عنوان مقاله:   
بررسی نقش سرنخ های بصری در تشخیص فریب با استفاده از هوش مصنوعی توضیح پذیر XAI در موقعیت های پرمخاطره

مقدمه و اهمیت موضوع:

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.

داده‌های استفاده شده در مقاله:

Real-Life Trial Dataset.

 این مجموعه داده به عنوان ارزشمندترین مجموعه داده در موقعیت‌های پرمخاطره شناخته می‌شود. شامل کلیپ‌های ویدئویی از محاکم واقعی دادگاه است.  نوع داده‌ها: شامل سرنخ‌های بصری (غیرکلامی) و کلامی است. نمونه‌های فریبکارانه و صادقانه در آن تقریباً متعادل هستند.

دسته‌بندی سرنخ‌های بصری: ۳۹ ویژگی بصری در ۷ دسته اصلی گروه‌بندی شده‌اند: دهان، چشم‌ها، نگاه، ابروها، سر، حرکات عمومی و دست.

فرآیند کلی روش‌شناسی و مدل‌های یادگیری ماشینی ایتنطوری است که شامل آماده‌سازی داده‌ها، انتخاب و تنظیم مدل‌های ML، و استفاده از تکنیک‌های XAI است. طیف متنوعی از مدل‌های ml شامل MLP, SVM, Decision Trees, Random Forests, Logistic Regression, KNN, Naive Bayes, LGBM, XGBoost, CatBoost بررسی شده است. و مدل Multi-layer Perceptron (MLP) بهترین عملکرد را داشت. دارای ویژگی های زیر است:

* دقت (Accuracy): ۸۸٪.
* Recall (نرخ مثبت واقعی): ۹۲.۸۶٪.
* Precision (ارزش پیش‌بینی مثبت): ۸۶.۶۷٪.
* AUC : ۸۴.۴۲٪.

تکنیک‌های XAI مورد استفاده:

* Permutation Importance: برای ارزیابی اهمیت هر ویژگی یا گروه با اندازه‌گیری کاهش عملکرد مدل هنگام به‌هم ‌ریختن مقادیر آن ویژگی/گروه.
* Partial Dependence Plots (PDP): برای نمایش رابطه بین یک یا دو ویژگی و خروجی پیش‌بینی شده مدل (مثلاً احتمال فریب).
* SHapley Additive exPlanations (SHAP): برای اختصاص سهم هر ویژگی به پیش‌بینی‌های منفرد (تفسیر محلی) و همچنین درک کلی اهمیت ویژگی‌ها در کل مجموعه داده (تفسیر سراسری).

معرفی تکنیک XAI نوآورانه این تحقیق که "Set-of-features Permutation Importance"

مثلاً در یک لحظه نمی‌توان هم اخم کرد و هم ابرو بالا انداخت، به‌هم‌ریختن ویژگی‌های تکی غیرواقعی است.

 هدف: ارزیابی تأثیر یک گروه کامل از ویژگی‌ها (مثل گروه ابروها) بر عملکرد مدل.

مهم ترین یافته های این تحقیق تحلیل‌های XAI (به خصوص Permutation Importance) است که به وضوح نشان داد که سرنخ‌های بصری مرتبط با حرکات ابروها (مانند بالا بردن ابرو و اخم کردن) مهمترین نقش را در پیش‌بینی فریب دارند. این یافته در مدل‌های برتر سازگار بود.

روش Set-of-features Permutation Importance نیز تأیید کرد که گروه "ابروها" تأثیرگذارترین گروه ویژگی‌ها بر عملکرد مدل MLP و SVM است (به جز CatBoost که در آن گروه دست مهمتر بود).

PDPs:

* اخم کردن (Frowning): دارای شیب مثبت در PDP، نشان‌دهنده افزایش تمایل مدل به پیش‌بینی فریب.
* بالا بردن ابروها (Raising eyebrows): دارای شیب منفی در PDP، نشان‌دهنده تمایل مدل به پیش‌بینی صداقت.
* شیب تند این ویژگی‌ها نشان‌دهنده تأثیر بالای آن‌هاست.

نتیجه کلی:

* XAI به ما کمک می‌کند تا درک عمیق‌تری از نحوه عملکرد مدل و تأثیر ویژگی‌های خاص داشته باشیم.
* این مطالعه نشان داد که حرکات ظریف ابروها می‌تواند سیگنال‌های بسیار قوی برای تشخیص فریب باشد، چیزی که شاید انسان‌ها به راحتی متوجه آن نشوند یا اهمیت آن را درک نکنند.
* یافته‌ها بر اهمیت شفافیت مدل‌های ML در حوزه‌های حساس تأکید می‌کنند. XAI ابزاری حیاتی برای رسیدن به این هدف است.
* این بینش‌ها می‌توانند به بهبود مدل‌های آینده و حتی شاید آموزش انسان‌ها برای تشخیص بهتر فریب کمک کنند.