MAE Vet Fak Derg, 7 (3): 195-200, 2022
DOI: 10.24880/maeuvfd.1133145

Research Article

Classification of digital dermatitis with image processing and machine learning methods

Kürşad Yiğitarslan[1], İsmail Kırbaş[2]

[1]Department of Surgery , Faculty of Veterinary Medicine, Burdur Mehmet Akif Ersoy University, Burdur, Türkiye
[2]Application and Research Center of Digital Technologies in Livestock Sector Joint, Burdur Mehmet Akif Ersoy University, Burdur, Türkiye

Key Words:
digital dermatitis
image classification
image processing
machine learning
supervised learning

Received : 20.06.2022
Accepted : 27.10.2022
Published Online : 31.12.2022
Article Code : 1133145

Correspondence:
K YİĞİTARSLAN
(kyigitarslan@mehmetakif.edu.tr)

ORCID
K YİĞİTARSLAN : 0000-0003-4416-1597
İ KIRBAŞ : 0000-0002-1206-8294

This study was supported within the scope of the "Diagnosis and treatment of foot diseases in dairy cattle" project, which is the 5th subproject of the main project titled "Increasing the Sectoral Competitiveness of the Province of Burdur: Integrated Development by Differentiating in Agriculture" under the coordination of Burdur Mehmet Akif Ersoy University, Agriculture and Livestock Development Project Coordinator.

ABSTRACT

In this study, it was aimed to perform the detection and grading of Digital Dermatitis disease, which is common in dairy cattle and causes serious economic losses, using artificial intelligence techniques in a computer environment with high accuracy without the need for any expert intervention.

Within the scope of the study, pictures of lesions caused by Digital Dermatitis were taken, and four groups were formed according to the degree of size. These examinations were performed on 168 cows of the Holstein breed, aged 4-7 years, whose lameness was detected on dairy farms located in the centre and districts of the Burdur region. The photographs obtained were first labelled according to the degree of disease by a faculty member specialised in podiatry. Afterwards, the tagged photographs were reproduced using artificial intelligence image augmentation techniques, and a sample of 1,000 datasets was carried out for each disease degree. The photographs that make up the dataset were processed using the Inception v3 deep learning algorithm, and 2,148 numerical features were extracted. Then, machine learning models were developed using six different machine learning algorithms to classify these features. The results obtained were examined in detail with the help of tables and graphics, and they showed that the developed artificial intelligence models could be used in the classification of Digital Dermatitis case photos with a cumulative accuracy value above 0.87.

INTRODUCTION

Bovine Digital Dermatitis (DD), known by different names such as papillomatous digital dermatitis, Mortellaro's disease, or hairy heel warts, is a very common infectious foot disease (Biemans et al., 2017; Bruijnis et al., 2012). Treponema spp. plays a role in the occurrence of this disease, which is frequently seen in the hind feet (Clegg et al., 2015; Sogstad et al., 2005). This disease, which was first described in Italy in 1974, was observed to cause serious lameness in the cows in the herd (Biemans et al., 2017). The prevalence of DD is reported to be significantly higher in dairy cows (32.2%) than in beef cows (10.8%) (Hesseling et al., 2019). Digital Dermatitis causes a decrease in milk production, reproductive performance and animal welfare in dairy cows, as well as an increase in treatment costs, and the average cost per case of digital dermatitis is US$ 132.96 (Cha et al., 2010).

Digital Dermatitis is a multifactorial disease, even if Treponema spp. is accepted as the primary pathogen in the formation of the disease. The wet and dirty walking path contributes to the development of this disease (Holzhauer et al., 2008; Trott et al., 2003). In addition, pathogens such as Bacteroidetes, Fusobacteria, Tenericutes, Firmicutes, Proteobacteria, and Actinobacteria play a secondary role in digital dermatitis (Hesseling et al., 2019).

Lesions that are painful and prone to bleeding can be observed above the interdigital space and near the heels, along the coronary band. Filiform papillae may develop in the lesioned area, and the lesions can be surrounded by hyperkeratotic skin with longer than normal hair (Biemans et al.2018). For the classification of these lesions, the scoring system developed by Döpfer et al. is used, and the lesions are scored at four different levels (Döpfer et al., 1997). Döpfer et al. named the cases with 0.5-4 cm diameter and less than 2 mm depth from the epithelial tissue as granulomatous lesions as 1st degree. When this lesion is seen as a classic ulceration area that is deeper than 2 mm from the epithelial level and can reach up to 7 cm in diameter, it is considered grade 2. The appearance of

Bu makaleye atıfta bulunmak için/ To cite this article:
Yiğitarslan K. Kırbaş İ. Classification of digital dermatitis with image processing and machine learning methods. MAE Vet Fak Derg. 2022; 7 (3) : 195-200.

the lesion covered with crust in the healing period was classified as grade 3, and when skin lesions were hyperkeratotic and proliferative, they were classified as grade 4.

Today, with the development of computer hardware technology and artificial intelligence techniques, it has become possible to create information systems that can learn and process images without human intervention. In particular, recent developments in Convolutional Neural Networks and deep learning algorithms have been ground-breaking for digitising and classifying images. In our study, DD disease is considered as a classification problem, and the development of models that can classify with high accuracy by using 6 different machine learning algorithms and the performance comparison of these models are discussed in detail.

In the scope of the study, the problem was not only reduced to the binary classification problem of separating diseased individuals from healthy individuals, but it also aimed to determine the degree of the disease by utilising techniques from artificial intelligence. This was accomplished by reducing the problem to its simplest form. In machine learning studies, binary classification problems are generally problems where high performance can be achieved easily. Nevertheless, in most cases, the classification performance of the model deteriorates noticeably as the number of classes grows. With the help of photographs and various techniques that utilise artificial intelligence, the purpose of this study is not only to identify the disease but also to ascertain the severity of the disease that has been identified using as much precision as possible.

MATERIALS and METHODS

To collect data for the study, 206 photographs of the lesions observed on the hind legs of 168 cows of DD disease, Holstein breed, aged 4-7 years, as determined by examinations conducted in the Burdur region, were taken, graded based on the severity of the lesions, and assigned to the appropriate class. Degree 1 represents the mildest form of the disease, while degree 4 represents the most severe case. After the photographs were grouped, there were 60 photographs in the 1st Degree, 71 in the 2nd Degree, 56 in the 3rd Degree, and 19 in the 4th Degree, respectively. Photographic examples of each classification are depicted in Figure 1.

In classification-based machine learning problems, the equal distribution of the amount of data in the classes that make up the data set among the classes is extremely critical to obtaining successful results. For this reason, to create a balanced data set and increase the number of examples, the data set consisting of original photographs was reproduced with 1,000 examples in each class using artificial intelligence techniques. 2,148 numerical features were obtained by using the Inception V3 algorithm, which is a widely used deep learning algorithm in image processing of the reproduced dataset. The numerical data set obtained after this transformation process was classified by running with AdaBoost (AB), Naive Bayes(NB), Stochastic Gradient Descent (SGD), Support Vector Machine (SVM), Random Forest (RF), and Logistic Regression (LR) algorithms, which are widely used in machine learning problems.

The Naive Bayes method is one of the oldest approaches to machine learning. It is founded on the calculation of probabilities and uses simple models. The SVM method involves drawing the border lines between the clusters that represent the classes. It is important to remember that the distance between the cluster centres and the elements should be kept to a minimum, but the distance along the border line between the clusters should be kept to a maximum. The Random Forest algorithm is an example of a method that combines the operations of multiple independent tree algorithms, which allows for higher levels of accuracy to be achieved. The Logistic Regression algorithm is a type of algorithm that was developed before the artificial neural network approach. It makes use of the easily derived sigmoid function, decreases the amount of error produced with each iteration, works quickly, and is simple to train.

A method known as stochastic gradient descent is an example of an iterative approach to optimising an objective function that must have suitable smoothness properties. Since it replaces the actual gradient that was calculated from the entire data set with an estimate of the gradient that was calculated from a randomly selected subset of the data, it can be regarded as a stochastic approximation of gradient descent optimization. This is due to the fact that it calculates the estimate of the gradient from the data. This reduces the extremely high computational burden, which allows for faster iterations in ex-

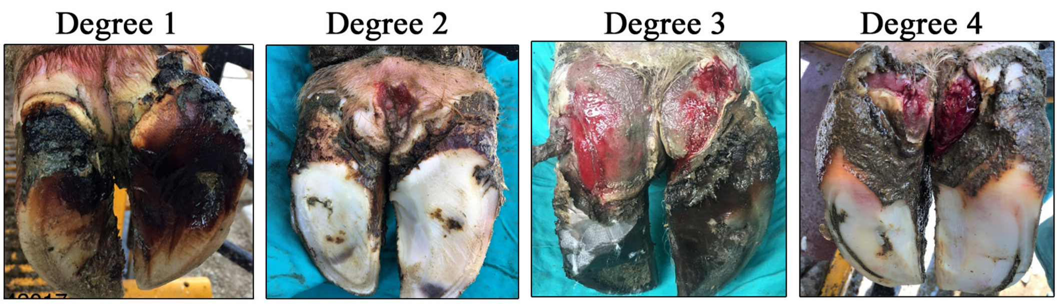

Figure 1. Photo examples representing the 4 different degrees of Digital Dermatitis.
Degree 1. Granulomatous lesion 0.5-4 cm in diameter and less than 2 mm deep from epithelial tissue
Degree 2. A classic ulceration area that is deeper than 2 mm and can reach 7 cm in diameter
Degree 3. lesion covered with crust and healing
Degree 4. skin lesions are hyperkeratotic and proliferative

change for a lower convergence rate. This benefit is especially noticeable when dealing with high-dimensional optimization problems.

AdaBoost, short for Adaptive Boosting, is a statistical classification meta-algorithm and can be used in conjunction with many other types of learning algorithms to improve performance. The final output of the boosted classifier is determined by combining the output of various weaker learning algorithms into a weighted sum. This sum represents the final output of the boosted classifier. AdaBoost is a classification meta-algorithm that, in order to improve performance, can be combined with a wide variety of other kinds of learning algorithms. Specifically, it can be used to improve neural network performance. The output of the other learning algorithms of weak learners is combined into a weighted sum that represents the final output of the boosted classifier. Usually, AdaBoost is presented for binary classification, although it can be generalised to multiple classes or bounded intervals on the real line.

Within the confines of the research project, the supervised learning method was implemented, and test and training data sets were generated by arbitrarily dividing the dataset into 80 percent and 20 percent. In the beginning, each model was initially trained with the help of the training dataset. The validation process utilised 10% of the training data set, and the success of the training was evaluated for each model independently. Figure 2 shows the stages of the classification process. Accordingly, the original data set was first reproduced by performing rotate, zoom in, and zoom out operations in the Python programming language to work with an equal number of samples. After augmentation, a data set was created by randomly selecting 1,000 samples for each class. The dataset created from augmented photographs has been subjected to visual feature extraction so that it can work with numerical data and can be run with widely used machine learning algorithms in the literature. Visual feature extraction is roughly the digitised version of the layer outputs obtained before the fully connected layer in the last layer of a deep learning algorithm.

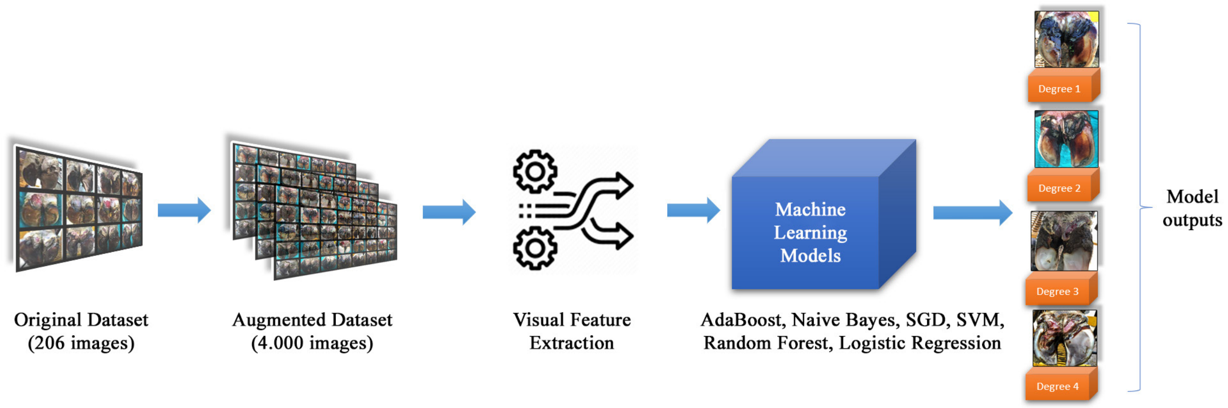

Figure 2. Stages of the artificial intelligence assisted medical photo classification process.

Table 1. Model performance metrics and equations for the classification metrics.

Performance Metric	Equation
Precision	$\frac{TP}{TP+FP}$
Recall	$\frac{TP}{TP+FN}$
Specificity	$\frac{TN}{TN+FP}$
Accuracy	$\frac{TP + TN}{TP + FN + TN + FP}$
F1 Score	$\frac{2. (precision.recall)}{precision + recall}$

TP: True positive, FP: False Positive, TN: True Negative, FN: False Negative

Figure 2 shows the steps of augmentation of the original data set, visual feature extraction, and then classification using machine learning algorithms.

There are many metrics commonly used in the literature to measure model performance. The top among these are Area Under Curve (AUC), Cumulative Accuracy (CA), F1-Score, Precision, Recall, and Specificity measurements. The performance metrics and formulas used in the study are given in Table 1.

RESULTS

The performance data of 6 machine learning models developed and trained within the scope of the study were compared using 6 classification metrics, and the obtained values are given in Table 2 and Figure 3 respectively.

Table 2. Numeric model performance results for all machine learning models.

Model	AUC	Cumulative Accuracy	F1	Precision	Recall	Specificity
AdaBoost	0.6358	0.4538	0.4537	0.4537	0.4538	0.8179
Naive Bayes	0.8167	0.5768	0.5780	0.5831	0.5768	0.8589
SGD	0.9073	0.8610	0.8599	0.8595	0.8610	0.9537
Random Forest	0.8009	0.5795	0.5799	0.5806	0.5795	0.8598
Logistic Regression	0.9778	0.8735	0.8731	0.8729	0.8735	0.9578
SVM	0.9806	0.8730	0.8724	0.8746	0.8730	0.9577

SGD: Stochastic Gradient Descent, SVM: Support Vector Machine, AUC: Area Under Curve.

Figure 3. Graphical model performance comparison for all machine learning models.
AUC: Area Under Curve, CA: Cumulative Accuracy.

Table 3. Numeric train and test time comparison for all machine learning models.

Model	Train time(s)	Test time (s)
AdaBoost	40.8760	2.4560
Naive Bayes	11.6720	2.7450
SGD	20.3750	4.2810
Random Forest	10.1840	1.8400
Logistic Regression	128.6910	2.2080
SVM	151.7910	21.0330

SGD: Stochastic Gradient Descent, SVM: Support Vector Machine

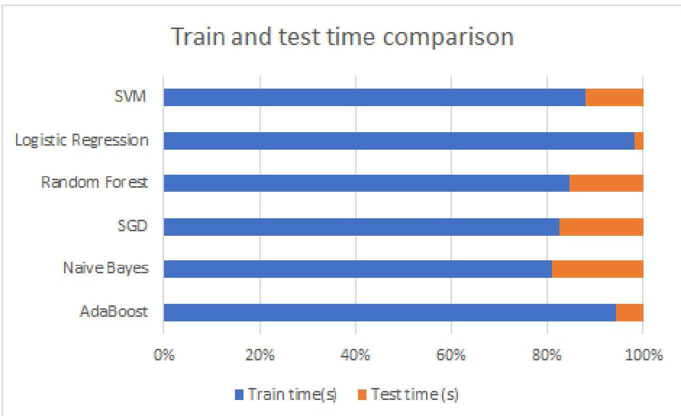

Figure 4. Graphical training and test time comparison for all machine learning models.

Accordingly, when the Cumulative Accuracy performance metric is taken as a basis, LR, SVM, and SGD models have achieved values above 0.86. These results show that all three models can produce a successful classification for this problem. When the models used in the study were evaluated on the precision metric, the LR, SVM, and SGD models showed significantly higher performance than the others. The same is valid for F1-score and recall metrics. Finally, when the specificity metric is considered, it is seen that the LR, SVM, and SGD models show the best performances, although there is no major difference between the models.

In addition to the performance of the model developed using machine learning techniques, the time spent for training and testing is also considered an important metric. In addition to the high classification performance of the developed model, it is desirable that the training and test run times be as short as possible. In Table 3, the training and testing times of the models covered in the study are given in seconds.

The numerical values given in Table 3 are visualized as a bar graph in Figure 4.

When the train and test time values are examined, the longest training time has emerged in the SVM and Logistic Regression models, respectively. The fastest trained model was the Naive Bayes model, with 11.67 seconds. While the longest test time was measured as 21 seconds in the SVM model, the model with the fastest test time was the Random Forest at 1.84 seconds.

DISCUSSION

It is stated that Digital Dermatitis disease, which is common in dairy cattle, is seen especially in Holstein cows (Demirkan et al., 2000). Even though there are cattle breeds like Simental and Montofon in the Burdur region, the fact that the cows diagnosed with DD are all of the Holstein breed lends credence to the study. Although there are cattle breeds such as Simental and Montofon in the Burdur region, the fact that the cows diagnosed with DD are Holstein breeds supports the study.

According to Hernandez et al.'s (2001) research, the aetiology of DD disease can be traced back to conditions such as inadequate hygienic conditions, improper care of the nails, and wet barn floors. The literature data are supported by the fact that comparable images were seen in the various places of business that were investigated as part of the scope of the study.

It is emphasised that DD disease is seen especially in the hind legs and the lesions are in the plantar region (Bassett et al., 2017). Similarly, the fact that DD disease was found in the hind legs and between the heels of cows in the study is consistent with the data that has been found in the previous research.

The review of the relevant literature revealed that there are no artificial intelligence medical image classification studies that have been developed for DD disease. This particular research stands out as an original piece of work due to the aforementioned consideration.

CONCLUSION

Within the scope of the study, the detection and evaluation of digital dermatitis disease, which is common and causes serious economic losses, through photographs was carried out with high accuracy by computer using artificial intelligence techniques.

In our application for the classification of medical images, 6 different machine learning models (AdaBoost, Naive Bayes, SGD, SVM, LR, and Random Forest) were developed and compared using 6 different classification metrics (AUC, CA, F1-score, precision, recall, and specificity). When the results are examined, it is seen that a very high classification success rate of 0.87 has been achieved. Accordingly, the most successful classification models stand out as LR, SVM, and SGD.

All three algorithms are classification algorithms with relatively short working and training times. In terms of the outcomes of the performance tests that were carried out, the differences between them are negligible at best. Although the success of the algorithms varies according to the problem and the data set, when the SVM algorithm is used for problems with high dimension input size, model training takes longer and computational resources are consumed more.

In situations where there is overlap between classes, the classification accuracy of the SVM algorithm is typically lower. The learning phase of the LR algorithm can typically be completed rapidly and with a reduced demand on the available resources of the computer. SGD was determined to be the model that required the least amount of time to train out of these three different models.

DECLARATIONS

EthicsApproval

The ethical approval decision is dated 16.03.2022 and numbered 869 taken from Burdur Mehmet Akif Ersoy University Local Ethics Committee of Animal Experiments.

Conflict of Interest

The authors declare that there have no conflict of interests.

Author ontribution

Idea, concept and design: KY, IK

Data collection and analysis: KY, IK

Drafting of the manuscript: KY, IK

Critical review: KY, IK

Data Availability

The data collected within the scope of the study has not been shared.

Acknowledgements

Not applicable

REFERENCES

1. Bassett, D. R., Toth, L. P., LaMunion, S. R., & Crouter, S. E. (2017). Step Counting: A Review of Measurement Considerations and Health-Related Applications. Sports Medicine. 47(7), 1303–1315. https://doi.org/10.1007/s40279-016-0663-1

2. Biemans, F., Bijma, P., Boots, N., & De Jong, M. (2017). Digital Dermatitis in dairy cattle: The contribution of different disease classes to transmission. Epidemics. 23, 76-84. https://doi.org/10.1016/j.epidem.2017.12.007

3. Bruijnis, M. R. N., Beerda, B., Hogeveen, H., & Stassen, E. N. (2012). Assessing the welfare impact of foot disorders in dairy cattle by a modeling approach. Animal: An International Journal of Animal Bioscience. 6(6), 962–970. https://doi.org/10.1017/S1751731111002606

4. Cha, E., Hertl, J. A., Bar, D., & Gröhn, Y. T. (2010). The cost of different types of lameness in dairy cows calculated by dynamic programming. Preventive Veterinary Medicine. 97(1), 1–8. https://doi.org/10.1016/j.prevetmed.2010.07.011

5. Clegg, S. R., Mansfield, K. G., Newbrook, K., Sullivan, L. E., Blowey, R. W., Carter, S. D., & Evans, N. J. (2015). Isolation of digital dermatitis treponemes from hoof lesions in Wild North American Elk (Cervus elaphus) in Washington State, USA. Journal of Clinical Microbiology. 53(1), 88–94. https://doi.org/10.1128/JCM.02276-14

6. Demirkan, I., Murray, R., & Carter, S. (2000). Skin diseases of the bovine digit associated with lameness. The Veterinary Bulletin. 70(2), 149–171.

7. Döpfer, D., Koopmans, A., Meijer, F. A., Szakáll, I., Schukken, Y. H., Klee, W., Bosma, R. B., Cornelisse, J. L., van Asten, A. J. A. M., & ter Huurne, A. A. H. M. (1997). Histological and bacteriological evaluation of digital dermatitis in cattle, with special reference to spirochaetes and Campylobacter faecalis. The Veterinary Record. 140(24), 620–623. https://doi.org/10.1136/vr.140.24.620

8. Garbarino, E. J., Hernandez, J. A., Shearer, J. K., Risco, C. A., & Thatcher, W. W. (2004). Effect of lameness on ovarian activity in postpartum holstein cows. Journal of Dairy Science. 87(12), 4123–4131. https://doi.org/10.3168/jds.S0022-0302(04)73555-9

9. Hernandez, J., Shearer, J. K., & Webb, D. W. (2001). Effect of lameness on the calving-to-conception interval in dairy cows. Journal of The American Veterinary Medical Association. 218(10), 1611–1614. https://doi.org/10.2460/javma.2001.218.1611

10. Hesseling, J., Legione, A. R., Stevenson, M. A., McCowan, C. I., Pyman, M. F., Finochio, C., Nguyen, D., Roic, C. L., Thiris, O. L., Zhang, A. J., van Schaik, G., & Coombe, J. E. (2019). Bovine digital dermatitis in Victoria, Australia. Australian Veterinary Journal. 97(10), 404–413. https://doi.org/10.1111/avj.12859

11. Holzhauer, M., Bartels, C. J. M., Döpfer, D., & van Schaik, G. (2008). Clinical course of digital dermatitis lesions in an endemically infected herd without preventive herd strategies. Veterinary Journal. 177(2), 222–230. https://doi.org/10.1016/j.tvjl.2007.05.004

12. Losinger, W.C. (2006). Economic impacts of reduced milk production associated with papillomatous digital dermatitis in dairy cows in the USA. The Journal of Dairy Research. 73(2), 244–256. https://doi.org/10.1017/S0022029906001798

13. Sogstad, A. M., Fjeldaas, T., Østerås, O., & Forshell, K. P. (2005). Prevalence of claw lesions in Norwegian dairy cattle housed in tie stalls and free stalls. Preventive Veterinary Medicine. 70(3-4), 191–209. https://doi.org/10.1016/j.prevetmed.2005.03.005

14. Trott, D. J., Moeller, M. R., Zuerner, R. L., Goff, J. P., Waters, W. R., Alt, D. P., Walker, R. L., & Wannemuehler, M. J. (2003).

15. Characterization of Treponema phagedenis-Like Spirochetes Isolated from Papillomatous Digital Dermatitis Lesions in Dairy Cattle. Journal of Clinical Microbiology. 41(6), 2522–2529. https://doi.org/10.1128/JCM.41.6.2522-2529.2003