Multimodal Detection of Alzheimer's Disease

Theodora Pavlidou

Marilena Papasideri

Δ.Π.Μ.Σ. ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ

ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

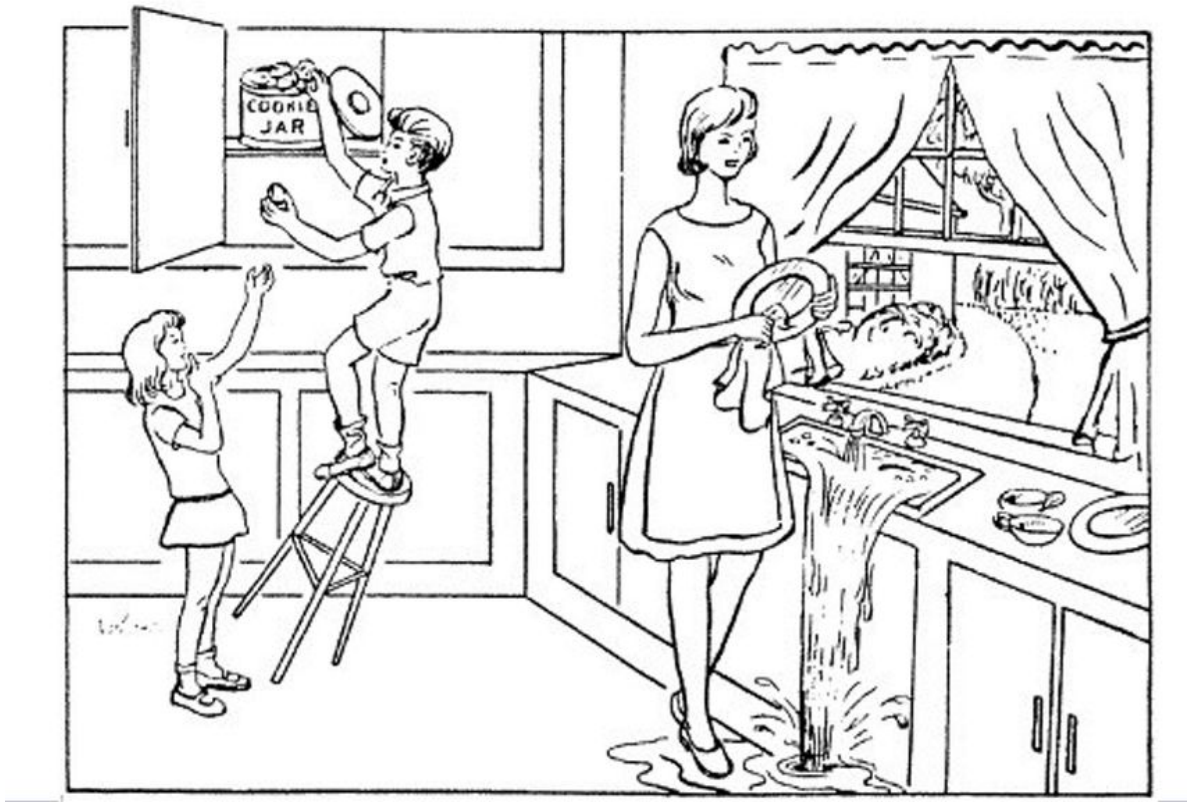
UNIVERSITY OF PIRAEUS

DEMOKRITOS

NATIONAL CENTRE FOR SCIENTIFIC RESEARCH

Project Scope & Dataset Integrity

- **Objective:** Multimodal framework for Alzheimer's Dementia (AD) detection using acoustic and linguistic markers.
- **Dataset:** ADRESS (DementiaBank Pitt Corpus).
- **Constraint Management:** N=108 (54 AD / 54 HC).
- **Task:** Standardized Boston Cookie Theft picture description.



Preprocessing & Feature Selection

Preprocessing:

- **Investigator Removal**: Termination of audio runs upon investigator (*INV) detection.
  - **Diarization**: pyannote.audio for automated speaker turn detection.
  - **Transcript-Based Segmentation**: Manual CHAT file parsing for high-fidelity patient speech isolation (*PAR).
- **Standardization**: 5-second fixed-length chunking for input consistency.

Feature Processing:

- **Correlation Filtering**: Removal of highly correlated features ($\rho > 0.95$) to reduce redundancy and multicollinearity.

Audio Feature Extraction

- **Extraction:** pyAudioAnalysis library with hierarchical temporal analysis
  - 50ms short-term windows
  - 2-second mid-term statistics (mean/std)

- **Feature Types:**
  - MFCCs (vocal tract characteristics)
  - Spectral descriptors (energy, centroid, rolloff, flux)
  - Delta coefficients (temporal dynamics)

- **Optimization**: Chroma features (52-D) excluded after ablation study (no performance gain, added noise)
- **Final Dimensionality:** 84-D optimized acoustic vector per segment

Text Feature Extraction

- Linguistic features extracted from .cha transcription files capture cognitive and speech patterns.

- Text Features:
  - **Filler Ratio**: Frequency of filler words (e.g., "um", "uh")
  - **Pause Ratio**: Frequency of non-lexical pauses (e.g., (.), (..))
  - **Repetition Ratio**: Frequency of word/phrase repetitions ([/])
  - **Error Ratio**: Frequency of phonological or lexical errors ([*])
  - **Correction Ratio**: Frequency of corrections ([: replacement])
  - **Self-Correction Ratio**: Proportion of utterances containing self-corrections([//])
  - **Words Per Minute**: Speech rate based on timestamp analysis

Modeling & Cost-Sensitive Learning

Classifiers:

- SVM (Linear/RBF)
- Random Forest (100 trees)
- XGBoost

Regularization:

- Low C=0.1 (SVM) and max_depth=5 (RF) to handle high-dimensional features with small sample size

Recall Prioritization:

- **Cost-Sensitive Learning:** Class weights 1:1.5 (SVM/RF), scale_pos_weight=2.5 (XGBoost)
- **Purpose:** Emphasize Recall over Precision, minimizing False Negatives is critical for early dementia screening
- **Trade-off:** Accept higher false positives to avoid missing positive cases

Validation & Fusion Strategy

- **Validation Integrity:**
  - **Leave-One-Out (LOOCV):** Used for raw recording pipelines.
  - **Leave-One-Group-Out (LOGO):** Critical for segmented data to prevent **Data Leakage** (ensuring all segments from one patient are kept within the same fold).

- **Fusion Logic:**
  - **Early Fusion:** Feature-level concatenation.
  - **Late Fusion:** Decision-level Soft Voting based on class probabilities from unimodal baselines.

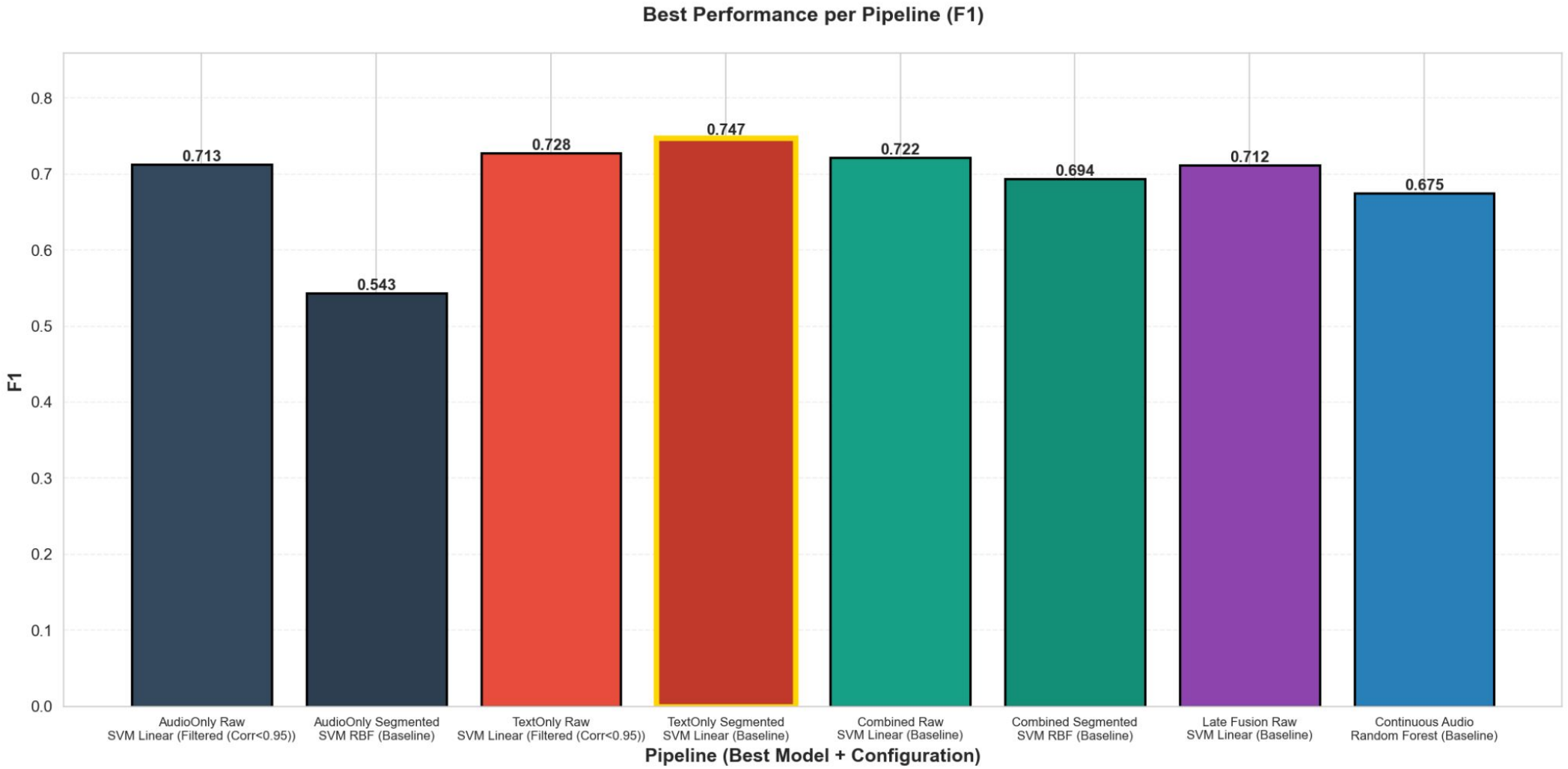
Pipelines

- Pipeline Diversity:
  - Developed 8 distinct pipelines to isolate the impact of segmentation, modality, and fusion.

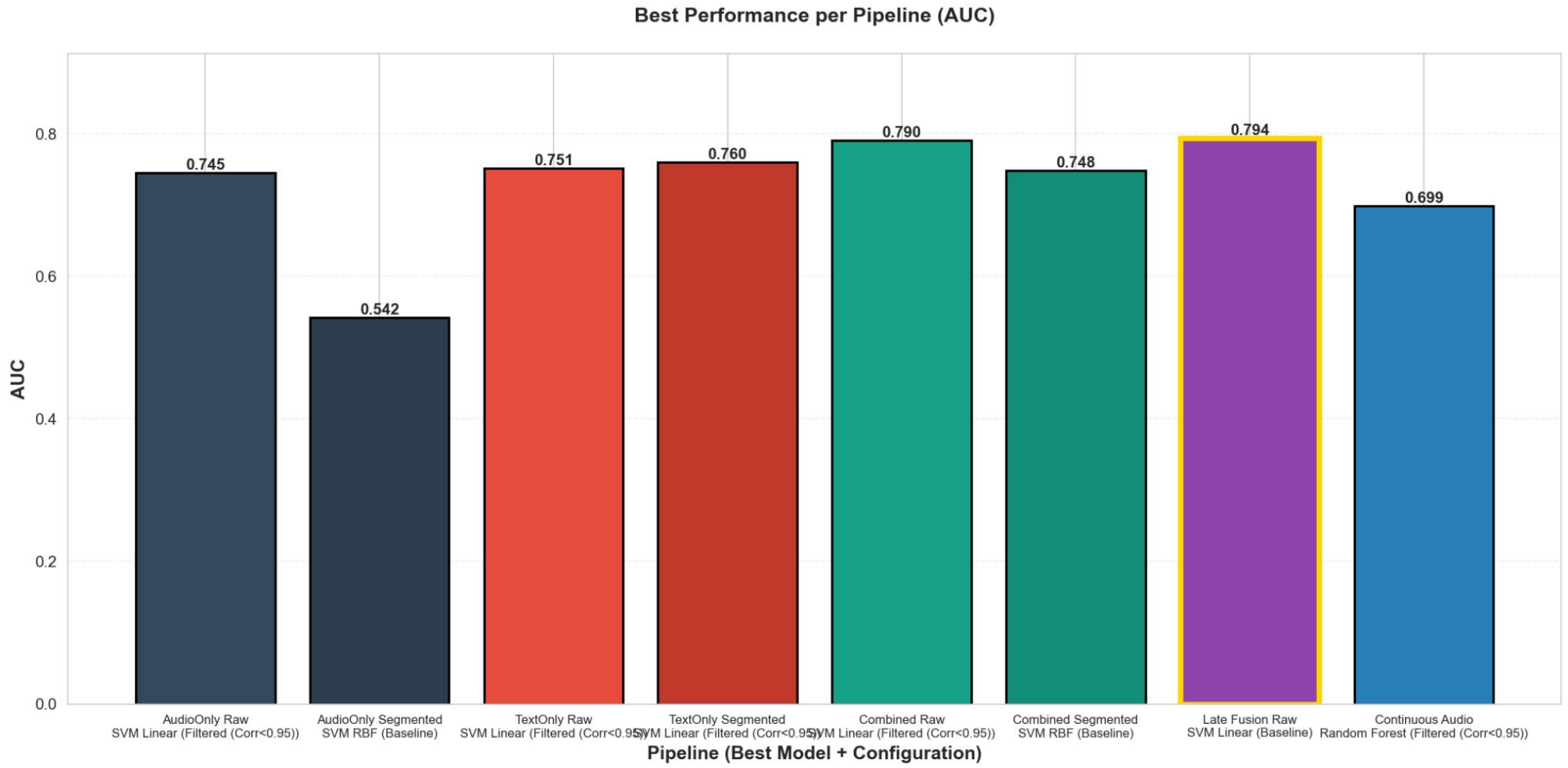
- Unimodal Pipelines:
  - **Audio-Only (Raw)**: Full-length audio with LOOCV validation
  - **Audio-Only (Segmented)**: 5-second segments with LOGO validation
  - **Text-Only (Raw)**: Linguistic features from full transcripts with LOOCV
  - **Text-Only (Segmented)**: Text features with LOGO validation

- Multimodal Fusion:
  - **Early Fusion (Raw)**: Feature concatenation at input level, LOOCV
  - **Early Fusion (Segmented)**: Feature concatenation, 5-second segments, LOGO
  - **Late Fusion (Raw)**: Decision-level weighted averaging, LOOCV

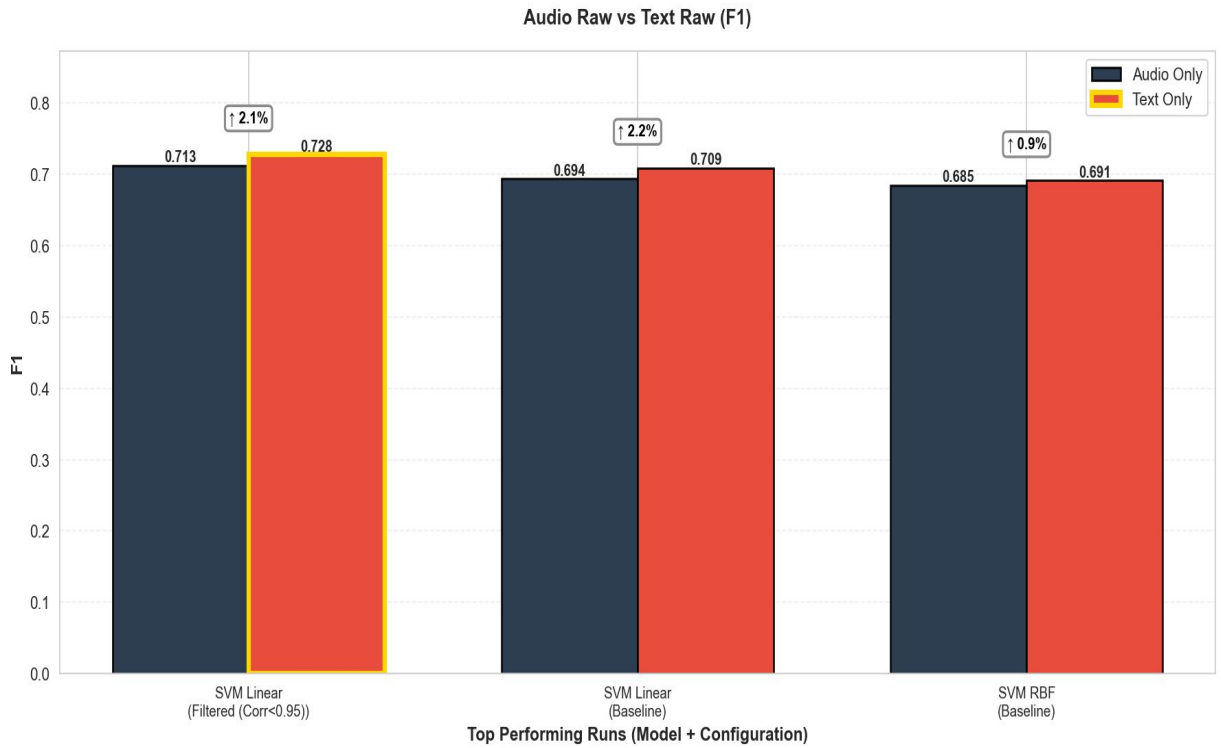
Best Performance Summary

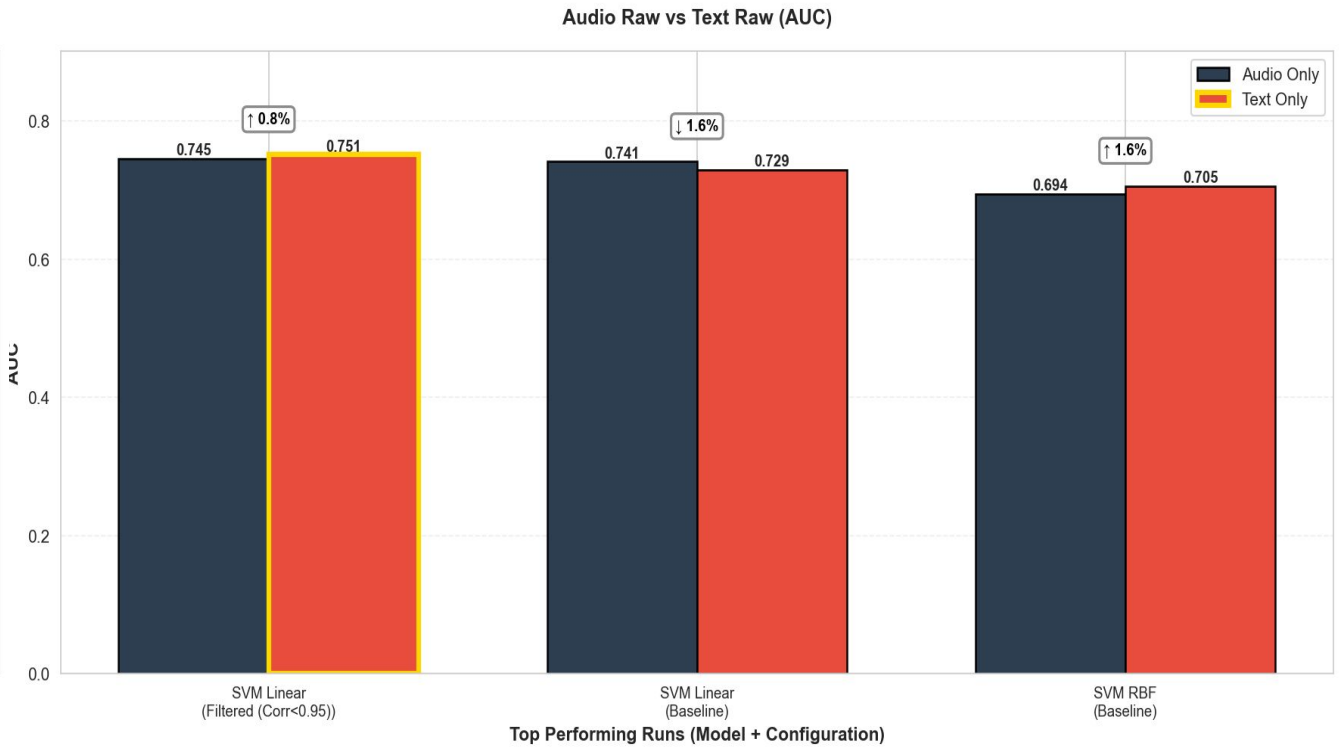


Best Performance Summary

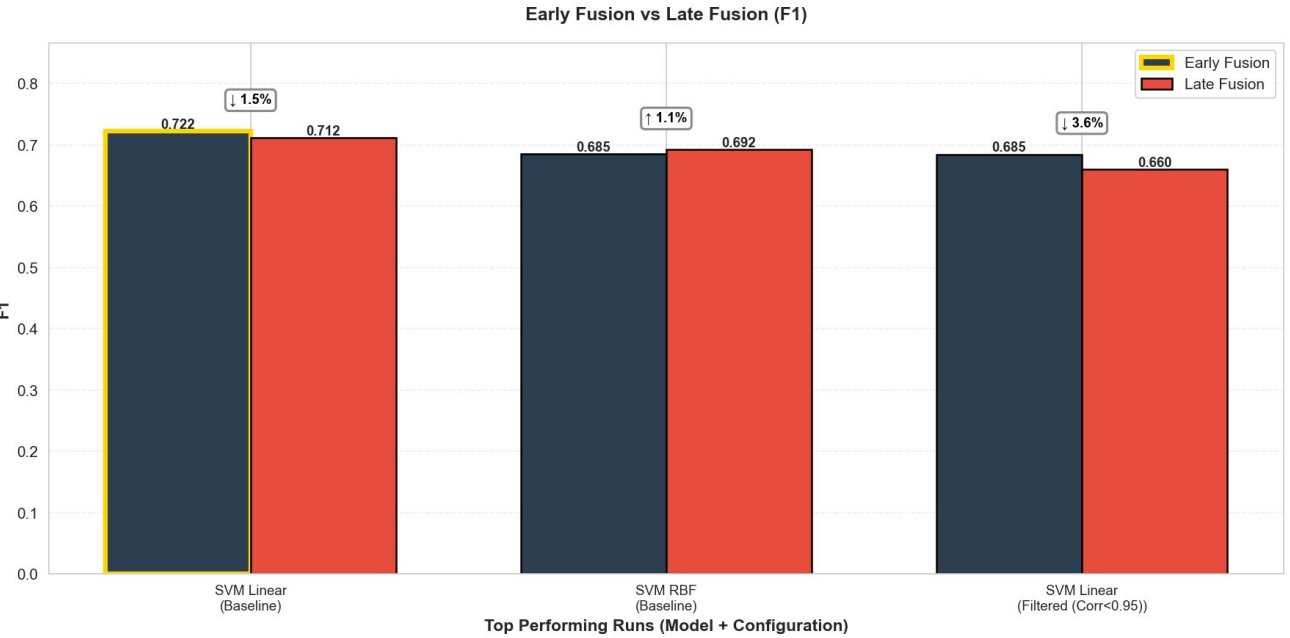


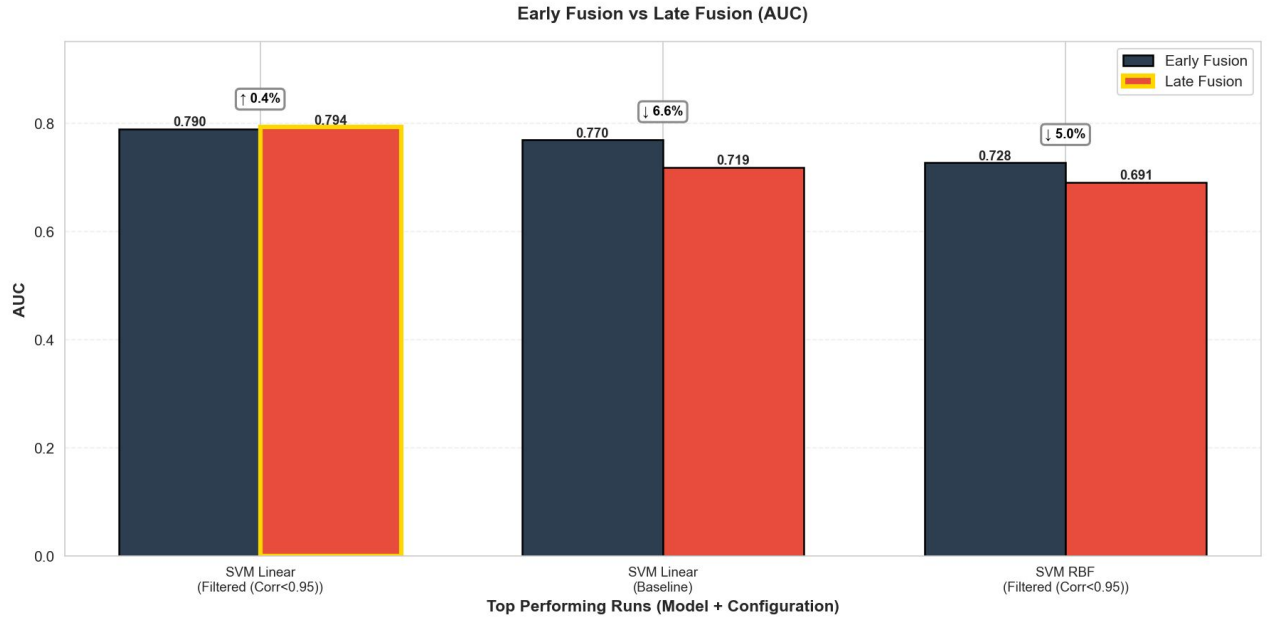
Modality Comparison



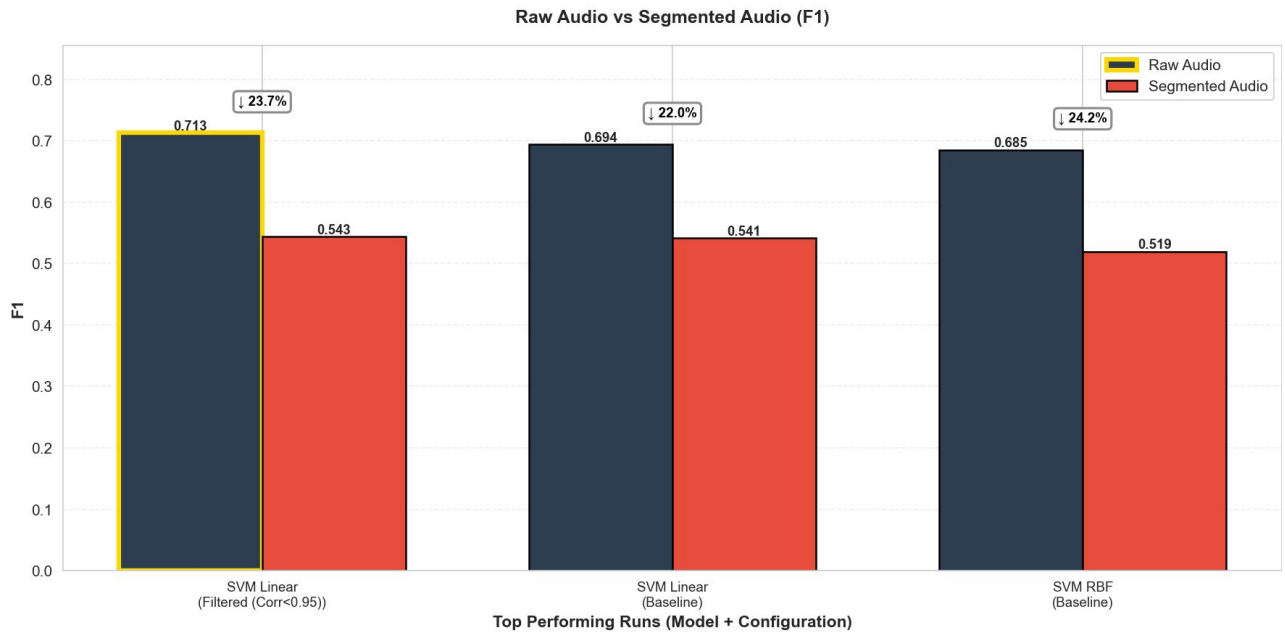


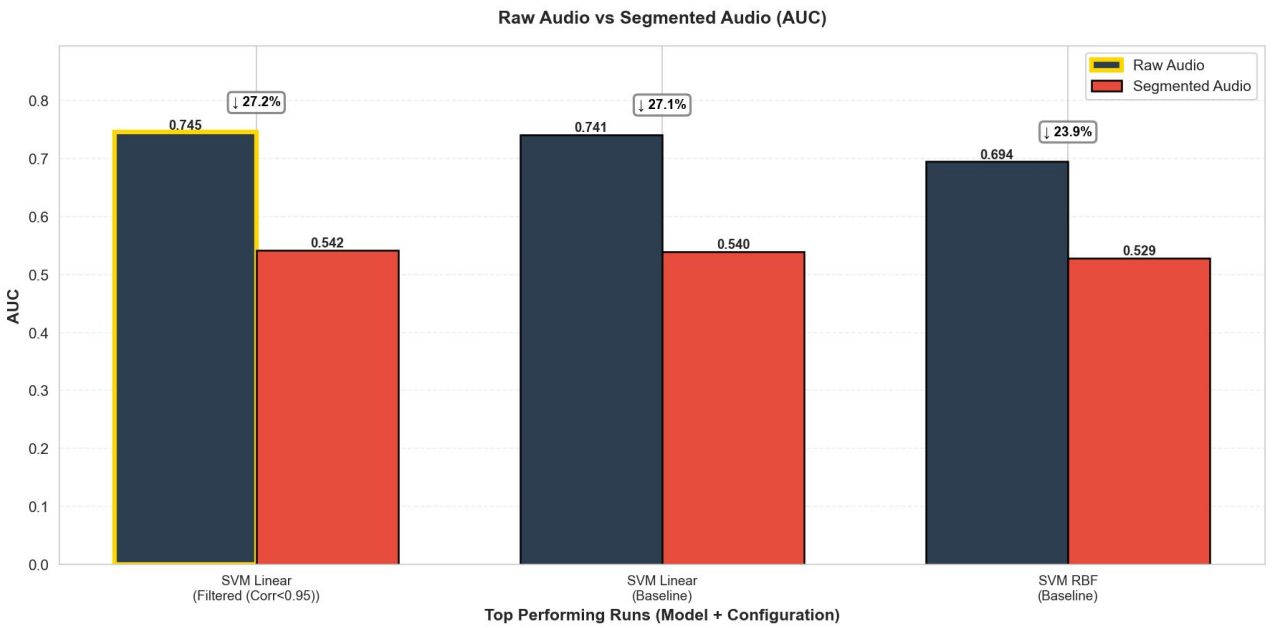
Fusion Strategy Comparison



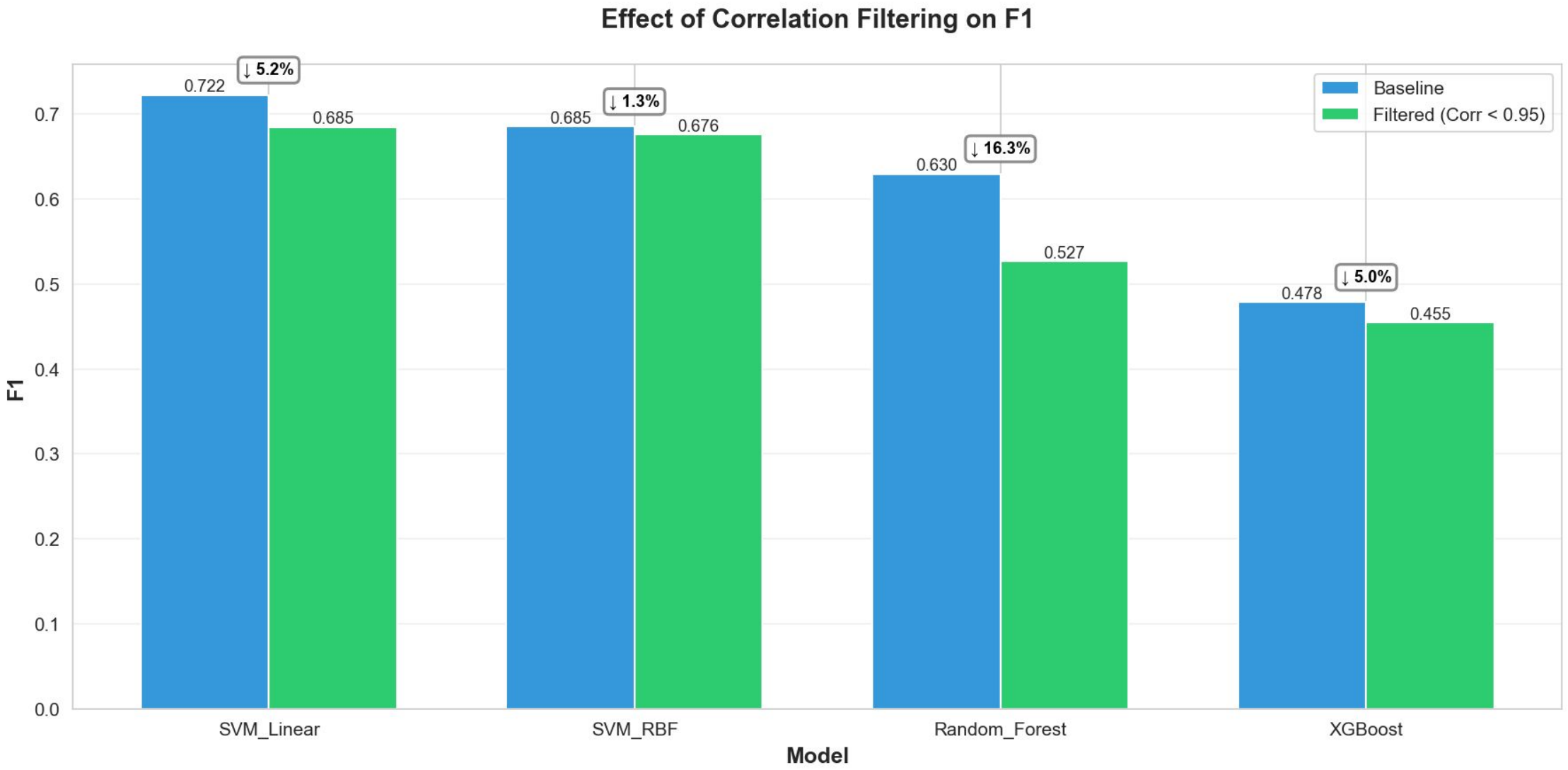


Impact of audio segmentation





Feature Correlation Filtering



Experimental Results and Conclusion

- **Late Fusion Superior:** Best AUC (0.794), 3.5% improvement over Early Fusion
- **Text Outperforms Audio:** 2-3% higher performance—linguistic markers more discriminative
- **Raw Audio Critical:** 21-28% better than segmented (avoids data leakage)
- **Validation Matters:** Subject-independent (LOOCV/LOGO) essential for realistic estimates
- **Optimal Setup:** Late Fusion + SVM Linear + Raw Audio → F1=0.712, AUC=0.794

Ευχαριστούμε για την προσοχή σας!
Ερωτήσεις;