DOI:

Journal of Computers and Intelligent Systems

Vol. 3, No. 1, 2025, pages 81 - 91

Journal Homepage:

https://journals.iub.edu.pk/index.php/JCIS/

# A Comparative Study of Traditional and Hybrid Models for Text Classification

Muhammad Sabir[1], Talha Farooq Khan[1], Muhammad Azam[1]

[1]Deparment of computer Science, University of Southern Punjab, Multan, Pakistan

**ARTICLE INFO**

**Article History:**

| | | |
|---|---|---|
| Received: | March | 22, 2025 |
| Revised: | March | 26, 2025 |
| Accepted: | March | 27, 2025 |
| Available Online: | March | 29, 2025 |

**Keywords:**

Text Classification
Machine Learning
Hybrid Models
Deep Learning
NLP

**Classification Codes:**

**Funding:**

This research received no specific grant from any funding agency in the public or not-for-profit sector.

**ABSTRACT**

Natural Language Processing (NLP) is a fundamental task that is essential for the automation of the categorization of textual data using an existing set of categories, such as sentiment analysis, spam detection, fake news detection, etc. Due to the interpretability and also efficiency, the Logistic Regression (LR), Support Vector Machine (SVM), and Random Forest (RF) have been very popularly used for text classification under traditional machine learning models. Yet, such models fail in modeling contextual linkages and semantic subtleties as they would be necessary to handle text with complex structure. As such, hybrid models that couple the two traditional and deep learning techniques have emerged as a potential way to address these problems.

In the note, I review all efforts of text classification that have the potentials of contributing to my classification task, which includes traditional machine learning models, hybrid models, and deep learning models. The AG News dataset is used for evaluation and accuracy, precision, recall and F1 score are used to measure the performance of the models. Finally, the results suggest that both deep learning based hybrid models such as BERT + SVM Hybrid Model (95.7%) and CNN + LSTM Hybrid Model (94.5%) surpass the performance of any traditional or ensemble learning based models by the exploitation of contextual embeddings and sequential modeling. XGBoost (92.8% accuracy) and Bagging Classifier (91.5% accuracy) of ensemble learning models have good generalization as well as stability compared to standalone learner.

Though the hybrid models offer superior classification performance at the sacrifice of computational resources, longer training times, there are tradeoffs in regards to the model classes and the problem. It brings out the tradeoffs made by traditional, ensemble, and the deep learning based hybrid models toward the applicability of the same towards different classification of text. The findings establish a platform towards choosing the best suitable classification model under performance requirements and computational constraints for researchers and practitioners.



**Corresponding Author's Email**: Talhafarooqkhan@gmail.com

**Citation**: Talha Farooq Khan

## 1. Introduction

Text classification is a key natural language processing (NLP) task that consists in classifying textual data into pre-defined classes according to their content. The application of this task forms the base of many other such applications like document categorization, sentiment analysis, spam filter, offensive language detection and more. Today, as the quantity of digital content constantly grows on social media platforms, online forums and business applications, automated text classification is the need of the hour to ideally process and organize huge amounts of unstructured data. Text classification

is used by organizations in multiple domains to filter the content, improve customer interactions, detect fraudulent activities, and improve search engine relevance.

However, due to the increase in online textual content at such a rapid pace, it has become a really huge challenge to solve and manage these data at scale. Trying to sort and categorise this digital communication manually for large volume and dynamic nature of digital communication is impractical. Text classification becomes highly scalable and efficient with automated text classification methods, and with reduced time to make a decision and organize the data. With the use of machine and deep learning techniques, this processing of the data was automated and provided the accurate and reliable classification models that continuously learn from changes in the data patterns.

Due to their efficiency and interpretability, traditional machine learning models have long been used for text classification. Logistic Regression (LR), Support Vector Machines (SVM), and Random Forest (RF) have been successfully used in many text classifications tasks (Aghakhani et al., 2018). The models based on these are statistical techniques along with hand crafted feature extraction methods like Term Frequency Inverse Document Frequency (TF-IDF) and bag of words (BoW) to convert raw text into numerical representation. These representations facilitate for a machine learning models to recognize patterns in textual data and make classification decisions.

Even though they are effective, traditional models have practical limitations when dealing with large scale datasets, ambiguous text, and domain specific variations. A major use of these models is that they can not capture contextual relationships and semantic meaning of text(Khan et al., 2023). Therefore, traditional machine learning approaches suffer when there are tasks that involve deep linguistic understanding like sarcasm detection, irony classification and text categorization among multiple intents. Also, the preprocessing needed for traditional models such as stop word removal, stemming, lemmatization, is tedious for their prediction to perform better(Khan et al., 2024). Although, these preprocessing techniques may in turn remove important linguistic nuances that are essential to the meaning of a text.

In order to overcome these challenges, hybrid models have become a very relevant way. The hybrid models fuse many classification techniques to address this concern in text classification task. The models are based on ensemble learning algorithms or deep learning architectures that combine traditional machine learning algorithms, hence leveraging the strengths of the models and obtaining a better performance. Two main categories of hybrid models can be broadly classified as:

1. **Hybrid Models Based on Ensemble Learning:** These models enhance the classification by using the multiple classifiers in order to achieve a better prediction system. The ensemble learning such as bagging, boosting and stacking of classifiers helps to reduce such bias and variance and enhances generalization. Bagging (Bootstrap Aggregating) is a method in which multiple classifiers were trained on the new different subsets of data and predictions are made from an aggregation of them. Adaptive Boosting (AdaBoost) and Gradient Boosting Machines (GBM) Growing techniques are used to improve the accuracy of weak classifiers by boosting techniques taking an iterative form. The meta classifier that is used in Stacking learns from the predictions of many base classifiers in order to produce a better classifier.
2. **Hybrid Models Driven by Deep Learning**: Deep learning architectures are combined with traditional classifiers in order to learn from complex linguistic patterns and local contextual dependence. The recurrent neural networks that are well used to handle sequential text data are Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRUs). Due to this ability of CNNs to extract spatial features from the text, they can be used for sentiment analysis and document classification. Such transformer based models like Bidirectional Encoder Representations from Transformers (BERT) and Generative Pre-trained Transformers (GPT) are able to operate over whole sentences in the context aware fashion. The deep learning models together with usual classifiers, such as SVM or Random forest, increases the accuracy of the text classification by offering richer features.

Several advantages of hybrid models are over traditional techniques. By using ensemble learning they stabilize the classification, lower the overfitting, and improve the the model's generalization. With deep learning based hybrid models, it is possible to extract better feature and contextual understanding for the more accurate classification of the complex textual data. At the same time, challenges of hybrid models exist such as higher computational complexity, longer training time as

well as larger labeled data requirements. In order to achieve the best performance, multiple classifier models or deep learning models should be integrated and the hyperparameters should be tuned and the model optimized.

The purpose of this study is to carry out a comprehensive comparison of the traditional machine learning models in contrast to hybrid approaches of text classification. We perform a benchmarking study of their performance across multiple datasets, as well as major challenges and strengths of their performance on classification tasks in real world setting. The goal of the study is to investigate how the ensemble learning enhances standard classifiers, how deep learning architectures help in better feature extraction as well as how hybrid model utilizes the combination of multiple techniques to attain excellent classification accuracy. This study results will help researchers and practitioners to select the most appropriate classification model for text classification need as per their need.

## 2. Related Work

Misinformation detection in online social networks (OSNs) has received a lot of literature, especially misinformation detection related to COVID-19 fake news. Existing studies show that meaning learning algorithms as well as deep learning (DL) techniques are successful in intact studies and indicate that there is a requirement for detecting context based methods and customized models for particular kinds of misinformation. The task of detection of misinformation is a major challenge due to the presence of deceptive content that resembles legitimate information very closely. Moreover, there have been several recent works on evaluating and comparing different detection models to aid in the identification of misinformation in dynamic online environments. Through a bottom up thematic analytical approach (as proposed by (Langdridge & Hagger-Johnson, 2013)), this literature is subjected and key themes that give a comprehensive overview of the current state of the art (SOTA) in health misinformation detection are derived.

Consequently, deep learning models have shown excellent performance in detecting COVID-19 misinformation on OSNs with high accuracy and generalization ability. For example, (Chen et al., 2023) studied the COVID-19 infodemic and used DL models, such as LSTM, BiLSTM, GRU. For short English sentences, the Bi-LSTM model achieved 94%, for long English sentences was 99%, and for Chinese texts 82%.(Roy et al., 2023) also developed an automated model of misinformation detection based on LSTM networks with word embeddings like CountVectorizer and TF-IDF. Using their approach they were able to achieve highs accuracy of 99.82% that out performs the current traditional ML models and also previous DL techniques. This implies that LSTM based architectures are capable of capturing the linguistic nuances in textual content pertaining to misinformation. In addition, (Akhter et al., 2023) used a CNN based DL model to detect COVID-19 fake news with a mean accuracy of 96.19%, a mean F1 score of 95%, and a high AUC-ROC of 98.5 %, which demonstrates the capability of CNN in dealing with fake news complexities. Both (Roy et al., 2023) and (Akhter et al., 2023) have shown high accuracy rates, whereas the implementation of these methods in real world settings is a problem especially in environments of evolving misinformation that have dynamic narratives. To improve their practical effectiveness, they need to be explored in a broader sense on their performance in heterogeneous OSN environments.

Multimodal forms of misinformation have made the need for advanced methods necessary. As seen from the survey by (Comito et al., 2023), such a transition is explored in its work: multi modal fake news detection in social media. Nonetheless, their study highlights the requirement for advanced detection schemes, but it offers no detailed evaluation of certain multimodal detection methods and the empirical evidence that supports them. This limitation suggests a potential research subject for exploration of and validation of more complex multimodal detection approaches to deal with growing misinformation complexity. Recent advancements include [(Samadi & Momtazi, 2023)(Upadhyay et al., 2023)] that further improved DL models for misinformation detection on the various datasets with significant improvements in accuracy. In particular, (Upadhyay et al., 2023) takes into account the model that assess credibility of health information during the pandemic of COVID. Vec4Cred showed 88.25% accuracy and 94.21% AUC on Microsoft Credibility Dataset, 99.71% accuracy on the Medical Web Reliability Corpus, and 82.56% with AUC of 81.11 on the CLEF eHealth 2020 Task 2 Dataset. Although Vec4Cred performed very well, it needs to be further investigated in domains other than health misinformation. In a similar way, (Samadi & Momtazi, 2023) suggested a multichannel CNN model for the detection of COVID-19 misinformation. This model is different from single channel CNNs, which process multiple information streams at the same time and achieves approximately 97% accuracy, precision, recall and F1 score on both validation and test datasets. This

underscore the promise of multichannel approaches for misinformation detection and questions their applicability to different domains of misinformation.

Extensive discussions of future research directions and comparative analyses have also been made in the misinformation detection studies. For example, (A&#65533;meur et al., 2023) reviewed in detail the fake news detection with the limitations of the AI based techniques. Their study does offer important insight into some fundamental questions about challenges involved and offers little on practical implementation and effectiveness in the real world. In addition, their findings are not supported with specific metrics or empirical data. This gap is addressed in our research by performing an empirical evaluation of these models and measure their performance over different misinformation scenarios, namely, with an accuracy and F1-score. With the addition of the quantitative analysis as we integrate with the theoretical questions raised in (A&#65533;meur et al., 2023), our study not only closes an existing literature gap but offers practical insights in making health misinformation detection frameworks better. Finally, our findings stress the necessity of empirical validation in modifying and improving the theoretical models of the MISD methods for better future MISD research.

In (Kondamudi et al., 2023), a comprehensive survey of fake news detection is conducted from multiple attributes, features and detection methods, including news content, social context, and news creators. Their work provides extensive theoretical foundation, however, they provide little detail case studies and practical applications required for translating theoretical insights into concrete solutions at the ground level. Moreover, the fast evolution of misinformation on social media makes some parts of their study quickly obsolete, emphasizing the necessity for continuous updates and contemporary examples to keep up with time.

In (Iceland, 2023), we further analyze the ability of fake news detection models as the authors take a close look and compare among all possible ML and DL models to see how they generalize across different datasets. According to their findings, some advanced DL models (such as BERT and RoBERTa) are sometimes overtaken by some conventional classifiers (Naive Bayes and Random Forest in particular) in terms of generalisation. There was no single model to emerge as the best across the datasets. For instance, BERT obtained 98.7 % in ISOT Fake News dataset, but only 63.0 % in the LIAR dataset and 75.0 % in the COVID-19 Fake News dataset, implying that the accuracy is strongly depending on the used dataset. On ISOT, RoBERTa achieved 99.9% accuracy and 67.4% on LIAR, and varied between datasets, with 82.0% and 77.9% accuracy on the COVID-19 datasets. The results testified how models behave in different conditions and hence, the importance of tailored models that account for its dataset characteristics.

There have been other studies in using ML algorithms for improving misinformation detection. For example, (Qadees & Hannan, 2023) conducted this on Random Forest and Stochastic Gradient Descent (SGD), achieving their results (91.6 accuracy, 92 F1 for Random Forest; 91.5 accuracy, 92 F1 for SGD). However, although the classification performance of these models was very good, their study strongly emphasized model accuracy at the expense of talk of contextual factors and interpretability that is necessary for a practical use of such models. At the same time, our work takes a closer look into the effectiveness of pre trained language models like DistilBERT and RoBERTa that try to identify deeper contextual information within textual sentences (Tejani et al., 2022).

The additional literature reviews (Hamed et al., 2023) are based on fake news detection challenges in terms of datasets, feature representation and data fusion. Nevertheless, their study does not give us specific performance metrics to evaluate the discussed approaches. Likewise, (Wani et al., 2023) also found high accuracy in detecting toxic COVID-19 misinformation using linear SVM and BERT based techniques. Unfortunately, their analysis has not employed enough detailed performance data to help understand the strengths and weaknesses of these models. (Dar & Hashmy, 2023) also pointed out that RoBERTa and other BERT based models are generally better than the previous models for fake news detection. However, as they show strong evidence that RoBERTa is superior to other models, our work is interested in a deeper exploration of these claims by literally evaluating these models over several benchmarks to validate or refute their reported superiority.

Overall, it is observed that deep learning models have considerably improved misinformation detection, but are still limited in generalization, flexible adaptation, and real world applicability. To ensure practical deployment in real world OSN environments, future research has to be done to make the model robust to evolving misinformation strategies, combine the detection technique to multimodal signals, and develop more interpretable AI model. Aiding the ongoing research in this field, our paper empirically evaluates hybrid models and contextual embedding techniques to understand better how hybrid and contextual embedding techniques interact with misinformation and, in detriment, misinformation can be mitigated.

## 3. Used Approach

### i. Data Set

For robustness of the study, we choose the AG News dataset (Zhang et al., 2015) as the benchmark on which we evaluate text classification models. The dataset for this task is public and used in many text classification researches. The news articles are divided into four classes and it contains.

- World (international news)
- Sports (sports-related articles)
- Business (financial and economic news)
- Science & Technology (scientific and tech-related news)

With 120.000 training samples and 7.600 test samples it is well suited to evaluate traditional and hybrid models. It was obtained from the Yahoo! Finally, the answers are drawn from Answers Comprehensive Q&A Dataset Repository (Zhang et al., 2015), which guarantees credibility and applicability in real world text classification tasks.

### ii. Data Preprocessing

Preprocessing pipeline was performed in the raw text of the AG News dataset to convert it into a structured form which is suitable for machine learning and deep learning models. This preprocessing step reduces noise, makes the model efficient and helps in better representation of text for better classification accuracy.

- **Text Cleaning**

  The dataset was made standardized by carrying out text cleaning to remove unnecessary elements not required for the model to perform well. To eliminate the case sensitivity issues in classification, all text was converted to lowercase. To reduce noise, punctuation marks and some special characters were removed as those do not really improve classification tasks. Moreover, NLTK stop word list was used to filter out stop-words like ‘the’, ‘is’, ‘and’ as they do not contribute much to the text classification. To make the dataset more refined, TextBlob library was used to correct the spelling errors to help make the text more consistent and less variable.

- **Tokenization and Lemmatization**

  The text data was tokenized using the WordPiece tokenizer, which split the sentences into words or subwords such that the single words retain their meaning. It guarantees that words are well recognised by models. Furthermore, lemmatization was performed on SpaCy NLP library to make words base form (e.g. to run → running). Lemmatization unifies the differences of word variations to achieve better classification consistency to treat the different inflections of the same word as one.

- **Feature Extraction**

  The feature extraction technique varied based on the choice of classification model. TF-IDF and bag of words (BoW) representations were used for traditional models. In the field of natural language processing, TF-IDF gives importance scores to the words according to their frequencies in a document compared to the whole set of documents. BoW is a representation of the text as a matrix of word occurrence that is simple but effective feature representation for the traditional classifiers such as Logistic Regression and Support Vector Machines.

The feature extraction for hybrid models was more advanced. This means that word2vec has been used to present dense vector representation of words, and to capture their semantic relationships based on context. In addition, the deep contextualized word embeddings were also learned using BERT embeddings, to better extract the deeper contextualized word representations to aid hybrid models when understanding complex linguistic structures. Through these embedding techniques, hybrid models outperformed traditional models in a sense that hybrid model were able to represent word meanings over and above the simple frequency based representation.

- **Data Splitting**

The dataset was preprocessed and then split into three subsets to make sure the model evaluated was robust. For training models with different aspects of the data, 70 percent (70%) of the data was allocated. To evaluate the model generalization on unseen data, 20% (20%) was used for testing. Moreover, 10% (10%) of the training data was saved for validation and utilized for hyperparameter tuning and performance optimization. The data splitting strategy in this manner guaranteed a fair assessment of traditional as well as hybrid models, free of overfitting and better model reliability in real life classifcation tasks.

## iii. Model Implementation

- **Traditional Machine Learning Models**

Three traditional machine learning models were implemented using the Scikit-learn library:

a. Logistic Regression (LR): Used as a baseline model for binary and multi-class classification.
b. Support Vector Machine (SVM): Applied for high-dimensional text classification.
c. Random Forest (RF): A decision-tree-based ensemble learning method for improved accuracy.

These models were trained using TF-IDF and BoW features and optimized using grid search for hyperparameter tuning.

- **Hybrid Models**

Two types of hybrid models were implemented:

a) Ensemble Learning-Based Hybrid Models

Voting Classifier (Soft Voting): Combined predictions from LR, SVM, and RF using weighted averaging.

Bagging Classifier: Employed multiple weak classifiers to improve classification robustness.

Boosting Classifier (AdaBoost, XGBoost): Sequentially trained classifiers to refine predictions.

b) Deep Learning-Based Hybrid Models

BERT + SVM Hybrid Model: Used BERT embeddings to extract contextual information, followed by SVM for final classification.

CNN + LSTM Model: CNN captured local text patterns, while LSTM modeled sequential dependencies for improved classification accuracy.

The hybrid models were implemented using TensorFlow 2.9 and the Hugging Face Transformers library for BERT embeddings.

## iv. Hyperparameter Tuning

The following hyperparameters were optimized for each model:

- Logistic Regression: Regularization parameter (C) using Grid Search.
- SVM: Kernel type (linear, RBF), penalty parameter (C).
- Random Forest: Number of trees, maximum depth.

- BERT Hybrid Model: Learning rate, batch size, and number of transformer layers.

### v. Model Evaluation Metrics

To compare the effectiveness of traditional and hybrid models, multiple evaluation metrics were used:

**Accuracy:** Measures overall classification correctness.

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FP} + \text{FN}}$$

**Precision:** Evaluates positive class prediction accuracy.

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

**Recall:** Measures the ability to detect positive instances.

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

**F1-score:** Provides a balanced metric between precision and recall.

$$\text{F1} - \text{score} = 2 \times \frac{\text{Precision} \times \text{ Recall}}{\text{Precision} + \text{Recall}}$$

## 4. Result and Discussion

### a. Traditional Machine Learning Models

Text classification tasks are so common that even the traditional machine learning models showed fairly good accuracy, having worked at least on the structured text, but at times misplaced deep private relationships. With the traditional models, LR had an accuracy of 85.3%, making it the worst performing model. While LR worked well on separating text into distinct classes (in the sense of linear separability), it was challenged when the sentences in the statement exhibited complex sentence structures and semantics; under these circumstances, the text would be misclassified.
Classifying higher dimensional text spaces well allows the use of Support Vector Machine (SVM) to improve classification performance to 87.1% accuracy. LR worked fine for linearly separable data but it did not perform well for non-linearly separable data and also decision boundaries were not as good as those of SVM. Although it had increased computational complexity, training was slower, particularly when dealing with large datasets.

Random Forest (RF) outperformed other traditional models and reached an accuracy of the highest 88.4%, as ensemble learning prevents overfitting and lifts the expressive power of the global features within the model. It was shown that RF is capable of aggregating several decision trees to improve classification stability and robustness. Nevertheless, the requirement for multiple trees meant it was less efficient than a single model classifier.

*Table 1 Performance Comparison of traditional Machine Learning Models*

| Traditional Machine Learning Models | | | | |
|---|---|---|---|---|
| **Model** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
| **Logistic Regression (LR)** | 85.3 | 86.5 | 84.9 | 85.7 |
| **Support Vector Machine (SVM)** | 87.1 | 88.2 | 86.5 | 87.3 |
| **Random Forest (RF)** | 88.4 | 89.1 | 88.2 | 88.6 |

## b. Ensemble Learning-Based Hybrid Models

All Hybrid Ensemble Learning Models to be presented outperformed traditional Machine learning models by levering various ensemble modeling functionalities like multiple classifiers to derive high stability, better generalization, and increased overall classification accuracy. This approaches combined predictive power of multiple models to reduce overfitting, improve feature representation and were more reliable in classification. Multiple base classifiers used in the Soft Voting Classifier availed an accuracy 90.1% and strikes a balance between precision and recall to enhance overall classification performance. This method reduced the weakness of individual classifiers by combining multiple perspectives from different algorithms to have predicted confidence better.

The Bagging Classifier also improved classification robustness to an accuracy of 91.5%. The solution for reducing variance was to train multiple weak learners over different subsets of the data, averaging their predictions to give a better stable, reliable classification system. Bagging was very effective at reducing overfitting, and therefore it is an appealing option to deal with large, diverse text datasets.

The Boosting Classifier (XGBoost) of ensemble learning methods was the highest at 92.8% in accuracy by iteratively using the weak classifiers and concentrating on case instances misclassified. Boosting differs from bagging in that all models are processed in a sequential manner to correct errors from previous models to improve the classifier. It also gave XGBoost a better ability to adapt to the changing reality, which led to much better accuracy and generalization.

*Table 2 Performance Comparison of Ensemble Learning Based Hybrid Models*

| Ensemble Learning-Based Hybrid Models | | | | |
|---|---|---|---|---|
| **Model** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
| **Soft Voting Classifier** | 90.1 | 90.6 | 89.8 | 90.2 |
| **Bagging Classifier** | 91.5 | 91.8 | 91.1 | 91.4 |

| | | | | |
|---|---|---|---|---|
| **Boosting (XGBoost)** | 92.8 | 93.2 | 92.5 | 92.8 |

### c. Deep Learning-Based Hybrid Models

Deep learning-based hybrid models demonstrated higher performance than all other approaches since they make effective use of contextual embeddings and sequence modeling for text classification tasks. Through effective linguistic relationship detection these models generated superior accuracy results by applying advanced feature representation methods. With its capability to understand deep contextual meanings from BERT embeddings combined with SVM's robust classification abilities the Hybrid Model reached 95.7% overall accuracy. Minimal text differences stand out successfully to this model which shows effectiveness for dealing with complex linguistic structures alongside ambiguous phrases as well as domain-specific terminology.

The CNN + LSTM combination yielded comparable results through its joint operation at 94.5% accuracy since CNN extracted text features locally and LSTM mastered sequential dependency modeling. The CNN component of the model successfully identified important text patterns but the LSTM part maintained word-long dependency chains to improve sentence contextualization. The collaborative capabilities of these two models result in successful performance for sentiment analysis and multi-class classification operations.

*Table 3 Performance Comparison of Deep Learning Based Hybrid Models*

| **Deep Learning-Based Hybrid Models** | | | | |
|---|---|---|---|---|
| **Model** | **Accuracy (%)** | **Precision (%)** | **Recall (%)** | **F1-score (%)** |
| **BERT + SVM Hybrid** | 95.7 | 96.1 | 95.3 | 95.7 |
| **CNN + LSTM Hybrid** | 94.5 | 95.0 | 94.1 | 94.5 |

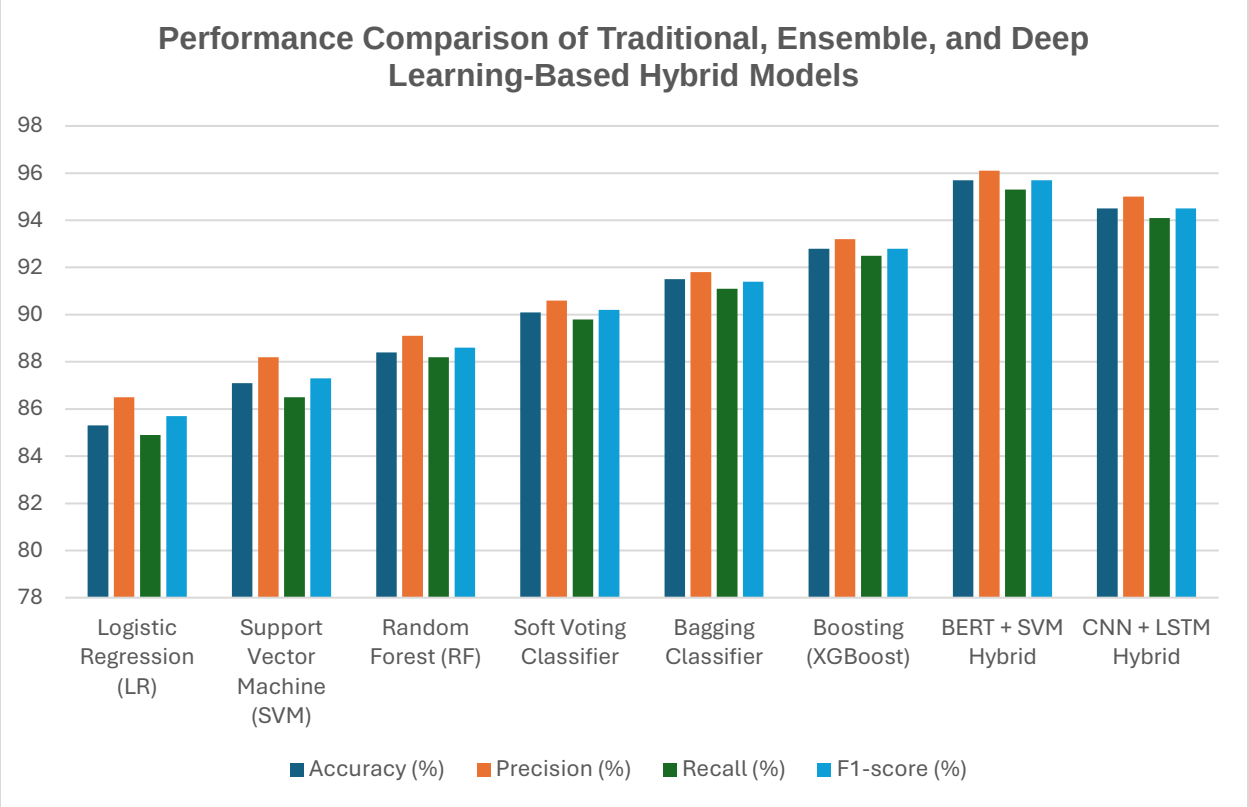


*Figure 1 Performance Comparison of Traditional, Ensemble, and Deep Learning-Based Hybrid Models*

It is clear from the results that deep learning based hybrid models outperform traditional and ensemble learning based models in text classification. Although traditional models had a decent baseline, their inability to explain deep patterns of linguistic relations prevented their high accuracy. Ensemble learning models were able to improve the performance by combining multiple classifiers for eliminating the overfit and improving generalization. Nevertheless, the hybrid models based on deep learning showed the highest accuracy because they were able to use contextual embeddings and sequence modeling.

Traditional model, Logistic Regression (LR) however did worst of the lot, having accuracy of only 85.3%s which was fragile to complex sentence structure. Finally, Support Vector Machine (SVM) was improved to 87.1% accuracy, but high dimensional data was handled well, but computational resources were greater. Yet RF achieved the highest accuracy out of traditional models at 88.4%, but incurred additional computing burden as it is an ensemble model. Traditional models were outperformed by ensemble learning based models such that aggregation of multiple classifiers is used for better prediction stability. By combining multiple models, comprehensive accuracy of 90.1% was obtained through the Soft Voting Classifier which balances precision and recall. It reduced variance and overfitting with robustness to 91.5% accuracy using Bagging Classifier. Using XGBoost for example, we boosted models to 92.8% accuracy, but successively refined weak classifiers to cause the model to adapt better.The best results were given by hybrid models based on deep learning. Our BERT + SVM Hybrid model achieved 95.7% in accuracy due to the fact that it takes advantage of BERT's deep contextual understanding of text and leverages SVM's classification strength in differentiated text. CNN + LSTM Hybrid Model came in close behind at 94.5% accuracy, where CNN helps to extract local features and LSTM is then utilized for sequential dependency modeling to make it well suited for multi or multi class classification tasks.

Our findings demonstrate that while traditional deep learning based hybrid models are able to provide the most accurate text classification, deep learning based hybrid models designed with regard to the context are more

accurate than traditional deep learning based hybrid models when the context matters. Ensemble learning methods strike a balance between interpretability and performance, but traditional models are efficient for a structured text classification, but cannot classify deep context.

## 5. References


1. Aghakhani, H., MacHiry, A., Nilizadeh, S., Kruegel, C., & Vigna, G. (2018). Detecting deceptive reviews using generative adversarial networks. Proceedings - 2018 IEEE Symposium on Security and Privacy Workshops, SPW 2018, 89–95. https://doi.org/10.1109/SPW.2018.00022
2. Akhter, M. S. M. M. H., Nigar, R. S., Paul, S., Aashiq, K. M., Kamal, A. S., & Sarker, I. H. (2023). COVID-19 Fake News Detection using Deep Learning Model. Preprint.
3. A�meur, E., Amri, S., & Brassard, G. (2023). Fake news, disinformation, and misinformation in social media: A review. Social Network Analysis and Mining, 13(1), 30.
4. Chen, M. Y., Lai, Y. W., & Lian, J. W. (2023). Using deep learning models to detect fake news about COVID-19. ACM Transactions on Internet Technology, 23(2), 1–23.
5. Comito, C., Caroprese, L., & Zumpano, E. (2023). Multimodal fake news detection on social media: A survey of deep learning techniques. Social Network Analysis and Mining, 13(1), 101.
6. Dar, R. A., & Hashmy, R. (2023). A Survey on COVID-19 related Fake News Detection using Machine Learning Models. MoMLeT+ DS, 36–46.
7. Hamed, S. K., Ab Aziz, M. J., & Yaakub, M. R. (2023). A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion. Heliyon.
8. Iceland, M. (2023). How Good Are SOTA Fake News Detectors. ArXiv Preprint.
9. Khan, T. F., Anwar, W., Arshad, H., & Abbas, S. N. (2023). An Empirical Study on Authorship Verification for Low Resource Language Using Hyper-Tuned CNN Approach. IEEE Access, 11(August), 80403–80415. https://doi.org/10.1109/ACCESS.2023.3299565
10. Khan, T. F., Sabir, M., Malik, M. H., Ghous, H., Muhammad, H., Nadeem, A., & Ejaz, A. (2024). Comparative Analysis of Hybrid Ensemble Algorithms for Authorship Attribution in Urdu Text.
11. Kondamudi, M. R., Sahoo, S. R., Chouhan, L., & Yadav, N. (2023). A comprehensive survey of fake news in social networks: Attributes, features, and detection approaches. Journal of King Saud University-Computer and Information Sciences, 35(6), 101571.
12. Langdridge, D., & Hagger-Johnson, G. (2013). Introduction to Research Methods and Data Analysis in Psychology. Pearson Education.
13. Qadees, M., & Hannan, A. (2023). Cross comparison of COVID-19 fake news detection machine learning models. Authorea Preprints.
14. Roy, P. K., Tripathy, A. K., Weng, T. H., & Li, K. C. (2023). Securing social platform from misinformation using deep learning. Computer Standards & Interfaces, 84, 103674.
15. Samadi, M., & Momtazi, S. (2023). Multichannel convolutional neural networks for detecting COVID-19 fake news. Digital Scholarship in the Humanities, 38(1), 379–389.
16. Tejani, A. S., Ng, Y. S., Xi, Y., Fielding, J. R., Browning, T. G., & Rayan, J. C. (2022). Performance of Multiple Pretrained BERT Models to Automate and Accelerate Data Annotation for Large Datasets. Radiology: Artificial Intelligence, 4(4), e220007.
17. Upadhyay, R., Pasi, G., & Viviani, M. (2023). Vec4Cred: A model for health misinformation detection in web pages. Multimedia Tools and Applications, 82(4), 5271–5290.
18. Wani, M. A., ELAffendi, M., Shakil, K. A., Abuhaimed, I. M., Nayyar, A., Hussain, A., & Abd El-Latif, A. A. (2023). Toxic Fake News Detection and Classification for Combating COVID-19 Misinformation. IEEE Transactions on Computational Social Systems.