Journal of Information and Communication Technology

Volume 17, Issue 66-67, Autumn and Winter 2025, pp. 69-86

A Self-supervised Sensors' Anomaly Detection Scheme in Industrial Control Systems Based on Ensemble Deep Learning

Armin Salimi-Badr[1], Athena Abdi[2*], Afshin Souzani[3]

[1] Department of Computer Science and Engineering, Shahid Beheshti University, Tehran, Iran
[2] Department of Computer Engineering, K.N.Toosi University of Technology, Tehran, Iran
[3] ICT Security, ICT Research Institute, Tehran, Iran

Received: 17 August 2024, Revised: 14 May 2025, Accepted: 02 September 2025
Paper type: Research

Abstract

In this paper, a self-supervised one-class sensors' anomaly detection approach based on ensemble deep learning for industrial control systems (ICS). ICSs were traditionally thoroughly separated from the internet and enterprise networks. However, technological advancements have allowed them to connect to the internet to improve the performance of their remote control. Although this connection provides many advantages for ICS, it causes vulnerabilities against cyber-attacks. Anomaly detection is a prominent process to mitigate faults along with the cyber-attacks. In this context, several anomaly detection methods are proposed that are mainly based on local and short-term analyses of the data. The proposed method employs an ensemble deep learning scheme based on combining various temporal, spatial, local, and global characteristics of the individual detection agents during the prediction process, simultaneously. The detection agents have a homogenous workflow with heterogenous prediction structures to consider various characteristics of the input signal. The considered structures of the proposed detection method are based on Long-Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and fully connected (Dense) encoder-decoder schemes. Each unit calculates a normal degree based on the prediction and reconstruction error for the input signal. The normal degree is calculated based on the statistics of the encoder-decoder error considering the correlations among spatial and temporal features. These structures execute in parallel and send their results to a weighted threshold gate voter to determine the final output. As a result, the combination of critical characteristics of the data is reflected in the final output, and its performance is enhanced in comparison to each detection agent. To evaluate the efficiency of the proposed method, several experiments on a simulated ICS are performed and the results demonstrate an average improvement of 14% in precision compared to related studies.

Keywords: Industrial Control System, Anomaly Detection, Deep Learning, Ensemble Learning, Correlation.

[*] Corresponding Author's email: a_abdi@kntu.ac.ir

دوفصلنامه
فناوری اطلاعات و ارتباطات ایران

سال هفدهم، شماره‌های ۶۶ و ۶۷، پاییز و زمستان ۱۴۰۴، صفحه ۶۹ الی ۸۶

ارائه یک روش خودنظارتی کشف ناهنجاری حسگرها در سامانه‌های کنترل صنعتی مبتنی بر یادگیری عمیق گروه‌محور

آرمین سلیمی‌بدر[۱]، آتنا عبدی[۲*]، افشین سوزنی[۳]

[۱] دانشکده مهندسی و علوم کامپیوتر، دانشگاه شهید بهشتی، تهران، ایران

[۲] دانشکده مهندسی کامپیوتر، دانشگاه صنعتی خواجه‌نصیرالدین طوسی، تهران، ایران

[۳] پژوهشکده امنیت، پژوهشگاه ارتباطات و فناوری اطلاعات، تهران، ایران

تاریخ دریافت: ۱۴۰۳/۰۵/۲۷ تاریخ بازبینی: ۱۴۰۴/۰۲/۲۴ تاریخ پذیرش: ۱۴۰۴/۰۶/۱۱

نوع مقاله: پژوهشی

چکیده

در این مقاله رویکرد خودنظارتی کشف ناهنجاری داده‌های حسگر مبتنی بر یادگیری عمیق گروه‌محور و یک دسته‌ای در کاربردهای کنترل صنعتی ارائه شده است. سامانه‌های کنترل صنعتی به‌صورت سنتی از شبکه منفک بوده‌اند اما با پیشرفت فناوری و به منظور افزایش کارایی در کنترل از راه دور به اینترنت متصل شده‌اند. این اتصال در کنار مزایای زیاد منجر به افزایش آسیب‌پذیری در برابر حملات سایبری شده است. کشف ناهنجاری یکی از فرایندهای شناخته شده مواجهه با اشکالات و حملات سایبری می‌باشد. بدین منظور رویکردهای کشف ناهنجاری متعددی ارائه شده‌اند که عموما مبتنی بر تحلیل محلی و کوتاه‌مدت داده‌ها می‌باشند. روش پیشنهادی با بکارگیری رویکردی گروه‌محور متشکل از چندین عامل تشخیص مبتنی بر روش‌های یادگیری عمیق مختلف، ویژگی‌های زمانی، مکانی، محلی و سراسری داده در فرایند پیش‌بینی را به‌صورت همزمان درنظر می‌گیرد. عوامل تشخیص دارای روند کاری همگن و ساختار پیش‌بینی ناهمگن می‌باشند تا هر یک بر اساس ویژگی ساختار مورد استفاده شاخصه‌های متفاوتی از سیگنال ورودی را مورد بررسی قرار دهند. ساختارهای درنظر گرفته شده در روش پیشنهادی برپایه حافظه بلندکوتاه مدت، شبکه عصبی پیچشی و شبکه عصبی تمام‌متصل می‌باشد. هر واحد تشخیص درجه هنجاربودن برای سیگنال ورودی بر مبنای تحلیل آماری خطای پیش‌بینی ساختار کدگذار-کدگشای خود با در نظر گرفتن همبستگی زمانی-مکانی بین ویژگی‌ها محاسبه می‌کند. این ساختارها به‌صورت موازی اجرا شده و رای‌گیری وزن‌دار آستانه‌محور با هدف مشخص کردن نتیجه نهایی از اجماع روش‌ها بر خروجی‌های عوامل تشخیص اعمال می‌شود. به‌این ترتیب، ترکیبی از ویژگی‌های اساسی داده در نتیجه نهایی منعکس می‌شوند و کارایی آن نسبت به تحلیل هر یک از عوامل تشخیص بهبود می‌یابد. به‌منظور بررسی قابلیت‌های روش پیشنهادی، آزمایش‌های متعددی بر بستر سامانه کنترل صنعتی شبیه‌سازی شده انجام گرفته است و نتایج بهبود دقت ۱۴٪ به‌طور میانگین نسبت به رویکردهای پیشین را نشان می‌دهد.

کلیدواژگان: سامانه‌های کنترل صنعتی، کشف ناهنجاری، یادگیری عمیق، یادگیری گروه‌محور، همبستگی.

* رایانامهٔ نویسنده مسؤول: a_abdi@kntu.ac.ir

۱- مقدمه

سامانه‌های کنترل صنعتی (ICS) با هدف پایش و نظارت زیرساخت‌های حیاتی صنعتی مانند تولید انرژی، تصفیه آب، صنایع هوشمند، پالایشگاه‌ها و غیره بکارگرفته می‌شوند. این سامانه‌ها از سه بخش اساسی شبکه تجاری سازمان، شبکه کنترلی و شبکه میدان تشکیل شده‌اند. شبکه تجاری شامل سرورهای ارتباطی متصل به اینترنت، شبکه کنترلی متشکل از سامانه مرکزی پایش و کنترل و شبکه میدان متشکل از حسگرها، محرکه‌ها و کنترلرهای منطقی برنامه‌پذیر می‌باشند[۳-۱]. اتصال سامانه‌های کنترل صنعتی به بستر اینترنت با هدف افزایش بهره‌وری و کاهش هزینه پایش و مدیریت مطرح شده است. این اتصال در کنار مزایای زیاد منجر به افزایش آسیب‌پذیری این سامانه‌ها بدلیل حملات سایبری می‌شود. این آسیب‌پذیری‌ها با پیشرفت فناوری افزایش یافته و منجر به ایجاد حملات متخاصمانه به این سامانه‌ها و کاهش امنیت آن‌ها در کاربردهای حیاتی می‌شود[۶-۴].

یکی از موثرترین رویکردهای مواجهه با چالش امنیتی سامانه‌های کنترل صنعتی، بکارگیری سامانه‌های تشخیص نفوذ (IDS) می‌باشد. سامانه تشخیص نفوذ به‌طور گسترده به بررسی فعالیت‌های شبکه بر اساس تحلیل لاگ‌ها و ترافیک شبکه می‌پردازد و در صورت تشخیص نفوذ بالفعل یا بالقوه، پیام هشدار ایجاد می‌کند. این ابزار معمولاً مبتنی بر تشخیص فعالیت مشکوک در قالب فعالیت‌های ناهنجار اقدام به تشخیص می‌کند. سامانه‌های تشخیص نفوذ در دو دسته مبتنی بر امضا و مبتنی بر رفتار دسته‌بندی می‌شوند[۳ و ۷]. سامانه‌های تشخیص نفوذ مبتنی بر امضا براساس پایگاه دانش مشخصی از حملات پیشین، حملات را به‌صورت ایستا شناسایی می‌کنند. درصورت بروز حمله جدید، این نوع از سامانه‌ها قابلیت تشخیص حمله را ندارد. از سوی دیگر رویکردهای مبتنی بر رفتار، سعی در استخراج مدل رفتار هنجار سیستم و شناسایی رفتارهای ناهنجار از طریق دسته‌بندی رخدادها را دارند. درنتیجه، توانایی استخراج مدل مناسب در سامانه‌های تشخیص نفوذ مبتنی بر ناهنجاری بسیار حائز اهمیت می‌باشد[۱۴-۸]. این واحد در واحد کنترل نظارتی و جمع‌آوری داده (SCADA) سامانه‌های کنترل صنعتی تعبیه می‌شود که قابلیت پردازشی مناسبی را فراهم می‌آورد.

به‌منظور استخراج رفتار هنجار سیستم لازم است تحلیل مناسبی بر داده‌ها صورت گیرد. باتوجه به حجم زیاد داده‌ها و تنوع رفتاری موجود، رویکردهای مبتنی بر یادگیری ماشین عملکرد توسعه‌یافته و مناسبی در این زمینه دارند. در این رویکردها آموزش به‌صورت خودکار از داده انجام می‌شود و الگوهای موجود شناسایی می‌شوند. از مهم‌ترین مزایای روش‌های مبتنی بر یادگیری می‌توان به هزینه کم عملیات، عدم وابستگی به درخواست‌های کاربردان، بهبود کارایی، کاهش زمان محاسبات، ایجاد دید بهتر از محیط و استخراج ویژگی‌های با اهمیت در عملیات اشاره کرد [۱۷-۱۵]. با پیشرفت فناوری و افزایش داده‌های، بکارگیری یادگیری عمیق و استخراج شبکه‌های عصبی مصنوعی مبتنی بر رفتار هنجار سیستم نتایج بهتری در سامانه‌های تشخیص نفوذ رفتار-محور نشان داده‌اند[۴ و ۱۹-۱۸].

در رویکردهای مبتنی بر یادگیری، آموزش به‌صورت خودکار از داده انجام می‌شود و الگوهای موجود شناسایی شده و بر آن اساس تصمیماتی گرفته می‌شوند. باتوجه به عدم وجود داده‌های برچسب‌دار زیاد در محیط‌های صنعتی، روش‌های کشف حملات عموما براساس یادگیری بدون نظارت و استخراج رفتار صحیح سامانه براساس مدل می‌باشد. شناسایی مدل رفتار هنجار سامانه، حملات و رفتارهای ناهنجار را مشخص می‌کند. طبق پژوهش‌های انجام شده با زیاد شدن حجم داده‌ها و نیاز به تحلیل برخط، یادگیری عمیق در این حیطه از عملکرد بسیار خوبی برخوردارند [۲۲-۲۰]. بزرگترین چالش رویکردهای کشف ناهنجاری مبتنی بر یادگیری عمیق پیشین، پیچیدگی محاسباتی و تمرکز بر حملات به‌صورت محلی می‌باشد. تلاش‌هایی به‌منظور بهبود حجم محاسباتی و درنظر گرفتن همبستگی داده‌های حمله در روش‌های ارائه شده انجام شده‌است اما عموما بر یک ساختار و مدل یادگیری متمرکز می‌باشند و قابلیت خودنظارتی در آن‌ها وجود ندارد.

هدف این مقاله ارائه روشی خودنظارتی مبتنی بر یادگیری عمیق گروه‌محور جهت کشف ناهنجاری حسگرها در سامانه‌های کنترل صنعتی می‌باشد. این روش با درنظر گرفتن موثرترین رویکردهای یادگیری عمیق شامل حافظه بلندکوتاه مدت، شبکه عصبی پیچشی و شبکه عصبی تمام متصل به‌صورت گروه‌محور سعی در بکارگیری مزایای آن‌ها در استخراج ویژگی و بهبود عملکرد کشف ناهنجاری دارد. معماری رویکردهای کشف ناهنجاری روش پیشنهادی مبتنی بر ساختار رمزگذار رمزگشا می‌باشد و با یادگیری رفتار داده به‌صورت بدون نظارت، ناهنجاری‌ها را به‌صورت خودنظارتی کشف می‌کنند. این رویکردها بر جنبه‌های زمانی، مکانی، محلی و سراسری داده‌ها تمرکز داشته و ترکیب گروه‌محور آن‌ها منجر به هم‌افزایی این قابلیت‌ها در یک مدل می‌شود.

به‌منظور اجماع روش‌های کشف ناهنجاری، نتایج شباهت‌سنجی این واحدها به یک رای‌گیر گیت آستانه ارسال می‌شود تا براساس قابلیت اطمینان هر واحد تصمیم آن در نتیجه نهایی اعمال گردد. نوآوری‌های اصلی این پژوهش به شرح زیر خلاصه شده است:

- ارائه یک رویکرد خودنظارتی گروه‌محور کشف ناهنجاری با درنظر گرفتن ویژگی‌های مکانی، زمانی و محلی داده‌های ثبت شده توسط حسگرهای سامانه اینترنت اشیا صنعتی؛
- ارائه رویکرد رای‌گیری حد آستانه وزن‌دار به‌منظور تعیین خروجی نهایی براساس قابلیت اطمینان روش‌های تجمیع شده؛
- درنظر گرفتن مقادیر متغیر قاب‌بندی پنجره تحلیل به‌منظور استخراج همبستگی‌های کوتاه و باند مدت زمانی.

در ادامه و در بخش دوم مطالعات پیشین بررسی می‌شود، در بخش سوم به مرور مفاهیم پایه پرداخته می‌شود، جزئیات روش پیشنهادی و آزمایش‌های تجربی و ارزیابی در فصل‌های چهارم و پنجم ارائه شده است. درنهایت و در بخش ششم به جمع‌بندی و روال آتی پرداخته شده است.

۲- پژوهش‌های پیشین

تجهیزات در سامانه‌های کنترل صنعتی با هدف تعامل با محیط فیزیکی به‌منظور جمع‌آوری اطلاعات و اعمال تحریک بکارگرفته می‌شوند. در این سیستم‌ها، داده و اطلاعات از محیط فیزیکی توسط حسگرها دریافت می‌شود و پس از پردازش و محاسبات، کنترل‌های لازم به محرکه‌ها جهت بازگشت به محیط فیزیکی داده می‌شود[۲۳]. با پیچیده شدن کاربردها و افزایش بکارگیری این سیستم‌های در ماموریت‌های ایمنی بحرانی، مسئله خطا و خرابی در آن‌ها بیش از پیش مطرح شده است. از سوی دیگر، با افزایش تعداد و حجم اطلاعات حسگرها به‌همراه پیچیدگی پردازش در این سیستم‌ها، مسئله مواجهه با خطاها اهمیت بیشتری یافته‌اند. درنتیجه شناسایی و مدیریت این ناهنجاری‌ها توسط عامل انسانی میسر نیست و لازم است به‌صورت خودکار و ماشینی انجام گیرد. پژوهش‌هایی با هدف کشف ناهنجاری‌ها انجام گرفته است که می‌توان آن‌ها را براساس ماهیت به روش‌های سنتی و داده‌محور دسته‌بندی کرد.

روش‌های کشف ناهنجاری سنتی عموما مبتنی بر رویکردهای آماری-احتمالی، به مدلسازی و تحلیل رفتار نرمال براساس تاریخچه براساس رویکردهای ریاضی می‌پردازند[۲۴و۲۵]. در این زمینه، اطلاعات جمع‌آوری شده توزیع احتمالی را شکل می‌دهند و از عملگرهای آماری مانند میانگین و واریانس برای کشف رفتارهای ناهنجار استفاده می‌شود. مشابهت داده حسگرها در مدل آماری استخراج شده، معیار کشف ناهنجاری در این روش‌ها می‌باشد. روش‌های مبتنی بر رویکردهای آماری، زنجیره مارکف، تحلیل سری زمانی، تطابق الگو و نظریه اطلاعات در این دسته‌بندی قرار می‌گیرند. این روش‌ها از دقت، سرعت و قطعیت مناسبی برخوردارند و عموما در روش‌های برخط مورد استفاده قرار می‌گیرند [۲۸-۲۶]. گرچه با پیچیده شدن کاربردها و افزایش حجم داده‌ها، پردازش‌های لازم و تخمین توزیع مناسب در این دسته از روش‌ها صورت نمی‌گیرد و در حملات پیچیده به نتیجه مناسب نمی‌رسند [۲۸و۲۹].

روش‌های کشف ناهنجاری داده محور، با بکارگیری رویکردهای یادگیری ماشین، یادگیری عمیق و یادگیری تقویتی سعی بر تحلیل داده و استخراج ناهنجاری دارند. قابلیت اطمینان و قطعیت مدل‌های داده‌محور از روش‌های سنتی کمتر است اما به‌دلیل عملکرد مناسب در داده‌های زیاد و استخراج ویژگی بیشتر مورد استفاده قرار می‌گیرند. در این زمینه، رویکردهای مبتنی بر یادگیری به‌صورت بانظارت، بدون نظارت و نیمه‌نظارتی ارائه شده است. در [۲۹]. روش‌های k نزدیک‌ترین همسایه، جنگل تصادفی و درخت تصمیم به‌عنوان راهکارهای بانظارت تشخیص ناهنجاری در واحد تشخیص نفوذ سامانه کنترل صنعتی بکارگرفته شده است. باتوجه به ناشناخته بودن حملات در محیط‌های کنترل صنعتی و پیچیده شدن الگوی آن‌ها با پیشرفت فناوری، تکیه بر رویکردهای بانظارت و استفاده از دانش قبلی به‌منظور برچسب‌دهی به داده‌ها کافی نمی‌باشد.

با گذشت زمان و افزایش حجم داده‌ها و پیشرفته شدن راهکارها در سامانه‌های اینترنت اشیا صنعتی، روش‌های یادگیری عمیق بعنوان موثرترین راهکارهای مواجهه با چالش‌ها کاندید شدند. در این رویکردها، هوشمندی محاسباتی از اطلاعات حسگرها فراهم می‌شود و فرایند تولید به‌طور موثر ادامه می‌یابد. در این روش‌ها عموما از ساختارهای خودرمزگذار، شبکه‌های عصبی پیچشی، شبکه بیز و بوتزمان استفاده شده است. یکی از نخستین سامانه‌های تشخیص نفوذ صنعتی مبتنی بر تحلیل سری زمانی به کمک شبکه عصبی حافظه بلند- کوتاه مدت (LSTM) و ترکیب آن با رویکردهای مبتنی بر امضا در [۳۰] ارائه شده است. یکی از چالش‌های مهم استفاده از این رویکرد توانایی محدود آن در مدیریت تصمیم‌گیری مبتنی بر تحلیل اطلاعات قبلی است. به‌نوعی این ساختار با استفاده از نورون‌های دروازه‌ای مبتنی بر

دروازه‌هایی نظیر فراموشی و انتخاب ورودی و غیره، می‌تواند همبستگی‌های زمانی طولانی را تا حدودی لحاظ کند. اما نهایتاً برای ارتباطات زمانی بسیار طولانی و غیرمارکفی، توانایی مناسبی را از خود نشان نمی‌دهد. در [۳۱] رویکردی مبتنی بر استفاده از شبکه‌های عصبی پیچشی (CNN) برای تشخیص ناهنجاری در شبکه‌های کنترل صنعتی ارائه شده است. این رویکرد همبستگی بین ویژگی‌ها و توزیع نامتوازن نمونه‌ها را درنظر می‌گیرد و به دلیل تنکی اتصالات و وجود وزن مشترک بین نورون‌های لایه‌های کانوولوشنی، متغیرهای مرتبط را با هدف افزایش سرعت کاهش می‌دهد. با این وجود، شبکه‌عصبی پیچشی به‌دلیل وجود لایه‌های متعدد و استفاده مکرر از عملگرهای ضرب و جمع و اعمال توابع غیرخطی زمان اجرای طولانی دارد که بر کارایی عملیات برخط در واحدهای کنترل صنعتی تاثیر منفی دارد. در در [۳۲] راهکارهایی مبتنی بر استفاده ترکیبی از روش‌های ذکر شده پیشنهاد شده است و در [۳۳] ساختار سلسله‌مراتبی و سبک مبتنی بر شبکه عصبی حافظه بلند و کوتاه مدت با هدف کشف و دسته‌بندی ناهنجاری سامانه‌های سایبرفیزیکی ارائه شده است.

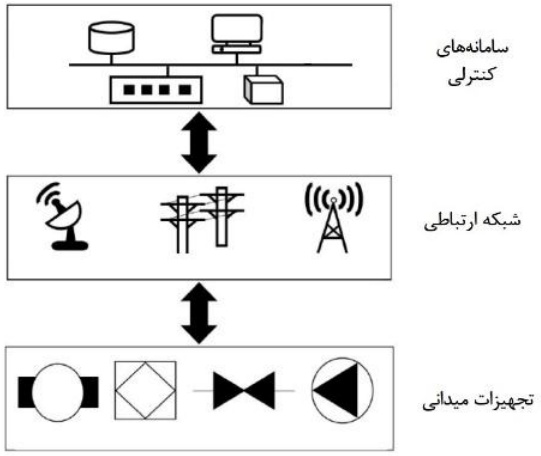
بکارگیری رویکردهای برخط و قابل اطمینان در واحد تشخیص نفوذ سامانه‌های کنترل صنعتی لازم می‌باشد. با افزایش حجم داده‌ها و تنوع رفتار ناهنجاری‌ها و حملات، بکارگیری رویکردهای داده محور مبتنی بر یادگیری کارایی بهتری ایجاد می‌کند. رویکردهای پیشین ارائه شده عموما بر ویژگی‌های زمانی یا مکانی داده‌ها متمرکز بوده‌اند و زمان پردازشی بالایی داشته‌اند. ارائه رویکرد کم هزینه و قابل اطمینان با درنظر داشتن ویژگی‌های مختلف داده و پویایی رفتار آن همچنان یکی از نیازمندی‌های اساسی این سامانه‌ها می‌باشد.

۳- مفاهیم پایه

در این بخش به مروری بر مفاهیم پایه‌ای که در روش پیشنهادی مورداستفاده قرار گرفته است پرداخته می‌شود.

۳-۱- سامانه‌های کنترل صنعتی

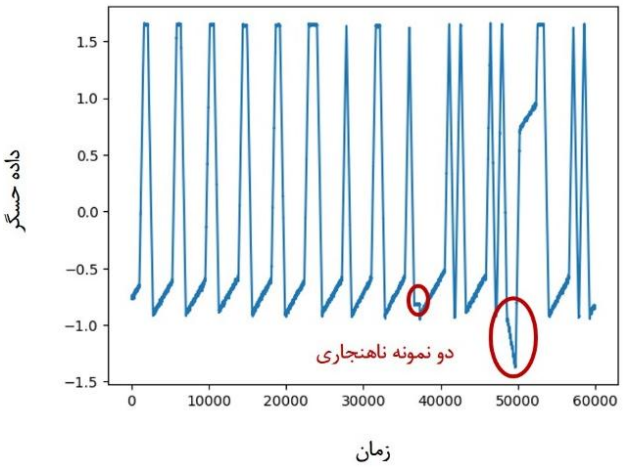
سامانه‌های کنترل صنعتی با هدف مدیریت فرایندهای فیزیکی در محیط‌های صنعتی بکار گرفته می‌شوند. این سامانه‌ها به‌طور متداول از سه شبکه تجاری سازمان، کنترلی و میدان تشکیل می‌شوند. شبکه تجاری سازمان شامل سرورهای ایمیل و وب و غیره بوده و معمولاً با شبکه اینترنت در ارتباط است. شبکه کنترلی از یک ایستگاه نظارت و کنترل مرکزی، تعدادی واسط انسان-ماشین (HMI) برای نمایش اطلاعات در مورد ایستگاه‌های راه‌دور و ارسال مستقیم دستورات به سیستم تشکیل شده است. شبکه میدان متشکل از تجهیزات فیزیکی با هدف جمع‌آوری داده و اعمال فرامین کنترلی می‌باشد. در این شبکه، حسگرها اطلاعات محیطی موردنیاز در پردازش را جمع‌آوری کرده و کنترلر منطقی قابل برنامه‌ریزی (PLC) اطلاعات حسگرها را به‌عنوان ورودی خوانده و دستورات برنامه‌ریزی شده واحد کنترل را هدایت می‌کند[۴،۲-۱]. ساختار و اجزای سامانه کنترل صنعتی براساس شبکه‌های شرح داده شده در شکل ۱ آمده است.



شکل ۱. ساختار سامانه کنترل صنعتی

۳-۲- ناهنجاری حسگرها

حسگرها با هدف جمع‌آوری اطلاعات محیطی در سامانه‌های کنترل صنعتی بکار گرفته می‌شوند. این اطلاعات به‌صورت سریال در زمان تولید می‌شوند و پس از انجام پردازش و محاسبات با هدف کنترل‌های لازم به محرکه‌ها جهت بازگشت به محیط فیزیکی داده می‌شود [۳۴]. بروز ناهنجاری در این داده‌ها به‌معنای خارج شدن از رفتار موردانتظار و معمول در یک نقطه یا مجموعه‌ای از نقاط است. در حالت کلی نرمال بودن داده‌های حسگر براساس محدوده اندازه‌گیری آن توسط تعریف حد آستانه مشخص می‌شود. خارج شدن محدوده داده از آستانه مشخص شده و قابل قبول سیستم، به‌معنای رخدادن یک ناهنجاری است که می‌بایست پیش از انتشار، مدیریت و اصلاح شود. لازم است ابتدا بر شناسایی و سعی در بررسی عاملی که باعث بروز خطا می‌شود تمرکز کرد تا از انتشار آن در سیستم جلوگیری شود. پس از شناسایی، فرایند تحمل‌پذیر کردن سیستم در برابر آن از طریق ایزوله کردن و حذف اثر اعمال می‌شود. شکل ۲ نمایی از رخداد ناهنجاری در اطلاعات یک حسگر در طول زمان را نشان می‌دهد.



شکل ۲. نمایی از رخداد ناهنجاری در مقادیر یک حسگر در زمان

برحسب نوع رفتار و مدت زمان اثر، ناهنجاری‌ها در انواع مختلف طبقه‌بندی می‌شوند. داده‌های ناهنجار فاصله‌ی زیادی با الگوی جامع بقیه‌ی داده‌ها دارند و ناشی از بروز خطا در سیستم یا رخداد نویز گذرا براثر پایین آمدن سطح انرژی حسگر می‌باشند. یکی از اساسی‌ترین چالش‌ها در فرایند تشخیص و مواجهه با ناهنجاری، شناسایی و تمیز بین این دو دسته از حیث نوع و زمان تاثیر بر سیستم و اعمال واکنش مناسب برای هریک در سیستم می‌باشد. بدین منظور درنظر گرفتن حدود آستانه مناسب برای تشخیص نوع ناهنجاری سیستم یا اجتناب از بروز ناهنجاری در سیستم روش مناسبی است. براساس دسته‌بندی داده‌ها و شناسایی نوع ناهنجاری، اعمال سازوکارهای کنترلی و پیشگیرانه مناسب منجر به رفع چالش و بهبود قابلیت اطمینان سیستم می‌گردد. لازم به ذکر است که عدم مواجهه صحیح با نویز، روال کشف ناهنجاری را به‌دلیل انتشار آن در سیستم تغییر داده و دشوارتر می‌سازد [۳۵].

۳-۳- ساختارهای یادگیری عمیق

یکی از عوامل تأثیرگذار در میزان موفقیت الگوریتم‌های یادگیری ماشین، چگونگی نمایش و بازنمایی داده‌ها می‌باشد. روش‌های گوناگون بازنمایی می‌توانند اطلاعات و دانش خاصی را درباره موضوع مورد بررسی آشکار یا نهان نمایند. یکی از ضعف‌ها و مشکلات روش‌های یادگیری فعلی، ضعف آن‌ها در استخراج و توسعه اطلاعات تفکیکی یک مجموعه داده به‌صورت خودکار است.

مهندسی ویژگی یک راه رایج برای جبران دانش خاص پیرامون یک موضوع خاص تحقیقاتی با استفاده از دانش بشری است که می‌تواند در برخی موارد کمک‌کننده باشد. با این وجود یک هدف در هوش مصنوعی، کاهش وابستگی سامانه مبتنی بر آن به انسان است. در حقیقت یک سامانه مبتنی بر هوش مصنوعی بایستی جهان اطراف را درک نموده و این میسر نیست مگر از طریق یادگیری چگونگی استخراج و آشکارسازی دانش نهفته در دل داده‌های خام سطح پایین‌تر.

یادگیری عمیق عبارت است از چگونگی ترکیب نگاشت‌های غیرخطی بر روی دادگان برای رسیدن به نمایشی کم‌حجم‌تر، چکیده‌تر و مفیدتر. در حقیقت در یادگیری عمیق، نگاشت‌های غیرخطی متوالی چنان با یکدیگر ادغام می‌شوند که در هر مرحله اطلاعاتی مفیدتر از چکیده‌سازی و ادغام اطلاعات مراحل قبل و به‌صورت سلسله‌مراتبی حاصل شود[۳۶].

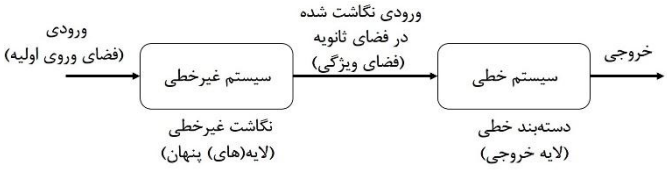
ساختار یادگیری عمیق عموماً مبتنی بر استفاده از شبکه‌های عصبی است که با اعمال نگاشت‌های غیرخطی اقدام به استخراج ویژگی‌های سطح بالاتر در لایه‌های عمیق (نزدیک به ورودی) خود و سپس تصمیم‌گیری مبتنی بر مقادیر این ویژگی‌ها در لایه خروجی خود می‌کنند. شکل ۳ مفهوم کلی استخراج ویژگی در یک ساختار عمیق را نمایش می‌دهد. در این شکل مجموعه‌ای از لایه‌ها در حکم ماشین غیرخطی وظیفه استخراج ویژگی‌های مناسب از دل داده اولیه را داشته و در ادامه، لایه خروجی در حکم یک ماشین خطی اقدام به تصمیم‌گیری نهایی می‌کند.

در ادامه این بخش به مروری بر کلیت سه روش یادگیری عمیق مورد استفاده در این روش شامل شبکه‌عصبی تمام‌متصل[1]، شبکه‌عصبی بلند-کوتاه مدت[2] و شبکه‌عصبی پیچشی[3] می‌پردازیم.

۳-۳-۱- شبکه‌عصبی تمام‌متصل

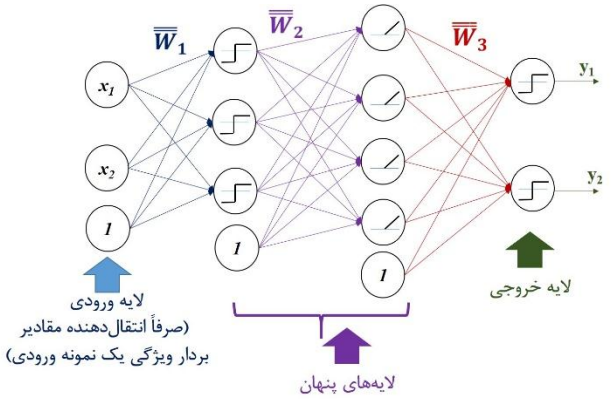
در شبکه‌عصبی تمام‌متصل، نورون‌های هر لایه خروجی همه نورون‌های لایه پیشین خود را به‌عنوان ورودی دریافت کرده و پس از اعمال بردار وزن و یک تابع غیرخطی، خروجی خود را تولید می‌کنند. مزیت این معماری آگاهی هر نورون از کلیه پردازش‌های لایه پیشین خود و داشتن توانایی در کشف همبستگی و ارتباط بین ویژگی‌های استخراج‌شده در لایه قبل برای تولید خروجی خود می‌باشد. این ارتباط محدود به ارتباط محلی یک ویژگی با ویژگی‌های اطراف خود نیست و در یک مسئله مبتنی بر سری زمانی (نظیر پایش خروجی حسگرها در مسئله تشخیص نفوذ در سامانه‌های کنترل صنعتی) از سوی دیگر، چالشی که این معماری با آن مواجه است وجود اتصالات بسیار زیاد و در نتیجه وزن‌های متعددی است که لازم است در فرایند یادگیری آموزش ببینند.

[1] Fully Connected
[2] Long-Short Term Memory (LSTM)
[3] Convolutional



شکل ۳. مفهوم استخراج ویژگی در ساختار عمیق

بنابراین این مدل با انفجار تعداد پارامترها مواجه است که با توجه به یادگیری ساختارهای عصبی مبتنی بر الگوریتم‌های جستجوی محلی (نظیر الگوریتم تنزل گرادیان[1]) می‌تواند به گیرکردن آن‌ها در کمینه محلی منجر شود[۳۶و۳۷]. شکل ۴ ساختار یک شبکه‌عصبی تمام‌متصل با دو لایه پنهان را نمایش می‌دهد. در این شکل ماتریس وزن هر لایه با $\overline{\overline{W}}_i$ نمایش داده شده است. هر لایه دارای تابع فعالیت غیرخطی خود می‌باشد (در این شکل، لایه پنهان اول دارای تابع فعالیت tanh، لایه پنهان دوم دارای تابع فعالیت ReLU و لایه خروجی دارای تابع فعالیت غیرخطی سیگموئید است.



شکل ۴. نمایی کلی از یک ساختار عصبی تمام‌متصل با دو لایه پنهان، دو بعد ورودی (x) و دو بعد خروجی (y)

۳-۳-۲- حافظه بلند-کوتاه‌مدت

شبکه‌های عصبی بازگشتی۲، دسته‌ای از شبکه‌های عصبی را تشکیل می‌دهند که به‌دلیل داشتن بازخورد داخلی، قادر به‌خاطرسپاری دنباله‌ای از حالات می‌باشند. به‌عبارت دیگر، شبکه‌های عصبی بازگشتی، در هر زمان یک گام از یک دنباله‌ی ورودی را مورد پردازش قرار داده و در لایه‌ی پنهان خود، یک بردار حالت نگهداری می‌کنند که به‌طور تلویحی اطلاعاتی از تاریخچه‌ی همه عناصر قبلی دنباله را در خود نگه می‌دارد [۳۸]. بنابراین، این ساختار برای کاربردهایی مبتنی بر دریافت یک دنباله، نظیر سری زمانی یک حسگر مناسب خواهد بود.

با وجود آنکه هدف اصلی شبکه‌های عصبی بازگشتی یادگیری وابستگی‌های زمانی درازمدت است، شواهد نظری و تجربی حاکی از پیچیدگی ذخیره‌سازی اطلاعات برای مدت‌زمان طولانی است [۳۹]. برای مواجهه با این مشکل شبکه‌های عصبی بازگشتی، تلاش‌هایی در جهت ارائه‌ی شبکه‌های عصبی با قابلیت حفظ وابستگی‌های زمانی دراز و کوتاه‌مدت از طریق افزودن یک حافظه‌ی صریح مدیریت‌شده انجام شده است. یکی از موفق‌ترین ساختارهای ارائه شده در این راستا، شبکه‌های عصبی بازگشتی موسوم به حافظه‌های بلند-کوتاه‌مدت۳ می‌باشند [۴۰].

حافظه‌های بلند-کوتاه‌مدت (LSTM)، دسته‌ای موفق از شبکه‌های عصبی بازگشتی می‌باشند که به‌نوعی در آن‌ها عنصر حافظه، توانایی فراموشی و به‌خاطرسپاری موارد مناسب قرار داده شده است. به‌عبارت دیگر، در این شبکه‌های عصبی، همه‌ی موارد گذشته به‌خاطر سپرده نمی‌شوند؛ بلکه موارد زائد فراموش شده و صرفاً موارد مناسب بلندمدت یا کوتاه‌مدت به‌خاطر سپرده می‌شوند. قابلیت فراموشی موارد زائد، توانایی یادگیری وابستگی‌های بلندمدت موجود در یک دنباله را به این شبکه‌ها منتقل می‌کند.

ایده‌ی اساسی در شبکه‌های LSTM، استفاده از واحدهای حافظه مجهز به نورون‌های خاصی با عنوان «نورون‌های دروازه‌ای[4]» می‌باشد. این نورون‌های دروازه‌ای عموماً به سه دسته تقسیم می‌شوند، نورون‌های دروازه‌ی ورودی، خروجی و فراموشی . این سه نورون دروازه، امکانات یک حافظه‌ی رقمی شامل خواندن، نوشتن و راه‌اندازی مجدد را فراهم می‌کنند. به‌عبارت دیگر، ورودی بلوک حافظه، پیش از بررسی در خروجی دروازه‌ی ورودی (که مقدار بین صفر و یک است) ضرب می‌شود تا میزان تأثیر ورودی حساب شود. خروجی پیش از ارائه شدن در خروجی نورون، در خروجی دروازه‌ی خروجی (که مقداری بین صفر و یک است) ضرب می‌شود و همچنین دروازه‌ی فراموشی، که به‌صورت یک عامل ضرب دیگر پیاده‌سازی می‌شود، میزان بخاطرسپاری و تغییر وضعیت نورون را تعریف می‌کند [۳۶و۴۱].

به‌طور کلی، ساختارهای بازگشتی، از واحدهای تکراری تشکیل می‌شوند. به‌طور مثال، شبکه‌ی عصبی بازگشتی معمولی، از یک واحد ساده تک نورون تشکیل شده که در طول زمان تکرار می‌شود. اما ساختار LSTM، خود از چهار لایه یا مسیر مجزا تشکیل می‌شود که در طول زمان تکرار می‌شوند. شکل ۵، کلیت

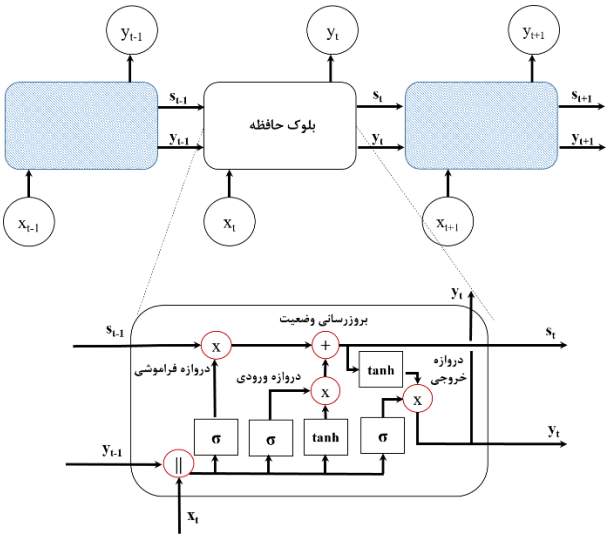
[1] Gradient descent
[2] Recurrent Neural Network (RNN)
[3] Long Short-Term Memory (LSTM)
[4] Gated neuron

ساختار شبکه‌عصبی LSTM را به‌صورت بازشده در طول زمان نمایش می‌دهد. بلوک‌های تکرار شونده در حقیقت، تکرار یک بلوک در طول زمان است. در این شکل، نماد σ، تابع سیگموئیدی با خروجی بین صفر و یک است. سیگنال x، سیگنال ورودی، سیگنال y، سیگنال خروجی و سیگنال s، سیگنال وضعیت سلول است. علامت || به‌معنای ادغام سیگنال ورودی و خروجی است.

در این شبکه، عنصر محوری سیگنال s_t است که در طول زمان تاریخچه اطلاعات ورودی را حفظ می‌کند و سیگنال وضعیت سلول نام دارد. میزانی از این سیگنال که لازم است در طول هر گام زمانی حفظ شود، به‌کمک دروازه‌ی فراموشی، مبتنی بر ورودی جدید و خروجی گام قبل، تعیین می‌شود. بنابراین امکان راه‌اندازی مجدد حافظه از طریق صفر کردن این سیگنال توسط دروازه‌ی فراموشی وجود دارد. چنانچه خروجی تابع سیگموئید دروازه‌ی فراموشی را در زمان t، f_t بنامیم، این مقدار از رابطه‌ی زیر بدست می‌آید:

$$f_t = \sigma(W_f . [y_{t-1}, x_t] + \theta_f) \quad (۱)$$

که در اینجا W_f، بردار وزن دروازه فراموشی، $[y_{t-1},x_t]$، بردار متشکل از ورودی جدید و خروجی گام قبل و θ_f حد آستانه‌ی دروازه‌ی فراموشی است. لازم است بردار وزن و حد آستانه بر اساس روش‌های یادگیری نظیر پس‌انتشار خطا در زمان، یاد گرفته شوند. با ضرب f_t که مقداری بین صفر و یک است ، در سیگنال s_t، میزانی از سیگنال وضعیت سلول که لازم است حفظ شود تعیین می‌شود.



شکل ۵. حافظه بلند-کوتاه‌مدت

لازم است سیگنال وضعیت سلول برای حفظ اطلاعات ورودی جدید، در هر گام به‌طور مناسب بروزرسانی شود. این بروزرسانی می‌تواند از طریق افزایش یا کاهش این سیگنال در بخش بروزرسانی وضعیت انجام شود. در دروازه‌ی ورودی، تعیین می‌شود که چه میزان از اصلاح باید انجام شود و میزان اصلاح به‌صورت مقداری در بازه [۱، ۱-] توسط تابع tanh مشخص می‌شود. نهایتاً این مقدار در یک جمع‌کننده، با میزان حفظ‌شده از سیگنال وضعیت سلول جمع می‌شود. به این ترتیب، میزانی از اطلاعات ورودی که لازم است در حافظه نگهداری شود، مشخص می‌شود. چنانچه میزان خروجی تابع سیگموئیدی دروازه‌ی ورودی، با i_t و میزان اصلاح خروجی از تابع سیگموئیدی دروازه‌ی ورودی با c_t مشخص شوند خواهیم داشت:

$$i_t = \sigma(W_i . [y_{t-1}, x_t] + \theta_i) \quad (۲)$$

$$c_t = tanh(W_c . [y_{t-1}, x_t] + \theta_c) \quad (۳)$$

که در اینجا Wi، بردار وزن دروازه ورودی، θi حد آستانه‌ی دروازه‌ی ورودی، Wc بردار وزن اصلاح و θc حد آستانه میزان اصلاح است. بنابراین میزان سیگنال وضعیت سلول به‌صورت زیر بروزرسانی می‌شود:

$$s_t = f_t \times s_{t-1} + i_t \times c_t \quad (۴)$$

در پایان لازم است خروجی و میزان آن توسط دروازه‌ی خروجی تعیین شود که مشابه با دو دروازه‌ی دیگر عمل می‌کند. خروجی، بر اساس سیگنال وضعیت سلول و مبتنی بر یک تابع tanh مشخص شده و میزان خروجی توسط دروازه‌ی خروجی تعیین می‌شود. چنانچه خروجی تابع سیگموئیدی دروازه‌ی خروجی که تعیین‌کننده‌ی میزان خروجی است را با ot نمایش دهیم خواهیم داشت:

$$o_t = \sigma(W_o . [y_{t-1}, x_t] + \theta_o) \quad (۵)$$

$$y_t = ot \times tanh(s_t) \quad (۶)$$

که در اینجا Wo، بردار وزن دروازه خروجی، [yt-1,xt]، بردار متشکل از ورودی جدید و خروجی گام قبل و θo حد آستانه‌ی دروازه‌ی خروجی است.

عموماً از قاعده پس‌انتشار خطا در زمان[1] برای آموزش این نوع شبکه‌عصبی استفاده می‌شود.

[1] Backporpagation through time (BPTT)

۳-۳-۳- شبکه‌عصبی پیچشی

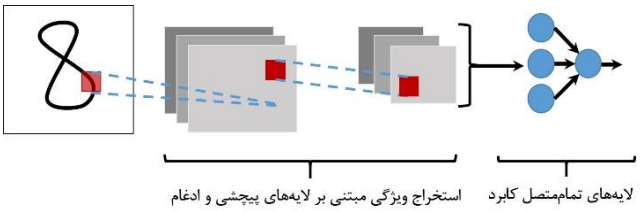
شبکه‌های عصبی پیچشی (کانوولوشنی) [۴۲] نوع خاصی از شبکه‌های عصبی را تشکیل می‌دهند که برای پردازش داده با توپولوژی شبکه‌ای مشخص، نظیر سری‌های زمانی (یک‌بعدی) در بازه‌های زمانی مشخص و تصویر (دوبعدی)، مورد استفاده قرار می‌گیرند. نام کانوولوشن نشئت‌گرفته از استفاده‌ی این شبکه از عملگر کانوولوشن بر روی داده‌های ورودی است. به‌عبارت دیگر، این شبکه‌های عصبی، حداقل در یکی از لایه‌های خود از عملگر کانوولوشن بجای ضرب ماتریسی کلی، استفاده می‌کنند. شبکه‌های عصبی کانوولوشنی، از ساختارهای تأثیرگذار بر روی یادگیری عمیق می‌باشند. چهار ایده‌ی اساسی در شبکه‌های کانوولوشنی عبارت‌اند از: اتصالات محلی، وزن‌های به‌اشتراک‌گذاشته‌شده، ادغام و استفاده از تعداد زیاد لایه.

معماری یک شبکه‌ی عصبی کانوولوشنی از دنباله‌ای از مراحل تشکیل شده است. مراحل اولیه‌ی متشکل دو نوع لایه می‌باشند: ۱- لایه‌ی کانوولوشن[1] و لایه‌ی ادغام[2]. هر واحد هر لایه‌ی کانوولوشنی به‌صورت محلی به بخشی از ویژگی‌های ورودی از لایه‌ی قبل متصل است و این ورودی‌ها را در یک وزن ورودی ضرب می‌کند (به مجموعه‌ی وزن‌ها بانک فیلتر گفته می‌شود). حاصل ضرب وزن در بخش محدود و محلی ویژگی‌های ورودی از یک تابع غیرخطی نظیر سیگموئید یا ReLU عبور داده می‌شود. همه‌ی واحدهای یک بخش دارای وزن‌های یکسان هستند (در حکم کانوولوشن و لغزاندن پنجره که نظیر آن در فیلترکردن تصاویر مورد استفاده قرار می‌گیرد). به هر بخش یک نگاشت ویژگی گفته می‌شود. نگاشت‌های ویژگی مختلف از بانک فیلتر مختلف استفاده می‌کنند [۳۸]. شکل ۶، نمایی از کلیه عملکرد شبکه عصبی پیچشی به همراه عملکرد فیلترها را در یک مثال نمایش می‌دهد.

پس از اعمال فیلترهای مختلف، نتایج لایه‌ی کانوولوشن در لایه‌ی ادغام، تجمیع می‌شود. این ادغام می‌تواند به‌صورت محاسبه‌ی بیشینه در یک نگاشت ویژگی مورد استفاده قرار گیرد. چندین لایه کانوولوشن، نگاشت غیر خطی و ادغام می‌تواند منجر به استخراج ویژگی‌های کمتر با مفهوم سطح بالاتر مختلف محلی شود که می‌توانند برای تصمیم‌گیری مناسب مورد استفاده قرار گیرند. برای یادگیری می‌توان از قانون پس‌انتشار خطا استفاده کرد. مزیت این ساختار، یکسان بودن وزن‌های یک بانک فیلتر و داشتن اتصالات محلی است؛ درواقع هر نورون در این لایه برخلاف مدل تمام‌متصل فقط به یک همسایگی در نورون‌های لایه قبل متصل است. این موضوع منجر به کاهش تعداد پارامترهای این مدل نسبت به رویکرد تمام‌متصل می‌شود.

۳-۳-۴- الگوریتم یادگیری

معمولاً در فرایند یادگیری وزن‌های شبکه‌های عصبی از جمله شبکه‌عصبی تمام‌متصل، LSTM و پیچشی از رویکردهای مبتنی بر پس‌انتشارخطا و تنزل گرادیان استفاده می‌شود. این رویکردها، رویکردهای بهینه‌سازی تکراری مبتنی بر جستجوی محلی برای کاهش خطای بین خروجی شبکه و خروجی به‌کمک مشتق زنجیره‌ای می‌باشند. بر اساس تحلیل بسط تیلور مرتبه اول، بردار گرادیان جهت رشد تابع را نشان می‌دهد. بنابراین تغییر در جهت قرینه آن منجر به کاهش تابع می‌شود. چنانچه هدف از یادگیری، کاهش تابع خطا باشد، می‌توانیم وزن‌های شبکه را در جهت قرینه بردار گرادیان تغییر دهیم (تنزل گرادیان). این رویکرد منجر به کاهش تدریجی خطا خواهد شد. یکی محبوب‌ترین الگوریتم‌های این خانواده که بسیار پراستفاده است الگوریتم ADAM می‌باشد که مبتنی بر تقریبی از مشتق دوم و تقعر تابع ارائه شده است [۳۷]. شبه‌کد این الگوریتم در شکل ۷ ارائه شده است.



شکل ۶. ساختار کلی شبکه‌عصبی کانوولشنی

1: Initialize parameters θ, first moment vector m, and second moment vector v
2: Initialize time step $t \leftarrow 0$
3: Set hyperparameters α (learning rate), β_1, β_2 (decay rates), ϵ (small constant)
4: **while** until θ converges **do**
5: $t \leftarrow t+1$
6: Compute gradient of the loss function $g_t \leftarrow \nabla_\theta \mathcal{L}(\theta)$
7: Update first moment estimate $m_t \leftarrow \beta_1 \cdot m_{t-1} + (1-\beta_1) \cdot g_t$
8: Update second moment estimate $v_t \leftarrow \beta_2 \cdot v_{t-1} + (1-\beta_2) \cdot g_t^2$
9: Correct bias in first moment $\hat{m}_t \leftarrow m_t / (1-\beta_1^t)$
10: Correct bias in second moment $\hat{v}_t \leftarrow v_t / (1-\beta_2^t)$
11: Apply update $\Delta\theta \leftarrow -\alpha \cdot \hat{m}_t / (\sqrt{\hat{v}_t} + \epsilon)$
12: Update parameters $\theta \leftarrow \theta + \Delta\theta$
13: **end while**

شکل ۷. الگوریتم بهینه‌سازی ADAM [۳۷]

[1] Convolution layer
[2] Pooling layer

۳-۴- یادگیری خودنظارتی

رویکردهای خودنظارتی[1] به‌منظور یادگیری ویژگی‌های تمیزدهنده مناسب سطح بالا از انبوده‌ی داده بدون برچسب ارائه شده‌اند. شکل ۸ نمایی کلی از رویکرد خودنظارتی را نمایش می‌دهد. در این فرایند، ابتدا در مرحله پیش‌تعلیم[2] به‌منظور استخراج ویژگی، یک وظیفه کمکی[3] تعریف می‌شود که به‌کمک داده بدون برچسب قابل حل است. در این مرحله، شبه‌برچسب‌هایی برای وظیفه کمکی تعریف می‌شوند. پس از تکمیل فرایند پیش‌تعلیم، مدل بدست آمده برای انجام وظیفه اصلی انتقال می‌یابد [۴۳-۴۴].

معمولاً ساختار در هنگام پیش‌تعلیم، مشابه رویکرد خودرمزگذار[4] از دو بخش کدگذار[5] و کدگشا[6] تشکیل می‌شود. در مرحله کدگذاری، داده ورودی به فضای ویژگی نگاشت می‌شود که بخش کدگشا به‌کمک آن بتواند وظیفه کمکی را حل کند. درواقع اطلاعات سطح بالایی از دل داده اصلی در این فضای ویژگی کدگذاری شده است. برخی وظایف کمکی متداول عبارتند از پرکردن بخشی از تصویر، پیش‌بینی قاب بعدی در ویدئو و بازسازی تصویر سه‌بعدی از روی تصاویر دوبعدی.

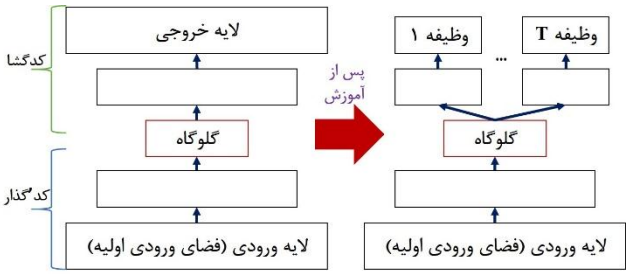
۴- روش پیشنهادی

۱-۴- بیان مسئله

روش پیشنهادی با هدف کشف ناهنجاری در حسگرهای سامانه‌های کنترل صنعتی که ناشی از حملات سایبرفیزیکی یا کارکرد نادرست است ارائه شده است. به‌دلیل حجم زیاد داده‌ها و نیاز به استخراج ویژگی‌های داده هنجار در سطوح مختلف، روش پیشنهادی مبتنی بر رویکردهای خودنظارتی و یادگیری عمیق گروه‌محور ارائه شده است. معلومات و اهداف روش پیشنهادی در ادامه بیان می‌شوند.

معلومات روش پیشنهادی:

- معماری زیرساخت سامانه کنترلی هدف
- داده‌های خروجی حسگرها در قالب سری زمانی
- الگوی رفتاری هنجار داده‌های حسگرها



شکل ۸. خط لوله متداول رویکردهای یادگیری خودنظارتی

اهداف روش پیشنهادی:

- تشخیص الگوی ناهنجار در داده‌های حسگرها
- استخراج الگوی هنجار مبتنی بر تحلیل همبستگی ویژگی‌های زمانی-مکانی داده‌های حسگرها
- ارائه راهکار گروه‌محور به‌منظور تمرکز بر ویژگی‌های مختلف با هدف بهبود اطمینان
- پیکربندی سامانه کنترلی هدف با حذف اثر ناهنجاری کشف شده و اعلام هشدار در بازه زمان مشخص

۲-۴- جزئیات روش پیشنهادی

روش پیشنهادی با هدف کشف ناهنجاری حسگرهای سامانه‌های کنترل صنعتی مبتنی بر یادگیری عمیق گروه‌محور ارائه شده است. برای تشخیص ناهنجاری در این روش، ابتدا داده‌های حسگرها پیش‌پردازش می‌شوند و سپس این داده‌ها به ساختارهای تصمیم‌گیر طراحی شده داده می‌شوند تا هنجار یا ناهنجار بودن آنها مشخص گردد. شکل ۹ معماری ساختار پیشنهادی را نشان می‌دهد. طبق این شکل، ابتدا داده‌های خام حسگرها به واحد پیش‌پردازش وارد می‌شوند و استانداردسازی، انتخاب ویژگی و قاب‌بندی زمانی بر آن‌ها اعمال می‌گردند. در واحد تشخیص، سیگنال‌های پیش‌پردازش شده به‌صورت موازی به سه عامل تشخیص از پیش تعلیم داده شده وارد می‌شوند. هریک از این عوامل تشخیص، داده ورودی را براساس عوامل مختلف زمانی، مکانی، محلی و سراسری تحلیل می‌کنند. مبتنی بر تحلیل خطای بازسازی، پیش‌بینی و همبستگی بین خطا در حسگرهای مختلف، قاب زمانی ورودی توسط هریک از عامل‌های تشخیص به‌عنوان هنجار یا ناهنجار معین می‌گردد. در نهایت، بین عوامل تشخیص رای‌گیری انجام می‌شود تا خروجی نهایی براساس تصمیم هر عامل مشخص گردد. براساس دقت یادگیری هر عامل، وزن اطمینان مشخصی برای هریک از عوامل تشخیص در رای‌گیر اعمال می‌شود. سازوکار رای‌گیری براساس رویکرد گیت آستانه درنظر گرفته شده است. در ادامه، جزئیات معماری پیشنهادی شرح داده می‌شود.

[1] Self-supervised learning
[2] Pre-training
[3] Pretext task
[4] Auto-encoder
[5] Encoder
[6] Decoder

شکل ۹. معماری ساختار تشخیص ناهنجاری پیشنهادی

۴-۲-۱- پیش‌پردازش

در ساختار روش پیشنهادی، داده ورودی از حسگرها به واحد پیش‌پردازش وارد می‌شود. این داده‌ها محدوده تغییرات متنوعی دارند که می‌بایست قبل از پردازش به قالب قابل یکسانی تغییر داده شوند. همچنین همه ویژگی‌های ورودی در فرایند تشخیص مفید نبوده، و برخی از آن‌ها فاقد تغییرات می‌باشند. بنابراین، در واحد پیش‌پردازش، داده‌های خام حسگرها با درنظر داشتن میانگین و انحراف از معیار استانداردسازی می‌شوند و در بازه مقادیر یکسانی قرار می‌گیرند. این رویکرد که در حوزه یادگیری ماشین به استانداردسازی معروف است، پراکندگی مقادیر ویژگی‌ها را در یکسان‌کردن دامنه هر یک لحاظ می‌کند. به‌علاوه برای داده‌های ورودی ویژگی‌های متفاوتی تعریف شده است که ناهنجاری در تمامی این ویژگی‌ها منعکس نشده است. درنتیجه به‌منظور بهبود فرایند یادگیری و تمرکز بر تشخیص ناهنجاری، ویژگی‌های ثابت شناسایی و حذف می‌شوند.

داده ورودی سامانه از حسگرها تامین می‌شود و در قالب سری زمانی می‌باشد پس لازم است قاب‌بندی درستی روی آن اعمال شود. اندازه قاب در کارایی روش پیشنهادی و پیش‌بینی اهمیت دارد، قاب کوچک مشخصه‌های وابستگی داده را بدرستی منعکس نمی‌کند و قاب بزرگ تحلیل پیچیده‌ای بر سیستم اعمال می‌کند. درنتیجه براساس آزمایش‌های تجربی و روش‌های پیشین، قاب‌بندی مناسبی بر داده ورودی اعمال می‌شود.

۴-۲-۲- واحد تشخیص و تصمیم

پس از پیش‌پردازش، داده وارد واحد تشخیص ناهنجاری و تصمیم می‌شود. در روش پیشنهادی این واحد از سه عامل تشکیل شده است که به‌صورت موازی داده ورودی را دریافت کرده و به تحلیل آن می‌پردازند. این عامل‌ها با بکارگیری رویکردهای یادگیری عمیق متفاوت که بر جنبه‌های زمانی، مکانی، سراسری و محلی تاکید دارند به بررسی و تحلیل داده حسگرها می‌پردازند. سپس نتیجه نهایی از اجماع این عامل‌ها در قالب رویکردی گروه‌محور و رای‌گیری وزن‌دار مبتنی بر اطمینان و گیت آستانه به‌دست می‌آید. نتیجه نهایی ناهنجار بودن یا نبودن قاب زمانی سری زمانی ورودی را نشان می‌دهد.

این واحد از سه عامل تشخیص تشکیل شده است که در کنار هم به‌صورت گروه‌محور وظیفه تشخیص را انجام می‌دهند. عوامل تشخیص دارای روند کاری همگن، اما ساختار ناهمگن هستند تا هر یک بر اساس ویژگی ساختار مورد استفاده شاخصه‌های متفاوتی از سیگنال ورودی را مورد بررسی قرار دهند. عملکرد کلی هر عامل تشخیص مبتنی بر معماری رمزگذار-رمزگشاست که بخش زمانی از یک قاب ورودی را دریافت کرده، به فضای کد برده و در فرایند کدگشایی به بخش باقیمانده زمانی از قاب در خروجی می‌رسد. این وظیفه کمکی پیش‌بینی بخشی از سیگنال ورودی منجر به کاهش فضای خروجی و تسهیل یادگیری می‌شود که به‌صورت خودنظارتی مبتنی بر داده هنجار تعلیم دیده است.

پس از بازسازی بخش دوم سیگنال ورودی (پیش‌بینی بخش دوم مبتنی بر بخش اول)، خروجی بدست آمده با بخش دوم از داده اصلی مقایسه شده و بردار خطای فرایند بازسازی تولید می‌شود. در زمان آموزش، بردارهای خطای داده هنجار محاسبه شده و جمعیت آماری خطای ساختار کدگذار-کدگشا را برای داده هنجار ایجاد می‌کند. خطای موجود در هر بعد، می‌تواند با خطای حاصل از بازسازی سایر ابعاد همبستگی داشته باشد. با فرض توصیف توزیع احتمال خطای بازسازی مبتنی بر یک توزیع گاوسی چندمتغیره که همبستگی ابعاد را نیز لحاظ می‌کند، این توزیع مبتنی بر بردار میانگین و ماتریس کواریانس خطا به‌صورت روابط زیر مبتنی بر قانون اعداد بزرگ[1] و قضیه حد مرکزی[2] قابل محاسبه خواهد بود:

$$e_i^{(k)} = x_i^{(k)} - \hat{x}_i^{(k)} \quad (۷)$$

$$m_i = \frac{1}{N}\sum_{k=1}^{N} e_i^{(k)} \quad (۸)$$

$$C_i = \frac{1}{N}\sum_{k=1}^{N} \left(e_i^{(k)} - m_i\right)^T \left(e_i^{(k)} - m_i\right) \quad (۹)$$

که در این روابط، $x_i^{(k)}$ و $\hat{x}_i^{(k)}$ به‌ترتیب بردار خروجی مطلوب و خروجی تولیدشده توسط عامل iام برای نمونه kام، $e_i^{(k)}$ بردار خطای بازسازی عامل iام برای نمونه kام، N تعداد نمونه‌های آموزشی، m_i بردار میانگین خطای بازسازی عامل iام و C_i ماتریس

[1] Law of Large Numbers
[2] Central Theorem Law

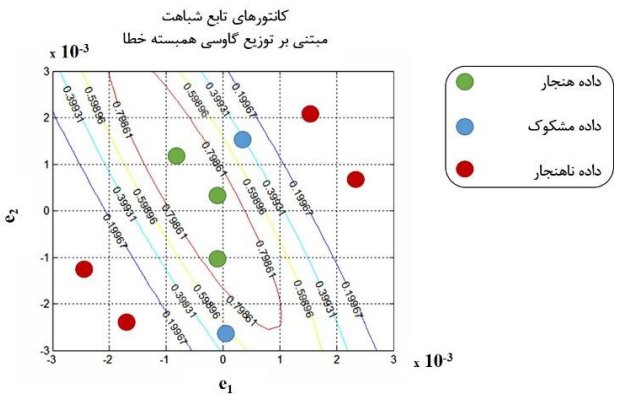
کواریانس عامل iام برای بردارهای خطا می‌باشد. در اینجا مبتنی بر فاصله ماهالانوبیس[1] به‌منظور در نظرگرفتن همبستگی ابعاد در شباهت‌سنجی، معیار شباهت زیر برای بردار خطای بازسازی تعریف می‌شود:

$$s_i(e_i) = \exp\left(-(\mathrm{e_i} - m_i)^T C_i^{-1}(\mathrm{e_i} - m_i)\right) \qquad (۱۰)$$

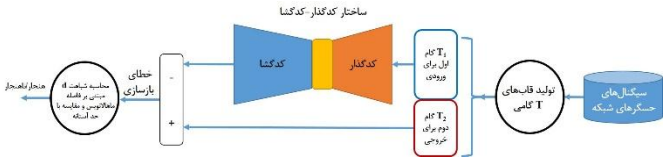
که s_i معیار شباهت است و مقداری در بازه [۰ ، ۱] دارد و هر چه بردار خطا ورودی به میانگین نزدیک‌تر باشد، مقداری نزدیک‌تر به ۱ دریافت می‌کند. برای تصمیم در مورد وجود ناهنجاری یک قاب این مقدار شباهت با یک حد آستانه از پیش تعریف شده مقایسه شده و در صورت بیشتر بودن شباهت از حد آستانه، داده هنجار و در غیراینصورت ناهنجار تلقی خواهد شد. شکل ۱۰ مفهوم نمونه‌های هنجار و ناهنجار در این روش را نمایش می‌دهد. همچنین شکل ۱۱ کلیت عملکرد هر عامل را نمایش می‌دهد.

اولین عامل دارای ساختار رمزگذار-رمزگشا مبتنی بر حافظه بلندکوتاه مدت می‌باشد. باتوجه به اینکه داده ورودی در قالب سری زمانی است، بکارگیری ساختار حافظه بلندکوتاه مدت عملکرد مناسبی در تخمین وابستگی‌های طولانی مدت و پیش‌بینی دنباله داده‌ها دارد. درنتیجه خاصیت‌های تحلیل زمانی، مکانی و سراسری محدودی بر پیش‌بینی داده‌ها اعمال می‌کند. در این ساختار داده‌ها با موارد پیش‌بینی شده مقایسه شده و درصورت عدم برابری رفتار ناهنجار گزارش می‌شود. عامل دوم مبتنی بر معماری تمام‌متصل است که بیشتر بر بررسی سراسری ویژگی‌ها متمرکز است. آخرین عامل، مبتنی بر ساختار پیچشی است (کدگذار پیچشی و کدگشای پیچشی ترانهاده) که مبتنی بر فیلترها، به بررسی محلی سیگنال‌های ورودی می‌پردازد.

به‌منظور شکل‌دهی رویکرد گروه‌محور پیشنهادی، نتایج بدست آمده از سه عامل تشخیص در واحد رای‌گیر مجتمع می‌شوند. رای‌گیر تعبیه‌شده نتایج را از عوامل تشخیص گرفته و وزن‌دهی می‌کند. وزن‌دهی به عوامل براساس پارامتر اطمینان مستخرج از معیار میانگین مربعات خطای یادگیری انجام می‌گیرد. پارامتر میانگین مربعات خطای یادگیری، معرف انحراف از معیار توزیع خطا می‌باشد که میزان پراکندگی خطا و عدم اطمینان روش را به‌خوبی نشان می‌دهد. بنابراین معکوس این معیار، شناسه مناسبی برای تعیین کارایی و اطمینان عامل‌های تشخیص می‌باشد که در روش پیشنهادی مورد استفاده قرار گرفته است.



شکل ۱۰. مفهوم نمونه‌های هنجار و ناهنجار مبتنی بر رابطه شباهت پیشنهادی



شکل ۱۱. معماری هر یک از عوامل تشخیص

همانگونه که شرح داده شد، خروجی عوامل تشخیص ناهنجاری، نتیجه تحلیل و بررسی آن‌هاست که بازه تغییرات صفر تا یک دارد. واحد رای‌گیر در روش پیشنهادی از نوع مبتنی بر آستانه درنظر گرفته شده است. این واحد، نتایج ضرب شده در وزن اطمینان هریک از عوامل را جمع کرده و با آستانه از پیش تعیین شده مقایسه می‌کند. اگر نتیجه بدست آمده از آستانه بیشتر بود، ناهنجاری رخ داده است. سازوکار واحد رای‌گیر به‌صورت زیر تعریف می‌شود:

$$Y = \begin{cases} 0, & \sum w_i \times s_i < \theta \\ 1, & \sum w_i \times s_i \geq \theta \end{cases} \qquad (۱۱)$$

در این رابطه، Y خروجی واحد تشخیص است که یک بودن آن رخداد ناهنجاری را نشان می‌دهد. همچنین پارامترهای w_i و s_i به‌ترتیب بیانگر وزن اطمینان و معیار شباهت بردار خطای بازسازی عوامل تشخیص طبق رابطه (۴) می‌باشند. پارامتر θ حد آستانه در نظر گرفته شده تصمیم برای رای‌گیر است که به‌صورت تجربی و در حین فرایند یادگیری عوامل تشخیص بدست می‌آید.

۵- نتایج

به‌منظور ارزیابی روش پیشنهادی در کشف ناهنجاری، در این بخش تنظیمات محیط شبیه‌سازی، مجموعه داده مورد استفاده، تحلیل ویژگی‌ها و قابلیت‌های روش پیشنهادی و مقایسه با

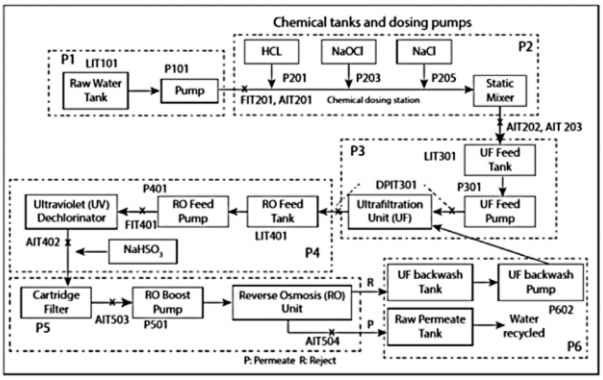
[1] Mahalanobis distance

رویکردهای پیشین مورد بررسی و آزمایش قرار گرفته است.

۵-۱- تنظیمات محیط آزمایش و مجموعه داده

روش پیشنهادی در زبان برنامه‌نویسی پایتون پیاده‌سازی شده و برای آموزش مدل‌های موازی آن از محیط Google Colab استفاده شده است. بستر آزمون یک سیستم پردازشی عام‌منظوره با پردازنده Intel Corei7 و حافظه 16GB بوده است که این پیکربندی در واحد کنترل نظارتی و جمع‌آوری داده (SCADA) سامانه‌های کنترل صنعتی قابل دستیابی است. باتوجه به اهمیت مجموعه داده در مطالعات مبتنی بر یادگیری عمیق، جمع‌آوری حجم داده زیاد از سامانه‌های کنترل صنعتی مستلزم آلوده کردن آن‌ها با هدف مشخص نمودن الگوی ناهنجاری است که منجر به خسارات زیادی می‌شود. درنتیجه در پژوهش‌های این حیطه عموما از مجموعه داده‌هایی که در بستر آزمایشی با درنظر گرفتن ویژگی‌های محیط صنعتی واقعی ساخته شده‌اند استفاده می‌شود. در این پژوهش از مجموعه داده SWAT[1] که مربوط به سامانه کنترلی تصفیه آب می‌باشد استفاده شده است[۴۵]. این مجموعه داده از ۵۱ حسگر و محرکه موجود در سامانه کنترل تصفیه آب در طی ۱۱ روز جمع‌آوری شده است. نمونه‌های جمع‌آوری شده، شامل خروجی حسگرها و وضعیت محرکه‌ها می‌باشد. این نمونه‌ها در دو دسته عادی و ناهنجار طبقه‌بندی می‌شوند که ۱۱/۹٪ از آن‌ها مربوط به بخش ناهنجار است.

سامانه کنترلی و نظارت بر پردازه‌های کنترلی صنعت تصفیه آب هدف در این مجموعه داده، ۷ روز عملکرد عادی و ۴ روز وضعیت تحت حمله برای جمع‌آوری نمونه‌ها را داشته است. داده‌ها از شش فرایند ورودی آب خالص اولیه (P1)، گندزدایی شیمیایی(P2)، فرافیلترسازی (P3)، دکلره به‌کمک لامپ‌های فرابنفش (P4)، خالص‌سازی (P5) و فرافیلترسازی غشا و تمیزسازی (P6) تشکیل شده است. این سامانه مجهز به یک بخش سایبری است که ارتباطات، تعامل انسانی، PLCها، SCADA و تاریخچه داده را مدیریت می‌کند. در این پژوهش، سناریوهای حملات انجام شده بر روی فرایند P1 مورد بررسی قرار گرفته‌اند. این فرایند شامل ۵ حسگر و محرکه 'MV101'، 'P101'، 'P102'، 'LIT101' و 'FIT101' می‌باشد که دو مورد آخر حسگر هستند. شکل ۱۲ بستر سامانه هدف را براساس اجزای شرح داده شده نشان می‌دهد.



شکل ۱۲. بستر آزمایش مجموعه داده SWAT [۴۵]

داده‌های هفت روز ابتدایی براساس رفتار هنجار سامانه است و در چهار روز بعدی ناهنجاری ناشی از ۴۱ نوع حمله لحاظ شده است. بر اساس بررسی داده حسگرها و با پیروی از پژوهش پیشین انجام شده در [۷،۴]، نمونه‌های اولیه مربوط به ۱۰۰،۰۰۰ گام اولیه حذف شده‌اند. این موارد مربوط به راه‌اندازی اولیه سامانه بوده و پایداری و قابلیت اتکای کافی را دارا نمی‌باشند. همچنین ۱۱ ویژگی که در طی زمان هیچ تغییراتی نداشته‌اند از روند یادگیری حذف شده‌اند. مجموعه داده مورد استفاده از دو بخش تشکیل شده است: ۱- ارتباطات عادی شبکه و ۲- ارتباطات تحت شرایط حملات. برای آموزش عامل‌های تشخیص از ارتباطات عادی شبکه استفاده شده است.

۵-۲- معیارهای ارزیابی

به‌منظور ارزیابی روش پیشنهادی معیارهای متداول دسته‌بندی شامل «دقت[2]» و «فراخوانی[3]» مورد استفاده قرار می‌گیرند که به‌صورت روابط زیر قابل تعریف می‌باشند:

$$Precision = \frac{TP}{TP + FP} \quad (۱۲)$$

$$Recall = \frac{TP}{TP + FN} \quad (۱۳)$$

که در این رابطه TP، تعداد مثب صحیح، FN تعداد منفی کاذب، TN تعداد منفی صحیح و FP تعداد مثبت کاذب است. همچنین رابطه امتیاز F_1 نیز برای تجمیع دو معیار فوق قابل تعریف است:

$$F_1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (۱۴)$$

همچنین معیار مورد استفاده برای آموزش شبکه‌های عصبی کدگذار-کدگشار (عامل‌های تشخیص) میانگین مجموع مربعات

[1] Secure Water Treatment

[2] Precision

[3] Recall

خطا می‌باشد که در رابطه زیر تعریف شده است:

$$MSE = \frac{1}{N}\sum_{k=1}^{N}\|\bar{y}_k^* - \bar{y}_k\|_2^2 \quad (۱۵)$$

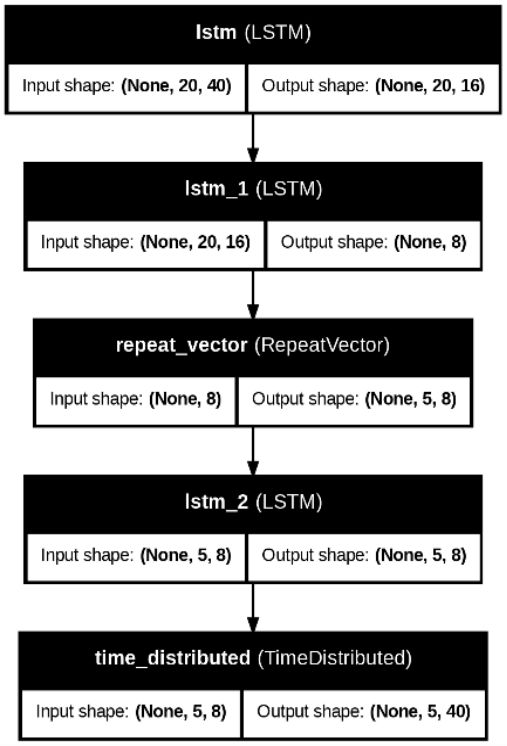
که در این رابطه N تعداد داده‌های آموزشی، $\bar{y}_k^*$ بردار خروجی مطلوب و $\bar{y}_k$ بردار خروجی شبکه عصبی می‌باشند.

۵-۳- آزمایشات و نتایج

به‌منظور ارزیابی روش پیشنهادی و بررسی قابلیت‌های آن در کشف ناهنجاری حسگرها، در این بخش دو دسته آزمایش درنظر گرفته شده است. اولین آزمایش به بررسی کارایی عوامل تشخیص پیشنهادی و تاثیر تجمیع گروه‌محور آن‌ها می‌پردازد. در دومین آزمایش، کارایی رویکرد پیشنهادی با روش‌های پیشین مقایسه می‌گردد.

۵-۳-۱- بررسی کارایی روش پیشنهادی در کشف ناهنجاری

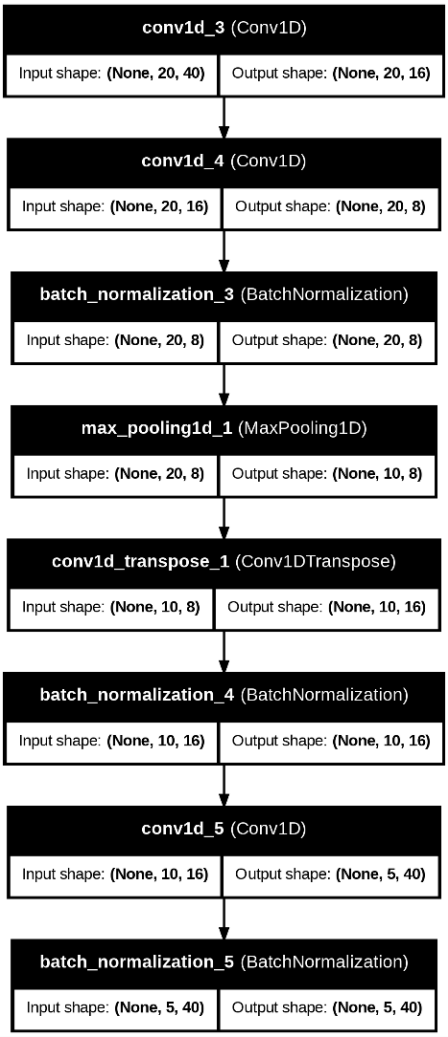
همانگونه که در بخش‌های پیشین شرح داده شد، روش کشف ناهنجاری پیشنهادی از ترکیبی گروه‌محور از سه عامل تشخیص می‌باشد. اولین عامل مبتنی بر ساختار عمیق حافظه بلندکوتاه مدت می‌باشد. شکل ۱۳ ساختار این عامل را نشان می‌دهد.



شکل ۱۳. ساختار عامل تشخیص مبتنی بر حافظه کوتاه بلندمدت

همانگونه که در شکل ۱۳ مشاهده می‌شود، ساختار عامل از دو بخش رمزگذار و رمزگشا با تعداد نورون محدود تشکیل شده است. در بخش رمزگذار، دولایه حافظه بلندکوتاه مدت ۱۶ و ۸ واحدی سیگنال ورودی ۴۰ بعدی در پنجره زمانی به طول ۲۰ را به فضای کد ۸ بعدی نگاشت می‌کنند. باتوجه به کمتر بودن طول پنجره خروجی (۵ واحد) بخش رمزگشا، صرفا از یک لایه ۸ واحدی به‌همراه لایه خروجی با ۴۰ نورون خطی تشکیل می‌شود.

عامل تشخیص دوم مبتنی بر یک کدگذار-کدگشای مبتنی بر لایه‌های پیچشی و ترانهاده پیچشی تشکیل شده است. این عامل در شکل ۱۴ نمایش داده شده است. تابع فعالیت لایه‌های مختلف به‌غیر از لایه خروجی تابع فعالیت ReLU است. بین لایه‌های پیچشی از لایه هنجارسازی دسته[1] به‌منظور استانداردسازی خروجی لایه‌ها استفاده شده است.



شکل ۱۴. ساختار عامل تشخیص مبتنی بر شبکه‌عصبی پیچشی و ترانهاده پیچشی

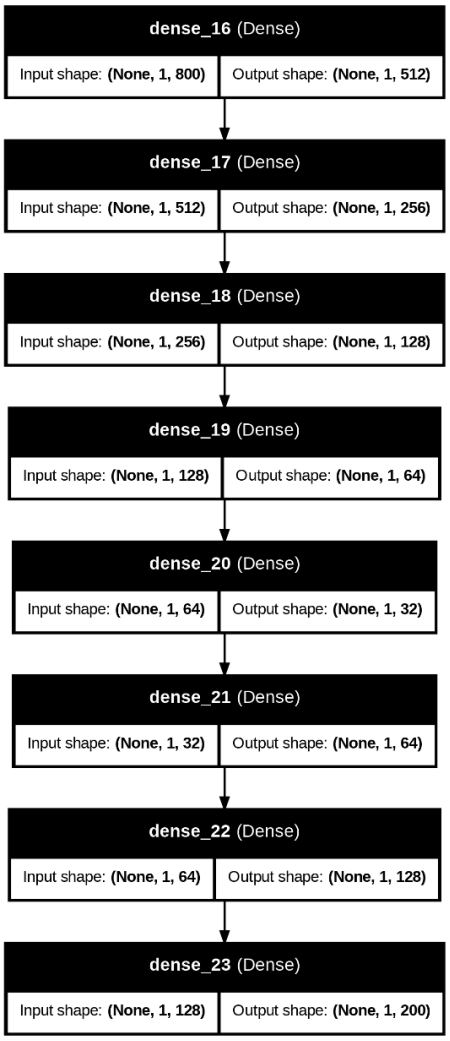
[1] Batch Normalization

نهایتاً شکل ۱۵ آخرین عامل تشخیص مبتنی بر ساختار تمام‌متصل را نمایش می‌دهد که ۲۰ واحد زمانی سیگنال ۴۰ بعدی را دریافت کرده و به یک بردار ویژگی ۸۰۰ بعدی تبدیل می‌کند. سپس بخش کدگذار، مبتنی بر لایه‌های تمام‌متصل با تابع فعالیت ReLU این بردار ۸۰۰بعدی را به فضای ویژگی ۳۲ بعدی در لایه گلوگاه می‌رساند. سپس بازسازی و پیش‌بینی ۵ واحد زمانی سیگنال ۴۰ بعدی (بردار ۲۰۰ بعدی) درخروجی انجام می‌شود. لایه خروجی مبتنی بر نورون‌های خطی است.

هر یک از روش‌های فوق با الگوریتم ADAM بر روی داده هنجار، آموزش داده می‌شوند. مشخصات فراپارامترهای یادگیری مورد استفاده در جدول ۱ ارائه شده است. پس از آموزش، برای هر یک از این رویکردها، آمارگان خطای بازسازی ۵ واحد زمانی مطابق با توضیحات ارائه‌شده در بخش ۴-۲-۲ در قالب یک توزیع گاوسی همبسته محاسبه می‌شود. برای تشخیص شباهت خطای بازسازی مطابق رابطه (۴) محاسبه شده و حد آستانه ۰٫۷ برای تشخیص به‌عنوان داده هنجار لحاظ می‌شود.

رویکرد گروه‌محور مبتنی بر استفاده از سامانه رأی‌گیر مشروح در بخش ۴-۲-۲ و رابطه (۵) استفاده می‌کند. مبتنی بر آموزش سه عامل تشخیص و RMSE آموزش هر روش، بردار وزن رابطه (۵) برای رأی‌گیر مطابق جدول ۲ خواهد بود.

جدول ۳ کارایی روش گروه‌محور پیشنهادی و سه عامل تشخیص (هریک به تنهایی) را با یکدیگر مقایسه می‌کند. همانگونه که در این جدول مشاهده می‌شود، روش گروه‌محور پیشنهادی تقریباً مزیت هر سه روش در کنار هم را دارد و دارای بالاترین مقدار دقت، فراخوانی و امتیاز F1 می‌باشد. به‌دلیل نامتوازن بودن مجموعه داده و اهمیت تشخیص صحیح رفتار ناهنجار، در مسئله هدف این پژوهش معیارهای فراخوانی و امتیاز F1 از اهمیت بیشتری برخوردار می‌باشند. بهبود این معیارها منجر به افزایش نرخ مثبت صحیح و حفظ تعادل آن با کاهش نرخ مثبت کاذب می‌شود. همچنین نتایج نشان می‌دهند که نمونه‌های ناهنجار مختلف لزوماً توسط یک تحلیل قابل تشخیص نبوده و برخی بر اساس هم‌افزایی روش‌های مختلف با توجه به جوانب مختلف سیگنال، قابل تشخیص هستند.



شکل ۱۵. ساختار عامل تشخیص مبتنی بر شبکه‌عصبی تمام‌متصل

جدول ۱. فراپارامترهای آموزش

فراپارامتر	مقدار
معیار آموزش	MSE
نرخ یادگیری اولیه	۰٫۰۱
اندازه دسته (Batch)	۱۲۸
تعداد تکرار یادگیری	۴۰

جدول ۲. وزن‌های رأی‌گیر

وزن عامل	مقدار
حافظه بلند-کوتاه‌مدت	۲٫۳۹
پیچشی-ترانهاده پیچشی	۳٫۲۵
تمام‌متصل	۴٫۳۴

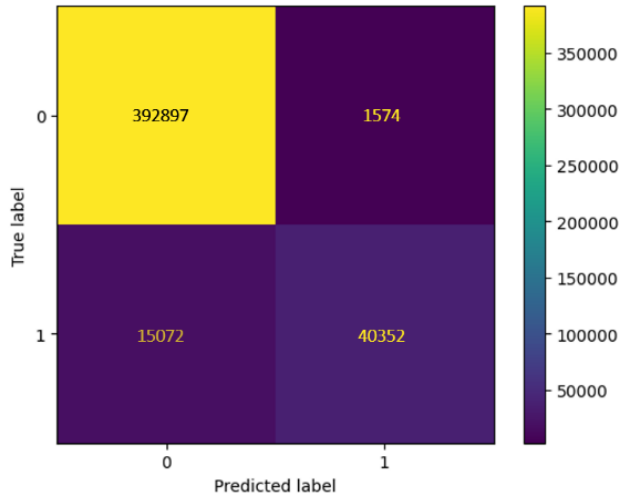
جدول ۳. مقایسه کارایی عامل‌های تشخیص بصورت مجزا و گروه محور

روش	دقت	فراخوانی	F1
پیچشی-ترانهاده پیچشی	۰٫۸۹	۰٫۶۸	۰٫۷۷
حافظه بلند-کوتاه‌مدت	۰٫۸۸	۰٫۶۵	۰٫۷۵
تمام‌متصل	۰٫۹۳	۰٫۶۵	۰٫۷۷
گروه‌محور	۰٫۹۶	۰٫۷۳	۰٫۸۳

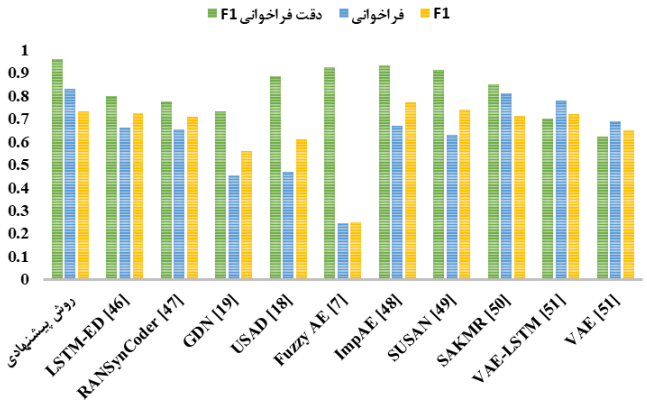
براساس نتایج ارائه شده، روش گروه‌محور پیشنهادی در این پژوهش معیارهای فراخوانی، امتیاز F1 و دقت را نسبت به روش‌های مجزای تشخیص به‌طور میانگین به‌اندازه ۱۰/۶٪، ۹/۷٪ و ۶/۷٪ بهبود داده است. همچنین به‌منظور بررسی جزئی‌تر روش پیشنهادی، ماتریس درهم‌ریختگی آن که دربرگیرنده تعداد تشخیص‌های مثبت و منفی کاذب و صحیح است در شکل ۱۶ نشان داده شده است که طبق آن، روش پیشنهادی عموم نمونه‌های هنجار را به‌درستی شناسایی کرده و در نمونه‌های ناهنجار عملکرد خوبی داشته است.

۵-۳-۲- مقایسه روش پیشنهادی با پژوهش‌های پیشین

به‌منظور ارزیابی قابلیت‌های روش پیشنهادی در کشف موثر ناهنجاری‌ها، عملکرد آن با پژوهش‌های پیشین مورد بررسی قرار گرفته است. در این راستا، روش‌های کشف ناهنجاری مبتنی بر تحلیل سری زمانی در سامانه‌های کنترل صنعتی هدف قرار داده شد و موثرترین آن‌ها برای مقایسه انتخاب شد. به‌منظور مقایسه کارایی، از معیارهای دقت، فراخوانی و F1 استفاده شده است. روش‌های برگزیده LSTM-ED [۴۶]، RanSynCoder [۴۷]،GDN [۱۹]، USAD [۱۸] ، Fuzzy-ED [۷]، ImpAE [۴۸]، SUSAN [۴۹]، SAKMR [۵۰]، VAE و VAE-LSTM [۵۱] می‌باشند که همگی مبتنی بر یادگیری عمیق بدون نظارت ارائه شده‌اند. شکل ۱۷ نمودار مقایسه روش پیشنهادی با ۱۰ روش برگزیده پیشین را در معیارهای بیان‌شده نشان می‌دهد. همانگونه که این شکل نشان می‌دهد، روش پیشنهادی دقت، فراخوانی و امتیاز F1 بالاتری نسبت به سایر پژوهش‌ها دارد. علت این امر، درنظر گرفتن موازی تحلیل‌های گوناگون زمانی، مکانی، محلی و سراسری در روش پیشنهادی است. همچنین همبستگی بین ابعاد به‌خوبی در این روش منعکس شده است. سایر روش‌ها عموما از یک رویکرد تحلیلی بهره برده‌اند و خاصیت اجماع در آن‌ها دیده نشده است.



شکل ۱۶. ماتریس درهم‌ریختگی روش گروه‌محور پیشنهادی



شکل ۱۷. مقایسه روش پیشنهادی با روش‌های پیشین

۶- جمع‌بندی و روال آتی

در این مقاله رویکرد خودنظارتی مبتنی بر یادگیری عمیق گروه‌محور به‌منظور کشف ناهنجاری در سامانه‌های کنترل صنعتی ارائه شد. سامانه‌های کنترل صنعتی شامل دستگاه‌ها، شبکه‌ها، سیستم‌ها و کنترل‌هایی است که برای بهره‌برداری فرآیندهای صنعتی استفاده می‌شوند. با پیشرفت فناوری و اتصال این سامانه‌ها به شبکه، حملات سایبرفیزیکی به‌عنوان یکی از آسیب‌پذیری‌های آنها مطرح است و رویکردهای کشف ناهنجاری با هدف مواجهه با این حملات در حین کار مطرح می‌شوند.

روش پیشنهادی در این مقاله، از دو مرحله پیش‌پردازش و تصمیم تشکیل شده است که در گام اول داده‌های خام حسگرها استانداردسازی و قاب‌بندی می‌شوند. در گام دوم، سه عامل تشخیص ناهنجاری مبتنی بر موثرترین رویکردهای یادگیری عمیق به‌صورت موازی بکارگرفته شده‌اند. این عامل‌ها با تکیه بر جنبه‌های گوناگون زمانی، مکانی، محلی و سراسری در فرایند پیش‌بینی دنباله، شباهت داده‌ها را با مقدار موردانتظار می‌سنجند و ناهنجار بودن آن را معین می‌کنند. عامل‌های لحاظ شده در روش پیشنهادی عامل‌های تشخیص ساختار یکپارچه رمزگذار رمزگشا دارند و از مدل‌های حافظه بلندکوتاه مدت، شبکه عصبی پیچشی و شبکه عصبی تمام‌متصل شکل داده شده‌اند. خروجی هر واحد تشخیص درجه هنجاربودن برای سیگنال ورودی بر مبنای تحلیل آماری خطای پیش‌بینی با استفاده از ساختار کدگذار-کدگشای خود و درنظر داشتن همبستگی زمانی-مکانی بین ویژگی‌ها می‌باشد. به‌منظور اجماع عوامل و استناج براساس آن‌ها،

نتایج عوامل تشخیص به یک رای‌گیر آستانه گیت وزن‌دار داده می‌شود که براساس قابلیت اطمینان هرواحد خروجی آن را در نتیجه نهایی تاثیر می‌دهد. واحد پیشنهادی در بخش تشخیص نفوذ سامانه کنترل نظارتی و جمع‌آوری داده که قابلیت پردازشی مناسبی دارد بکارگرفته می‌شود.

به‌منظور ارزیابی روش پیشنهادی، آزمایش‌های متعددی بر بستر سامانه کنترل صنعتی شبیه‌سازی شده تصفیه آب انجام شده است. در این آزمایش‌ها کارایی هریک از عوامل تشخیص و ترکیب گروه‌محور آن‌ها از نظر معیارهای دقت، فراخوانی و امتیاز F1 گزارش شده است. همچنین روش پیشنهادی با پنج پژوهش برگزیده پیشین که بهترین نتایج را داشته‌اند مقایسه شده است. طبق نتایج بدست آمده، روش پیشنهادی دقت روش‌های پیشین را به‌طور میانگین ۱۴٪ افزایش داده است. به‌عنوان روال آتی، اعمال یادگیری بر آستانه رای‌گیر و همچینن بکارگیری روش‌های یادگیری فعال در واحد تشخیص به‌منظور واردکردن بازخورد در تشخیص ناهنجاری پیشنهاد می‌شود.

مراجع

[1] E. Knapp, Industrial Network Security: Securing critical infrastructure networks for smart grid, SCADA, and other Industrial Control Systems. Elsevier, 2024

[2] R. Radvanovsky and J. Brodsky, Handbook of SCADA. Boca Raton Crc Press, 2016.

[3] B. Zhu and S. Sastry, "Scada-specific intrusion detection/prevention systems: a survey and taxonomy," in Proceedings of the 1st Workshop on Secure Control Systems (SCS), 2010.

[4] B. Kim, M. Alawami, E. Kim, S. Oh, J. Park, H. Kim, "A comparative study of time series anomaly detection models for industrial control systems," *Sensors*, vol. *23*, p. 1310, January 2023.

[5] M. Nawrocki, M, T. Schmidt, M. Wählisch, "Uncovering Vulnerable Industrial Control Systems from the Internet Core," In Proceedings of the IEEE/IFIP Network Operations and Management Symposium, Budapest, Hungary, 20–24 April 2020.

[6] A. Di Pinto, Y. Dragoni, A. Carcano, "The First ICS Cyber Attack on Safety Instrument Systems," In Proceedings of the Black Hat USA, Las Vegas, NV, USA, 4–9 August 2018.

[7] K. D. Gupta, K. Singhal, D. K. Sharma, N. Sharma, and S. J. Malebary, "Fuzzy Controller-empowered Autoencoder Framework for anomaly detection in Cyber Physical Systems," Computers & Electrical Engineering, vol. 108, p. 108685, May 2023.

[8] D. Pliatsios, P. Sarigiannidis, T. Lagkas, and A. Sarigiannidis, "A survey on SCADA systems: secure protocols, incidents, threats and tactics," IEEE Communications Surveys & Tutorials, vol. 22, pp.1942-1976, April 2020.

[9] Y. Yang, K. McLaughlin, T. Littler, S. Sezer, B. Pranggono, and H. Wang, "Intrusion detection system for IEC 60870-5-104 based scada networks," in Proceeding IEEE Power & Energy Society General Meeting, 2013.

[10] S. Alem, D. Espes, L. Nana, E. Martin, F. De Lamotte, "A novel bi-anomaly-based intrusion detection system approach for industry 4.0," Future Generation Computer Systems, vol. 145, pp.267-283, August 2023.

[11] F. Skopik, I. Friedberg, and R. Fiedler, "Dealing with advanced persistent threats in smart grid ict networks," in Proceeding Innovative Smart Grid Technologies Conference (ISGT), 2014.

[12] I. Friedberg, F. Skopik, G. Settanni, and R. Fiedler, "Combating advanced persistent threats: From network event correlation to incident detection," Computers & Security, vol. 48, pp. 35–57, 2015.

[13] F. Zhang, H. Kodituwakku, J. Hines, J. Coble, " Multilayer Data-Driven Cyber-Attack Detection System for Industrial Control Systems Based on Network, System, and Process Data," IEEE Transactions on Industrial Informatics, vol. 15, pp.4362-4369, January 2019.

[14] GR. MR, N. Somu, A. Mathur, "A Multilayer Perceptron Model for Anomaly Detection in Water Treatment Plants," International Journal of Critical Infrastructure Protection, vol. 31, p. 100393, December 2020.

[15] R. Khalil, N. Saeed, M. Masood, Y. Fard, M. Alouini, T. Al-Naffouri, "Deep learning in the industrial internet of things: Potentials, challenges, and emerging applications," IEEE Internet of Things Journal, vol. 8, pp. 11016-11040, 2021.

[16] H. Mao, M. Alizadeh, I. Menache, S. Kandula, "Resource management with deep reinforcement learning," in Proceedings of the 15th ACM Workshop on Hot Topics in Networks, 2016.

[17] Y. Lu, S. Chai, Y. Suo, F. Yao, C. Zhang, "Intrusion detection for Industrial Internet of Things based on deep learning," Neurocomputing, vol. 564, 2024.

[18] J. Audibert, P. Michiardi, F. Guyard, S. Marti, M. Zuluaga, "USAD: Unsupervised Anomaly Detection on Multivariate Time Series," In Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2020.

[19] A. Deng, B. Hooi, "Graph Neural Network-Based Anomaly Detection in Multivariate Time Series," In Proceedings of the AAAI Conference on Artificial Intelligence, 2021.

[20] Z. Li, Y. Zhao, J. Han, Y. Su, R. Jiao, X. Wen, D. Pei, "Multivariate Time Series Anomaly Detection and Interpretation using Hierarchical Inter-Metric and Temporal Embedding," In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021.

[21] A. Koay, R. Ko, H. Hettema, K. Radke, "Machine learning in industrial control system (ICS) security: current landscape, opportunities and challenges," Journal of Intelligent Information Systems, vol. 60, pp. 377-405, 2023.

[22] M. Nankya, R. Chataut, R. Akl, "Securing industrial control systems: components, cyber threats, and machine learning-driven defense strategies," Sensors, vol. 23, p. 8840, 2023.

[23] L. Yuan, X. Ya, C. Long, P. Guojun, Y. Danfeng "Deep Learning-Based Anomaly Detection in Cyber-Physical Systems: Progress and Opportunities," ACM Computing Surveys, vol. 54, pp. 1-36, 2021.

[24] W. Hilal, S. Gadsden, J. Yawney, "Financial fraud: a review of anomaly detection techniques and recent advances," Expert systems With applications, vol. 193, p. 116429, 2022.

[25] A. Sgueglia, A. Sorbo, C. Visaggio, G. Canfora, 'A systematic literature review of IoT time series anomaly detection solutions,' Future Generation Computer Systems, Vol. 134, PP. 170-186, 2022.

[26] A.Cook, G. Mısırlı, Z. Fan, "Anomaly Detection for IoT Time-Series Data: A Survey," IEEE Internet of Things Journal, December 2019.

[27] L. Erhan, M. Ndubuaku, M. Di Mauro, W. Song, M. Chen, G. Fortino, O. Bagdasar, A. Liotta, 'Smart anomaly detection in sensor systems: A multi-perspective review', Information Fusion,2020.

[28] A. Blázquez-García, A. Conde, U. Mori, J. Lozano, "A review on outlier/anomaly detection in time series data," ACM computing surveys (CSUR), vol. 54, pp. 1-33, 2021.

[29] M. Van Onsem, D. De Paepe, S. Hautte, P. Bonte, V. Ledoux, A. Lejon, S. Van Hoecke, "Hierarchical pattern matching for anomaly detection in time series," Computer Communications, vol. 193, pp. 75-81, 2022.

[42] E. Sisinni, A. Saifullah, S. Han, U. Jennehag, M. Gidlund, “Industrial internet of things: Challenges, opportunities, and directions,” IEEE transactions on industrial informatics, vol. 14, pp. 4724-4734, 2018.

[43] J. Yu, H. Yin, X. Xia, T. Chen, J. Li and Z. Huang, "Self-Supervised Learning for Recommender Systems: A Survey," in *IEEE Transactions on Knowledge and Data Engineering*, vol. 36, pp. 335-355, Jan. 2024.

[44] J. Gui, T. Chen, J. Zhang, Q. Cao, Z. Sun, H. Luo, D. Tao, “A Survey on Self-supervised Learning: Algorithms, Applications, and Future Trends,” IEEE Transactions on Pattern Analysis and Machine Intelligence, June 2024.

[45] A. Mathur, N. Tippenhauer, “SWaT: A Water Treatment Testbed for Research and Training on ICS Security,” In Proceedings of the International Workshop on Cyber-Physical Systems for Smart Water Networks, 2016.

[46] M, Macas, W. Chunming, “Enhanced Cyber-Physical Security through Deep Learning Techniques,” In Proceedings of the CPS Summer School PhD Workshop, 2019.

[47] A. Abdulaal, Z. Liu, T. Lancewicki, “ Practical Approach to Asynchronous Multivariate Time Series Anomaly Detection and Localization,” In Proceedings of the ACM SIGKDD Conference on Knowledge Discovery & Data Mining, 2021.

[48] M. Aslam, A. Tufail, L. Chandratilak De Silva, R. Anna Awg Haji Mohd Apong, A. Namoun. "An improved autoencoder-based approach for anomaly detection in industrial control systems." Systems Science & Control Engineering, vol. 12, 2024.

[49] A. Gómez, L. Fernández Maimó, A. Huertas Celdrán, F. García Clemente. "SUSAN: A Deep Learning based anomaly detection framework for sustainable industry," Sustainable Computing: Informatics and Systems, vol. 37, 2023.

[50] S. Tang, Y. Ding, M. Zhao, H. Wang, "SAKMR: Industrial control anomaly detection based on semi-supervised hybrid deep learning," Peer-to-Peer Networking and Applications, vol. 17, 2024.

[51] L. Pinto, L. Herrera, Y. Donoso, J. Gutierrez. "Enhancing Critical Infrastructure Security: Unsupervised Learning Approaches for Anomaly Detection." International Journal of Computational Intelligence Systems, vol. 17, 2024.

[30] C. Feng, T. Li and D. Chana, "Multi-level Anomaly Detection in Industrial Control Systems via Package Signatures and LSTM Networks," 2017 47th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), 2017.

[31] Y. Lai, J. Zhang, and Z. Liu, “Industrial Anomaly Detection and Attack Classification Method Based on Convolutional Neural Network,” Security and Communication Networks, vol. 2019, pp. 1–11, Sep. 2019.

[32] M. Kravchik and A. Shabtai, “Detecting Cyber Attacks in Industrial Control Systems Using Convolutional Neural Networks,” Proceedings of the 2018 Workshop on Cyber-Physical Systems Security and PrivaCy, 2018.

[33] A. Abdi, A. Ghasemi-Tabar, "ARAD: Automated and Real-Time Anomaly Detection in Sensors of Autonomous Vehicles Through a Lightweight Supervised Learning Approach," *IEEE Access*, vol. 12, pp. 90432-90441, 2024

[34] L. Yuan, X. Ya, C. Long, P. Guojun, Y. Danfeng “Deep Learning-Based Anomaly Detection in Cyber-Physical Systems: Progress and Opportunities,” ACM Computing Surveys, vol. 54, pp. 1-36, 2021.

[35] Y. Wu, H. Dai, H. Tang, H, “Graph neural networks for anomaly detection in industrial internet of things,” IEEE Internet of Things Journal, vol. 9, pp. 9214-9231, 2021.

[36] Y. LeCun, Y. Bengio, and G. Hinton, “Deep Learning,” Nature, vol. 521, pp. 436-444, 2015.

[37] A. Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. “O’Reilly Media, Inc.,” 2022.

[38] H. Mao, M. Alizadeh, I. Menache, S. Kandula, “Resource management with deep reinforcement learning,” in Proceedings of the 15th ACM Workshop on Hot Topics in Networks. ACM, 2016, pp. 50-56.

[39] Y. Lu, S. Chai, Y. Suo, F. Yao, C. Zhang, “Intrusion detection for Industrial Internet of Things based on deep learning,” Neurocomputing, vol. 564, 2024.

[40] Y. LeCun, Generalization and network design strategies, Technical Report, CRG-TR-89-4, University of Toronto, 1989.

[41] M.T. Jones, A beginner’s guide to artificial intelligence, machine learning, and cognitive computing, Technical Report, IBM, 2017.