

# Sistem cu Inteligență Artificială pentru Detectia Automată a Vorbirii Înșelătoare

Proiect IAECV 2024-2025

## **Studenti:**

Baron Lucian

Diaconu Adrian-Gabriel

Universitatea Națională de Știință și Tehnologie Politehnica București  
Facultatea de Electronică, Telecomunicații și Tehnologia Informației

Mai 2025

# Cuprins

1 Introducere

2 Metodologie

3 Rezultate Experimentale și Evaluare

4 Concluzii

# Introducere

# Obiectivul Proiectului

## Obiectiv General

Dezvoltarea unui sistem de învățare automată pentru detecția automată a minciunilor din vorbire, mai precis, diferențierea între vorbirea înșelătoare și cea sinceră.

# Obiectivul Proiectului

## Obiectiv General

Dezvoltarea unui sistem de învățare automată pentru detecția automată a minciunilor din vorbire, mai precis, diferențierea între vorbirea înșelătoare și cea sinceră.

## Componentele Sistemului

- Bloc de extragere de trăsături.
- Clasificator (model cu învățare automată).

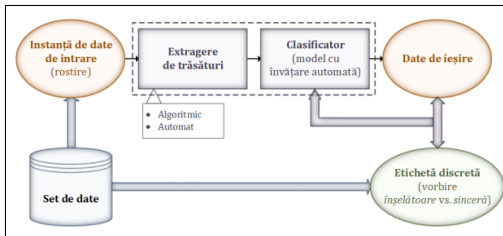


Figura 1: Schema bloc a sistemului de detecție.

# Descrierea Setului de Date

## Conținut

- 121 înregistrări audio în limba engleză (procese judiciare SUA).
- 61 înregistrări etichetate ca înșelătoare, 60 ca sincere.
- Durata totală: 56 min; Durata medie/înregistrare: 28 s.
- 56 subiecți distincți (22F, 34M).

## Format și Preprocesare

- Audio: format WAV (necomprimat), în subdirectorul "extrAudio".
- Preprocesare: 16 kHz, normate în amplitudine, PCM 16 biți.
- Denumiri: trial\_lie\_xxx.wav, trial\_truth\_xxx.wav.

## Adnotări

- Conțin: start, sfârșit, identitate și gen vorbitor per rostire.
- Total rostiri: 929 (463 înșelătoare, 466 sincere) - clase echilibrate.

## Metodologie

# Extragerea Trăsăturilor

## Extragere Algoritmică

- Rostirile împărțite în cadre de 25ms cu 10ms suprapunere (Hamming).
- Trăsături la nivel de cadru: Frecvența fundamentală ( $F_0$ ), primii 13 coeficienți Mel-cepstrali (MFCC), etc..
- Funcții statistice (medie, dev. std.) pentru valori la nivel de rostire.
- Normalizare z-score per vorbitor distinct.



# Extragerea Trăsăturilor

## Extragere Algoritmică

- Rostirile împărțite în cadre de 25ms cu 10ms suprapunere (Hamming).
- Trăsături la nivel de cadru: Frecvența fundamentală ( $F_0$ ), primii 13 coeficienți Mel-cepstrali (MFCC), etc..
- Funcții statistice (medie, dev. std.) pentru valori la nivel de rostire.
- Normalizare z-score per vorbitor distinct.

## Extragere Automată (pentru CNN)

- Extragere de spectrograme din rostiri.
- DFT 512 puncte, păstrare [0, 8] kHz (primele 257 puncte), scală liniară frecvență, logaritmare amplitudini (dB).
- Completare cu zerouri (padding) pentru dimensiune finală uniformă a spectrogramelor.

# Modele de Învățare Automată

Patru tipuri de modele de învățare automată sunt dezvoltate și evaluate:

- **Mașini cu Vectori Suport (SVM)**

- Intrări: Trăsături extrase.
- Hiperparametri:  $C$ , funcția kernel, etc. (10 configurații).

- **Ansambluri de Arbori de Decizie (Random Forest - RF)**

- Intrări: Trăsături extrase.
- Hiperparametri: Număr arbori, adâncime maximă, funcție criteriu, etc. (10 configurații).

- **Rețele Neuronale Complet Conectate (FCNN)**

- Intrări: Trăsături extrase.
- Hiperparametri: Număr straturi ascunse, număr neuroni/strat, funcții activare, etc. (100 configurații).

- **Rețele Neuronale Convoluționale (CNN)**

- Intrări: Spectrograme (extragere automată de trăsături).
- Hiperparametri: Număr straturi convoluționale, număr filtre, dimensiune filtre, etc. (20 configurații).

# Metodologie Experimentală și Validare

## Validare Încrucișată (Cross-Validation)

- 5 fold-uri (5-fold cross-validation).
- 80% date pentru antrenare, 20% pentru validare.
- Asigurare distribuție proporțională a claselor (înșelătoare/sincere).
- Asigurare independență relativ la vorbitori (speaker independence).
- Împărțire proporțională a vorbitorilor după gen.

# Metodologie Experimentală și Validare

## Validare Încrucișată (Cross-Validation)

- 5 fold-uri (5-fold cross-validation).
- 80% date pentru antrenare, 20% pentru validare.
- Asigurare distribuție proporțională a claselor (înșelătoare/sincere).
- Asigurare independență relativ la vorbitori (speaker independence).
- Împărțire proporțională a vorbitorilor după gen.

## Evaluarea Modelelor

- Măsură de performanță principală: Acuratețea (Accuracy).
- Pentru FCNN/CNN: ultimul strat cu un neuron și funcție sigmoidă, funcție de cost entropie încrucișată binară, optimizator Adam.
- Tehnici avansate: early stopping, dropout, batch normalization, regularizare L1/L2.

## Rezultate Experimentale și Evaluare

# Procesul de Evaluare și Centralizarea Rezultatelor

## Generarea și Colectarea Datelor de Performanță

- Toate experimentele de antrenare și validare (pe cele 5 fold-uri) pentru fiecare configurație de model (SVM, RF, FCNN, CNN), precum și rezultatele individuale (acuratețe per fold, parametri) sunt stocate în fișierul `training_results.csv`.
- Scriptul `evaluate_models.py` procesează acest fișier pentru a calcula performanțele agregate.

# Procesul de Evaluare și Centralizarea Rezultatelor

## Generarea și Colectarea Datelor de Performanță

- Toate experimentele de antrenare și validare (pe cele 5 fold-uri) pentru fiecare configurație de model (SVM, RF, FCNN, CNN), precum și rezultatele individuale (acuratețe per fold, parametri) sunt stocate în fișierul `training_results.csv`.
- Scriptul `evaluate_models.py` procesează acest fișier pentru a calcula performanțele agregate.

## Metrica Principală

- Performanța modelelor este evaluată folosind **acuratețea medie** obținută prin validare încrucișată pe cele 5 fold-uri.
- Deviația standard a acurateței indică stabilitatea performanței modelului.

# Performanța Modelelor - Top Configurații (CNN și FCNN)

## Model Tip: CNN

Rang	Parametri Cheie Config. CNN	Acc. Medie
1.	CB: [(16,(5,5),(2,2)), (32,(3,3),(2,2))], Dense: [64], LR: 0.0005, DO: 0.0, Act: relu, BN: True, L2: 0.0	<b>0.6803 ± 0.0540</b>
2.	CB: [(32,(3,3),(2,2)), (64,(3,3),(2,2))], Dense: [64,32], LR: 0.0001, DO: 0.0, Act: relu, BN: True, L2: 0.0	0.6555 ± 0.0563
3.	CB: [(32,(3,3),(2,2)), (64,(3,3),(2,2))], Dense: [64], LR: 0.001, DO: 0.0, Act: relu, BN: True, L2: 0.0	0.6436 ± 0.0705

## Model Tip: FCNN

Rang	Parametri Cheie Config. FCNN	Acc. Medie
1.	HU: [128], Act: relu, DO: 0.5, L2: 0.0, LR: 0.0005, BN: True	<b>0.5773 ± 0.0445</b>
2.	HU: [64,64,32], Act: relu, DO: 0.2, L2: 0.001, LR: 0.001, BN: True	0.5732 ± 0.0576
3.	HU: [128,64], Act: relu, DO: 0.0, L2: 0.001, LR: 0.001, BN: True	0.5710 ± 0.0496

Abrevieri: HU=hidden\_units, Act=activation, DO=dropout\_rate, L2=l2\_reg, LR=learning\_rate, BN=use\_batch\_norm



# Performanța Modelelor - Top Configurații (RF și SVM)

## Model Tip: RF (Random Forest)

Rang	Parametri Config. RF	Acc. Medie
1.	n_est: 200, depth: 10, crit: gini, split: 5	<b>0.4652 ± 0.0439</b>
2.	n_est: 300, depth: 15, crit: gini, leaf: 3	0.4628 ± 0.0667
3.	n_est: 100, depth: 10, crit: gini	0.4628 ± 0.0652

## Model Tip: SVM (Support Vector Machine)

Rang	Parametri Config. SVM	Acc. Medie
1.	C: 1, kernel: rbf, gamma: 0.1	<b>0.4753 ± 0.0766</b>
2.	C: 10, kernel: rbf, gamma: 0.1	0.4701 ± 0.0653
3.	C: 10, kernel: sigmoid	0.4664 ± 0.0428

# Cel Mai Performant Model General

## Configurația Optimă a Modelului CNN

- **Arhitectură Convoluțională:** [(16 filtre, kernel (5,5), pool (2,2)), (32 filtre, kernel (3,3), pool (2,2))]
- **Straturi Dense:** [64 neuroni]
- **Rată Dropout:** 0.0 (fără dropout)
- **Rată de Învățare:** 0.0005
- **Funcție de Activare:** ReLU
- **Normalizare per Lot:** True
- **Regularizare L2:** 0.0 (fără L2)
- **Epoci maxime:** 30 (cu Early Stopping, patience=10)
- **Dimensiune Lot:** 4

## Rezultate Obținute

- **Acuratețe Medie (5-Fold CV): 0.6803**
- **Deviație Standard Acuratețe: 0.0540**

## Concluzii

# Concluzii și Direcții Viitoare

## Sumar Rezultate

- Modelele CNN au demonstrat cea mai bună performanță generală, cu o acuratețe medie de **0.6803** pentru configurația optimă.
- Modelele FCNN au urmat, cu acurateți medii în jur de 0.57, în timp ce SVM și RF au avut performanțe mai modeste, sub 0.50.
- Acest lucru sugerează că trăsăturile extrase automat de CNN-uri din spectrograme au fost mai informative pentru această sarcină comparativ cu trăsăturile algoritmice utilizate de celelalte modele.

# Concluzii și Direcții Viitoare

## Sumar Rezultate

- Modelele CNN au demonstrat cea mai bună performanță generală, cu o acuratețe medie de **0.6803** pentru configurația optimă.
- Modelele FCNN au urmat, cu acurateți medii în jur de 0.57, în timp ce SVM și RF au avut performanțe mai modeste, sub 0.50.
- Acest lucru sugerează că trăsăturile extrase automat de CNN-uri din spectrograme au fost mai informative pentru această sarcină comparativ cu trăsăturile algoritmice utilizate de celelalte modele.

## Pași Următori și Îmbunătățiri Posibile

- Explorarea unor arhitecturi CNN mai complexe sau pre-antrenate.
- Augmentarea setului de date (dacă este posibil) sau utilizarea tehnicilor de transfer learning.
- Analiza erorilor pentru a înțelege cazurile dificile pentru cel mai bun model.
- Investigarea altor metrici de performanță pe lângă acuratețe

Vă mulțumim pentru atenție!

---