

DARIUSZ R. MAŃKOWSKI¹**WITOLD KOZIROK**²**MONIKA JANASZEK**³¹ Pracownia Ekonomiki Nasiennictwa i Hodowli Roślin, Zakład Nasiennictwa i Nasionoznawstwa
Instytut Hodowli i Aklimatyzacji Roślin — Radzików² Katedra Przetwórstwa i Chemii Surowców Roślinnych, Wydział Nauki o Żywności
Uniwersytet Warmińsko-Mazurski — Olsztyn³ Zakład Mechaniki i Techniki Ciepłej
Katedra Podstaw Inżynierii, Wydział Inżynierii Produkcji
Szkoła Główna Gospodarstwa Wiejskiego — Warszawa

Wstępne badania nad zastosowaniem cyfrowej analizy obrazu oraz analizy funkcji dyskryminacji do oceny jakości browarnej wybranych odmian jęczmienia jarego

Introductory study on application of digital image analysis and discriminant function analysis in assessment of malting quality of spring barley cultivars

Ziarno jęczmienia jest najważniejszym surowcem do produkcji słoju piwowarskiego. Jego jakość jest w głównej mierze determinowana jednolitością odmianową partii ziarna kierowanej do słodowania. Każda odmiana jęczmienia browarnego ma swoje charakterystyczne właściwości, które bada się w mikrosłodowaniach według kryteriów Europejskiej Konwencji Browarnej (EBC). Kierując do słodowania jęczmień określonej odmiany szacuje się na podstawie badań wydajność i jakość słoju. Jakość odmiany ocenia się według wskaźnika Q , gdzie wartość Q poniżej 5 określa odmiany pastewne, a Q bliskie 9 określa odmiany o bardzo dobrych właściwościach browarnych. Dotychczas oceny jakości browarnej ziarniaków jęczmienia dokonywano metodami laboratoryjnymi. Zastosowanie komputerowej analizy obrazu mogłoby znacznie usprawnić proces oceny jakości ziarna oraz obniżyć koszty tego procesu. W pracy podjęto próbę wykorzystania komputerowej analizy obrazu ziarniaków oraz analizy funkcji dyskryminacji do zaklasyfikowania ziarna 11 odmian jęczmienia jarego do określonych grup jakości. Do analizy wybrano parametry barwy modelu RGB oraz parametry geometryczne ziarniaków. Analizę funkcji dyskryminacji przeprowadzono w pakiecie SAS 9.1. Zaprezentowano kolejne etapy tej analizy oraz składnię programów w języku 4GL, umożliwiającą jej wykonanie.

Słowa kluczowe: analiza funkcji dyskryminacji, cyfrowa analiza obrazu, jakość browarna, jęczmień jary, model RGB, SAS

Klasyfikacja AMS 2000: 62–07, 62H20, 62H30, 62H35

Barley grain is the main source for production of malt. Varietal purity is essential for production of good quality malt. Each malting barley cultivar has a description of characters prepared on the basis of micromalting tests conducted according to the criteria of the European Brewing Convention (EBC). Before sending barley grain of a certain cultivar to a brewery, both the efficiency and quality of malt are to be evaluated. Quality of a cultivar is assessed according to the Q index where Q value below 5 specifies feed cultivars and Q value near 9 specifies very good malting cultivars. Hitherto malting barley quality assessment has been conducted using laboratory methods. Digital image analysis might facilitate the barley quality assessment procedure and reduce its costs. Discriminant analysis and digital image analysis were used in classification of grain of 11 spring barley cultivars to the quality groups. Discriminant analysis was based on morphological features and color features characteristic for RGB color space. Discriminant analysis was conducted using SAS 9.1. Successive steps of the analysis and procedures syntax in the 4GL programming language have been presented.

Key words: digital image analysis, discriminant function analysis, malting quality, RGB model, SAS, spring barley

AMS 2000 Classification: 62-07, 62H20, 62H30, 62H35

WSTĘP

Jakość słodu jęczmiennego jest uwarunkowana wieloma cechami surowca, spośród których ogromne znaczenie mają cechy odmianowe jęczmienia browarnego. Przy ocenie wartości browarnej odmian jęczmienia szczególne znaczenie ma struktura bielma ziarniaków. Opracowano wiele metod oceny przydatności ziarna jęczmienia do słodowania. Rozważa się również wykorzystanie w tym celu techniki cyfrowej analizy obrazu, w wyniku której uzyskuje się szereg parametrów barwy oraz cech geometrycznych ziarniaków i bielma jęczmienia. Ilość możliwych do uzyskania w tej analizie parametrów utrudnia wybór tych, które decydują o przypisaniu ziarna jęczmienia do określonej klasy jakości browarnej.

Z powodzeniem wykorzystano cyfrową analizę obrazu do identyfikacji odmianowej ziarna roślin zbożowych, w tym również do identyfikacji odmian jęczmienia jarego (Majumdar i Jayas 2000 a, 2000 b, 2000 c; Kozirok, 2005).

Podjęto również badania nad zastosowaniem komputerowej analizy obrazu w ocenie jakości ziarna zbóż. Stosując analizę obrazu w ocenie cech jakościowych ziarna pszenicy ozimej stwierdzono, występowanie istotnych i silnych korelacji pomiędzy parametrami geometrycznymi a wyróżnikami jakości ziarna pszenicy ozimej (Majewska, 2004).

Celem niniejszej pracy była próba zastosowania cyfrowej analizy obrazu i analizy funkcji dyskryminacji w ustaleniu cech, które istotnie różnicują jakość browarną ziarna jęczmienia jarego oraz prezentacja wykorzystania pakietu SAS 9.1 (SAS Institute Inc., 2004 a, 2004 b, 2004 c) do przeprowadzenia analizy funkcji dyskryminacji. Do analiz wybrano te cechy pochodzące z analizy obrazu, które umożliwiają identyfikację odmianową ziarna jęczmienia jarego (Kozirok, 2005). Praca ma na celu pokazanie możliwości komputerowej analizy obrazu oraz przybliżenie idei analizy dyskryminacyjnej.

PRZEGLĄD LITERATURY

Jęczmień jest zbożem szeroko wykorzystywanym w słodownictwie. O przydatności ziarna jęczmienia do produkcji słodu decyduje przede wszystkim jego skład chemiczny

oraz wyrównanie ziarna, które oznacza się jako procentowy udział ziarna pozostającego na sitach o wymiarach oczek $2,8 \text{ mm} \times 25 \text{ mm}$ oraz $2,5 \text{ mm} \times 25 \text{ mm}$. Jedynie te dwie frakcje ziarna wykorzystuje się w technologii produkcji słoðu (Kunze, 1999; Błażewicz i Liszewski, 2003).

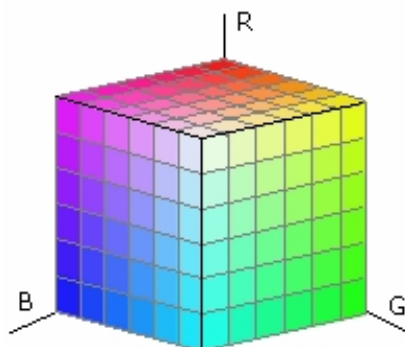
Ostateczne zaklasyfikowanie ziarna jęczmienia browarnego do określonych grup jakościowych następuje po wyznaczeniu wskaźnika wartości browarnej Q (Molina-Cano, 1987). Jest to średnia ważona następujących parametrów: ekstraktywności słoðu, liczby Kolbacha, stopnia ostatecznego odfermentowania, lepkości brzezki i siły diastatycznej (Molina-Cano, 1987; Projekt Eureka, 2005).

Wiele spośród cech wpływających na jakość browarną ziarna jęczmienia jest uwarunkowana genetycznie i silnie zależna od odmiany (Projekt Eureka, 2005). Obecnie stosuje się dwie metody klasyfikacji odmianowej ziarna jęczmienia. Pierwsza opiera się na elektroforetycznym rozdziale białek hordeinowych na żelu poliakryloamidowym w środowisku kwaśnym, druga metoda opiera się na wizualnej obserwacji cech morfologicznych opisujących jej fenotyp (Drzewiecki i in., 2000; Lewis i Young, 2001; Projekt Eureka, 2005). Alternatywę dla tych metod stanowi analiza obrazu, a w szczególności barwy i cech geometrycznych ziarniaków oraz bielma jęczmienia browarnego. Dotychczasowe badania potwierdziły, iż zarówno barwa, jak i cechy geometryczne ziarniaków i bielma mogą być cechami zależnymi od odmiany (Kozirok, 2005).

Majumdar i Jayas (2000 a, 2000 b, 2000 c), w badaniach nad odmianami pszenicy, jęczmienia, żyta i owsa, pobranych z 20 obszarów północnej Kanady, porównali efektywność klasyfikacji gatunkowej i odmianowej, przeprowadzonej na podstawie cech morfologicznych (w szczególności cech geometrycznych) i parametrów barwy powierzchni ziaren, uzyskanych dzięki cyfrowej analizie obrazu, rejestrowanego kamerą CCD. Autorzy wykazali, że klasyfikacja przeprowadzona wyłącznie na podstawie cech morfologicznych była mniej trafna w porównaniu do klasyfikacji odmian, opartej zarówno na cechach morfologicznych, jak i parametrach barwy ziarniaków.

Cechy geometryczne ziarniaków są typowymi cechami ilościowymi i mierzalnymi. Również parametry barwy można przedstawić za pomocą liczb dzięki zastosowaniu tak zwanych matematycznych modeli barw.

Jednym z modeli przestrzeni barw jest model RGB opracowany w 1931 roku przez CIE — Międzynarodową Komisję Oświeceniową (rys. 1). Jest to model liniowy oparty o addytywne mieszanie barw. Wszystkie barwy w przestrzeni RGB można uzyskać przez zmieszanie w odpowiednich proporcjach trzech głównych składowych: R — barwy czerwonej (Red), odpowiadającej wrażeniu wywołanemu przez promieniowanie o długości fali 700 nm, G — barwy zielonej (Green), odpowiadającej wrażeniu wywołanemu przez promieniowanie o długości fali 546,1 nm oraz B — barwy niebieskiej (Blue), odpowiadającej wrażeniu wywołanemu przez promieniowanie o długości fali 435,8 nm. Barwę achromatyczną uzyskuje się w wyniku addytywnego zmieszania jednakowych ilości tych składowych (Mielicki, 1997). Model RGB jest wykorzystywany w komputerowej analizie obrazu cyfrowego. W technice cyfrowej każdy kanał barwy jest opisany liczbą 8 bitową, co pozwala na uzyskanie 256 (2^8) poziomów jasności. Tak zbudowany model RGB obejmuje około 16 milionów kolorów (Hebisz, 2003).



Rys. 1. Przestrzeń barw RGB
Fig. 1. The RGB color cube

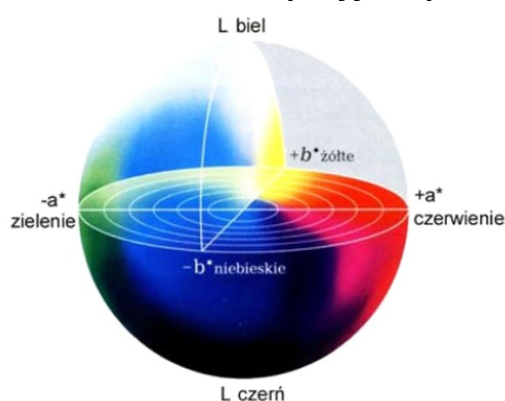
Innym szeroko stosowanym modelem, odpowiadającym indywidualnemu odczuciu barwy jest model CIE La^*b^* . Model ten odwzorowuje uporządkowaną, trójwymiarową przestrzeń barw i stanowi kolorymetryczny standard, obowiązujący na całym świecie. Barwę opisuje się w nim za pomocą trzech współrzędnych: L — jasności (Lightness), a^* — barwy zielonej do czerwonej, b^* — barwy niebieskiej do żółtej. Wszystkie barwy o jednakowej jasności leżą na powierzchni kołowej, wyznaczonej osiami a^* i b^* (rys. 2). Model La^*b^* oparty jest na systemie barw przeciwstawnych, a więc takich, które wzajemnie się wykluczają. Inaczej mówiąc, barwa nie może być jednocześnie czerwona i zielona ani też żółta i niebieska. Dodatnie wartości a^* odpowiadają więc czerwieni, a ujemne zieleni. Natomiast dodatnie wartości b^* odpowiadają żółci, a ujemne barwie niebieskiej. Jasność L zmienia się w kierunku pionowym. Parametry tego modelu w pełni definiują barwę, a jej różnicę można przedstawić zależnością:

$$\Delta E = \sqrt{(\Delta L)^2 + (\Delta a^*)^2 + (\Delta b^*)^2}$$

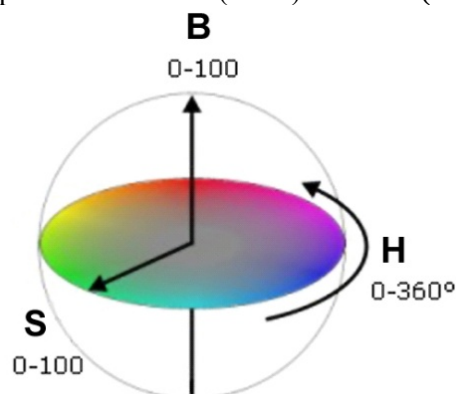
Dzięki ΔE można precyzyjnie określić jak bardzo dwie barwy różnią się od siebie. Model La^*b^* ma ogromne znaczenie zarówno w grafice komputerowej, jak i w praktyce drukarskiej, ponieważ znajduje zastosowanie w najnowszych systemach elektronicznego sterowania regulacją farby w procesie druku, a także przy produkcji wzorców przemysłowych, które z kolei umożliwiają porównanie zakresów barwowych różnych urządzeń wejściowych i wyjściowych oraz ich kalibrację. Podstawową zaletą tego modelu jest to, że korekcja barw jest wolna od wpływów subiektywnego ich odczuwania i zmiennego oświetlenia (Pastuszek, 2000).

Inną propozycją przestrzeni barw jest model HSB (rys. 3), w którym zastosowano bardziej intuicyjny opis barwy. Podstawą tego modelu jest sposób postrzegania koloru przez człowieka. Model HSB wykorzystuje do opisu każdej barwy trzy podstawowe atrybuty. Odcień (Hue), który odpowiada długości fali światła odbitego lub przechodzącego przez obiekt. Jego miarą jest położenie na standardowym kole barw, wyrażone w stopniach od 0° do 360° . Nasycenie (Saturation), oznaczające siłę koloru, czyli stosunek szarości do czystego odcienia, wyrażane w procentach od 0% (szary) do 100%

(czysty kolor, pełne nasycenie). Jaskrawość (Brightness), która określa względną jaskrawość koloru i zazwyczaj jest wyrażana w procentach od 0% (czern) do 100% (biel).



Rys. 2. Przestrzeń barw $L^*a^*b^*$
Fig. 2. The $L^*a^*b^*$ color space



Rys. 3. Przestrzeń barw HSB
Fig. 3. The HSB color space

Analizę funkcji dyskryminacji (Fisher, 1936; Huberty, 1975; Klecka, 1980; Hand, 1981; Krzyśko, 1990) wykorzystuje się do rozstrzygania, które zmienne różnicują dwie lub więcej naturalnie wyłaniających się grup. Znajduje ona zastosowanie w wielu dziedzinach nauki, w tym także w naukach przyrodniczych, między innymi w ocenie zmienności genetycznej roślin uprawnych (McElroy i in., 2002; Vaylay i Santen, 2002) oraz w ocenie zmienności innych cech roślin uprawnych, np. plonowania (Jaynes i in., 2003; Weber i Zalewski, 2004).

MATERIAŁ I METODY

Do badań wykorzystano ziarno dominującej frakcji (powyżej 2,5 mm) 11 odmian jęczmienia jarego, które na podstawie laboratoryjnej oceny wartości browarnej w roku 2001 (Lista Odmian COBORU, 2001) zostały zakwalifikowane do różnych klas jakości browarnej (tab. 1). Kwalifikację odmian do określonej klasy jakości browarnej przeprowadzono w dwóch wariantach: wariant I — C dla $Q < 5$, B dla $Q \in (5; 6)$, A dla $Q \in (6; 8)$ oraz E dla $Q \geq 8$ (Projekt Eurequa, 2005); II — C dla $Q < 5,0$, B dla $Q \in (5,0; 6,5)$, A dla $Q \in (6,5; 8,0)$ oraz E dla $Q \geq 8,0$. Dla zachowania równej liczebności w klasach, w analizach, w przypadku II wariantu, nie uwzględniono odmian Brenda i Rasbet. Odmiany wzięte do badań pochodziły zarówno z polskich hodowli, jak i zagranicznych.

Tabela 1

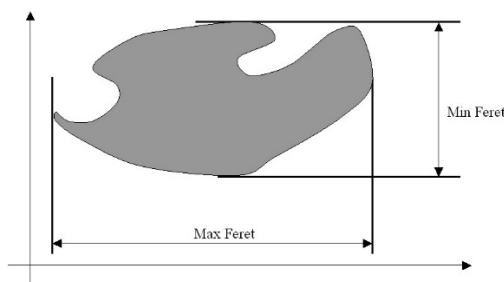
Wykaz wybranych odmian jęczmienia jarego
List of chosen spring barley cultivars

Odmiana Cultivar	RO*	RW**	Hodowca Breeder	Q***	Klasa**** Class****	
					wariant I variant I	wariant II variant II
Atol	1997	...	Hodowla Roślin Strzelce Sp. z o. o.	6,15	A	B
Barke	2001	...	Saatzucht Josef Breun GbR	8,00	E	E
Brenda	1998	...	Sekundo Saatucht GmbH	7,90	A	—
Granal	2001	...	Poznańska Hodowla Roślin Sp. z o. o.	8,30	E	E
Orlik	1990	2000	Hodowla Roślin Smolice Sp. z o. o.	6,70	A	A
Poldek	1999	...	Małopolska Hodowla Roślin Sp. z o. o.	6,20	A	B
Polo	1992	2003	Hodowla Roślin Strzelce Sp. z o. o.	5,75	B	B
Rasbet	1998	...	Hodowla Roślin Smolice Sp. z o. o.	7,95	A	—
Scarlett	1999	...	Saatzucht Josef Breun GbR	7,45	A	A
Sezam	2000	...	Hodowla Roślin Szelejowo Sp. z o. o.	8,10	E	E
Stratus	1999	...	Hodowla Roślin Strzelce Sp. z o. o.	6,55	A	A

* — Rok wpisania do rejestru; The year of registration ** — Rok wykreślenia z rejestru; The year of removing from crop plants register; *** — Q — wartość browarna; Q — malting quality **** — Klasa jakości browarnej; Malting class

Pomiar barwy oraz cech geometrycznych w płaszczyźnie przekrojów podłużnych jęczmienia browarnego wykonano na trzystu ziarniakach każdej z badanych odmian przy zastosowaniu systemu cyfrowej analizy obrazu, pracującej w oparciu o oprogramowanie LUCIA ver. 4.61 (*Laboratory Universal Computer Image Analysis*). Stanowisko pomiarowe systemu cyfrowej analizy obrazu składało się z kamery Panasonic GP-KR222E o rozdzielczości 1044 dpi, dwupunktowego oświetlenia światłowodowego BOB OM 100x1 o natężeniu oświetlenia 60 kLx, zakresie widma na wyjściu światłowodu 410–700nm i temperaturze barwowej 3000°K, komputera z kartą akwizycji obrazu VFG, oprogramowania oraz monitora. W oparciu o dostępne w oprogramowaniu algorytmy przetwarzania i analizy obrazu wyznaczono zespół cech geometrycznych oraz barwy, wyznaczając wartości składowych barwy RGB, atrybutów barwy HSB oraz parametrów La^*b^* . Oznaczono następujące cechy geometryczne ziarniaków jęczmienia:

- AREA — powierzchnia rzutu na płaszczyznę;
- ELONGATION — wydłużenie ziarniaka;
- EQDIAMETER — średnica koła o polu powierzchni równym polu rzutu ziarniaka;
- LENGTH — długość ziarniaka;
- MAX_FERET i MIN_FERET — największa i najmniejsza wartość średnicy Fereta. Średnicą Fereta nazywamy: dla osi 0x — największą odległość pomiędzy wartościami współrzędnych x_i punktów (x_i, y_i) obiektu, a dla osi 0y — największą odległość pomiędzy wartościami współrzędnych y_i punktów (x_i, y_i) obiektu (rys. 4.) (Zieliński i Strzelecki, 2002);
- PERIMETER — obwód ziarniaka;
- VOLEQCYLINDER — współczynnik cylindryczności ziarniaka;
- VOLEQSPHERE — współczynnik sferyczności ziarniaka;
- W_2 — stosunek szerokości do długości, tak zwany współczynnik baryłkowatości;
- WIDTH — szerokość ziarniaka.



Rys. 4. Średnice Fereta
Fig. 4. Feret's diameters

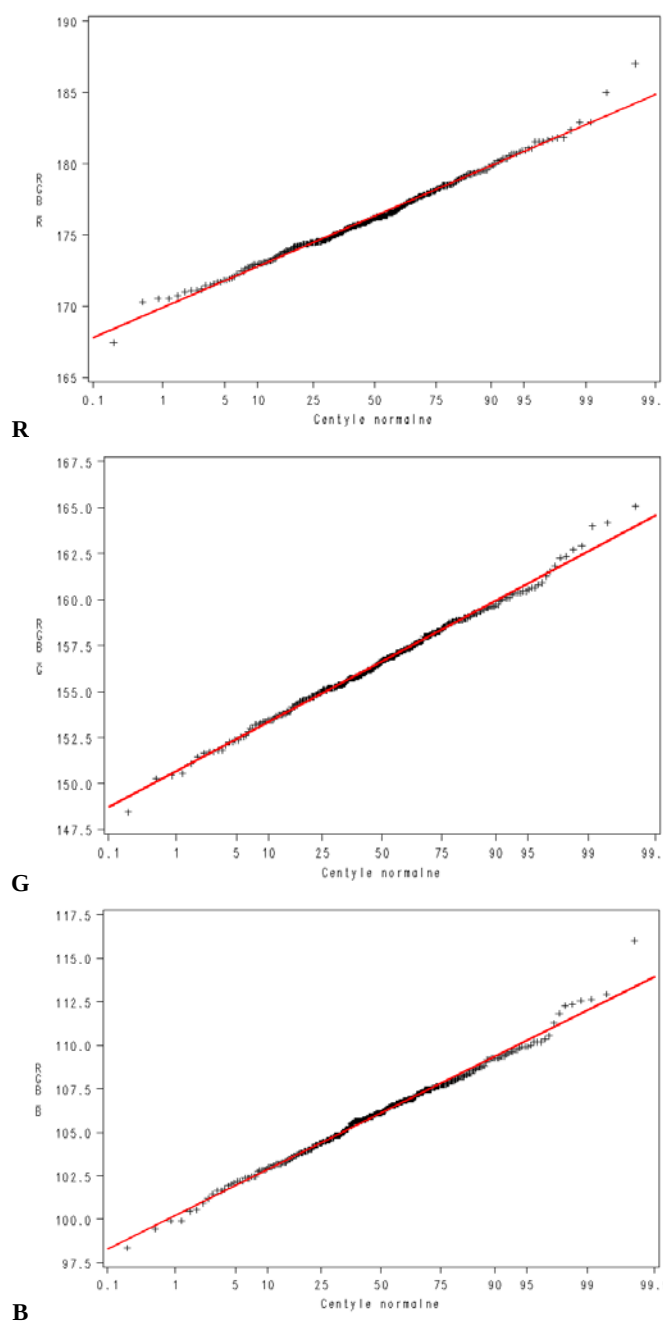
Do klasyfikacji ziarna do grup jakościowych można wykorzystać kilka metod, między innymi modelowanie neuronowe, a w szczególności autoasocjacyjną sieć Hopfielda lub samoorganizującą sieć Kohonena (Osowski, 1996; Trajer i Sałat, 2005), metodę SVM (Osowski, 2002; Trajer i Sałat, 2005) oraz analizy statystyczne (analiza dyskryminacyjna). Do klasyfikacji odmian jęczmienia jarego do grup jakościowych w niniejszej pracy wykorzystano analizę funkcji dyskryminacji. Do weryfikacji wyestymowanego modelu wykorzystano dane oryginalne, na których model ten został zbudowany oraz dane pochodzące z bootstrapu.

Metoda bootstrapu została zaproponowana przez Bradleya Efrona w 1979 roku. Stosowana jest między innymi przy estymacji nieznanymi, bądź trudnymi do wyznaczenia rozkładów statystyk poprzez wielokrotne próbkowanie z próby pierwotnej (tak zwany *Resampling*) zgodnie z procesem generującym te dane. W niniejszej pracy zastosowano metodę bootstrapu polegającą na wylosowaniu z próby pierwotnej po 2400 obserwacji dla każdej klasy jakości browarnej jęczmienia jarego (łącznie po 7200 obserwacji). Losowania wykonano za pomocą makroprogramu arkusza kalkulacyjnego Ms Excel napisanego w języku VBA (*Visual Basic for Application*). Losowania wykonano oddzielnie dla obu wariantów klasyfikacji jakości browarnej.

Analizy statystyczne wykonano z wykorzystaniem pakietu SAS 9.1 (SAS Institute Inc., 2004 a, 2004 b, 2004 c). Przeprowadzono test normalności Shapiro-Wilka oraz analizę funkcji dyskryminacji dla badanych parametrów ziarniaków jęczmienia jarego i ich wpływu na jego jakość browarną. Powyższe analizy są dostępne w module SAS/STAT pakietu SAS 9.1 (SAS Institute Inc., 2004 c; Shapiro i Wilk, 1965; Khattre i Naik, 200; O'Rourke i in., 2005).

OMÓWIENIE WYNIKÓW I DYSKUSJA

Barwę ziarniaków wyrażono za pomocą parametrów charakterystycznych dla modeli barwy RGB, La^*b^* i HSB. W analizie funkcji dyskryminacji zmienne opisujące powinny mieć rozkład zbliżony do rozkładu normalnego. Wybór odpowiedniego modelu barwy uzależniono od wyniku testu normalności rozkładów parametrów barwy.



Rys. 5. Wykresy probabilistyczne dla porównania rozkładów parametrów barwy RGB_R, RGB_G i RGB_B z rozkładem normalnym
Fig. 5. Probability plot for comparison of distributions of color parameters RGB_R, RGB_G and RGB_B with normal distribution

Do testowania normalności rozkładu wykorzystano test Shapiro-Wilka, ponieważ liczba obserwacji nie przekraczała 2000 (Shapiro i Wilk, 1965). Użyto w tym celu procedurę UNIVARIATE. Składnia tej procedury ma następującą postać:

```
PROC UNIVARIATE DATA=JECZMIEN_BARWA NORMALTEST;
  PROBLOT RGB_R RGB_G RGB_B / NORMAL (MU=EST SIGMA=EST
    COLOR=BLUE W=2);
RUN;
```

Analiza wyników testu normalności dla parametrów barwy ziarniaków jęczmienia jarego wykazała, że w przypadku wszystkich odmian, jedynie parametry modelu RGB charakteryzują się rozkładem zbliżonym do rozkładu normalnego (nieistotne statystyki testowe uniemożliwiły odrzucenie hipotezy zerowej o normalności rozkładu cechy) (tab. 2; rys. 5). Do analizy funkcji dyskryminacji wykorzystano więc parametry barwy w modelu RGB.

Tabela 2

Wyniki testu normalności Shapiro-Wilka dla parametrów barwy ziarniaków wybranych odmian jęczmienia jarego

Results of Shapiro-Wilk's normality test for kernels color parameters of chosen spring barley cultivars

Odmiana Cultivar	Wartości testu Shapiro-Wilka Shapiro-Wilk test values								
	RGB			La*b*			HSB		
	R	G	B	L	a	b	H	S	B
Atol	0,9891	0,9857	0,9860	0,9929	0,9478	0,8923*	0,3875*	0,9895	0,9862
Barke	0,9820	0,9821	0,9827	0,9758	0,8943*	0,8911*	0,4511*	0,8906*	0,9834
Brenda	0,9940	0,9952	0,9957	0,9961	0,8441*	0,9948	0,7656*	0,8956*	0,8966*
Granal	0,9867	0,9870	0,9914	0,9827	0,9779	0,9902	0,6738*	0,8913*	0,9893
Orlik	0,9837	0,9840	0,9862	0,9749	0,9477	0,9888	0,6940*	0,8919*	0,9848
Poldek	0,9852	0,9864	0,9925	0,9817	0,8917*	0,9875	0,8057*	0,8942*	0,9877
Polo	0,9902	0,9930	0,9947	0,8907*	0,9470	0,9898	0,7974*	0,8909*	0,8940*
Rasbet	0,9917	0,9940	0,9918	0,8914*	0,9871	0,9677	0,8061*	0,9723	0,8955*
Scarlett	0,9926	0,9906	0,9927	0,9897	0,9126	0,9735	0,8534*	0,9851	0,8921*
Sezam	0,9891	0,9930	0,9943	0,8916*	0,8962*	0,8931*	0,7892*	0,9901	0,8946*
Stratus	0,9873	0,9904	0,9923	0,9867	0,9862	0,8917*	0,6174*	0,9899	0,8919*

* — Istotne przy $\alpha = 0,05$; Significant at $\alpha = 0,05$

Analizę funkcji dyskryminacji przeprowadzono w dwóch etapach. W pierwszym etapie wyodrębniono grupę zmiennych, które najlepiej odzwierciedlały różnice między klasami jakości browarnej. W tym celu wykorzystano procedurę STEPDISC. Analizę przeprowadzono metodą krokową. W każdym kroku testowano zarówno wprowadzenie nowej zmiennej do modelu, jak i usunięcie zmiennej spośród zmiennych znajdujących się już w modelu (Jennrich, 1977; Klecka, 1980; Khattre i Naik, 2000). Test wprowadzenia zmiennej do modelu i usunięcia zmiennej z modelu przeprowadzono na poziomie istotności $\alpha = 0,01$. Składnia tego proc-stepu ma postać następującą:

```
PROC STEPDISC DATA=JECZMIEN_BROWARNY METHOD=STEPWISE
  SLENTRY=0.01 SLSTAY=0.01;
  CLASS KLASA_BR;
  VAR RGB_R RGB_G RGB_B AREA LENGTH WIDTH W2 MAX_FERET MIN_FERET
    EQDIAMETER VOLEQSPHERE VOLEQCYLINDER PERIMETER ELONGATION;
RUN;
```

```

PROC STEPDISC DATA=JECZMIEN_BROWARNY2 METHOD=STEPWISE
  SLENTRY=0.01 SLSTAY=0.01;
CLASS KLASA_BR_2;
VAR RGB_R RGB_G RGB_B AREA LENGTH WIDTH W2 MAX_FERET MIN_FERET
  EQDIAMETER VOLEQSPHERE VOLEQCYLINDER PERIMETER ELONGATION;
RUN;

```

Procedura ta dla każdej zmiennej dodawanej bądź usuwanej z modelu wyznacza cząstkowy współczynnik determinacji (R^2) oraz statystykę F na podstawie jednokierunkowej analizy kowariancji. Zmienne do modelu są dodawane, gdy ich tolerancja oraz tolerancje wszystkich zmiennych, które już w modelu się znajdują są większe niż założono, domyślnie jest to wartość 1×10^{-8} . Tolerancja dla zmiennej wchodzącej jest wyznaczana jako $1 - R^2$ z regresji tej zmiennej względem zmiennych, znajdujących się już w modelu. Tolerancja dla zmiennych znajdujących się w modelu jest wyznaczana jako $1 - R^2$ z regresji każdej zmiennej względem pozostałych zmiennych w modelu i zmiennej wchodzącej do modelu. W podobny sposób odbywa się usuwanie zmiennych z modelu. W tej procedurze przeprowadzany jest również test lambda Wilksa, test śladu Pillai (Pillai, 1985; Khattre i Naik, 2000) oraz wyznaczana jest przeciętna kwadratowa korelacja kanoniczna (ASCC — *Average Squared Canonical Correlation*). Jeżeli do modelu nie można już dodać żadnej zmiennej, oraz gdy z modelu nie można już żadnej zmiennej usunąć, następuje zakończenie procedury STEPDISC oraz wyświetlenie podsumowania (tab. 3).

Tabela 3

Podsumowanie selekcji krokowej procedury STEPDISC
STEPDISC procedure stepwise selection summary

Krok Step	Liczba zmiennych Number of variables	Wstawiono do modelu Entered to model	Usunięto z modelu Removed from model	Cząstkowy R^2 Partial R^2	Statystyka F F-statistic	Lambda Wilks	ASCC
variant I — variant I							
1	1	MAX_FERET	—	0,1297	245,61**	0,87033**	0,06483**
2	2	ELONGATION	—	0,0675	119,27**	0,81159**	0,09732**
3	3	VOLEQCYLINDER	—	0,0379	64,87**	0,78085**	0,11449**
4	4	WIDTH	—	0,0373	63,80**	0,75173**	0,13026**
5	5	RGB_R	—	0,0326	55,42**	0,72725**	0,14337**
6	6	RGB_G	—	0,0216	36,36**	0,71153**	0,15183**
7	7	RGB_B	—	0,0171	28,69**	0,69934**	0,15959**
8	8	LENGTH	—	0,0132	22,03**	0,69009**	0,16516**
9	9	VOLEQSPHERE	—	0,0301	51,00**	0,66934**	0,17748**
10	10	EQDIAMETER	—	0,0320	54,29**	0,64794**	0,19127**
11	11	MIN_FERET	—	0,0098	16,24**	0,64160**	0,19542**
12	10	—	ELONGATION	0,0018	2,89 ^{NS}	0,64273**	0,19468**
variant II — variant II							
1	1	RGB_B	—	0,1514	240,61**	0,84859**	0,07571**
2	2	MAX_FERET	—	0,0943	140,41**	0,76854**	0,12281**
3	3	ELONGATION	—	0,0299	41,52**	0,74557**	0,13634**
4	4	VOLEQSPHERE	—	0,0236	32,60**	0,72795**	0,14675**
5	5	EQDIAMETER	—	0,1052	158,25**	0,65139**	0,19212**
6	6	RGB_G	—	0,0225	31,04**	0,63671**	0,20089**
7	7	RGB_R	—	0,0291	40,30**	0,61819**	0,21206**
8	8	VOLEQCYLINDER	—	0,0154	21,00**	0,60869**	0,21844**
9	9	WIDTH	—	0,0055	7,49**	0,60532**	0,22068**

** — Istotne przy $\alpha = 0,01$; Significant at $\alpha = 0.01$

^{NS} — Nieistotne; Not significant

Zmienne wymienione w tabeli 3 w sposób istotny różnicują klasy jakości browarnej, dla zmiennych, które nie zostały uwzględnione w tej tabeli, nie stwierdzono istotnego wpływu na jakość browarną. Zmienne, które wykazują interakcję ze zmiennymi włączonymi do modelu, przez co same pogarszają jakość modelu i są z niego usuwane (tab. 3, zmienna ELONGATION, wariant I, krok 12).

Powyższa analiza pozwoliła na wyselekcjonowanie zmiennych, które istotnie dyskryminują przynależność badanych odmian do klas jakości browarnej jęczmienia jarego. W przypadku wariantu I tymi cechami są: MAX_FERET, VOLEQCYLINDER, WIDTH, RGB_R, RGB_G, RGB_B, LENGTH, VOLEQSPHERE, EQDIAMETER oraz MIN_FERET. Nie stwierdzono natomiast by współczynnik W_2 , AREA, ELONGATION i PERIMETER mogły dyskryminować przynależność badanych odmian do wyznaczonych klas. W przypadku wariantu II do modelu włączone zostało dziewięć zmiennych (RGB_B, MAX_FERET, ELONGATION, VOLEQSPHERE, EQDIAMETER, RGB_G, RGB_R, VOLEQCYLINDER, WIDTH), pozostałe zmienne nie wpływają istotnie na klasyfikację ziarna jęczmienia jarego do klas jakości browarnej.

W drugim etapie za pomocą procedury DISCRIM wyznaczono kryteria dyskryminacji, czyli tak zwaną funkcję dyskryminacji. Funkcja dyskryminacji jest determinowana przez ogólną odległość kwadratową (Rao, 1973). Kryterium klasyfikacji może opierać się na indywidualnych wewnątrzgrupowych macierzach kowariancji lub na wspólnej macierzy kowariancji. Składnia tej procedury jest następująca:

```
PROC DISCRIM DATA=JECZMIEN_BROWARNY DISTANCE TESTDATA=JB_BOOTST;
  CLASS KLASA_BR;
  VAR RGB_R RGB_G RGB_B LENGTH WIDTH MAX_FERET MIN_FERET
      EQDIAMETER VOLEQSPHERE VOLEQCYLINDER;
  PRIORS EQUAL;
RUN;
PROC DISCRIM DATA=JECZMIEN_BROWARNY2 DISTANCE TESTDATA=JB2_BOOTST;
  CLASS KLASA_BR_2;
  VAR RGB_R RGB_G RGB_B WIDTH MAX_FERET ELONGATION
      EQDIAMETER VOLEQSPHERE VOLEQCYLINDER;
  PRIORS EQUAL;
RUN;
```

W wyniku zastosowania tej procedury otrzymano zestawienia zawierające liczebność obserwacji w grupach, wyznaczone wagi, udział liczby obserwacji z danej grupy w ogólnej liczbie obserwacji oraz wyznaczone prawdopodobieństwo *a priori* dla każdej grupy (tab. 4).

W tym przypadku przyjęto jednakowe prawdopodobieństwa dla każdej z grup (PRIORS EQUAL). Otrzymano również następujące informacje o macierzach kowariancji: ranga macierzy kowariancji (w tym przypadku = 10 dla wariantu I oraz 9 dla wariantu II) oraz logarytm naturalny wyznacznika macierzy kowariancji (w tym przypadku = -10,54507 dla wariantu I oraz -4,93531 dla wariantu II). Zastosowana opcja DISTANCE dała w wyniku macierz odległości D^2 Mahalanobisa pomiędzy klasami jakości jęczmienia oraz test F istotności tych odległości (tab. 5).

Tabela 4

Zestawienie informacji o grupach
Class level information

Grupa Class	Liczebność obserwacji Number of observations	Waga Weight	Udział Proportion	Prawdopodobieństwo <i>a priori</i> Prior probability
wariant I — variant I				
A	2100	2100	0,636364	0,333333
B	300	300	0,090909	0,333333
E	900	900	0,272727	0,333333
wariant II — variant II				
A	900	900	0,333333	0,333333
B	900	900	0,333333	0,333333
E	900	900	0,333333	0,333333

Tabela 5

Macierz odległości Mahalanobisa
Mahalanobis distance matrix

Grupa Class	A		B		E	
	D^2	Statystyka F F-statistic	D^2	Statystyka F F-statistic	D^2	Statystyka F F-statistic
wariant I — variant I						
A	0	—	2,15882	56,51431**	1,54418	97,01758**
B	2,15882	56,51431**	0	—	4,54162	101,90745**
E	1,54418	97,01758**	4,54162	101,90745**	0	—
wariant II — variant II						
A	0	—	1,30327	64,97008**	2,06350	102,86894**
B	1,30327	64,97008**	0	—	1,80182	89,82384**
E	2,06350	102,86894**	1,80182	89,82384**	0	—

** — Istotne przy $\alpha = 0,01$ ** — Significant at $\alpha = 0.01$.

Wyznaczono liniowe funkcje rozpoznawczą dla zmiennej klasyfikującej (tab. 6). W tym przypadku zmienna klasyfikująca grupuje obserwacje w trzech klasach, otrzymano więc trzy funkcje liniowe.

Dla każdej obserwacji wyliczono wartości trzech funkcji liniowych, następnie klasyfikowano ją do tej grupy, dla której wyliczona funkcja przyjęła najwyższą wartość.

Procedura DISCRIM wykonuje test wyznaczonych funkcji dla wszystkich obserwacji. Możliwe jest również wykonanie testu na podstawie nowej grupy danych. Dokładność modelu określa się na podstawie liczby obserwacji, prawidłowo zaklasyfikowanych do grup, przez liniową funkcję rozpoznawczą (tab. 7, tab. 8).

Weryfikację wyestymowanego modelu wykonano w oparciu o dane oryginalne, na których zbudowano model, oraz o dane pochodzące z bootstrapu. W bootstrapie wylosowano z próby podstawowej (danych oryginalnych) po 2400 obserwacji dla każdej z klas jakości browarnej ziarna jęczmienia jarego (łącznie po 7200 obserwacji dla obu wariantów).

Zastosowanie uzyskanej liniowej funkcji rozpoznawczej do klasyfikacji danych wejściowych nie dało zadowalających wyników. Spośród obserwacji dla odmian klasy A do tej klasy zostało zakwalifikowane w przypadku wariantu I tylko 65,33% dla danych oryginalnych i 41,08% dla danych z bootstrapu; w przypadku wariantu II uzyskano

odpowiednio 65,56% dla danych oryginalnych i 45,96% dla danych z bootstrapu; w klasie B w przypadku wariantu I było to 77,00% dla danych oryginalnych i 40,63% dla danych z bootstrapu; w przypadku wariantu II uzyskano odpowiednio 61,33% dla danych oryginalnych i 20,75% dla danych z bootstrapu; natomiast w klasie E w przypadku wariantu I — 52,78% dla danych oryginalnych i 36,21% dla danych z bootstrapu; w przypadku wariantu II uzyskano odpowiednio 53,56% dla danych oryginalnych i 48,17% dla danych z bootstrapu. Błędnie zostało zaklasyfikowane w przypadku wariantu I 35% obserwacji oryginalnych i 61% obserwacji pochodzących z bootstrapu; w przypadku wariantu II — 40% obserwacji oryginalnych i 62% obserwacji z bootstrapu. Może to świadczyć o tym, że parametry zastosowane do budowy funkcji rozpoznawczej, choć są wystarczające do identyfikacji odmianowej (Koziorok, 2005), to nie są wystarczające do opisanja jakości browarnej ziarna jęczmienia jarego.

Tabela 6

Liniowa funkcja rozpoznawcza Linear discriminant function			
Zmienna — Variable	A	B	E
wariant I — variant I			
Stała — Constant	-5952,00000	-5947,00000	-5877,00000
RGB_R	-0,48095	-0,57907	-0,35246
RGB_G	10,49419	10,46977	10,27787
RGB_B	-10,92610	-10,82889	-10,79758
LENGTH	-37,80227	-34,17880	-36,15294
WIDTH	-557,39524	-522,40872	-544,64166
MAX_FERET	-39,03031	-37,22094	-36,68625
MIN_FERET	4,08237	14,86063	4,00120
EQDIAMETER	3810,00000	3780,00000	3772,00000
VOLEQSPHERE	-92,03175	-91,60322	-91,20641
VOLEQCYLINDER	14,68923	13,67355	14,39240
wariant II — variant II			
Stała — Constant	-7115,00000	-7083,00000	-7014,00000
RGB_R	-1,10021	-1,12394	-0,99935
RGB_G	11,84777	11,82660	11,62074
RGB_B	-9,70164	-9,75638	-9,59343
WIDTH	-125,95297	-118,58388	-119,93672
MAX_FERET	-256,46454	-257,53183	-255,33331
ELONGATION	420,79151	422,99591	422,97941
EQDIAMETER	4168,00000	4159,00000	4132,00000
VOLEQSPHERE	-88,88913	-88,47041	-87,90084
VOLEQCYLINDER	2,60349	2,18154	2,27637

Pomimo, tak słabych efektów zastosowania komputerowej analizy obrazu do klasyfikacji ziarna jęczmienia jarego do klas jakości browarnej, w przyszłości, po dopracowaniu metody i wybraniu lepszych parametrów pozyskanych z analizy obrazu, prawdopodobnie uda się zastąpić używaną obecnie metodę laboratoryjną oceny jakości, metodą tańszą, mniej uciążliwą, a przede wszystkim szybszą. Wskaźnik *Q* jest wskaźnikiem kompleksowym, stąd też i zależności pomiędzy barwą i cechami geometrycznymi ziarniaków a wartością *Q* mogą być niewystarczające do opracowania nowej i pewnej metody.

Tabela 7

Podsumowanie klasyfikacji obserwacji do grup za pomocą wyznaczonych funkcji liniowych, wariant I
Resubstitution summary using linear discriminant function, variant I

Grupa Class	A		B		E		Razem — Total	
	liczność number of observations	%	liczność number of observations	%	liczność number of observations	%	liczność number of observations	%
dane oryginalne — original data								
A	1372	65,33	428	20,38	300	14,29	2100	100,00
B	54	18,00	231	77,00	15	5,00	300	100,00
E	315	35,00	110	12,22	475	52,78	900	100,00
Razem Total	1741	52,76	769	23,30	790	23,94	3300	100,00
Średnio — błędy Mean — errors		34,67		23,00		47,22		34,96
Prawdopodobieństwo <i>a priori</i> Prior probability		33,33		33,33		33,33		
bootstrap								
A	986	41,08	734	30,58	680	28,33	2400	100,00
B	933	38,88	975	40,63	492	20,50	2400	100,00
E	955	39,79	576	24,00	869	36,21	2400	100,00
Razem — Total	2874	39,92	2285	31,74	2041	28,35	7200	100,00
Średnio — błędy Mean — errors		58,92		59,38		63,79		60,69
Prawdopodobieństwo <i>a priori</i> Prior probability		33,33		33,33		33,33		

Tabela 8

Podsumowanie klasyfikacji obserwacji do grup za pomocą wyznaczonych funkcji liniowych, wariant II
Resubstitution summary using linear discriminant function, variant II

Grupa Class	A		B		E		Razem — Total	
	liczność number of observations	%	liczność number of observations	%	liczność number of observations	%	liczność number of observations	%
dane oryginalne — original data								
A	590	65,56	211	23,44	99	11,00	900	100,00
B	235	26,11	552	61,33	113	12,56	900	100,00
E	173	19,22	245	27,22	482	53,56	900	100,00
Razem Total	998	36,96	1008	37,33	694	25,70	2700	100,00
Średnio — błędy Mean — errors		34,44		38,67		46,44		39,85
Prawdopodobieństwo <i>a priori</i> Prior probability		33,33		33,33		33,33		
bootstrap								
A	1103	45,96	263	10,96	1034	43,08	2400	100,00
B	943	39,29	498	20,75	959	39,96	2400	100,00
E	1034	43,08	210	8,75	1156	48,17	2400	100,00
Razem — Total	3080	42,78	971	13,49	3149	43,74	7200	100,00
Średnio — błędy Mean — errors		54,04		79,25		51,83		61,71
Prawdopodobieństwo <i>a priori</i> Prior probability		33,33		33,33		33,33		33,33

W badaniach należałoby skupić się na poszukiwaniu zależności pomiędzy parametrami pozyskanymi przy wykorzystaniu komputerowej analizy obrazu a poszczególnymi cechami decydującymi o jakości browarnianej ziarna jęczmienia.

WNIOSKI

1. Z uwzględnionych w badaniach modeli barwy (RGB, La^*b^* oraz HSB) jedynie parametry barwy modelu RGB charakteryzowały się rozkładem zbliżonym do rozkładu normalnego.
2. Stwierdzono, że cechami istotnie dyskryminującymi przynależność badanych odmian jęczmienia jarego do klas jakości browarnej są: MAX_FERET, VOLEQCYLINDER, WIDTH, RGB_R, RGB_G, RGB_B, LENGTH, VOLEQSPHERE, EQDIAMETER oraz MIN_FERET dla I wariantu klasyfikacji jakości oraz RGB_B, MAX_FERET, ELONGATION, VOLEQSPHERE, EQDIAMETER, RGB_G, RGB_R, VOLEQCYLINDER oraz WIDTH dla wariantu II klasyfikacji jakości ziarna jęczmienia jarego.
3. Wyznaczone liniowe funkcje rozpoznawcze poprawnie zakwalifikowały średnio dla wariantu I — 65% obserwacji oryginalnych i 39% obserwacji pochodzących z bootstrapu oraz dla wariantu II — 60% obserwacji oryginalnych i 38% obserwacji z bootstrapu.

LITERATURA

- Błażewicz J., Liszewski M. 2003. Ziarno jęczmienia nagiego odmiany 'Rastik' jako surowiec do produkcji słodów typu pilzneńskiego. *Acta Scientiarum Polonorum, Technologia Alimentaria*, 2 (1): 63 — 74.
- Drzewiecki J., Fornal Ł., Filipowicz A., Konopka S. 2000. Comparison of computer image analysis and urea soluble proteins electrophoresis for the identification of malting barley varieties. *Plant Varieties and Seeds*, 13: 159 — 171.
- Fisher R. A. 1936. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7: 179 — 188.
- Hand D. J. 1981. *Discrimination and classification*. New York, John Wiley & Sons Inc.
- Hebisz T. 2003. Multimedia i grafika komputerowa. Modele barw. Łódź, Instytut Sterowania i Systemów Informatycznych.
- Huberty C. J. 1975. Discriminant analysis. *Review of Educational Research*, 45: 543 — 598.
- Jaynes D. B., Kaspar T. C., Colvin T. S., James D. E. 2003. Cluster analysis of spatiotemporal corn yield in an Iowa field. *Agron. J.* 95: 574 — 586.
- Jennrich R. I. 1977. Stepwise discriminant analysis. In: *Statistical methods for digital computers*. Enslein K., Ralston A., Wilf H. (eds.). New York, John Wiley & Sons Inc.
- Khattre R., Naik D. N. 2000. *Multivariate data reduction and discrimination with SAS software*. SAS Publishing INC., John Wiley & Sons Inc., New York, USA.
- Klecka W. R. 1980. *Discriminant analysis*. Sage University Papers Series on Quantitative Applications in the Social Sciences: 07—019, Beverly Hills, CA: Sage Publications.
- Kozirok W. 2005. Badania nad zastosowaniem cyfrowej analizy obrazu w ocenie jakości ziarna jęczmienia browarnego. Praca Doktorska. Uniwersytet Warmińsko-Mazurski w Olsztynie.
- Krzyśko M. 1990. *Analiza dyskryminacyjna*. Wydawnictwa Naukowo-Techniczne, Warszawa.
- Lewis M. J., Young T. W. 2001. *Piwowarstwo*. Warszawa, PWN.
- Lista Odmian Roślin Rolniczych 2001. COBORU, Słupia Wielka.

- Majewska K. M. 2004. Badania wyróżników jakości technologicznej ziarna pszenicy w aspekcie jego cech geometrycznych. Rozprawy i monografie. Wydawnictwo UWM, Olsztyn.
- Majumdar S., Jayas D. S. 2000 a. Classification of cereal grains using machine vision: I. Morphology models. ASAE. vol. 43(6): 1669 — 1675.
- Majumdar S., Jayas D. S. 2000 b. Classification of cereal grains using machine vision: II. Color models. ASAE. vol. 43(6): 1677 — 1680.
- Majumdar S., Jayas D. S. 2000 c. Classification of cereal grains using machine vision: IV. Combined morphology, color and texture models. ASAE. vol. 43(6): 1689 — 1694.
- McElroy J. S., Walter R. H., Santen van E. 2002. Patterns of variation in *Poa annua* populations as revealed by canonical discriminant analysis of life history trials. Crop Sci., 42: 513 — 517.
- Mielicki J. 1997. Zarys wiadomości o barwie. Łódź, Fundacja rozwoju polskiej kolorystyki.
- Molina-Cano J. L. 1987. The EBC barley and malt committee index for the evaluation of malting quality in barley and its use in breeding. Plant Breeding, 98: 249 — 256.
- O'Rourke N., Hatcher L., Stepanski E. J. 2005. A step-by step approach to using SAS for univariate and multivariate statistics. 2nd Edition. SAS Publishing Inc., Jon Wiley & Sons Inc., New York, USA.
- Osowski S. 1996. Sieci neuronowe w ujęciu algorytmicznym. Wydawnictwa Naukowo-Techniczne, Warszawa.
- Osowski S. 2002. Sieci neuronowe typu SVM w zastosowaniu do klasyfikacji wzorców. Przegląd Elektrotechniczny, 78(2): 29 — 36.
- Pastuszak W. 2000. Barwa w grafice komputerowej. Warszawa, PWN.
- Pillai K. C. S. 1985. Multivariate analysis of variance (MANOVA). ESS, 6: 20 — 29.
- Rao C. R. 1973. Linear statistical inference and its applications. New York, John Wiley & Sons Inc.
- SAS Institute, Inc., 2004 a. BASE SAS 9.1.3 Procedures Guide. SAS Publishing, SAS Institute Inc., Cary, NC, USA.
- SAS Institute, Inc., 2004 b. SAS 9.1 Companion for Windows. SAS Publishing, SAS Institute Inc., Cary, NC, USA.
- SAS Institute, Inc., 2004 c. SAS/STAT 9.1 User's Guide. SAS Publishing, SAS Institute Inc., Cary, NC, USA.
- Shapiro S. S., Wilk M. B. 1965. An analysis of variance test of normality (complete samples). Biometrika, 52: 591 — 611.
- Trajer J., Sałat R. 2005. Sztuczne sieci neuronowe w klasyfikacji warzyw i identyfikacji ich cech. W: Trajer J., Jaros M. (red.). Zastosowanie metod sztucznej inteligencji do oceny zmian jakości wybranych warzyw w procesach ich suszenia i przechowywania. Wydawn. SGGW, Warszawa: 37 — 54.
- Vaylay R., Santen van E. 2002. Application of canonical discriminant analysis for the assessment of genetic variation in tall fescue. Crop Sci. 42: 534 — 539.
- Weber R., Zalewski D. 2004. Plonowanie odmian pszenicy ozimej w zróżnicowanych środowiskach. Biuletyn IHAR, 233: 17 — 28.
- Zieliński K. W., Strzelecki M. 2002. Komputerowa analiza obrazu biomedycznego: wstęp do morfometrii i patologii ilościowej. Warszawa, PWN.
- Strony internetowe:
 Projekt Eurequa; weryfikacja: 2005 — <http://www.eurequa.pl/>
 Teoria postrzegania barw; weryfikacja: 2005 — <http://semmix.pl/color/models/mo112.htm>.