

مقدمه

IoT به عنوان یک شبکه به هم پیوسته و توزیع شده از سیستم های تعبیه شده در نظر گرفته می شود که از طریق تکنولوژی های ارتباطی بی سیم یا سیمی ارتباط برقرار می کنند [۱]. همچنین به عنوان شبکه اشیا فیزیکی یا چیزهایی که با محاسبات، ذخیره سازی و قابلیت های ارتباطی محدود توانمند شده اند و همچنین با الکترونیک (مانند حسگرها و فعال کننده ها)، نرم افزار و اتصال شبکه که این اشیا را قادر به جمع آوری، پردازش و تبادل داده می کند، تعریف می شود .

اشیا در IoT به اشیا زندگی روزمره ما از دستگاه های هوشمند خانگی مانند لامپ هوشمند، آداپتور هوشمند، کنترل هوشمند، یخچال هوشمند، فر هوشمند، AC، سنسور دما، آشکارساز دود، دوربین IP گرفته تا دستگاه های پیچیده تر مانند دستگاه های تشخیص فرکانس رادیویی (RFID)، آشکارسازهای ضربان قلب، شتاب سنج، حسگرها اشاره دارد .

عنوان مقاله ما یادگیری ماشین در امنیت اینترنت اشیا هست که در ادامه به بررسی آن می پردازیم.

تعریف مسئله و هدف اصلی مقاله

داده های تولید شده توسط دستگاه های IoT عظیم هستند و بنابراین، تکنیک های جمع آوری، ذخیره سازی و پردازش داده های سنتی ممکن است در این مقیاس کار نکنند. علاوه بر این، مقدار مطلق داده ها می تواند برای الگوها، رفتارها، پیش بینی ها و ارزیابی نیز مورد استفاده قرار گیرد. علاوه بر این، ناهمگونی داده های تولید شده توسط IoT جبهه دیگری را برای مکانیزم های پردازش داده فعلی ایجاد می کند .

ضرورت تحقیق

برای کنترل ارزش داده‌های ایجاد شده توسط IoT، به مکانیزم‌های جدیدی نیاز است. در این زمینه، یادگیری ماشین **machin learning** می‌تواند به ماشین‌ها و دستگاه‌های هوشمند کمک کند تا دانش مفید را از داده‌های تولید شده توسط انسان استنباط کنند. همچنین می‌تواند به عنوان توانایی یک دستگاه هوشمند برای تغییر یا خودکار کردن موقعیت یا رفتار براساس دانش که به عنوان یک بخش ضروری برای یک راه‌حل IoT در نظر گرفته می‌شود، تعریف شود. این حقیقت که اینترنت اینترنت از فن‌آوری‌های توانا مانند شبکه سازی تعریف شده نرم‌افزار (SDN)، محاسبات ابری (CC) و محاسبات مه استفاده می‌کند نیز چشم‌انداز تهدیدها برای مهاجمان را افزایش می‌دهد.

ML می‌تواند به ماشین‌ها و دستگاه‌های هوشمند کمک کند تا دانش مفید حاصل از داده‌های تولید شده انسان را استنباط کنند. همچنین می‌تواند به عنوان توانایی یک دستگاه هوشمند برای تغییر یا خودکار کردن وضعیت یا رفتار مبتنی بر دانش تعریف شود که به عنوان یک بخش ضروری برای راه‌حل IoT محسوب می‌شود.

شاخه های حل پیشین:

شاخه حل جاری

ML می‌تواند در اینترنت برای ارائه خدمات هوشمند به کار رود. با این حال، در این مقاله، ما بر روی برنامه‌های ML در ارائه خدمات امنیتی و حریم خصوصی به شبکه‌های اینترنت متمرکز می‌شویم. امنیت و حریم خصوصی دو عامل اصلی در درک تجاری خدمات و برنامه‌های اینترنت IoT هستند. اینترنت کنونی یک زمین‌بازی فریبنده برای حملات امنیتی است که از **hacks** ساده تا سطح شرکت وجود دارد. شکاف‌های امنیتی هماهنگ که به طور زیان‌آور بر صنایع مختلف از قبیل مراقبت‌های بهداشتی و کسب‌وکار تاثیر می‌گذارد. محدودیت‌های تجهیزات اینترنت و محیطی که آن‌ها در آن فعالیت می‌کنند، چالش‌های اضافی برای امنیت هر دو برنامه و وسایل ایجاد می‌کنند.

Gaps in the Existing Security Solution for IoT Networks

شکاف‌ها در راه‌حل امنیت موجود برای شبکه‌های IoT. برای تحقق موفقیت آمیز اینترنت IoT مهم است که ریشه‌های مسایل امنیت و حریم خصوصی را آنالیز کنیم. به طور دقیق‌تر واژه اینترنت IoT از فن‌آوری‌های موجود به پایین پرتاب شده‌است و بنابراین ضروری است که بدانیم آیا چالش‌های امنیتی در اینترنت اشیا جدید یا ترکیبی از ارث از تکنولوژی‌های قدیمی هستند یا خیر.

نگرانی اولیه از اینترنت اشیا - محدودیت‌های مربوط به منابع است که مانع محدودیت‌ها هستند. علاوه بر این، راه‌حل‌هایی برای مسایل امنیت و حریم خصوصی در اینترنت به طراحی لایه متقاطع و الگوریتم‌های بهینه نیاز دارند. به عنوان مثال به دلیل محدودیت‌های محاسباتی، وسایل اینترنت IoT ممکن است به نژاد جدیدی از رمزنگاری بهینه نیاز داشته باشند.

D. Machine Learning: A Solution to IoT Security Challenges

یادگیری ماشینی: راه‌حلی برای چالش‌های امنیتی IoT. یادگیری ماشینی به روش‌های هوشمندی اشاره دارد که برای بهینه‌سازی معیارهای عملکرد با استفاده از داده‌های نمونه یا تجربیات گذشته از طریق یادگیری مورد استفاده قرار می‌گیرند ML همچنین توانایی (برای دستگاه‌های هوشمند) برای یادگیری بدون برنامه‌ریزی صریح را ممکن می‌سازد.

یادگیری ماشینی زمانی به کار می‌رود که تخصص انسان وجود نداشته باشد و یا نمی‌تواند مورد استفاده قرار گیرد، مانند هدایت یک مکان خصمانه که در آن انسان‌ها قادر به استفاده از تخصص خود نیستند، برای مثال رباتیک، تشخیص گفتار و غیره. همچنین در شرایطی که در آن انسان‌ها قادر به استفاده از تخصص خود نیستند. علاوه بر این، در سیستم‌های هوشمند عملی استفاده می‌شود، به عنوان مثال، گوگل از ML برای تجزیه و تحلیل تهدیدات در برابر نقاط پایانی تلفن همراه و برنامه‌های کاربردی در حال اجرا بر روی اندروید استفاده می‌کند. همچنین از آن برای شناسایی و حذف بدافزار از هندس‌های آلوده استفاده می‌شود. بنابراین، تکنیک‌های ML نیاز به راهنمایی و اصلاح برای مدل دارد اگر پیش‌بینی نادرست ایجاد شود. برعکس، در یادگیری عمیق (DL)، نوع جدیدی از ML، این مدل می‌تواند دقت پیش‌بینی را خودش تعیین کند. تولید مقدار زیادی از داده‌ها که توسط روش‌های ML و DL برای آوردن هوش به سیستم‌ها مورد نیاز است. علاوه بر این، سودمندی داده‌های تولید شده توسط IoT بهتر با تکنیک‌های ML و DL استفاده می‌شود که

سیستم‌های IoT را قادر به تصمیم‌گیری‌های آگاهانه و هوشمند می‌سازد ML و DL عمدتاً برای امنیت، حریم خصوصی، تشخیص حمله و تجزیه و تحلیل malware استفاده می‌شوند .

برخی از برنامه‌های مربوط به امنیت دنیای واقعی ML به شرح زیر هستند:

- ✓ تشخیص چهره در پزشکی قانونی: ژست، نورپردازی، انسداد (عینک، ریش)، آرایش، سبک مو و غیره .
- ✓ تشخیص هویت برای رمزنگاری امنیتی: سبک‌های مختلف نگارش .
- ✓ شناسایی کد مخرب: شناسایی کد مخرب در برنامه‌ها و نرم‌افزار.
- ✓ انکار توزیع شده سرویس (DDoS) تشخیص: حملات DDoS به زیرساخت از طریق تجزیه و تحلیل رفتار.

ادبیات تحقیق حاضر امنیت در IoT را با بررسی راه‌حل‌های سنتی موجود و راه‌حل‌های ارائه‌شده از طریق تکنولوژی‌های نوظهور جدید پوشش می‌دهد. اگرچه ML و DL در چند نظرسنجی پوشش داده شده‌اند اما اطلاعات کلی در مورد کاربرد جامع ML و DL کمیاب است. برای پر کردن شکاف‌ها، ما یک بررسی جامع از تکنیک‌های ML و DL مورد استفاده در امنیت IoT انجام می‌دهیم .

حملات فیزیکی

در حملات فیزیکی، مهاجمان به دستگاه‌ها دسترسی مستقیم دارند و جنبه‌های مختلف دستگاه‌ها را دستکاری می‌کنند. برای دسترسی به دستگاه‌های فیزیکی، مهندسی اجتماعی یکی از برجسته‌ترین روش‌ها است که در آن حمله‌کنندگان به دستگاه‌ها دسترسی دارند و حمله واقعی را انجام می‌دهند که در محدوده آسیب فیزیکی به دستگاه تا استراق‌سمع، کانال‌های جانبی، و دیگر حملات مرتبط قرار دارد

علاوه بر این، برای شروع حملات فیزیکی، حمله‌کنندگان باید در نزدیکی دستگاه‌ها / سخت‌افزار با مقاصد مختلف مانند تخریب فیزیکی سخت‌افزار، محدود کردن طول عمر آن، به خطر انداختن مکانیزم ارتباطی، دستکاری در منبع انرژی و غیره باشند. همچنین شایان‌ذکر است که حملات فیزیکی ممکن است سنگ بنای سایر حملات باشند، به عنوان مثال از کار انداختن زنگ خطر در خانه می‌تواند منجر به سرقت یا سایر آسیب‌های مرتبط در محیط خانه هوشمند شود. تخریب گره مخرب به شبکه نیز می‌تواند باعث حمله انسان در میانه شود که به مهاجم این امکان را می‌دهد تا امتیازات را افزایش داده و حملات دیگری را آغاز کند. علاوه بر این، چنین دستکاری در دستگاه‌ها نیز ممکن است مهاجمان را قادر به ایجاد تغییراتی در جداول مسیریابی و کلیدهای امنیتی کند که بر ارتباط با لایه‌های بالاتر تاثیر خواهد گذاشت. دیگر حملات فیزیکی شامل مسدود کردن فرکانس‌های رادیویی است که ارتباط در محیط IoT را انکار می‌کند.

مسائل فیزیکی (PHY) و امنیت لایه پیوند

IoT فن آوری های ارتباطی مختلف را در لایه های پایین تر پشته پروتکل TCP / IP ترکیب می کند و بنابراین یک شبکه ناهمگن پیچیده فراهم می کند. این تکنولوژی ها عبارتند از:

اما محدود به Zig های، WSN، MANET، WiFi، RFID، NFC و غیره نمی شود و علاوه بر آن این تکنولوژی ها مسائل امنیتی خاص خود را دارند. همانطور که قبلا ذکر شد، ناهمگونی در لایه فیزیکی IoT معرفی می شود و سپس اصلاحات مختلفی در لایه لینک داده ایجاد می شود، برای مثال طراحی کانال خاص و غیره، بسته به تکنولوژی لایه فیزیکی اساسی. برای این منظور، مکانیزم های امنیتی IoT باید ناهمگونی در لایه پیوند داده و فیزیکی را در بر بگیرد. علاوه بر این، تشخیص اختلال در سخت افزار نیز از اهمیت بالایی برخوردار است و باید برای جلوگیری از ناهنجاری در لایه بالایی کنترل شود. شایان ذکر است که حتی برای تشخیص نفوذ در لایه های بالایی، به عنوان مثال حمله مسیریابی، بردارهای حمله بسیاری وجود دارند برای این منظور، IEEE یک استاندارد شناخته شده به نام IEEE ۸۰۲،۱۱ ۱۵،۴ را اجباری کرده است تا اجازه دهد دستگاه محدود به طور موثر ارتباط برقرار کند

حملات لایه شبکه

در سطح شبکه، حملات با هدف مسیریابی، داده ها و تجزیه و تحلیل ترافیک، جعل و راه اندازی حمله با دخالت انسان انجام می شوند. علاوه بر این، حملات سیبیل نیز در لایه شبکه امکان پذیر است که در آن هویت های جعلی / هویت های سیبیل برای ایجاد توهم در شبکه مورد استفاده قرار می گیرند

از سوی دیگر، نفوذ از طریق ابزارهای مختلف، راهی را برای مهاجم به سیستمی فراهم می کند که در آن حمله کنندگان می توانند حملات دیگری را انجام دهند و بنابراین، امنیت شبکه برای مهار حملات در مراحل اولیه ضروری است. در لایه شبکه، مهاجم همچنین می تواند یک گره در معرض خطر را برای استفاده از آن به عنوان گره ارسال جعلی و ایجاد حفره سینکویل نفوذ کند.

حملات لایه حمل و نقل

لایه حمل و نقل مسئول فرآیند تحویل است که در آن پروتکل‌های حمل و نقل امکان تبادل داده‌ها را فراهم می‌کنند. در زمینه IoT، مسائل امنیتی لایه حمل و نقل سنتی همچنان ادامه دارند.

جدی‌ترین حمله در این لایه، حمله انکار سرویس است که شبکه را خفه می‌کند و منجر به انکار سرویس به برنامه‌های کاربردی می‌شود. لازم به ذکر است که به دلیل ماهیت IoT، پروتکل‌های TCP و UDP سنتی با دستگاه‌های منابع محدود مقیاس بندی نمی‌شوند، و بنابراین نسخه‌های سبک پروتکل‌های حمل و نقل در مقالات پیشنهاد شده‌اند

لازم به ذکر است که به دلیل ماهیت IoT، پروتکل‌های TCP و UDP سنتی با دستگاه‌های منابع محدود مقیاس بندی نمی‌شوند، و بنابراین نسخه‌های سبک پروتکل‌های حمل و نقل در مقالات پیشنهاد شده‌اند با این حال، امنیت این پروتکل‌ها برای کاهش حملات DDOS و DoS در IoT از اهمیت ویژه‌ای برخوردار است.

حملات لایه برنامه

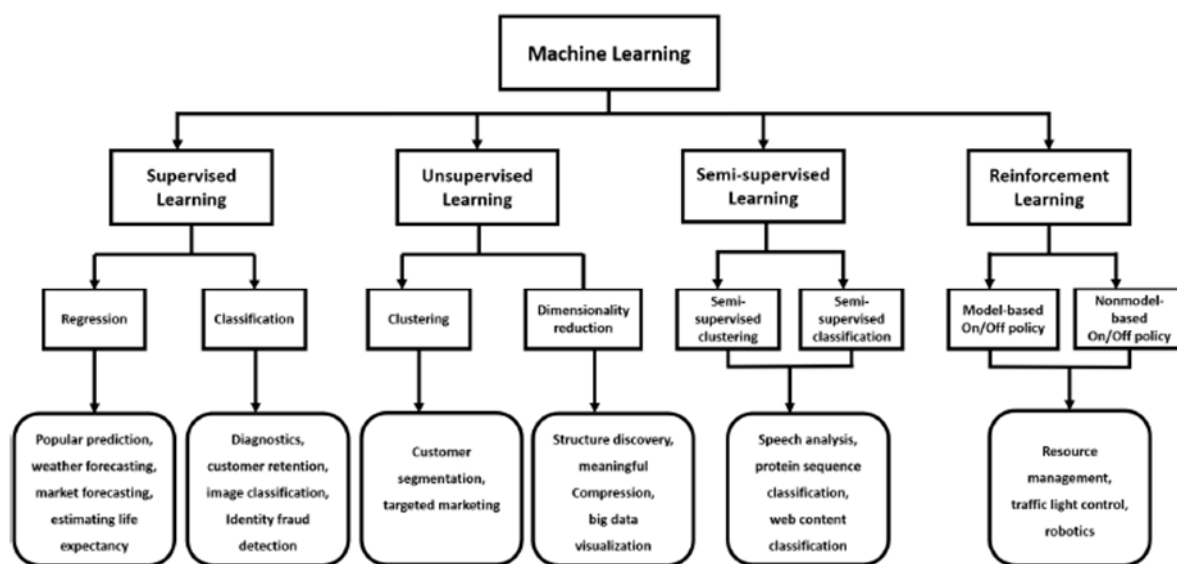
برنامه‌های کاربردی IoT اهداف نسبتاً سودمندی برای حمله‌کنندگان هستند، زیرا راه‌اندازی حملات در سطح برنامه‌ها نسبتاً آسان است. برخی از حملات شناخته‌شده شامل حملات سرریز بافر، حملات بدافزار، انکار خدمات، فیشینگ، بهره‌برداری از آسیب‌پذیری‌های WebApp، حملات رمزنگاری، حملات کانال جانبی، و حملات انسان در موضوع هستند اما محدود به آن‌ها نمی‌شوند. تکنیک‌های موجود برای کاهش مکانیزم سرریز بافر شامل تجزیه و تحلیل کد استاتیک و دینامیک و دیگر مکانیزم‌های پیچیده مانند اشکال‌زدایی نمادین است. با این حال، به دلیل محدودیت منابع نمی‌توان از این تکنیک‌ها در IoT استفاده کرد. برنامه‌های کاربردی IoT همچنین در اثر سرریز بافر در معرض تزریق کد مخرب قرار دارند و سایر آسیب‌پذیری‌ها مانند تزریق SQL، کدنویسی در محل متقابل، ارجاع به شی و ... نیز در معرض

توزیع کد مخرب قرار دارند. پروژه امنیت برنامه وب باز (OWASP ۱۰) آسیب پذیری برتر را شناسایی کرد که باعث حملات مختلف به برنامه ها می شوند .

IOT SECURITY AND MACHINE LEARNING

الگوریتم های یادگیری ماشین پایه

الگوریتم های ML می توانند به چهار دسته طبقه بندی شوند؛ الگوریتم های نظارت شده، بدون نظارت، نیمه نظارت شده و یادگیری تقویتی



یادگیری بدون نظارت: در یادگیری بدون نظارت، محیط تنها ورودی هایی را بدون اهداف مطلوب فراهم می کند. به داده های برجسته دار نیاز ندارد و می تواند تشابه بین داده های برجسته دار نشده را بررسی کند و داده ها را به گروه های مختلف طبقه بندی کند .

یادگیری با نظارت و تکنیک‌های بدون نظارت عمدتاً بر مشکلات تجزیه و تحلیل داده‌ها تمرکز دارند در حالی که یادگیری تقویتی برای مقایسه و مشکلات تصمیم‌گیری ترجیح داده می‌شود. این طبقه‌بندی و انتخاب تکنیک ML به ماهیت داده‌های موجود بستگی دارد .

الگوریتم‌هایی مانند SVR ، شبکه‌های عصبی، و نایو بیس برای مدل‌سازی پیش‌بینی استفاده می‌شوند.

یادگیری نیمه نظارت شده: در دو نوع قبلی، یا هیچ لیبل برای تمام مشاهدات در مجموعه داده وجود ندارد یا لیبل برای تمام مشاهدات وجود دارد . در بسیاری از موقعیت‌های عملی، هزینه برچسب زدن بسیار بالا است، زیرا برای انجام این کار به متخصصان ماهر انسانی نیاز است. بنابراین، در غیاب برچسب‌ها در اکثر مشاهدات، اما در تعداد کمی از آن‌ها، الگوریتم‌های نیم عمقی بهترین گزینه برای ساخت مدل هستند.

یادگیری تقویتی: در یادگیری تقویتی (RL)، هیچ نتیجه خاصی تعریف نشده است و نماینده پس از تعامل با محیط از بازخورد یاد می‌گیرد . برخی اقدامات را انجام می‌دهد و براساس پاداش به دست آمده تصمیم‌گیری می‌کند . یک نماینده را می‌توان برای انجام اقدامات خوب یا مجازات برای اقدامات بد و استفاده از معیارهای بازخورد برای به حداکثر رساندن پاداش‌های بلند مدت پاداش داد . همچنین انتخاب تابع پاداش مناسب بسیار مهم است زیرا موفقیت و شکست عامل به پاداش کل انباشته بستگی دارد

در ادامه، همان طور که در جدول ۳ نشان داده شده است، الگوریتم‌های مختلف ML با تمرکز بر مسائل امنیتی و حریم خصوصی اساسی در شبکه‌های IOT را مورد بحث قرار می‌دهیم . به طور دقیق‌تر، ما احراز هویت، تشخیص و کاهش حمله، حملات انکار سرویس توزیع شده (DDoS)،

تشخیص ناهنجاری و نفوذ، و تجزیه و تحلیل malware را در نظر می‌گیریم .

Machine Learning Algorithm	Description
Naive Bayes	It is the classification algorithm used with binary and multi-class environment. It is named as "Naive", as over-simplified assumptions are made for the calculation of probabilities for specific hypothesis. All the attributes are assumed to be conditionally independent instead of calculating the actual values [74].
K-Nearest Neighbour	It is simple and effective supervised learning algorithm and is used for associating new data points to the existing similar points by searching through the available dataset. The model is trained and grouped according to some criteria and incoming data is checked for similarity within K neighbours [75].
K-Means Algorithm	The most commonly used well know technique is K -means clustering algorithm belonging to the unsupervised category of ML family. K -Means clustering is used to classify or group devices based on attributes or parameters, into K number of groups, where K is a positive integer number and its value has to be known for the algorithm to work [76].
Random Forest and Decision Tree (DT)	It is a supervised learning method. It defines a model by implementing certain rules inferring from the data features. Afterwards, this model is used to predict the value of new targeted variable. Decision tree is used in classification and as well as regression problems. Essentially, these trees are used to split dataset into several branches based on certain rules [77].
Support Vector Machines (SVM)	SVM is a supervised ML algorithm with low computational complexity, used for classification and regression. It has the ability to work with binary as well as with multi-class environments [78], [79]. It classifies input data into n dimensional space and draws $n - 1$ hyperplane to divide the entire data points into groups.
Recurrent Neural Networks (RNN)	This is a supervised learning algorithm used to develop a cascaded chain of decision units for solving the complex problems [80]. It essentially constructs network with certain number of inputs to trigger outputs. Various types of neural networks have been proposed in the literature, e.g. Multi-Layer Perceptron (MLP), Convolutional Neural Networks (CNNs), and Recurrent Neural Networks (RNNs) [81], [82], [83].
Principal Component Analysis	It is an unsupervised ML algorithm and is multivariate technique for data compression. It performs dimensionality reduction in large data sets and extracts useful information in the form of set comprised of orthogonal variables known as "principal components". These components are organized in an increasing order of variance where first component is associated with highest variance of the data and it continues to the last. The least variance components having least information can be discarded [84].
Q-Learning	It is used for resource scheduling in spectrum management in addition to security in IoT. Q-learning belongs to reinforcement learning (RL) class of the ML. In RL, an agent learns by trial and error that how its actions effect the environment. It estimates the reward after each action and moves to the new state accordingly [76]. It will get reward for good action and penalty for bad actions.
Deep Learning	It is essentially a feed Forward Neural Network (NN) in which each neuron is connected to another layer and no connection exists within the layer. The term deep learning refers to multiple layers holding multiple levels of perception such that each layer receives input from the previous layer and feeds the result to following layer [85].

ML برای ایجاد مدل‌ها استفاده می‌شود، که برای طراحی، تست، و آموزش مجموعه داده‌ها استفاده می‌شود. این الگوریتم‌های ML برای تشخیص الگوها و شباهت‌های ممکن در مجموعه داده‌های بزرگ استفاده می‌شوند و می‌توانند در داده‌های جدید پیش‌بینی کنند .

ارزیابی و بررسی

در این مطالعه ما یادگیری ماشین در امنیت اینترنت اشیا را مورد بحث و بررسی قرار دادیم همچنین درباره ی انواع حملات مختلف در لایه های مختلف را مورد بحث و بررسی قرار دادیم.همچنین ما در مورد چالش‌های امنیتی و حریم خصوصی در IoT ، بردارهای حمله، و الزامات امنیتی بحث کرده‌ایم.

جمع بندی

امنیت IoT و حریم خصوصی از اهمیت بالایی برخوردار هستند و نقشی محوری در تجاری سازی فن آوری IoT ایفا می کنند. راه حل های امنیتی و حریم خصوصی سنتی از تعدادی از مسائل رنج می برند که به ماهیت پویای شبکه های IoT مربوط می شوند. تکنیک های ML و DL و DRL به طور خاص تر می توانند برای قادر ساختن دستگاه های IoT برای انطباق با محیط پویای خود مورد استفاده قرار گیرند. این تکنیک های یادگیری می توانند عملیات خود سازماندهی را پشتیبانی کنند و همچنین عملکرد کلی سیستم را با یادگیری و پردازش اطلاعات آماری از محیط بهینه کنند) برای مثال کاربران انسانی و دستگاه های (IoT). این تکنیک های یادگیری ذاتا توزیع شده اند و نیازی به ارتباط متمرکز بین دستگاه و کنترلر ندارند. با این حال، مجموعه داده های مورد نیاز برای الگوریتم های ML و DL هنوز هم کمیاب هستند، که باعث تعیین معیار می شوند

کارهار آینده

اندازه برای همه مناسب نیست: تکنیک های DL بسیار خاص کاربرد هستند که در آن یک مدل آموزش دیده برای حل یک مساله ممکن است نتواند برای مساله دیگری در دامنه مشابه به خوبی عمل کند. این مدل ها معمولا نیاز به آموزش مجدد با داده های مربوطه دارند تا برای مشکلات مشابه دیگر مورد استفاده قرار گیرند.

شبکه‌های عصبی عمیق مانند یک جعبه سیاه عمل می‌کنند، همانطور که نمی‌دانیم چگونه هر مدل DL با دستکاری پارامترها و داده‌های ورودی به نتیجه می‌رسد. همانند مغز انسان، غیر ممکن است که بدانیم مغز چگونه کار می‌کند و راه حل به دست آمده نتیجه نورون‌های جاسازی شده در لایه‌های به هم پیوسته پیچیده است. بنابراین، پیش‌بینی لایه دقیق برای شکست احتمالی دشوار است و از این رو برای کاربردهایی که صحت سنجی در آنها مهم است نامناسب می‌شود.