به نام خدا

Explainable AI for Unraveling the Significance of Visual Cues in High Stakes Deception Detection

نام استاد

استاد امیر احمدی

دانشجویان:

ریحانه محبی

حنانه عباسی

مرتضی رسولی

لیست مطالب

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. چالش هایی که در یادگیری ماشین باهاش مواجه می شوند عبارت اند 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

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 .

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 ازXAI ، 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؟

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

نتایج مدل ها

Classifier	Accuracy	Precision	Recall	AUC
MLP	88.00%	86.67%	92.86%	84.42%
SVM	80.00%	80.00%	85.71%	81.17%
XGBoost	80.00%	80.00%	85.71%	88.96%
CatBoost	76.00%	75.00%	85.71%	68.83%
Random Forest	72.00%	76.92%	71.43%	76.36%
Decision Tree	72.00%	73.33%	78.57%	66.23%
Logistic Regression	68.00%	75.00%	64.29%	70.13%
KNN	68.00%	71.43%	71.43%	66.08%
Naive Bayes	60.00%	66.67%	57.14%	62.99%
LGBM	56.00%	61.54%	57.14%	58.44%

نظر تیم

این تحقیق به بررسی هوش مصنوعی قابل توضیح (XAI) در تشخیص فریب در محیط‌های پرخطر مانند دادگاه‌ها می‌پردازد. مدل‌های یادگیری ماشین، با وجود دقت بالا، اغلب مانند "جعبه سیاه" عمل می‌کنند که شفافیت تصمیماتشان را کاهش می‌دهد. برای رفع این مشکل، تحقیق از تکنیک‌های XAI شامل Permutation Importance، PDP، و SHAPاستفاده کرده است تا اهمیت نشانه‌های بصری مانند اخم و بالا بردن ابرو را در تشخیص فریب روشن کند. مدل MLPبا دقت ۸۸٪ و Recall ۹۲/۸۶٪ عملکرد برتری نشان داد و یافته‌ها حاکی از آن است که اخم بیشتر با فریب مرتبط است، در حالی که بالا بردن ابرو نشانه صداقت است .این پژوهش بر شفافیت مدل‌های هوش مصنوعی برای پذیرش اخلاقی و کاربرد در سیستم‌های انسان در حلقه (Human-in-loop AI) تأکید دارد. محدودیت‌های مطالعه شامل عدم اتوماسیون استخراج نشانه‌های بصری است که می‌تواند در تحقیقات آینده بهبود یابد.