دانشگاه صنعتی اصفهان

دانشکده مهندسی برق و کامپیوتر

کاربرد های نوین شبکه های عصبی و یادگیری عمیق

مجتبی ملائی، سامان اصغری، مهتا میرزائی، کوروش جمشیدی

استاد درس

دکتر مریم بقولی زاده

۱۴۰۴

چکیده

یادگیری عمیق[1] به‌عنوان یکی از شاخه‌های پرکاربرد هوش مصنوعی، توانسته است در حوزه‌هایی مانند پزشکی، امنیت سایبری[2] و تحلیل محتوای دیجیتال، نتایج قابل توجهی به دست آورد. در یک پژوهش، یک شبکه عصبی کانولوشنی[3] ۹ لایه برای تشخیص خودکار انواع ضربان‌های قلبی از روی سیگنال الکتروکاردیوگرام[4] طراحی شد. این مدل توانست پنج نوع ضربان مختلف (از جمله ضربان‌های نارسایی بطنی و فوق‌بطنی) را حتی در شرایط نویزی با دقتی در حدود ۹۴٪ شناسایی کند. این مدل با استفاده از داده‌های تقویت‌شده و پاک‌سازی‌شده آموزش دید و توانست ابزار مناسبی برای غربالگری سریع آریتمی‌ها باشد.

در حوزه رسانه‌های دیجیتال نیز، مطالعه‌ای به بررسی روش‌های تشخیص ویدیوهای جعلی (جعل عمیق)[5] با استفاده از دو معماری معروف Xception و MobileNet پرداخت. این مدل‌ها بر اساس داده‌های مجموعه‌ی FaceForensics++ آموزش دیدند که شامل ویدیوهایی با چهار روش مختلف ساخت جعل عمیق است. نتایج نشان داد که مدل‌ها دقتی بین ۹۱٪ تا ۹۸٪ دارند و استفاده از یک مکانیزم رأی‌گیری میان آن‌ها باعث افزایش دقت نهایی سیستم در تشخیص ویدیوهای جعلی شد.

در حوزه امنیت سایبری نیز، یک مدل سریع و سبک برای تشخیص نشانی‌های فیشینگ[6] معرفی شد که بر پایه‌ی شبکه عصبی کانولوشنی در سطح کاراکتر عمل می‌کند. این روش برخلاف روش‌های سنتی نیازی به استخراج محتوای صفحات وب یا استفاده از سرویس‌های شخص ثالث ندارد. مدل با یادگیری الگوهای متنی موجود در نشانی وب، توانست بدون نیاز به ویژگی‌های دستی، نشانی‌های مخرب را با دقتی بالاتر از ۹۵٪ شناسایی کند و در چند پایگاه داده‌ی معتبر نیز دقتی بالاتر از روش‌های موجود به ثبت رساند.

در مجموع، این مطالعات نشان می‌دهند که یادگیری عمیق با بهره‌گیری از معماری‌های مناسب و داده‌های غنی، می‌تواند در حل مسائل پیچیده در حوزه‌های حیاتی و چالش‌برانگیز عملکرد مؤثری داشته باشد.

واژه‌های کلیدی:
تشخیص فیشینگ، مهندسی ویژگی‌های نشانی، نهفته‌سازی کاراکتری، یادگیری عمیق، تشخیص ضربان قلب، آریتمی، بیماری‌های قلبی، شبکه عصبی کانولوشنی، سیگنال الکتروکاردیوگرام

[1] Deep Learning
[2] Cybersecurity
[3] Convolutional Neural Network - CNN
[4] Electrocardiogram - ECG
[5] Deepfake
[6] Phishing URLs

فصل ۱

مقدمه

با گسترش سریع فناوری‌های مبتنی بر یادگیری عمیق[۱]، کاربردهای متنوعی در حوزه‌های مختلف از جمله امنیت سایبری[۲]، تحلیل داده‌های زیستی، و پردازش سیگنال‌های پزشکی ایجاد شده است. در این میان، سه چالش مهم و نوظهور شامل تشخیص محتوای جعلی (جعل عمیق)[۳]، شناسایی وب‌سایت‌های فیشینگ[۴]، و تحلیل خودکار نوار قلب (الکتروکاردیوگرام)[۵] به‌عنوان موضوعات کلیدی پژوهش‌های اخیر مطرح شده‌اند.

جعل‌های عمیق، که با بهره‌گیری از شبکه‌های مولد تخاصمی[۶] تولید می‌شوند، قابلیت شبیه‌سازی چهره و صدای افراد با دقت بالا را دارند و تهدیدی جدی برای اعتماد عمومی، انتخابات، و امنیت اطلاعات محسوب می‌شوند. در این زمینه، مدل‌های سبک و قدرتمند مانند Xception و MobileNet با استفاده از مجموعه‌داده‌ی FaceForensics++ جهت تشخیص ویدیوهای جعلی آموزش دیده‌اند و با ترکیب خروجی آن‌ها، مکانیزم رأی‌گیری برای افزایش دقت سیستم پیشنهاد شده است.

در حوزه امنیت سایبری، حملات فیشینگ همچنان از مهم‌ترین روش‌های سوءاستفاده مهاجمان برای سرقت اطلاعات شخصی کاربران محسوب می‌شود. بسیاری از روش‌های سنتی مانند لیست‌های سیاه یا تحلیل مبتنی بر خدمات شخص ثالث، در مقابله با حملات روز صفر[۷] کارایی کافی ندارند. به همین دلیل، در این پژوهش از شبکه عصبی کانولوشنی[۸] در سطح کاراکتر برای تحلیل مستقیم رشته‌ی نشانی[۹] استفاده شده است؛ روشی که مستقل از زبان، سریع، و بدون نیاز به مهندسی ویژگی دستی عمل می‌کند.

در حوزه پزشکی نیز بیماری‌های قلبی–عروقی، به‌ویژه آریتمی‌ها که ناشی از اختلال در سیستم الکتریکی قلب هستند، همچنان عامل اصلی مرگ‌ومیر در جهان‌اند. تشخیص دقیق این اختلالات از روی سیگنال‌های

[1] Deep Learning
[2] Cybersecurity
[3] Deepfake
[4] Phishing Websites
[5] Electrocardiogram - ECG
[6] Generative Adversarial Networks - GANs
[7] Zero-day Attacks
[8] Convolutional Neural Network - CNN
[9] URL

الکتروکاردیوگرام نیازمند تحلیل دقیق فرم موج و تفکیک انواع ضربان‌ها است. در این راستا، مدل پیشنهادی ما از یک شبکه عصبی پیچشی عمیق برای طبقه‌بندی خودکار ۵ نوع ضربان غیرعادی بهره می‌برد و توانسته است دقت بالایی را در داده‌های دارای نویز و بدون نویز ثبت کند.

ترکیب این سه مطالعه نشان می‌دهد که یادگیری عمیق با بهره‌گیری از معماری‌های بهینه‌ی کانولوشنی می‌تواند راهکارهای دقیق، سریع و مقیاس‌پذیر برای حل مسائل پیچیده در حوزه‌های پزشکی، امنیت دیجیتال و رسانه فراهم آورد.

فصل ۲

یک مدل CNN برای تشخیص ضربان قلب

۲-۱ مقدمه

بیماری‌های قلبی‌عروقی (CVD) [۱] عامل اصلی مرگ‌ومیر در سراسر جهان هستند. طبق گزارش سازمان جهانی بهداشت[۲]، در سال ۲۰۱۵ حدود ۱۷.۷ میلیون نفر به‌علت این بیماری‌ها جان باختند. CVD به‌طور کلی به سه دسته تقسیم می‌شود: اختلالات الکتریکی (آریتمی)[۳]، اختلالات گردش خون و بیماری‌های ساختاری قلب. تمرکز این پژوهش بر آریتمی‌هاست که به اختلالات الکتریکی قلب مربوط می‌شود.

آریتمی‌ها می‌توانند به شکل ضربان آهسته، سریع یا نامنظم ظاهر شوند و به دو دسته تهدیدکننده[۴] و غیر تهدیدکننده حیات تقسیم می‌شوند. تشخیص آن‌ها از طریق بررسی نوار قلب (ECG) و طبقه‌بندی ضربان‌ها بر اساس فرم سیگنال صورت می‌گیرد. بر اساس استاندارد AAMI، آریتمی‌های غیر تهدیدکننده در پنج کلاس[۵] N، S، V، F و Q قرار می‌گیرند.

تفاوت‌های بارز در شکل سیگنال‌های ECG برای هر نوع آریتمی باعث دشواری در شناسایی دقیق آن‌ها می‌شود. ارزیابی دستی نوار قلب ممکن است با خطای انسانی همراه باشد. بنابراین، توسعه سیستم‌های تشخیص کامپیوتری (CAD)[۶] با استفاده از یادگیری ماشین مورد توجه قرار گرفته است. در روش‌های کلاسیک نیاز به استخراج و انتخاب ویژگی به‌صورت دستی وجود دارد که ممکن است موجب بیش‌برازش[۷] شود.

در مقابل، یادگیری عمیق[۸] این امکان را می‌دهد که مدل به‌صورت خودکار ویژگی‌ها را از داده‌های خام استخراج کند. پژوهش‌های مختلف نشان داده‌اند که مدل‌های مبتنی بر یادگیری عمیق دقت بالاتری در

[1] Cariovascular Diseases
[2] World Health Organization
[3] Arrhythmia
[4] Arrhythmia
[5] Class
[6] Computer-Aided Diagnosis
[7] Overfit
[8] Deep Learning

طبقه‌بندی[1] ECG دارند. در این تحقیق، مدلی بر پایه شبکه‌های عصبی کانولوشنی[2] برای شناسایی ۵ نوع ضربان غیرعادی ECG معرفی شده است. این کار ادامه‌ی پژوهش‌های پیشین ما در زمینه تشخیص آریتمی و بیماری‌های قلبی با استفاده از CNN است.

۲-۱-۱ پایگاه داده ECG

سیگنال‌های ECG از پایگاه داده[3] Arrhythmia MIT-BIH گرفته شده‌اند که شامل ۴۸ ضبط نیم‌ساعته از ۴۷ فرد می‌باشد. سیگنال‌ها با فرکانس ۳۶۰ هرتز ثبت شده‌اند و طول هر ضربان ۲۶۰ نمونه است. این داده‌ها توسط حداقل دو متخصص قلب تفسیر و تأیید شده‌اند. مجموعه‌ای از داده‌ها بدون فیلتر (مجموعه A) و مجموعه‌ای با حذف نویز[4] (مجموعه B) تهیه شده‌اند.

۲-۲ روش پیشنهادی

۲-۲-۱ پیش‌پردازش

برای حذف نویز و خط مبنا، از فیلتر wavelet با تابع Daubechies سطح ۶ استفاده شده است.

۲-۲-۲ تولید داده مصنوعی

ضربان‌های ECG در مجموعه‌های A و B استخراج و حول نقطه R تقسیم‌بندی شده‌اند. سپس نرمال‌سازی[5] Z-score انجام شده تا مشکل مقیاس دامنه و افست[6] برطرف شود. برای رفع عدم تعادل بین کلاس‌ها[7]، از داده مصنوعی استفاده شده است. ضربان‌های کلاس N دست‌نخورده باقی مانده و سایر کلاس‌ها با داده‌های مصنوعی هم‌تراز شده‌اند. پس از افزایش، تعداد کل ضربان‌ها به ۹۶۰,۴۵۲ رسید.

۲-۲-۳ شبکه عصبی کانولوشنی (CNN)

CNN نوعی شبکه عصبی عمیق است که به‌دلیل ساختار خاص خود نسبت به چرخش[8] و انتقال مقاوم است. معماری مدل پیشنهادی شامل ۹ لایه تشکیل شده از سه لایه کانولوشن، سه لایه pooling و سه لایه fully connected. توابع فعال‌سازی از نوع LeakyReLU هستند و لایه نهایی با تابع Softmax، پنج کلاس N، S، V، F و Q را خروجی می‌دهد.

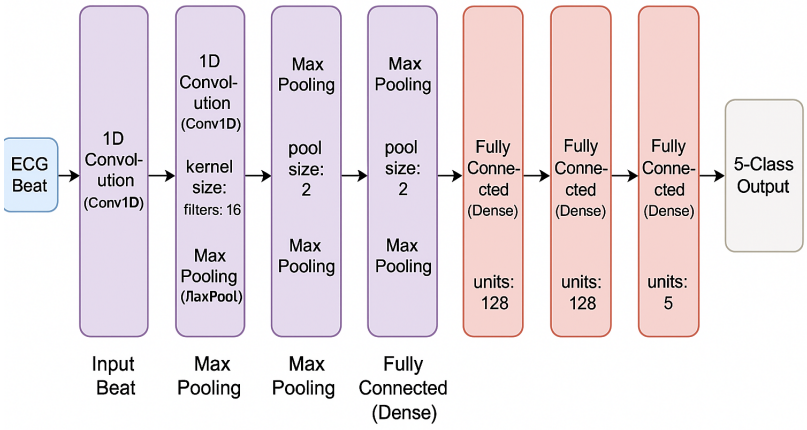
---

[1] Clustering
[2] CNN: Convolutional Neural Network
[3] Database
[4] Noise
[5] Normalization
[6] Off-set
[7] Class Imbalance
[8] Rotatin

۲-۲-۴ جزئیات معماری و آموزش

لایه‌های کانولوشن دارای کرنل‌هایی[1] با اندازه‌های ۳، ۴ و ۴ هستند. stride برای کانولوشن و pooling به‌ترتیب ۱ و ۲ در نظر گرفته شده‌اند. آموزش مدل با الگوریتم backpropagation انجام شده و پارامترهای آموزش عبارت‌اند از: نرخ یادگیری 3×10^{-3}، ضریب منظم‌سازی 0.2، و مومنتوم[2] 0.7.

آموزش در ۲۰ تکرار[3] صورت گرفته و اعتبارسنجی[4] پس از هر تکرار انجام شده است. همچنین از اعتبارسنجی متقاطع ۱۰ بخشی استفاده شده که میانگین دقت، حساسیت و ویژگی برای ارزیابی نهایی گزارش شده‌اند.



شکل ۲-۱: معماری شبکه عصبی

۲-۳ نتایج و بحث

الگوریتم پیشنهادی شبکه عصبی کانولوشنی بر روی یک سیستم دارای دو پردازنده Intel Xeon 2.40GHz و ۲۴ گیگابایت رم آموزش داده شد. مدت‌زمان آموزش برای هر تکرار به‌طور میانگین حدود ۹۵۷۳ ثانیه برای مجموعه‌ی A (بدون حذف نویز) و ۹۵۸۶ ثانیه برای مجموعه‌ی B (با حذف نویز) بود. پیاده‌سازی الگوریتم در نرم‌افزار MATLAB انجام شده است.

جداول 4 و 5 ماتریس سردرگمی[5] حاصل از طبقه‌بندی ضربان‌های قلبی در مجموعه‌های A و B را نشان می‌دهد. در مجموعه‌ی A کمتر از ۱۲٪ و در مجموعه‌ی B کمتر از ۱۰٪ خطای طبقه‌بندی مشاهده شده است.

[1] Kernel
[2] momentum
[3] Epochs
[4] Validation
[5] Confusion Matrix

کمترین دقت پیش‌بینی مثبت (PPV) مربوط به کلاس N بوده که به‌ترتیب ۸۵٪ و ۸۷٪ ثبت شده است. خلاصه‌ای از عملکرد مدل در جدول 6 آورده شده است.

از آنجا که عدم تعادل داده‌ها بر دقت طبقه‌بندی تأثیر منفی دارد، داده‌های مصنوعی تولید شدند تا تعداد نمونه‌ها در تمام کلاس‌ها یکسان شود. نتایج نشان دادند که مدل آموزش‌دیده با داده‌های متعادل عملکرد بسیار بهتری نسبت به مدل آموزش‌دیده با داده‌های نامتوازن دارد. به‌طور مثال، در داده‌های نامتوازن، دقت PPV کلاس F تا حدود ۳۵٪ کاهش یافته است. نتایج این آزمایش در جداول پیوست A1 و A2 گزارش شده‌اند.

مقایسه‌ی نتایج دو مجموعه‌ی A و B نشان می‌دهد که مدل CNN حتی بدون حذف نویز نیز عملکرد قابل قبولی دارد. این نشان می‌دهد که مدل توانایی یادگیری فیلترهایی برای حذف نویز را به‌طور خودکار دارد.

۲-۳-۱ مزایای مدل پیشنهادی

- به‌طور کامل خودکار است و نیازی به استخراج یا انتخاب ویژگی ندارد.
- به کیفیت سیگنال ECG حساس نیست.
- از اعتبارسنجی متقاطع ده‌تایی استفاده شده که باعث افزایش پایداری مدل شده است.

۲-۳-۲ محدودیت‌ها

- نیاز به زمان آموزش طولانی، سخت‌افزار قوی (GPU) و هزینه‌ی محاسباتی بالا دارد.
- برای آموزش مؤثر، نیاز به حجم بالایی از داده است.

با این‌حال، پس از اتمام آموزش، طبقه‌بندی ضربان‌های قلبی به‌سرعت انجام می‌شود. این سیستم می‌تواند در محیط‌های بالینی و حتی مناطق محروم به‌عنوان ابزاری کمکی برای پزشکان مورد استفاده قرار گیرد.

۲-۴ نتیجه‌گیری

در این پژوهش، یک روش یادگیری عمیق برای شناسایی و طبقه‌بندی خودکار انواع مختلف ضربان‌های قلبی ECG ارائه شده است که در تشخیص آریتمی قلبی بسیار حیاتی است. مدل CNN توسعه‌یافته قادر به طبقه‌بندی ۵ نوع مختلف از ضربان‌های قلبی است و می‌تواند به‌عنوان بخشی از یک سیستم تشخیص کمک‌پزشکی (CAD) برای تشخیص سریع و قابل‌اعتماد به‌کار گرفته شود.

این مدل پتانسیل استفاده در محیط‌های بالینی را دارد و می‌تواند به عنوان ابزار کمکی به پزشکان در تفسیر سیگنال‌های ECG کمک کند. همچنین پیاده‌سازی آن در کلینیک‌ها، چه به‌صورت آنلاین و چه آفلاین، برای

غربالگری سریع تعداد زیادی از نوار قلب‌ها، می‌تواند زمان انتظار بیماران را کاهش داده، بار کاری پزشکان را کم کند و هزینه‌های پردازش سیگنال ECG در بیمارستان‌ها را کاهش دهد.

در مطالعات آتی، نویسندگان قصد دارند مدل پیشنهادی را با آموزش CNN برای تشخیص دنباله‌های زمانی ضربان‌های قلبی توسعه دهند. توالی، الگوهای وقوع و پایداری پنج کلاس N، S، V F و Q می‌توانند در سه دسته‌ی اصلی سبز، زرد و قرمز قرار گیرند که به‌ترتیب نشان‌دهنده وضعیت طبیعی، غیرطبیعی و بالقوه خطرناک فعالیت الکتریکی قلب هستند.

همچنین، برنامه‌ریزی شده است تا عملکرد مدل CNN در مواجهه با داده‌های متعادل‌شده[1] و داده‌هایی با سطوح مختلف نویز بررسی گردد.

[1] de-skewed

فصل ۳

تشخیص جعل عمیق[1] با استفاده از یادگیری عمیق[2]

۳-۱ مقدمه

با پیشرفت چشمگیر فناوری‌های هوش مصنوعی، تولید محتوای جعلی، به‌ویژه ویدیوهای جعلی موسوم به جعل عمیق، به یکی از چالش‌های جدی در حوزه‌های امنیتی، سیاسی و اجتماعی تبدیل شده است. این ویدیوها عمدتاً با بهره‌گیری از شبکه‌های مولد تخاصمی[3] تولید می‌شوند و می‌توانند ظاهر، حرکات و حتی صدای افراد را با دقت بالایی بازسازی کنند.

در آستانهٔ انتخابات ریاست‌جمهوری ایالات متحده در سال ۲۰۲۰، نگرانی‌ها دربارهٔ استفاده از جعل‌های عمیق برای تأثیرگذاری بر افکار عمومی افزایش یافت. در پاسخ، بسترهایی[4] مانند فیسبوک[5] و اینستاگرام[6] سیاست‌هایی برای حذف محتوای دست‌کاری‌شده توسط هوش مصنوعی اتخاذ کردند. با این حال، تشخیص سریع و دقیق این ویدیوها مستلزم بهره‌گیری از سامانه‌های خودکار و مبتنی بر یادگیری عمیق است.

در این پژوهش، دو معماری قدرتمند در حوزهٔ بینایی ماشین یعنی Xception و MobileNet به‌منظور تشخیص خودکار ویدیوهای جعل عمیق به‌کار گرفته شده‌اند. این مدل‌ها با استفاده از مجموعه‌دادهٔ FaceForensics++ که شامل ویدیوهای جعلی تولیدشده با چهار روش رایج است، آموزش دیده‌اند. به‌منظور افزایش دقت نهایی، یک سازوکار رأی‌گیری میان خروجی این مدل‌ها طراحی و پیاده‌سازی شده است.

---

[1] DeepFake
[2] Deep Learning
[3] GAN
[4] Platforms
[5] Facebook
[6] Instagram

۳-۲ کارهای مرتبط

تکنیک‌های تولید جعل عمیق به‌طور کلی به دو دستهٔ اصلی **تعویض چهره**[۱] **و بازسازی حالات چهره**[۲] تقسیم می‌شوند.

۳-۲-۱ تعویض چهره

در این روش، چهرهٔ فردی با چهرهٔ فرد دیگری جایگزین می‌شود. یکی از اولین نمونه‌های کاربردی این تکنیک توسط کاربران فضای مجازی و با استفاده از ساختارهای سادهٔ رمزگذار_رمزگشا[۳] ارائه شد. پس از آن، روش‌های پیشرفته‌تری همچون Faceswap-GAN معرفی شدند که با افزودن ضررهای ادراکی و متضاد[۴]، دقت بازسازی در نواحی حساس مانند چشم و دهان را افزایش دادند. همچنین، تکنیک‌هایی مانند Fast Face-swap با بهره‌گیری از انتقال سبک[۵] و شبکه‌های پیش‌آموزش‌دیده مانند VGG19[۶]، توانستند کنترل بیشتری بر ویژگی‌های ظاهری و محتوایی اعمال کنند.

۳-۲-۲ بازسازی حالات چهره

در این دسته، چهرهٔ فرد در تصویر حفظ می‌شود اما حرکات و حالات آن از فرد دیگری الگوبرداری می‌شود. پروژه‌هایی مانند Face2Face با بهره‌گیری از مدل‌سازی سه‌بعدی[۷] و تطبیق کلیدهای چهره[۸]، توانستند حرکات چهره را در زمان واقعی بازسازی کنند[۹]. همچنین NeuralTextures با استفاده از بافت‌های آموخته‌شده[۱۰]، ناحیهٔ دهان را با دقت بالا بازسازی کرد.

۳-۲-۳ روش‌های تشخیص جعل عمیق

در گذشته، روش‌های سنتی تشخیص جعل عمیق متکی بر ویژگی‌هایی همچون نرخ پلک‌زدن، ناهماهنگی نور، یا تحلیل جرم‌شناسی دیجیتال[۱۱] بودند. اما این ویژگی‌ها به‌راحتی توسط الگوریتم‌های سازندهٔ جعل عمیق اصلاح‌پذیر هستند. در نتیجه، تمرکز پژوهش‌ها به سمت مدل‌های یادگیری عمیق، به‌ویژه شبکه‌های عصبی کانولوشنی[۱۲]، سوق یافته است. این مدل‌ها قادرند ویژگی‌های غیرقابل تشخیص برای انسان را از داده‌ها استخراج و تحلیل کنند.

---

[1] Face Swapping
[2] Face Reenactment
[3] Encoder-Decoder
[4] Perceptual and Adversarial Losses
[5] Style Transfer
[6] VGG19 Pre-trained Network
[7] 3D Modeling
[8] Facial Landmark Matching
[9] Real-time Reconstruction
[10] Learned Textures
[11] Digital Forensics
[12] Convolutional Neural Networks (CNN)

۳-۳ مجموعه‌داده

برای آموزش و ارزیابی مدل‌ها، از مجموعه‌دادهٔ FaceForensics++ استفاده شد. این مجموعه شامل ۱۰۰۰ ویدیوی واقعی و ۴۰۰۰ ویدیوی جعلی تولیدشده با روش‌های Deepfakes، Face2Face، FaceSwap و NeuralTextures است. ویدیوها با نرخ فشرده‌سازی ۲۳ برابر[1] و به‌صورت h۲۶۴[2] ذخیره شده‌اند. پردازش داده‌ها در زیرساخت SPARTAN متعلق به دانشگاه ملبورن انجام شده است.

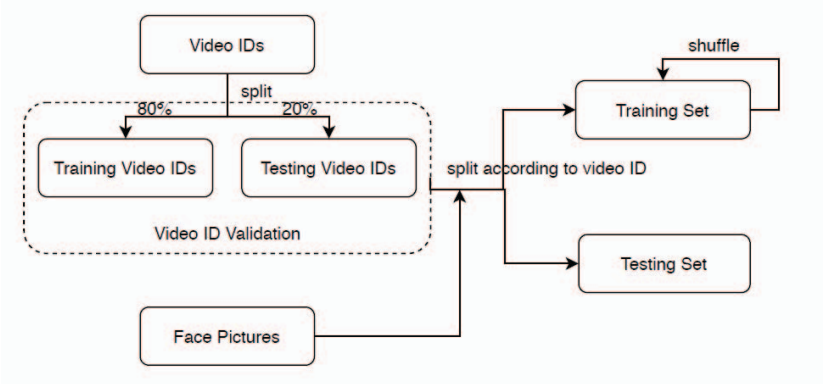
فرایند تقسیم داده‌ها به‌گونه‌ای انجام گرفت که از نشت داده[3] بین بخش آموزش و آزمون جلوگیری شود. فریم‌ها از ویدیوها استخراج، چهره‌ها شناسایی و برش خورده، و نهایتاً به اندازهٔ مورد نیاز مدل‌ها تغییر اندازه داده شدند.

۴-۳ روش‌شناسی

مراحل اصلی این پژوهش عبارتند از:

۱-۴-۳ پیش‌پردازش و استخراج چهره‌ها

با استفاده از OpenCV، ویدیوها به فریم‌های جداگانه تقسیم و سپس چهره‌ها با طبقه‌بند Haar[4] شناسایی شدند. فقط بزرگ‌ترین ناحیهٔ چهره در هر فریم استخراج و به اندازهٔ موردنظر برای مدل‌ها تغییر داده شد: ۲۹۹×۲۹۹ برای Xception و ۲۲۴×۲۲۴ برای MobileNet.



شکل ۳-۱: روندنمای پیش‌پردازش

[1] Compression Ratio of 23:1
[2] H.264 Video Encoding
[3] Data Leakage
[4] Haar Cascade Classifier

۳-۴-۲ معماری مدل‌ها

Xception: یک معماری مبتنی بر کانولوشن‌های جداشونده که همبستگی مکانی و کانالی را به‌طور مستقل مدل می‌کند. این شبکه از ۳۶ لایه کانولوشن تشکیل شده و در این پروژه از نسخهٔ پیاده‌سازی‌شده در Keras استفاده شده است.

MobileNet: معماری سبک‌وزن با استفاده از کانولوشن‌های جداشونده برای استفاده در دستگاه‌های با منابع محدود. این مدل نیز از نسخهٔ آماده در Keras بهره گرفته است.

۳-۴-۳ تنظیم محیط آزمایش

برای اجرای مراحل آموزش، محیطی مجازی با نسخه‌های سازگار از CUDA و TensorFlow در خوشهٔ[1] SPARTAN راه‌اندازی شد.

۳-۴-۴ تنظیمات آموزش

هر مدل برای یک روش خاص جعل عمیق آموزش داده شد. تنظیمات آموزشی به‌شرح زیر است:

- بهینه‌ساز[2]: Adam
- تابع هزینه[3]: Binary Cross-Entropy
- اندازه دسته[4]: ۳۲
- تعداد دوره‌ها[5]: ۱۰
- بارگذاری تصاویر با مولد[6] جهت مدیریت حافظه

۳-۵ نتایج

نتایج نشان دادند که:

- مدل Xception دقتی بین ۹۳٪ تا ۹۸٪ ارائه داد.
- مدل MobileNet دقتی بین ۹۱٪ تا ۹۶٪ داشت.

---
[1] Cluster
[2] Optimizer
[3] Loss function
[4] Batch size
[5] Epochs
[6] Generator

- هر دو مدل در مواجهه با ویدیوهای NeuralTextures عملکرد ضعیف‌تری داشتند.
- دقت مدل‌ها روی روش‌های دیگر جعل عمیق (که با آن‌ها آموزش ندیده بودند) به‌طور قابل توجهی کاهش یافت.

۳-۶ سازوکار رأی‌گیری

برای بهبود عملکرد نهایی، نتایج چهار مدل به‌صورت رأی‌گیری ترکیب شد. اگر هر مدلی بیش از نیمی از فریم‌های یک ویدیو را جعلی تشخیص می‌داد، رأی آن مدل «جعلی» محسوب می‌شد. سپس اگر حتی یک مدل رأی «جعلی» می‌داد، کل ویدیو به‌عنوان جعلی در نظر گرفته می‌شد. این سازوکار به‌ویژه در مواردی که فقط یک مدل قادر به شناسایی جعلی بودن ویدیو بود (مانند ویدیوهای NeuralTextures) مؤثر واقع شد.

۳-۷ بحث

نتایج پژوهش نشان داد که آموزش مدل‌ها به‌صورت اختصاصی برای هر نوع جعل عمیق منجر به افزایش دقت می‌شود، اما توانایی تعمیم مدل‌ها محدود باقی می‌ماند. ترکیب مدل‌ها به‌صورت رأی‌گیری این مشکل را تا حدی برطرف کرده است، ولی همچنان در برابر روش‌های جدید (مانند StyleGAN2) چالش‌هایی باقی می‌ماند.

۳-۸ نتیجه‌گیری

در این پژوهش، یک سامانهٔ تشخیص جعل عمیق مبتنی بر یادگیری عمیق با استفاده از دو معماری Xception و MobileNet طراحی و ارزیابی شد. با آموزش مدل‌ها روی مجموعه‌دادهٔ FaceForensics++ و طراحی سازوکار رأی‌گیری، عملکرد کلی سیستم در شناسایی انواع رایج جعل عمیق بهبود یافت.

پیشنهادهای آینده شامل:

- آموزش مدل برای تکنیک‌های جدید مانند StyleGAN2
- تحلیل ویژگی‌های ناحیه‌ای چهره
- استفاده از توالی فریم‌ها و اطلاعات زمانی
- توسعهٔ رابط کاربری ساده برای استفاده عمومی

فصل ۴

تشخیص فیشینگ[۱] توسط شبکه‌های کانولوشنی[۲]

۴-۱ مقدمه

فیشینگ به عنوان یک حمله سایبری[۳] شناخته می‌شود که در آن مهاجمان سعی می‌کنند کاربران را فریب دهند تا اطلاعات حیاتی و شخصی نظیر جزئیات کارت اعتباری و رمزهای عبور را فاش کنند. این نوع حملات معمولاً از طریق ایمیل‌ها، پیام‌های فوری یا تماس‌های تلفنی آغاز می‌شوند. یکی از روش‌های رایج فیشرها[۴] طراحی وب‌سایت‌های فریبنده است که تقلیدی از وب‌سایت‌های قانونی (مانند PayPal یا eBay) بوده و بر روی دامنه‌های هک شده میزبانی می‌شوند. تشخیص تفاوت بین صفحات وب قانونی و تقلیدی برای چشم انسان دشوار است. با دسترسی کاربر به سایت شبیه‌سازی شده، اطلاعات حیاتی با استفاده از اسکریپت‌ها[۵] به سرقت می‌رود. جرایم فیشینگ هر ساله به دلیل رشد سریع کاربران تجارت الکترونیک در حال افزایش است.

طبق گزارش گروه کاری ضد فیشینگ (APWG)، تعداد کل سایت‌های فیشینگ شناسایی شده در سه‌ماهه اول سال ۲۰۱۹ به ۱۸۰,۷۶۸ مورد رسید که افزایش قابل توجهی نسبت به سه‌ماهه چهارم ۲۰۱۸ (۱۳۸,۳۲۸ مورد) و سه‌ماهه سوم ۲۰۱۸ (۱۵۱,۰۱۴ مورد) نشان می‌دهد. فیشینگ به دلیل خسارات گسترده به صنایع هدف مانند پرداخت و مؤسسات مالی، به یک مشکل جدی تبدیل شده است.

روش‌های سنتی تشخیص فیشینگ، مانند استفاده از لیست‌های سیاه[۶]، در برابر حملات روز صفر[۷] که URL های[۸] فیشینگ جدید هنوز در لیست سیاه ثبت نشده‌اند، کارایی ندارند. بسیاری از تکنیک‌های تشخیص مبتنی بر اکتشافی، ویژگی‌ها را از محتوای صفحات وب و خدمات شخص ثالث استخراج می‌کنند. با این حال، استفاده از خدمات شخص ثالث مانند رتبه صفحه یا معیارهای ترافیک شبکه، می‌تواند زمان‌بر باشد و منجر به کندی

[1] Phishing
[2] CNN: Convulotional Neural Networks
[3] Cyber Attack
[4] phishers
[5] scripts
[6] Black Lists
[7] Zero Days
[8] Uniform Resource Locator

فرآیند طبقه‌بندی شود. تکنیک‌های یادگیری ماشین نیز برای بررسی URL صفحات وب با مجموعه‌های ویژگی دست‌ساز استفاده شده‌اند. با این حال، این روش‌ها به مهندسی ویژگی دستی نیاز دارند و در شناسایی حملات فیشینگ نوظهور کارآمد نیستند.

با پیشرفت‌های اخیر در تکنیک‌های یادگیری عمیق[1]، بسیاری از مدل‌های مبتنی بر یادگیری عمیق نیز برای بهبود عملکرد طبقه‌بندی معرفی شده‌اند. یادگیری عمیق می‌تواند ویژگی‌ها را به صورت خودکار از داده‌های خام استخراج کند و نیازی به دانش قبلی متخصصان امنیت سایبری ندارد. در این مقاله، ما یک مدل مبتنی بر یادگیری عمیق را برای تشخیص URL فیشینگ پیشنهاد می‌کنیم. به طور خاص، ما از شبکه‌های عصبی کانولوشنی در سطح کاراکتر استفاده می‌کنیم. این رویکرد به‌ویژه برای URL ها مفید است، زیرا آن‌ها اغلب حاوی کلمات بی‌معنی هستند و مهاجمان می‌توانند با تغییرات کوچک کاراکتری (مانند "www.icbc.com" به "www.lcbc.com")، URL های فیشینگ را شبیه به URL های قانونی کنند.

مزایای اصلی مدل پیشنهادی ما عبارتند از:

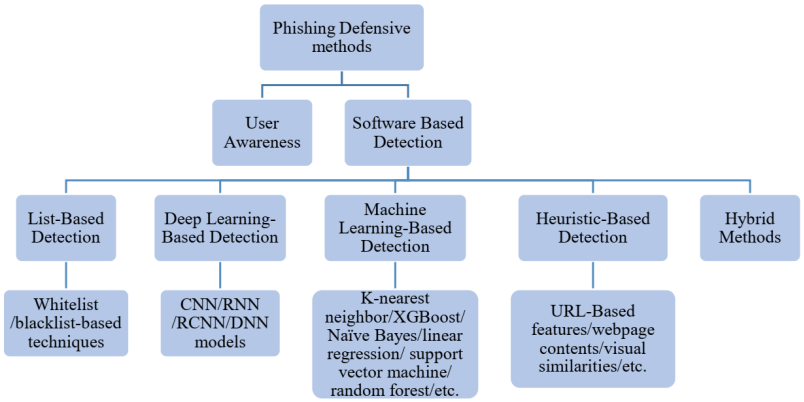
- **عدم وابستگی به سرویس‌های شخص ثالث:** مدل ما فقط از URL وب‌سایت استفاده می‌کند و نیازی به لیست‌های سیاه یا سفید[2]، رتبه صفحه یا معیارهای ترافیک شبکه ندارد. زمان تشخیص برای طبقه‌بندی هر URL تنها ۰/۴۷ میلی‌ثانیه است.
- **استقلال از زبان:** از آنجا که ویژگی‌ها از رشته URL استخراج و بر اساس یک واژگان کاراکتری از پیش تعریف‌شده جاسازی می‌شوند، مدل ما برای وب‌سایت‌ها با محتوای هر زبانی مؤثر است.
- **قابلیت تشخیص وب‌سایت‌های جدید (حملات روز صفر):** به دلیل استفاده از ویژگی‌های جاسازی در سطح کاراکتر، مدل ما می‌تواند به راحتی برای URL های جدید تعمیم یابد و سایت‌های فیشینگ جدید را شناسایی کند که قبلاً به عنوان فیشینگ طبقه‌بندی نشده‌اند.
- **عدم نیاز به کارشناسان امنیت سایبری:** CNN به طور خودکار ویژگی‌ها را برای نمایش URL تشخیص می‌دهد و نیازی به مهندسی ویژگی‌های پیچیده و دستی توسط کارشناسان در طول فرآیند یادگیری ندارد.

۲-۴ کارهای مرتبط

به طور کلی، تشخیص فیشینگ می‌تواند از طریق روش‌های مبتنی بر لیست، یادگیری ماشین، اکتشافی یا یادگیری عمیق انجام شود. با این حال، مشکل فیشینگ آن‌قدر پیچیده است که هیچ راه‌حل قاطعی برای مقابله مؤثر

[1] Deep Learning
[2] White Lists

با همه تهدیدات وجود ندارد؛ بنابراین، اغلب چندین تکنیک برای جلوگیری از حملات خاص به کار گرفته می‌شوند. روش‌های حفاظتی به دو گروه اصلی تقسیم می‌شوند: افزایش دانش کاربر و استفاده از نرم‌افزارهای اضافی (شکل ۳-۱ را ببینید).



شکل ۴-۱: مروری بر تکنیک‌های تشخیص فیشینگ

۴-۲-۱ تشخیص مبتنی بر لیست

تکنیک‌های تشخیص فیشینگ مبتنی بر لیست به دو دسته لیست سفید و لیست سیاه تقسیم می‌شوند. لیست سفید شامل URL ها و آدرس‌های IP قانونی است که برای اعتبارسنجی یک URL مشکوک استفاده می‌شود. تکنیک‌های مبتنی بر لیست سیاه به طور گسترده‌ای در نوار ابزارهای ضد فیشینگ در دسترس عموم مانند Google Safe Browsing استفاده می‌شوند. اگرچه این روش‌ها می‌توانند دقت نسبتاً بالایی داشته باشند، اما نگهداری یک لیست جامع از URL های فیشینگ دشوار است، زیرا URL های جدید هر روز ایجاد می‌شوند. نقطه ضعف اصلی این روش‌ها این است که نمی‌توانند حملات روز صفر را شناسایی کنند.

۴-۲-۲ تشخیص مبتنی بر اکتشافی

اساس تکنیک‌های تشخیص مبتنی بر اکتشافی، ایجاد ویژگی‌های سایت فیشینگ بر اساس بسیاری از ویژگی‌های دست‌ساز، مانند ویژگی‌های مبتنی بر URL، محتوای صفحه وب و شباهت‌های بصری وب‌سایت است. این روش‌ها مانند Cantina و Cantina+ از موتورهای جستجو و قوانین اکتشافی برای تشخیص استفاده می‌کنند. با این حال، برخی از این روش‌ها به زبان انگلیسی محدود بوده یا به دلیل وابستگی به موتورهای جستجو و سرویس‌های شخص ثالث، زمان‌بر هستند.

۴-۲-۳ تشخیص مبتنی بر یادگیری ماشین

برای مقابله با محدودیت‌های زمانی و محاسباتی در جمع‌آوری ویژگی‌ها، در سال‌های اخیر از تکنیک‌های یادگیری ماشین (مانند K-nearest Neighbor، XGBoost، Naïve Bayes، رگرسیون خطی، SVM و جنگل تصادفی[1]) برای تشخیص فیشینگ استفاده شده است. این روش‌ها ویژگی‌ها را از URL ها استخراج کرده و سپس طبقه‌بندی می‌کنند. با این حال، این روش‌ها نمی‌توانند با اطلاعات جدیدی که در مجموعه آموزشی وجود ندارند، مقابله کنند.

۴-۲-۴ تشخیص مبتنی بر یادگیری عمیق

به دلیل موفقیت یادگیری عمیق در پردازش زبان طبیعی، برخی از این تکنیک‌ها (مانند CNN، RNN، RCNN و DNN) اخیراً برای تشخیص فیشینگ به کار گرفته شده‌اند. اگرچه تکنیک‌های یادگیری عمیق به دلیل زمان آموزش گسترده کمتر مورد استفاده قرار گرفته‌اند، اما اغلب دقت بیشتری را ارائه می‌دهند و ویژگی‌ها را به طور خودکار از داده‌های خام استخراج می‌کنند. به عنوان مثال، مدل PDRCNN از ترکیب LSTM و CNN برای استخراج ویژگی‌های جهانی و محلی URL استفاده می‌کند. برخی دیگر از مدل‌ها نیز ویژگی‌های چندبعدی را از URL ها استخراج می‌کنند.

۴-۲-۵ روش‌های ترکیبی

تکنیک‌های تشخیص ترکیبی به ترکیب بیش از یکی از تکنیک‌های قبلی برای دستیابی به عملکرد بهتر در تشخیص سایت‌های فیشینگ متکی هستند. به عنوان مثال، Yang و همکاران از الگوریتم‌های یادگیری عمیق (CNN-LSTM) برای استخراج ویژگی‌های URL و سپس ترکیب آن‌ها با ویژگی‌های آماری URL، کد صفحه وب و متن صفحه وب برای طبقه‌بندی با الگوریتم یادگیری ماشین (XGBoost) استفاده می‌کنند.

۴-۳ روش‌شناسی پیشنهادی

مهاجمان معمولاً URL های فیشینگ را به گونه‌ای ایجاد می‌کنند که شبیه وب‌سایت‌های قانونی به نظر برسند. ایده اصلی این مقاله، تشخیص سریع وب‌سایت‌های فیشینگ با استفاده از ویژگی‌های سبک وزن است. این امر با استخراج ویژگی‌ها تنها از URL و بدون بازدید از محتوای وب‌سایت حاصل می‌شود.

[1] Random Forest

۴-۳-۱ مؤلفه‌های URL

URL شامل مؤلفه‌هایی نظیر پروتکل (مثل http, https)، آدرس IP یا دامنه میزبان، مسیر منبع، و ترکیب نام دامنه سطح دوم و سطح بالا است که به آن domain host گفته می‌شود. شکل URL در فرآیند تشخیص نقش کلیدی دارد.

۴-۳-۲ طراحی مدل

مدل پیشنهادی بر پایه‌ی شبکه عصبی کانولوشنی در سطح کاراکتر است که هدف آن یادگیری ساختار ترتیبی URL ها است.

ویژگی‌های جاسازی کاراکتر

جاسازی در سطح کاراکتر از یک واژگان ۹۵ کاراکتری (شامل حروف بزرگ و کوچک، ارقام و علائم خاص) بهره می‌گیرد. هر کاراکتر توسط رمزگذاری یک‌داغ[1] به بردار m-بعدی نگاشت می‌شود. برای تطابق با معماری CNN، طول همه‌ی URL ها به ۲۰۰ کاراکتر استانداردسازی شده است (padding).

ساختار CNN

ساختار CNN شامل:

- یک لایه جاسازی برای کاهش ابعاد ماتریس sparse
- هفت لایه کانولوشنی با تابع فعال‌سازی ReLU برای استخراج ویژگی
- سه لایه کاملاً متصل جهت تحلیل ویژگی‌های عمیق
- یک لایه خروجی با دو گره[2] و تابع softmax برای تعیین برچسب نهایی URL به عنوان فیشینگ یا معتبر

۴-۳-۳ ویژگی‌های URL

چهار گروه ویژگی مورد استفاده قرار گرفت:

۱. ویژگی‌های سطح کاراکتر با جاسازی

۲. ویژگی‌های TF-IDF سطح کاراکتر

۳. ویژگی‌های دست‌ساز مبتنی بر اجزای URL

[1] Hot-one
[2] Node

۴. ویژگی‌های شمارشگر سطح کاراکتر (count vector)

۴-۳-۴ الگوریتم‌های طبقه‌بندی

برای مقایسه عملکرد، از الگوریتم‌هایی نظیر DNN، Random Forest، Regression، Naïve Bayes Logistic، XGBoost، CNN استفاده شد. معیار اصلی، دقت تشخیص فیشینگ بود.

۴-۴ آزمایش و تحلیل نتایج

۴-۴-۱ مجموعه داده

چهار مجموعه داده مورد استفاده قرار گرفت:

- D1: شامل ۳۱۸،۶۴۲ URL جمع‌آوری شده؛
- D2، D3، D4: مجموعه‌های داده مرجع استخراج‌شده از پژوهش‌های پیشین مانند OpenPhish و PhishTank.

۴-۴-۲ ارزیابی عملکرد بر روی D1

تمام گروه‌های ویژگی (FG1 FG4) روی D1 با طبقه‌بندهای مختلف تست شدند. بهترین نتیجه مربوط به CNN با FG2 بود (دقت: 95.13% :F1).

۴-۴-۳ ارزیابی بر روی D3

FG1 و FG2 به ترتیب با طبقه‌بندهای کلاسیک و مدل‌های عمیق ارزیابی شدند. CNN با FG۲ بهترین عملکرد را داشت (دقت: 95.41%).

۴-۴-۴ ارزیابی بر روی D2

مدل CNN با FG2 دقت 98.58% را بدست آورد که بالاتر از سایر مدل‌های یادگیری عمیق مانند RNN و VDCNN بود.

۴-۴-۵ مقایسه با روش‌های موجود

مدل پیشنهادی در مقایسه با روش‌های Sahingoz، Rao و Le عملکرد بهتری داشت و روی D2 به دقت 98.58% رسید.

۵-۴ نتیجه‌گیری

مدل مبتنی بر CNN با ورودی در سطح کاراکتر، توانست بدون نیاز به ویژگی‌های مهندسی‌شده و صرفاً از طریق تحلیل ساختار ،URL عملکرد مناسبی در تشخیص فیشینگ داشته باشد. مزیت این روش، استقلال از دسترسی به وب و مناسب بودن برای پیاده‌سازی سمت کلاینت است. با این حال، زمان آموزش بالا و ضعف در تشخیص URL های کوتاه یا دارای واژگان معمول از جمله نقاط ضعف آن است.

در آینده، هدف توسعه مدل برای بهره‌گیری از محتوای HTML و کدهای سمت کاربر به منظور تقویت دقت تشخیص خواهد بود.

فصل ۵

نتیجه‌گیری

در این پژوهش، کاربرد یادگیری عمیق در سه حوزه حیاتی شامل تشخیص محتوای جعلی (جعل عمیق)، شناسایی حملات فیشینگ، و تحلیل سیگنال‌های قلبی مورد بررسی و پیاده‌سازی قرار گرفت. برای هر مسئله، معماری‌هایی متناسب انتخاب شد که نشان‌دهنده توانایی بالای شبکه‌های عصبی در استخراج ویژگی‌های مؤثر از داده‌های خام و افزایش دقت تشخیص هستند.

در حوزه تشخیص جعل عمیق، مدل‌های Xception و MobileNet به‌طور مستقل و ترکیبی عملکرد مناسبی در شناسایی ویدیوهای جعلی حاصل از چهار تکنیک رایج نشان دادند. ترکیب این مدل‌ها با سازوکار رأی‌گیری منجر به افزایش دقت کلی و کاهش خطاهای تک‌مدل‌ها شد. آینده این حوزه می‌تواند شامل آموزش مدل‌های جدید برای روش‌های نوین تولید جعل عمیق و تحلیل فریم‌های زمانی به‌جای تصاویر ایستا باشد.

در زمینه مقابله با فیشینگ[1]، استفاده از شبکه عصبی کانولوشنی[2] در سطح کاراکتر توانست بدون تکیه بر خدمات شخص ثالث و مهندسی ویژگی دستی، نشانی‌های[3] جعلی را با دقت مناسب تشخیص دهد. این روش، زمان پاسخ بسیار کمی دارد و برای پیاده‌سازی سمت کلاینت مناسب است. به‌منظور بهبود عملکرد، پیشنهاد می‌شود از محتوای HTML صفحات و کدهای سمت کاربر نیز در طراحی مدل‌های آینده بهره گرفته شود.

در حوزه پزشکی، مدل کانولوشنی توسعه‌یافته توانست ۵ نوع ضربان غیرعادی قلب را با دقت بالا طبقه‌بندی کند و به‌عنوان ابزار پشتیبان تصمیم در سیستم‌های تشخیص یاری‌رسان پزشکی (CAD[4]) مطرح گردد. این سیستم نه‌تنها موجب کاهش خطاهای انسانی می‌شود، بلکه توانایی استفاده در محیط‌های بالینی برای غربالگری سریع سیگنال‌های ECG را نیز دارد. گسترش آینده شامل تحلیل دنباله‌های زمانی ضربان، بررسی عملکرد مدل در داده‌های متعادل‌شده و نویزی و همچنین تفکیک بیماران در سه سطح خطر خواهد بود.

---

[1] Phishing
[2] Convolutional Neural Network
[3] URL
[4] Computer-Aided Diagnosis

در مجموع، نتایج حاصل از این مطالعه چندجانبه نشان می‌دهد که یادگیری عمیق، در کنار معماری‌های مناسب و انتخاب ویژگی‌های درست، می‌تواند راه‌حلی اثربخش و کاربردی برای حل مسائل پیچیده در حوزه‌های امنیتی، زیستی و اجتماعی باشد.

مراجع

[1] A. Lal, P. Kumar, and S. Halder, “Heartbeat classification based on deep convolutional neural network,” in *2023 International Conference on Networking and Communications (ICNWC)*, pp. 1–4, 2023.

[2] D. Pan, L. Sun, R. Wang, X. Zhang, and R. O. Sinnott, “Deepfake detection through deep learning,” in *2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT)*, pp. 134–143, 2020.

[3] A. Aljofey, Q. Jiang, Q. Qu, M. Huang, and J.-P. Niyigena, “An effective phishing detection model based on character level convolutional neural network from url,” *Electronics*, vol. 9, no. 9, 2020.