



پیش‌بینی وقوع ورم پستان در گاو شیری با استفاده از الگوریتم‌های یادگیری ماشینی و داده‌های بیماری‌های حوالی زایش و شاخص‌های تولیدی شیر

سیده زهرا صحرانی^۱، مرتضی کیوانلو^۲، بابک خرمیان^۲، احسان الله افشاری صفوی^۲، محمد رحمانی منش^۳

^۱ دانش آموخته دکتری عمومی دامپزشکی دانشکده دامپزشکی دانشگاه سمنان

^۲ گروه علوم درمانگاهی، دانشکده دامپزشکی، دانشگاه فردوسی مشهد، مشهد، ایران

^۳ گروه آموزشی مهندسی نرم افزار کامپیوتر، دانشکده مهندسی برق و کامپیوتر، دانشگاه سمنان، ایران

mkeivanlou@um.ac.ir

زمینه و نوع مطالعه: از آنجا که ورم پستان یکی از پرهزینه‌ترین بیماری‌ها در صنعت گاو شیری است، شناسایی زودهنگام گاوهای در معرض خطر می‌تواند نقش مهمی در مدیریت اقتصادی گله ایفا کند. مطالعه حاضر با هدف ارزیابی قدرت پیش‌بینی الگوریتم‌های یادگیری ماشینی در پیشگویی وقوع ورم پستان بر اساس داده‌های رایج ثبت‌شده در گله، به صورت یک مطالعه گذشته‌نگر انجام شد.

هدف: هدف مقایسه عملکرد چند الگوریتم یادگیری ماشینی در پیش‌بینی وقوع ورم پستان بالینی و تحت بالینی در دوره‌های مختلف پس از زایش، بر اساس متغیرهای مرتبط با بیماری‌های حوالی زایش و شاخص‌های تولید شیر بود.

روش کار: داده‌های مربوط به ۲۹۰۵ گاو هلشتاین یک گله صنعتی طی دو سال جمع‌آوری شد. متغیرهای مستقل شامل شکم زایش، وقوع بیماری‌های حوالی زایش (سخت‌زایی، جفت ماندگی، متريت، ورم پستان و ادم پستانی در ماه اول) و شاخص‌های شیر ماه اول (میزان تولید، درصد چربی و پروتئین و SCC) بودند. سه مدل پیش‌بینی با متغیرهای وابسته مختلف طراحی شد: وقوع ورم پستان بالینی در ماه‌های ۲ تا ۸، وقوع ورم پستان تحت بالینی در ماه‌های ۲ تا ۸، و وقوع ورم پستان تحت بالینی در ماه دوم پس از زایش. از الگوریتم‌های یادگیری ماشینی شامل Logistic regression، Decision Tree، Support Vector Machine و Random Forest و همچنین الگوریتم Artificial Neural Networks جهت پیشگویی متغیر وابسته بر اساس متغیر مستقل مذکور استفاده گردید.

نتایج: در پیش‌بینی ورم پستان تحت بالینی در ماه دوم، الگوریتم درخت تصمیم با بالاترین مقدار AUC (۰/۷۹) و الگوریتم جنگل تصادفی با بالاترین صحت (۰/۹۳) عملکرد بهتری نشان دادند. با کاهش فاصله زمانی بین متغیرهای پیش‌بین و وقوع رویداد، قدرت پیش‌بینی مدل‌ها به طور قابل توجهی افزایش یافت. در مدل‌های با افق پیش‌بینی طولانی‌تر (ماه‌های ۲ تا ۸)، عملکرد مدل‌ها کاهش ولی همچنان شبکه عصبی مصنوعی و درخت تصمیم نتایج نسبتاً قابل قبولی ارائه دادند.

نتیجه‌گیری: الگوریتم‌های یادگیری ماشینی، به ویژه درخت تصمیم و شبکه عصبی مصنوعی، توانایی خوبی در پیش‌بینی کوتاه‌مدت وقوع ورم پستان تحت بالینی بر اساس داده‌های ساده و در دسترس گله دارند.

واژگان کلیدی: ورم پستان؛ بیماری‌های دوره انتقال؛ شاخص‌های تولید شیر؛ گاو شیری؛ یادگیری ماشینی.

Predicting the Occurrence of Mastitis in Dairy Cows Using Machine Learning Algorithms Based on Peri-Parturient Disorders and Milk Production Indices

Seyedeh Zahra Sahraei¹, Morteza Keywanloo², Babak Khoramian², Ehsanallah Afshari Safavi²,
Mohammad Rahmanimanesh³

¹ DVM Graduate, Faculty of Veterinary Medicine, Semnan University, Semnan, Iran

² Department of Clinical Sciences, Faculty of Veterinary Medicine, Ferdowsi University of Mashhad, Mashhad, Iran

³ Department of Computer Software Engineering, Faculty of Electrical & Computer Engineering, Semnan University, Iran

Background and Study type: Mastitis is one of the important diseases in the dairy industry. Early identification of at-risk cows can play a significant role in the economic management of the herd. The present study was conducted as a retrospective study with the aim of evaluating the predictive power of machine learning algorithms in forecasting the occurrence of mastitis based on commonly recorded herd data.

Objectives: This research aimed to compare the performance of several machine learning algorithms in predicting the occurrence of clinical and subclinical mastitis at different periods after calving, based on variables related to peri-parturient disorders and milk production indices.

Methods: Data from 2905 Holstein cows in a large industrial herd over two years were collected. Independent variables included parity, occurrence of peri-parturient disorders (dystocia, retained placenta, metritis, mastitis, and mammary edema in the first month), and first-month milk indices (production yield, fat and protein percentage, and somatic cell count (SCC)). Three prediction models with different dependent variables were designed: occurrence of clinical mastitis in months 2 to 8, occurrence of subclinical mastitis (SCC >200,000) in months 2 to 8, and occurrence of subclinical mastitis in the second month. Five machine learning algorithms, including Logistic Regression, Decision Tree, Support Vector Machine, Random Forest, and Artificial Neural Networks, were trained and evaluated using metrics such as Accuracy, Area Under the Curve (AUC), Precision, and Recall.

Results: In the third model (predicting subclinical mastitis in the second month), the Decision Tree algorithm showed the highest AUC value (0.79), and the Random Forest algorithm showed the highest Accuracy (0.93). Overall, as the time gap between the predictor variables (first month) and the event occurrence (second month) decreased, the predictive power of the models increased significantly. In models with longer prediction horizons (months 2 to 8), performance decreased, but Artificial Neural Networks and Decision Tree still provided relatively acceptable results.

Conclusions: Machine learning algorithms, particularly Decision Tree and Artificial Neural Networks, demonstrate good capability for the short-term prediction of subclinical mastitis occurrence based on simple and readily available herd data.

Keywords: Mastitis; Peri-parturient disorders; Milk production indices; Dairy cow; Machine learning.