

Zastosowanie metody MARS do klasyfikowania zabiegów inseminacyjnych u bydła mlecznego*

Wilhelm Grzesiak, Piotr Sablik, Daniel Zaborski,
Agata Żukiewicz, Andrzej Dybus, Iwona Szatkowska

Zachodniopomorski Uniwersytet Technologiczny w Szczecinie,
Katedra Nauk o Zwierzętach Przeżuwających,
ul. Dr. Judyma 10, 46-746 Szczecin

W niniejszych badaniach zaprezentowano klasyfikację zabiegów inseminacyjnych u krów za pomocą funkcji klasyfikacyjnych (CF) oraz metody MARS – Multivariate Adaptive Regression Splines (wielowymiarowa adaptacyjna regresja krzywych sklejanych). Zabiegi inseminacyjne podzielono na dwie kategorie – prawidłowe, kończące się zapłodnieniem za pierwszym bądź drugim zabiegiem („DOBRZE”) oraz problematyczne, wymagające do zapłodnienia więcej niż dwóch zabiegów („ZŁE”). Jakość przygotowanych klasyfikatorów była dobra, zarówno dla czułości jak i specyficzności i przekraczała 80%. W prognozowaniu kategorii zabiegu inseminacyjnego metoda MARS okazała się bardziej precyzyjna w przewidywaniu problematycznych inseminacji (czułość) w porównaniu do prognoz funkcji klasyfikacyjnych i ostrożniejsza w przewidywaniu zabiegów prawidłowych (specyficzność), co jest istotne z hodowlanego punktu widzenia. Spośród zmiennych determinujących rodzaj inseminacji, największe znaczenie dla obydwu klasyfikatorów posiadała zmienna oznaczająca okres międzyciążowy oraz okres międzywycieleniowy. Z pozostałych zmiennych powtarzały się kondycja w momencie inseminacji oraz przeciętna kondycja krów, zawartość białka i wydajność tłuszczu w mleku. Otrzymane prognozy można wykorzystywać do odpowiedniego przygotowania krów do inseminacji.

SŁOWA KLUCZOWE: MARS / bydło mleczne / inseminacja / klasyfikacja

Powszechnie wiadomo, że precyzyjne wykrycie rui, odpowiedni moment unasięniania, a także właściwa technika wykonania zabiegu przez inseminatora stanowią podstawowy warunek skutecznej inseminacji [6, 11, 13, 14]. Jak podaje Thibier [27], na świecie tylko 20% krów zachodzi w ciążę po pierwszym zabiegu inseminacyjnym. Na skuteczność tego zabiegu wpływa wiele czynników, między innymi: jakość i ilość

*Projekt finansowany z grantu KBN No N311 017 31/3536

nasienia buhaja [12], wiek [26] i kondycja krów [2], wydajność mleka [23] i inne. W związku z czym, mimo poprawnego wykonania zabiegu inseminacyjnego, jego skuteczność może być wątpliwa, co generuje wymierne straty ekonomiczne [9, 25].

Do klasyfikacji można wykorzystać wiele różnych metod, takich: jak analiza dyskryminacji z funkcjami klasyfikacyjnymi (CF), regresja logistyczna czy drzewa klasyfikacyjne, bądź sztuczne sieci neuronowe. Ciekawym podejściem do tej kwestii wydaje się zastosowanie metody MARS – Multivariate Adaptive Regression Splines, zaproponowanej przez Friedmana [7]. Metodę tę określić można jako ważoną sumę wybranych tzw. funkcji sklejanych (zwanymi też funkcjami bazowymi) do odwzorowania zmienności poszczególnych zmiennych objaśniających. Jest to uogólniony model addytywny. Analizie podlegają wszystkie obserwacje poszczególnych zmiennych objaśniających, a obszar zmienności cechy jest dzielony na przedziały, w których zmienna ta ma różny wpływ na analizowane zjawisko. Zmienna uwzględniana jest w modelu z różnym współczynnikiem wagowym i różnym znakiem, w zależności od tego czy znajduje się powyżej czy poniżej pewnej wartości progowej. To rozróżnienie dokonuje się przy udziale funkcji sklejanych, a w zasadzie przy wykorzystaniu iloczynu tensorowego tych krzywych do opisanego zależności między zmiennymi objaśniającymi. W metodzie tej istnieje możliwość uwzględnienia interakcji między zmiennymi objaśniającymi. Przy tworzeniu odpowiedniego modelu MARS w pierwszej kolejności definiowany jest model ogólny z wieloma funkcjami sklejanymi, których nadmiar prowadzi do zbyt dokładnego dopasowania modelu do danych. Potem następuje „przycinanie”, czyli eliminacja z modelu funkcji nadmiarowych. Szukanie najlepszego predyktora oraz miejsca jego wiązania jest procesem iteracyjnym.

Wykorzystanie tej metody nie jest wcale incydentalne, nawet w zagadnieniach związanych z szeroko rozumianą biologią, aczkolwiek w zagadnieniach *stricto* zootechnicznych zastosowań takich nie znaleziono. Metodę MARS zastosowano, między innymi: do identyfikacji czynników genetycznych i pozagenetycznych odpowiedzialnych za choroby [30]; do opracowania nowych metod ulepszających predykcję rozmieszczenia gatunków, na podstawie danych dotyczących ich występowania [3], do przewidywania chromatograficznej retencji peptydów [24]; do mapowania preferencji konsumentów batoników serowych [29] i innych.

Celowym wydaje się więc podjęcie badań umożliwiających klasyfikację krów, w stosunku do których skuteczność zabiegu może okazać się problematyczna. W niniejszej pracy zostaną także sprawdzone możliwości predykcyjne metody MARS na tle typowej analizy dyskryminacyjnej z funkcjami klasyfikacyjnymi [10]. Funkcje klasyfikacyjne (CF) są stosowane dość powszechnie w różnych zagadnieniach klasyfikacyjnych, np. do detekcji dystocji u bydła mięsnego [21], do detekcji mastitis u krów [18, 28] czy też do klasyfikowania typów genetycznych bydła na podstawie wybranych cech mleczności [16].

Metody MARS i CF zostaną sprawdzone w praktycznym zastosowaniu. Dobry model klasyfikacyjny może stanowić podporę dla hodowcy, ułatwiając podjęcie różnego rodzaju przedsięwzięć zmierzających do poprawy skuteczności zapłodnienia poszcze-

gólnych krów, a tym samym podnieść ogólną płodność w stadzie i, co za tym idzie, efektywność gospodarstwa.

Materiał i metody

W badaniach wykorzystano 918 pełnych rekordów zawierających dane o zabiegach inseminacyjnych krów z wysokim (ok. 86%) udziałem genów rasy holsztyńsko-fryzyskiej (hf), w dwóch sezonach użytkowania. Krowy przebywały w jednej z zachodniopomorskich obór wolnostanowiskowych, z możliwością korzystania przez cały rok z wybiegów. Przeciętna wydajność mleka za 305 dni laktacji wynosiła 9428 kg mleka. Charakterystykę badanego stada przedstawiono w tabeli 1.

Tabela 1 – Table 1

Charakterystyka badanego stada
Characterization of the examined herd

Wyszczególnienie Specification		Zbiór – Set	
		L (zbiór uczący) (learning set)	T (zbiór testowy) (test set)
n		786	150
Udział genów rasy h.f. (%)	\bar{x}	85,47	85,68
Participation of HF genes (%)	Sd	9,77	8,86
Wiek (miesiące)	\bar{x}	28,6	29,5
Age (months)	Sd	11,9	12,2
Okres międzyciążowy (dni)	\bar{x}	144,7	148,4
Calving-to-conception interval (dni)	Sd	69,8	66,2
Okres międzywycieleniowy (dni)	\bar{x}	427,7	431,5
Calving interval (days)	Sd	72,6	73,9
Długość trwania ciąży (dni)	\bar{x}	280,1	280,5
Length of pregnancy (days)	Sd	6,5	6,7
BCS _{AI} ¹⁾ (pkt. – points)	\bar{x}	2,77	2,72
	Sd	0,38	0,36
BCS _M ²⁾ (pkt. – points)	\bar{x}	2,85	2,80
	Sd	0,32	0,27
Wydajność mleka (kg)	\bar{x}	9360	9778
Milk yield (kg)	Sd	2076	2252
Wydajność tłuszczu (kg)	\bar{x}	409,8	426,3
Fat yield (kg)	Sd	94,6	98,2
Zawartość tłuszczu (%)	\bar{x}	4,41	4,40
Fat content (%)	Sd	0,50	0,51
Wydajność białka (kg)	\bar{x}	311,4	324,5
Protein yield (kg)	Sd	63,9	69,0
Zawartość białka (%)	\bar{x}	3,36	3,34
Protein content (%)	Sd	0,21	0,20

BCS_{AI}¹⁾ – kondycja krowy oceniana w trakcie inseminacji – cow condition during insemination

BCS_M²⁾ – przeciętna kondycja krowy oceniana po porodzie, w trakcie inseminacji oraz w trakcie trwania ciąży (3 pomiary) – the average condition of cow, as evaluated after parturition, during insemination and during pregnancy (3 measurements)

W analizach uwzględniano krowy, u których nie stwierdzono w okresie inseminowania poważnych schorzeń. Jako zmienne klasyfikacyjne wybrano: HF% – udział genów hf w genotypie krowy (%); LAKTACJA – kolejną laktację przed zacieleniem (do 4 włącznie); SEZON – sezon inseminacji (jesiennie-zimowy oraz wiosennie-letni); MIESIĄC – miesiąc zacielenia (oznaczenia kalendarzowe); WIEK – wiek zacielanej krowy (w miesiącach); PŁEĆ – płeć cielęcia (oznaczana jako: jałówka – 0; byczek – 0,5; dwie jałówki – 1; jałówka i byczek – 1,5 oraz dwa byczki – 2); OMC – okres międzyciążowy (dni); OMW – okres międzywycieleniowy (dni); CIAŻA – długość trwania ciąży (dni); BCS_{AI} – kondycja krowy oceniana w trakcie zabiegu (Body Condition Score), określana według zmodyfikowanej skali 5-punktowej (za optymalną uznano 3,50 pkt. i przy wartościach większych odejmowano od 3,50 wielokrotność 0,25; jeżeli krowę wyceniono na 3,75 pkt. przypisywano jej 3,25 pkt; jeżeli krowę wyceniono na 4,00 pkt. otrzymywała 3,00; 4,25 – 2,75; 4,50 – 2,50 i 4,75 – 2,25 i 5,00 – 2,00 (na bazie [4, 5])); BCS_M – przeciętna kondycja krowy oceniana po porodzie, w trakcie zabiegu inseminacji oraz w trakcie trwania ciąży (3 pomiary); MLEKO – średnią wydajność mleka krowy z poprzedniej laktacji (kg); TŁUSZCZ – średnią wydajność tłuszczu w mleku (kg); TŁUSZCZ% – średnią zawartość tłuszczu w mleku (%); BIAŁKO – średnią wydajność białka w mleku (kg); BIAŁKO% – średnią zawartość białka w mleku (%).

Założono, na podstawie uzyskanych w gospodarstwie informacji, że wykrywanie rui jak również sam zabieg inseminacji przeprowadzane były w odpowiednim czasie i zgodnie z wymaganiami, a co najmniej podejście do wszystkich krów było jednakowe. Wybrane zmienne diagnostyczne stanowią więc dodatkowy element rozpoznawczy dla zmiennej wyjściowej, oznaczającej kategorię zabiegu inseminacyjnego u krowy: „DOBRZE” – krowy zacielały się po pierwszej, ewentualnie drugiej inseminacji; „ŹLE” – krowy zacielały się po więcej niż dwóch inseminacjach (od 3 do 11) – tabela 2.

Wszystkie 918 zabiegów inseminacyjnych podzielone zostało losowo na zbiór uczący (768 zabiegów) oraz zbiór testowy (150 zabiegów). Zbiór testowy, tworzyły przypadki nieuwzględnione wcześniej przy przygotowywaniu modelu MARS, a także przy estymacji wskaźników klasyfikacyjnych. Wyniki przeprowadzonych zabiegów ze zbioru testowego posłużyły do sprawdzenia zdolności predykcyjnej badanych modeli. Charakterystyka tych zbiorów uzupełniona została w tabeli 2.

Tabela 2 – Table 2

Podział zabiegów inseminacyjnych wg przyjętych kryteriów (L – zbiór uczący, T – zbiór testowy)
Division of inseminations according to the established criteria (L – learning set, T – test set)

Zbiór Set	"DOBRZE" "GOOD"	"ŹLE" "POOR"
L	500 (65,10%)	268 (34,90%)
T	86 (57,34%)	64 (42,66%)

Liniowa funkcja klasyfikacyjna była ważoną sumą analizowanych zmiennych predykcyjnych (X_i).

$$\bar{y} = d_0 + d^T \cdot X_i$$

gdzie:

\bar{y} – to zmienna przewidywana;

d_0 – wartości początkowe;

d^T – wektor współczynników klasyfikacji;

X_i – wektor zmiennych opisujących cechy predykcyjne X_j w i -tym obiekcie.

Wartość funkcji klasyfikacji, wyznaczonej dla rozpatrywanych obiektów, przyrównywano do wzorca określając w ten sposób przynależność obiektów do określonej klasy. Jakość kryteriów klasyfikacyjnych poszczególnych zmiennych oceniono za pomocą istotności statystyki F obliczonej na podstawie współczynnika lambda Wilksa [20].

$$\Lambda = \prod_{i=m+1}^t (1 - r_d^2)$$

gdzie:

Λ – to współczynnik lambda Wilksa;

$m = 1, 2, \dots, t-1$, gdzie t oznacza rząd macierzy międzygrupowej sumy kwadratów;

r_d^2 – to współczynnik determinacji pomiędzy zmiennymi kanonicznymi, o wariacjach równych 1, będącymi kombinacjami liniowymi najlepiej reprezentującymi zmienne pomiarowe i klasyfikacyjne (r maksymalne).

Sprawdzono także, czy są spełnione założenia analizy klasyfikacyjnej – normalność rozkładu poszczególnych zmiennych (testem Shapiro-Wilka) oraz homogeniczność wariacji (testem Levene'a).

W metodzie MARS obszar zmienności dzielony był na przedziały, w których poszczególne zmienne objaśniające mają różny wpływ na zmienną klasyfikacyjną (rodzaj zabiegu inseminacyjnego). Zmienne objaśniające (predykcyjne) wykorzystywane były w modelu z różnymi współczynnikami wagowymi. Zmienna klasyfikacyjna była kodowana na wielokrotną zmienną zero-jedynkową. Ogólny model przybrał postać:

$$\bar{y} = \beta_0 + \sum_{m=1}^M \beta_m h_m(X)$$

gdzie:

\bar{y} – to zmienna przewidywana;

b_0 – to współczynnik stałej funkcji bazowej,

b_m – to współczynnik m -tej funkcji bazowej;

$h_m(X)$ – to m -ta funkcja bazowa, która może być pojedynczą funkcją sklejaną lub efektem interakcji dwóch lub więcej funkcji sklejanych oraz M jest numerem funkcji bazowej włączonej do modelu, a X oznacza zmienne predykcyjne.

Przy tworzeniu modelu badano do 150 funkcji sklepanych, z których w rekurencyjnym procesie przygotowania były eliminowane te, których usunięcie nie powodowało spadku dobroci dopasowania. Uwzględniano interakcje od 6-go, do 1-go stopnia. Miara jakości modelu był współczynnik błędu resztowego GCV (Generalized Cross Validation error), którego minimalizacja umożliwia szybką optymalizację modelu.

$$GCV(K) = \frac{1}{N} \sum_{i=1}^N \left(\frac{y_i - \bar{y}}{1 - \frac{r + cK}{N}} \right)^2$$

gdzie:

y_i – to wartość rzeczywista;

\bar{y} – wartość prognozowana

N – liczba przypadków w zbiorze uczącym;

r – liczba niezależnych funkcji bazowych;

K – liczba węzłów;

c – współczynnik „kary”, im większa wartość c tym więcej funkcji bazowych będzie usuniętych z modelu.

Najlepszy model został wykorzystany do prognozowania zacielenia u krów ze zbioru testowego i sprawdzenia jakości tego prognozowania.

Jakość analizowanych sieci opisano przeciętną wielkością błędu RMS dla danych uczących:

$$RMS = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y})^2}{n}}$$

gdzie:

n – liczba obserwacji;

y_i – wartości rzeczywiste;

\hat{y} – wartości prognozowane przez dany model.

Jakość efektów przeprowadzonych klasyfikacji określona została za pomocą czułości (sensitivity) – warunkowego prawdopodobieństwa prawdziwie pozytywnego P_{TP} i specyficzności (specificity) – warunkowego prawdopodobieństwa prawdziwie negatywnego P_{TN} .

$$P_{TP} = \frac{A}{A + C} \quad P_{TN} = \frac{D}{B + D}$$

gdzie:

A – liczba poprawnie rozpoznanych inseminacji kategorii „ŻLE”;

C – liczba inseminacji rozpoznanych jako inseminacje kategorii „DOBRZE”, a w rzeczywistości były to inseminacje kategorii „ŻLE”;

B – liczba inseminacji rozpoznanych jako inseminacje kategorii „ŹLE”, a w rzeczywistości były to inseminacje kategorii „DOBRZE”;
D – liczba poprawnie rozpoznanych inseminacji kategorii „DOBRZE” (tab. 3).

Tabela 3 – Table 3

Zestawienie obserwowanych i przewidywanych kategorii zabiegów inseminacyjnych
 Specification of the observed and predicted categories of inseminations

Kategorie inseminacji przewidywanych Prediction of AI	Kategorie inseminacji obserwowanych Observation of AI		Razem Total
	"ŹLE" "POOR"	"DOBRZE" "GOOD"	
	"ŹLE" "POOR"	A	
"DOBRZE" "GOOD"	C	D	C+D
Razem – Total	A+C	B+D	A+B+C+D

Zastosowana została także miara określająca ogólne prawdopodobieństwo błędu (tj. prawdopodobieństwo odpowiedzi prawdziwie pozytywnych lub prawdziwie negatywnych P_G):

$$P_G = \frac{A + D}{A + B + C + D}$$

Różnice między prawdopodobieństwami poszczególnych modeli liczone za pomocą testu dwóch proporcji dla nierównych wariancji.

Istotność wkładu poszczególnych zmiennych dla funkcji klasyfikacyjnych (CF) określono za pomocą wielkości oraz istotności współczynnika tolerancji (T_c)

$$T_c = 1 - r^2_{X_i/X_{p-1}}$$

gdzie:

$r^2_{X_i/X_{p-1}}$ – to wielowymiarowy współczynnik determinacji pomiędzy jedną ze zmiennych losowych X_i , a pozostałymi $p-1$ zmiennymi losowymi wektora zmiennych diagnostycznych [1].

Dla metody MARS zmienne uporządkowano w zależności do ilości odwołań dla każdej zmiennej (NoA).

Wyniki i dyskusja

Jakość klasyfikatorów

Dla metody MARS wybrano model, w którym uwzględnionych zostało 100 funkcji sklejanych oraz interakcje 4-go stopnia, posiadający najniższą wartością błędu resztowego $GCV = 0,2021$. Z charakterystyki CF przedstawionej w tabeli 4 wynika, że naj-

większy wkład w determinację kategorii zabiegów inseminacyjnych miały pierwsze cztery zmienne o największej wartości i istotności współczynnika Λ . Zmienne, dla których wartość F była mniejsza od 1, wskazane zostały jako zbędne.

Tabela 4 – Table 4

Charakterystyka parametrów funkcji klasyfikacyjnej
Characterization of the classification function parameters

Zmienne Variables	Λ	F	p	Tc
OMC ¹⁾	0,6179	90,96	0,0000	0,58
OMW ²⁾	0,5793	37,88	0,0000	0,58
BCS _{AI} ³⁾	0,5613	13,06	0,0003	0,65
BCS _M ⁴⁾	0,5577	8,17	0,0044	0,61
BIAŁKO% PROTEIN%	0,5542	3,25	0,0719	0,92
LAKTACJA LACTATION	0,5539	2,84	0,0923	0,67
HF%	0,5536	2,43	0,1192	0,89
TŁUSZCZ FAT	0,5531	1,83	0,1760	0,64
SEZON SEASON	0,5517	0,16	0,6935	0,95
MIESIĄC MONTH	0,5518	0,03	0,8665	0,97
WIEK AGE	0,5513	0,70	0,4019	0,09
PŁEĆ SEX	0,5516	0,27	0,6025	0,99
CIĄŻA PREGNANCY	0,5512	0,88	0,3489	0,90
MLEKO MILK	0,5517	0,10	0,7562	0,20
TŁUSZCZ% FAT%	0,5517	0,11	0,7403	0,64
BIAŁKO PROTEIN	0,5518	0,03	0,8565	0,22

OMC¹⁾ – okres międzyciążowy – calving-to-conception interval

OMW²⁾ – okres międzywycieleniowy – calving interval

BCS_{AI}³⁾ – kondycja krwi w trakcie inseminacji – cow condition during insemination

BCS_M⁴⁾ – przeciętna kondycja krwi oceniana po porodzie, w trakcie inseminacji oraz w trakcie trwania ciąży (3 pomiary) – The average condition of cow, as evaluated after parturition, during insemination and during pregnancy (3 measurements)

Ogólny błąd RMS dla metody MARS wynosił 0,29, natomiast dla CF był nieco wyższy – 0,35. Pewne różnice w jakości można dostrzec przy analizie czułości i specyficzności uzyskanych na podstawie zbioru uczącego (tab. 5). Zauważyć można, że większą czułością (PTR), czyli zdolnością do rozpoznawania krów z inseminacjami należącymi do kategorii „ŻLE”, charakteryzowała się metoda MARS (0,85), podczas gdy czułość CF wynosiła 0,60, co przekłada się na odpowiednio 72 i 106 krów z kło-

potami inseminacyjnymi, których „nie zauważyły” klasyfikatory (tab. 6). Obserwowane prawdopodobieństwa (P_{TP}) dla MARS i CF były wyższe niż uzyskane, przy najlepszej dokładności dla CF, przez Montgomery i wsp. [18], przy detekcji mastitis (czułość ok. 0,45). Natomiast Morrison i wsp. [22], przy klasyfikacji dystocji u bydła za pomocą CF, uzyskali czułość dla zbioru uczącego wynoszącą 0,78, a więc wyższą niż w niniejszej pracy dla CF.

Tabela 5 – Table 5

Wartości czułości (P_{TP}) i specyficzności (P_{TN}) dla MARS i CF (L – zbiór uczący, T – zbiór testowy)
Values of sensitivity (P_{TP}) and specificity (P_{TN}) for MARS and CF (L – learning set, T – test set)

	Zbiór Set	n	P_{TP}	P_{TN}	P_G
MARS	L	768	0,85 ^A	0,86 ^A	0,85
	T	150	0,88 ^A	0,88 ^A	0,88
CF	L	768	0,60 ^A	0,93 ^A	0,82
	T	150	0,63 ^A	0,98 ^A	0,83

A – $P \leq 0,01$

Tabela 6 – Table 6

Macierz klasyfikacji MARS i CF
Classification matrix for MARS and CF

Kategorie inseminacji przewidywane Prediction of AI	Kategorie inseminacji obserwowane Observation of AI			
	"ŹLE" "POOR"		"DOBRCZE" "GOOD"	
	MARS	CF	MARS	CF
Zbiór uczący – Learning set				
"ŹLE" – "POOR"	227	162	72	36
"DOBRCZE" – "GOOD"	41	106	428	464
Zbiór testowy – Test set				
"ŹLE" – "POOR"	56	40	10	2
"DOBRCZE" – "GOOD"	8	24	76	84

Prawdopodobieństwo klasyfikowania krów z inseminacjami kategorii „DOBRCZE” (P_{TN}), czyli specyficzność, najwyższe było dla CF (0,93). Wartość ta różniła się statystycznie istotnie od prawdopodobieństw dla MARS wynoszącego 0,86, przy czym wartości te były podobne do uzyskanych przez Morrison i wsp. [22], przy detekcji dystocji za pomocą CF, lecz niższe niż uzyskane przez Montgomery i wsp. [18] przy detekcji mastitis (0,97). Z hodowlanego punktu widzenia, zdecydowanie bardziej korzystna jest zdolność do poprawnego wskazywania krów z trudnymi wycieleniami, co przemawia za metodą MARS.

Ogólne prawdopodobieństwo (PG), charakteryzujące dokładność klasyfikacji (stosunek poprawnie zaklasyfikowanych przypadków do ogólnej liczby przypadków), wynosiło 0,82 dla CF oraz 0,85 dla MARS. Wartości te były zbliżone do uzyskanych przez Morrison i wsp. [22] przy klasyfikacji dystocji u krów. Zaznaczyć trzeba jednak, że przy tak znacznej liczebności już niewielkie różnice charakteryzują się statystyczną istotnością.

Trzeba także podkreślić, że niektóre ze zmiennych objaśniających charakteryzowały się znacznymi odstępstwami od rozkładu normalnego (wiek zacielanej krowy, kondycja krów w trakcie zabiegu, średnia wydajność mleka krowy z poprzedniej laktacji, zawartość tłuszczu w mleku, jak również sezon, kolejna laktacja i pleć cielęcia). Ważniejsze jednak było złamanie założeń dotyczących jednorodności wariancji i kowariancji tych zmiennych. Ze statystycznego punktu widzenia wykorzystanie takiego modelu w przypadku CF nie byłoby wskazane, natomiast przy zastosowaniu metody MARS te założenia nie są wymagane, aczkolwiek wpływ takich zmiennych na jakość tego modelu nie jest bez znaczenia.

Prognozy

Tak przygotowane klasyfikatory wykorzystano do klasyfikowania inseminacji na podstawie danych niedostępnych przy estymacji współczynników klasyfikacyjnych CF i wyborze modelu MARS. Wyniki tych klasyfikacji przedstawiono w tabeli 5 (pogrubioną czcionką) oraz w tabeli 6 dla zbioru testowego. Przewidywanie krów z inseminacjami kategorii „ŻLE” (P_{TP}) dla MARS (0,88) było statystycznie istotnie wyższe ($P < 0,01$) niż dla CF (0,63), co w praktyce przełożyć można na 8 krów, których model MARS nie uwzględnił, jako mogących mieć kłopoty z inseminacją, podczas gdy CF „przegapiła” aż 24 takie krowy (tab. 6). Specyficzność MARS była niższa niż CF o prawie 0,10. Czulości i specyficzności uzyskane przez Morrison i wsp. [22] wynosiły, odpowiednio: 0,57 oraz 0,91 i były dość podobne do wartości uzyskanych w niniejszej pracy dla CF. W innych badaniach Morrison i wsp. [21] stwierdzili dla CF większą czulość (0,87) niż specyficzność (0,81) w prognozowaniu dystocji u krów, jednakże wykorzystali te same dane do utworzenia modelu i następnie sprawdzenia jego zdolności prognostycznych.

Zmienne o największym wkładzie determinacyjnym

Uporządkowanie poszczególnych zmiennych klasyfikacyjnych według ich wkładu w determinację kategorii zabiegu inseminacyjnego przedstawiono w tabeli 7. Dla CF oraz dla MARS najważniejszą zmienną determinacyjną był okres międzyciążowy (OMC), czyli okres między wycieleniem a kolejnym zacieleniem oraz okres międzywycieleniowy (OMW), czyli czas między jednym a drugim wycieleniem. Wydaje się to zrozumiałe gdyż im dłuższe są te okresy tym większe kłopoty z zacieleniem. Natomiast, co do kolejności pozostałych zmiennych, to efekt jest różny w zależności od klasyfikatora. Dla MARS są to: długość ciąży, zawartość białka w mleku, wiek zacielanej krowy, a także przeciętna kondycja krów (BCS_M); natomiast dla CF: kondycja krów (BCS_{AI}, BCS_M); zawartość białka w mleku; udział genów HF w genotypie krowy. Moreira i wsp. [19] uważają, że kłopoty z zapłodnieniem mają krowy, których kondycja w momencie inseminacji jest niższa od 2,50. Może to być następstwem niskiej kondycji

krów przy wycieleniu, co przekłada się na opóźnienie wznowienia aktywności jajników [17]. W niniejszej pracy kondycja w momencie inseminacji jest związana także z przeciętną kondycją krów (BCSM), będącą wypadkową 3 pomiarów (po porodzie, w trakcie inseminacji oraz w trakcie trwania ciąży). W każdym razie nie jest to takie jednoznaczne, gdyż np. Garnsworthy [8] uważa, że kondycja po porodzie nie wiąże się z ewentualnymi kłopotami przy inseminacji, natomiast Kozdrowski i wsp. [15] uważają, że większe kłopoty z zacieleniem mają krowy z wyższym wskaźnikiem BCS (ponad 3,3).

Tabela 7 – Table 7

Analiza zmiennych objaśniających o największym wkładzie determinacyjnym dla metody MARS oraz dla CF
An analysis of explanatory variables with the greatest determination contribution for the MARS method and CF

Wyszczególnienie Specification	MARS		CF	
	ranga rang	NoA ¹⁾	ranga rang	Tc
HF%	12	3	6	0,89
LAKTACJA LACTATION	13	2	7	0,67
SEZON SEASON	14	0	9	0,95
MIESIĄC MONTH	15	0	10	0,97
WIEK AGE	5	15	11	0,09
PLEĆ SEX	16	0	12	0,99
OMC ²⁾	1	24	1	0,58
OMW ³⁾	2	20	2	0,58
CIAŻA PREGNANCY	3	16	13	0,90
BCS ⁴⁾	7	10	3	0,65
BCS _M ⁵⁾	6	14	4	0,61
MLEKO MILK	9	5	14	0,20
TŁUSZCZ FAT	8	8	8	0,64
TŁUSZCZ% FAT%	11	3	15	0,64
BIAŁKO PROTEIN	10	4	16	0,22
BIAŁKO% PROTEIN%	4	16	5	0,92

NoA¹⁾ – liczba odwołań – number of references

OMC²⁾ – okres międzyciążowy – calving-to-conception interval

OMW³⁾ – okres międzywycieleniowy – calving interval

BCS_A⁴⁾ – kondycja krowy w trakcie inseminacji – cow condition during insemination

BCS_M⁵⁾ – przeciętna kondycja krowy oceniana po porodzie, w trakcie unasieniania oraz w trakcie trwania ciąży (3 pomiary) – the average condition of cow, as evaluated after parturition, during insemination and during pregnancy (3 measurements)

Trudno wskazać pozycje piśmiennictwa ukazujące powiązanie wyszczególnionych zmiennych z skutecznością zabiegu inseminacyjnego. Z przeprowadzonych ponad 60 lat temu badań Tanabe i Salisbury [26] wynika, że z wiekiem krów odsetek skutecznych inseminacji rośnie. Nebel i McGilliard [23] wskazują na różne powiązania skuteczności zabiegów inseminacyjnych z mlecznością krów, jednakże nie są to wskazania jednoznaczne. Pozostałe zmienne charakteryzowały się wartościami wskazującymi na ich niewielki wkład w determinację skutecznego zabiegu inseminacyjnego. Trzeba jednak mieć na uwadze, że wobec ewentualnych współzależności pomiędzy zmiennymi, przedstawione w tabeli 6 efekty mogą nie odzwierciedlać rzeczywistych powiązań i oddziaływań determinacyjnych poszczególnych zmiennych.

Podsumowując należy stwierdzić, że przeprowadzone analizy pokazały przydatność badanych klasyfikatorów do detekcji skutecznych inseminacji u bydła mlecznego. W szczególności metoda MARS, w porównaniu do CF, wykazała pewną przewagę. Więcej trafnych detekcji inseminacji kategorii „DOBRZE” (PFN) dokonywane było przez CF, natomiast metoda MARS nie była w tym względzie tak „optymistyczna”. Metoda MARS dawała bardziej „ostrożne” klasyfikacje, tzn. bardziej precyzyjnie wskazywała krowy z trudnymi inseminacjami, a z większą ostrożnością podchodziła do krów, które teoretycznie kłopotów przy inseminacji mieć nie powinny, w przeciwieństwie do metody CF. Ta ostrożność jest w tym wypadku zaletą, ponieważ zbyt optymistyczne czy też pochopne wskazywanie krowy bez problemów z inseminacją, w warunkach hodowli może okazać się zawodne. W niniejszych badaniach pewną rolę, w detekcji poszczególnych kategorii inseminacji odegrała zmienna oznaczająca kondycję krowy. Można zasugerować hodowcom ocenę kondycji krowy w celu bardziej precyzyjnego wskazania krowy z odpowiednią kategorią zabiegu inseminacyjnego. Wykorzystanie metody MARS może stanowić podporę do podjęcia pewnych profilaktycznych zabiegów w stosunku do krów, które mogą mieć kłopoty z inseminacją. Przyczyni się to do zwiększenia płodności w stadzie, a co za tym idzie zwiększenia efektów ekonomicznych gospodarstwa.

PIŚMIENNICTWO

1. DOBOSZ M., 2001 – Wspomagana komputerowo statystyczna analiza wyników badań. AOW EXIT, Warszawa.
2. DOMEcq J.J., SKIDMORE A.L., LLOYD J.W., KANEENE J.B., 1997 – Relationship Between Body Condition Scores and Conception at First Artificial Insemination in a Large Dairy Herd of High Yielding Holstein Cows. *Journal of Dairy Science* 80, 113-120.
3. ELITH J., GRAHAM C.H., ANDERSON R.P., DUDIK M., FERRIER S., GUI SAN A., HJMANS R.J., HUETT MANN F., LEATHWICK J.R., LEHMANN A., LI J., LOHMANN L.G., LOISELLE B.A., MANION G., MORITZ C., NAKAMURA M., NAKAZAWA Y., OVERTON J. McC., PETERSON A.T., PHILLIPS S.J., RICHARDSON K.S., SCACHETTI-PEREIRA R., SCHAPIRE R.E., SOBERON J., WILLIAMS S., WISZ M.S., ZIMMERMANN N.E., 2006 – Novel methods improve prediction of species' distributions from occurrence data. *Ecography* 29, 129-151.
4. FERGUSON J.D., GALLIGAN D.T., THOMSEN N., 1994 – Principal descriptors of body condition in Holstein dairy cattle. *Journal of Dairy Science* 77, 2695-2703.

5. FLAMENBAUM I., WOLFENSON D., KUNZ P.L., MAMAN M., BERMAN A., 1995 – Interactions between body condition at calving and cooling of dairy cows during lactation in summer. *Journal of Dairy Science* 78, 2221-2229.
6. FOOTE R.H., 1979 – Time of artificial insemination and fertility in dairy cattle. *Journal of Dairy Science* 62, 355.
7. FRIEDMAN J., 1991 – Multivariate Adaptive Regression Splines. *Annals of Statistics* 19, 11-67.
8. GARNSWORTHY P.C., 1998 – The effect of energy reserves at calving on performance of dairy cows. Nutrition and Lactation in the Dairy Cow. 1st ed. Butterworths, London, England, 157.
9. GONZALEZ-RECIO O., PEREZ-CABAL M.A., ALENDA R., 2004 – Economic Value of Female Fertility and its Relationship with Profit in Spanish Dairy Cattle. *Journal of Dairy Science* 87, 3053-3061.
10. HAND D., 1981 – Discrimination and Classification. Wiley, Chichester.
11. JANKOWSKA M., 2002 – Double insemination of cows during one oestrus. *Electronic Journal of Polish Agricultural Universities* 5 (1).
12. JAŚKOWSKI J.M., SZENFELD J., 1999 – Wpływ ilości i jakości nasienia oraz techniki unasielenia na wyniki zacieleni krow. *Medycyna Weterynaryjna* 55, 3, 160-162.
13. JAŚKOWSKI J.M., 2000 – Wybrane zagadnienia z rozrodu bydła. Nowoczesny chów bydła. Magrol, Bydgoszcz.
14. KONDRACKI S., 2002 – Od czego zależy skuteczność inseminacji krow? *Chów bydła* 12, 14-16.
15. KOZDROWSKI R., TWARDON J., DEJENKA G.J., DZIĘCIOŁ M., 2006 – Influence of estrus intensity and level of physical constitution on results of artificial insemination in cattle. *Medycyna Weterynaryjna* 62 (9), 1038-1040.
16. LEOTTA R., 2004 – Use of linear discriminant analysis to characterise three dairy cattle breeds on the basis of several milk characteristics. *Italian Journal of Animal Science* 3, 377-383.
17. MARKUSFELD O., GALON N., EZARA E., 1997 – Body condition score, health, yield and fertility in dairy cows. *The Veterinary Record* 141, 67-72.
18. MONTGOMERY M.E., WHITE M.E., MARTIN S.W., 1987 – A comparison of discriminant analysis and logistic regression for the prediction of coliform mastitis in dairy cows. *Canadian Journal of Veterinary Research* 51(4), 495-498.
19. MOREIRA F., RISCO C., PIRES M.F.A., AMBROSE J.D., DROST M., DeLORENZO M., THATCHER W.W., 2000 – Effect of body condition on reproductive efficiency of lactating dairy cows receiving a timed insemination. *Theriogenology* 53, 1305-1319.
20. MORRISON D.F., 1990 – Wielowymiarowa analiza statystyczna. Warszawa, PWN 329.
21. MORRISON D.G., HUMES P.E., KEITH N.K., GODKE R.A., 1985 – Discriminant Analysis for Predicting Dystocia in Beef Cattle. I. Comparison with Regression Analysis. *Journal of Animal Science* 60, 608-616.
22. MORRISON D.G., HUMES P.E., KEITH N.K., GODKE R.A., 1985 – Discriminant Analysis for Predicting Dystocia in Beef Cattle. II. Derivation and Validation of a Prebreeding Prediction Model. *Journal of Animal Science* 60, 617-621.
23. NEBEL R.L., MCGILLIARD M.L., 1993 – Interactions of high milk yield and reproductive performance in dairy cows. *Journal of Dairy Science* 76 (10), 3257-3268.
24. PUT R., VANDER HEYDEN Y., 2007 – The evaluation of two-step multivariate adaptive regression splines for chromatographic retention prediction of peptides. *Proteomics* 7, 1664-1677.

25. SØRENSEN J.T., ØSTERGAARD S., 2003 – Economic consequences of postponed first insemination of cows in a dairy cattle herd. *Livestock Production Science* 79, 145-153.
26. TANABE T., SALISBURY G.W., 1946 – The influence of age on breeding efficiency of dairy cattle in artificial insemination. *Journal of Dairy Science* 23 (6), 337-344.
27. THIBIER M., 2002 – World statistics for artificial insemination in cattle. *Livestock Production Science* 74, 203-212.
28. WHITE M.E., GLICKMAN L.T., BARNES-PALLESEN F.D., PEARSON E.G., MONTGOMERY M.E., ARMSTRONG D., WICKENDEN R.P., HICKEY G., 1986 – Discriminant analysis of the clinical indicants for bovine coliform mastitis. *Cornell Veterinarian* 4, 335-341.
29. XIONG R., MEULLENET J-F., 2004 – Application of Multivariate Adaptive Regression Splines (MARS) to the Preference Mapping of Cheese Sticks. *Journal of Food Science* 69, SNQ 131-139.
30. YORK T., EAVES L., 2001 – Common disease analysis using multivariate adaptive regression splines (MARS): genetic analysis workshop 12 Simulated Sequence Data. *Genetic Epidemiology* 21, 649-654.

Wilhelm Grzesiak, Piotr Sablik, Daniel Zaborski,
Agata Żukiewicz, Andrzej Dybus, Iwona Szatkowska

Application of MARS method in classification of inseminations of dairy cattle

S u m m a r y

In this study, classification of inseminations in cows using classification functions (CF) and MARS method is presented. The inseminations were divided into two categories: normal, resulting in conception after one or two services („GOOD”) and problematic, requiring more than two services for conception („POOR”). The quality of the prepared classifiers was good, both for sensitivity and specificity, and exceeded 80%. In prediction of the insemination class, the MARS method appeared to be more precise in predicting problematic inseminations (sensitivity) compared to predictions of CF and more conservative in predicting normal inseminations (specificity), which is significant from a breeding point of view. Among the variables determining insemination class, calving-to-conception interval and calving interval were the most significant for both classifiers. The condition at insemination and average condition of cows, protein content and fat yield in milk were repeated among the remaining variables. The obtained prognoses may be used to prepare cows appropriately for insemination.