ÚJ MÉLYTANULÁSI MÓDSZEREK AZ OPTIKAI FELISMERÉS TERÜLETÉN

Doktori (Ph.D.) értekezés tézisei

készítette: Rádli Richárd Bence

Témavezető: Dr. Czúni László

Pannon Egyetem

Műszaki Informatikai Kar

Informatikai Tudományok Doktori Iskola

2026

1. Bevezetés

Jelen disszertáció három, egymástól elkülönülő számítógépes látási problémát tárgyal, melyek mindegyike jelentős gyakorlati alkalmazhatósággal rendelkezik a vizuális objektum felismerés, anomáliadetekció és a modellek tanításának területén.

A termékek vizuális ellenőrzése számos gyártási folyamatban gyakran elkerülhetetlen. Mivel a rendellenes elemek gyakran hiányoznak a tanítás során, a felügyelet nélküli anomáliaészlelés a legkézenfekvőbb megközelítés egy hibadetektáló rendszer készítéséhez. Ennek megfelelően az értekezés első felében az autoencoder hálózatok (AE) tovább fejlesztésének különböző módjait vizsgáltam. Összehasonlítottam és módosítottam egy meglévő SSIM-AE-t [2] a hálózat konvolúciós rétegeinek számának növelésével, valamint az alapmodell denoising autoencoderré alakításával. Továbbá megvizsgáltam a Spatial Transformer Network-ök (STN-ek) [4] integrációját is. Míg a konvolúciós rétegek és az STN súlyok egyidejű tanítása hatástalannak bizonyult a rekonstrukció szempontjából, észrevettem, hogy az STN előzetes tanítása a képorientáció normalizálására jelentősen javítja az autoencoder hibadetektálási teljesítményét.

A dolgozat második részében egy új megközelítést mutatok be mély neurális modellek osztályozási feladatokra történő felépítésére, iteratív, pszeudo-inverz optimalizációs technikát alkalmazva. A mély tanulási neurális hálózatok jelentős teljesítmény javulást mutatnak a sekély hálózatokhoz képest összetett problémák esetén. Fő hátrányaik a memóriaigény, az eltűnő gradiens probléma, valamint a legjobb létező súlyok és egyéb paraméterek megtalálásának időigényes folyamata. Mivel számos alkalmazás (például a folyamatos tanulás) gyors tanítást igényelne, az egyik lehetséges megoldás a nagyon gyorsan tanítható alhálózatok alkalmazása.

A stratégia mély, véletlenszerű struktúrák felépítését foglalja magában egymást követő fázisokban, ahol szignifikáns a valószínűsége, hogy az osztályozási teljesítmény javulni fog, vagy legalábbis nem fog csökkeni, a hálózat mélységének növekedésével.

A generált struktúra súlyait különálló fázisokban határozom meg, a *backpropagation* alkalmazása nélkül. Továbbá egy új *pruning* és súlycsere módszert vezettem be a hálózat szerkezetének és klasszifikálási teljesítményének további finomítása érdekében. A bemutatott módszer tanítási ideje – a legtöbb teszt esetén – jellemzően töredéke a népszerű gradiens alapú optimalizáló módszereknek, mint például az Adamnek vagy az SGD-nek.

A gyógyszerek használata világszerte széles körben és egyre növekvő mértékben elterjedt, különösen az idősek körében. Egyes országokban a felnőttek 66%-a használ vényköteles gyógyszereket, ami az életkorral növekszik: az 50 és 64 év közöttiek 75%-a, míg a 80 éves és idősebbek 91%-a érintett [3]. A szakpolitikai szervezetek állítása szerint a gyógyszer kiadagolási hibák a leggyakoribbak az egészségügyi ellátásban. Ennek eredményeként nagy az érdeklődés a gyógyszerhasználati folyamat biztonságának javítása iránt. Az optikai gyógyszer felismerés hasznos megoldás lehet ezen hibák minimalizálására, végső ellenőrzésként a betegeknek történő gyógyszerkiadáskor vagy önkiszolgálás esetén. Annak ellenére, hogy az AutoML megoldások könnyű kezelhetőséget biztosítanak a gyógyszerészek számára, ezen megoldások megbízhatatlannak bizonyultak klinikai környezetekben, ami okot szolgáltat robusztusabb és pontosabb alternatívák létrehozására. Ezért harmadik hozzájárulásom a gyógyszerészek munkájának támogatását hivatott segíteni a gyógyszerkiadás folyamata során.

Ezen a területen a fő hozzájárulásom egy már meglévő, gyógyszer felismeréshez alkalmazott többcsatornás, kétfázisú neurális hálózat továbbfejlesztése. Ez a módszer metrikus tanulást alkalmaz, amely hatékony módszer a kevés példával rendelkező objektumfelismerésre, ezáltal a gyógyszerek azonosítására. A korábban gyógyszer felismerési versenyt megnyerő többcsatornás neurális hálózat architektúráját [5] alapul véve módosítottam ezen modellt, így biztosíva, hogy az a gyógyszer jellemzők széles skálájából tanuljon. Részletesebben, az eredeti OCR ágat egy LBP ággal helyettesítettem, és a második fázisban alkalmazott *multi-head attention* rétegek

hatását is megvizsgáltam. A kutatás egyik legfontosabb eredménye egy új, továbbfejlesztett *triplet* veszteségfüggvény, amely szó vektor-beágyazásokat vagy bizonyos vizuális jellemzőket integrál annak érdekében, hogy a gyógyszerek közötti hasonlóságot beépítse a modellekbe. Ez az innováció pontosabb beágyazásokat eredményez, és felülmúlja a hagyományos triplet veszteségfüggvények teljesítményét. A javasolt megközelítést egy új gyógyszer adathalmaz felhasználásával validáltam, amelyet kollégáim segítségével hoztam létre [6].

2. A kutatási módszertana és felhasznált eszközök

A disszertációban bemutatott eredmények és javaslatok egységes kutatásmódszertani keretet követnek valamennyi téziscsoport esetében. Minden vizsgálatot egy gyakorlati, valós probléma motivált, amely a további elemzések alapjaként szolgált. Ezt követően átfogó szakirodalmi áttekintést végeztem, különös tekintettel a korszerű (state-of-the-art) módszerekre és azok korlátaira az adott alkalmazási területen.

Ezen megállapításokra építve új, illetve továbbfejlesztett megoldásokat terveztem és valósítottam meg, kifejezetten az egyes problémák sajátos kihívásaihoz igazítva. A folyamat magába foglalta különféle egyedi modellek fejlesztését, meglévő architektúrák adaptálását és finomhangolását, valamint saját adathalmazok létrehozását is. Emellett minden kísérlethez saját fejlesztésű kódbázist hoztam létre, ezáltal biztosítva a reprodukálhatóságot és a kiértékelések rugalmasságát.

Az egyes téziscsoportok esetében szisztematikus kiértékeléseket és teszteket végeztem, amelyek során a javasolt módszereket alaposan összehasonlítottam a baseline, valamint korszerű megközelítésekkel. Az értékelési eljárások során törekedtem fairnek és következetesnek lenni, amely jellemzően standardizált adatfelosztást, megfelelő teljesítménymutatók alkalmazását, valamint ismételt futtatásokat vagy keresztvalidációt foglalt magába.

A kísérleteket egy dedikált munkaállomáson végeztem, amely a következő hardverkonfigurációval rendelkezett: AMD Ryzen 5 5600X processzor, 16 GB videómemóriával ellátott NVIDIA RTX 5000 grafikus kártya, valamint 32 GB DDR4 rendszermemória. Operációs rendszernek a Windows 11 Pro-t és az Ubuntu 24.04 LTS-t választottam.

A szoftverkörnyezet Python 3.8-3.11 alapú volt, numerikus számításokhoz (NumPy), mélytanuláshoz (PyTorch) és számítógépes látáshoz (OpenCV) használt könyvtárakkal. A paraméterkereséshez a RayTune-t alkalmaztam. További hasznos könyvtárak, mint például a Pandas, Matplotlib, Scikit-Image és Scikit-Learn is

felhasználásra kerültek. A fejlesztés a PyCharm Professional fejlesztői környezetben történt.

Az **I. Tézis csoporthoz** tartozó keretrendszer két részre tagolódott. Az első részben négy autoencoder architektúrát értékeltem ki: egy standard SSIM-AE-t, ennek egy továbbfejlesztett változatát (SSIM-AEe) megnövelt konvolúciós rétegekkel, egy maszk blokkokkal betanított denoising autoencodert (SSIM-DAE), és annak megfelelő továbbfejlesztett verzióját (SSIM-DAEe).

Az autoencoderek teljesítményének kiértékeléséhez három adathalmazt használtam: a két textúrákat tartalmazó képi adathalmazt (Texture 1 és Texture 2) [2], amelyek mindegyike 100 hibátlan és 50 hibás képet tartalmazott, ez utóbbi képek különböző hibatípusokat mutattak be. Bevonásra került egy egyedi, általam készített CPU-t tartalmazó adathalmaz is, amely 60 darab CPU hátlap képből állt. Ezt az adatkészletet 48 tanító és 12 tesztképre osztottam fel, ahol a tesztképek az utóbbi három különböző hibaszimuláción estek át:

- CPUa: vizuálisan realisztikus megjelenésű kiegészítő komponenseket adtam a tesztképekhez;
- CPUc: a tesztképeket szintetikus szennyeződéssel terheltem;
- CPUm: egyes komponenseket (például lábakat vagy kondenzátorokat) eltávolítottam.

Ezeket a műveleteket mind a 12 tesztképen elvégeztem. A tanításhoz a hiperparaméter-beállítások az eredeti cikkben leírtakat követték.

A kísérletek második fázisához az MVTec adathalmazból származó Screw adathalmazt [1] adtam hozzá az előbb bemutatott Texture 1 és 2-höz. Ezen adathalmaz 320 hibátlan képet tartalmazott, mindegyik példány 1024×1024 méretű volt. A hibadetektálási kiértékelésekhez a *Thread-top* hiba alosztályt használtam, amely 23 képet tartalmazott. Négy különböző variációt vizsgáltam meg a tanítást illetően:

1. Referenciaként egy alap AE-t tanítottam be.

2. Egy AE és egy STN együttesen került betanításra, mindkettő véletlenszerű súlyokkal inicializálva.

3. Az STN-t előzetesen betanítottam a bemeneti minták orientációjának normalizálására, majd ezt követően az AE tanítása során az STN lokalizációs hálózatának paramétereit rögzítettem.

4. Ugyanaz, mint a 3. eset, de az STN súlyainak finomítását engedélyeztem az AE betanítása során.

Mindkét kísérleti fázisban a Structural Similarity Index Measure-t (SSIM) és a Mean Squared Errort (MSE) számítottam ki az autoencoderek rekonstrukciós pontosságának bemutatására.

A **II. Tézis csoport** kísérletei során 6 módszert hasonlítottam össze: két népszerű gradiens alapú optimalizálót – az Adamet és az SGD-t –, egy autoencoder alapú ELM-et, melynek neve H-ELM, és a javasolt módszerem három változatát.

Az UCI Machine Learning Repository 13 adathalmazát használtam fel. A kiválasztás során öt képi adathalmazt és nyolc hagyományos, nem képi adathalmazt választottam, amelyek mindegyike osztályozási feladatokat tartalmaz. Az adathalmazokat 70-15-15 arányban osztottam fel tanító-, validációs- és teszthalmazokra. Minden hálózati konfigurációban a validációs halmazt használtam a hiperparaméterek hangolására. Ezenkívül az Adam és SGD esetében a validációs halmazt használtam a tanítás során alkalmazott early stopping mechanizmus alkalmazására. Az előfeldolgozás során minden adathalmazt normalizáltam a jellemzők 0 és 1 közötti skálázásával.

Minden hálózattal minden adathalmazon tíz tesztet futtattam le, és átlagoltam az eredményeket. Az alkalmazott metrikák a tanítási és tesztelési pontosság, a tanítási és tesztelési F1-érték, valamint a tanítási idő voltak.

A **III. Tézis csoportban** először meglévő AutoML platformok és a népszerű objektum detektáló modell, a YOLO11 teljesítményét vizsgáltam a klinikai gyógyszerfelismerés területén, kettő kontrollált adathalmazon, melyeket én hoztam létre, továbbá valós környezetben gyűjtött adathalmazokon is. Bár olyan platformok, mint a Google Vertex AI, a Microsoft Azure Custom Vision és az Amazon Rekognition Custom Labels bizonyos tesztekben magas pontosságot értek el, azonban teljesítményük klinikai környezetben következetlennek és megbízhatatlannak bizonyultak.

E korlátok motiváltak arra, hogy tovább fejlesszek egy többcsatornás, kétfázisú neurális hálózatot, amelyet a fent említett platformokkal hasonlítottam össze két általam készített, laboratóriumi körülmények között gyűjtött adathalmazon, one-shot learning szerű tesztelési metódussal (ami azt jelenti, hogy a modell minden osztályhoz csak egyetlen referenciaképet kapott a tesztelés során). Ebben a felállásban a modell 76 gyógyszerosztályon tanult, majd 30 korábban nem látott osztályon és azok megfelelő képein került kiértékelésre.

Később a javasolt módszer teljesítményét két további adathalmazon is elemeztem: a nyilvánosan elérhető CURE képi adathalmazon, valamint egy saját, OGYEIv2 nevű adathalmazon, amelyet kollégáim segítségével készítettem.

Fontos megjegyezni, hogy ezekben a kísérletekben a CURE adathalmaz esetében egy- és kétoldalas teszteket is végeztem, míg az OGYEIv2 esetében csak kétoldalas teszteket. A kétoldalas teszt a gyógyszer mindkét oldalát ugyanahhoz a kategóriához tartozónak tekinti, míg az egyoldalas teszt a két oldalt két külön osztályként kezeli, ezáltal duplázva az osztályok számát.

A kiértékelés statisztikai robusztusságának növelése érdekében 5-fold cross validation-t alkalmaztam. Az értékelési folyamat során a metrikus tanulási modelleknél alkalmazott hagyományos módszertant követtem: a referencia- és felhasználói képeket metrikus beágyazási eljárásnak vetettem alá, majd a generált vektorokat euklideszi távolság segítségével hasonlítottam össze egymással a legjobb

egyezés megtalálása érdekében. Top-1 és Top-5 pontosságokat számítottam, átlagoltam, majd százalékos értékekké alakítottam mind az öt foldra nézve.

Két felhasználási esetet vizsgáltam a referenciahalmaz kiválasztásával kapcsolatban:

1. Hagyományosan a referenciahalmaz az adathalmaz összes gyógyszerosztályát tartalmazza. Ebben az esetben az inferencia során a lekérdezésekhez a felhasználói osztályok egyik foldját választottam ki.

2. A második típusú referenciahalmazban csak azokat az osztályokat szerepeltettem, amelyek a tényleges lekérdezési foldban vannak jelen, ezáltal biztosítva, hogy a modell azonos felhasználói és referenciaosztályokat hasonlítson össze.

A modell teljesítményének további elemzése érdekében további vizsgálatokat végeztem, szisztematikusan eltávolítva bizonyos alhálózatokat, a doménspecifikus jellemzők hatásának vizsgálata céljából. Továbbá kísérleteket is végeztem, amivel az egyes csatornák egyéni hozzájárulását értékeltem ki. Egy másik vizsgálat során feltártam a különböző fényforrások - felső vagy oldalsó - fontosságát a tabletta felismerésben.

3. Tézis csoportok

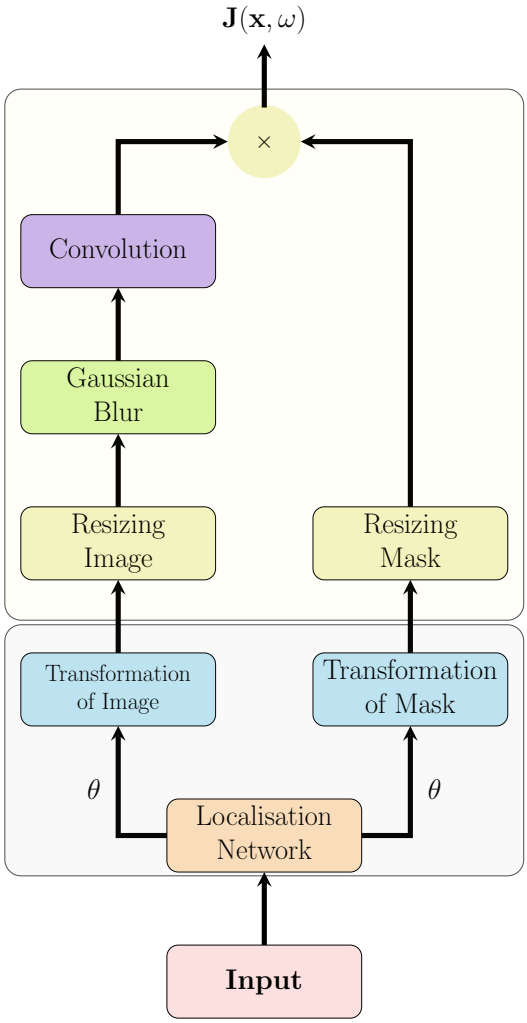
I. Tézis csoport *Autoencoder architektúrák továbbfejlesztése a vizuális hibadetektálási képességek javítása érdekében*

1.1 Ha egy autoencoder (AE) nagyobb pontossággal képes (anomália mentes) képeket reprodukálni, az nem feltétlenül jelenti azt, hogy jobb az anomália észlelésben.

Különböző SSIM AE-k teljesítményét vizsgáltam hibadetektálás céljából ismétlődő és fél-ismétlődő struktúrákon három adathalmazon. Bizonyos esetekben egy modell annak ellenére mutathat gyengébb anomália észlelési teljesítményt a Receiver Operating Characteristic görbe alatti terület (ROC AUC) tekintetében, hogy magasabb Structural Similarity Index Measurement (SSIM) és alacsonyabb Mean Squared Error (MSE) értéket ér el, mint egy másik modell. Ez a megfigyelés arra enged következtetni, hogy noha az SSIM és az MSE metrikák a jobb képhűséget és rekonstrukciós pontosságot jelzik, ezek nem mutatnak következetes összefüggést az anomáliaészlelés teljesítményével, amit a ROC AUC mutató tükröz.

1.2 Egy keretrendszert javasoltam annak bemutatására, hogy a Spatial Transformer Networks (STN-ek) betaníthatók a bemeneti képek orientációjának normalizálására. Az AE-k és STN-ek kombinációja javíthatja az AE-k anomáliadetektálási hatékonyságát.

Vizsgálataim során megfigyeltem, hogy az AE-k konvolúciós rétegeinek egyidejű tanítása az STN-ek súlyaival együtt nem eredményezett kielégítő rekonstrukciókat a dekóder részéről. Ehelyett a megközelítésem az STN kifejezett betanítására összpontosított, hogy normalizálja a bemeneti képek orientációját (ahogy az az 1. ábrán is látható) Ennek a stratégiának a hatékonyságát három különböző bemeneti adathalmazon értékeltem rekonstrukciós hiba és szabványos anomáliadetektálási metrikák segítségével.



1. ábra. STN egyedi veszteségfüggvénnyel

Kapcsolódó publikációk: [RR4] és [RR5].

II. Tézis csoport *Gyors és skálázható mély véletlenszerű neurális hálózatok*

2.1 Egy olyan keretrendszert javasoltam, amelyben mély véletlenszerű struktúrák egymást követő fázisokban épülnek fel. Az így létrejövő hálózat betanítása jelentősen gyorsabb, mint a hiba visszaterjesztési algoritmuson (backpropagation) alapú tanulás, miközben azonos számú neuron esetén a legtöbb tesztesetben pontosabb osztályozást eredményez.

Egy olyan keretrendszert fejlesztettem ki, amely lehetővé teszi tetszőleges számú rejtett réteg hozzáadását egy egyrétegű, előrecsatolt (feed-forward) neurális hálózathoz. Minden fázisban az új neuronok három kiegészítő stratégia alkalmazásával kerülnek beépítésre:

(a) a korábban betanított kimeneti neuronok közvetlen újrafelhasználásával;

(b) a korábban tanult kimeneti neuronok perturbálása additív Gauss-zaj hozzáadásával;

(c) véletlenszerű mintavételezésű ortogonális súlyokkal.

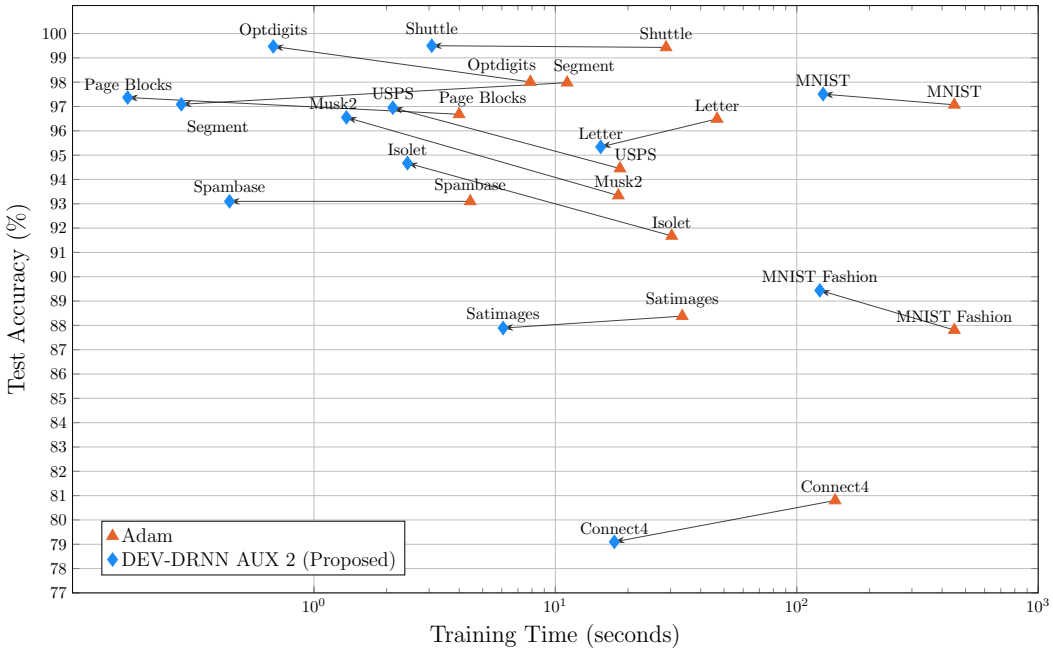
Ez a megközelítés lehetővé teszi a paramétertér hatékony feltérképezését, miközben megőrzi a korábbi fázisokban tanult hasznos reprezentációkat. Fontos megjegyezni, hogy ebben a keretrendszerben a neuronok száma rétegről rétegre exponenciálisan csökken.

2.2 Az előző keretrendszerre építve egy skálázható módszert javasoltam a modell legkevésbé hozzájáruló neuronjainak eltávolítására, majd ezek helyettesítésére a segéd hálózat(ok) leginkább hozzájáruló neuronjaival, ezzel tovább javítva az előző eredményeket.

Az általam javasolt iteratív finomítási módszer javítja a hálózat általános osztályozási teljesítményét azáltal, hogy megőrzi a kedvező véletlenszerű kapcsolatokat, miközben szisztematikusan azonosítja és lecseréli az

alulteljesítőket. A megközelítés segéd hálózatokat alkalmaz az értékeléshez, kiválasztva a leghatékonyabb kapcsolatokat, majd integrálva azokat az eredeti hálózatba. A módszer teljes mértékben skálázható, lehetővé téve tetszőleges számú segéd hálózat létrehozását és további rétegekre való kiterjesztését.

Az alap- és a továbbfejlesztett változatokat összehasonlítottam a két legelterjedtebb gradiensalapú optimalizálóval, az Adam és az SGD algoritmusokkal, valamint egy autoencodereket és a FISTA optimalizálót alkalmazó módszerrel 13 adathalmazon, ahol is versenyképes eredményeket mutattak. Egy teljesen összekapcsolt neurális hálózat (Adam optimalizálóval) és a javasolt DEV-DRNN AUX 2 vizuális összehasonlítása a 2. ábrán ábrán látható.



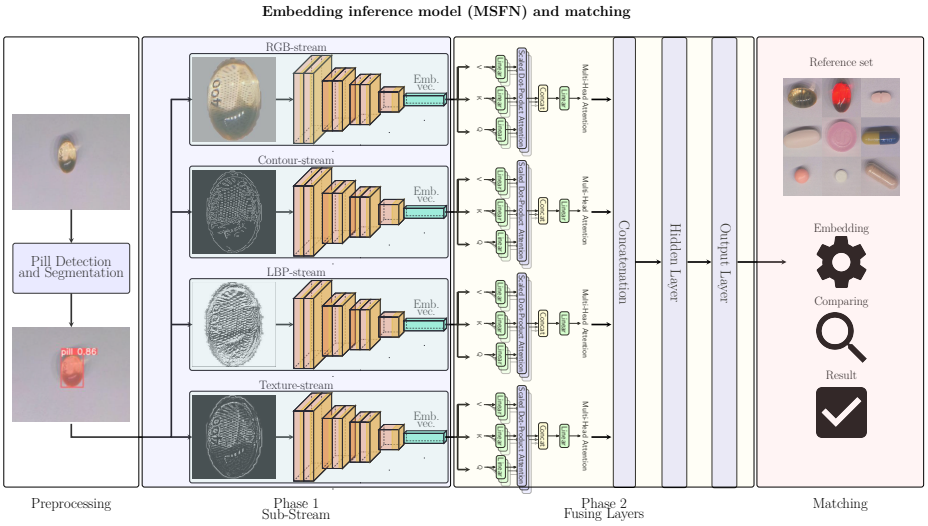
2. ábra. Az Adam optimalizálóval tanított teljesen összekapcsolt hálózatok, valamint a javasolt DEV-DRNN AUX 2 módszer teljesítményének összehasonlítása 13 adathalmazon (10 kísérlet átlagaként). A tanítási idő minden esetben rövidebb, míg 13 esetből 9-ben a tesztpontosság azonos vagy magasabb a javasolt megoldás esetén. Megjegyzendő, hogy a tanítási idő logaritmikus skálán van ábrázolva, ami jól szemlélteti a javasolt módszer jelentős előnyét.

Kapcsolódó publikációk: [RR3] és [RR6].

III. Tézis csoport *Metrika alapú gyógyszerfelismerés szöveges és vizuális jellemzők, valamint multi-head attention mechanizmus segítségével*

3.1 Egy többcsatornás, kétfázisú neurális hálózatot javasoltam, amely multi-head attention mechanizmust alkalmaz.

Egy többcsatornás, kétfázisú neurális hálózatot hoztam létre, amely az első fázisban az EfficientNetV2-Small hálózatot alkalmazza gerinchálózatként, míg a második fázisban multi-head attention modulokat vezet be, ahogy az a 3. ábrán is látható. A konvolúciós neurális hálózatok (CNN-ek) és a multi-head attention mechanizmus együttes alkalmazása javíthatja a meglévő keretrendszerek gyógyszerfelismerési képességeit. Más megközelítésekkel ellentétben ez a módszer nem igényel szemantikus hálózatokat, például OCR-t, ezáltal hatékonyabbá és függetlenebbé válik. Az értékelést két adathalmazon végeztem el, ahol a javasolt hálózat felülmúlta a konkurens hálózat teljesítményét.



3. ábra. A javasolt többcsatornás fúziós hálózat architektúrája.

3.2 A modell pontossága javítható magas szintű jellemzők bevitelével – legyenek azok szabad szövegből nyert információk vagy vizuális megjelenésből származó jellemzők – a többcsatornás metrikus modellek tanítása során.

Az előző architektúrát továbbfejlesztve két módszert mutattam be a modell osztályozási pontosságának növelése érdekében. Diszkriminatív jellemzőket generáltam: az egyiket természetes nyelvi szöveges forrásokból, a másikat pedig nagy felbontású referenciaképek vizuális leíróiból. Ezen magas szintű jellemzők integrálásával a modell hatékonyabban képes megkülönböztetni a minták közötti vizuális eltéréseket, javítva ezzel az összesített diszkriminatív teljesítményét.

3.3 Az általam javasolt módszert összehasonlítottam különböző AutoML platformokkal és a YOLO11-gyel, és bemutattam annak kiemelkedő teljesítményét egy one-shot learning teszten, kontrollált gyógyszerfelismerési adathalmazokon. Bár maga a kiértékelés erre a tesztre korlátozódott, az eredmények azt sugallják, hogy a javasolt módszer szélesebb körű alkalmazásban is megőrizné előnyeit.

Három széles körben használt AutoML platformot – a Google Vertex AI-t, a Microsoft Azure Custom Visiont és az Amazon Rekognition Custom Labelst – valamint a YOLO11 objektumdetektort tanítottam be és értékeltem ki egyedi gyógyszer-adathalmazokon, a tanítási és tesztelési osztályok szegregálása nélkül. Az AutoML platformok közül a Google Vertex AI érte el a legmagasabb tesztpontosságot, azonban mindegyik platform jelentős teljesítménybeli ingadozást mutatott klinikai tesztelési körülmények között. A saját módszerem előnyeinek bemutatására a többcsatornás, kétfázisú neurális hálózatot két kontrollált teszthalmazon értékeltem ki, ahol a tanítási és tesztelési osztályok teljes mértékben elkülönültek, és inferencia során osztályonként csak egy referencia-kép állt rendelkezésre. Ebben a one-shot learning tesztben a modellem kiemelkedően jobb eredményeket ért el, mint a többi megközelítés.

Kapcsolódó publikációk: [RR1], [RR2], [RR7], [RR8] és [RR9].

4. A tudományos eredmények alkalmazása

Ez a kutatás gyakorlati megoldásokat mutat be, amelyek különböző Ipar 4.0 technológiákra és iparágakra, például a gyártásra és az egészségügyre alkalmazhatóak.

Az **I. tézis csoportban** kifejlesztett autoencoder-hálózatok sokoldalúnak és hatékonynak bizonyultak különféle hibadetektálási alkalmazásokban. Megjegyzendő, hogy ezek a hálózatok közvetlenül beépítésre kerültek egy rendszerbe a 2018-1.3.1-VKE-2018-00048 (2019–21) számú projekt részeként, amelyet a Foxconn Hungary Kft. és az Eclipse Automation Hungary Kft. közreműködésével valósítottunk meg.

A **II. tézis csoportban** bemutatott módszer kiválóan alkalmazható olyan esetekben, ahol gyors tanításra van szükség. A tézisben foglalt módszer egy rendkívül hatékony osztályozómodult kínál, amely könnyen integrálható végső döntési rétegként nagyobb és összetettebb hálózati architektúrákba.

A **III. tézis csoportban** részletezett gyógyszerfelismerési módszerek közvetlenül alkalmazhatók valós környezetben. A megoldásaim sikeresen beépítésre kerültek egy nővérkocsi rendszerébe a 2020-1.1.2-PIACI-KFI-2021-00296 (2022–24) projekt keretében, az IVM Zrt.-vel együttműködésben. Ez a termék elnyerte a "Minőség-Innováció 2023" Nemzeti Díjat is. Ezen túlmenően a Pécsi Egyetemmel folytatott együttműködés keretében további felhasználási lehetőségeket vizsgálunk a kifejlesztett metrika-tanulási hálózat számára.

5. Kapcsolódó publikációk

Nemzetközi folyóiratok

[RR1] Rádli, R., Vörösházi, Z., & Czúni, L. (2024). Metric-based pill recognition with the help of textual and visual cues. *IET Image Processing*, 18(14), 4623–4638.

[RR2] Ashraf, AR., Rádli, R., Vörösházi, Z., & Fittler, A. (2026). Code-Based Versus AutoML Methods for Pill Recognition in Clinical Settings: Comparative Performance Study. *JMIR Medical Informatics*, 14 (1), e79160

[RR3] Rádli, R., & Czúni, L. (2026). Faster and Accurate: Developmental Deep Randomized Networks. *Neurocomputing*, Under review.

Konferencia kiadványok

[RR4] Rádli, R., & Czúni, L. (2021). About the Application of Autoencoders for Visual Defect Detection. In *29. International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Václav Skala - UNION Agency, pp. 181-188.

[RR5] Rádli, R., & Czúni, L. (2022). Improving the Efficiency of Autoencoders for Visual Defect Detection with Orientation Normalization. In *Proceedings of the 17th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2022) - Volume 4: VISAPP*, SciTePress, pp. 651-658.

[RR6] Rádli, R., & Czúni, L. (2023). Deep Randomized Networks for Fast Learning. In *International Conference on Learning and Intelligent Optimization*, Springer International Publishing, pp. 121-134.

[RR7] Rádli, R., Vörösházi, Z., & Czúni, L. (2023). Multi-stream pill recognition with attention. In *2023 IEEE 12th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, IEEE, Vol. 1, pp. 942-946.

[RR8] Rádli, R., Vörösházi, Z., & Czúni, L. (2023). Pill Metrics Learning with Multihead Attention. In *Proceedings of the 15th International Joint Conference*

on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KDIR, SciTePress, pp. 132-140.

[RR9] Rádli, R., Vörösházi, Z., & Czúni, L. (2024). Word and Image Embeddings in Pill Recognition. In *In Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 3: VISAPP*, SciTePress, pp. 729-736.

Egyéb hivatkozások

[1] Paul Bergmann, Michael Fauser, David Sattlegger, and Carsten Steger. Mvtec ad–a comprehensive real-world dataset for unsupervised anomaly detection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9592–9600, 2019.

[2] Paul Bergmann, Sindy Löwe, Michael Fauser, David Sattlegger, and Carsten Steger. Improving unsupervised defect segmentation by applying structural similarity to autoencoders. *arXiv preprint arXiv:1807.02011*, 2018.

[3] Lee Shirey Emily Ihara, Laura Summer. Prescription Drugs, September 2002. `https://hpi.georgetown.edu/rxdrugs/`.

[4] Max Jaderberg, Karen Simonyan, Andrew Zisserman, et al. Spatial transformer networks. *Advances in Neural Information Processing Systems*, 28:2017–2025, 2015.

[5] Suiyi Ling, Andreas Pastor, Jing Li, Zhaohui Che, Junle Wang, Jieun Kim, and Patrick Le Callet. Few-shot pill recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9789–9798, 2020.

[6] Richárd Rádli, József Bene, and Zsolt Vörösházi. OGYEIv2, 2024. `https://www.kaggle.com/dsv/8417535`.