KAGGLE SYSTEMS DESIGNS – NFL CONTACT DETECTION

ANDERSON DAVID ARENAS GUTIERREZ - 20231020030
NICOLAS FELIPE PULIDO SUAREZ – 20231020045
NAHIN JOSE PEÑARANDA MEJIA – 20231020032

UNIVERSIDAD DISTRITAL
FRANCISCO JOSÉ DE CALDAS

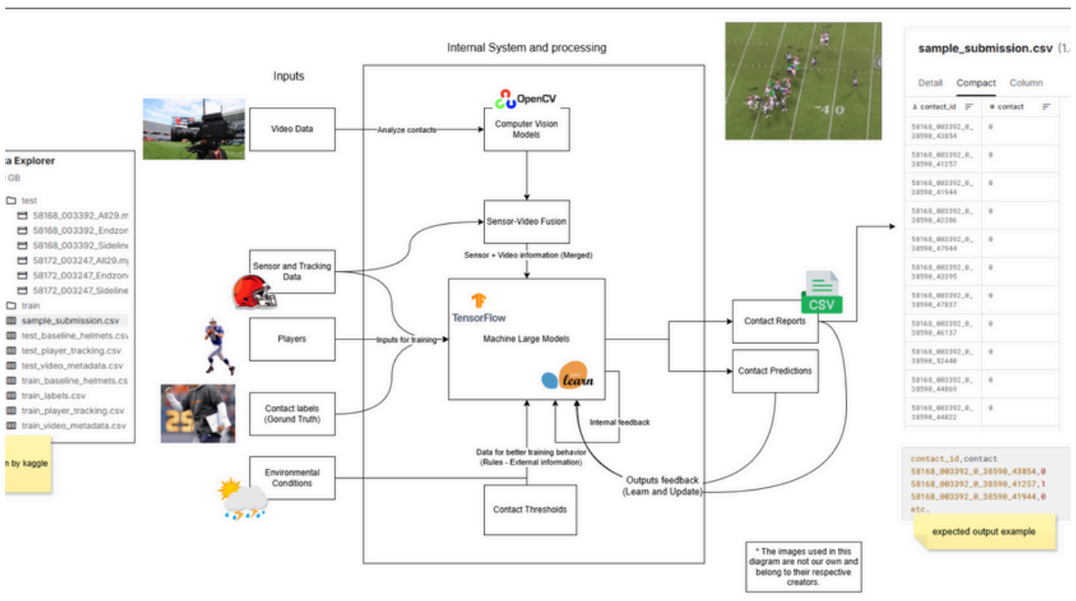
INTRODUCTION

Detecting player-to-player and player-to-ground contacts from tracking data can enhance safety and officiating. We propose a lightweight, CPU-friendly RandomForest approach on engineered kinematic features.

GOAL

Build and validate a RandomForest pipeline that predicts contact events using only position, speed, and acceleration data. Deliver an end-to-end workflow from raw CSVs to a Kaggle-ready submission.

PROPOSED SOLUTION



Compute relative features (Δx, Δy, distance, speed and acceleration differences) for each contact event. Train a class-weighted RandomForest with hyperparameter tuning via 5-fold cross-validation.

EXPERIMENTS

We split data 80/20 (train/validation) stratified by contact label and tuned tree depth and leaf size. Class weights balanced rare contact events, and the pipeline ran in minutes on a standard CPU.

RESULTS

Validation Accuracy ≈ 85 %; Precision ≈ 82 %; Recall ≈ 60 %. These metrics confirm that a RandomForest on tracking features is both effective and efficient.

```
Validation Accuracy: 0.9911
Validation Precision: 0.7804
Validation Recall: 0.4869
```

CONCLUSIONS

A RandomForest on numerical features provides a strong, resource-light baseline for contact detection. Future work could explore boosted trees, temporal models, and integrating video features.

BIBLIOGRAPHY

- Kaggle, “NFL Player Contact Detection” competition page.
- Pedregosa, F. et al., “Scikit-learn: Machine Learning in Python,” Journal of Machine Learning Research, 2011.
- Peñaranda Mejía, N. J., Pulido Suárez, N. F., Arenas Gutiérrez, A. D., “Workshop 1: Kaggle Competition Analysis,” Universidad Distrital Francisco José de Caldas, 2025.
- Peñaranda Mejía, N. J., Pulido Suárez, N. F., Arenas Gutiérrez, A. D., “Workshop 2: System Design and Architecture,” Universidad Distrital Francisco José de Caldas, 2025.