Comparative Study of Deep Learning Models for News Text Classification Using CNN, LSTM, and BiGRU

A Research Presentation

Mohammad Varaliya, Abhishek Rane, Ayush Bhalgat, Shreyash Patekar, Dhruv Panchal

K. J. Somaiya Institute of Management, Somaiya Vidyavihar University

2023

Abstract & Overview

- Text classification as a **fundamental challenge** in the digital age
- Comparative analysis of **CNN, LSTM, and BiGRU** on News Category Dataset
- Evaluation of 5 models: 2 traditional baselines (Logistic Regression, SVM) and 3 deep learning variants
- **BiGRU achieved superior performance** with 80.94% accuracy and 80.42% weighted F1-score
- Highlights importance of bidirectional sequential context in text representation

Research Motivation & Problem Statement

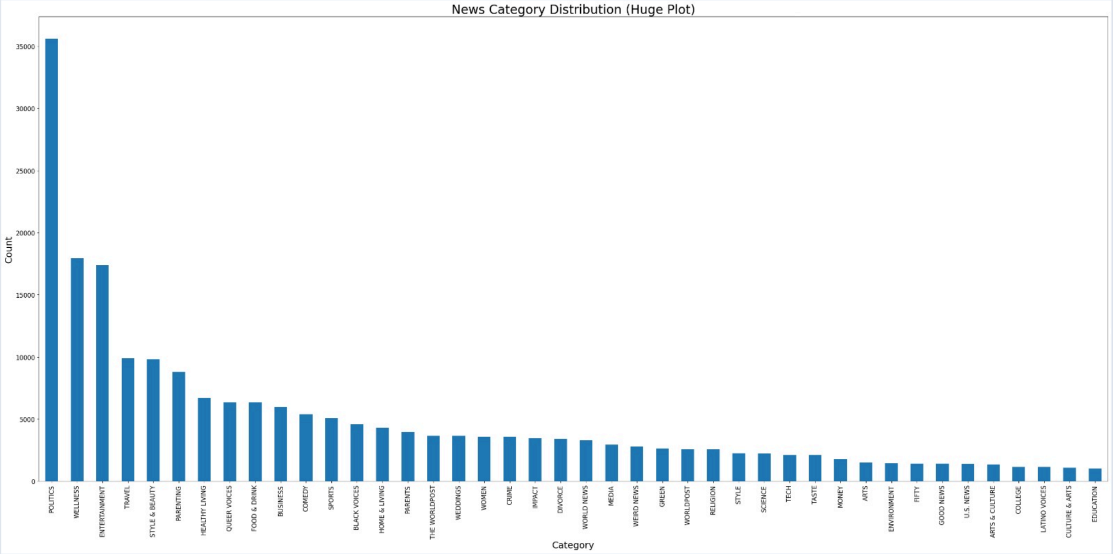
- Exponential growth of online news platforms demands efficient automated systems
- News classification enables personalized content delivery and recommendation engines
- Traditional approaches depend on manually engineered features that fail to capture semantic depth
- Deep learning models learn hierarchical feature representations directly from raw text
- Different architectures make different tradeoffs: CNN (local patterns), LSTM (sequential dependencies), BiGRU (bidirectional context)

Objectives and Research Questions

- **Primary Objective:** Compare deep learning architectures for news text classification
- **Research Question 1:** How do CNN, LSTM, and BiGRU perform on multi-class news classification?
- **Research Question 2:** What are the convergence patterns and training dynamics of each architecture?
- **Research Question 3:** Which architecture provides the best balance of computational efficiency and classification accuracy?
- **Hypothesis:** Bidirectional sequential processing (BiGRU) will outperform unidirectional approaches

Dataset Description

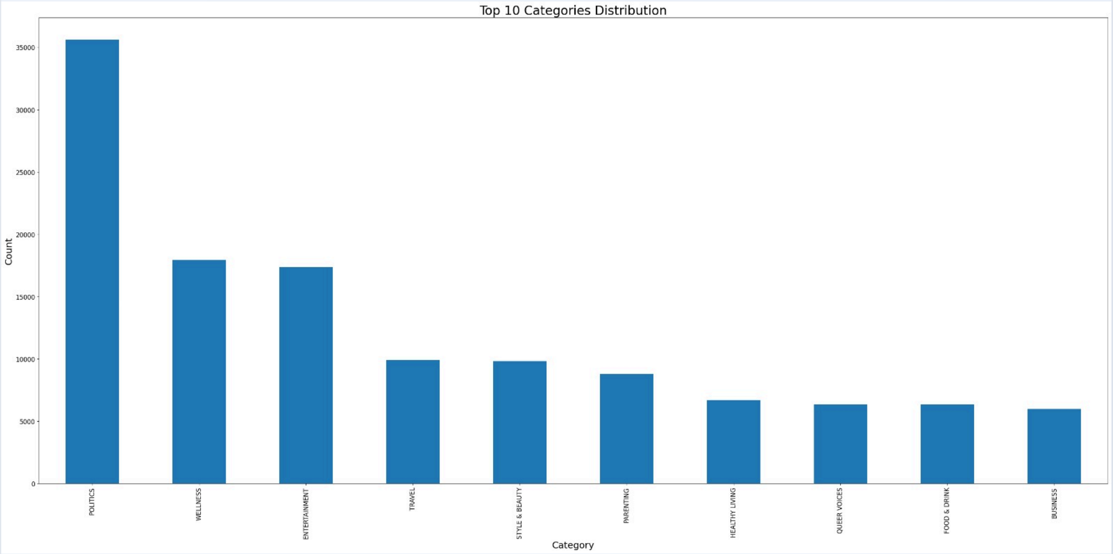
- **Source:** News Category Dataset from Huffington Post (Kaggle)
- **Original size:** 209,527 news articles with complete metadata
- **Final corpus:** 124,787 articles from top 10 most frequent categories
- **Data split:** 72% training (90,157), 13% validation (15,911), 15% test (18,719)
- **Text representation:** Combined headline and short description fields



News Category Distribution (Huge Plot)

Exploratory Data Analysis

- **Text preprocessing:** Lowercase conversion, special character removal, stopword removal, tokenization, lemmatization
- **Sequence preparation:** Standardized to maximum length of 100 tokens
- **Feature representation:** TF-IDF for traditional models, sequence indices for deep learning models
- **Category distribution:** Natural imbalance preserved with POLITICS dominating (~35,000 samples)
- **Data integrity checks:** Verified no null values and maintained category distribution



Top 10 Categories Distribution

Methodology Overview

Data Collection → Preprocessing → Feature Representation → Model Training → Evaluation

- **Traditional models:** Logistic Regression and SVM with TF-IDF features
- **Deep learning models:** CNN, LSTM, and BiGRU with sequence embeddings
- **Evaluation metrics:** Accuracy, F1-Score, Loss, Confusion Matrix
- **Optimization:** Adam optimizer with categorical cross-entropy loss
- **Hyperparameters:** 128-dimensional embeddings, maximum sequence length of 100 tokens
- **Data split:** 72% training, 13% validation, 15% test

Model Architectures

CNN

- Extracts local phrase-level features through convolutions
- Multiple filter sizes (3, 4, 5-grams) capture different n-gram patterns
- Global max pooling across each filter's output
- Computationally efficient with parallel processing

LSTM

- Captures long-term dependencies through gating mechanism
- Three gates: input, forget, and output gates
- Sequential processing maintains state across sequences
- Prone to vanishing gradients and convergence issues

BiGRU

- Processes sequences bidirectionally for complete context
- Simplified gating (reset and update gates) vs LSTM
- More computationally efficient than LSTM
- Robust convergence with stable training dynamics

Model Evaluation Metrics

Accuracy

Proportion of correctly classified instances out of total instances

80.94%

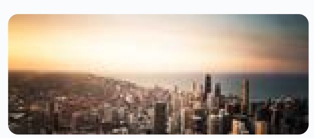
F1-Score

Harmonic mean of precision and recall, weighted for class imbalance

```
F1 = 2 × (Precision × Recall) / (Precision + Recall)
```

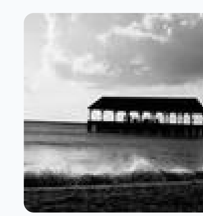
Loss

Measure of model error during training (categorical cross-entropy)



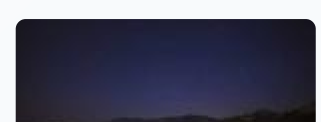
Confusion Matrix

Table showing actual vs. predicted classifications for each category



Training Dynamics

Convergence patterns, overfitting detection, stability across epochs



Results - CNN

- **Performance:** 79.76% accuracy, 79.16% F1-Score
- **Training time:** ~8 minutes (20 epochs)
- **Convergence:** Smooth, stable training curve
- **Key insight:** CNNs capture local phrase patterns overlooked by bag-of-words approaches
- **Limitation:** Restricted receptive field struggles with long-range dependencies

CNN Accuracy and Loss

CNN Accuracy and Loss

79.76%	79.16%	8 min
Accuracy	F1-Score	Training Time

Results - LSTM

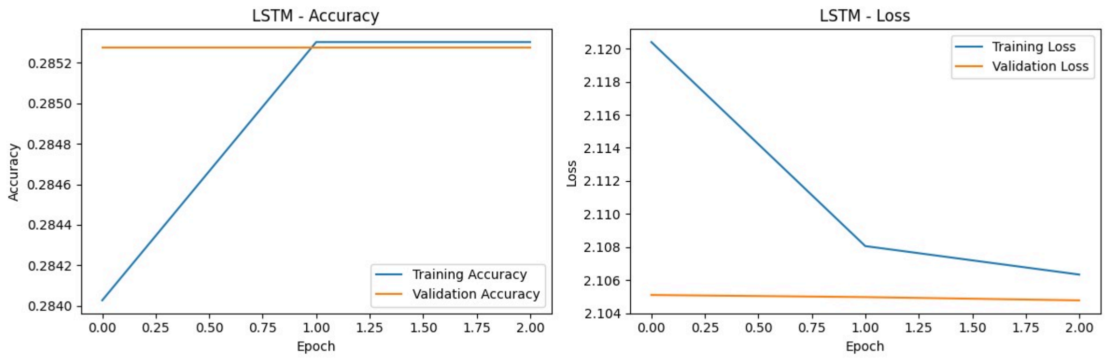
- **Performance:** 28.53% accuracy, 12.66% F1-Score
- **Training time:** ~15 minutes (20 epochs)
- **Convergence:** Severe divergence; validation accuracy remained below 30%
- **Root causes:** Hyperparameter sensitivity, convergence difficulties, recurrent dropout effects
- **Key insight:** Theoretical advantages do not guarantee practical success

LSTM performance was shockingly poor, achieving accuracy barely above random chance (10% for 10 classes)

28.53%
Accuracy

12.66%
F1-Score

15 min
Training Time



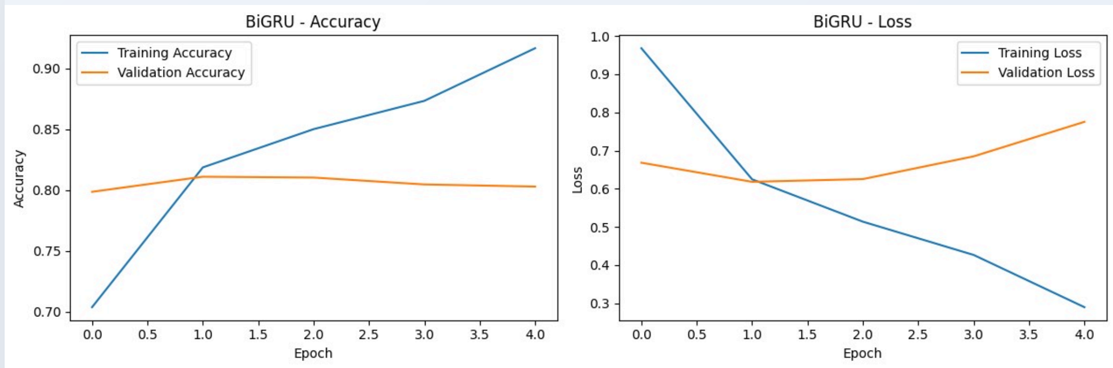
LSTM Accuracy and Loss

Results - BiGRU

- **Performance:** 80.94% accuracy, 80.42% F1-Score (best results)
- **Training time:** ~10 minutes (20 epochs)
- **Convergence:** Smooth, stable training curve with clear improvement trajectory
- **Advantages:** Bidirectional context, simplified gating, computational efficiency
- **Category performance:** Politics achieved particularly high accuracy (4,854/5,400)

> **BiGRU achieved the highest performance among all models tested, outperforming CNN by 1.2% and baselines by 3%**

80.94%	80.42%	10 min
Accuracy	F1-Score	Training Time



BiGRU Accuracy and Loss

Confusion Matrix - BiGRU

Comparison and Insights

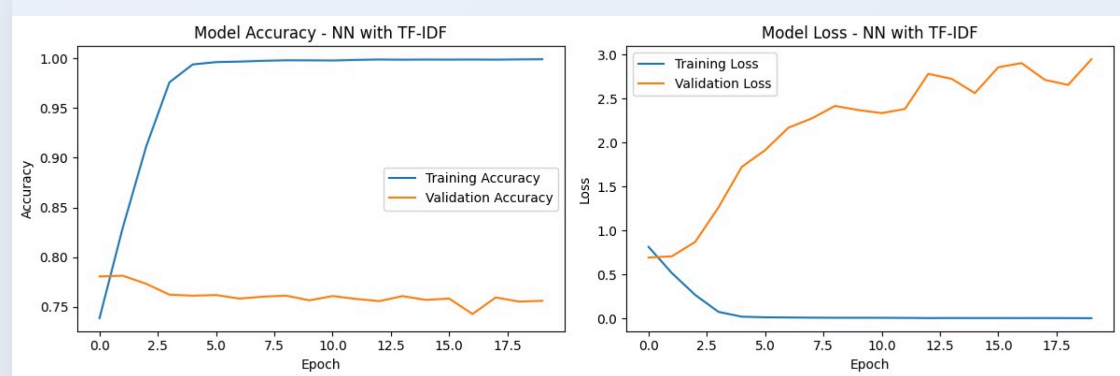
- **BiGRU achieved highest performance:** 80.94% accuracy, 80.42% F1-score
- **CNN improved over baselines** by approximately 2% accuracy
- **LSTM performed poorly** due to convergence challenges
- **Traditional baselines** provide formidable competition (Logistic Regression: 77.91%, SVM: 77.79%)
- **Key insight:** Bidirectional sequential context is crucial for news classification

Model Accuracy Comparison

Model Accuracy Comparison

> **BiGRU's 3% improvement over baselines represents approximately 2,360 additional articles correctly classified out of 18,719 test samples**

Model	Accuracy	F1-Score
BiGRU	**80.94%**	**80.42%**
CNN	79.76%	79.16%
Logistic Regression	77.91%	76.90%
SVM	77.79%	76.83%
LSTM	**28.53%**	**12.66%**



Model Accuracy - NN with TF-IDF

Conclusion & Future Work

- **Key finding:** BiGRU outperforms all other architectures for news text classification
- **Bidirectional sequential context** is crucial for capturing semantic relationships in news text
- **Traditional baselines** provide formidable competition that deep learning must meaningfully surpass

Future Improvements

Pretrained embeddings (GloVe, Word2Vec)	Transformer-based models (BERT, RoBERTa)
Ensemble methods	Hyperparameter optimization

Practical recommendation: BiGRU as a strong starting point balancing performance, training efficiency, and implementation simplicity

References & Acknowledgments

Key References

- Kim, Y. (2014). Convolutional neural networks for sentence classification. EMNLP.
- Cho, K., et al. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. EMNLP.
- Schuster, M., & Paliwal, K. K. (1997). Bidirectional recurrent neural networks. IEEE Transactions on Signal Processing.
- Conneau, A., et al. (2016). Very deep convolutional networks for text classification. EACL.

Acknowledgments

- Prof. Mayura Nagar for mentorship and guidance
- Open-source community for PyTorch, TensorFlow, and supporting libraries
- Kaggle and Huffington Post for the News Category Dataset
- K. J. Somaiya Institute of Management for computing resources

Thank You!