, victims may find themselves deceived, either by providing sensitive personal information or falling victim to financial schemes. These fraudulent activities targeting job seekers not only take advantage of them but also erode trust in genuine hiring procedures[2]. In the dynamic landscape of online job platforms, the persistent challenge of fake job postings has prompted the development of various models and algorithms to address this pressing issue. Traditional machine learning (ML) algorithms have been employed with varying degrees of success, yet the ever-evolving strategies of those behind fraudulent activities demand a more sophisticated approach [4]. In parallel, deep learning algorithms have demonstrated remarkable capabilities in pattern recognition and anomaly detection, making them promising candidates for improving the accuracy of fraudulent job detection.

Several studies have already explored the effectiveness of both traditional ML algorithms and deep learning architectures in distinguishing genuine job listings from deceptive ones[3]. These comparative analyses have shed light on the strengths and limitations of each approach. However, recognizing the need for a comprehensive solution that transcends the confines of individual methodologies. While existing models often rely solely on either traditional ML algorithms or deep learning methods, our deep learning approach takes a step beyond the comparative analysis proposing a solution for data balancing using loss function. Our holistic solution that overpass the gap between traditional ML and deep learning as shown in Fig.1.

The emphasis on hybrid deep learning acknowledges the complementary nature of these approaches, offering a adaptive framework for tackling the challenges posed by fake job detection. As we navigate through this project, we anticipate contributing not only to the ongoing discourse on deceptive job practices but also setting a precedent for the efficacy of hybrid deep learning models in addressing complex issues within the realm of cybersecurity and online integrity. The rise of fake job postings represents a significant threat to the integrity of the job market and the well-being of job seekers. Identifying and combatting these fraudulent listings requires a concerted effort from all stakeholders involved. By remaining vigilant, leveraging technology, and fostering collaboration, we can work towards creating a safer and more transparent job marketplace for all with this project.

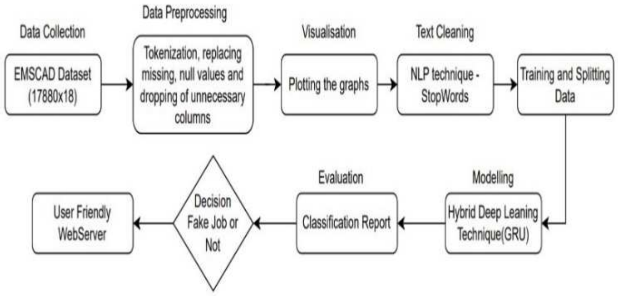

Fig. 1. Graphical Representation of Fake Job Detection using Deep Learning

A. Motivation and Contribution

Despite concerted efforts over the years, the arsenal of machine learning methods employed for fake job posting detection has encountered limitations. These approaches, while initially promising, have often been confined by their reliance on static feature sets. Their inability to swiftly adapt to the constantly evolving tactics employed by perpetrators of fake job postings has resulted in compromised detection accuracy. As a consequence, the prevalence of false positives and negatives has surged, exacerbating the challenges faced by job seekers in navigating the intricate web of job marketplaces. In response to this pressing predicament, our project emerges as a beacon of innovation, introducing a groundbreaking solution in the form of a hybrid Deep Learning method. This novel approach leverages the capabilities of advanced deep architectures to redefine the landscape of fake job posting detection. By transcending the limitations of traditional methodologies, our initiative aims to elevate accessibility and substantially enhance the accuracy of detecting fraudulent job listings. At the core of our project lies the development of a robust and sophisticated model. This model, meticulously designed and trained, is envisioned not merely as an abstract concept but as a practical tool to be deployed as a user-friendly webserver. This user-friendly interface empowers job seekers by swiftly and decisively distinguishing between genuine job openings and deceptive ones, thereby fortifying their defenses against potential exploitation within the complex realm of job marketplaces. To validate the efficacy and reliability of our proposed method, our journey includes rigorous experimentation and assessment. Leveraging the industry-standard EMSCAD dataset, our experiments are meticulously designed to comprehensively evaluate the performance of our approach. Through a meticulous performance analysis, we aim to benchmark our novel hybrid Deep Learning method against existing approaches, showcasing its superiority in mitigating the proliferation of fake job postings while minimizing false outcomes. Ultimately, the realization of our endeavor culminates in the deployment of our refined model onto a webserver accessible to job seekers worldwide. This tangible tool serves as a bulwark against the rising tide of fraudulent job postings, championing the cause of job seekers by offering an efficient, reliable, and trustworthy solution within the competitive and often treacherous job marketplace. Our initiative stands as a testament to the commitment to restore integrity and credibility in the digital realm of job searching, empowering individuals to navigate their career paths with confidence and assurance.

II. Related Work

Deep learning, a powerful subset of machine learning, has significantly advanced various scientific and technological fields by enabling systems to learn complex patterns [10, 15]. Various power deep learning algorithms such as Convolutional Neural networks [24], Recurrent Neural Networks [15], and Attention models [18] have been utilized in image, text, and sequence data. In rapidly evolving job market, technological advancements have provided job seekers with unprecedented access to diverse employment opportunities, response to the rising number of job seekers, there has been an unfortunate increase in fraudulent activities, where scammers exploit the situation by posting fake job opportunities [22]. Recognizing this challenge, developers have come up with a proactive solution. They are working on creating predictive models that can identify and flag fake job postings. This initiative aims to empower job seekers by providing them with tools to be aware of potential job scams and make informed decisions in their job search. Table-1 gives the summary of existing methods for fake job detection.

Detecting employment scams will help job-seekers identify genuine job offers from legitimate companies. To address this issue, various machine learning algorithms have been suggested. Experimental findings show that the Random Forest classifier performs better by achieving an accuracy of 98.27%. While [2] utilizes the EMSCAD dataset, experimenting with both traditional machine learning algorithms (SVM, KNN, Naive Bayes, Random Forest, MLP) and a deep learning model (Deep Neural Network). Results indicate that the Random Forest Classifier emerges as the top performer among traditional methods, boasting a high classification accuracy. Remarkably, the Deep Neural Network achieves a competitive accuracy rate of 98%. While in [3] a new method is introduced for identifying fraudulent job postings using Bidirectional LSTM networks. Experimental results demonstrate that the model exhibits higher precision and recall rates compared to alternative techniques. This superiority is attributed to the Bi-LSTM's capacity to capture complex patterns and contextual details accurately, achieving a precision of 98%. [4] introduces an-automated tool for detecting fraudulent job postings on the internet using machine learning. In this the experimental results favours with Random Forest Classifier with accuracy of 96%.[5] addresses the global issue of job scams using the EMSCAD dataset, to further enhance the model's performance, focusing on their impact and the challenge of identifying fake job posts. Employing a Linear Regression achieved a notable 97% accuracy. [6] introduces an automated tool for detecting fraudulent job postings on the internet using machine learning. It employs various classifiers, comparing their results to identify the most effective employment scam detection model.

TABLE I. Summary of the existing methods for job authenticity assessments.

Author Name	Algorithm	Precision	Recall	F1 Score	Accuracy

Nagaraju et al., 2021 [1]	Random Forest	0.67	0.64	0.72	0.96
Akshata et al., 2022 [2]	Random Forest	1.00	0.54	0.70	0.97
Pillai et al., 2023 [3]	Bi-LSTM	0.89	0.83	0.86	0.98
Amaar et al., 2022 [4]	Linear Regression	1.00	0.39	0.56	0.97
Shawni et al., 2020 [5]	Ensemble Classifier (Random Forest Classifier)	-	-	0.97	0.98
Nguyen et al., 2021 [6]	SVM	0.89	0.89	0.56	0.92
Sultana et al., 2021 [7]	Decision Tree	0.93	0.95	0.93	0.96

While [7] employs various data mining techniques and classification algorithms, including KNN[19], decision trees, support vector machines, naive Bayes classifier, random forest classifier, multilayer perceptron, and deep neural networks, to determine whether a job posting is genuine or fraudulent. Among these, the deep neural network classifier stands out for its exceptional performance in this classification task, achieving an accuracy of approximately 98% in identifying fraudulent job postings. While [23] provides a comparative analysis of classifier performance, focusing on deep learning and conventional machine learning techniques. It concludes that among the conventional machine learning methods examined, in terms of deep learning models, the DNN (fold 9) and Deep Neural Network exhibit the highest average classification accuracy. In [11], the suggested method attained an accuracy of 98.27% through the utilization of gradient boosting, surpassing the performance of existing techniques by a significant margin. [12] is a case study study investigated the effectiveness of organization, job description, and compensation type characteristics in detecting online job recruitment fraud, finding that XGB-based models outperformed others. The results of the [13] show that the Random Forest classifier performed exceptionally well compared to other classifiers, achieving an accuracy rate of 97% in classification tasks. The utilization of the random forest classifier in [14] resulted in a notable accuracy rate of 98%, surpassing other algorithms and reinforcing the commitment to enhancing online hiring practices by mitigating the risks of fraud and deception in employment. [25] addresses the significant challenge of predicting fraudulent job postings achieving an impressive classification accuracy of approximately 98% using a deep neural network classifier. In [16] deploying and maintaining a Bidirectional LSTM model for fake job detection involves careful evaluation, deployment, and ongoing monitoring. The study [17] compared SVM and gradient-boosting algorithms for fake job detection, with SVM achieving the highest accuracy after parameter tuning. Future research could explore data class balancing, additional features, ensemble algorithms, and cross-validation methods for improved generalization.

III. Materials and Methods

A. Dataset and Features

The first step towards developing a model is to collect the dataset for online job authenticity detection. In this work, we used the benchmark Employment Scam Aegean Dataset (EMSCAD) dataset. It contains 17,880 genuine job advertisements. Following are the 17 features in EMSCAD dataset:

- Job id: A unique identifier for a job posting.
- Title: A official designation or title for a specific position.
- Location: The geographical area where the company is located.
- Department: The organizational unit within the company to which the job pertains.
- Salary range: Range of salary offered for the job posting.
- Company profile: A concise description providing an overview of the company.
- Description: A comprehensive outline detailing the job responsibilities and requirements.
- Requirements: A list enumerating the essential skills and qualifications necessary for the job.
- Benefits: The perks or advantages offered to employees by the company.
- Telecommuting: A binary indicator denoting whether the job allows for telecommuting.
- Has company logo: A binary variable indicating whether the company logo is included in the job posting.
- Has questions: A binary variable indicating whether the job posting contains questions for applicants.
- Employment type: Specifies whether the job is full-time and any required experience.
- Required experience: The level of experience necessary for candidates applying for the job.
- Required education: The minimum educational qualification required for the job.
- Industry: The sector or industry to which the job belongs.
- Function: The specific role or function of the job within the company.

B. Proposed Method

The proposed Job Authenticity Assessment System leverages advanced deep learning techniques, primarily Bidirectional Gated Recurrent Units (Bi-GRUs), for the identification of fake jobs as shown in figure-2. The system follows a streamlined process involving data acquisition, pre-processing, and model training on EMSCAD dataset, in addition to this benchmark dataset EMSCAD. The raw textual data, sourced from EMSCAD, undergoes essential pre-processing steps such as tokenization and vectorization to facilitate feature extraction. The dataset is meticulously divided into three subsets: training, validation, and testing, guaranteeing an equitable distribution of real and fraudulent job postings to tackle class imbalance efficiently. We used various deep learning architectures with different configurations and hyper parameters to improve the accuracy of online job authenticity detection.

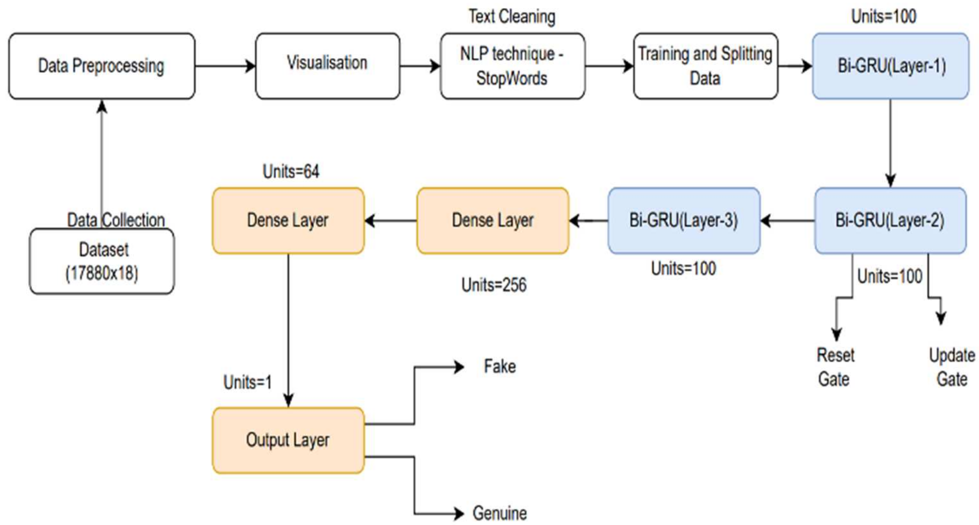

Fig. 2. Proposed System of Online Job Authenticity Detection

C. Bidirectional Long Short-Term Memory

Bi-directional Long Short-Term Memory (Bi-LSTM) enhances LSTM networks which follows RNN architecture by incorporating information from both past and future time steps, thus expanding its functionality. [8]. LSTMs, a variant of recurrent neural networks, are tailored to efficiently handle sequential data. Their proficiency makes them ideal for tasks like natural language processing (NLP), time series analysis, and speech recognition. Key Concepts of Bi-LSTM:

LSTM (Long Short-Term Memory) networks, created by Hochreiter and Schmidhuber, are sophisticated sequential neural networks within deep learning. They address limitations found in traditional RNNs and other machine learning algorithms. Implementation of LSTM in Python is facilitated by the Keras library.

LSTM models address this issue by extending their memory, allowing them to grasp and understand long-term dependencies within inputs. This memory extension empowers LSTM models with the capability to retain and learn information over extended durations. In the context of LSTMs, their memory is referred to as a "gated" cell. The term "gate" reflects the model's decision-making ability to either retain or discard specific memory information. This gating mechanism enables [26] LSTMs to make informed choices about what information is valuable for long-term retention. During the training process, weight values are assigned to information, guiding the model in determining whether to preserve or discard it. Consequently, an LSTM model excels at capturing essential features from inputs and retains this information for an extended period. The decision-making process, driven by assigned weight values, ensures that the model learns to discern the significance of information, optimizing its ability to preserve or remove information based on its importance.

1) Gates in LSTM: LSTMs use three types of gates to control the flow of information:

2) Forget Gate: Decides which information from the previous state to discard and is formulated as (1).

3) Input Gate: Determines which information to update in the cell state and is formulated as (2).

4) Output Gate: Controls which information to output from the current state and is computed as (3).

$$Forget\ Gate, f_t = \sigma\left(X_t * U_f + H_{t-1} * W_f\right) \tag{1}$$

$$\text{Ινπυτ Γατε}, i_t = \sigma(X_t * U_i + H_{t-1} * W_i) \tag{2}$$

$$Output\ Gate,\ o_t = \sigma(X_t * U_o + H_{t-1} * W_o) \tag{3}$$

$$\text{Χελλ Στατε}, c_t = f_t * c_{t-1} + i_t * c_t' \tag{4}$$

$$c_t' = \sigma_c(X_t * U_c + H_{t-1} * W_c) \tag{5}$$

$$\text{Ηιδδεν Στατε}, h_t = o_t * \sigma_c(c_t) \tag{6}$$

Where,
σ represents $sigmoid\ function, \sigma_c$ denotes $tanh,\ X_t$ zepresents input to the current timestamp, H_{t-1}represents hidden state at the previous timestamp, U_f represents weight associated with input, W_f represents Weight matrix associated with hidden state, U_i represents weight matrix of input, $W_i\ represents$ weight matrix of input associated with hidden

state U_o represents Weight matrix of output, W_o represents Weight matrix of output associated with hidden state.

Bidirectional LSTMs analyze input sequences in two directions: forward and backward. This enables the model to gather insights from both past and future contexts simultaneously at each time step.

The LSTM architecture comprises two LSTM layers: one processes the sequence from start to finish, while the other processes it in reverse. Typically, the outputs from both directions are merged or combined to generate the ultimate output.

D. Bidirectional GRU

Bidirectional Gated Recurrent Units (Bi-GRU) is an extension of the Gated Recurrent Unit (GRU) architecture.

1) GRU (Gated Recurrent Unit): GRU is a variant of Recurrent Neural Networks architecture that is similar to LSTM (Long Short-Term Memory) [9]. It was introduced to simplify the architecture of traditional LSTMs. Like LSTM, GRU (Gated Recurrent Unit) is crafted to grasp prolonged relationships within sequential data. It achieves this by employing a gating mechanism, which regulates how information flows within the network.

2) *Gates in GRU:* There are two main types of gates in a GRU as opposed to three gates in an LSTM :

a) *Reset Gate:* The Reset Gate Manages the relevance of past information for the current context i.e., the hidden state and is computed as (7).

b) *Update Gate:* The Update Gate Controls the blend of old and new information i.e., the hidden state and is computed as (8).

$$Reset\ Gate, r_t = \sigma(X_t * U_r + H_{t-1} * W_r) \quad (7)$$

$$Update\ Gate, u_t = \sigma(X_t * U_u + H_{t-1} * W_u) \quad (8)$$

$$Candidate\ Hidden\ State, H_t' = \sigma_c(X_t * U_g + (r_t . H_{t-1}) * W_g) \quad (9)$$

$$Hidden\ State, H_t = U_t . H_{t-1} + (1 - U_t) . H_t' \quad (10)$$

Where, X_t represents input to the current timestamp, H_{t-1} represents a hidden state at the previous timestamp, U_r represents Weight associated with reset gate, W_r represents Weight matrix associated with reset state, U_u represents Weight associated with update gate, W_u represents Weight matrix of input associated with update state.

Bidirectional GRU analyzes input sequences in both forward and backward directions, merging insights from both. [21] This architecture involves two GRU layers: one processes sequences from start to finish, the other in reverse. Their outputs are usually merged to create the final output. This bidirectional method enables the model to grasp context from both past and future steps, enhancing its understanding of sequential data.

E. Handling data Imbalance in job authenticity assessment

From this Bi-GRU with 3 layer we have got good accuracy, but then it contains imbalance data as shown in figure-3 i.e., 17014 genuine posting and 866 fake postings. So, to handle this problem we have used the focal loss function.

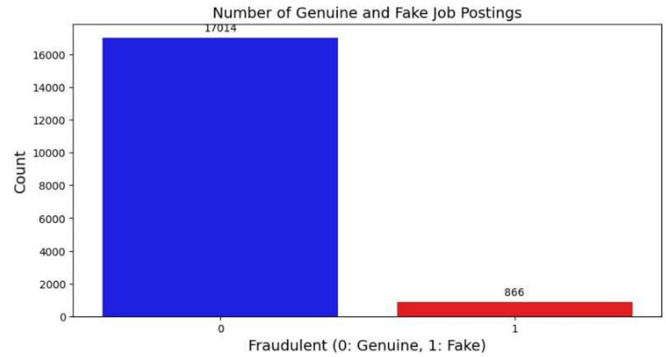

Fig. 3. Graph of number of genuine and fake job postings

Focal Loss function is a specialized loss function primarily used in tasks involving imbalanced datasets, such as object detection or segmentation. It aims to address the issue of class imbalance by assigning higher weights to misclassified examples belonging to the minority class, thereby focusing more on difficult-to-classify examples. This helps in improving the performance of models trained on imbalanced datasets by penalizing misclassifications of minority classes more heavily, thus emphasizing learning from challenging instances. Also it helps stabilizes the training process especially when dealing with high imbalance data.

Focal loss function introduces a tunable focusing parameter it is denoted by gamma and alpha which regulates the reduction speed of the loss for accurately classified examples. [20]. This gamma also allows fine tuning the balance between emphasizing hard examples and maintaining stability during training. It mostly uses the alpha value in between 0 and 1. Gamma is a crucial parameter in Focal Loss that determines how quickly the loss decreases as the model becomes more confident in its predictions. By setting gamma values higher than 0, the loss for accurately classified examples is reduced proportionally. γ adjusts the emphasis on hard examples during training, with higher values prioritizing difficult cases, aiding in addressing class imbalance, while balancing the risk of over-fitting.

$$FL(p_t) = -\alpha_t (1 - p_t)^{\gamma} \log(p_t) \quad (11)$$

Where p_t denotes Predicted probability of true class, α_t represents Balanced factor, γ represents Focusing parameter that controls rate of at which the modulating factor decreases.

IV. Experimental Setup And Results

A. Evaluation Metrics

The evaluation metrics are present in the library sklearn. metrics. The different evaluation metrics which are precision, recall, f1 score, and accuracy.

1) Precision: Precision assesses the model's capability to correctly identify true positives and avoid false positives.

2) Recall: Recall evaluates the model's effectiveness in capturing all positive instances while minimizing false negatives.

3) F1 Score: F1-Score provides a balanced measure between precision and recall, combining both metrics into a single value.

4) *Accuracy:* Accuracy gauges the overall correctness of the model's predictions across all classes or categories.

$$Precision = \frac{True\ Positives}{True\ Positives+False\ Positives} \tag{12}$$

$$Recall = \frac{True\ Positives}{True\ Positives+False\ Negatives} \tag{13}$$

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision+Recall} \tag{14}$$

$$Accuracy = \frac{No.of\ Correct\ Predictions}{Total\ No.Of\ Predictions} \tag{15}$$

B. *Experimental Set-Up*

The experiments were conducted on Google Colaboratory, utilizing its high computational resources including 25 GB of RAM and a T4 GPU. Multiple experiments were carried out with various model configurations in an effort To boost the model's accuracy, an initial approach involved utilizing a Bidirectional Long Short-Term Memory (Bi-LSTM) structure. However, in pursuit of further enhancement, experimentation was conducted by implementing a Bidirectional Gated Recurrent Unit (Bi-GRU) architecture. To further enhance the system's robustness and to prevent overfitting, dropout layers, early stopping and model checkpointing mechanisms are strategically introduced.

Subsequently, the issue of class imbalance within the dataset was addressed (discussed in section III.C). We conducted experiments with different gamma values ranging from 0 to 5.

C. *Results*

As we conducted several experiments on our model using Bi-LSTM, Bi-GRU with multiple layers and for handling imbalance of data, focal loss is used. The Table II and Fig. 4 are the results of all the experiments that are performed.

TABLE II. PERFORMANCE METRICS (PRECISION, RECALL, F1 SCORE) OF DIFFERENT MODELS WITH DIFFERENT CONFIGURATIONS.

ALGORITHM	PRECISION	RECALL	F1 SCORE	ACCURACY
BI-LSTM	0.83	0.70	0.76	0.96
BI-GRU WITH SINGLE LAYER	0.76	0.76	0.76	0.97
BI-GRU WITH THREE LAYERS	0.86	0.72	0.84	0.98
BI-LSTM WITH FOCAL LOSS	0.86	0.22	0.35	0.96
BI-GRU WITH FOCAL LOSS	0.85	0.84	0.85	0.986

Fig. 4. Performance Comparison Graph of different algorithms.

When the experiment is done using Bi-LSTM the accuracy is 0.96, with Bi-GRU with single layer it is 0.97 and Bi-GRU with three layers the accuracy is 0.98. Although we achieved a comparable performance using Bi-GRU, to handle the class imbalance a loss function focal loss is used for better generalisation and performance improvement. By using focal loss function with Bi-GRU, the accuracy is 0.986 as shown in Table III.

C. *Comparison with existing methods*

TABLE III. COMPARISON OF EVALUATION METRICS WITH EXISTING METHODS

Author Name	Algorithm	Precision	Recall	F1 Score	Accuracy
Nagaraju et al., 2021 [1]	Random Forest	0.67	0.64	0.72	0.96
Akshata et al., 2022 [2]	Random Forest	1.00	0.54	0.70	0.97
Pillai et al., 2023 [3]	Bi-LSTM	0.89	0.83	0.86	0.98

Amaar et al., 2022 [4]	Linear Regression	1.00	0.39	0.56	0.97
Shawni et al., 2020 [5]	Ensemble Classifier (Random Forest Classifier)	-	-	0.97	0.98
Nguyen et al., 2021 [6]	SVM	0.89	0.89	0.56	0.92
Sultana et al., 2021 [7]	Decision Tree	0.93	0.95	0.93	0.96
Proposed System	Bi-GRU with 3 layers using focal loss function	0.85	0.84	0.85	0.986

V. Conclusion and Future scope

In our study, we introduced a real-time system for detecting fake job postings using deep learning methods, specifically Bi-LSTM and Bi-GRU architectures. Initially, we implemented a Bi-LSTM model and achieved a baseline accuracy of 96%. Later, we explored the Bi-GRU architecture, which showed better performance, reaching 97% accuracy with a single layer. Further experimentation with three layers of Bi-GRU led to a significant improvement, reaching 98% accuracy. To handle class imbalance in the dataset, we introduced focal loss, a specialized loss function. Incorporating focal loss helped alleviate the impact of class imbalance, resulting in a notable performance boost, with accuracy reaching 98.6%.

Looking ahead, we plan to investigate additional features to enhance our Fake Job Detection system's predictive performance. This includes leveraging advanced natural language processing techniques to extract more detailed features from job postings. Additionally, we will explore the use of advanced deep learning architectures like Recurrent Neural Networks (RNNs) to better capture dependencies and sequential patterns in job descriptions. Through the adoption of RNNs, we aim to further improve the model's predictive accuracy.

References

[1] Anita, C. S., Nagarajan, P., Sairam, G. A., Ganesh, P., & Deepakkumar, G. (2021). Fake job detection and analysis using machine learning and deep learning algorithms. Revista Geintec-Gestao Inovacao e Tecnologias, 11(2), 642-650.

[2] Khandagale, Priya, Akshata Utekar, Anushka Dhonde, and S. S. Karve. "Fake Job Detection using Machine Learning." International Journal for Research in Applied Science & Engineering Technology (IJRASET), 10 (5) 1826 (2022).

[3] Pillai, A. S. (2023). Detecting Fake Job Postings Using Bidirectional LSTM. arXiv preprint arXiv:2304.02019.

[4] Amaar, A., Aljedaani, W., Rustam, F., Ullah, S., Rupapara, V., & Ludi, S. (2022). Detection of fake job postings by utilizing machine learning and natural language processing approaches. Neural Processing Letters, 1-29.

[5] Dutta, Shawni, and Samir Kumar Bandyopadhyay. "Fake job recruitment detection using machine learning approach." International Journal of Engineering Trends and Technology 68, no. 4 (2020): 48-53.

[6] Vo, M. T., Vo, A. H., Nguyen, T., Sharma, R., & Le, T. (2021). Dealing with the class imbalance problem in the detection of fake job descriptions. Computers, Materials & Continua, 68(1), 521-535.

[7] Habiba, S. U., Islam, M. K., & Tasnim, F. (2021, January). A comparative study on fake job post prediction using different data mining techniques. In 2021 2nd international conference on robotics, electrical and signal processing techniques (ICREST) (pp. 543-546). IEEE.

[8] Siami-Namini, S., Tavakoli, N., & Namin, A. S. (2019, December). The performance of LSTM and BiLSTM in forecasting time series. In 2019 IEEE International conference on big data (Big Data) (pp. 3285-3292). IEEE.

[9] Wang, Q., Xu, C., Zhou, Y., Ruan, T., Gao, D., & He, P. (2018, December). An attention-based Bi-GRU-CapsNet model for hypernymy detection between compound entities. In 2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM) (pp. 1031-1035). IEEE.

[10] Sofi, M. A., & ArifWani, M. (2021, March). Improving prediction of amyloid proteins using secondary structure based alignments and segmented-PsSm. In 2021 8th international conference on computing for sustainable global development (INDIACom) (pp. 87-92). IEEE

[11] Jagadeesh, C., Kshirsagar, P. R., Sarayu, G., Gouthami, G., & Manasa, B. (2021). Artificial intelligence based Fake Job Recruitment Detection Using Machine Learning Approach. Journal of Engineering Sciences, 12, 0377-9254.

[12] Mehboob, A., & Malik, M. S. I. (2021). Smart fraud detection framework for job recruitments. Arabian Journal for Science and Engineering, 46(4), 3067-3078.

[13] Nandini, T., Chandrika, S. G., Mounika, P., & Kumar, V. S. (2023). Developing a model to detect fraudulent job postings: fake vs. real. Int J Recent Develop Sci Technol, 7, 52-59.

[14] Anbarasu, V., Selvakani, S., & Vasumathi, M. K. (2024). Fake Job Prediction Using Machine Learning. ubiquity, 13(1), 06-14.

[15] Sofi, M. A., & Wani, M. A. (2023). RiRPSSP: A unified deep learning method for prediction of regular and irregular protein secondary structures. Journal of Bioinformatics and Computational Biology, 21(01), 2350001.

[16] Luz, A., & Daniel, S. (2024). Training a Bidirectional LSTM model for fake job detection.

[17] Rofik, R., Hakim, R. A., Unjung, J., Prasetiyo, B., & Muslim, M. A. (2024). Optimization of SVM and Gradient Boosting Models Using GridSearchCV in Detecting Fake Job Postings. MATRIK: Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 23(2), 407-418.

[18] Sofi, M. A., & Wani, M. A. (2022, March). Improving prediction of protein secondary structures using attention-enhanced deep neural networks. In 2022 9th International Conference on Computing for Sustainable Global Development (INDIACom) (pp. 664-668). IEEE.

[19] Guo, G., Wang, H., Bell, D., Bi, Y., & Greer, K. (2003). KNN model-based approach in classification. In On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE: OTM Confederated International Conferences, CoopIS, DOA, and ODBASE 2003, Catania, Sicily, Italy, November 3-7, 2003. Proceedings (pp. 986-996). Springer Berlin Heidelberg.

[20] Lin, T. Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal loss for dense object detection. In Proceedings of the IEEE international conference on computer vision (pp. 2980-2988).

[21] Li, X., Ma, X., Xiao, F., Xiao, C., Wang, F., & Zhang, S. (2022). Time-series production forecasting method based on the integration of Bidirectional Gated Recurrent Unit (Bi-GRU) network and Sparrow Search Algorithm (SSA). Journal of Petroleum Science and Engineering, 208, 109309.

[22] Laboratory of Information and Communication Systems, University of the Aegean, Samos, Greece. EMSCAD Employment Scam Aegean Dataset, 2016. Available online: http://icsdweb.aegean.gr/emscad (accessed on 22 December 2023)

[23] Swetha, K., Reddy, M. T., Sravani, K., Subramanyam, B., & Kumar, S. A. (2023). FAKE JOB DETECTION USING MACHINE LEARNING APPROACH. Journal of Engineering Sciences, 14(02).

[24] Sofi, M. A., & Wani, M. A. (2022). Protein secondary structure prediction using data-partitioning combined with stacked convolutional neural networks and bidirectional gated recurrent units. International Journal of Information Technology, 14(5), 2285-2295.

[25] Singh, V. R., Sampras, P., & Dhage, A. (2023). Fake Job Post Prediction Using Data Mining. Journal of Scientific Research and Technology, 39-47.

[26] Yu, Y., Si, X., Hu, C., & Zhang, J. (2019). A review of recurrent neural networks: LSTM cells and network architectures. Neural computation, 31(7), 1235-1270.