Detecting Fake News Images Using a Hybrid CNN-LSTM Architecture

Dina R. Salem[1], Abdullah A. Abdullah[2], AbdAllah A. AlHabshy[3], Kamal A. ElDahshan[4]
School of Computer Science, Canadian International College – CIC, Cairo, Egypt[1, 2]
Department of Mathematics, Faculty of Science, Al-Azhar University, Cairo, Egypt[3, 4]

***Abstract*—In today's digital world, images have become a double-edged tool in the dissemination of news; as much as they contribute to enriching honest content and communicating information effectively, they are increasingly being used to mislead the public and spread fake news. The ease of manipulating images and taking them out of their original context, or even creating them entirely with advanced techniques, gives them tremendous power in lending false credibility to false narratives, taking advantage of the human eye's tendency to believe what it sees and the image's superior ability to directly evoke emotions. These misleading images, which are often difficult to debunk with the naked eye, spread at lightning speed across digital platforms, allowing fake news to reach and influence large audiences before it can be verified. However, they tend to generate inaccurate reports. This study proposes a model architecture to detect fake news images. Machine learning and deep learning algorithms were used. The deep learning models are used depending on conventional neural nets (CNN), long short-term memory (LSTM) and a hybrid model that combines CNN and LSTM frameworks on Google Cloud. The hybrid model was able to categorize news with better accuracy than using each model individually. The model was tested and trained on a dataset for classifying fake news images. We used different evaluation metrics (precision, recall, F1 metric, etc.) to measure the efficiency of the model.**

Keywords—Fake news images; machine learning; deep learning; cloud computing; CNN; LSTM

I. Introduction

The rapid expansion of social media platforms has dramatically increased the dissemination speed and reach of news content worldwide [1]. While this democratization of information access has many benefits, it has also facilitated the widespread propagation of fake news—deliberately fabricated or misleading information designed to deceive readers [2]. Fake news poses significant threats to public opinion, political stability, and social trust, particularly when it spreads unchecked on online platforms [3].

Detecting fake news has become a critical research challenge in artificial intelligence, with numerous studies focusing on automated methods to identify misinformation. Traditionally, many approaches have concentrated on textual analysis using natural language processing (NLP) techniques and machine learning classifiers [4][5].

However, news articles frequently contain multimodal content, including images and videos, which can be equally manipulated to mislead audiences [6]. Despite this, visual content analysis for fake news detection has received comparatively less attention, even though fake images are often used to support fabricated stories and increase their perceived credibility [7].

Recent advances in deep learning have shown promise in addressing this challenge by leveraging convolutional neural networks (CNNs) to extract spatial features from images and recurrent neural networks, such as long short-term memory (LSTM) networks, to capture temporal or sequential dependencies in data. Hybrid architectures combining CNN and LSTM models have demonstrated improved performance in various multimedia classification tasks by integrating spatial and sequential feature extraction capabilities [8][9].

This study proposes a hybrid CNN-LSTM architecture for detecting fake news images, aiming to leverage the strengths of both models to improve classification accuracy. We evaluate our approach on a publicly available dataset from Kaggle, comparing the hybrid model's performance against standalone CNN and LSTM models. Our results confirm that the hybrid model achieves superior accuracy in identifying fake news images, highlighting the potential of combined deep learning architectures in multimodal misinformation detection.

Moreover, hybrid models generally outperform single-architecture approaches by integrating the strengths of both components. Recent studies indicate that combining CNNs with LSTMs yields better detection rates than traditional methods reliant on separate evaluations of text or images [10]. This synergy enhances not only accuracy but also the interpretability of predictions, fostering user trust.

Additionally, these systems provide flexibility in adapting to various media types, essential as fake news increasingly incorporate multimedia elements [11]. Hybrid models ensure adaptability in addressing diverse forms of misinformation effectively.

To lessen the impact of bogus news before it becomes viral, early detection is essential. Machine learning-backed proactive detection techniques can lessen the negative social and psychological effects of misleading narratives [12].

The study is organized as follows: Section II tackles the literature review in some detail. The proposed model architecture is presented in Section III. Section IV presents the experiments and the results with discussion. The study concludes in Section V.

II. Literature Review

This section reviews recent and relevant literature on fake news detection, focusing on the transition from traditional approaches to deep learning techniques, the emergence of hybrid models such as CNN-LSTM, and the growing role of cloud computing in scalable deployment.

A. *Text-Based Fake News Detection*

Traditional approaches to fake news detection have primarily focused on analyzing the linguistic features and semantic content of news articles. Early methods often relied on feature engineering and machine learning classifiers. For instance, Alhindi, T., Kalita, J., and Sajjad, H. (2020) [13] classified fake news using Support Vector Machines (SVM) and TF-IDF features on the LIAR dataset. While their approach provided reasonable accuracy by capturing word importance, it struggled to understand complex contextual nuances and did not scale well to real-world scenarios due to restricted semantic representation and dataset diversity.

More recently, deep learning models have shown superior capabilities in capturing intricate textual patterns. Gupta, R., et al. (2021) [14] utilized Bidirectional Encoder Representations from Transformers (BERT) for enhanced textual feature extraction and fake news classification. Their model demonstrated high performance on several benchmark datasets by leveraging BERT's contextual embeddings. However, as a purely text-based method, it inherently fails to account for the crucial visual elements that often accompany and influence the perception of news, particularly when key information is conveyed through embedded text or manipulated images.

Another significant line of work explored sequential models for temporal analysis of text. Ma, J., et al. (2019) [15] proposed using Recurrent Neural Networks (RNNs) to model the propagation patterns of rumors over time. While effective in identifying how narratives evolve, their reliance solely on textual content means they cannot detect misinformation where visual cues, such as text embedded in images, play a primary role in conveying the false message, thus limiting their applicability to multimodal misinformation.

B. *Image-Based Fake News Detection*

Parallel to text-based methods, considerable research has focused on analyzing visual content to identify fake news. These methods often aim to detect image manipulation, inconsistencies, or to extract visual cues indicative of deception. Wang, Y., et al. (2022) [16] developed a CNN-based framework for identifying visual inconsistencies in news images, leveraging various image forensic techniques. Their method achieved high accuracy in detecting manipulated images. However, it does not explicitly analyze any textual content present within the image, which is a critical oversight given that many misleading images rely on superimposed text to convey a deceptive narrative rather than pure visual manipulation.

Similarly, Zhang, L., et al. (2020) [17] proposed a deep learning model to classify images as real or fake based on extracted semantic features using pre-trained CNNs (e.g., ResNet). While effective in capturing high-level visual representations, their approach often struggles with cases where the image itself is authentic but carries a false message through external text added to it or is presented in a misleading context that can only be understood through accompanying textual cues. Their method thus has limited utility for detecting fake news where the visual medium's core deceptive element resides within embedded text, rather than the image's inherent authenticity.

Li, S., et al. (2023) [18] focused on identifying manipulated visual elements in news images using advanced CNN architectures and attention mechanisms to highlight suspicious regions. While their approach was robust against various manipulation techniques, its primary limitation lies in its inability to infer the veracity of the news article or message when the image itself is unaltered but the embedded text (or overall context) is misleading.

C. *Multimodal Fake News Detection*

Recognizing the limitations of unimodal approaches, a growing body of research has explored multimodal strategies that integrate information from both textual and visual modalities to enhance fake news detection. These approaches aim to capture the complex interdependencies between different data types.

Liu, X., et al. (2021) [19] proposed a multimodal framework that combines a CNN for image feature extraction and a GRU (Gated Recurrent Unit) for text features, fusing them at a later stage for classification. Their model showed improved performance compared to unimodal baselines by leveraging complementary information. However, their textual analysis was primarily focused on accompanying article text rather than the direct processing of text embedded within the images, which presents distinct challenges due to its unique formatting and often succinct nature.

Pan, W., et al. (2022) [20] developed a co-attention network that learns cross-modal interactions between text and images to better align information for fake news detection. While innovative in capturing intricate relationships, their method might not explicitly distinguish between general image features and the specific textual content within the image, potentially overlooking the direct semantic impact of embedded text. This limits their ability to robustly detect misinformation where the key deceptive element is the overlaid text.

Despite these advancements, a significant challenge remains in effectively integrating deep learning models that can specifically process and understand the unique characteristics of text directly embedded within images for fake news detection. Existing multimodal approaches often treat images as mere visual contexts or focus on detached textual descriptions, leaving a critical gap in analyzing the inherent deceptive power of image-text composites. Our proposed hybrid CNN-LSTM model directly addresses this limitation by meticulously extracting visual features via CNNs and simultaneously processing the sequential, embedded textual information using LSTMs, providing a more granular and effective approach to multimodal fake news detection.

In conclusion, the literature reveals several gaps: limited hybrid model integration, inconsistent usage of scalable deployment platforms, and a lack of comprehensive testing across diverse datasets. These gaps justify the need for a cloudoptimized hybrid CNN-LSTM model that can robustly handle fake news detection at scale.

While these multimodal studies have made significant strides, many do not explicitly focus on the challenging domain of text directly embedded within images, which is a distinct modality requiring specialized processing, particularly given the potential for OCR inaccuracies. The existing literature often treats images as general visual context or focuses on visual manipulation, rather than the intrinsic textual information they carry. This gap forms the primary motivation for our current study, aiming to provide a robust deep learning solution for fake news detection that specifically leverages and integrates textual content extracted from images [21].

III. Methodology

The proposed methodology investigates the most famous state-of-the-art deep learning algorithms for Fake news detection; like deep learning techniques, they have the advantage of their ability to capture semantic features from textual data automatically. The three different combinations of deep neural networks, namely, CNN, LSTM and CNN-LSTM classification, have been performed. The proposed methodology is shown in Fig. 1.

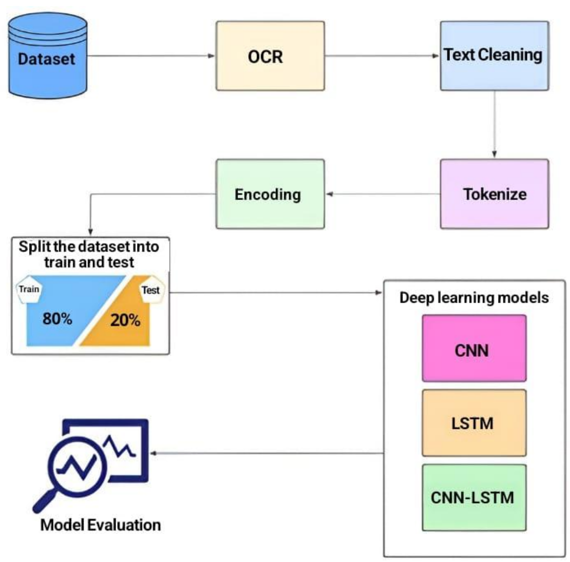


Fig. 1. System architecture.

A. Dataset

The Fake News Classification Image Dataset, publicly available on Kaggle [22]. This dataset comprises both textual and visual information, crucial for multimodal fake news detection.

- Dataset Source: Kaggle (Fake News Classification Image).
- Data Partitioning: The dataset was split into training, validation, and test sets.
  - Training Samples: 3600
  - Validation Samples: 450
  - Test Samples: 255
- Class Distribution: The dataset appears to have a relatively balanced distribution of "Fake" (class 0) and "Real" (class 1) news samples, as evidenced by the support values in the classification reports (e.g., Test set 126 for class 0, 129 for class 1).

B. Data Preprocessing

This effective preprocessing of both image and textual data is paramount for deep learning models. The following steps were applied:

1) Optical Character Recognition (OCR): Textual content embedded within the images was extracted using an OCR technique. This process is critical for converting visual text into a machine-readable format. The approach aligns with methodologies seen in advanced OCR systems, which aim to extract characters from optical images into editable formats, potentially leveraging tools such as PaddleOCR for pseudo ground truth acquisition as discussed in relevant literature [23].

2) Text cleaning: Following OCR extraction, the raw text underwent a meticulous cleaning process to standardize its format and reduce noise, which is crucial for subsequent natural language processing. This involved converting all extracted text to lowercase to ensure uniformity. Additionally, non-alphanumeric characters, such as punctuation and special symbols, were systematically removed to retain only relevant linguistic elements (letters, numbers, and spaces). Finally, multiple consecutive whitespace characters were consolidated into single spaces, and leading or trailing whitespace was trimmed, resulting in a clean and normalized textual representation [24].

3) Tokenization and padding: To prepare the cleaned textual data for input into the neural networks, a two-stage process of tokenization and padding was performed. Initially, a tokenizer was fitted on the entire collection of cleaned extracted text, encompassing data from the training, validation, and test sets. This process built a vocabulary, with a predefined size of 512 words, mapping each unique word to a numerical index. Subsequently, the cleaned text from each dataset (training, validation, and test) was converted into sequences of these numerical indices. To ensure that all input sequences had a uniform length, a max_length was determined by identifying the longest sequence across all datasets. All shorter sequences were then padded with zeros at the end ('post' padding) until they reached this max_length. This standardization is vital for batch processing in deep learning models [25].

4) Label encoding: The categorical labels associated with each image, 'real' and 'fake', were transformed into numerical representations suitable for binary classification. Specifically, 'real' news instances were mapped to the numerical value of 1, and 'fake' news instances were mapped to 0. This encoding converted the target variables into a format directly usable by the model's output layer [26].

Following the completion of the preprocessing stages, the data is divided into testing and training sets in an 80:20 ratio. Following that, the features are suitably reshaped to satisfy the input specifications of the various deep learning architectures (CNN, LSTM and CNN-LSTM models).

C. Model Architectures

This study investigates three distinct deep learning architectures: a Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM) network, and a Hybrid CNN-LSTM model. All models utilize the preprocessed textual data as input.

1) Convolutional Neural Network (CNN): The CNN model is designed for feature extraction from sequential data, leveraging its ability to capture local patterns. The architecture is as shown in Fig. 2:

- Embedding Layer: input_dim=512 (vocabulary size), output_dim=16 (embedding dimensions).
- Conv1D Layer: filters=8, kernel_size=3, activation='relu'.
- GlobalAveragePooling1D Layer: Reduces each feature map to a single value, capturing the most significant features.
- Dropout Layer: rate=0.5 for regularization, preventing overfitting.
- Dense Layer: units=1, activation='sigmoid' for binary classification output.

The model's input shape is configured to handle sequences of varying batch sizes and the determined max_length.

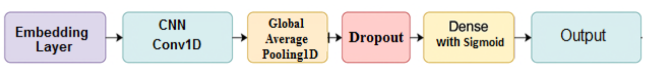


Fig. 2. Workflow of the CNN model.

2) Long Short-Term Memory (LSTM): The LSTM model is particularly suited for capturing long-range dependencies in sequential data. The architecture is as shown in Fig. 3:

- Embedding Layer: input_dim=512, output_dim=16.
- Bidirectional LSTM Layer: units=8 A bidirectional setup processes the sequence in both forward and backward directions, enhancing context understanding.
- Dropout Layer: rate=0.5.
- Dense Layer: units=1, activation='sigmoid'.

The model's input shape is set to accommodate sequences of varying batch sizes and the max_length.

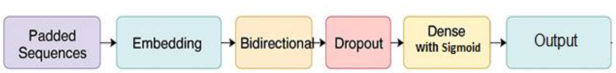


Fig. 3. Workflow of the LSTM model.

3) A Hybrid (CNN-LSTM) model: This hybrid model combines the strengths of both CNNs for local feature extraction and LSTMs for capturing sequential dependencies. This architecture aims to leverage the benefits of both approaches for improved fake news detection as shown in Fig. 4:

- Embedding Layer: input_dim=512, output_dim=128.
- First Conv1D Layer: filters=256, kernel_size=3, activation='relu', padding='same'.
- First MaxPooling1D Layer: pool_size=2.
- Second Conv1D Layer: filters=128, kernel_size=3, activation='relu', padding='same'.
- Second MaxPooling1D Layer: pool_size=2.
- Bidirectional LSTM Layer: units=128, return_sequences=False (outputs only the last hidden state).
- First Dense Layer: units=128, activation='relu'.
- Output Dense Layer: units=1, activation='sigmoid'.

The model's input shape is built to handle sequences of varying batch sizes and the max_length.

Fig. 4. Workflow of the CNN-LSTM model.

D. Training Details

All three models (CNN, LSTM, and Hybrid) were trained using consistent parameters to ensure fair comparison:

- Loss Function: binary_crossentropy was chosen as the loss function, suitable for binary classification tasks.
- Metrics: Model performance was evaluated based on accuracy.
- Epochs: Each model was trained for 25 epochs.
- Batch Size: A batch_size of 32 was used during training.
- Validation Data: The validation set (val_x, val_y) was used to monitor performance and prevent overfitting during training.
- Execution Environment: The training was conducted on Google Colab, leveraging its computational resources to facilitate efficient model training and experimentation.

The training process for each model was initiated, and the total training time was recorded to assess computational efficiency.

IV. Experiments and Results

This section details the experimental setup, presents the quantitative results of the trained models, and discusses their performance in classifying fake news from images with embedded text. The evaluation focuses on comparing the proposed hybrid CNN-LSTM model against standalone CNN and LSTM models.

A. Hyperparameters of Each Model

As indicated in Table I, we examine the key hyperparameters that govern their architecture and training process. These consist of the number and layers types, activation functions, optimizer settings, batch size, dropout rates, and the output layer configurations. The following table summarizes the hyperparameters for each model, highlighting the differences in their architectures and the settings used to train them. This comparison provides insights into how each model was structured and trained to fake news detection task.

TABLE I. COMPARISON BETWEEN HYPEPARAMETERS OF EACH MODEL

Hyperparameter	CNN Model	LSTM Model	Hybrid (CNN-LSTM) Model
Layers	Embedding, Conv1D, GlobalAveragePooling1D, Dropout, Dense	Embedding, Bidirectional LSTM, Dropout, Dense	Embedding, Conv1D, MaxPooling1D, Conv1D, MaxPooling1D, Bidirectional LSTM, Dense, Dense
Number of Layers	5	4	8
Activation Function	ReLU (Conv1D), Sigmoid (Output Dense)	Sigmoid (Output Dense)	ReLU (Conv1D, Dense), Sigmoid (Output Dense)
Optimizer (Adam)	Adam (Learning Rate: 0.001)	Adam (Learning Rate: 0.0005)	Adam (Learning Rate: 0.0001)
Loss Function	Binary Cross-Entropy	Binary Cross-Entropy	Binary Cross-Entropy
Batch Size	32	32	32
Epochs	25	25	25
Dropout	0.5 (after GlobalAveragePooling1D)	0.5 (after Bidirectional LSTM)	Not explicitly used
No. of Filters (CNN)	8 (Conv1D)	N/A	256 (1st Conv1D), 128 (2nd Conv1D)
Kernel Size (CNN)	3 (Conv1D)	N/A	3 (both Conv1D layers)
Pooling Layer (CNN)	GlobalAveragePooling1D	N/A	MaxPooling1D (pool_size=2), used twice
LSTM Units	N/A	8 (Bidirectional LSTM)	128 (Bidirectional LSTM)
Input Shape	16	16	128
Output Layer	(None, max_length)	(None, max_length)	(None, max_length)

B. Evaluation Metrics

To thoroughly assess the performance of each classification model, a suite of widely recognized metrics was employed. These metrics provide insights into different aspects of model accuracy and robustness in distinguishing between "Fake" and "Real" news. The metrics are derived from the Confusion Matrix, which quantifies True Positives (TP), True Negatives (TN), False Positives (FP), and False Negatives (FN) [27][28]:

- Accuracy: Represents the proportion of correctly classified instances out of the total instances.
- Precision: For a given class, it is the ratio of correctly predicted positive observations to the total predicted positive observations. It quantifies the ability of the model to avoid false alarms.
- Recall (Sensitivity): For a given class, it is the ratio of correctly predicted positive observations to all observations in that class. It quantifies the ability of the model to find all positive instances.
- F1-Score: The harmonic mean of Precision and Recall. It provides a single score that balances both precision and recall, especially useful in cases of imbalanced class distributions.
- ROC AUC (Receiver Operating Characteristic Area Under the Curve): Measures the ability of the model to distinguish between classes. A higher AUC indicates better discriminatory power across various classification thresholds.
- PR AUC (Precision-Recall Area Under the Curve): Focuses on the trade-off between precision and recall for different thresholds. It is particularly informative for imbalanced datasets, where it provides a more robust evaluation than ROC AUC.
- MCC (Matthews Correlation Coefficient): A single metric that considers all four values in the confusion matrix. It is generally regarded as a balanced measure which can be used even if the classes are of very different sizes. A value of +1 represents a perfect prediction, 0 an average random prediction, and -1 an inverse prediction.

C. Results Evaluation

In this section, we display the outcomes of the models that were used on the Fake News Classification image Dataset. We used CNN, LSTM and Hybrid models for this dataset. Accuracy, precision, recall, F1-score, confusion matrix, classification report, accuracy curve, and loss curve are among the performance measures. These measurements give a thorough evaluation of each model's performance on the Fake News Classification Image Dataset, providing important information about how well they work for fake news detection.

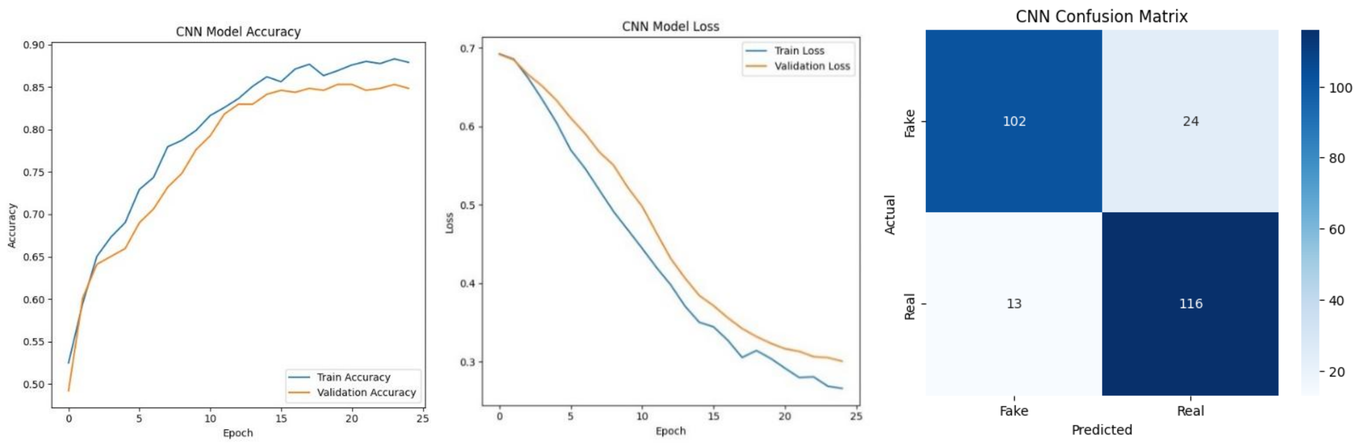
1) CNN Model: Table II, shows the evaluation metrics of the CNN model demonstrating better performance.

TABLE II. EVALUATION METRICS OF CNN MODEL

Metric	CNN Model
Accuracy	0.8667
Precision (Fake)	0.9259
Recall (Fake)	0.7937
F1-Score (Fake)	0.8547
Precision (Real)	0.8231
Recall (Real)	0.9380
F1-Score (Real)	0.8768
ROC AUC	0.9476
PR AUC	0.9559
MCC	0.7403

These results validate the CNN model's capacity to reliably detect fake news.

Fig. 5 represents fake and real image news using the CNN Accuracy Curve in (a), Loss Curve in (b) and Confusion Matrix in (c). As shown in Fig. 5(a), the results of the a model accuracy on training and validation data during training both curves exhibit a steady increase, converging at a high accuracy level, which indicates the model's ability to generalize well. According to the Loss Curve in Fig. 5(b), the decreasing trend in both curves signifies effective model optimization, with the validation loss closely following the training loss, suggesting minimal overfitting. Also in Fig. 5(c), the results of the confusion matrix the high values along the diagonal indicate strong predictive accuracy, while the off-diagonal values represent misclassifications. The intensity of the color reflects the frequency of predictions, highlighting the model's effectiveness in distinguishing between classes.



(a) (b) (c)

Fig. 5. Performance evaluation of the CNN model: (a) Accuracy curve, (b) Loss curve and (c) Confusion matrix.

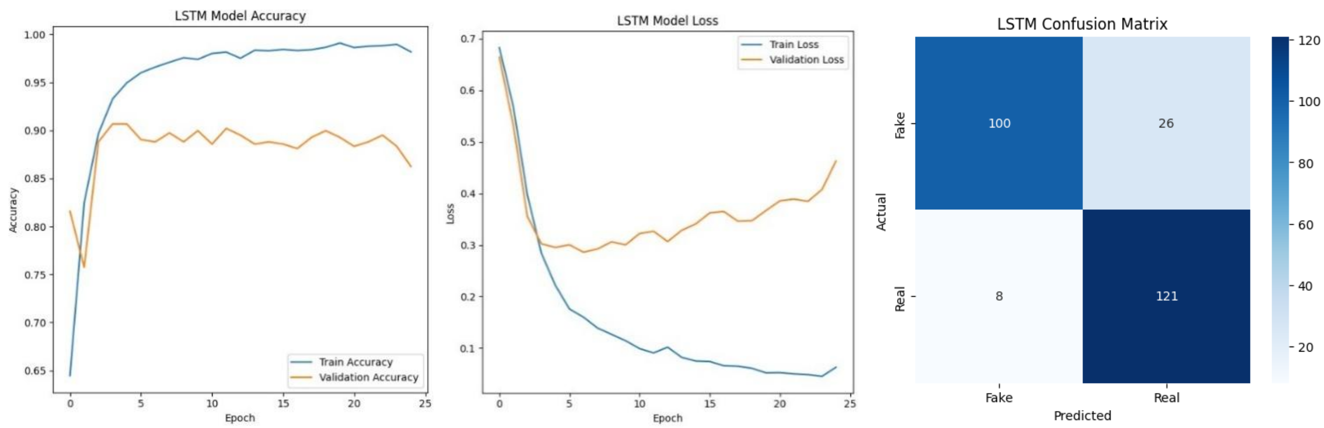
2) LSTM Model: Table III, shows the evaluation metrics of implemented LSTM model. It demonstrated strong classification capabilities.

TABLE III. EVALUATION METRICS OF LSTM MODEL

Metric	LSTM Model
Accuracy	0.8549
Precision (Fake)	0.8870
Recall (Fake)	0.8095
F1-Score (Fake)	0.8465
Precision (Real)	0.8286
Recall (Real)	0.8992
F1-Score (Real)	0.8625
ROC AUC	0.9493
PR AUC	0.9577
MCC	0.7121

Fig. 6 represents fake and real image news using the LSTM Accuracy Curve in (a), Loss Curve in (b) and Confusion Matrix in (c). As shown in Fig. 6(a), the upward trend in both curves indicates the model's improved performance as training progresses, with the training and testing accuracies converging at a high value. According to the Loss Curve in Fig. 6(b), the decreasing trend in both curves signifies effective model optimization, with the validation loss closely following the training loss, suggesting minimal overfitting. Also in Fig. 6(c), the confusion matrix visualizes the model's classification performance across different categories. The strong diagonal dominance indicates high classification accuracy, while the off-diagonal elements represent misclassifications.

3) A Hybrid (CNN-LSTM) Model: Table IV, shows the evaluation matrices of the hybrid model. It demonstrated strong classification capabilities. The model's capacity for generalization was enhanced and overfitting was lessened with the aid of batch normalization and dropout. The hybrid model showed its strong performance in fake news detection tasks by outperforming the baseline accuracy.



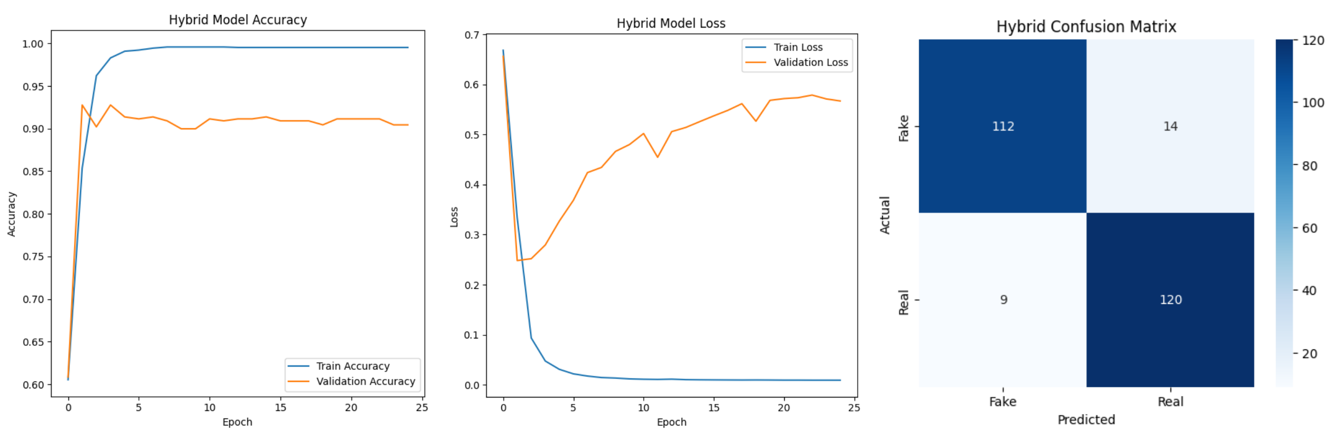
(a) (b) (c)

Fig. 6. Performance evaluation of the LSTM model: (a) Accuracy curve, (b) Loss curve and (c) Confusion matrix.

TABLE IV. EVALUATION METRICS OF THE CNN-LSTM MODEL

Metric	Hybrid Model
Accuracy	0.9098
Precision (Fake)	0.9256
Recall (Fake)	0.8889
F1-Score (Fake)	0.9069
Precision (Real)	0.8955
Recall (Real)	0.9302
F1-Score (Real)	0.9125
ROC AUC	0.9673
PR AUC	0.9713
MCC	0.8201

Fig. 7 represents fake and real image news using the CNN-LSTM Accuracy Curve in (a), Loss Curve in (b) and Confusion Matrix in (c). As shown in Fig. 7(a), this plot displays the training and testing accuracy over multiple epochs. The upward trend in both curves indicates the model's improved performance as training progresses, with the training and testing accuracies converging at a high value. According to the Loss Curve in Fig. 7(b), the graph illustrates the training and testing loss across epochs. The decreasing loss curves reflect effective model learning, with the test loss closely following the training loss, suggesting good generalization with minimal overfitting. Also in Fig. 7(c), the confusion matrix visualizes the model's classification performance across different categories. The strong diagonal dominance indicates high classification accuracy, while the off-diagonal elements represent misclassifications.



(a) (b) (c)

Fig. 7. Performance evaluation of the hybrid (CNN-LSTM) model: (a) Accuracy curve, (b) Loss curve and (c) Confusion matrix.

D. Discussion

This study evaluates and compares the performance of three deep learning models – CNN, LSTM and CNN-LSTM model for fake news detection using the Fake News Classification Image Dataset. The results give important insights into the effectiveness of these models and their suitability for specific scenarios in Artificial intelligence applications.

1) Performance on fake news classification image dataset: For the Fake News Classification Image dataset, the proposed hybrid LSTM-CNN model outperformed the standalone CNN and LSTM models in detecting fake news. While CNN is effective in extracting local features and LSTM captures sequential dependencies, each model has limitations when used independently. By combining their strengths, the hybrid architecture achieved better generalization and higher classification accuracy. On the Fake News Classification Image dataset, the hybrid model reached an accuracy of 90.98%, surpassing CNN (86.67%) and LSTM (85.49%), with improved sensitivity and specificity. These results confirm the hybrid model's effectiveness in handling the complexity of fake news detection.

This improvement is consistent across several critical metrics:

- ROC AUC (0.9673) and PR AUC (0.9713): The significantly higher values for the Hybrid model in both ROC AUC and PR AUC suggest excellent discriminatory power and strong performance even in potentially imbalanced classification scenarios.
- Matthews Correlation Coefficient (MCC): With an MCC of 0.8201, the Hybrid model exhibits a strong correlation between predicted and true labels, indicating a highly reliable classification performance across both classes.
- F1-Scores: The Hybrid model achieved the highest F1-Scores for both "Fake" (0.9069) and "Real" (0.9125) classes, highlighting its balanced ability to maintain high precision and recall for both positive and negative classifications.
- Recall (Fake) and Specificity (Fake): The Hybrid model shows a notable improvement in Recall for the "Fake" class (0.8889) and Specificity (0.8889), which is crucial for effectively identifying actual fake news and minimizing false positives (classifying real news as fake). While the LSTM model had a slightly higher Recall for "Real" news, the Hybrid model demonstrates a better overall balance and a higher true negative rate.

2) Comparative insights: The comparison of results highlights the CNN-LSTM model's robustness, particularly for datasets with complex temporal or sequential patterns, such as the Fake News Classification Image dataset. While the CNN and LSTM models also performed well, their slight limitations in handling temporal dependencies make them less effective than Hybrid in such scenarios.

3) Implications and practical applications: The results clearly highlight the crucial importance of selecting suitable models based on the characteristics of datasets used in fake news detection. The Convolutional Neural Network (CNN)'s ability to handle feature-based patterns makes it a preferable choice in contexts where local spatial relationships within text (such as linguistic constructions and phrases) are critically important. Conversely, the Long Short-Term Memory (LSTM) model, with its superior capacity to capture sequential dependencies, proves valuable for datasets where temporal or contextual patterns across the text play a key role in distinguishing fake news. The hybrid model proved to be the strongest due to its effective utilization of the strengths of both CNN and LSTM. Furthermore, this study underscores the practical feasibility of these models in real-world fake news detection applications, supported by their high performance and adaptability. The insights gained can guide future developments in fake news detection systems, particularly in confronting the ever-increasing volume of malicious content online.

V. Conclusion and Future Work

In order to detect fake image news, this study suggested a hybrid deep learning model that combines the strengths of CNN and LSTM. The convolutional layers are adept at extracting local features and patterns (e.g., n-gram features from text sequences), while the Bidirectional LSTM layers excel at capturing long-range dependencies and contextual information in sequences from both forward and backward directions. This synergy, further bolstered by the integration of OCR-extracted text from images, allows the hybrid model to learn richer and more discriminative representations from the combined textual input derived from multimodal data, leading to superior fake news detection. The model was implemented in a cloud computing environment.

According to the experimental results, the hybrid model achieved a high performance of 90.98% and outperformed the results of running individually the CNN and LSTM models, in terms of accuracy, sensitivity, and specificity. This performance was resulted by the successful extraction of both local features and long-term dependencies made possible by the combination of both models.

While the achieved results in detecting fake news images using a hybrid CNN-LSTM architecture are highly promising, this study is not without certain limitations. Specifically, the Fake News Classification Image Dataset utilized may still present constraints in terms of its diversity, complexity, or its full representation of the constantly evolving deceptive visual content found in real-world scenarios. Future work could explore more challenging, imbalanced, or real-world datasets derived from diverse sources to further assess the robustness of the proposed hybrid model against evolving fake news dissemination tactics. Additionally, given the resource-intensive nature of deep learning models, there is a need to explore lightweight alternatives or model optimization techniques for deployment in resource-constrained environments or on devices with limited computational capacity.

Furthermore, the model's capabilities can be refined to include not only images but also video content that may contain fake information or be AI-generated, addressing the growing challenge of multimodal disinformation.

References

[1] F. Pierri and S. Ceri, “False news on social media: A data-driven survey”, ACMSIGMOD Record, vol. 48, no. 2, pp. 18–27, 2019.

[2] S.Kumar andN. Shah, “False information on web and socialmedia:Asurvey,”, ArXiv, vol. abs/1804.08559, 2018.

[3] M. Choudhary, S. Jha, Prashant, D. Saxena, and A. K. Singh, “A review of fake news detection methods using machine learning,” in 2021 2nd International Conference for Emerging Technology, INCET 2021, 2021.

[4] T. Traylor, J. Straub, Gurmeet, and N. Snell, “Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator,” in Proceedings - 13th IEEE International Conference on Semantic Computing, ICSC 2019, 2019.

[5] Shu, K., Wang, S., and Liu, H, “Beyond news contents: The role of social context for fake news detection”, ACM Transactions on Information Systems, 39(3), 1–31, 2021.

[6] Saadi, S. M., and Al-Jawher, W. A. M, “Image Fake News Prediction Based on Random Forest and Gradient-boosting Methods”, Journal Port Science Research, 6(4), 2023.

[7] Z. Liang, “Fake News Detection Based on Multimodal Inputs,” Computers, Materials and Continua, vol. 75, no. 2, 2023.

[8] Thakre, N., Thakare, V., Sawarkar, A., Mankar, T., and Dhande, S “fake news detection using convolution neural network on social platforms”, Journal of Research Administration, Society of Research Administrators International, 2020.

[9] Jaybhaye, S. M., Badade, V., Dodke, A., Holkar, A., and Lokhande, P, “ Fake News Detection using LSTM based deep learning approach”, ITM Web of Conferences, 56, 03005, 2023.

[10] D. G. Dev, V. Bhatnagar, B. S. Bhati, M. Gupta, and A. Nanthaamornphong, "LSTMCNN: A hybrid machine learning model to unmask fake news," PMC10844048, 2024.

[11] P. Meel and D. K. Vishwakarma, "HAN, image captioning, and forensics ensemble multimodal fake news detection," S0020025521002826, 2021.

[12] Pavan M N, Prasad, P. R., Gowda, T., Vibhakar, T. S., and Shidnal, S, “ Fake News Detection using Machine Learning”, International Research Journal of Engineering and Technology (IRJET), 7(12), 2395-0056, 2020.

[13] Alhindi, T., Kalita, J., & Sajjad, H, “Contextual Fake News Detection Using Machine Learning Techniques”. In Proceedings of the IEEE International Conference on Big Data,2020.

[14] Gupta, R., et al, “ BERT-Based Textual Feature Extraction for Fake News Detection”, Expert Systems with Applications, 178, 115000, 2021.

[15] Ma, J., et al, “Detecting Rumors on Twitter with Tree-structured Recursive Neural Networks”, Journal of Information Science, 45(4), 483-494, 2019.

[16] Wang, Y., et al, “A CNN-based Framework for Visual Misinformation Detection through Image Forensics”, Journal of Visual Communication and Image Representation, 82, 103350, 2022.

[17] Zhang, L., et al,“Deep Learning for Image-based Fake News Detection”, Neurocomputing, 381, 230-241,2020.

[18] Li, S., et al, “Attention-Based Deep Learning for Manipulated Image Detection in News”, Multimedia Tools and Applications, 82(5), 7891-7910, 2023.

[19] Liu, X., et al, “Multimodal Fake News Detection: A Unified Framework with Hierarchical Attention”, Information Sciences, 574, 21-34, 2021.

[20] Pan, W., et al, “Cross-Modal Attention Network for Fake News Detection”, IEEE Transactions on Multimedia, 24, 1-13, 2022.

[21] Lin, S. Y., Chen, Y. C., Chang, Y. H., Lo, S. H., and Chao, K. M, “Text-image multimodal fusion model for enhanced fake news detection”, Science Progress, 107(4), 1-31, 2024

[22] The datasets used during the current study are available in Kaggle https://www.kaggle.com/datasets/vivek1coder/fake-news-classification-image/data

[23] Haoran Wei, Chenglong Liu, et al, “General OCR Theory: Towards OCR-2.0 via a Unified End-to-end Model”, arXiv:2409.01704, 2024.

[24] Vasu Agarwal, H Parveen Sultana, Srijan Malhotra, and Amitrajit Sarkar, “Analysis of classifiers for fake news detection”, Procedia Computer Science, 165:377‑383, 2019.

[25] Pritika Bahad, Preeti Saxena, and Raj Kamal, “Fake news detection using bi-directional lstmrecurrent neural network”, Procedia Computer Science, 165:74–82, 2019.

[26] S.-Y. Lin, Y.-C. Chen, Y.-H. Chang, S.-H. Lo, and K.-M. Chao, “Text–image multimodal fusion model for enhanced fake news detection,” Sci Prog, vol. 107, no. 4, p. 00368504241292685, 2024,

[27] Khalil, A., Jarrah, M., and Aldwairi, M, “Hybrid Neural Network Models for Detecting Fake News Articles”, Human-Centric Intelligent Systems, 4, 136-146, 2024.

[28] First, B. M., Shanmugavadive, K., Sathishkumar, V. E., Maruthappa, M., Subramanian, M., and Thinakaran, R, “Attention Mechanism-Based CNN-LSTM for Abusive Comments Detection and Classification in Social Media Text”, International Journal of Advanced Computer Science & Applications, 15(10), 2024.