Analysis of Models for Audio Deepfake Detection

Kartik Dua

05/04/2025

1. Support Vector Machine (SVM) with Handcrafted Features

- **Alignment with Project Requirements:**
  - SVM is a linear classifier effective for binary classification tasks, such as distinguishing between authentic and fake audio samples.
  - Utilizes manually extracted features like Mel-Frequency Cepstral Coefficients (MFCCs).
  - Suitable for projects with limited computational resources and smaller datasets.
- **Computational Efficiency:**
  - Moderate computational requirements; more complex than Logistic Regression but still manageable on standard hardware.
  - Training is relatively straightforward and requires less time compared to deep learning models.
- **Generalization Capabilities:**
  - Performance heavily depends on the quality of handcrafted features.
  - May struggle to capture complex patterns inherent in audio deepfakes, leading to limited generalization.
- **Interpretability:**
  - Offers clear insights into the decision-making process by highlighting the importance of each feature.
  - Easier to interpret compared to deep learning models.
- **Practical Considerations:**
  - Manual feature extraction is labor-intensive and requires domain expertise.

– Risk of underfitting due to the model's simplicity.

Relevant Research Papers:

- Shaaban, O., Yildirim, R., & Alguttar, A. (2023). *Audio Deepfake Approaches*. Available at: https://www.researchgate.net/publication/356015648_A_Deep_Learning_Framework_for_Audio_Deepfake_Detection
- Hamza, A., Javed, A. R., Iqbal, F., & Borghol, R. (2022). *Deepfake Audio Detection via MFCC Features Using Machine Learning*. Available at: https://www.researchgate.net/publication/366489016_Deepfake_Audio_Detection_via_MFCC_features_using_Machine_Learning

2. Siamese Convolutional Neural Network (SCNN)

- **Alignment with Project Requirements:**
  - Designed to determine similarity between pairs of inputs, making it adept at distinguishing between authentic and fake audio samples.
  - Balances performance and computational efficiency, suitable for moderate-sized datasets.
- **Computational Efficiency:**
  - Moderate complexity allows effective training within resource constraints.
  - Requires more resources than SVM but remains feasible.
- **Generalization Capabilities:**
  - Effectively captures complex patterns and generalizes well to unseen data.
  - Demonstrated efficacy in audio deepfake detection tasks.
- **Interpretability:**
  - Provides similarity scores, though understanding exact contributing features is more challenging.
  - Trade-off between improved performance and interpretability is acceptable.
- **Practical Considerations:**
  - Leverages existing research and methodologies effectively.
  - Implementation allows building upon proven techniques and adapting them to specific contexts.

Relevant Research Papers:

- Nekadi, R. (2020). *Siamese Network-Based Multi-Modal Deepfake Detection*. University of Missouri-Kansas City. Available at: https://mospace.umsystem.edu/xmlui/handle/10355/74345

3. Deep Residual Network (ResNet)

- **Alignment with Project Requirements:**
  - ResNet is a deep learning model capable of capturing intricate patterns in data, making it suitable for complex tasks like audio deepfake detection.
  - Well-suited for projects with access to large datasets and substantial computational resources.
- **Computational Efficiency:**
  - High computational requirements due to deep architecture.
  - Training requires significant time and powerful hardware, such as GPUs or TPUs.
- **Generalization Capabilities:**
  - High capacity to learn complex representations, reducing the risk of underfitting.
  - However, prone to overfitting if not properly regularized, especially with limited data.
- **Interpretability:**
  - Low interpretability; difficult to discern how specific features contribute to predictions.
  - Often considered a "black box" model.
- **Practical Considerations:**
  - Implementation complexity is high, requiring expertise in deep learning frameworks.
  - Risk of overfitting necessitates careful tuning and validation.

Relevant Research Papers:

- Chen, T., Zhang, Z., Wang, Z., & Li, J. (2017). *ResNet for Audio Deepfake Detection*. Available at: https://arxiv.org/abs/1705.07663