

# لیست مطالب

مقدمه و زمینه تحقیق

هدف مقاله و سوال اصلی

داده ها و روش شناسی

بررسی نتایج

نظر تیمی



# مقدمه



تشخیص دروغگویی در انسان‌ها در حوزه‌های حساس و پرمخاطره مثل اجرای قانون، امنیت، و دادگاه‌ها پیامدهای قابل توجهی دارد. تشخیص دقیق برای تضمین عدالت و امنیت حیاتی است. اما انسان‌ها در تشخیص فریب دقت پایینی دارند، به طور متوسط حدود ۵۴٪ است. بهتر است از یادگیری ماشینی (ML) و هوش مصنوعی برای افزایش دقت و کارایی سیستم‌های تشخیص فریب استفاده کنند. مدل‌های ML می‌توانند طیف وسیعی از سرنخ‌ها (بصری، کلامی، روانشناختی) را تحلیل کنند. چالش‌هایی که در یادگیری ماشین باهاش مواجه می‌شوند عبارت‌اند از: مدل‌های یادگیری ماشینی اغلب مانند یک "جعبه سیاه" عمل می‌کنند. دقت پیش‌بینی آن‌ها ممکن است بالا باشد، اما نمی‌توانند دلیل تصمیم خود را توضیح دهند. اما در حوزه‌های حساس مانند تشخیص فریب، فهمیدن اینکه چرا یک مدل به نتیجه‌ای رسیده، بسیار مهم است. این شفافیت برای اعتمادپذیری، شناسایی سوگیری‌ها، و بهبود مدل ضروری است. برای بهبود این راه حل از هوش مصنوعی توضیح‌پذیر (XAI) استفاده کنند. یعنی هدف XAI این است که تصمیمات مدل‌های ML را برای انسان قابل درک و اعتماد کند.

# هدف اصلی و سوال مقاله

## سوال مقاله

مسئله کلیدی این است که مدل‌های یادگیری ماشین تشخیص فریب، با وجود دقت بالا، به دلیل عملکرد "جعبه سیاه"، منطق پشت تصمیمات خود را توضیح نمی‌دهند. این موضوع، درک و اعتماد به آن‌ها را در شرایط پرخطر دشوار می‌سازد. بنابراین، سؤال اصلی این است: چگونه می‌توان با استفاده از XAI، فرآیند تصمیم‌گیری مدل‌های تشخیص فریب با ریسک بالا را شفاف‌سازی کرد و به طور خاص، نقش و اهمیت نشانه‌های بصری مختلف را در این فرآیند کشف نمود؟ کدام نشانه‌های بصری بیشترین ارتباط را با فریب دارند؟



## هدف مقاله

هدف اصلی این مطالعه، شکستن ماهیت "جعبه سیاه" مدل‌های یادگیری ماشین در تشخیص فریب با ریسک بالا و روشن کردن اهمیت نشانه‌های بصری در تصمیم‌گیری این مدل‌ها با استفاده از هوش مصنوعی قابل توضیح (XAI) است. در نهایت، پژوهش به دنبال افزایش شفافیت و قابلیت توضیح‌پذیری مدل‌ها برای بهبود اعتماد و کاربرد آن‌ها توسط انسان در زمینه‌های مهم مانند عدالت و امنیت است.



# داده های تحقیق



**REAL-LIFE TRIAL DATASET** مجموعه داده شامل کلیپ‌های ویدیویی دادگاه‌های عمومی است که نمونه‌های واقعی فریب در شرایط حساس را نشان می‌دهد. دارای ۱۲۱ کلیپ متوازن بین اظهارات صادقانه و فریبکارانه، شامل متهمان و شاهدان. شامل نشانه‌های صوتی، بصری و متنی است، اما تمرکز اصلی بر نشانه‌های بصری است. 39 ویژگی بصری استخراج شده‌اند و در ۷ دسته گروه‌بندی شده‌اند: دهان، چشم‌ها، نگاه، ابروها، سر، حرکات و دست.

# تکنیک های استفاده شده در مقاله

## تکنیک های ML

Multi-layer Perceptron (MLP)  
SVM (Support Vector Machine)  
Decision Trees  
Random Forests  
Logistic Regression  
KNN (K-Nearest-Neighbours)  
Naive Bayes  
LGBM  
XGBoost  
CatBoost



## تکنیک های XAI

PERMUTATION IMPORTANCE  
SINGLE-FEATURE PERMUTATION IMPORTANCE  
PARTIAL DEPENDENCE PLOTS (PDP)  
SINGLE FEATURE IMPACT  
FEATURE INTERACTION IMPACT  
SHAPLEY ADDITIVE EXPLANATIONS (SHAP)  
LOCAL INTERPRETABILITY  
GLOBAL INTERPRETABILITY

