



ANOMÁLIADETEKTÁLÁS MESTERSÉGESINTELLIGENCIA-MÓDSZEREKKEL – ÁTTEKINTÉS

ANOMALY DETECTION WITH ARTIFICIAL INTELLIGENCE METHODS – AN OVERVIEW

Sebestyén Pál György,¹ Hangan Lia-Anca,² Czakó Zoltán³

¹ Kolozsvári Műszaki Egyetem, Automatika és Számítástechnika Kar, Számítástechnika Tanszék, Kolozsvár, Románia, gheorghe.sebestyen@cs.utcluj.ro

² Kolozsvári Műszaki Egyetem, anca.hangan@cs.utcluj.ro

³ Kolozsvári Műszaki Egyetem, zoltan.czako@cs.utcluj.ro

Abstract

Nowadays, more and more human activities depend on computer-based automated systems. Fully automated (robotized) production lines, energy distribution infrastructures and other urban services or environmental surveillance systems are just some examples of cyber-physical systems that depend entirely on automated control systems. In these cases a significant challenge is to identify abnormal behaviors of the supervised or controlled systems, in order to avoid malfunction or sometimes catastrophic events. Our main research goal was to evaluate the potential of adapting and using AI techniques in the field of anomaly detection. We also developed a platform, called AutomaticAI, which can help specialists in different domains to identify the best approaches to solve a given anomaly detection problem. The platform can select the best AI algorithm and parameter configuration for a given set of data containing normal and abnormal data. The tool was used successfully in a variety of domains, from cyber-physical systems to the medical domain.

Keywords: *anomaly detection, artificial intelligence, outlier detection.*

Összefoglalás

Napjainkban egyre több emberi tevékenység függ a számítógépen alapuló automatizált rendszerektől. A teljesen automatizált (robotizált) gyártóvonalak, az energiaelosztó infrastruktúrák és más városi szolgáltatások vagy környezetvédelmi rendszerek csak néhány példa a kiberfizikai rendszerekre, amelyek teljes mértékben az automatizált vezérlőrendszerektől függenek. Ezekben az esetekben jelentős kihívást jelent a rendszerek rendellenes viselkedésének azonosítása a hibás működés vagy néha katasztrofális események elkerülése érdekében. Fő kutatási célunk az volt, hogy felmérjük a mesterségesintelligencia-technikák adaptálásának és használatának lehetőségeit az anomáliadetektálás területén. Ennek érdekében, fejlesztettünk egy platformot, melynek segítségével a különböző területeken dolgozó szakemberek hatékony megoldást találhatnak az anomáliák felderítésére. A platform tartalmaz egy eszközt, amely automatikusan kiválasztja a legjobb észlelési algoritmust egy adott adatkészlethez. Az eszközt számos orvosi és ipari területen sikeresen teszteltük.

Kulcsszavak: *anomáliadetektálás, mesterséges intelligencia, kiugró értékek detektálása.*

1. Bevezetés

1.1. Háttér és motiváció

A negyedik ipari forradalom (Ipar 4.0) korszakában különféle kiberfizikai rendszerek vesznek körül, mint például robotizált ipari folyamatok,

energia- és egyéb városi szolgáltatáselosztó infrastruktúrák, autonóm közlekedési eszközök, intelligens épületek és városok. Mindezeket a rendszereket számítógépek, számítógépes programok vagy számítógépes eszközök felügyelik és vezérlik. Néha az életünk biztonsága szempontjából

kritikus szolgáltatások és funkciók (pl. autonóm autók) teljes mértékben az ilyen autonóm rendszerektől függenek. De mi történik, ha valami elromlik, például egy érzékelő nem a megfelelő értéket adja meg, egy kritikus funkcióért felelős vezérlőberendezés meghibásodik, vagy a vezérelt rendszer viselkedése olyan módon változik, amelyet a tervezési fázisban nem láttak előre? Hogyan észleli ezeket az eseteket egy automatizált (pl. számítógépes) felügyeleti rendszer, majd hogyan reagál egy ilyen anomáliára?

A múltban, amikor a felügyeleti feladatokat emberek (pl. folyamatmérnökök) végezték, egy alapképés elegendő volt az ilyen esetek észleléséhez és az azokra való reagáláshoz. Egy automatizált rendszerben az anomáliák detektálása kötelező funkciója kellene, hogy legyen a vezérlőrendszernek. Ezért az elmúlt években az anomáliák detektálása fontos kutatási témává vált, és számos megoldást javasoltak és teszteltek.

Az anomáliák különböző típusúak lehetnek, a nagyon egyszerűektől (pl. egy megengedett intervallumot meghaladó érték), amelyek egyszerű küszöbérték-meghatározási vagy szűrési módszerekkel megoldhatók, egészen az összetettekig, amelyeket még egy emberi megfigyelő is nehezen definiálhat és észlelhet (pl. az éghajlatváltozás hatásai). Ez utóbbi esetben mesterségesintelligencia-módszerek használhatók a rendellenes viselkedés modellezésére és felismerésére.

1.2. Anomáliatípusok

Az "anomália" egyszerű definíciója a következő lenne: olyan adat vagy (rendszer) viselkedés, amely nagyon különbözik a többi adattól vagy észlelt viselkedéstől. Hawkins 1980-ban részletesebb meghatározást adott: „az anomália olyan megfigyelés, amely annyira különbözik más megfigyelésektől, hogy felveti a gyanút, hogy azt egy másik mechanizmus hozta létre” (pl. a rendszer meghibásodása vagy egy pénzügyi folyamatba való mesterséges beavatkozás).

Vannak más fogalmak is, amelyek hasonló vagy közeli jelentéssel bírnak az anomália szóval: abnormális (viselkedés), kiugró vagy deviáns.

Az anomália felismerésével kapcsolatban számos kérdés merül fel:

- mekkora legyen az eltérés ahhoz, hogy anomáliának lehessen minősíteni?
- az egyes adatokban jelen lévő természetes zajt nem szabad összetévesztetni az anomáliával;
- az anomália lehet nézőpont kérdése, vagy attól függ, hogy milyen kontextusban keletkezett

(pl. egy ázsiai populációban lévő európai anomáliának tekinthető);

- az emberek az oktatás és a tapasztalat alapján természetes képességgel rendelkeznek az anomália felismerésére, még akkor is, ha nincs racionális magyarázat rá; a számítógépek nem rendelkeznek ezzel a „természetes” képességgel.

Ami a lehetséges anomáliák típusait illeti, egy párat említhetünk:

- olyan adat/érték egy jel alakulásában, amely meghalad egy normálisnak tekintett intervallumot;
- a jel lineáris alakulásának változása;
- néhány sztochasztikus érték, amely nem követ egy bizonyos ismert valószínűségi eloszlást (pl. Gauss-eloszlás);
- néhány több paraméteres megfigyelés, amely nem tartozik semmilyen előre meghatározott normál esetnek minősített osztályhoz;
- bizonyos eltérések a jel periodicitásában (pl. aritmiák egy ECG-jelben).

1.3. Az anomáliadetektálás alkalmazási területei

Az anomáliák automatikus felismerése fontos feladat [1–4] az olyan új trendek alkalmazása esetén, mint az IoT, az IIoT és az Industry 4.0. Fontos a biztonság szempontjából kritikus kiberfizikai rendszerek esetében, ahol a rendszernek hibátűrő és öngyógyító megoldásokat kell tartalmaznia.

Az egészségügyben a különböző betegségek és ezekhez kapcsolódó tünetek az emberi test abnormális viselkedésének tekinthetők. Egy páciens fiziológiai paramétereinek változása által különböző betegségek előjeleit lehet felismerni. Általános anomáliadetektáló eszközök használhatók ebben az esetben is.

Az anomáliadetektálás minden olyan alkalmazásban hasznos, amely hatalmas mennyiségű adat feldolgozásával foglalkozik (ún. „big data”-alkalmazások). Ebben az esetben a manuális elemzések időkorlátok miatt nem kivitelezhetők; viszont néhány kiugró adat (anomália) befolyásolhatja az automatizált elemzés eredményét. Ezért az anomáliadetektálást a rendszer a hibás adatok kiküszöbölésére használhatja.

Automatizált anomáliadetektálásra van szükség azokban az esetekben is, amikor a változások annyira finomak, hogy az emberi érzékszervek nem észlelik őket. Például az iparban az elektromechanikai eszközök meghibásodásának korai felismerése a megelőző karbantartás alapja lehet.

Alább felsoroltunk egy pár olyan területet, ahol az anomáliadetektálás jelentős szerepet játszhat:

- gazdasági és pénzügyi alkalmazásokban, csatlások vagy gazdasági tendenciák változásainak felderítésére,
- ipari folyamatokban, hibásan működő eszközök vagy infrastruktúrák észlelésére, megelőző karbantartásra, riasztások generálására,
- az orvostudományban betegségek és tünetek kimutatására vagy kóros viselkedés kimutatására, pszichésen sérült személyek esetén,
- a környezet monitorizálásában, ahol paraméter-változásokat, szennyeződést, légszennyezést lehet észlelni,
- számítástechnikai eszközök és hálózatok elleni kibernetikus támadások felderítésére.

A paradoxon az, hogy ugyanaz az anomáliadetektálási technika számos tartományra és alkalmazástípusra alkalmazható. Ezért úgy gondoljuk, hogy szükség van egy általános anomáliadetektálási eszközre, amely a különböző tartományokban alkalmazható.

2. Az anomáliadetektálási technikák taxonómiája

Mivel sokféle anomália és anomáliadetektálási technika létezik, úgy gondoltuk, hogy hasznos lenne egyfajta besorolást vagy taxonómiát definiálni [5]. A szakirodalmat elemezve számos kritériumot azonosítottunk, amelyek felhasználhatók az osztályozáshoz:

- az anomália jellege alapján,
- az elemzett adatok jellege alapján,
- az értékelési módszerek alapján,
- az alkalmazások jellege alapján.

Az első kritérium szerint a következő típusú anomáliák azonosíthatók:

- egyponthoz tartozó anomáliák vagy kiugró értékek – olyan értékek, amelyek jelentősen eltérnek a többi értéktől; ilyenek például a rossz mérések által generált értékek, a közvetítés alatt történő hibák vagy túlzott zaj miatt eltérő értékek;
- kontextuális anomália – ahol egy érték/adat nagyon különbözik egy közelebbi kontextustól – egy összetettebb anomália, amely figyelembe veszi az érték szomszédságát; példák: eltérések a tözsdén, nulla alatti hőmérsékleti értékek nyáron;
- anomáliák az idősorokban – ahol hirtelen változások észlelhetők egymásután következő értékek között; pl. eltérések egy EKG-jelben.

Az elemzett adatok jellege alapján figyelembe vehetjük a megfigyelésben szereplő attribútumok

számát és a megfigyelések között fennálló korreláció típusát. Az első szempontból egyváltozós megfigyeléseink vannak (pl. hőmérséklet-változások), és többváltozós megfigyeléseink (pl. EEG-jelek). Az anomáliadetektáló eljárás ennek alapján egy vagy több változót vesz figyelembe. A második nézet alapján lehetnek:

- statisztikai korrelációkkal rendelkező adatok; az anomália ebben az esetben egy olyan megfigyelés, amely megszegi az adott korrelációt (pl. Gauss-szórás);
- időalapú korreláció – ahol az időben mintavételezett értékek között van valamilyen összefüggés (pl. folyamat-paraméterek);
- térbeli korreláció – ahol a szomszédos csomópontokban mért értékek valamilyen módon függnek egymástól (pl. egy adott régióban elszórt érzékelő által végzett hőmérsékletmérések);
- funkcionális korreláció – ahol a fizikai és kémiai törvények szigorú korrelációt határoznak meg a mért értékek között (pl. feszültség, áram és teljesítmény egy elektromos infrastruktúrában).

Az anomáliaészlelési módszerek kiválasztása esetén figyelembe kell venni a bemeneti adatkészletben jelen lévő ilyen típusú összefüggéseket.

Az anomáliák detektálásához különböző ismert általánosabb jellegű algoritmusokat és módszereket lehet alkalmazni. Némelyikük rendszerelemzési és jelfeldolgozási módszereken alapul, mint például a min-max, a küszöbalapú detekció, az interpoláció, a rendszerazonosítás, a Fourier- vagy a Laplace-transzformációk. Más módszereket a mesterséges intelligencia inspirált, mint például az osztályozás és a klaszterezés (pl. KNN, Kmeans, SVM, random forest stb.) vagy a neurális hálózatok. Ebben az esetben egy adott osztályozó egy tanulási folyamaton keresztül különbséget tehet a normál és a rendellenes megfigyelések között.

És végül, az alkalmazások típusa alapján megkülönböztethetünk olyan technikákat, amelyek offline módban alkalmazhatók, ahol nincsenek valós idejű korlátozások, és bármilyen összetett módszer alkalmazható, és az online mód technikáit, ahol szigorú időkorlátozások nehezítik a bonyolultabb módszerek használatát.

3. Egy szoftverplatform anomáliadetektáláshoz

Annak érdekében, hogy általános eszközt kínáljunk az anomáliadetektáláshoz, kifejlesztettünk egy platformot, amely tartalmazza mindazokat a hasznos eljárásokat, amelyek szükségesek az adatok feldolgozásában és elemzésében [6, 7].

Miért egy általános platform egy alkalmazott és adaptált megoldás helyett? Az ilyen megközelítés motivációi a következők voltak:

- számos észlelési technika létezik, és nehéz eldönteni a kezdetektől fogva, hogy melyik lesz a legjobb egy adott adatkészlethez;
- az anomáliadetektálás mellett általában más adat-előfeldolgozó és vizualizációs eszközökre is szükségünk van, amelyek egy platformban előgyártott formában találhatók;
- előfordulhat, hogy a különböző területek (fizika, kémia, biológia, környezet- vagy földtudományok) szakemberei nem rendelkeznek elegendő ismerettel a mesterségesintelligencia- vagy a jelfeldolgozó területeken, de anomáliadetektáló eszközökre van szükségük.

A platform egy konfigurálható eszköz, amelyben a felhasználó meghatározhat egy adatfeldolgozási folyamatot, amely különböző előfeldolgozási, besorolási és vizualizációs eljárásokat tartalmaz, amelyek egy adott adatkészlethez vagy alkalmazáshoz vannak igazítva. A platformot eredetileg anomáliadetektálásra fejlesztettük ki, de más célokra is hasznos lehet, ahol az adatkészleteket osztályozni kell.

3.1 AutomaticAI – eszköz a legjobb osztályozó algoritmus kiválasztásához

Ha csak a mesterséges intelligencia módszereit és azok variációit vesszük figyelembe, akkor azt tapasztaljuk, hogy nagyon sokféle választási lehetőség van, és sajnos nincs egyedülálló legjobb megoldás minden lehetséges alkalmazáshoz. A klasszikus (manuális) kutatási megközelítésben korlátozott számú módszert tesztelhetünk és hasonlíthatunk össze. Végül ezek közül választjuk ki a legjobb módszert. De lehetnek más módszerek is (amelyeket időkorlátok miatt nem teszteltünk), amelyek még jobban teljesíthetnek.

Ugyancsak gondot okoz és időt igényel az adott módszer optimális paramétereinek beállítása is; a kapott eredmények drasztikusan eltérhetnek a beállítási paraméterek értéke miatt. A kutatónak ki kell választania a legjobb módszert egy többdimenziós keresési térből, ami nagyon időigényes folyamat lehet [8].

Ennek a kiválasztási és paraméterezési problémának a megoldására javasoltunk és implementáltunk egy automatikus módszert, amely a legkülönbözőbb lehetőségek közül próbálja megtalálni a legjobban teljesítő módszert [9].

Megoldásunk egy részecske-raj-optimalizálásra (particle swarm optimization) [10, 11] és szimu-

lált lágyítási módszerekre (simulated annealing) épülő optimalizálási technikán alapul.

Elvileg a módszer a következőképpen működik:

- első lépésben számos olyan részecskét definiálunk, amelyek különböző besorolási technikákat képviselnek különböző beállításokban;
- a következő ismétlődő lépésben, több korszakon keresztül a legjobb algoritmust és annak legjobb beállítását a részecske-raj-optimalizálási technikával választjuk ki; minden algoritmus betanítása az adott adatkészleten történik, és egy teljesítményfüggvény határozza meg, hogy melyik rendelkezik a legjobb minőségi paraméterekkel. A következő korszakban megmarad a legjobban teljesítő algoritmuskészlet, és mutációk (pl. paraméter-változások) révén új részecskék jönnek létre. A szimulált lágyítással a mutációs fok (pl. paraméter-változások) csökken, ahogy a szimulációs korszakok elmúlnak;
- végül a módszer bemutatja a legjobban teljesítő osztályozó és egyben anomáliadetektáló algoritmust és azok beállításait, valamint azok minőségi mérőszámait (pl. pontosság, precizitás, visszahívás, F1, zavart mátrixok).

A különböző területek különböző adatkészletein végzett kísérletek igazolták a javasolt módszer hatékonyságát; az automatizált módszer minden esetben észszerű időn belül olyan megoldást talált, amely más kutatócsoportok által javasolt módszerekkel összehasonlítható, minőségi paramétereket adott. A kísérletek azt is kimutatták, hogy nincs egyedülálló legjobb megoldás az összes vizsgált esetre. Ezért a legjobb megoldás megtalálásának folyamata szükséges lépés.

4. Kísérletek az AutomaticAI platformmal

Ebben a fejezetben számos kísérletet mutatunk be különböző területeken, amelyeket a javasolt AutomaticAI eszköz hatékonyságának validálása érdekében hajtottak végre.

4.1 Anomáliadetektálás a vízügyi infrastruktúrákban

Ebben a kísérletben az volt a cél, hogy találjunk egy automatikus módszert a víz minőségének meghatározására, és riasztásokat generáljunk szennyeződés esetén.

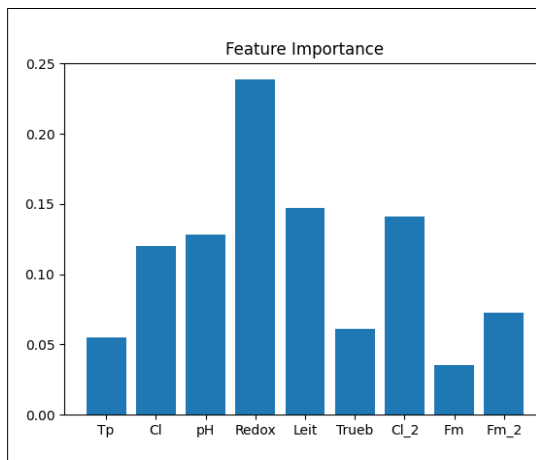
Ebben a kísérletben egy GECCO 2017 nevű adatkészletet használtunk [12], amely körülbelül 100 000 megfigyelést tartalmaz, és mindegyik megfigyelés 9 paramétert tartalmaz, amelyek relevánsak a víz minősége szempontjából (pl. hőmérséklet, pH, vezetőképesség, klórtartalom, zavarosság stb.).

1. táblázat. A vízminőség osztályozásához kiválasztott algoritmusok (beállított paraméterek: *class_weight= 'balanced', max_depth=42, n_estimators=130*)

Algoritmus	F1-paraméter
RANDOM FOREST	99,92%
EXTRA TREES CLASSIFIER	99,81%
DECISION TREE	99,19%
MLP	99,49%
KNN	99,47%
One-Class SVM	81,46%
SGD CLASSIFIER	50,36%
LOGISTIC REGRESSION	49,66%
PASSIVE AGGRESSIVE CLASSIFIER	45,36%
RIDGE CLASSIFIER	37,34%

Az AutomaticAI algoritmus alkalmazása révén a legjobban teljesítő algoritmusokat kaptuk, a legjobb beállítással. Az **1. táblázat** az első 10 osztályozási algoritmus eredményeit mutatja be. Úgy tűnik, hogy ebben a kísérletben a „Random forest” bizonyult a legjobbnak a szennyezett, illetve tiszta víz felismerésében, a Ridge-osztályozó pedig a legrosszabbnak.

A platformban levő egyéb módszerekkel kimutatható, hogy melyik paraméternek volt a legnagyobb befolyása a döntésben. Az **1. ábra** ezt a függőséget mutatja be.



1. ábra. Jellemző fontosság a vízszondák osztályozásában

4.2. Koronavírus-fertőzés felismerése anomáliadetektálási módszerrel

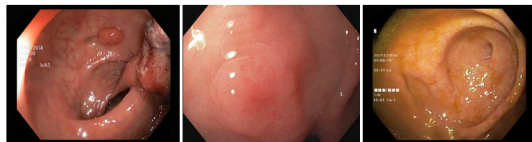
Ebben a kísérletben az volt a célunk, hogy a koronavírus-fertőzést a hagyományos vérvizsgálatból mutassuk ki anélkül, hogy speciális tesztekre (pl. PCR-tesztekre) lenne szükség **[13]**. Az ilyen megközelítés szükségessége nyilvánvaló volt a világjárvány első szakaszában, amikor a PCR-tesztek ritkák voltak.

Ebből a célból találtunk egy adatkészletet, amely 5000 beteg vérvizsgálatát tartalmazza. Egy külön oszlopban a mintákat Covid-fertőzöttnek vagy nem fertőzöttnek minősítették, egy specifikus teszt alapján. A vérminták tipikus paramétereit tartalmazták (pl. leukociták, neutrofilek, limfociták, monociták, eozinofilek, kor stb.). Az AutomaticAI-módszer alkalmazása után a detektálásra kiválasztott legjobb megoldás a Random Forest-osztályozó volt, amelynek paraméterei: *criterion='gini', min_sample_split=8* és *n_estimator=155*). Ennek az osztályozónak a minőségi paraméterei az adott adatkészlethez a következők voltak: Accuracy=98%, Precision=94% és Recall=97,8. Ezek az értékek összehasonlíthatók a speciális PCR-Covid-teszt eredményeivel.

4.3. Rák kimutatása vastagbélképekből

Ebben a kísérletben az volt a cél, hogy a kolorektális képeket úgy osztályozzák, mint amelyek rákos polipokat tartalmaznak, vagy sem. A **2 ábrán** látható mennyire nehéz különbséget tenni rákos polipot tartalmazó és nem tartalmazó képek között: az első két kép (balról jobbra) rákot tartalmaz, a harmadik meg nem.

Ebben a kísérletben egy előre betanított konvolúciós neurális hálózatot, a ResNet50-et használtuk, hogy releváns jellemzőket nyerjünk ki a képekből. Ezután az AutomaticAI segítségével meghatároztuk, hogy melyik osztályozó algoritmus teljesít a legjobban a rákképződést tartalmazó/nem tartalmazó képek megkülönböztetésében. Ebben a kísérletben a „Ridge classifier”-algoritmus bizonyult a legjobbnak, a következő minőségi paraméterekkel: Accuracy = 98.33%, Precision = 100% és a Recall = 76.64%.



2. ábra. Kolonoszkópiás képek; az első kettő rákkal, és az utolsó anélkül

Érdekes megemlíteni, hogy az első bemutatott kísérletben a Ridge Classifier volt a leggyengébben teljesítő algoritmus.

4.4. Anomáliadetektálás nyelőcső-mérésekben

Ebben a kísérletben [14, 15] akusztikus 2D-s képeket elemeztünk amelyeket a High Resolution Manometry (HRM) nevű eszköz segítségével kaptunk (3. ábra). Ezt az eszközt a nyelőcső rendellenes működésének kimutatására használják.

A cél ebben az esetben kettős volt:

- a szonda a nyelőcsőben való rossz helyzetének azonosítása,
- a nyelési folyamattal kapcsolatos betegségek osztályozása.

A HRM-készülékről kapott képeket minden esetben képzett szakemberek jelölték meg, mint rossz vagy jó pozícióban levő szonda. A jó pozíció esetén az orvos meghatározta a betegség típusát (ha létezett). Ezeket a szakember által jelölt képeket használtuk fel a tanulási folyamatban.

Ebben az esetben a nehézséget az jelentette, hogy viszonylag kevés címkézett kép áll rendelkezésre a tanuláshoz, és hogy néha még az emberi megfigyelő számára is nehéz volt különbséget tenni a szonda helytelen elhelyezése és egy adott betegség között. Emiatt a minőségi paraméterek nem olyan jók, mint az előző kísérletekben, de

ebben az összefüggésben még mindig nagyon jók. A szonda pozicionálási hibáinak észlelése esetén a kapott paraméterek: Precision 90,67%, az F1-paraméter pedig 84,21% volt. A betegségek osztályozásához az orvosi szakirodalomban található klasszikus (pontos) algoritmust használtuk.

5. Jövőbeli munka

Annak érdekében, hogy anomáliadetektálási platformunkhoz hozzáférést biztosítsunk a különböző területeken dolgozó szakemberek szélesebb kategóriájá számára, a platformot cloud-infrastruktúrán kívánjuk üzembe helyezni. Az első verzió sikeresen üzembe lett helyezve egy magán cloudba, de a nyílt hozzáférésű szolgáltatásokhoz publikus cloud-ra van szükség. A cloudra való üzembe helyezés kihasználhatja a magasabb számítási erőforrások előnyeit, ami hozzájárulhat az anomáliadetektálási algoritmus automatikus kiválasztásához és betanításához szükséges idő csökkentéséhez.

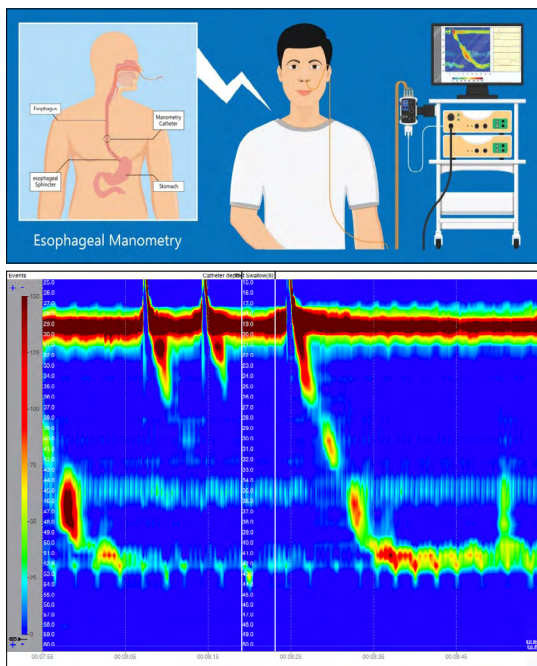
Emellett diverzifikálni kívánjuk azokat a domíniumokat, amelyekben az anomáliadetektálást és a platformunkat használják. Szándékunkban áll új mesterségesintelligencia-algoritmusokat, valamint jelfeldolgozási, -észlelési módszereket is hozzáadni a platformhoz.

6. Következtetések

Amint az ebben a tanulmányban látható, az anomáliák fontos szerepet játszanak az alkalmazások élettartamában. Az automatikus anomáliadetektálás egyre inkább egy automata rendszer kötelevő funkciója, főleg azokban az esetekben, amikor a rendszer egy kritikus szerepet tölt be életünkben. Fontos a hibatűrő és öngyógyító rendszerek számára is. Az anomáliadetektálási technikák különböző orvosi területeken is alkalmazhatók annak érdekében, hogy a betegségre utaló jeleket intelligens eszközökkel hatékonyabban felismerjük.

Egy adott terület szakemberének hatékony eszközre van szüksége az adatkészletek anomáliáinak azonosításához és kiküszöböléséhez. Ebből a célból kifejlesztettünk egy konfigurálható platformot, amely több, az adatelemzés és az anomáliadetektálás folyamatához szükséges eljárást tartalmaz.

A platform legfontosabb része az AutomaticAI-eszköz, amely automatikusan kiválasztja a legjobban teljesítő osztályozási algoritmust egy adott adatkészlethez. Ennek az automatikus módszernek a hatékonyságát és sokoldalúságát számos kísérlet bizonyította, amelyeket különböző terü-



3. ábra. HRM-mérőeszköz és kapott 2D-kép

leteken végeztek: víz-infrastruktúra, orvosi képek elemzése, Covid-felismerés stb.

Szakirodalmi hivatkozások

- [1] Chandola V., Banerjee A., Kumar V.: *Anomaly Detection: A Survey*. ACM Computing Surveys, 41/3. (2009) 1–58.
<https://dl.acm.org/doi/10.1145/1541880.1541882>
- [2] Agrawal S., Agrawal J.: *Survey on Anomaly Detection using Data Mining Techniques*. Procedia Computer Science, 60/1. (2015) 708–713.
<https://doi.org/10.1016/j.procs.2015.08.220>
- [3] Gupta M., Gao J., Aggarwal C. C., Han J.: *Outlier Detection for Temporal Data: A Survey*. IEEE Transactions on Knowledge and Data Engineering, 26/9. (2014) 2250–2267.
<https://doi.org/10.1109/TKDE.2013.184>
- [4] Hodge V. J., Austin J.: *A Survey of Outlier Detection Methods*. Kluwer Academic Publishers, 2004.
- [5] Czako Z., Sebestyén Gy., Hangan A.: *Colorectal image classification with transfer learning and auto-adaptive artificial intelligence platform*. In: Trends and Innovations in Information Systems and Technologies 28. Springer International Publishing, 2020. 534–543.
- [6] Czako Z., Sebestyén Gy., Hangan A.: *Automatic-AI-A hybrid approach for automatic artificial intelligence algorithm selection and hyperparameter tuning*. Expert Systems with Applications, 182. 2021.
<https://doi.org/10.1016/j.eswa.2021.115225>
- [7] Czako Z., Sebestyén Gy., Hangan A.: *Artificial Intelligence Algorithms Selection and Tuning for Anomaly Detection*. Computational Intelligence: International Joint Conference, IJCCI 2018 Sevilla, Spain. Revised Selected Papers, 2021.
- [8] Adankon M. M., Cherirt M.: *Model Selection for LS-SVM: Application to Handwriting Recognition*. Pattern Recognition, 42/12. (2009) 3264–3270.
<https://doi.org/10.1016/j.patcog.2008.10.023>
- [9] Sebestyén Gy., Hangan A., Czako Z., Kovács Gy.: *A taxonomy and platform for anomaly detection*. 2018 IEEE International Conference on Automation, Quality and Testing, Robotics (AQTR), 2018.
<https://doi.org/10.1109/AQTR.2018.8402710>
- [10] Pradeepmon T., Panicker V., Sridharan R.: *Parameter selection of discrete particle swarm optimization algorithm for the quadratic assignment problems*. Procedia Technology, 25. (2016) 998–1005.
<https://doi.org/10.1016/j.protcy.2016.08.199>
- [11] Simsek A., Kara R.: *Using Swarm Intelligence Algorithms to Detect Influential Individuals for Influence Maximization in Social Networks*. Expert Systems with Applications, 114. (2018) 224–236.
<https://doi.org/10.1016/j.eswa.2018.07.038>
- [12] Sebestyén Gy., Hangan A., Czako Z.: *Anomaly detection in water supply infrastructure systems*. 23rd International Conference on Control Systems and Computer Science (CSCS), 2021.
<https://doi.org/10.1109/CSCS52396.2021.00064>
- [13] Czako Z., Sebestyén Gy., Hangan A.: *COVID-19 Preliminary Patient Filtering based on Regular Blood Tests using Auto-Adaptive Artificial Intelligence Platform*. IEEE 16th International Conference on Intelligent Computer Communication and Processing (ICCP), 2020.
<https://doi.org/10.1109/ICCP51029.2020.9266277>
- [14] Popa Ș. L., Surdea-Blaga T., Dumitrașcu D. L., Sebestyén Gy., et al.: *Automatic Diagnosis of High-Resolution Esophageal Manometry Using Artificial Intelligence*. Journal of Gastrointestinal & Liver Diseases, 31/4. (2022) 383–389.
<https://doi.org/10.15403/jgld-4525>
- [15] Surdea-Blaga T., Sebestyén Gy., Czako Z., Hangan A., et al.: *Automated Chicago Classification for Esophageal Motility Disorder Diagnosis Using Machine Learning*. Sensors, 22/14. (2022).
<https://doi.org/10.3390/s22145227>