

Állatorvostudományi Egyetem

Állatorvostudományi Doktori Iskola

Marek József Klinikai és
Élelmiszerlánc-biztonsági
Állatorvostudományok Doktori Program



Gépi tanulás alkalmazása, fejlesztése az
állattermék-előállítási folyamatok értékeléséhez

PhD-értekezés tézisei

Nagy Sára Ágnes

2023

Társtémavezetők

Dr. Solymosi Norbert
Bioinformatikai Központ
Állatorvostudományi Egyetem

Dr. Gábor György
Androvet Kft.

Bírálóbizottsági tagok

Dr. Cseh Sándor
egyetemi tanár
Állatorvostudományi Egyetem
elnök

Dr. Rátky József
egyetemi tanár
Állatorvostudományi Egyetem
belső tag (tartalék elnök)

Dr. Pollner Péter
egyetemi docens
Semmelweis Egyetem
bíráló (külső)

Dr. Maróti-Agóts Ákos
egyetemi docens
Állatorvostudományi Egyetem
bíráló (belső)

Dr. Négyessy László
tudományos tanácsadó
Eötvös Loránd Kutatási Hálózat,
Wigner Fizikai Kutatóközpont
külső tag

Dr. Tenk Miklós
egyetemi docens
Állatorvostudományi Egyetem
belső tag

Dr. Erdélyi Károly
tudományos főmunkatárs
Eötvös Loránd Kutatási Hálózat,
Állatorvostudományi Kutatóintézet
külső tag

Dr. Mackei Máté
egyetemi adjunktus
Állatorvostudományi Egyetem
belső tag (titkár)

Dr. Bodor András
tudományos munkatárs
Eötvös Loránd Kutatási Hálózat,
Wigner Fizikai Kutatóközpont
tartalék külsős tag

Készült 7 példányban, ez a(z) számú példány.

1. A doktori értekezés előzményei és célkitűzései

A népesség növekedése, a növekvő minőségi igények, a klimatikus viszonyok megváltozása egyre nagyobb kihívást jelent a mezőgazdaságban. A piac felől érkező folyamatosan változó minőségi és mennyiségi igények miatt az agrárium változásra kényszerül. Ezzel szemben az agráriumban dolgozók aránya folyamatosan csökken. Ennek következtében a gazdaságok állománylétszáma egyre növekvő tendenciát mutat. A mezőgazdasági dolgozók és a rájuk bízott állatlétszám aránytalansága miatt többek között az állategészségügyi feladatok (megfigyelés, ellátás) elvégzése is hiányt szenved. Számos szerző és mezőgazdasági szereplő az automatizációban és a precíziós állattartásban (precision livestock farming, PLF) látja a probléma megoldását. A precíziós állattartás során nagy mennyiségű adat keletkezik, aminek használható információvá alakításában „big data” elemzési módszereket használunk, melyek közül az egyik legelőreutatóbb a mesterséges intelligencia (artificial intelligence, AI).

Doktori munkámban egyrészt tanulmányoztam a PLF módszer tanába illeszthető az AI körébe tartozó, gépi tanulási modelleket az állattenyésztés folyamatában, másrészt a PLF hatáskörén túl az állatitermék-előállítás több pontján is alkalmaztam ezen eljárásokat.

Ehhez először a tejelő tehenészetekben fontos tömegtakarmány, a szilázs antimikrobiális rezisztenciagén-tartalmát (antimicrobial resistance gene, ARG) vizsgáltam. Feltételeztük, hogy a szilázsban a fermentáció során az ARG-k feldúsulnak az erjesztett tejtermékekhez hasonlóan. Mivel ennek leírása a szilázsra vonatkozóan nem található meg a szakirodalomban, elhatároztuk,

hogy szabadon hozzáférhető, újgenerációs szekvenálási shotgun metagenom-adatkészleteket vizsgálunk egy egységesített bioinformatikai eljárással. A cél szilázból származó ARG-k diverzitásának, bakteriális rokonságának és mobilitási potenciáljának vizsgálata volt.

Kutatásom céljai közt szerepelt továbbá Holstein-fríz fajtájú tehenek testkondíció pontjának (BCS) automatikus becslésére alkalmas algoritmusok fejlesztése. Azt vizsgáltam, hogy a felügyelt gépi tanulás, különösen a mély konvolúciós neurális hálózat (CNN) alapú Detectron2 modellek milyen pontossággal használhatók az általam becsült BCS-osztályok predikciójára egy egyszerű RGB-kamerával, tehenekről készült képek alapján.

Valamint tanulmányoztam szubklinikai tőgygyulladás (SM) korai detekciójának megbízhatóságát javító eljárásokat. Munkánkban azt vizsgáltuk, hogy a tejelő tehenekre vonatkozó, automatikusan gyűjtött, laktációval, a tej jellemzőivel kapcsolatos paraméterek és egy könnyen elérhető, indirekt szomatikus sejtszám (SCC) kimutatási módszer a California Mastitis Test (CMT) kombinációjával hogyan javítható az SCC-növekedés predikciója. Ebből a célból mesterséges neurális hálózatokat tanítottunk be az elérhető, automatizáltan rögzített paraméterek felhasználásával az SCC-emelkedés klasszifikációjára. Majd az így létrehozott modellt kombináltuk CMT-vel.

Továbbá akár az állati takarmányokban, akár az állati-eredetű élelmiszerekben előforduló, azokon keresztül fertőzni képes baktériumok növekedési potenciáljának vizsgálata szintén kutatásom tárgyát képezték. A baktériumszaporodás mértéke patogénitási és élelmiszerbiztonsági jelentőséggel bír. Ezért a baktériummennyiség időegység alatt való megnövekedésének vizsgálata fontos adatokat nyújthat.

Manuálisan annotált, szilárd táptalajú tenyészeteken tanítottunk konvolúciós neurális hálózatokat (CNN) a baktériumtelepek mi-

nél pontosabb detekciója céljából. A betanított CNN-ekkel, független, *Staphylococcus aureus* tenyészetekről készült képsorozatokon végzett predikciók alapján becsültük a baktériumtelepek méretét és növekedési sebességét.

2. Az értekezés új tudományos eredményei

1. TÉZIS: A vizsgált szilázmintákban talált ARG-k közül feltételezhetően 53 mobilis, melyből 10 iMGE-hez, 41 plazmidhoz és kettő profághoz köthető.

A vizsgált minták több, mint harmadában találtunk ARG-t. Számos baktériumfaj genomjában olyan ARG-eket találtunk, melyek már ismertek voltak a szakirodalmi adatokból, azonban több ARG fajspecifikus társulását először írtuk le.

Az *ant(6)*-la *A. amylophilus*ban, a *DfrG* *Gracibacillus* spp.-ben, az *ant(6)*-la *L. manihotivorans*ban, a *PoxTA* *L. paracasei*, *L. plantarum* és *L. brevis* fajokban, az *ant(9)*-la *L. acidipiscis*ben, a *VatE* *L. plantarum*ban, az *erm(44)v* *S. carnosus* és *S. pseudoxylosus* fajokban, az *aph(3')*-IIIa *T. halophilus*ban, az *ant(6)*-la *W. paramesenteroides*ben való leírására munkánkban került először sor.

Számos gén mobilitási tulajdonságai is újdonságnak számítottak. A *poxTA* gén és egy iMGE együttes előfordulására *L. paracaseiben*, valamint az *IsaE* gén profágokkal való társulására *S. pseudoxylosus* és *S. suis* baktériumokban nem találtunk példát a szakirodalomban.

Továbbá tudomásunk szerint még nem publikáltak a követke-

ző gének plazmidhoz kötött előfordulásáról az alábbi baktériumokban: az *aph(3')*-IIa gén *B. subtilis*ban; *dfrG* *Gracilibacillus* sp. SCU50-ben; *ant(9)-Ia* *L. acidipiscis*ben; *IsaE* *S. suis*ban; *qnrS1* *Cronobacter* sp. JZ38-ban, *erm(44)v* *S. carnosus*ban és *S. pseudoxylosus*ban, a *poxtA* *L. plantarum*ban és *L. brevis*ben a *vat(E)* *L. plantarum*ban, az *aph(3')*-IIIa *T. halophilus*ban.

2. TÉZIS: Az ANN predikciója a Cohen-féle kappa és a pontosság tekintetében a humán pontozókéhoz hasonló szintet ér el. A 3 osztályon betanított neurális háló predikciója jelentős arányban mutatott erős egyezést, ami jobb volt, mint a szakirodalomban talált humán értékelők közötti egyezés.

Az általunk alkalmazott neurális háló közel olyan megbízhatósággal képes BCS-t prediktálni, mint egy humán szakértő. Az elérhető egyezés 0 hibatarományban minimális, 0,25 esetén minimális vagy gyenge, míg 0,5-ös hibataromány esetén erős.

Ezzel szemben a gyakorlati pontszám-tartományokkal (sovány, normál, kövér) pontosabb osztályozás lehetséges, mint a területen alkalmazott nagy felbontású skálán.

Vizsgálataink alapján a farról készített felvételeken használt átfedésben lévő, de méretükben jelentősen eltérő annotációs boxok nem adnak lényegi eltérést a predikció minőségében.

Az ANN súlyaira vonatkozó eredmények nyilvánosan elérhetők. A modellünk szabadon felhasználható hasonló problémakörben előtanított modellként, meggyorsítva ezzel a tanítási folyamatot.

3. TÉZIS: ANN segítségével az SCC-emelkedést 0,54-es specificitással és 0,77-es szenzitivitással tudtuk kimutatni. Ha az általunk létrehozott ANN-t CMT-vel sorozatosan kombináljuk 20%-os prevalenciánál 55%-os, 30%-os prevalenciánál 39%-os pozitív prediktív érték emelkedést érhetünk el az individuális CMT-hez képest.

Kutatásunkban fejés során automatizáltan képződő adatokra (elektromos vezetőképesség, tej mennyisége, stb.) alapozott ANN szubklinikai tőgygyulladás detekciójára való képességét vizsgáltuk a pozitív prediktív érték lehetőség szerinti legmagasabb szinten tartásával. Az ANN predikcióit összevetve a ground truth-szal az eljárásunk szenzitivitása 0,54, a specificitása 0,77 lett. Továbbá vizsgálatunk során arra jutottunk, hogy az általunk létrehozott modell predikciós képessége jelentősen javul CMT-vel sorozatkombinációban.

4. TÉZIS: A referenciaközleményben leírt statisztikai megközelítéssel számítva, a CNN-ünk a növekedési sebességet 0,2%-os hibával reprodukálta.

Munkánkban bemutattuk, hogy ANN segítségével lehetséges a baktérium telepszámlálás automatizálása és azok növekedési dinamikájának elemzése. Modellünk segítségével az első telepek megjelenési ideje a kontroll és antibiotikummal kezelt csoportban közelített egymáshoz (a korábbi 8,4 h helyett 3,8 h). A növekedési sebesség tekintetében azt találtuk, hogy a kontroll csoport esetében csökken a szomszédok számának növekedésével. Ezzel szemben a rifampicinnel kezelt csoport esetében a

szomszédal nem rendelkezők kisebb sebességgel növekszenek, mint a szomszédal rendelkezők, ugyanakkor a szomszédok számának növekedése egyértelmű sebességcsökkenést jelez. Mindkét csoport növekedéséről elmondható, hogy 150 telepszámnál kisebb tenyészetekben mindig nagyobb a sebesség (kontroll: $7,7 \mu m/h$, rifampicin: $2,6 \mu m/h$), mint az összes tenyészet együtteséből számított érték.

A modell működéséről elmondható, hogy míg a kisebb telepeken a prediktált határoló boxok szűkebbek, jobban megközelítik a telepek határát, addig a nagy telepeknél jelentősen eltérhetnek attól. A tanító halmazban használt nagyobb számú hosszán inkubált telepről készült felvétellel a jövőben javítható.

3. Publikációs jegyzék

Makrai, L., Fodróczy, B., **Nagy, S.Á.**, Czeiszing, P., Csabai, I., Szita, G. and Solymosi, N., 2023. Annotated dataset for deep-learning-based bacterial colony detection. *Scientific Data* 10. 497.

Nagy, S.Á., Csabai, I., Varga, T., Póth-Szebenyi, B., Gábor, G. and Solymosi, N., 2023. Neural network-aided milk somatic cell count gain prediction. PREPRINT

Nagy, S.Á., Csabai, I., Varga, T., Póth-Szebenyi, B., Gábor, G. and Solymosi, N., 2023. Mesterséges neurális hálózatok az állatiternék-előállításban. *Magyar Állatorvosok Lapja* 145. 309-319.

Nagy, S.Á., Kilim, O., Csabai, I., Gábor, G. and Solymosi, N., 2023. Impact evaluation of score classes and annotation regions in deep learning-based dairy cow body condition prediction. *Animals*, 13(2), p.194.

Nagy, S.Á., Makrai, L., Csabai, I., Tózsér, D., Szita, G. and Solymosi, N., 2023. Bacterial colony size growth estimation by deep learning. *bioRxiv*, pp.2023-04. PREPRINT

Nagy, S.Á., Tóth, A.G., Papp, M., Kaplan, S. and Solymosi, N., 2022. Antimicrobial resistance determinants in silage. *Scientific Reports*, 12(1), p.5243.

Nagy, S.Á., Tózsér, D., Szombath, G., Baranyi, D., Reibling, T., Biksi, I. and Solymosi, N., 2017. Statisztikai ellenőrző diagramok az állatiternék-előállításban. *Magyar Állatorvosok Lapja* 139. 515–523.