

Object classification using X-ray images

Klasyfikacja obiektów na podstawie ich zdjęć rentgenowskich

Piotr Nowosad* , Małgorzata Charytanowicz

Department of Computer Science, Lublin University of Technology, Nadbystrzycka 36B, 20-618 Lublin, Poland

Abstract

The main aim of the presented research was to assess the possibility of utilizing geometric features in object classification. Studies were conducted using X-ray images of kernels belonging to three different wheat varieties: Kama, Canadian and Rosa. As a part of the work, image processing methods were used to determine the main geometric grain parameters, including the kernel area, kernel perimeter, kernel length and kernel width. The results indicate significant differences between wheat varieties, and demonstrates the importance of their size and shape parameters in the classification process. The percentage of correctness of classification was about 92% when the k-Means algorithm was used. A classification rate of 93% was obtain using the K-Nearest Neighbour and Support Vector Machines. Herein, the Rosa variety was better recognized, whilst the Canadian and Kama varieties were less successfully differentiated.

Keywords: object classification; geometric features; image processing; X-ray imaging

Streszczenie

Głównym celem artykułu było zbadanie możliwości wykorzystania cech geometrycznych obiektów w procesie ich klasyfikacji. Materiał badawczy stanowiły zdjęcia rentgenowskie ziaren trzech odmian pszenicy: kama, kanadyjskiej i rosa. W ramach pracy opracowano metody pozwalające na wyznaczenie cech geometrycznych obiektów znajdujących się na obrazach cyfrowych, takich jak długość, szerokość, średnica, pole i obwód. Otrzymane wyniki wykazały istotne różnice pomiędzy parametrami charakteryzującymi kształt i wielkości poszczególnych odmian pszenicy i możliwość ich zastosowania w procesie klasyfikacji. Procent poprawnie zaklasyfikowanych ziaren za pomocą algorytmu k-średnich wynosił 92%. Nieco lepsze wyniki, rzędu 93%, uzyskano za pomocą metod K-najbliższych sąsiadów i wektorów wspierających. Najlepiej rozróżnialną odmianą okazała się rosa w porównaniu do odmian kanadyjskiej i kama.

Słowa kluczowe: klasyfikacja obiektów; cechy geometryczne; przetwarzanie obrazów; obrazowanie rentgenowskie

*Corresponding author

Email address: piotr.nowosad@pollub.edu.pl (P. Nowosad)

©Published under Creative Common License (CC BY-SA v4.0)

1. Wstęp

Dynamiczny rozwój technik informacyjnych pozwolił na opracowanie nowych metod gromadzenia, przetwarzania i analizy informacji. Możliwe stała się praktyczne zastosowanie wielu algorytmów, które dotychczas ze względu na ograniczenia sprzętowe i czasowe nie mogły być stosowane. Nowoczesne technologie pozwoliły również na wykorzystanie obrazu jako źródła danych. Metody przetwarzania i analizy obrazów cyfrowych stanowią obecnie jedną z ważniejszych gałęzi informatyki [1, 2].

Procesem łączącym wstępne przetwarzanie obrazu z jego analizą jest segmentacja, polegająca na wyodrębnianiu z obrazu jednorodnych obszarów według przyjętego z góry kryterium. Uważana jest ona za jedną z najtrudniejszych i najważniejszych operacji, umożliwiającą identyfikację obiektów oraz ich opis ilościowy i jakościowy. Z drugiej strony połączenie statystycznych metod analizy danych z technikami komputerowymi pozwoliło na rozwój działu statystyki zwanego wielowymiarową analizą klasyfikacyjną i analizą skupień [3, 4]. Systemy uczące się są obecnie podstawą inteligentnej analizy danych [5]. Wyróżniamy systemy uczące się pod nadzorem, wśród których warto wspomnieć o różnego typu klasyfikatorach, oraz systemy uczące się bez nadzoru, których zadaniem jest wykry-

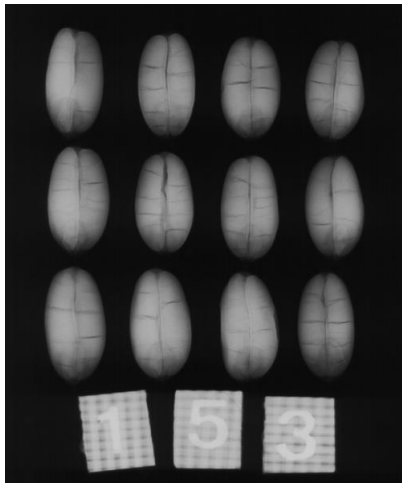
wanie wewnętrznej struktury zbioru danych. Przykładem jest tu analiza skupień [6, 7]. Rezonans magnetyczny, tomografia komputerowa, skanery i systemy wizyjne umożliwiły uzyskanie obrazów wysokiej jakości, stanowiących cenne źródło informacji w naukach technicznych, medycznych i przyrodniczych [8-10].

Celem niniejszej pracy jest zbadanie możliwości wykorzystania cech geometrycznych obiektów do ich klasyfikacji. Materiał badawczy stanowią zdjęcia rentgenowskie ziaren pszenicy. Zastosowanie metod przetwarzania obrazów cyfrowych do tego typu zagadnień jest często stosowane w praktyce. W pracy [11] przedstawiono metodę oceny zmian właściwości ziarna w zależności od warunków zewnętrznych [11]. Za pomocą progowania i filtracji z obrazów ziaren uzyskano obrazy binarne, które wykorzystano do pozyskania danych w celu przeprowadzenia dalszych analiz. Zagadnienie wyznaczenia cech geometrycznych odmian ziarna pszenicy, na podstawie których możliwa jest ich dyskryminacja, omówione zostało w artykułach [12-14]. Do identyfikacji parametrów geometrycznych wykorzystano obrazy cyfrowe ziaren. Następnie opracowano metodykę opartą na przekształcaniach filtracyjnych i metodach segmentacji, umożliwiających automatyczne wykonywanie pomiarów. Prezentowane rozwiązania wymagają zwykle posiadania specjalistycznego oprogramowania lub przygotowania własnego w oparciu

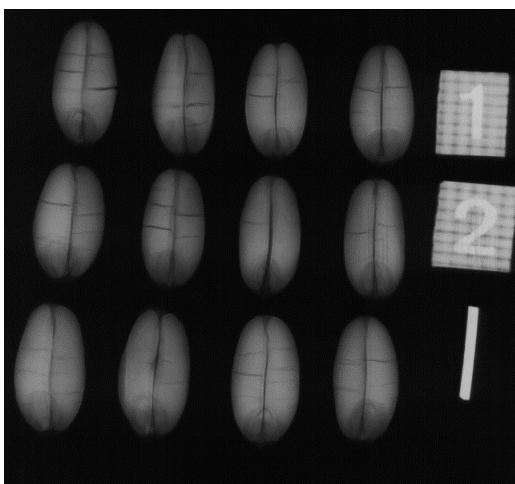
o istniejące biblioteki zawierające funkcje do wykonywania podstawowych operacji na obrazach cyfrowych [15-18]. Z tego względu w ramach badań opracowany został autorski program do wyznaczania cech geometrycznych obiektów znajdujących się na obrazach cyfrowych.

2. Materiał badawczy

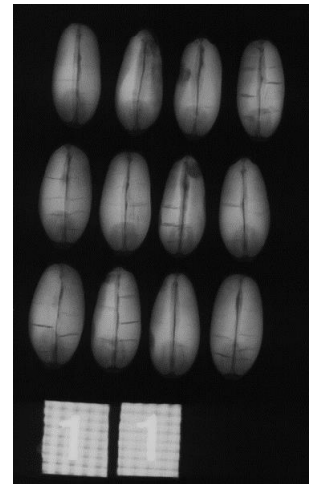
Materiał badań stanowiły obrazy rentgenowskie ziaren trzech odmian pszenicy: kama, kanadyjskiej i rosa, wybranych losowo ze zbioru kombajnowego. Do uzyskania zdjęć rentgenowskich został użyty aparat rentgenowski ELEKTRONIKA Model 25. Każdy obraz zawierał 12 ziaren. W celu wykonania pomiarów cech geometrycznych obrazy rentgenowskie zostały zeskanowane skanerem Epson Perfection V700 o rozdzielczości 600 dpi i zapisane w plikach graficznych o 24-bitowej głębi bitowej. Ich wielkość nie przekraczała 300 KB. Przykładowe obrazy ziaren poszczególnych odmian pszenicy, w skali 2:5 w stosunku do obrazów oryginalnych, zostały przedstawione na Rysunkach 1 – 3.



Rysunek 1: Materiał badawczy – odmiana kama



Rysunek 2: Materiał badawczy – odmiana kanadyjska

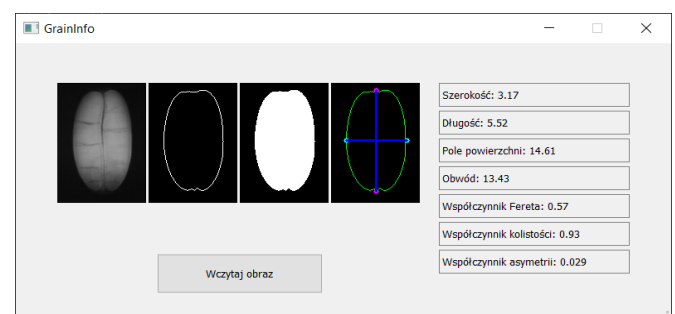


Rysunek 3: Materiał badawczy – odmiana rosa

W celu dokonania pomiarów każde ziarno zapisane zostało jako odrębny plik graficzny. Wyodrębniono 144 obrazy ziaren, po 48 dla każdej odmiany. Do wyznaczenia rzeczywistych wymiarów ziaren użyty został wzorzec o długości 5 mm, odpowiadający 114 pikselom na obrazie cyfrowym. Na tej podstawie określono skalę, która wynosiła 0,04386.

3. Wyznaczania cech geometrycznych obiektów

Na potrzeby prowadzonych badań opracowany został autorski program GrainInfo, napisany w środowisku Microsoft Visual Studio. Analiza obrazu została wykonywana z wykorzystaniem biblioteki OpenCV, natomiast interfejs graficzny użytkownika i operacje matematyczne są wykonywane za pomocą biblioteki Qt [19-22]. Program umożliwia wizualizację struktury wewnętrznej ziaren i wyznaczanie podstawowych parametrów geometrycznych obiektów. Główne okno programu przedstawione zostało na Rysunku 4.



Rysunek 4: Główne okno programu GrainInfo

Program GrainInfo umożliwia automatyczny pomiar podstawowych parametrów charakteryzujących kształt i wielkość ziaren. Jednostką pomiaru są milimetry. Wykonane pomiary obejmują:

- pole ziarna, traktowane jako pole rzutu ziarna,
- obwód ziarna, traktowany jako obwód rzutu ziarna,
- szerokość ziarna,
- długość ziarna,

- współczynnik Fereta,
- współczynnik kolistości,
- współczynnik asymetrii.

Współczynnik Fereta jest to iloraz maksymalnej średnicy obiektu w poziomie do maksymalnej średnicy w pionie. Współczynnik kolistości obliczany jest według zależności:

$$V_c = \frac{4\pi A}{D^2} \quad (1)$$

gdzie A oznacza pole powierzchni, natomiast D obwód obiektu. Współczynnik asymetrii zdefiniowany został jako iloraz wartości bezwzględnej różnicy pola lewej i prawej części rzutu ziarna do pola rzutu całego ziarna.

4. Analiza danych

Podstawowe statystyki opisowe badanych parametrów przedstawia Tabela 1. Jednostką pomiaru są milimetry.

Tabela 1: Podstawowe statystyki opisowe badanych parametrów

Odmiana kama ($n = 48$)					
Zmienna	Średnia	Mediana	Min	Max	Odch.std
szerokość	3,145	3,180	2,807	3,465	0,168
długość	5,596	5,614	5,044	6,228	0,236
pole	14,189	14,339	11,321	16,784	1,251
obwód	13,271	13,355	11,754	14,518	0,574
wsp. Fereta	0,562	0,560	0,520	0,610	0,023
wsp. kolist.	0,928	0,930	0,870	0,970	0,024
wsp. asym.	0,027	0,022	0,002	0,080	0,022
Odmiana kanadyjska ($n = 48$)					
Zmienna	Średnia	Mediana	Min	Max	Odch.std
szerokość	2,727	2,719	2,500	3,026	0,118
długość	5,312	5,307	4,825	5,789	0,191
pole	11,434	11,358	9,649	13,560	0,738
obwód	12,368	12,281	11,228	13,684	0,513
wsp. Fereta	0,514	0,510	0,460	0,560	0,024
wsp. kolist.	0,851	0,855	0,680	0,920	0,042
wsp. asym.	0,036	0,035	0,001	0,105	0,024
Odmiana rosa ($n = 48$)					
Zmienna	Średnia	Mediana	Min	Max	Odch.std
szerokość	3,581	3,575	2,982	3,991	0,236
długość	6,310	6,294	5,658	6,974	0,300
pole	18,173	18,242	13,294	21,993	1,890
obwód	15,916	15,987	13,246	17,807	0,882
wsp. Fereta	0,568	0,560	0,510	0,630	0,032
wsp. kolist.	0,833	0,850	0,600	0,950	0,080
wsp. asym.	0,020	0,018	0,001	0,053	0,013

Największymi wymiarami charakteryzowała się odmiana rosa, natomiast dla odmiany kanadyjskiej otrzymano najmniejsze wartości szerokości, długości, pola i obwodu ziaren. Współczynniki Fereta, kolistości i asymetrii były mniej zróżnicowane. Współczynnik Fereta był zbliżony dla odmian kama i rosa, dla odmiany kanadyjskiej uzyskał wartość najmniejszą. Największy współczynnik kolistości uzyskała odmiana kama, dla pozostałych dwóch odmian współczynnik ten był zbliżony.

Wnioskowanie statystyczne przeprowadzono na poziomie istotności $\alpha = 0,05$. Do weryfikację hipotezy

o zgodności rozkładu z rozkładem normalnym wykorzystany został test istotności Shapiro-Wilka. W każdej grupie rozkład badanych parametrów geometrycznych nie różnił się istotnie od rozkładu normalnego (otrzymano $p > 0,1$). Z tego względu do porównania wartości średnich przedstawionych parametrów zastosowano parametryczne testy istotności. Następnie zbadano możliwość wykorzystania analizy skupień do wydzielenia jednorodnych grup ziaren pod względem ich parametrów geometrycznych. Jakość klasyfikacji zbadano naiwną metodą Bayesa, metodą K -najbliższych sąsiadów oraz metodą wektorów wspierających.

4.1. Analiza wariancji

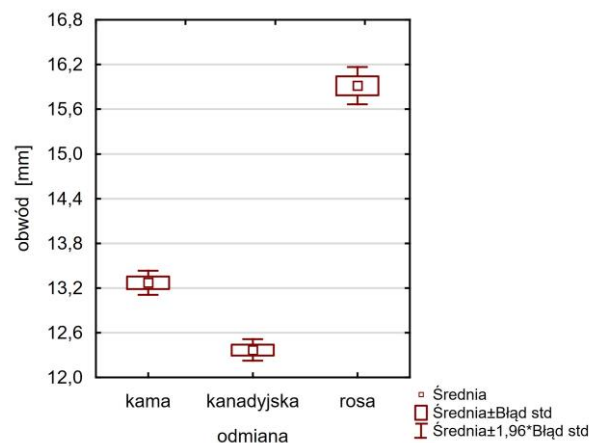
Hipotezę o równości wartości średnich badanych parametrów geometrycznych trzech odmian pszenicy zweryfikowano za pomocą metody analizy wariancji. Jednorodność wariancji sprawdzono testem Levene'a. Ponieważ wariancje badanych parametrów w badanych grupach różniły się istotnie ($p < 0,001$) zastosowano test F Welcha uwzględniający nierówne wariancje. W przypadku odrzucenia hipotezy o równości średnich danego parametru do stwierdzenia istotnych różnic pomiędzy grupami wykorzystano test Tukeya.

Wartość testu F Welcha analizy wariancji wraz z wartością istotności p przedstawia Tabela 2.

Tabela 2: Wyniki analizy wariancji

Zmienna	F (2,141)	p Welcha
szerokość	287,316	$p < 0,00001$
długość	187,926	$p < 0,00001$
pole	303,144	$p < 0,00001$
obwód	287,882	$p < 0,00001$
współczynnik Fereta	64,102	$p < 0,00001$
współczynnik kolistości	79,998	$p < 0,00001$
współczynnik asymetrii	7,201	$p < 0,0006$

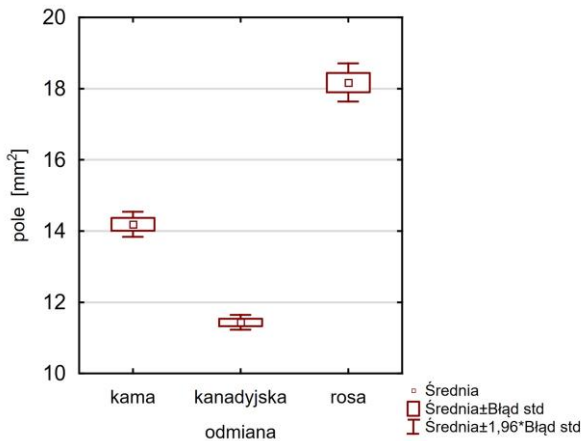
Na podstawie przeprowadzonego testu F Welcha dla każdego parametru stwierdzono statystycznie istotne różnice pomiędzy wartościami średnimi w badanych grupach odmian pszenicy: kama, kanadyjska i rosa. Wyniki testu Tukeya potwierdziły istotne statystycznie różnice pomiędzy poszczególnymi grupami.



Rysunek 5: Obwód ziaren dla trzech odmian pszenicy: istotne różnice pomiędzy odmianami kama – kanadyjska ($p < 0,0001$), kama – rosa ($p < 0,0001$), kanadyjska – rosa ($p < 0,0001$)

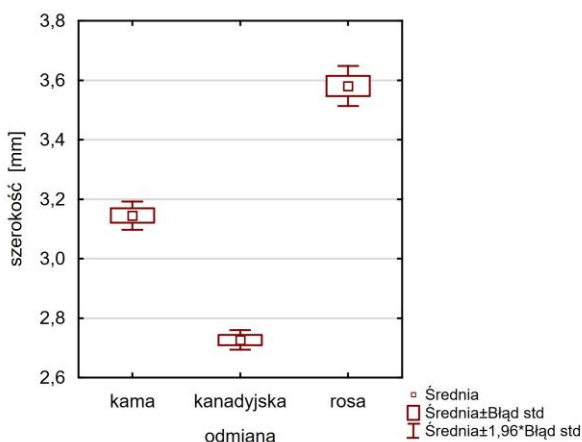
Skategoryzowane wykresy typu ramka-wąsy ilustrujące różnice pomiędzy wartościami średnimi badanych parametrów przedstawione są na Rysunkach 5 – 11.

Wszystkie trzy odmiany różnią się istotnie pod względem obwodu ziaren. Najmniejsze wartości charakteryzują odmianę kanadyjską, największe – odmianę rosa. Wartości średnie i odchylenia standardowe odmian kama i kanadyjskiej wynoszą odpowiednio $13,271 \pm 0,574$ mm i $12,368 \pm 0,513$ mm, natomiast dla odmiany rosa uzyskano wartości $15,916 \pm 0,882$ mm. Odmiana rosa bardziej różni się wymiarami od pozostałych dwóch odmian.



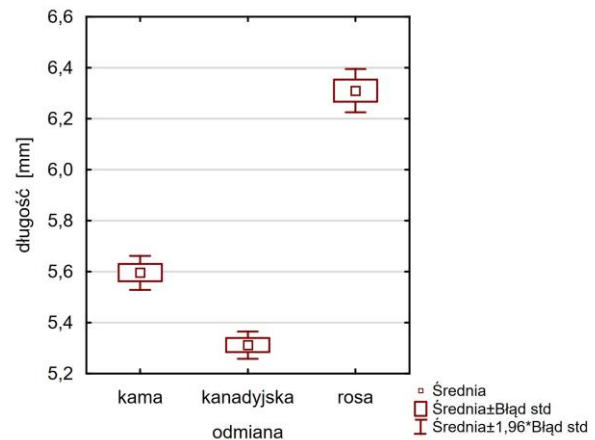
Rysunek 6: Pole ziaren dla trzech odmian pszenicy: istotne różnice pomiędzy odmianami kama – kanadyjska ($p < 0,0001$), kama – rosa ($p < 0,0001$), kanadyjska – rosa ($p < 0,0001$)

Wszystkie trzy odmiany odróżnia istotnie pole ziaren. Najmniejsze wartości charakteryzują odmianę kanadyjską, największe – odmianę rosa. Podobnie jak w przypadku obwodu, odmiana rosa różni się w większym stopniu. Wartości średnie i odchylenia standardowe dla odmian kama i kanadyjskiej wynoszą $14,189 \pm 1,251$ mm i $11,434 \pm 0,738$ mm, natomiast dla odmiany rosa uzyskano wynik $18,173 \pm 1,890$ mm.



Rysunek 7: Szerokość ziaren dla trzech odmian pszenicy: istotne różnice pomiędzy odmianami kama – kanadyjska ($p < 0,0001$), kama – rosa ($p < 0,0001$), kanadyjska – rosa ($p < 0,0001$)

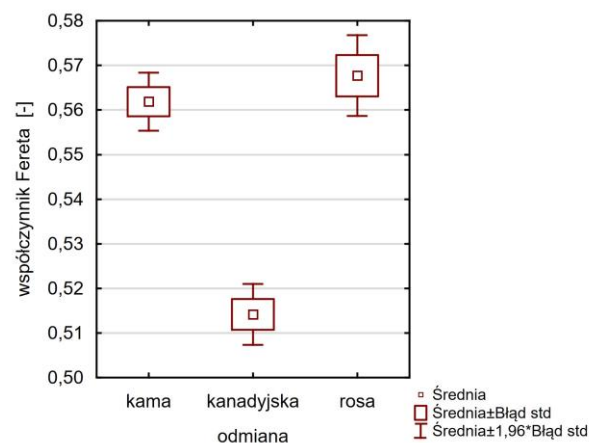
Wszystkie trzy odmiany różnią się istotnie pod względem szerokości ziaren. Najmniejsze wartości charakteryzują odmianę kanadyjską ($2,727 \pm 0,118$), nieco większe odmianę kama ($3,145 \pm 0,168$), a największe – odmianę rosa ($3,581 \pm 0,236$).



Rysunek 8: Długość ziaren dla trzech odmian pszenicy: istotne różnice pomiędzy odmianami kama – kanadyjska ($p < 0,0001$), kama – rosa ($p < 0,0001$), kanadyjska – rosa ($p < 0,0001$)

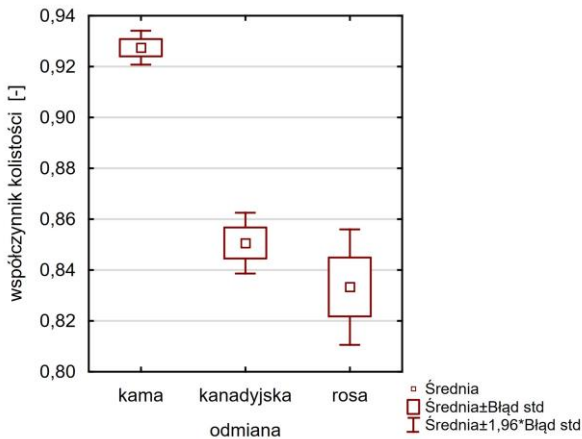
Wszystkie trzy odmiany różnią się istotnie pod względem długości ziaren. Najmniejsze wartości charakteryzują odmianę kanadyjską ($5,312 \pm 0,191$), nieco większe odmianę kama ($5,312 \pm 0,191$), a największe – odmianę rosa ($6,310 \pm 0,300$).

Przeprowadzone analizy wykazały, że ziarna badanych odmian pszenicy: kama, kanadyjskiej i rosa różnią się istotnie pod względem wielkości. Stwierdzono istotne różnice pomiędzy wartościami średnimi obwodu, pola, szerokości i długości ziaren. Największymi wymiarami charakteryzuje się odmiana rosa, najmniejszymi – odmiana kanadyjska.



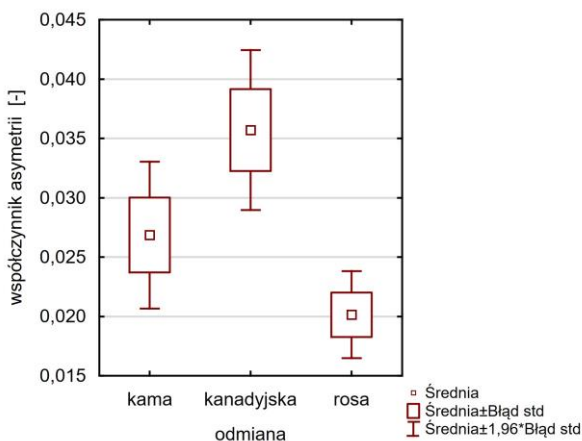
Rysunek 9: Współczynnik Fereta dla trzech odmian pszenicy: istotne różnice pomiędzy odmianami kama – kanadyjska ($p < 0,0001$), kanadyjska – rosa ($p < 0,0001$)

Współczynnik Fereta dla odmiany kanadyjskiej jest istotnie niższy od współczynników Fereta odmian kama i rosa.



Rysunek 10: Współczynnik kolistości dla trzech odmian pszenicy: istotne różnice pomiędzy odmianami kama – kanadyjska ($p < 0,0001$), kama – rosa ($p < 0,0001$)

Współczynnik kolistości dla odmiany kama jest istotnie wyższy od współczynników kolistości odmian kanadyjskiej i rosa.



Rysunek 11: Współczynnik asymetrii dla trzech odmian pszenicy: istotne różnice pomiędzy odmianami kanadyjska – rosa ($p < 0,0005$)

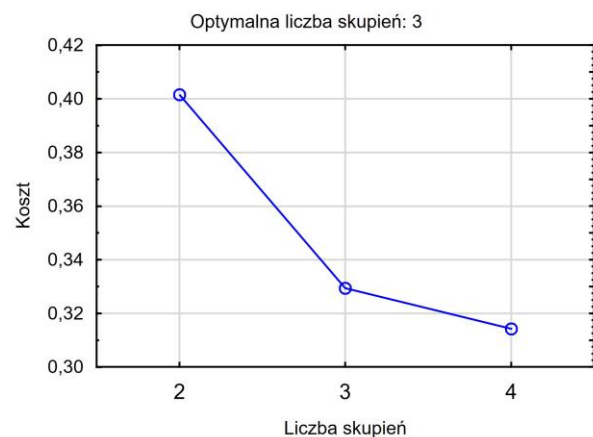
Współczynnik asymetrii dla odmiany kanadyjskiej jest istotnie wyższy od współczynnika asymetrii odmiany rosa.

Współczynnik Fereta różnicuje istotnie odmianę kanadyjską od odmian kama i rosa, współczynnik asymetrii odmianę rosa od kama i kanadyjskiej, a współczynnik kolistości odmianę kama od kanadyjskiej i rosa. Współczynnik kolistości osiąga największe wartości dla ziaren odmiany kama, co oznacza, że ich kształt w większym stopniu jest zbliżony do koła niż pozostałych dwóch odmian.

4.2. Analiza skupień

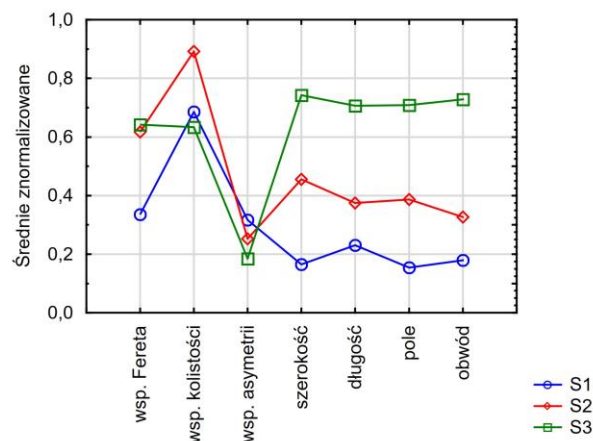
W celu zbadania, czy na podstawie podstawowych cech geometrycznych ziaren pszenicy możliwy jest ich podział na poszczególne odmiany, podjęto próbę wyodrębnienia jednorodnych grup za pomocą uogólnionej analizy skupień metodą k -średnich. Wykorzystano 144 elementy (ziarna) opisane za pomocą siedmiu zmien-

nych ilościowych (szerokość, długość, pole, obwód, współczynnik Fereta, współczynnik kolistości i współczynnik asymetrii). Jako miarę odległości wybrano odległość euklidesową, początkowe centra skupień zostały wybrane tak, by zmaksymalizować odległości skupień. Liczba skupień wyznaczana jest automatycznie, bez odgórnej ingerencji w analizę. Do jej oceny wykorzystany został sprawdzian krzyżowy, w którym liczba grup została ustalana za pomocą doboru najlepszego podziału danych na segmenty na podstawie minimalizacji odległości elementów skupienia od jego środka. Automatycznie została wybrana poprawna liczba skupień równa trzy. Wyniki sprawdzianu krzyżowego ilustruje wykres osypiska przedstawiony na Rysunku 12.



Rysunek 12: Uogólniona analiza skupień: wykres osypiska

Analiza wykresu potwierdza poprawę precyzji podziału na segmenty w przypadku zwiększenia ich liczby z dwóch do trzech. Kolejne zwiększanie liczby segmentów nie przynosi już tak znacznej poprawy. Na Rysunku 13 przedstawiono znormalizowane średnie zmiennych ilościowych dla skupień.



Rysunek 13: Wykres znormalizowanych wartości średnich parametrów geometrycznych ziaren dla trzech skupień wyznaczonych metodą k -średnich

Liczba ziaren każdego skupienia oraz ich przynależność do poszczególnych odmian pszenicy zawiera Tabela 3. Skupienie pierwsze zawiera 47 ziaren odmiany kanadyjskiej, 1 ziarno odmiany rosa i 7 ziaren odmiany kama. Skupienie drugie jest prawie jednorodne pod względem odmiany: zawiera 45 ziaren odmiany rosa i 1 ziarno odmiany kama. Skupienie trzecie zawiera 40 ziaren odmiany kama, 1 ziarno odmiany kanadyjskiej i 2 ziarna odmiany rosa. Wymiary ziaren różnicują wyznaczone skupienia. Skupienie 1 zawiera ziarna o najmniejszych wymiarach, skupienie 2 zawiera ziarna o największych wymiarach. Otrzymany podział na grupy potwierdzają również wyniki dla współczynników Fereta, asymetrii i kolistości.

Tabela 3: Liczba ziaren należących do poszczególnych skupień z podziałem na odmiany

Skupienie	Odmiana			Razem
	kanadyjska	rosa	kama	
Skupienie 1	47	1	7	55
Skupienie 2	0	45	1	46
Skupienie 3	1	2	40	43
Razem	48	48	48	144

Poprawnie zaklasyfikowano 132 ziarna, co stanowi niemal 92% ich całkowitej liczby. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 12. Odmiana rosa w większym stopniu odróżnia się od odmiany kanadyjskiej i kama.

Przeprowadzone analizy potwierdziły możliwość dokonania podziału zbioru ziaren na trzy grupy odpowiadające poszczególnym odmianom pszenicy bez odgórnej ingerencji w analizę.

4.3. Metody klasyfikacji

W celu zbadania możliwości identyfikacji ziaren na podstawie ich cech geometrycznych wykorzystano trzy klasyczne metody klasyfikacji:

- naiwną metodą Bayesa,
- metodę K -najbliższych sąsiadów
- metodę wektorów nośnych.

W pierwszym kroku badań dokonano losowego podziału zbioru danych na zbiór uczący i zbiór testowy w proporcjach:

- 90% : 10%,
- 80% : 20%,
- 70% : 30%.

W celu porównania metod klasyfikacji wykorzystano taki sam podział na zbiór uczący i testowy. Liczebności zbiorów: uczącego i testowego z podziałem na odmiany dla każdego podziału przedstawia Tabela 4.

Tabela 4: Liczebności zbiorów uczących i testowych z uwzględnieniem podziału na trzy odmiany

Podział	Liczba ziaren		Razem
	Zbiór uczący	Zbiór testowy	
90% : 10%	129 (3 x 43)	15 (3 x 5)	144 (3 x 48)
80% : 20%	114 (3 x 38)	30 (3 x 10)	144 (3 x 48)
70% : 30%	102 (3 x 34)	42 (3 x 14)	144 (3 x 48)

Każda odmiana zawierała 48 ziaren. Po dokonaniu podziału liczba ziaren każdej odmiany w zbiorze uczącym była taka sama. Odpowiednio zbiory testowe zawierały po 5, 10 i 14 ziaren każdej odmiany. W pierwszym przypadku liczebność zbioru testowego wynosiła 15, w drugim – 30, w trzecim 42.

Tabele 5-7 zawierają wyniki klasyfikacji uzyskane naiwną metodą Bayesa dla wyodrębnionych zbiorów testowych o liczebności.

Tabela 5: Wyniki klasyfikacji: naiwna metoda Bayesa, zbiór testowy: 10%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	5 (71,4%)	2 (28,6%)	7
kanadyjska	3 (100%)	0 (0%)	3
rosa	5 (100%)	0 (0%)	5
Razem	13 (87%)	2 (13%)	15

W wyniku klasyfikacji naiwną metodą Bayesa wykonaną dla zbioru testowego stanowiącego 10% wszystkich ziaren 2 ziarna odmiany kanadyjskiej zostały błędnie zaklasyfikowane jako ziarna odmiany kama. Pozostałych 13 ziaren zaklasyfikowano poprawnie, co stanowi 87% liczby ziaren zbioru testowego.

Tabela 6: Wyniki klasyfikacji: naiwna metoda Bayesa, zbiór testowy: 20%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	10 (83,3%)	2 (16,7%)	12
kanadyjska	8 (100%)	0 (0%)	8
rosa	10 (100%)	0 (0%)	10
Razem	28 (93%)	2 (7%)	30

Dla zbioru testowego stanowiącego 20% wszystkich ziaren 2 ziarna odmiany kanadyjskiej zostały błędnie zaklasyfikowane jako ziarna odmiany kama. Poprawnie zaklasyfikowano 28 ziaren, co stanowi 93% liczby ziaren zbioru testowego. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 2, jest to 7% liczby ziaren.

Tabela 7: Wyniki klasyfikacji: naiwna metoda Bayesa, zbiór testowy: 30%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	11 (84,6%)	2 (15,4%)	13
kanadyjska	12 (100%)	0 (0%)	12
rosa	14 (82,4%)	3 (17,6%)	17
Razem	37 (88%)	5 (12%)	42

Dla zbioru testowego stanowiącego 30% wszystkich ziaren 2 ziarna odmiany kanadyjskiej zostały błędnie zaklasyfikowane jako ziarna odmiany kama oraz 3 ziarna odmiany kama zostały błędnie zaklasyfikowane jako ziarna odmiany rosa. Poprawnie zaklasyfikowano 37 ziaren, co stanowi 88% ziaren zbioru testowego zawierającego 42 ziarna. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 5.

Tabele 8-10 zawierają wyniki klasyfikacji uzyskane metodą K -najbliższych sąsiadów z parametrem $K = 1$

dla zbioru testowego o liczebności 15, 30 i 42 ziarna odpowiednio.

Tabela 8: Wyniki klasyfikacji: metoda K -najbliższych sąsiadów, zbiór testowy: 10%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	5 (71,4%)	2 (28,6%)	7
kanadyjska	3 (100%)	0 (0%)	3
rosa	5 (100%)	0 (0%)	5
Razem	13 (87%)	2 (13%)	15

W wyniku klasyfikacji metodą K -najbliższych sąsiadów wykonaną dla zbioru testowego stanowiącego 10% wszystkich ziaren 2 ziarna odmiany kanadyjskiej zostały błędnie zaklasyfikowane jako ziarna odmiany kama. Poprawnie zaklasyfikowano 13 ziaren, co stanowi 87% całkowitej liczby ziaren zbioru testowego. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 2.

Tabela 9: Wyniki klasyfikacji metoda K -najbliższych sąsiadów, zbiór testowy: 20%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	10 (90,9%)	1 (9,1%)	11
kanadyjska	9 (100%)	0 (0%)	9
rosa	10 (100%)	0 (0%)	10
Razem	29 (93%)	1 (7%)	30

Dla drugiego zbioru testowego stanowiącego 20% wszystkich ziaren 1 ziarno odmiany kanadyjskiej zostało błędnie zaklasyfikowane jako ziarno odmiany kama. Poprawnie zaklasyfikowano 29 ziaren, co stanowi 93% ich liczby dla zbioru testowego. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 1 co stanowi 7% liczby ziaren.

Tabela 10: Wyniki klasyfikacji: K -metoda najbliższych sąsiadów, zbiór testowy: 30%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	13 (92,9%)	1 (7,1%)	14
kanadyjska	13 (100%)	0 (0%)	13
rosa	14 (82,4%)	1 (17,6%)	15
Razem	40 (95,2%)	2 (4,8%)	42

Dla trzeciego zbioru testowego stanowiącego 30% wszystkich ziaren 1 ziarno odmiany kanadyjskiej zostało błędnie zaklasyfikowane jako ziarno odmiany kama oraz 1 ziarno odmiany kama zostało błędnie zaklasyfikowane jako ziarno odmiany rosa. Poprawnie zaklasyfikowano 40 ziaren, co stanowi 95,2% liczby ziaren zbioru testowego. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 2 co stanowi 4,8% liczby ziaren.

Tabele 11-13 zawierają wyniki klasyfikacji uzyskane metodą wektorów wspierających dla rozpatrywanych zbiorów testowych o liczebności 15, 30 i 42 ziarna odpowiednio.

W wyniku klasyfikacji metoda wektorów wspierających, wykonaną dla zbioru testowego stanowiącego 10% wszystkich ziaren 2 ziarna odmiany kanadyjskiej

zostały błędnie zaklasyfikowane jako ziarna odmiany kama.

Tabela 11: Wyniki klasyfikacji: metoda wektorów wspierających, zbiór testowy: 10%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	5 (71,4%)	2 (28,6%)	7
kanadyjska	3 (100%)	0 (0%)	3
rosa	5 (100%)	0 (0%)	5
Razem	13 (87%)	2 (13%)	15

Poprawnie zaklasyfikowano 13 ziaren, co stanowi 87% ich całkowitej liczby. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 2. Wyniki były identyczne jak w przypadku metody K -najbliższych sąsiadów.

Tabela 12: Wyniki klasyfikacji metoda wektorów wspierających, zbiór testowy: 20%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	10 (83,3%)	2 (16,7%)	12
kanadyjska	8 (100%)	0 (0%)	8
rosa	10 (100%)	0 (0%)	10
Razem	28 (93%)	2 (7%)	30

Dla zbioru testowego stanowiącego 20% wszystkich ziaren 2 ziarna odmiany kanadyjskiej zostały błędnie zaklasyfikowane jako ziarna odmiany kama. Poprawnie zaklasyfikowano 28 ziaren. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 2. Wyniki były porównywalne z wynikami otrzymanymi naiwną metodą Bayesa i gorsze niż w przypadku metody K -najbliższych sąsiadów.

Tabela 13: Wyniki klasyfikacji: metoda wektorów wspierających, zbiór testowy: 30%

Odmiana	Liczba ziaren		Razem
	Trafne	Błędne	
kama	13 (86,7%)	2 (13,3%)	15
kanadyjska	12 (100%)	0 (0%)	12
rosa	14 (93,3%)	1 (6,7%)	15
Razem	39 (93%)	3 (7%)	42

Dla zbioru testowego stanowiącego 30% wszystkich ziaren 2 ziarna odmiany kanadyjskiej zostały błędnie zaklasyfikowane jako ziarna odmiany kama. Ponadto jedno ziarno odmiany kama zostało błędnie zaklasyfikowane jako ziarno odmiany rosa. Poprawnie zaklasyfikowano 39 ziaren, co stanowi 93% liczby ziaren zbioru testowego. Liczba niepoprawnie zaklasyfikowanych ziaren wynosi 2 co stanowi 7%. Wyniki były gorsze niż w przypadku metody K -najbliższych sąsiadów.

Na podstawie badań wykonanych naiwną metodą Bayesa oraz metodami K -najbliższych sąsiadów i wektorów wspierających można stwierdzić, iż wśród metod klasyfikacji najgorsze wyniki uzyskano dla naiwnej metody Bayesa. Metoda K -najbliższych sąsiadów metoda okazała się najlepszym klasyfikatorem. Dla zbioru testowego stanowiącego 10% całkowitej liczby obiektów wyniki były porównywalne, dla zbioru stanowiącego 20% ogólnej liczby ziaren najlepszą okazała się

metoda K -najbliższych sąsiadów. W trzecim teście, przeprowadzonym na zbiorze testowym o liczności 42 ziaren, stanowiącym 30% całkowitej liczby obiektów metodą Bayesa błędnie zaklasyfikowanych zostało 5 ziaren, metodą wektorów wspierających 3 ziarna, natomiast metodą K -najbliższych sąsiadów błędnie zaklasyfikowano tylko 2 ziarna.

5. Podsumowanie i wnioski

W niniejszej pracy przeprowadzone badania dotyczące wykorzystania podstawowych cech geometrycznych do klasyfikacji obiektów. Zastosowana w tym celu technika rentgenowska okazała się niedrogą i nieinwazyjną metodą obrazowania struktury wewnętrznej materiału badawczego. Badania przeprowadzone na podstawie obrazów ziaren pszenicy potwierdziły możliwość rozróżnienia odmian pszenicy na podstawie cech geometrycznych ziaren. W wyniku przeprowadzenia analizy skupień metodą k -średnich poprawnie zaklasyfikowanych zostało 132 ziarna, co stanowi 92% ich całkowitej liczby. W badaniach dotyczących podstawowych metod klasyfikacji procent poprawnie zaklasyfikowanych ziaren wynosił 93%. Najlepsze wyniki uzyskano metodą K -najbliższych sąsiadów.

Literatura

- [1] R. C. Gonzalez, R. E. Woods, Digital Image Processing, Prentice-Hall Inc., New Jersey, 2002.
- [2] R. Tadeusiewicz, Komputerowa analiza i przetwarzanie obrazów, Wydawnictwo Fundacji Postępu Telekomunikacji, Kraków 1997.
- [3] B. Mirkin, Clustering: A Data Recovery Approach, Chapman and Hall/CRC, 2012.
- [4] D. F. Morrison, Multivariate Statistical Methods, Brooks/Cole Thomson Learning, Belmont, California, 2005.
- [5] M. Romaniuk, O. Hryniewicz, Interval based, nonparametric approach for resampling of fuzzy numbers. 2019, Soft Computing, 23 (14), 5883–5903.
- [6] J. Koronacki, J. Ćwik, Statystyczne systemy uczące się, WNT 2008.
- [7] M. Krzyśko, W. Wołyński, T. Górecki, M. Skorzybut, Systemy uczące się. WNT, Warszawa, 2008.
- [8] H. Czachor, M. Charytanowicz, S. Gonet, J. Niewczas, G. Józefaciuk, L. Lichner, Impact of long term mineral and organic fertilization on water stability, wettability and porosity of aggregates of two silt loamy soils. 2015, European Journal of Soil Science, 66 (3), 577–588.
- [9] M. Klatka, E. Grywalska, M. Partyka, M. Charytanowicz, E. Kiszczak-Bochyńska, J. Roliński: Th17 and Treg cells in adolescents with Graves' disease. Impact of treatment with methimazole on these cell subsets, 2014, Autoimmunity, 47 (3), 201–211.
- [10] P. Kulczycki, P. A. Kowalski, Bayes Classification for Nonstationary Patterns, International Journal of Computational Methods, 2015, 12, ID 1550008.
- [11] T. Guz, Z. Kobus, E. Kusińska, R. Nadulski, Morphometric features of rye caryopses stored in a silo, Inżynieria Rolnicza Agricultural Engineering, 2012, 1 (4), 71–79.
- [12] P. Zapotoczny, Discrimination of wheat grain varieties using image analysis and neural networks. Part I. Single kernel texture. Journal of Cereal Science, 2012, 54, 60–68.
- [13] M. Charytanowicz, J. Niewczas, P. Kulczycki, P. A. Kowalski, S. Łukasik, Discrimination of Wheat Grain Varieties Using X-ray Images. Information Technologies in Biomedicine, E. Pietka, P. Badura, J. Kawa, W. Więclawek (red.), Advances in Intelligent Systems and Soft Computing, Springer, 2016, 471, 39–50.
- [14] M. Charytanowicz, P. Kulczycki, P. A. Kowalski, S. Łukasik, R. Czabak-Garbacz, An Evaluation of Utilizing Geometric Features for Wheat Grain Classification using X-ray Images. Computers and Electronics in Agriculture, 2018, 144, 260–268.
- [15] J. Niewczas, A. Strumiłło, Szczypiński, P. Makowski, W. Woźniak, Computer system for analysis of x-ray image of wheat grains, International Agrophysics 1999.
- [16] G. D. Jasmin, Shape based Object Classification Rusing Knowledge Vector Code International Journal of Innovative Research in Computer and Communication Engineering, 2017, 5 (7), 13440.
- [17] D. Zhang, G. Lu, Review of shape representation and description techniques. Pattern Recognition, 2004, 37, 1–19.
- [18] M. Zhu, T. J. Hastie, Feature Extraction for Nonparametric Discriminant Analysis. Journal of Computational and Graphical Statistics, 2003, 12(1), 101–120.
- [19] Visual Studio 2017 – Now Ready for Your Windows Application Development Needs, <https://blogs.windows.com/windowsdeveloper/2017/03/07/visual-studio-2017-now-ready-windows-application-development-needs/>, [22.04.2020].
- [20] Qt Software, <http://doc.qt.io/>, [01.04.2020].
- [21] C++ Programming Language, <https://www.techopedia.com/definition/26184/c-programming-language>, [22.04.2020].
- [22] About OpenCV, <https://opencv.org/about/>, [22.04.2020].