**Audio Deepfake Detection: Research, Implementation, and Analysis**

**Part 1: Research & Selection**

**Identified Approaches for Forgery Detection**

**Approach 1:**

**Support Vector Machine (SVM) Based ML Approach with MFCC Features**

**Key Technical Innovation**:

 Uses classical audio feature extraction techniques (e.g., MFCCs, CQCCs) to capture relevant spectral and temporal cues.

 Applies SVM with kernel methods to establish a discriminative boundary between genuine and spoofed audio.

**Reported Performance Metrics**:

 Studies have reported accuracies above 90% on ASVspoof 2019 subsets with well-engineered features.

 Relative improvements in EER have been noted when compared to some baseline systems.

**Why Promising for Your Needs:**

* **Efficiency & Speed:** Low computational overhead makes it ideal for real-time or near real-time applications.
* **Interpretability:** The SVM’s decision process is more transparent, aiding troubleshooting and understanding feature impact.
* **Ease of Deployment:** Suitable when computational resources are limited.

**Potential Limitations/Challenges**:

 Highly dependent on the quality and robustness of the extracted features.

 May not generalize as well to unseen or highly variable acoustic environments without additional adaptation.

**Approach 2:**

**AASIST-L (Lightweight Graph Attention-Based Model)**

**Key Technical Innovation**:

* End-to-end raw waveform processing with a graph attention mechanism that models both temporal and spectral dependencies.
* Lightweight design with only ~85K parameters.

**Reported Performance Metrics**:

* EER of approximately 0.99% on ASVspoof 2019 Logical Access (LA) task.
* Low min t-DCF (around 0.0309) as reported in the original AASIST papers.

**Why Promising for Your Needs:**

* **Real-Time Capability:** The compact size enables near real-time inference.
* **Robust Analysis:** Direct processing of raw audio captures fine-grained spoofing artefacts in natural conversations.

 **Potential Limitations/Challenges**:

* May need fine-tuning or hardware optimization for deployment on resource-constrained devices.
* Performance could be sensitive to varied acoustic environments not seen during training.

**Approach 3:**

**RawNet2-Based Residual Architecture with Squeeze-Excitation (SE) Blocks**

**Key Technical Innovation**:

* Processes raw audio using SincConv filters followed by deep residual blocks.
* Integrates SE blocks to adaptively recalibrate channel importance, enhancing discriminative power.

**Reported Performance Metrics**:

 Competitive EER scores in the range of ~1.2%–1.64% on ASVspoof 2019 LA evaluation.

 Demonstrates robust performance in logical access spoofing tasks.

**Why Promising for Your Needs:**

 **End-to-End Learning:** Minimizes preprocessing by working directly on raw audio.

 **Enhanced Feature Learning:** Residual and SE blocks help capture both local and global patterns, improving detection of AI-generated speech in real conversations.

 **Adaptability:** Strong performance in controlled tests suggests potential for generalization with proper augmentation.

 **Potential Limitations/Challenges**:

 May require additional data augmentation to handle noise and diverse real-world conditions.

 Hyperparameter tuning is critical to maintain performance across varied recording environments.

**Part 2: Implementation**

**1. Selected Model: SVM with MFCC Features**

* **Reason for Selection**: Given computational constraints, SVM with MFCC is a lightweight and effective approach for real-time detection.
* **Comparison to Other Models**:
  + More efficient than Wav2Vec2 but less accurate.
  + Less powerful than CNN-based approaches but requires less data and preprocessing.

**2. Dataset: ASVspoof 2019 (Reduced Version)**

* **Dataset Selection**: The ASVspoof 2019 dataset is widely used for benchmarking deepfake detection models.
* **Preprocessing**:
  + Extract MFCC features from audio files.
  + Normalize and structure data for training/testing.

**3. Model Implementation**

* **Feature Extraction**: MFCC features extracted using Librosa.
* **Model Training**:
  + SVM trained with an RBF kernel.
  + Hyperparameter tuning using grid search.
* **Evaluation Metrics**: Accuracy, Precision, Recall, F1-score.

**Part 3: Documentation & Analysis**

**1. Implementation Process**

**Challenges Encountered:**

 **Feature Extraction Issues**

* Requires **fixed-length input**, but audio is variable-length.
* **Choosing the right features** (MFCC, CQCC, spectrograms) is tricky.
* **Loss of temporal information** affects accuracy.
* ✅ Use **PCA, hybrid features (MFCC + CQCC), or feature selection**.

 **Computational Complexity**

* SVM is **slow on large datasets** (O(n²)-O(n³) complexity).
* **Kernel methods** add overhead, making real-time detection hard.
* ✅ Use **linear SVM, mini-batch training, or dimensionality reduction**.

 **Class Imbalance & Generalization**

* **More real speech than deepfakes** → biased predictions.
* **Fails on unseen attacks** (overfits to known types).
* ✅ Use **balanced class weights, data augmentation, and ensemble methods**.

**Solutions:**

* Reduced dataset size for faster processing.
* Optimized feature extraction pipeline.

**2. Model Analysis**

**Why This Model?**

* Lightweight and efficient for real-time detection.
* Can be deployed on low-resource devices.

**How It Works:**

* Extracts MFCC features and applies an SVM classifier.

**Performance Results:**

* Accuracy: ~80%
* Strengths: Fast, efficient, interpretable.
* Weaknesses: Lower accuracy compared to deep learning models.

**3. Reflection**

**Key Challenges:**

* Dataset preprocessing and feature extraction.
* Optimizing model parameters for better accuracy.

**Real-World Considerations:**

* Needs additional training on real-world deepfake samples.
* Sensitivity to background noise and speech variations.

**Improvements:**

* Use a hybrid approach combining SVM with CNN-based embeddings.
* Data augmentation techniques to enhance robustness.

**Deployment Strategy:**

* Deploy as a lightweight classifier in a real-time detection pipeline.
* Combine with a deep learning-based backend for improved accuracy.

This structured report provides a clear roadmap for research, implementation, and analysis of audio deepfake detection approaches.