Glossary

Accuracy: The proportion of correct predictions made by a model out of all predictions, typically calculated as the ratio of correctly classified instances to the total number of instances.

Adam Optimizer: An adaptive optimization algorithm used to update neural network weights during training by combining ideas from momentum and adaptive learning rates to improve convergence efficiency.

Adversarial Content: Inputs (e.g., images or text) that are intentionally designed or modified to mislead or deceive machine learning models, often to cause incorrect predictions.

Attention Mechanism: A neural network component that allows models to dynamically focus on the most relevant parts of the input data when making predictions, improving the modeling of long-range dependencies.

Artificial Intelligence (AI): A field of computer science focused on developing systems capable of performing tasks that typically require human intelligence, such as reasoning, learning, perception, and decision-making.

Automated Detection: The use of computational models or algorithms to identify patterns or classify data (e.g., detecting fake news) without human intervention.

Batch size: The number of training samples processed before the model's internal parameters are updated during training.

Benchmark dataset: A standardized dataset used to evaluate and compare the performance of different machine learning models.

BERT: A transformer-based language model (Bidirectional Encoder Representations from Transformers) that learns contextual relationships between words in text using bidirectional training.

Binary classification: A machine learning task where inputs are classified into one of two possible categories.

Class imbalance: A situation in which some classes in a dataset have significantly more samples than others, which can bias model training.

Classification performance: The effectiveness of a model in correctly assigning class labels, often measured using metrics such as accuracy, precision, recall, and F1 score.

Computer vision: A field of artificial intelligence that enables computers to interpret and analyze visual information from images or videos.

Convolutional Neural Network (CNN): A deep learning architecture designed for processing grid-like data such as images, using convolutional layers to automatically learn spatial features.

Cross-entropy loss: A loss function commonly used in classification tasks that measures the difference between predicted probabilities and true labels.

Cross-modal fusion: The process of combining information from different data modalities, such as text and images, to improve model performance.

Cross-modal inconsistencies: Conflicts or mismatches between different modalities (e.g., image and text) that may indicate misleading or fake content.

Cross-modal learning: A learning approach where information from one modality is used to improve understanding in another modality.

Cross-dataset evaluation: The evaluation of a model on a dataset different from the one it was trained on to assess generalization.

Dataset bias: Systematic biases present in a dataset that can lead a model to learn skewed or non-generalizable patterns.

DeBERTa: A transformer-based language model (Decoding-enhanced BERT with disentangled attention) that improves upon BERT by enhancing attention mechanisms and representation learning.

Deepfake: Synthetic media, typically images or videos, that have been manipulated using AI to replace or alter a person's appearance or actions.

Deepfake technologies: Techniques and algorithms used to create highly realistic synthetic media, often leveraging deep learning models such as GANs.

Domain shift: A change in the data distribution between training and testing environments that can reduce model performance.

EfficientNet: A family of convolutional neural networks that balance model depth, width, and resolution to achieve high accuracy with fewer parameters.

EfficientViT: A lightweight Vision Transformer variant designed to improve computational efficiency while maintaining strong performance.

Encoder: A component of a neural network that transforms input data into a compact feature representation.

Embedding: A numerical representation of data (such as words or images) in a continuous vector space that captures semantic meaning.

Fine-tuning: The process of further training a pretrained model on a specific task or dataset to adapt it to a new domain.

Fake news: False or misleading information presented as legitimate news, often with the intent to deceive.

Fakeddit dataset: A large-scale multimodal dataset of Reddit posts used for fake news detection research.

Feature representation: A transformed version of raw input data that captures important characteristics for machine learning tasks.

F1 score: A metric that combines precision and recall into a single score, calculated as their harmonic mean.

Fine-grained classification: A classification task that distinguishes between multiple closely related categories or subtle differences in data.

Generalization: The ability of a model to perform well on unseen data beyond the training set.

GPU (Graphics Processing Unit): Specialized hardware designed to accelerate parallel computations, commonly used in training deep learning models.

ImageNet: A large-scale image dataset commonly used for pretraining computer vision models.

InceptionResNetV2: A deep convolutional neural network architecture that combines Inception modules with residual connections for improved performance.

LIAR dataset: A dataset of short political statements labeled for truthfulness, commonly used in fake news detection research.

Learning rate: A hyperparameter that controls the step size at which a model's weights are updated during training.

LSTM (Long Short-Term Memory): A type of recurrent neural network designed to model sequential data and capture long-term dependencies.

Machine Learning (ML): A subset of artificial intelligence focused on enabling systems to learn patterns from data and make predictions or decisions.

Multimodal learning: A learning approach that integrates information from multiple data modalities, such as text and images.

Multimodal fusion: The process of combining features from different modalities into a unified representation for prediction.

Model architecture: The structural design of a machine learning model, including its layers and connections.

Model robustness: The ability of a model to maintain performance under variations, noise, or adversarial conditions.

Natural Language Processing (NLP): A field of artificial intelligence focused on enabling computers to understand, interpret, and generate human language.

Normalization: A preprocessing step that scales input data to a standard range or distribution to improve model training.

Overfitting: A modeling error that occurs when a model learns noise or details from the training data instead of general patterns.

Optimizer (Adam): An optimization algorithm that updates model parameters using adaptive learning rates and momentum-based techniques.

Precision: The proportion of true positive predictions among all positive predictions made by a model.

Pretrained model: A model that has been previously trained on a large dataset and can be adapted to new tasks through fine-tuning.

Pretraining: The initial training phase where a model learns general features from a large dataset before being adapted to a specific task.

Recall: The proportion of true positive predictions among all actual positive instances.

Regularization: Techniques used during training to prevent overfitting by constraining model complexity.

ResNet: A deep convolutional neural network architecture that uses residual connections to enable the training of very deep networks.

State-of-the-art (SOTA): The highest level of performance achieved by current models or methods on a specific task.

Swin Transformer: A hierarchical Vision Transformer that uses shifted windows to efficiently capture local and global image features.

Tokenization: The process of splitting text into smaller units such as words or sub words for processing by a model.

Transfer learning: A technique where knowledge gained from one task or dataset is applied to another related task.

Transformer (architecture): A neural network architecture based on attention mechanisms, widely used in NLP and vision tasks for modeling relationships in data.

Training / validation / test split: The division of a dataset into subsets for training a model, tuning hyperparameters, and evaluating performance.

Unimodal vs multimodal: The distinction between models that use a single data type (unimodal) and those that combine multiple data types (multimodal).

Vision Transformer (ViT): A transformer-based architecture that processes images as sequences of patches to learn visual representations.

Appendix

Table A1. Evaluation metrics of fine-tuned NLP models to classify fake news using unimodal approach (textual information only)

Model	**Train/Test/ Val**	**Label**	**Precision**	**Recall**	**F1**	**Samples**
Albert	Train	Fake	0.96	0.75	0.84	49,908
		Real	0.79	0.97	0.87	50,001
	Validation	Fake	0.95	0.71	0.81	34,396
		Real	0.68	0.94	0.79	23,161
	Test	Fake	0.95	0.71	0.81	34,162
		Real	0.68	0.94	0.79	23,333

BERTweet	Train	Fake	0.90	0.86	0.88	49,908
		Real	0.87	0.90	0.88	50,001
	Validation	Fake	0.92	0.85	0.88	34,396
		Real	0.79	0.88	0.83	23,161
	Test	Fake	0.92	0.85	0.88	34,162
		Real	0.79	0.88	0.84	23,333
DeBERTa	Train	Fake	0.97	0.97	0.97	49,908
		Real	0.97	0.97	0.97	50,001
	Validation	Fake	0.93	0.87	0.90	34,396
		Real	0.82	0.89	0.90	23,161
	Test	Fake	0.93	0.88	0.90	34,162
		Real	0.83	0.89	0.86	23,333
Universal Sentence Encoder	Train	Fake	0.83	0.78	0.80	49,908
		Real	0.79	0.84	0.82	50,001
	Validation	Fake	0.88	0.78	0.83	34,396
		Real	0.71	0.84	0.77	23,161

	Test	Fake	0.92	0.85	0.88	34,162
		Real	0.79	0.88	0.84	23,333
ELECTRA	Train	Fake	0.94	0.92	0.93	49,908
		Real	0.92	0.94	0.93	50,001
	Validation	Fake	0.93	0.86	0.89	34,396
		Real	0.81	0.89	0.85	23,161
	Test	Fake	0.93	0.86	0.89	34,162
		Real	0.81	0.89	0.85	23,333
FBERT	Train	Fake	0.97	0.98	0.98	49,908
		Real	0.98	0.97	0.98	50,001
	Validation	Fake	0.94	0.87	0.87	34,396
		Real	0.83	0.87	0.85	23,161
	Test	Fake	0.93	0.89	0.93	34,162
		Real	0.81	0.91	0.85	23,333
DistilBERT	Train	Fake	0.98	0.98	0.98	49,908
		Real	0.98	0.98	0.98	50,001

	Validation	Fake	0.92	0.84	0.88	34,396
		Real	0.79	0.89	0.83	23,161
	Test	Fake	0.92	0.85	0.88	34,162
		Real	0.79	0.88	0.84	23,333
RoBERTa	Train	Fake	0.92	0.95	0.94	49,908
		Real	0.95	0.92	0.93	50,001
	Validation	Fake	0.89	0.89	0.89	34,396
		Real	0.83	0.84	0.83	23,161
	Test	Fake	0.89	0.89	0.89	34,162
		Real	0.83	0.84	0.83	23,333

Table A2. Evaluation metrics of fine-tuned computer vision models to classify fake news images using unimodal approach (images only)

Model	**# of Params (M = Million)**	**Input Size**	**Train/ Test/Val**	**Label**	**Preci -sion**	**Rec -all**	**F1**	**Samples**

ConvNeXt-XL	350.1M	224x 224	Train	Fake	0.76	0.78	0.77	49,908
				Real	0.77	0.75	0.76	50,001
			Validation	Fake	0.79	0.76	0.78	34,396
				Real	0.67	0.70	0.68	23,161
			Test	Fake	0.79	0.77	0.78	34,162
				Real	0.67	0.70	0.69	23,333
DenseNet201	20.2M	224x 224	Train	Fake	0.51	0.48	0.49	49,908
				Real	0.51	0.53	0.52	50,001
			Validation	Fake	0.60	0.52	0.56	34,396
				Real	0.41	0.49	0.45	23,161
			Test	Fake	0.60	0.52	0.56	34,162
				Real	0.41	0.49	0.45	23,333
EfficientNet-B7	66.7M	600x 600	Train	Fake	0.51	0.45	0.48	49,908
				Real	0.41	0.49	0.45	50,001
			Validation	Fake	0.60	0.53	0.56	34,396
				Real	0.41	0.48	0.44	23,161

			Test	Fake	0.60	0.53	0.57	34,162
				Real	0.41	0.48	0.44	23,333
EfficientNetV2-B3	14.5M	300x300	Train	Fake	0.50	0.23	0.31	49,908
				Real	0.50	0.77	0.61	50,001
			Validation	Fake	0.59	0.25	0.35	34,396
				Real	0.41	0.75	0.53	23,161
			Test	Fake	0.60	0.25	0.35	34,162
				Real	0.40	0.76	0.35	23,333
Inception-ResNetV2	55.9M	299x299	Train	Fake	0.50	0.50	0.50	49,908
				Real	0.50	0.50	0.50	50,001
			Validation	Fake	0.59	0.59	0.59	34,396
				Real	0.40	0.41	0.40	23,161
			Test	Fake	0.59	0.53	0.56	34,162
				Real	0.41	0.47	0.44	23,333
InceptionV3	23.9M	299x299	Train	Fake	0.50	0.51	0.51	49,908
				Real	0.50	0.50	0.50	50,001

			Validation	Fake	0.60	0.55	0.58	34,396
				Real	0.41	0.46	0.43	23,161
			Test	Fake	0.48	0.43	0.52	34,162
				Real	0.52	0.47	0.51	23,333
MobileNetV2	3.5M	224x 224	Train	Fake	0.50	0.43	0.46	49,908
				Real	0.50	0.57	0.53	50,001
			Validation	Fake	0.60	0.54	0.57	34,396
				Real	0.40	0.46	0.43	23,161
			Test	Fake	0.59	0.54	0.57	34,162
				Real	0.41	0.46	0.43	23,333
ResNet50V2	25.6M	224x 224	Train	Fake	0.55	0.47	0.51	49,908
				Real	0.54	0.62	0.57	50,001
			Validation	Fake	0.65	0.48	0.55	34,396
				Real	0.44	0.61	0.51	23,161
			Test	Fake	0.64	0.47	0.54	34,162
				Real	0.44	0.61	0.51	23,333

ResNet101V2	44.7M	224x224	Train	Fake	0.57	0.50	0.53	49,908
				Real	0.55	0.62	0.59	50,001
			Validation	Fake	0.67	0.50	0.57	34,396
				Real	0.55	0.62	0.59	23,161
			Test	Fake	0.66	0.50	0.57	34,162
				Real	0.46	0.63	0.53	23,333
ResNet152V2	60.4M	224x224	Train	Fake	0.60	0.64	0.62	49,908
				Real	0.41	0.37	0.39	50,001
			Validation	Fake	0.60	0.64	0.62	34,396
				Real	0.41	0.37	0.39	23,161
			Test	Fake	0.51	0.62	0.56	34,162
				Real	0.51	0.40	0.45	23,333
VGG16	138.4M	224x224	Train	Fake	0.50	0.47	0.48	49,908
				Real	0.50	0.53	0.52	50,001
			Validation	Fake	0.60	0.49	0.54	34,396
				Real	0.40	0.51	0.45	23,161

			Test	Fake	0.60	0.50	0.54	34,162
				Real	0.41	0.51	0.45	23,333
Xception	22.9M	299x 299	Train	Fake	0.50	0.40	0.45	49,908
				Real	0.50	0.61	0.55	50,001
			Validation	Fake	0.60	0.56	0.58	34,396
				Real	0.40	0.42	0.42	23,161
			Test	Fake	0.60	0.56	0.58	34,162
				Real	0.41	0.45	0.43	23,333
ConvNeXt-Small	28.6M	224x 224	Train	Fake	0.64	0.52	0.57	49,908
				Real	0.58	0.43	0.46	50,001
			Validation	Fake	0.43	0.39	0.40	34,396
				Real	0.37	0.60	0.42	23,161
			Test	Fake	0.41	0.63	0.60	34,162
				Real	0.60	0.56	0.58	23,333
EfficientNet-V2L	119.0M	480x 480	Train	Fake	0.50	0.59	0.61	49,908
				Real	0.51	0.43	0.32	50,001

			Validation	Fake	0.52	0.49	0.42	34,396
				Real	0.48	0.47	0.60	23,161
			Test	Fake	0.41	0.56	0.38	34,162
				Real	0.54	0.63	0.64	23,333
ViT	-	-	Train	Fake	1.00	1.00	1.00	49,908
				Real	1.00	1.00	1.00	50,001
			Validation	Fake	0.73	0.85	0.79	34,396
				Real	0.71	0.54	0.61	23,161
			Test	Fake	0.73	0.85	0.79	34,162
				Real	0.71	0.54	0.62	23,333
DeiT	-	-	Train	Fake	1.00	1.00	1.00	49,908
				Real	1.00	1.00	1.00	50,001
			Validation	Fake	0.74	0.87	0.80	34,396
				Real	0.74	0.54	0.62	23,161
			Test	Fake	0.74	0.87	0.80	34,162
				Real	0.75	0.55	0.63	23,333

EfficientViT	-	-	Train	Fake	0.99	0.96	0.97	49,908
				Real	0.96	0.99	0.98	50,001
			Validation	Fake	0.84	0.80	0.82	34,396
				Real	0.72	0.77	0.75	23,161
			Test	Fake	0.84	0.80	0.82	34,162
				Real	0.73	0.78	0.75	23,333
Swin Transformer	-	-	Train	Fake	0.60	0.55	0.58	49,908
				Real	0.59	0.63	0.61	50,001
			Validation	Fake	0.69	0.54	0.60	34,396
				Real	0.48	0.64	0.55	23,161
			Test	Fake	0.70	0.54	0.61	34,162
				Real	0.49	0.65	0.56	23,333

Table A3. Evaluation metrics of fine-tuned multimodal fusion models combining the NLP and CV approaches (images and text)

Model	Train/Test/Val	Label	Precision	Recall	F1	Samples
DeBERTa & EfficientViT	Train	Fake	0.91	0.93	0.92	49907
		Real	0.93	0.91	0.92	50000
	Validation	Fake	0.86	0.93	0.90	36022
		Real	0.88	0.77	0.82	23320
	Test	Fake	0.86	0.94	0.90	35812
		Real	0.89	0.77	0.82	23507
DenseNet201 & Albert	Train	Fake	0.96	0.90	0.93	49907
		Real	0.91	0.96	0.93	50000
	Validation	Fake	0.88	0.89	0.89	36022
		Real	0.83	0.81	0.82	23320
	Test	Fake	0.88	0.89	0.89	35812
		Real	0.83	0.81	0.82	23507
DistilBERT & DeiT	Train	Fake	0.95	0.93	0.94	49907
		Real	0.93	0.95	0.94	50000
	Validation	Fake	0.84	0.92	0.88	36022

		Real	0.85	0.72	0.78	23320
	Test	Fake	0.84	0.92	0.88	35812
		Real	0.86	0.73	0.79	23507
EfficientNetV2B3 & RoBERTa	Train	Fake	0.99	0.96	0.97	49907
		Real	0.96	0.99	0.97	50000
	Validation	Fake	0.90	0.90	0.90	36022
		Real	0.85	0.85	0.85	23320
	Test	Fake	0.90	0.91	0.91	35812
		Real	0.86	0.85	0.86	23507
EfficientNetB7 & distilbert	Train	Fake	1.00	1.00	1.00	49907
		Real	1.00	1.00	1.00	50000
	Validation	Fake	0.91	0.90	0.91	36022
		Real	0.85	0.87	0.86	23320
	Test	Fake	0.91	0.91	0.91	35812
		Real	0.86	0.87	0.86	23507
ELECTRA &	Train	Fake	0.99	0.96	0.97	49907

MobileNetV2		Real	0.96	0.99	0.97	50000
	Validation	Fake	0.93	0.88	0.91	36022
		Real	0.83	0.90	0.87	23320
	Test	Fake	0.93	0.88	0.91	35812
		Real	0.84	0.91	0.87	23507
InceptionResNet V2 & DeBERTa	Train	Fake	0.98	0.94	0.96	49907
		Real	0.95	0.98	0.96	50000
	Validation	Fake	0.92	0.90	0.91	36022
		Real	0.85	0.87	0.86	23320
	Test	Fake	0.91	0.90	0.91	35812
		Real	0.85	0.87	0.86	23507
Swin Transformer & RoBERTa	Train	Fake	0.93	0.93	0.93	49907
		Real	0.93	0.93	0.93	50000
	Validation	Fake	0.76	0.96	0.85	36022
		Real	0.89	0.53	0.66	23320
	Test	Fake	0.75	0.96	0.85	35812

		Real	0.90	0.52	0.66	23507
VGG16 & BERTweet	Train	Fake	0.99	0.97	0.98	49907
		Real	0.97	0.99	0.98	50000
	Validation	Fake	0.90	0.90	0.90	36022
		Real	0.85	0.84	0.84	23320
	Test	Fake	0.90	0.90	0.90	35812
		Real	0.85	0.85	0.85	23507