University of Veterinary Medicine Budapest

Doctoral School of Veterinary Science

Application and development of machine learning for
the evaluation of animal production

PhD thesis

Sára Ágnes Nagy

2023

Supervisors

Dr. Norbert Solymosi
Centre for Bioinformatics
University of Veterinary Medicine
Budapest

Dr. György Gábor
Androvet Ltd.

Members of Evaluation Comittee

Dr. Sándor Cseh
professor
University of Veterinary Medicine
Budapest
chairman

Dr. József Rátky
professor
University of Veterinary Medicine
Budapest
inner member (reserve chairman)

Dr. Péter Pollner
associate professor
Semmelweis University

opponent (outside)

Dr. Ákos Maróti-Agóts
associate professor
University of Veterinary Medicine
Budapest
opponent (inner)

Dr. László Négyessy
science advisor
Eötvös Loránd Research Network,
Wigner Research Centre for
Physics
outside member

Dr. Miklós Tenk
associate professor
University of Veterinary Medicine
Budapest *inner member*

Dr. Károly Erdélyi
senior research fellow
Eötvös Loránd Research Network,
Veterinary Medical Research
Institute
outside member

Dr. Máté Mackei
assistant lecturer
University of Veterinary Medicine
Budapest *inner member*
(secretariat)

Dr. András Bodor
research fellow
Eötvös Loránd Research Network,
Wigner Research Centre for
Physics
reserve outside member

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

1. Background and objectives of the doctoral thesis

The growing population, increasing quality demands and changing climatic conditions represent an increasing challenge for agriculture. The constantly changing quality and quantity demands from the market are forcing agriculture to change. At the same time, the proportion of people working in agriculture is constantly falling. As a consequence, the average herd size is increasing. The disproportion between the number of farm workers and the number of animals they are responsible for is also leading to a lack of animal health services (monitoring and care). Automation and precision livestock farming (PLF) are seen by many authors and farmers as the solution to this problem. Precision livestock farming produces large amounts of data, which are transformed into meaningful information using 'big data' analysis, including one of the most advanced, the artificial intelligence (AI).

In my doctoral thesis, I have studied machine learning models in animal production that can be integrated into the PLF methodology, and I have also applied these techniques to several areas of animal production beyond the scope of PLF.

First, I investigated the antimicrobial resistance gene (ARG) content of silage, which is an important forage for dairy cows. It was assumed that ARGs are accumulated in silage during fermentation in a similar way to fermented dairy products. Since this is not described for silage in the literature, we decided to investigate freely available next-generation sequencing shotgun metagenome datasets using a standardised bioinformatics approach. The aim was to examine the diversity, bacterial identity and mobility potential of silage-derived ARGs.

The objectives of my study were also to develop algorithms for

the automatic estimation of the body condition score (BCS) of Holstein-Friesian cows. I investigated the accuracy of supervised machine learning, in particular deep convolutional neural network (CNN)-based Detectron2 models, for the prediction of my estimated BCS classes from a simple RGB camera image of cows.

I also studied procedures to improve the reliability of early detection of subclinical mastitis (SM).

In our work, we investigated how the combination of automatically collected parameters related to lactation, milk characteristics and an easily accessible indirect somatic cell count (SCC) detection method, the California Mastitis Test (CMT), can improve the prediction of SCC growth in dairy cows.

For this purpose, we trained artificial neural networks using accessible, automatically recorded parameters to classify SCC increase. We then combined the resulting model with CMT.

In addition, the growth potential of pathogenic bacteria occur in feed and food has also been the subject of my research. The degree of bacterial growth correlates with pathogenicity and food safety properties. Therefore, the study of bacterial growth over a unit of time can provide important data.

We have trained convolutional neural networks (CNNs) on manually annotated solid medium cultures to detect bacterial colonies as accurately as possible. The trained CNNs' predictions of bacterial colony size and growth rate were estimated from images of independent cultures of *Staphylococcus aureus*.

2. New scientific results of the thesis

THESIS 1: Of ARGs found in the silage samples we analysed, 53 were presumed to be mobile, of which 10 were linked to iMGE, 41 to plasmids and two to prophages.

ARGs were found in more than one third of the samples studied. In the genomes of several bacterial species, we found ARGs that were already known from the literature, but this is the first time that species-specific associations of several ARGs have been described.

We found *ant(6)-Ia* in *A. amylophilus*, *DfrG* in *Gracibacillus* spp., *ant(6)-Ia* in *L. manihotivorans*, *PoxtA* in *L. paracasei*, *L. plantarum* and *L. brevis*, *ant(9)-Ia* in *L. acidipiscis*, *VatE* in *L. plantarum*, *erm(44)v* in *S. carnosus* and *S. pseudoxylosus*, *aph(3')-IIIa* in *T. halophilus*, *ant(6)-Ia* in *W. paramesenteroides* were described for the first time in our work.

The mobility properties of several genes were also novel. The co-occurrence of the *poxtA* gene with an iMGE in *L. paracasei* and the association of the *lsaE* gene with prophages in *S. pseudoxylosus* and *S. suis* were not found in the literature.

Furthermore, to our knowledge, the plasmid-bound occurrence of the following genes in the following bacteria has not yet been published: the *aph(3')-IIa* gene in *B. subtilis*; *dfrG* in *Gracilibacillus* sp. SCU50; *ant(9)-Ia* in *L. acidipiscis*; *lsaE* in *S. suis*; *qnrS1* in *Cronobacter* sp. JZ38, *erm(44)v* in *S. carnosus* and *S. pseudoxylosus*, *poxtA* in *L. plantarum* and *L. brevis* in *vat(E) L. plantarum*, *aph(3')-IIIa* in *T. halophilus*.

THESIS 2: The ANN's prediction reaches a level of accuracy and Cohen's kappa similar to that achieved by human scorers. The prediction of the neural network trained on the 3 classes showed a significant rate of strong agreement, which was better than the agreement found between human scorers in the literature.

Our neural network can predict BCS with nearly the same reliability as a human expert. The achievable agreement is minimal for an error range of 0, minimal or weak for 0.25, and strong for an error range of 0.5.

In contrast, more accurate classification is possible with the practical score ranges (lean, normal, obese) than with the high-resolution scale used in the field.

Based on our tests, the use of overlapping bounding boxes of various sizes in the photos taken from the posterior area do not provide a significant difference in prediction quality.

The results on ANN weights are publicly available. Our model can be freely used as a pre-trained model for similar problems, thereby facilitating the learning process.

THESIS 3: Using ANN, we could detect SCC elevation with a specificity of 0.54 and a sensitivity of 0.77. Compared to individual CMT, the serial combination resulted in a 55% increase in PPV at a prevalence of 20% and 39% at a prevalence of 30%.

In our study, we investigated the ability of ANNs based on automatically obtained data (electrical conductivity, milk volume, etc.) to detect subclinical mastitis by keeping the positive predictive value as high as possible. Comparing the predictions of ANN with the ground truth, the sensitivity and specificity of our method were 0.54 and 0.77 respectively.

Furthermore, our study concluded that the predictive ability of our model is significantly improved with CMT in serial combination.

THESIS 4: Based on the statistical approach described in the reference publication, our CNN reproduced the growth rate with an error of 1%.

In our work, we demonstrated that ANN can be used to automatise bacterial colony counts and analyse their growth rate dynamics. Using our model, the time of first colony appearance in the control and antibiotic-treated groups has been approximated (3.8 h instead of 8.4 h).

Regarding the growth rate, we found that for the control group it decreases with the increase in the number of neighbours. In contrast, in the rifampicin-treated group, those without a neighbour grow at a lower rate than those with a neighbour, while the increase in the number of neighbours indicates a clear decrea-

se in growth. For both groups, growth rates in cultures with less than 150 colonies are always higher (control: 7.7 $\mu m/h$, rifampicin: 2.6 $\mu m/h$) than the growth rate of all cultures combined.

In terms of model performance, while in smaller colonies the predicted bounding boxes are narrower and more closely match the boundaries of the colonies, in large colonies they can deviate significantly. The larger number of long incubated colonies used in the training set can be improved in the future.

3. Publication list

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 állatitermék-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., Tôzsér, 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., Tőzsér, D., Szombath, G., Baranyi, D., Reibling, T., Biksi, I. and Solymosi, N., 2017. Statisztikai ellenőrző diagramok az állatitermék-előállításban. Magyar Állatorvosok Lapja 139. 515–523.