**چکیده:**

این مطالعه به بررسی چالش تشخیص فریب، به ویژه در محیط‌های پرریسک، می‌پردازد. در حالی که یادگیری ماشین توانایی تشخیص فریب را از طریق تحلیل نشانه‌های رفتاری و بصری افزایش داده است، مدل‌های "جعبه سیاه" اغلب فاقد شفافیت لازم برای درک منطق تصمیماتشان هستند. این مقاله با استفاده از تکنیک‌های هوش مصنوعی قابل توضیح (XAI) مدل-آگنستیک (Permutation Importance، Partial Dependence Plots، و SHapley Additive exPlanations) و یک تکنیک جدید به نام "set-of-features permutation importance"، اهمیت نشانه‌های بصری را در تشخیص فریب در مجموعه داده Real-Life Trial بررسی می‌کند.

**تم‌ها و ایده‌های اصلی:**

1.

**اهمیت هوش مصنوعی قابل توضیح (XAI) در تشخیص فریب با ریسک بالا:**

◦

مدل‌های یادگیری ماشین در تشخیص فریب به دلیل توانایی تحلیل داده‌های پیچیده پیشرفت‌های قابل توجهی داشته‌اند.

◦

اما مدل‌های یادگیری ماشین اغلب به عنوان "جعبه سیاه" عمل می‌کنند و پیش‌بینی‌هایی با دقت بالا ارائه می‌دهند بدون اینکه دلیل تصمیماتشان را توضیح دهند. این عدم شفافیت، قابلیت درک و اعتماد به مدل را کاهش می‌دهد.

◦

**"این عدم شفافیت نیاز به ادغام هوش مصنوعی قابل توضیح را برای قابل درک و قابل اعتماد کردن تصمیمات مدل‌ها ضروری می‌سازد."**

◦

در زمینه‌های پرریسک مانند اجرای قانون و دادرسی قضایی، درک منطق پشت تصمیمات مدل‌های تشخیص فریب برای اطمینان از عدالت و امنیت حیاتی است.

◦

XAI امکان درک و اعتماد متخصصان (مانند افسران تحقیقاتی) به تصمیمات مدل را فراهم می‌کند و پاسخگویی را افزایش می‌دهد.

2.

**تکامل روش‌های تشخیص فریب و نقش یادگیری ماشین:**

◦

روش‌های سنتی (مانند پلی‌گراف) مبتنی بر ارزیابی‌های روانشناختی و اندازه‌گیری‌های فیزیولوژیکی هستند و دارای محدودیت‌هایی مانند نیاز به همکاری فرد و احتمال سوگیری یا فریب‌کاری هستند.

◦

روش‌های غیرتهاجمی (مانند تحلیل کلامی، بصری، رفتاری و آکوستیک) جایگزین کم‌تداخلی را ارائه می‌دهند.

◦

تحلیل نشانه‌های غیرتهاجمی اغلب برای انسان بسیار پیچیده است و نیاز به استفاده از یادگیری ماشین را توجیه می‌کند.

◦

یادگیری ماشین به طور قابل توجهی سیستم‌های تشخیص فریب را در تحلیل داده‌های پیچیده مانند ظرافت‌های زبانی و حرکات چهره بهبود بخشیده است.

3.

**مجموعه داده Real-Life Trial:**

◦

این مطالعه بر روی مجموعه داده Real-Life Trial تمرکز دارد که به عنوان ارزشمندترین مجموعه داده پرریسک شناخته می‌شود.

◦

**"انتخاب آن ناشی از نیاز حیاتی به تحلیل فریب در محیط‌های پرریسک بود، جایی که پیامدهای فریب می‌تواند به طور قابل توجهی بر نتایج محاکمه و تصمیمات قضایی تأثیر بگذارد."**

◦

این مجموعه داده شامل 121 کلیپ ویدیویی از اظهارات فریبکارانه و صادقانه جمع‌آوری شده از ضبط‌های دادگاه‌های عمومی است.

◦

مجموعه داده Real-Life Trial چندوجهی است و نشانه‌های کلامی و غیرکلامی را شامل می‌شود.

◦

برای تحلیل، داده‌ها پیش‌پردازش شدند، شامل رونویسی محتوای صوتی، حاشیه‌نویسی رفتارهای غیرکلامی با استفاده از طرح کدگذاری MUMIN، و گروه‌بندی 39 ویژگی بصری به هفت دسته (دهان، چشم‌ها، نگاه، ابروها، سر، حرکات و دست).

◦

نحوه گروه‌بندی ویژگی‌ها در این مجموعه داده، الهام‌بخش توسعه تکنیک جدید "set-of-features permutation importance" بود.

4.

**کاربرد تکنیک‌های هوش مصنوعی قابل توضیح (XAI) مدل-آگنستیک:**

◦

این مطالعه از سه تکنیک مدل-آگنستیک XAI استفاده می‌کند: Permutation Importance، Partial Dependence Plots (PDP) و SHapley Additive exPlanations (SHAP). این تکنیک‌ها به درک چگونگی تأثیر ویژگی‌ها بر پیش‌بینی مدل کمک می‌کنند.

◦

**Permutation Importance:** اهمیت یک یا گروهی از ویژگی‌ها را با اندازه‌گیری کاهش عملکرد مدل هنگام جابجایی تصادفی مقادیر آن‌ها ارزیابی می‌کند. \* Single-feature Permutation Importance: تأثیر مستقل هر ویژگی را می‌سنجد. \* **"Novel Proposed Set-of-features Permutation Importance":** این تکنیک جدید، اهمیت گروه‌هایی از ویژگی‌ها (مانند گروه ابروها) را به طور کلی ارزیابی می‌کند تا از ایجاد ورودی‌های غیرواقعی ناشی از جابجایی ویژگی‌های تکی در گروه‌هایی که تنها یک ویژگی باید فعال باشد، جلوگیری کند. **"این تکنیک جدید به واقعیت‌گرایانه‌تر به نشانه‌های بصری مجموعه داده Real Life Trial می‌پردازد."**

◦

**Partial Dependence Plots (PDP):** رابطه بین یک یا دو ویژگی و نتیجه پیش‌بینی شده مدل را به صورت گرافیکی نشان می‌دهد و تأثیر آن‌ها را بر پیش‌بینی روشن می‌کند (تأثیر تک ویژگی و تعامل ویژگی).

◦

**SHapley Additive exPlanations (SHAP):** سهم هر ویژگی را در پیش‌بینی‌های تکی (تفسیر محلی) و همچنین اهمیت کلی ویژگی‌ها در کل مجموعه داده (تفسیر جهانی) اندازه‌گیری می‌کند. SHAP نشان می‌دهد که آیا یک ویژگی پیش‌بینی را به سمت فریب یا حقیقت سوق می‌دهد.

5.

**یافته‌های کلیدی در مورد نشانه‌های بصری مرتبط با فریب:**

◦

از میان مدل‌های یادگیری ماشین ارزیابی شده (MLP، SVM، Decision Trees، Random Forests، Logistic Regression، KNN، Naive Bayes، LGBM، XGBoost، و CatBoost)، Multi-layer Perceptron (MLP) بهترین عملکرد را با دقت 88% و recall 92.86% نشان داد.

◦

تحلیل با استفاده از تکنیک‌های XAI به طور مداوم نشان داد که **"نشانه‌های بصری مرتبط با حرکات ابرو بیشترین نشانگر رفتار فریبکارانه هستند."**

◦

به طور خاص، **"نتایج تجربی نشان می‌دهد که نشانه‌های بصری مرتبط با حرکات ابرو بیشترین نشانگر رفتار فریبکارانه هستند."**

◦

**"تحلیل اهمیت جایگشتی نشان می‌دهد که ویژگی‌های مرتبط با حرکات ابرو—به طور خاص بالا بردن و اخم کردن—مشارکت‌کنندگان قابل توجهی در پیش‌بینی‌های مدل هستند."** این یافته در تمام چهار مدل برتر (MLP، SVM، XGBoost، و Catboost) پایدار بود.

◦

PDPها نشان دادند که اخم کردن تمایل به سیگنال فریب دارد، در حالی که بالا بردن ابروها نشانگر حقیقت است. شیب PDP نشان‌دهنده تأثیر ویژگی بر پیش‌بینی بود، با شیب‌های تند برای "Frown" و "Raise" که تأثیر بالای آن‌ها را نشان می‌داد.

◦

تحلیل تعامل ویژگی‌ها نشان داد که هنگامی که "Frown" فعال و "Raise" غیرفعال است، پیش‌بینی به سمت فریب متمایل می‌شود و برعکس. این تأثیرات با PDPهای تک ویژگی مطابقت داشتند.

◦

Global SHAP برای MLP نشان داد که "Frowning" هنگام فعال بودن تأثیر قابل توجهی دارد، در حالی که "Raising" بدون توجه به وضعیتش اهمیت نشان می‌دهد.

6.

**"Human-in-loop AI" و هوش مصنوعی قابل اعتماد:**

◦

شفافیت و قابلیت تفسیر مدل‌های یادگیری ماشین برای ادغام موفق "human-in-loop AI" در زمینه تشخیص فریب ضروری است.

◦

XAI با قابل درک کردن تصمیمات مدل، به متخصصان اجازه می‌دهد تا به سیستم‌های هوش مصنوعی اعتماد کرده و از آن‌ها استفاده کنند.

◦

**"کار ما اهمیت شفاف‌تر و قابل توضیح‌تر کردن مدل‌های یادگیری ماشین را تأکید می‌کند، و در نتیجه سودمندی آن‌ها را برای هوش مصنوعی با انسان در حلقه و پذیرش اخلاقی افزایش می‌دهد."**

**حقایق و ایده‌های مهم از نظر نگارنده:**

•

نیاز فوری به شفافیت در مدل‌های یادگیری ماشین در زمینه‌های پرریسک تشخیص فریب.

•

مجموعه داده Real-Life Trial به عنوان یک منبع با ارزش برای مطالعه فریب در محیط‌های واقعی و پرخطر.

•

معرفی تکنیک جدید "set-of-features permutation importance" که با ماهیت گروه‌بندی شده نشانه‌های بصری در مجموعه داده‌های واقعی مطابقت دارد.

•

یافته اصلی این مطالعه که حرکات ابرو (اخم کردن و بالا بردن) قوی‌ترین نشانه‌های بصری مرتبط با فریب هستند.

•

برتری مدل MLP در این مطالعه برای تشخیص فریب بر اساس نشانه‌های بصری.

•

نقش حیاتی XAI (Permutation Importance، PDP، SHAP) در ارائه بینش‌های قابل درک در مورد چگونگی تصمیم‌گیری مدل‌ها.

•

تأکید بر اهمیت human-in-loop AI و هوش مصنوعی قابل اعتماد در کاربردهای عملی تشخیص فریب.

**محدودیت‌ها و کارهای آینده:**

•

فقدان اتوماسیون در استخراج ویژگی‌های بصری از افراد.

•

تمرکز انحصاری بر نشانه‌های بصری بدون در نظر گرفتن انواع دیگر نشانه‌ها (مانند کلامی، آکوستیک).

•

کارهای آینده باید به این محدودیت‌ها پرداخته و سایر انواع نشانه‌ها را نیز در نظر بگیرند تا سیستم‌های تشخیص فریب جامع‌تری توسعه یابند.