**شناسایی وب‌سایت فیشینگ بر مبنای شبکه عصبی پیچشی عمیق و یادگیری جمعی الگوریتم جنگل تصادفی**

**چکیده**

حملات فیشینگ، به یکی از بزرگترین و تاثیرگذارترین حملات سایبری تبدیل شده که هر ساله منجر به صدها میلیون دلار زیان و میلیون‌ها نقض داده می‌شود. در حال حاضر، روش‌های ضدفیشینگ نیازمند متخصصانی است تا ویژگی‌های سایت‌های فیشینگ را استخراج نموده و از خدمات ثالث برای شناسایی سایت‌های فیشینگ استفاده کنند. این روش‌ها دارای چند محدودیت است، که یکی از محدودیت‌ این است که استخراج خصوصیات فیشینگ نیازمند تخصص بوده و زمان‌بر است. دوم اینکه، استفاده از خدمات ثالث باعث تاخیر در شناسایی سایت‌های فیشینگ می‌شود. از این رو، این مقاله به تعریف روش یکپارچه شناسایی وب‌سایت فیشینگ بر مبنای شبکه‌ها عصبی پیچشی (CNN)[[1]](#footnote-1) و الگوریتم جنگل تصادفی (RF)[[2]](#footnote-2) می‌پردازد. این روش می‌تواند به پیش‌بینی صحت آدرس‌های اینترنتی، بدون دسترسی به محتوای وب یا استفاده از خدمات ثالث بپردازد. روش پیشنهادی از تکنیک‌های تعبیه کاراکتر برای تبدیل آدرس‌های اینترنتی (URL) به ماتریس‌هایی با اندازه ثابت، استخراج ویژگی در سطوح مختلف با استفاده از مدل‌های شبکه‌ها عصبی پیچشی، دسته‌بندی ویژگی‌های چندسطحی با استفاده از دسته‌بندی‌کننده‌های متعدد الگوریتم جنگل تصادفی، در نهایت، نتایج پیش‌بینی خروجی با استفاده از روش «برنده مالک همه چیز است»، استفاده می‌کند. در مجموعه داده‌های ما، میزان صحت 99.35% با استفاده از مدل پیشنهادی حاصل می‌شود. نرخ صحت 99.26% که براساس داده مبنا بدست می‌آید، بسیار بالاتر از مدل حداکثری کنونی است.

**کلیدواژه**: آدرس اینترنتی، شناسایی فیشینگ، یادگیری عمیق، الگوریتم جنگل تصادفی، یادگیری جمعی

1. Convolutional neural networks [↑](#footnote-ref-1)
2. Random forest (RF) [↑](#footnote-ref-2)