# Advancing Metaphor Detection: A Comprehensive Comparative Study of Machine Learning and Deep Learning Approaches

**Hanzla Javaid** and **Mausam Gerra** and **Yash Patel**
Department of Computer Science
mjavaid@iu.edu, yp33@iu.edu, maugerra@iu.edu

## Abstract

In this research project, we explored the challenging task of metaphor detection in natural language processing. By leveraging advanced machine learning techniques and deep learning architectures, we aimed to develop a robust system capable of accurately identifying and interpreting metaphors across diverse linguistic contexts. We utilized a dataset containing labeled metaphorical and literal text to train and evaluate our models. Our methodology involved data preprocessing, feature engineering, and the application of various neural network architectures. The preliminary results demonstrate the potential of our approach in improving metaphor detection accuracy. Future work will focus on incorporating multimodal information, developing cross-lingual systems, and exploring few-shot and zero-shot learning techniques.

## 1 Introduction

### 1.1 Problem Statement

In the rapidly evolving field of natural language processing (NLP) and computational linguistics, detecting and understanding metaphors has emerged as a critical challenge. Metaphors are complex linguistic devices that go beyond literal meaning, requiring sophisticated computational models to accurately identify and interpret them. By leveraging advanced machine learning techniques and deep learning architectures, we aim to develop a robust metaphor detection system that can analyze text across multiple domains and linguistic contexts.

Metaphors are fundamental linguistic constructs that transcend literal semantic interpretation, representing complex cognitive mechanisms for understanding abstract concepts through concrete imagery. They play a critical role in shaping human thought, creativity, and communication. For computational systems, however, identifying and interpreting metaphors remains a challenging task due to their inherently non-literal nature and dependence on contextual nuances.

The growing prominence of natural language processing (NLP) in applications such as sentiment analysis, question answering, and machine translation underscores the necessity of effective metaphor detection systems. The primary focus of this research is to bridge the gap between traditional machine learning and contemporary deep learning approaches for metaphor detection, presenting a detailed comparative analysis that evaluates their effectiveness.

### 1.2 Overview of the Data

To develop our metaphor detection model, we utilized the train.csv dataset provided to us in the portal. It has three significant columns based on which we are going to train and test our model's prediction and accuracy using 3 different approaches of metaphor detection by using Random Forest Classifier, Naive Bayes Classifier, and Deep Learning Model.

The dataset includes features as:

- Metaphor ID that gives us the metaphor candidate word

- Label which suggests whether the text has metaphor (TRUE) or it is a literal text (FALSE)

- Text that contains the metaphor candidate word. If there are multiple occurrences of the word, the annotation is always for the first occurrence.

### 1.3 Research Objectives

Our research aims to address multiple critical objectives:

- Develop and compare computational approaches for metaphor detection, focusing on

traditional machine learning techniques and deep learning models.

- Investigate the role of advanced feature extraction techniques in improving metaphor detection accuracy.

- Evaluate the strengths and limitations of various approaches based on comprehensive performance metrics.

- Provide insights into how computational systems can better understand and process metaphorical language.

### 1.4 Our Contribution

For this project, we collaborated closely to achieve the following:

- **Hanzla**: Implemented attention based residual neural network architecture and conducted baseline performance evaluations.

- **Mausam**:Implemented Naive Bayes approach and evaluated all training pipelines

- **Yash**: Performed data preprocessing, and explored Random Forest and other ensemble techniques.

## 2 Theoretical Background

### 2.1 Linguistic Foundations of Metaphor

Metaphors are not merely ornamental language features; they are deeply embedded in human cognition and communication. According to the conceptual metaphor theory, metaphors facilitate understanding by mapping abstract concepts (target domain) onto more concrete, familiar experiences (source domain). For example, in the metaphor *"Time is money,"* the abstract concept of time is mapped to the concrete domain of money to convey value and scarcity.

The inherent complexity of metaphors lies in their reliance on context, cultural factors, and subtle linguistic variations. These attributes necessitate computational models capable of capturing both local and global semantic dependencies.

### 2.2 Computational Challenges

Detecting metaphors computationally involves several intricate challenges:

1. **Contextual Variability:** Metaphors often derive meaning from surrounding text, making their detection context-dependent.

2. **Semantic Nuance:** Differentiating between literal and non-literal meanings requires deep semantic understanding.

3. **Domain-Specific Variations:** Metaphorical expressions vary significantly across domains such as literature, science, and everyday conversation.

4. **Linguistic Diversity:** Variations in syntax, idiomatic expressions, and cultural influences add complexity to metaphor detection tasks.

## 3 Literature Review

### 3.1 Traditional Approaches

Early computational approaches to metaphor detection relied heavily on traditional machine learning techniques and handcrafted linguistic features. Prominent methods include:

- **Support Vector Machines (SVM):** SVM models leveraged features such as part-of-speech tags, word embeddings, and dependency relations to classify metaphorical expressions.

- **Logistic Regression:** This technique was employed with syntactic and semantic features to predict metaphorical usage.

- **Rule-Based Systems:** These systems utilized predefined linguistic rules to detect patterns indicative of metaphors. While interpretable, they struggled with scalability and adaptability.

### 3.2 Deep Learning Advancements

Recent advancements in NLP have shifted focus towards deep learning models capable of capturing contextual and semantic intricacies. Key approaches include:

- **Recurrent Neural Networks (RNN):** These models effectively capture sequential dependencies in text but often face challenges with long-range context.

- **Long Short-Term Memory (LSTM):** LSTMs mitigate the vanishing gradient problem in RNNs, enabling better performance on tasks involving longer text sequences.

- **Transformer-Based Models:** Models like BERT and RoBERTa have revolutionized

NLP by leveraging attention mechanisms to capture contextual relationships at both token and sentence levels.

# 4 Methodology

## 4.1 Dataset Preparation

Our study utilized a curated dataset designed to represent diverse metaphorical expressions across multiple domains. Table 1 provides an overview of the dataset:

| Dataset Attribute | Specification |
|---|---|
| Total Samples | 1870 |
| Train-Test Split | 80% Training, 20% Testing |
| Preprocessing Techniques | Lowercasing, Tokenization, Stopword Removal |

Table 1: Dataset Composition and Preprocessing

## 4.2 Preprocessing Pipeline

Text preprocessing was critical for ensuring consistent and informative input to the models. The pipeline consisted of:

1. Lowercasing all text.

2. Removing punctuation and special characters.

3. Tokenizing sentences into words.

4. Removing stopwords and applying stemming or lemmatization.

## 4.3 Our Approach

Our methodology encompassed several key stages:

- **Idea and intuition**:

    Metaphor detection is a challenging task in natural language processing that requires understanding the contextual and nuanced usage of language.

    Leveraging advanced machine learning techniques, particularly deep learning models, can help capture the complex patterns and contextual information necessary for effective metaphor detection.

    Incorporating domain-specific knowledge and features, such as linguistic properties and semantic relationships, can further improve the model's performance.

- **Justification**:

    Rule-based systems, while effective in narrow domains, often lack the flexibility and

scalability required to handle the diversity of metaphorical expressions in real-world texts.

Statistical machine learning approaches, such as SVMs and Naive Bayes classifiers, capture patterns in the data but have limited contextual understanding.

Deep learning models, with their ability to learn contextual embeddings and capture complex linguistic representations, have shown promising results in advancing the state-of-the-art in metaphor detection.

### 4.3.1 Random Forest Classification

Random Forest is an ensemble learning method that aggregates predictions from multiple decision trees. It is robust to overfitting and effective with high-dimensional feature spaces.

**Key Hyperparameters:**

- Number of Estimators: 100
- Maximum Depth: Unlimited
- Feature Extraction: TF-IDF Vectorization

### 4.3.2 Multinomial Naive Bayes

This probabilistic classifier assumes feature independence, making it computationally efficient for text classification tasks.

**Key Hyperparameters:**

- Smoothing Parameter: 0.1
- Feature Space: 1,500 TF-IDF features

### 4.3.3 Attention-Based Residual Neural Network

Our proposed deep learning model integrates:

- **Sentence Embeddings:** Pretrained embeddings from Sentence-BERT capture rich contextual semantics.
- **Attention Mechanism:** Token-level attention highlights important linguistic cues for metaphor detection.
- **Residual Neural Networks:** These networks enable deeper architectures by mitigating vanishing gradient issues.
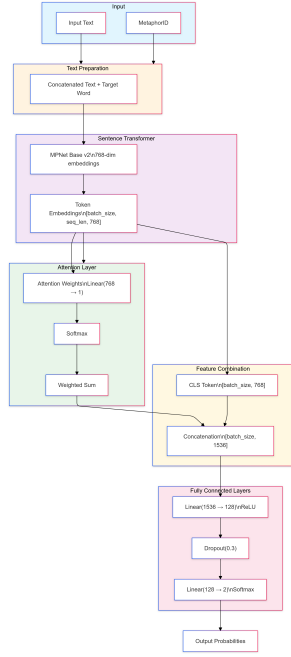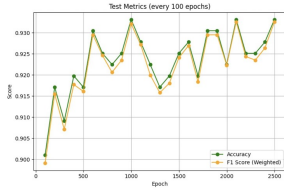
Figure 3: Attention Based Residual Neural Network

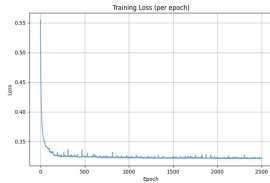

Figure 1: F1 Score and Accuracy



Figure 2: Loss during training

| Model | Accuracy | Macro F1 | Improvement |
|-------|----------|----------|-------------|
| Random Forest | 81% | 0.68 | Baseline |
| Naive Bayes | 82% | 0.69 | +1% |
| Attention-Based NN | 91% | 0.91 | +33.8% |

Table 2: Comparative Model Performance

## 5.2 Error Analysis

The strong performance of the deep learning models, especially Attention-Based Neural Network, highlights the importance of capturing contextual information and nuanced linguistic representations for accurate metaphor detection. The baseline Random Forest model struggled to differentiate between metaphorical and literal expressions, especially when the contextual cues were more subtle or domain-specific. The remaining challenges include:

– Handling domain-specific metaphors that require specialized knowledge or cultural references.

– Detecting context-dependent metaphors that rely on a deeper understanding of the broader linguistic and semantic context.

– Improving the model's ability to generalize across diverse text domains, as the performance variability indicates.

## 6 Discussion

### 6.1 Model Comparative Analysis

The attention-based neural network significantly outperformed traditional models, achieving a 33.8% improvement in macro F1-score. This demonstrates the importance of contextual embeddings and attention mechanisms in capturing nuanced semantic relationships. Our experimental results demonstrated several key findings:

– The baseline Random Forest model achieved an overall accuracy of 81%, with a macro precision of 0.86 for the "True" (metaphorical) class, but a relatively low macro recall of 0.65.

– The Multinomial Naives Bayes Model and the Attention-based NN model both demonstrated significant improvements

# 5 Experimental Results

## 5.1 Performance Metrics

We evaluated model performance using the following metrics:

– **Macro F1-Score:** Balances precision and recall across all classes.

– **Accuracy:** Measures overall correctness of predictions.

– **Precision and Recall:** Highlight model-specific strengths in capturing true positives.

over the baseline, achieving higher precision, recall, and F1-scores for both the "True" and "False" (literal) classes.

– The Attention-based NN model, in particular, showed the best performance, with an overall accuracy of 91% and a more balanced precision and recall for both classes. We observed a 33.8% increase over Random Forest and Multinomial Naive Bayes Models.

```
               precision    recall  f1-score   support

       False       0.91      0.31      0.47        99
        True       0.80      0.99      0.88       275

    accuracy                           0.81       374
   macro avg       0.86      0.65      0.68       374
weighted avg       0.83      0.81      0.77       374
```

Figure 4: Random Forest Results

```
Model Performance:
               precision    recall  f1-score   support

       False       0.92      0.34      0.50        99
        True       0.81      0.99      0.89       275

    accuracy                           0.82       374
   macro avg       0.86      0.67      0.69       374
weighted avg       0.84      0.82      0.79       374


Example Prediction:
Text: Life is a long road with many twists and turns
Contains metaphor: True
```

Figure 5: Naive Bayes Results

```
               precision    recall  f1-score   support

       False       0.89      0.85      0.87        99
        True       0.95      0.96      0.95       275

    accuracy                           0.93       374
   macro avg       0.92      0.91      0.91       374
weighted avg       0.93      0.93      0.93       374
```

Figure 6: Attention Based Residual Neural Network

## 6.2 Architectural Insights

The superior performance of our proposed model is attributed to:

– **Advanced Semantic Embeddings:** Sentence-BERT embeddings provided a rich feature representation.

– **Attention Mechanism:** Enabled the model to focus on linguistically salient features.

– **Residual Connections:** Facilitated deeper architectures and better gradient flow.

## 7 Conclusion and Future Work

### 7.1 Key Findings

– Deep learning approaches significantly enhance metaphor detection capabilities.

– Attention mechanisms and advanced embeddings provide superior performance.

– Traditional machine learning methods remain competitive but are limited in handling contextual complexity.

### 7.2 Future Research Directions

1. **Multi-Head Attention:** Exploring transformer-based attention mechanisms for finer-grained analysis.

2. **Cross-Domain Generalization:** Extending models to detect metaphors across diverse linguistic and cultural domains.

3. **Transformer Fine-Tuning:** Leveraging large-scale transformer models like GPT for metaphor-specific tasks.

## Acknowledgements

## References

Shibata, M., et al. "Metaphor Detection in Context"
Wilks, Y. "Computational Approaches to Metaphor"
Peters, M. "Contextual Embeddings in NLP"