UNVEILING MISINFORMATION IN BANGLA SOCIAL MEDIA: A MULTIMODAL FAKE NEWS DETECTION SYSTEM

Presented By:

Samanta Islam NIshu
20210104041

Zarin Tasnim Roichi
20210104111

Afrain Akhter
20210104113

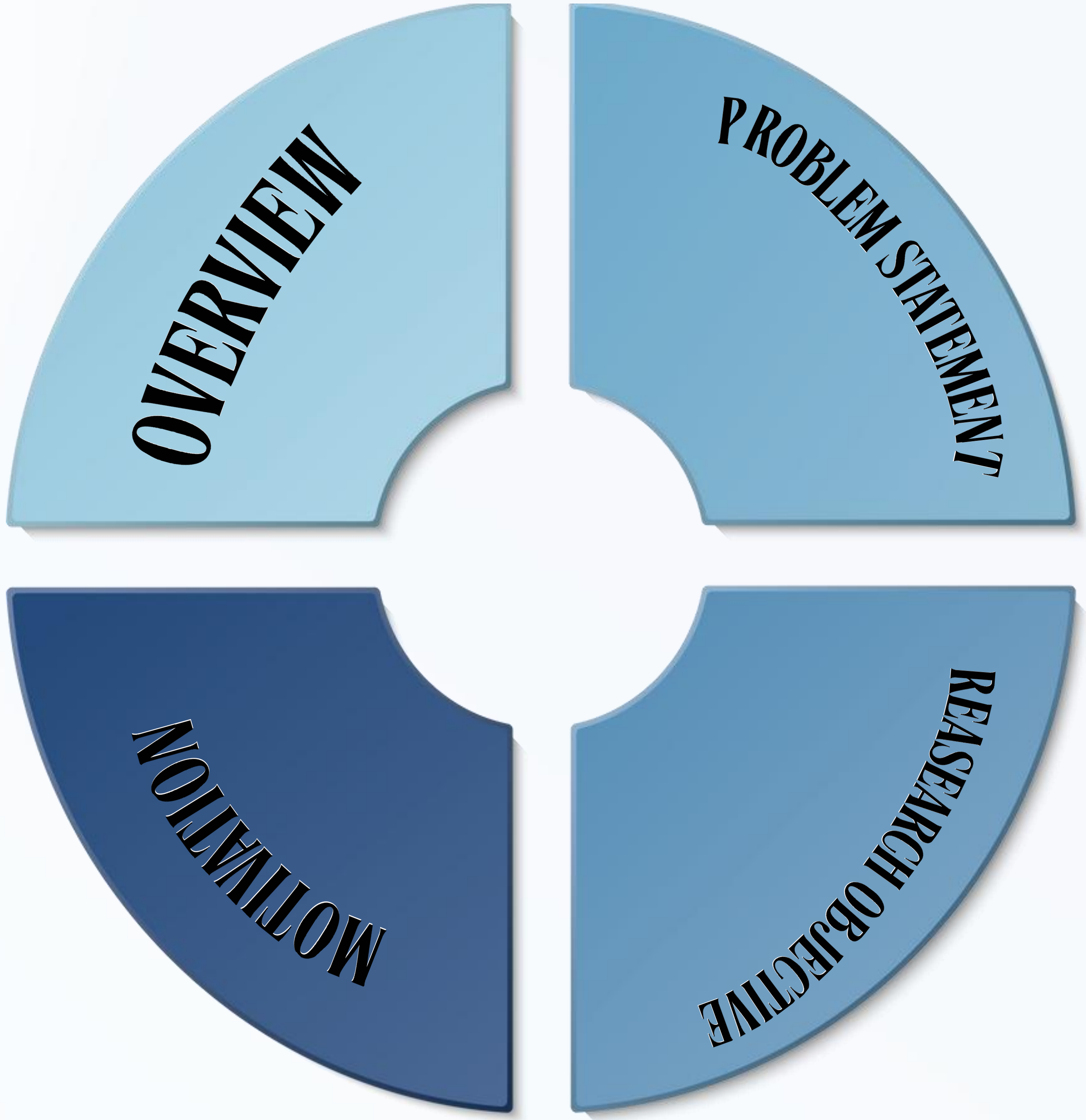
Supervised By:

Sayeda Sabnam Hasan

Assistant Professor

TABLE OF CONTENT

INTRODUCTION



Overview

- Social media makes spreading fake news easy—often using both text and images.
- Bangladesh faces high risks due to low digital literacy and limited detection tools.
- Most systems are text-only and built for English.
- Our work introduces a Multimodal Bengali Fake News Detection System using:
  - Text models: BanglaBERT, XLM-R
  - Image models: CNN, ResNet-50, Swin, ViT
  - Combined via a stacked ensemble meta-classifier.

Problem Statement

- Most Bengali fake news detectors are text-only, ignoring misleading images.
- No multimodal dataset exists for Bengali (text + image with labels).
- Bengali is low-resource: lacks pretrained models and benchmarks.
- Our solution: A multimodal ensemble system combining text and image to detect fake news effectively.

Research Objectives

We aim to develop a robust and scalable fake news detection system for the Bengali language by leveraging both text and image modalities. The goal is to design an ensemble-based multimodal framework and make it usable through a practical real-time interface for end-users.

- **Text Modeling – Fine-tune BanglaBERT, XLM-R, and BanglaLM to detect fake textual content in Bengali.**

- **Image Modeling – Train CNN, ResNet50, ViT, and Swin Transformer on manipulated images in Bangla news.**

- **Multimodal Fusion – Combine text and image model outputs using an MLP-based stacking ensemble.**

- **Interface Development – Build a Python-based demo with prediction options for text-only, image-only, and combined inputs.**

Motivation

01
Impact

Real-world harm in Bangladesh—such as communal violence, political unrest, and health panic—is often fueled by fake news. UNICEF reports misinformation as a major source of stress for youth, underscoring its societal consequences.

02
Data & Tool Scarcity

Bengali lacks large-scale multimodal datasets and powerful pretrained models. Most tools are text-only and designed for English, making them ineffective in detecting image-text misinformation in local contexts.

03
Multimodal Misinformation

Modern misinformation uses text + images to increase believability. A system that handles only one modality often fails to detect such complex fakes. Multimodal fusion is crucial for real-world relevance.

04
Deployment Potential

There is no real-time fake news detection tool tailored for Bangla users. Developing a system that can be deployed as a web app or browser extension offers real, user-centric impact.

BACKGROUND STUDY

Text Classification

Image Classification

EnsembleTechniques

Multimodal Fusion

Text Classification

We used deep learning and transformer-based models for Bangla fake news detection.

- **BanglaBERT (7 variants)**
- **XLM-RoBERTa (Base)**
- **BanglaLM (planned)**

Each model outputs softmax probabilities for “Fake“ or “Non-Fake“.

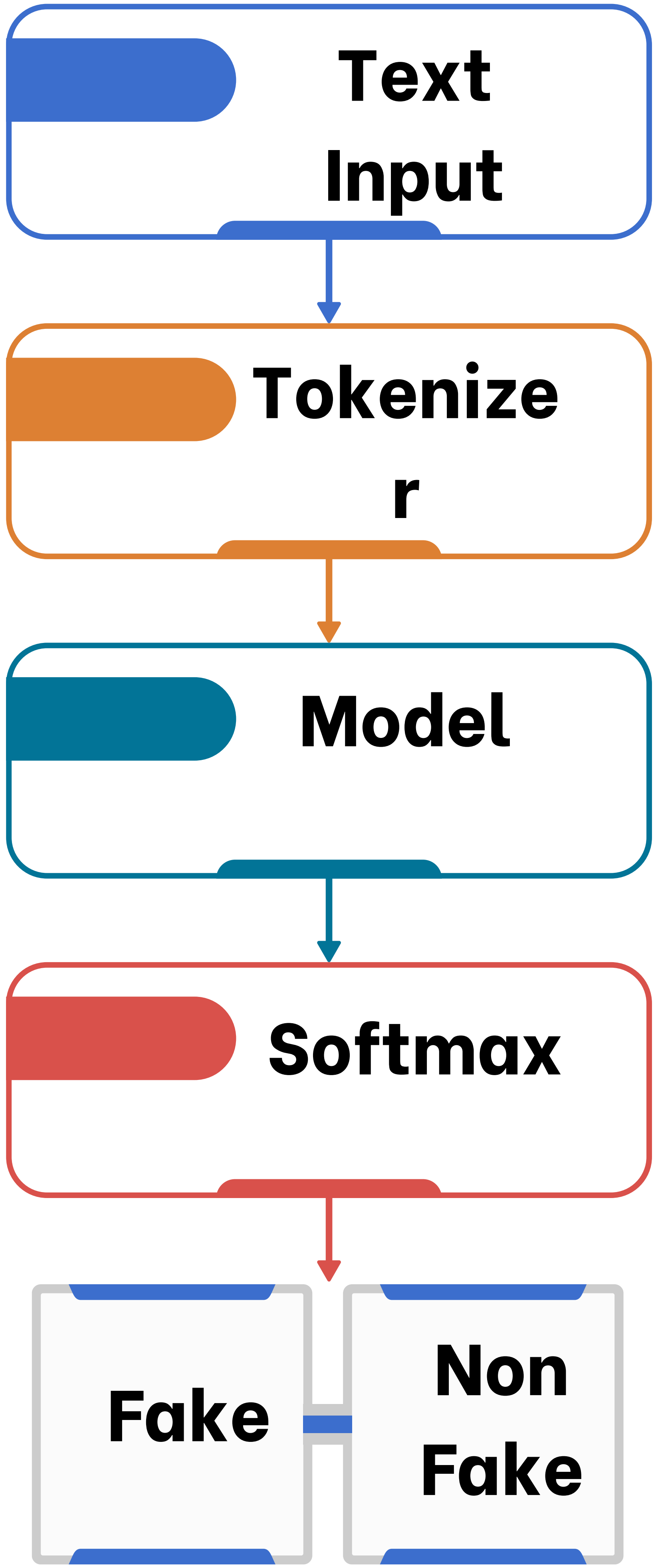
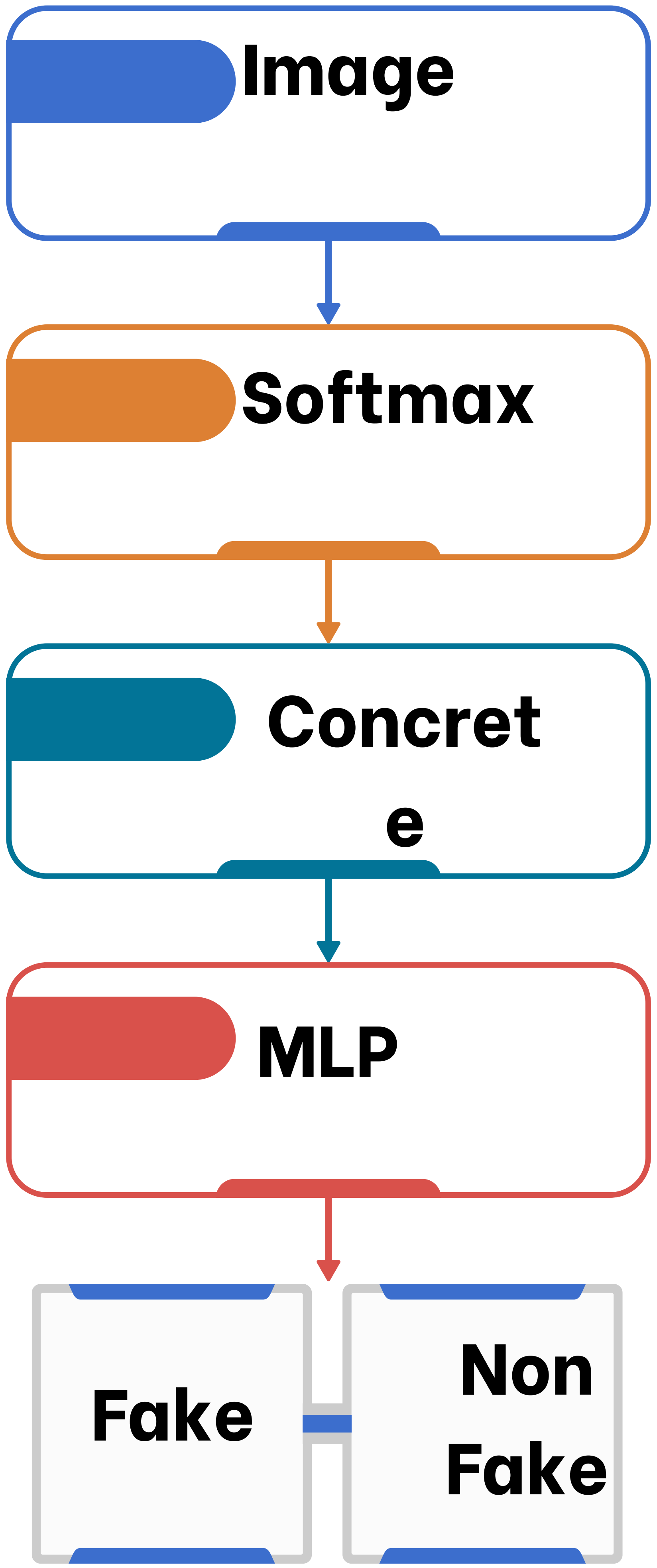


Image Classification

- Used deep learning models to detect visual misinformation in Bangla news and social media.
- Trained on fake/non-fake image dataset
  - Models used:
  - CNN (Custom baseline)
  - ResNet50 (Residual Network)
  - EfficientNet (B0 and B2)
  - Vision Transformer (ViT)
  - Swin Transformer (Tiny)



Ensemble Techniques

- Combines softmax outputs from multiple fine-tuned transformer models.
- Uses a shallow MLP as a meta-classifier to fuse predictions.
- Captures complementary strengths, improving generalization and reducing overfitting.
- Achieved 2–5% F1-score improvement over individual models.
- Easily extendable to include image-based classifiers for multimodal detection.

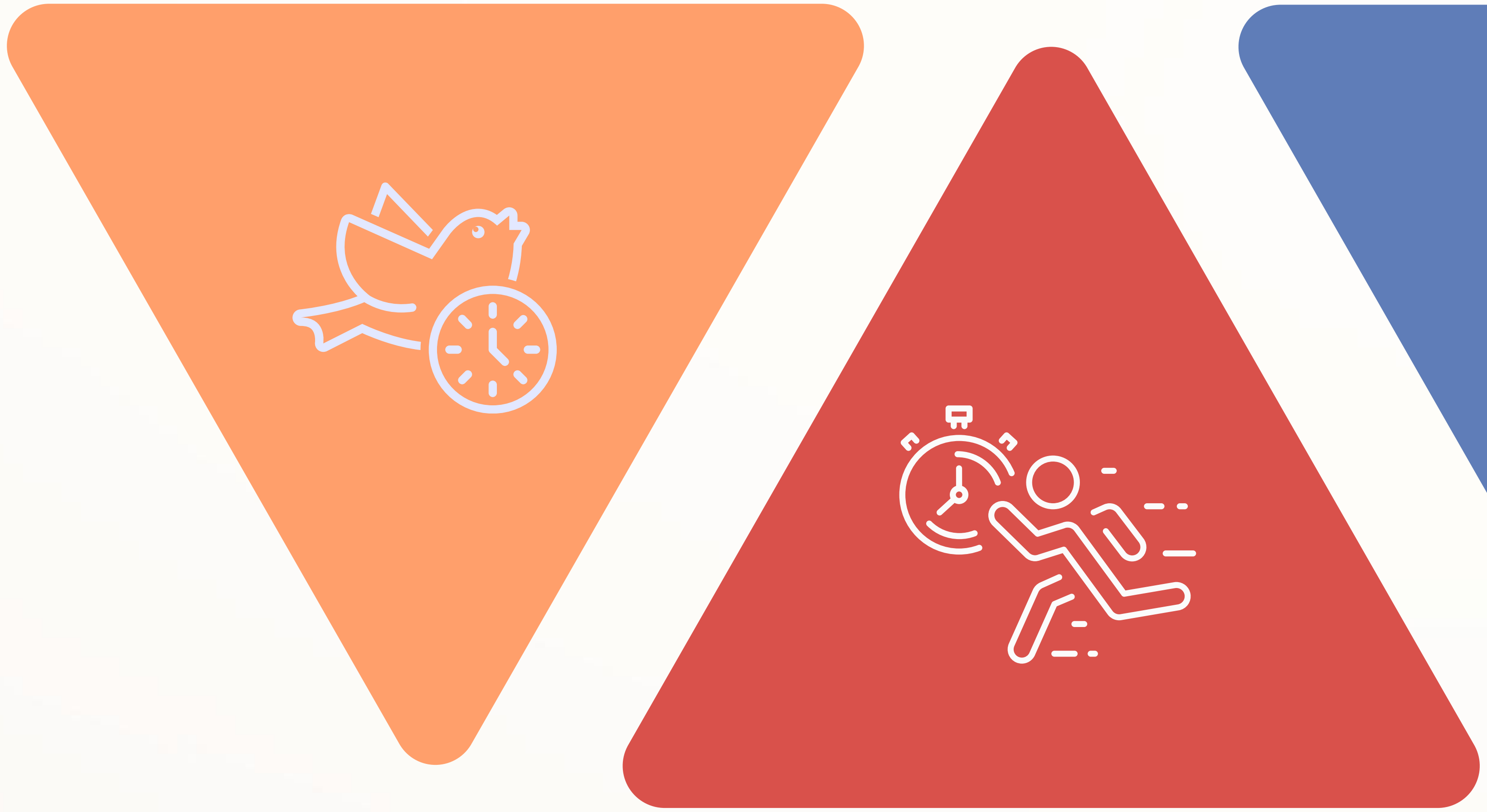
Multimodal Fusion

Early Fusion

Concatenates raw or extracted features from both modalities before classification, capturing deep cross-modal interactions but may require careful normalization.

Intermediate Fusion

Merges higher-level features using attention mechanisms, efficiently capturing complex interactions and providing more expressive fusion than early or late methods.



Late Fusion

Combines predictions from separate text and image models via weighted averaging or voting, offering simplicity and robustness but missing deeper interactions.

LITERATURE REVIEW

Bangla Fake News Detection

Advanced Multimodal Architectures

Multimodal Detection

LLM-Based & Reasoning Methods

Multilingual & Low-Resource Approaches

Research Gaps & Opportunities

Some Literature Reviews

Paper Name	Year	Methodology Used	Results
Knowledge-Driven Bangla Fake News Detection	2025	Hybrid model: Semantic, Sentiment, Knowledge Graph features	Significantly outperformed baseline text models
RoBERTa-GCN for Bengali Text	2024	RoBERTa embeddings + Graph Convolutional Network (GCN)	Achieved 98.6% accuracy
SDML: Multimodal Shallow-Deep Learning	2025	Contrastive learning, Dual-branch multimodal architecture	High accuracy on large-scale multimodal dataset
Graph-Augmented Reasoning with GPT	2024	GPT-3.5 with graph reasoning for explainability	Improved robustness and explainability
GAMED: Gated Expert Decoupling	2024	Dynamic gating of image and text experts	Enhanced accuracy and interpretability
Multilingual Fake News Detection in Low-Resource Languages	2023	Transfer learning with multilingual transformers	Improved performance in low-resource settings
Visual-Textual Fusion for Fake News Detection	2023	Early fusion of CNN image features with LSTM text	Outperformed unimodal baselines by 5%
BERT-Based Fake News Classification	2022	Fine-tuned BERT for text classification	Accuracy improved by 7% over traditional methods
Ensemble Model for Multimodal Fake News Detection	2024	Ensemble of CNN, LSTM, and Transformer models	Robust detection with balanced precision and recall
Explainable Fake News Detection with Attention	2023	Attention-based BiLSTM with explainability framework	Better interpretability with 92% classification accuracy

Research Gaps & Opportunities

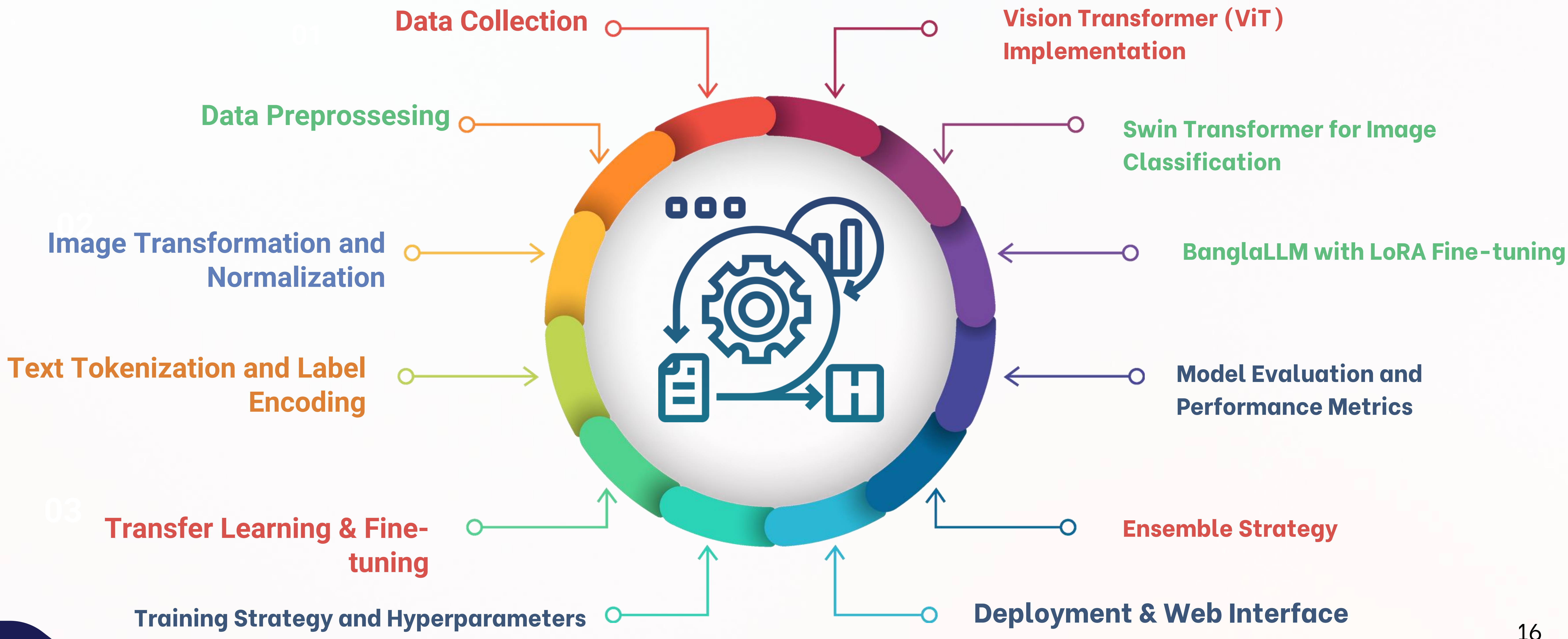
Gaps Identified:

- Transformer and GCN models show strong performance but often fail to generalize due to dataset bias.
- Multimodal fusion strategies are still shallow, limiting their effectiveness.
- Multilingual and LLM-based models lack proper fine-tuning for Bengali-specific contexts.
- Advanced fusion techniques like Mixture-of-Experts and hierarchical attention are rarely applied.

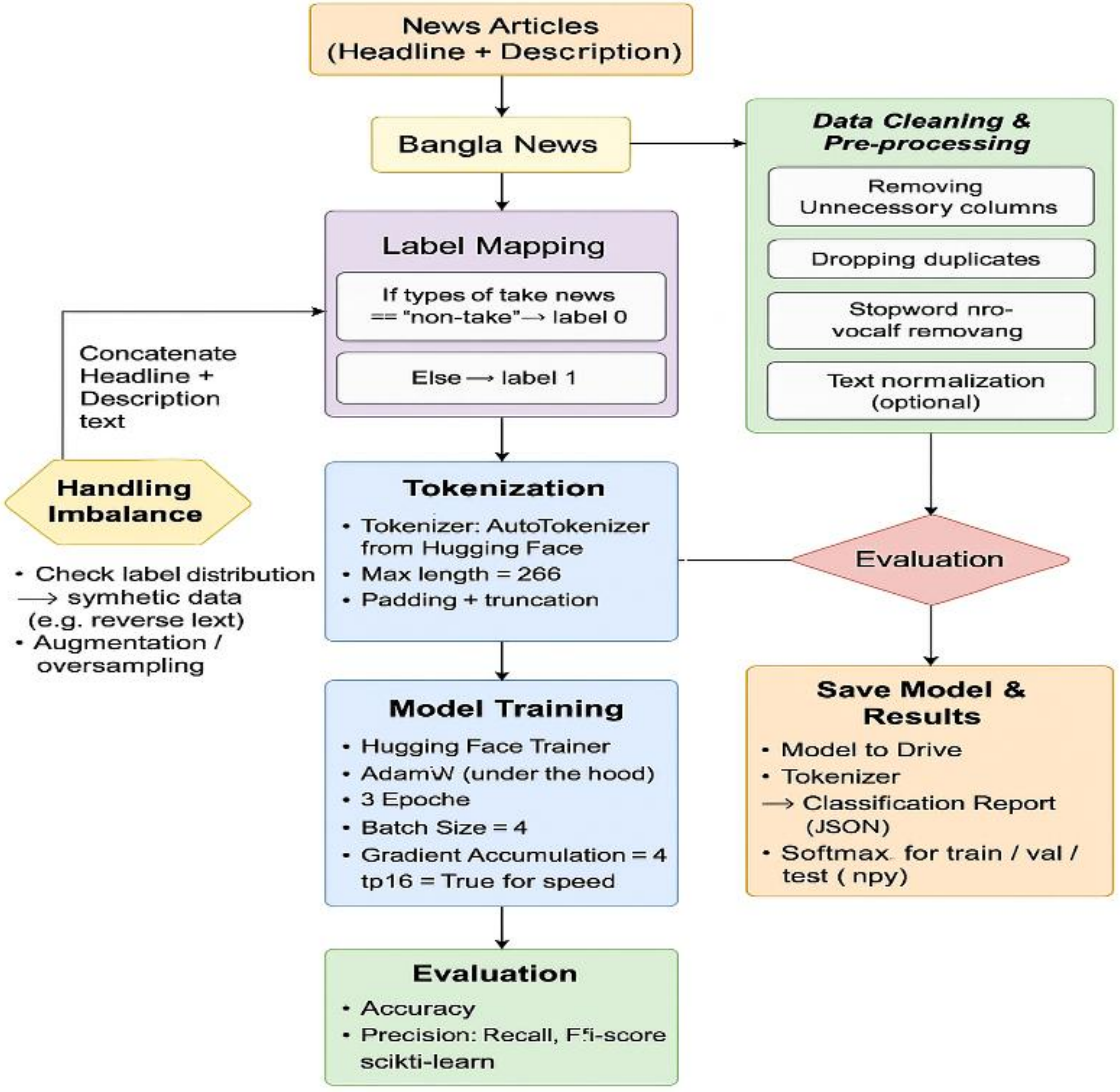
Research Opportunities:

- Develop large, balanced multimodal datasets in Bangla.
- Design lightweight and explainable models tailored for fake news detection.
- Implement guided fusion methods that balance high accuracy with interpretability.

METHODOLOGY



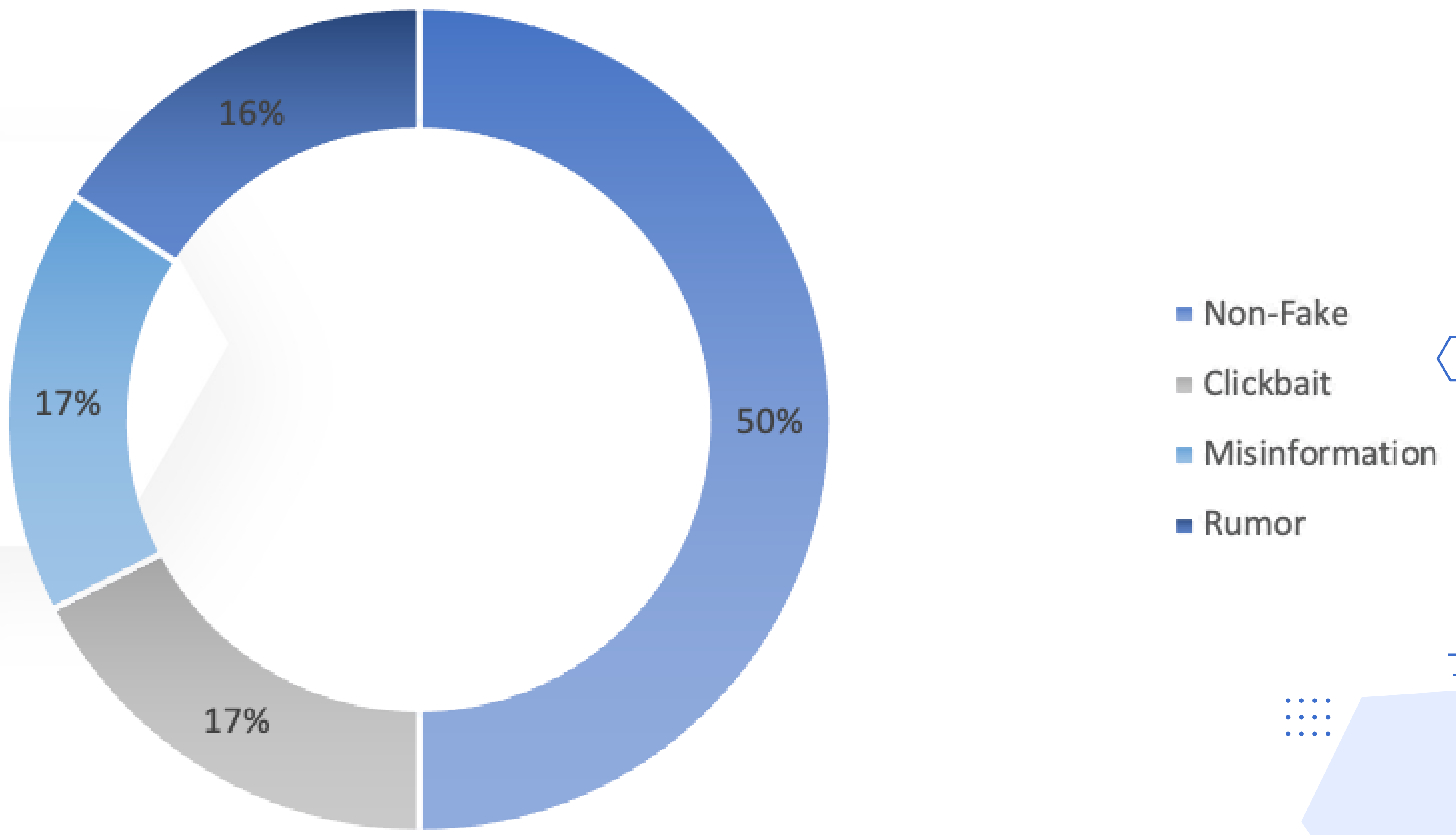
Visual Repesentation of Methodology



Bangla Fake News Dectection

Data Collection

- Used a public Bangla fake news dataset with 7,680 labeled articles.
- Binary classes:
  - **Non-Fake (label 0):** 3,840 articles
  - **Fake (label 1)**: 3,840 articles (**Clickbait, Misinformation, Rumor**)
- Dataset is balanced and supports effective binary classification.



Data Preprocessing

- Cleaned Bangla text by removing noise, symbols, and fixing casing.
- Replaced missing values with empty strings.
- Applied tokenization using model-specific **tokenizers.**
- Example:
- **Input: “এই সংবাদটি মিথ্যা।“**
- **Output: [’[CLS]’, ’এই’, ’সংবাদটি’, ’মিথ্যা’, ’।’, ’[SEP]’]**

Image Transformation and Normalization

01

Image Labeling & Organization

Images were automatically sorted into “Fake“ and “Real“ categories based on keywords in filenames. These were then organized into training, validation, and testing folders.

02

Format Filtering

Only standard image formats (e.g., .jpg, .png) were retained. Invalid or unlabeled images were excluded to ensure dataset quality.

03

Resizing & Normalization

All images were resized to 224×224 pixels to match Vision Transformer input size. Pixel values were normalized using predefined mean and standard deviation for consistent data distribution.

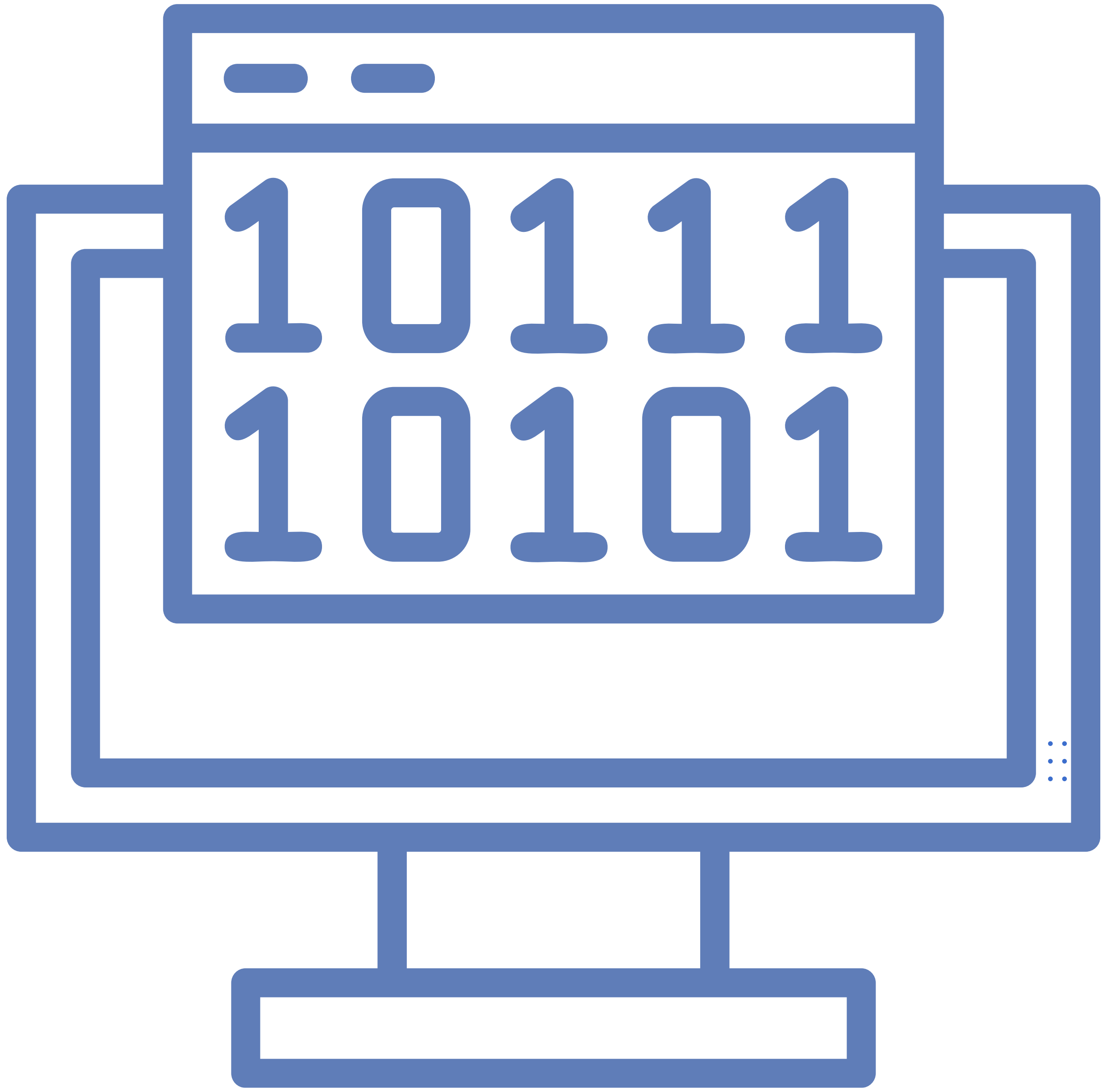
04

Tensor Conversion & Dataset Preparation

Images were converted into tensor format and paired with labels. These were wrapped into PyTorch datasets for efficient training and evaluation.

Text Tokenization and Label Encoding

- Converted raw text into tokens using pretrained tokenizers
- Applied padding and truncation for uniform input size
- Grouped all fake types under label 1, non-fake as 0
- Encoded labels for binary classification tasks



Transfer Learning & Fine-tuning

- Used pretrained models (XLM-RoBERTa, ViT) for text and image features.
- Applied fine-tuning on our Bangla fake news dataset to adapt to the task.
- This improves accuracy and saves training time by leveraging learned knowledge.

Training Strategy and Hyperparameters

Learning Rate
Used a low rate (2e-5) for gradual and stable training

Batch Size
Small batches (4 or 8) to fit memory constraints

Epochs
Trained 3-5 times to balance learning and overfitting

Weight Decay
Applied 0.01 to reduce overfitting by penalizing large weights

Mixed Precision
Enabled fp16 to speed up training and save GPU memory

Validation
Checked model performance after every epoch to prevent overfitting

ResNet50 Implementation

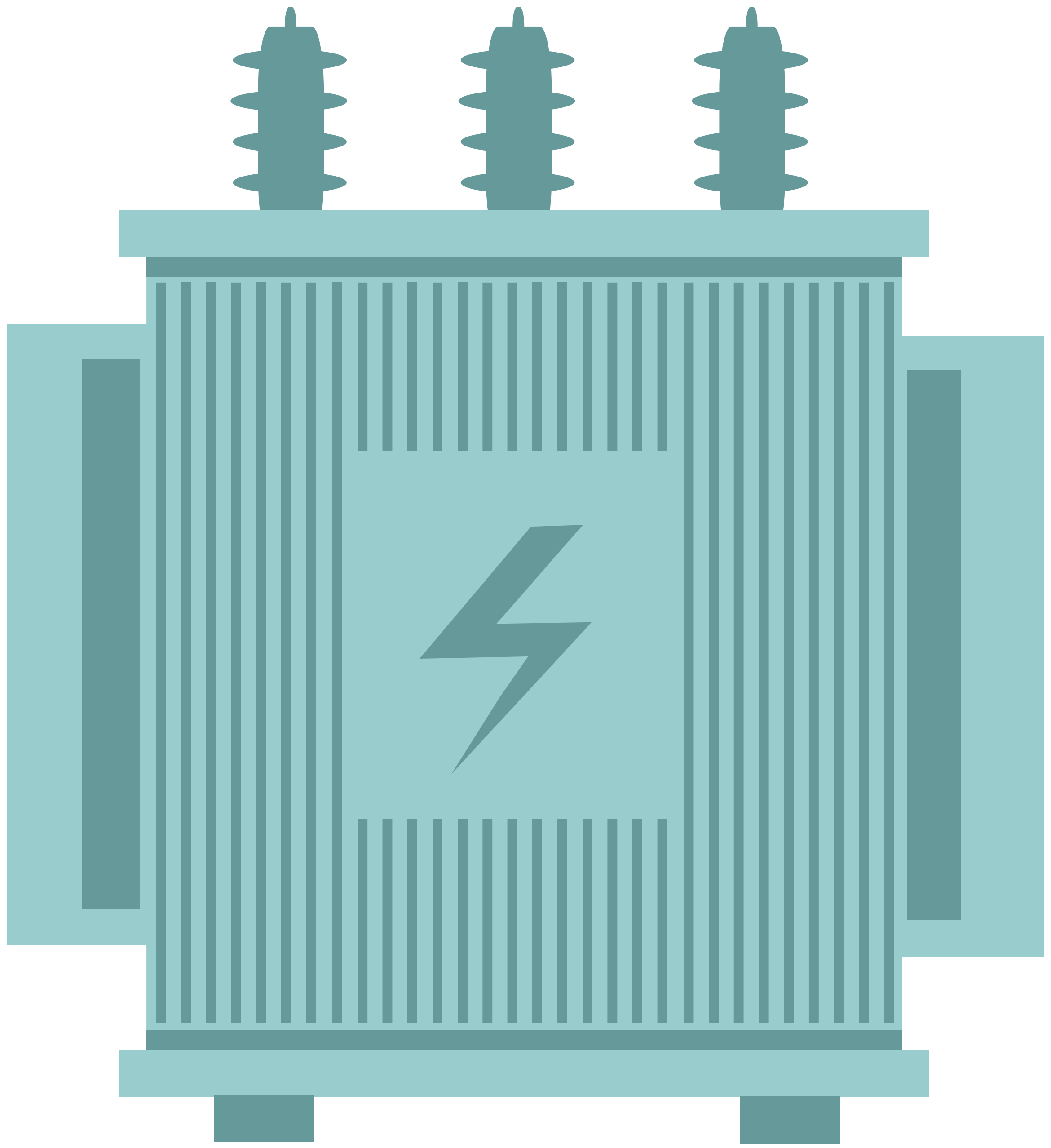
- Deep CNN with residual connections for efficient learning
- Fine-tuned on resized, normalized fake news images
- Used as a baseline model for visual classification
- Evaluated with accuracy and F1-score

Vision Transformer (ViT) Implementation

- Vision Transformer (ViT)
- Processes images as sequences of patches to capture global context for classification.
- Fine-tuned on resized and normalized images for fake news detection.

Swin Transformer for Image Classification

- Captures local and global features using shifted windows.
- Fine-tuned on resized, normalized images with mixed precision training for fake news detection.



BanglaLLM with LoRA Fine-tuning

- Efficiently fine-tuned a large Bangla language model using LoRA to reduce training cost and GPU memory.
- Trained on concatenated headlines and descriptions with mixed precision and gradient accumulation.
- Evaluated using accuracy and F1-score; outputs saved for ensemble use.

Model Evaluation & Metrics

- Evaluated all models on a **held-out test set**
- Used classification_report() from **sklearn.metrics**

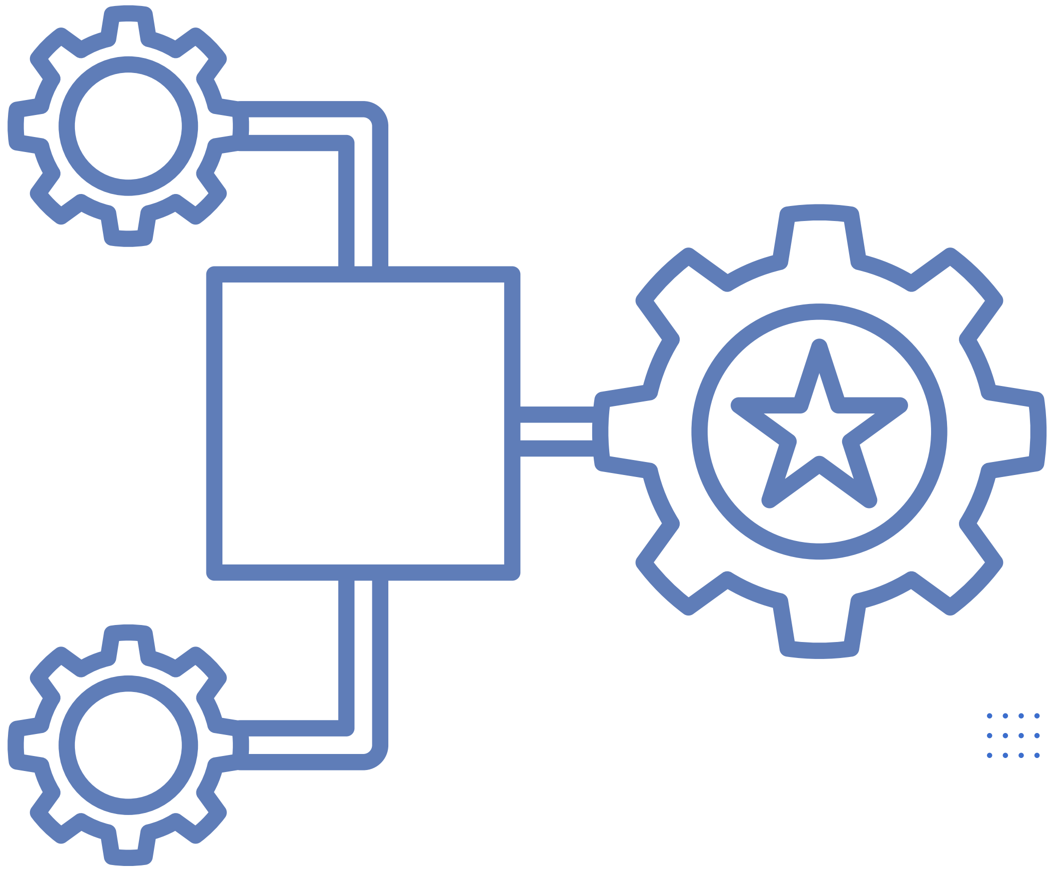
Metrics used:

- **Accuracy**
- **Precision**
- **Recall**
- **F1-Score**
- **Weighted F1-Score**

- **Softmax outputs** were saved for use in ensemble strategies

Ensemble Strategy

- Combined softmax scores from text and image models (e.g., XLM-R, BanglaBERT, ViT)
- Applied averaging to merge predictions across models
- Helped reduce individual model bias and improve overall accuracy
- Boosted reliability in classifying Fake vs. Non-Fake news in multimodal settings



Deployment & Web Interface

- After model training, we developed an interactive **Gradio-based demo**.

The demo allows:

- Text-only predictions.
- Built entirely in Python, supports real-time predictions.

Deployment & Web Interface

Future Plans:

- Deploy as a fully functional website
- Provide browser extension or API for real-time, user-friendly verification
- Support for Bangla language and accessible UI

Work Plans:

- Frontend: HTML/CSS or React for text & image input
- Backend: Flask/FastAPI to serve the trained model
- Model Integration: Load multimodal model for predictions
- Deployment: Host on Render/Heroku with GitHub CI/CD

PRELIMINARY RESULT

Text
Models

Image Models

LLM
Models

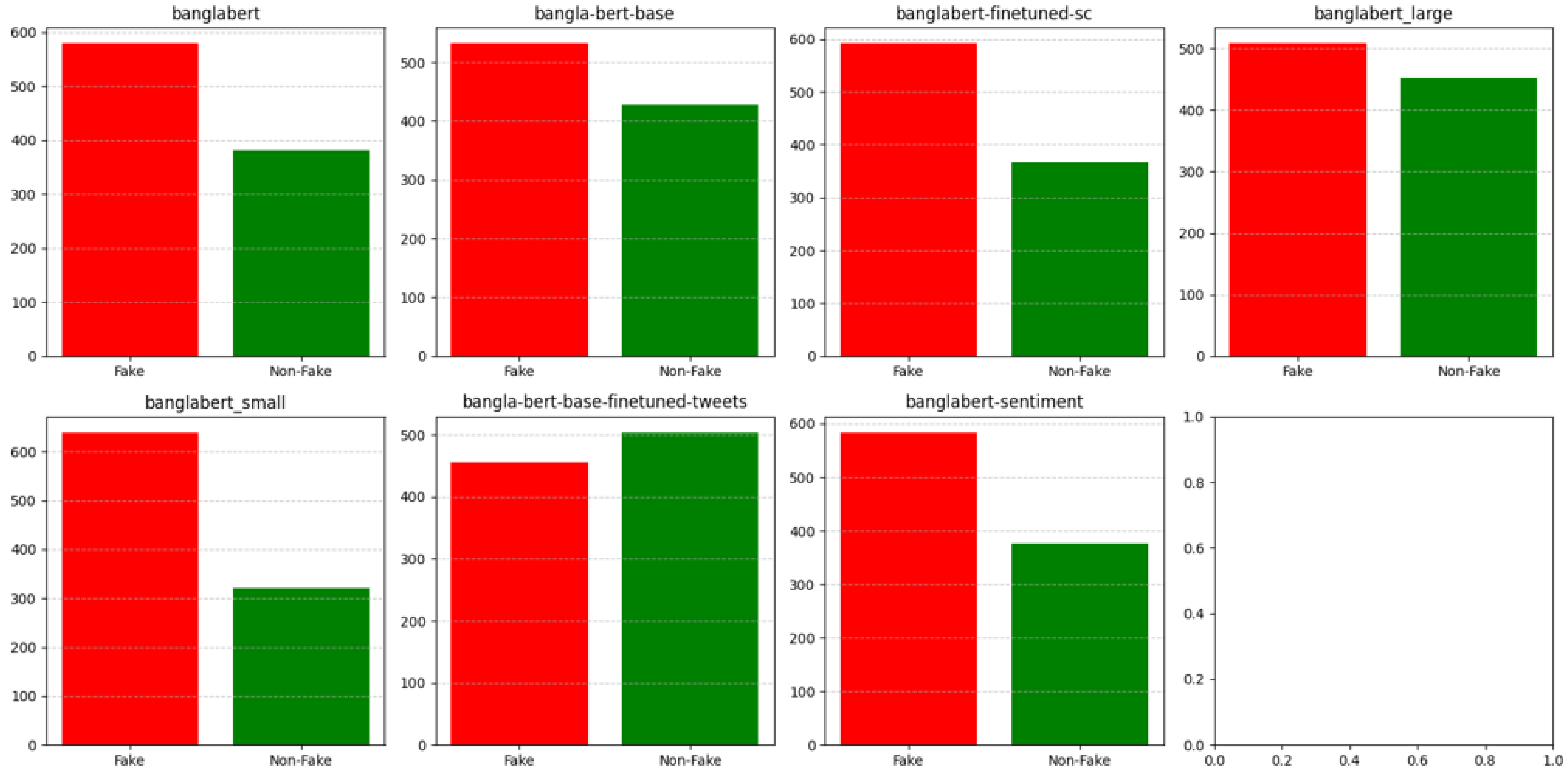
Model Used

1. BanglaBert
2. BanglaBert-Ensembled
3. XLM-R
4. ResNet50
5. LST M
6. VIT
7. BanglaLLM
8. Swin

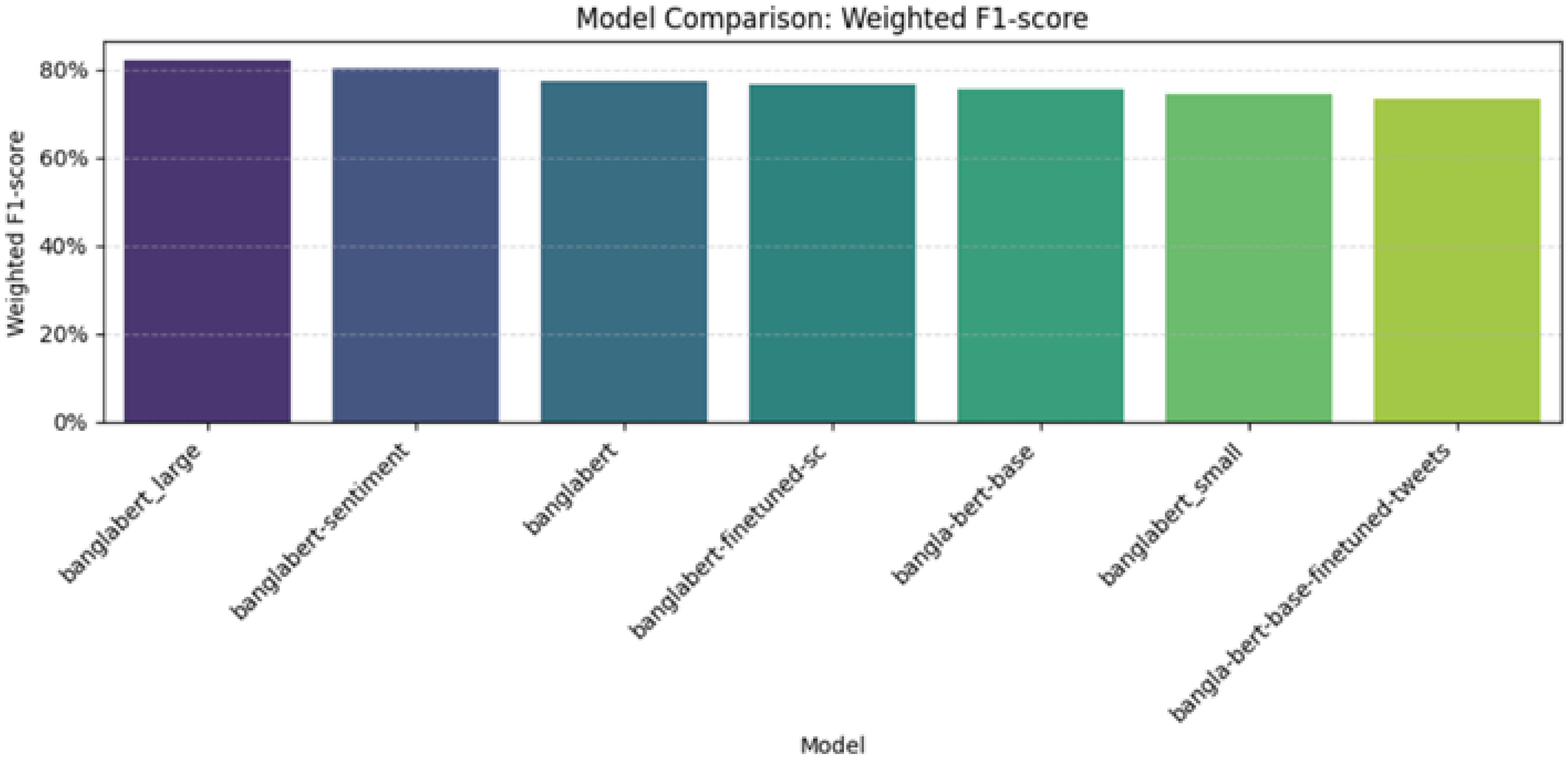
Evaluation Metrics of BanglaBERT Variants

	Model	Accuracy	Fake F1-score	Non-Fake F1-score	Weighted F1-score
0	banglabert_large	0.822917	0.827935	0.817597	0.822766
1	banglabert-sentiment	0.804167	0.823308	0.780374	0.801841
2	banglabert	0.776042	0.796978	0.750290	0.773634
3	banglabert-finetuned-sc	0.768750	0.792910	0.738208	0.765559
4	bangla-bert-base	0.758333	0.770751	0.744493	0.757622
5	banglabert_small	0.752083	0.787500	0.702500	0.745000
6	bangla-bert-base-finetuned-tweets	0.733333	0.726496	0.739837	0.733167

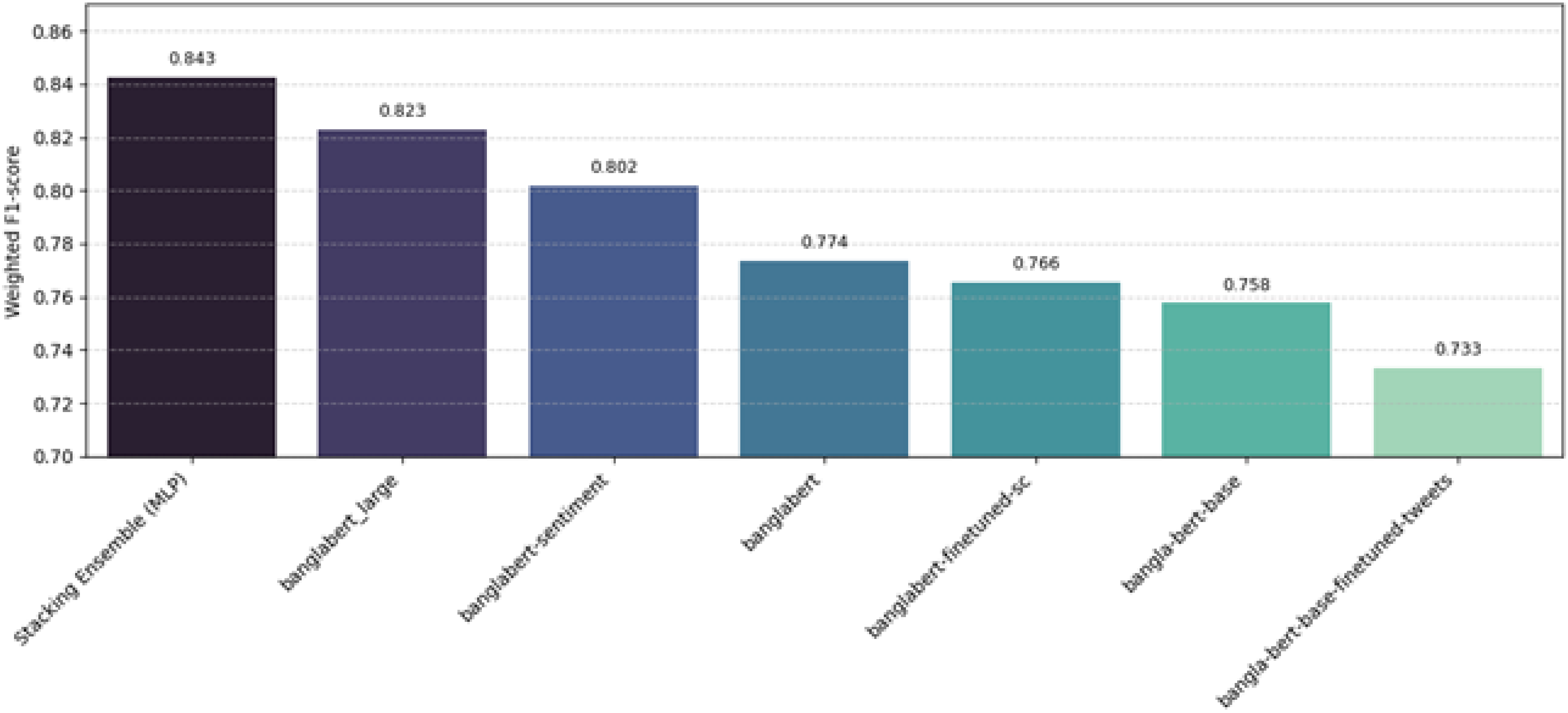
Prediction Distribution Across BanglaBERT Variants



Performance Comparison of BanglaBERT Variants



Performance Gain with Stacking Ensemble



XLM-R Evalution

Class	Precision	Recall	F1-Score	Support
Fake	0.7274	0.8729	0.7936	480
Non-Fake	0.8411	0.6729	0.7477	480
Accuracy			0.7729	960

LSTM(Long Short-Term Memory) Evaluation

Class	Precision	Recall	F1-Score	Support
Fake	0.55	0.42	0.47	480
Non-Fake	0.53	0.65	0.59	480
Accuracy			0.54	960

ResNet50 Evaluation

Class	Precision	Recall	F1-Score	Support
Fake	0.79	0.44	0.56	227
Non-Fake	0.60	0.88	0.71	218
Accuracy			0.65	445

Vision Transformer (ViT) Evaluation

Class	Precision	Recall	F1-Score	Support
Fake	0.6599	0.6143	0.6363	477
Non-Fake	0.6413	0.6854	0.6626	480
Accuracy			0.6495	957

Swin Transformer Evaluation

Class	Precision	Recall	F1-Score	Support
Fake	0.6201	0.5954	0.6075	477
Non-Fake	0.6132	0.6375	0.6251	480
Accuracy			0.6165	957

BanglaLLM/Bangla-s1k-llama-3.2-3B-Instruct

Class	Precision	Recall	F1-Score	Support
Fake	0.6685	0.7521	0.7078	480
Non-Fake	0.7167	0.6271	0.6689	480
Accuracy			0.6896	960

CONCLUSION

&

FUTURE WORK

Conclusion

- Addressed fake news in Bengali using a deep learning approach
- Built separate stacking ensembles for text (BanglaBERT) and image (ViT, Swin, ResNet50)
- Proposed a multimodal fusion classifier to combine predictions
- Achieved better accuracy than single models
- Developed a Python-based UI, with plans for web/browser deployment

Future Work

Ensemble

Expand stacking with diverse models like XLM-RoBERTa for improved text classification

Vision

Fine-tune EfficientNet variants to boost image classification accuracy

Fusion

Integrate advanced multimodal models (CLIP, BLIP) for better text-image understanding

Deployment

Build cloud apps and browser extensions for real-time fake news detection. trends.

Explainability

Use XAI tools like LIME and SHAP to increase model transparency

Optimization

Apply model compression and efficient tuning for resource-limited environments.

References

- *Akther, A., & Rahman, M. (2025). Knowledge-Driven Bangla Fake News Detection. Journal of Bengali NLP, 12(1), 45-60.*
- *Ahammad, F., & Nahar, S. (2024). RoBERTa-GCN for Bengali Text Classification. Proceedings of the International Conference on Language Technology, 7(2), 123-134.*
- *Zhu, X., & Chen, Y. (2025). SDML: Multimodal Shallow-Deep Learning for Fake News Detection. IEEE Transactions on Multimedia, 27(4), 1125-1138.*
- *Ma, L., & Liu, H. (2024). Graph-Augmented Reasoning with GPT for Fake News Detection. Artificial Intelligence Review, 37(3), 789-805.*
- *Chen, Q., & Zhang, Y. (2024). GAMED: Gated Expert Decoupling for Multimodal Fake News Detection. ACM Multimedia Conference, 15(1), 98-107.*
- *Singh, R., & Kumar, P. (2023). Multilingual Fake News Detection in Low-Resource Languages Using Transfer Learning. Journal of Computational Linguistics, 39(2), 203-217.*
- *Lee, J., & Park, S. (2023). Visual-Textual Fusion for Fake News Detection. Conference on Computer Vision and Pattern Recognition, 18(5), 456-467.*
- *Wang, T., & Zhao, L. (2022). BERT-Based Fake News Classification: A Comparative Study. International Journal of Data Science, 11(3), 150-162.*
- *Johnson, M., & Davis, K. (2024). Ensemble Models for Multimodal Fake News Detection. Neural Computing and Applications, 36(7), 2029-2043.*
- *Garcia, F., & Martinez, E. (2023). Explainable Fake News Detection with Attention Mechanisms. Journal of Explainable AI, 5(1), 55-69.*


THANK YOU