

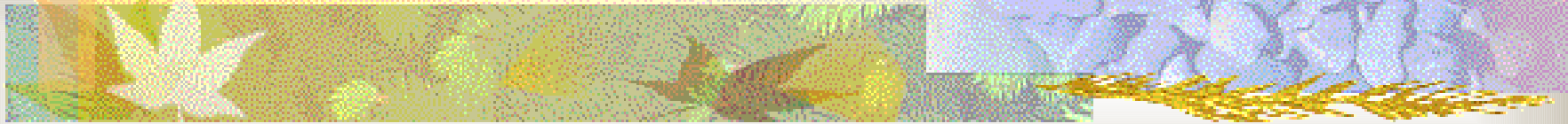
Efektywność klasyfikatorów łączonych z pojedynczych procedur dyskryminacyjnych we wspomaganie rozpoznawania chorób

Kierownik projektu:

Dr Małgorzata Ćwiklińska-Jurkowska

Katedra Podstaw Teoretycznych Nauk Biomedycznych i Informatyki

Medycznej





Cel pracy

Porównanie efektywności różnych metod łączenia klasyfikatorów na podstawie wielowymiarowych zbiorów danych, opisujących pacjenta

Zbadanie zależności efektywności klasyfikatorów łączonych od zróżnicowania

Wykorzystanie wyników do wspomagania rozpoznawania chorób



Material

zbiory danych, opisujących pacjenta
(anonimowe)

z klinik

dostępne w bazach internetowych

Liczebność grup waha się od 150 do 6000
pacjentów w poszczególnych
zagadnieniach medycznych



Zastosowane klasyfikatory podstawowe

**Metody Bayesowskie parametryczne (liniowa,
kwadratowa)**

**Metody Bayesowskie nieparametryczne (jądrowa,
najbliższego sąsiada)**

Dyskryminacja logistyczna

Drzewa klasyfikacyjne

CART (Breiman 1984)

QUEST tress (Loh & Shih 1997)

Sieci neuronowe



Macierz profilu decyzji $DP(x)$

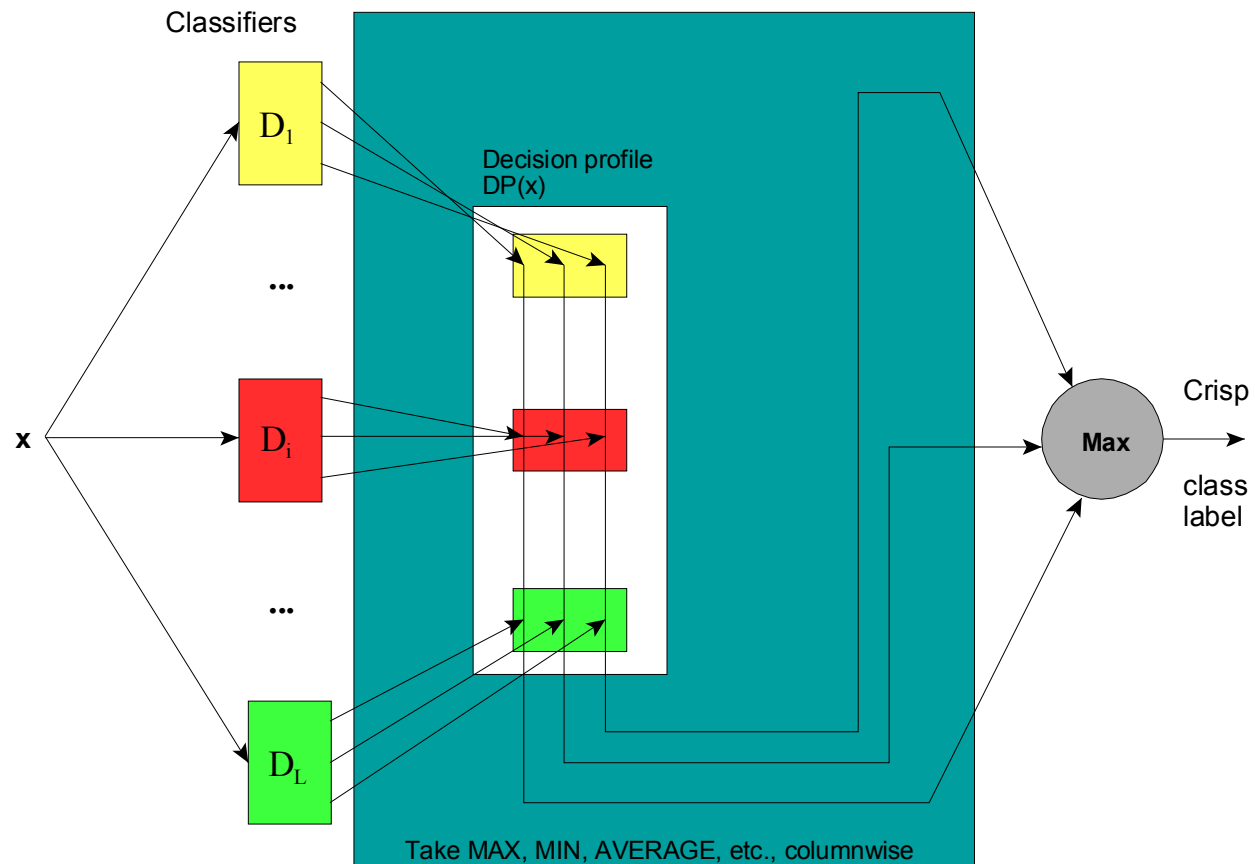
L klasyfikatorów

c grup

dla obserwacji x

$$DP(x) = \begin{bmatrix} d_{11}(x) & \dots & d_{1j}(x) & \dots & d_{1c}(x) \\ \dots & \dots & \dots & \dots & \dots \\ d_{i1}(x) & \dots & d_{ij}(x) & \dots & d_{ic}(x) \\ \dots & \dots & \dots & \dots & \dots \\ d_{L1}(x) & \dots & d_{Lj}(x) & \dots & d_{Lc}(x) \end{bmatrix}$$

Łączenie klasyfikatorów





Stosowane metody łączenia

Ze względu na zestaw cech: Metody
stacked (ten sam zestaw cech) i
parallel (różne podzbiory cech)

Ze względu na użycie w drugim etapie
zbioru uczącego:

Metody fixed i
trained

Fig. 2

The example explaining simple fusion stacked method
on the basis of one observation (number 1216) from testing set

obs1216		PROBABILITIES A POSTERIORI				
CONSTITUENT CLASSIFIERS		GROUP 1	GROUP 2	GROUP3		
		FROM GROUP			INTO GROUP	
LINEAR	PP	2	1E-05	0,06687	0,93312	3
QUADRATIC	PE	2	0,99996	4E-05	0	1
NORMAL KERNEL EQ COV MATR		2 DP(x)	0	0,86533	0,13467	2
NORMAL KERNEL NOT EQ COV MATR		2	0	1	0	2
6 NEAREST NEIGHBOUR		2	0	0,66667	0,33333	2

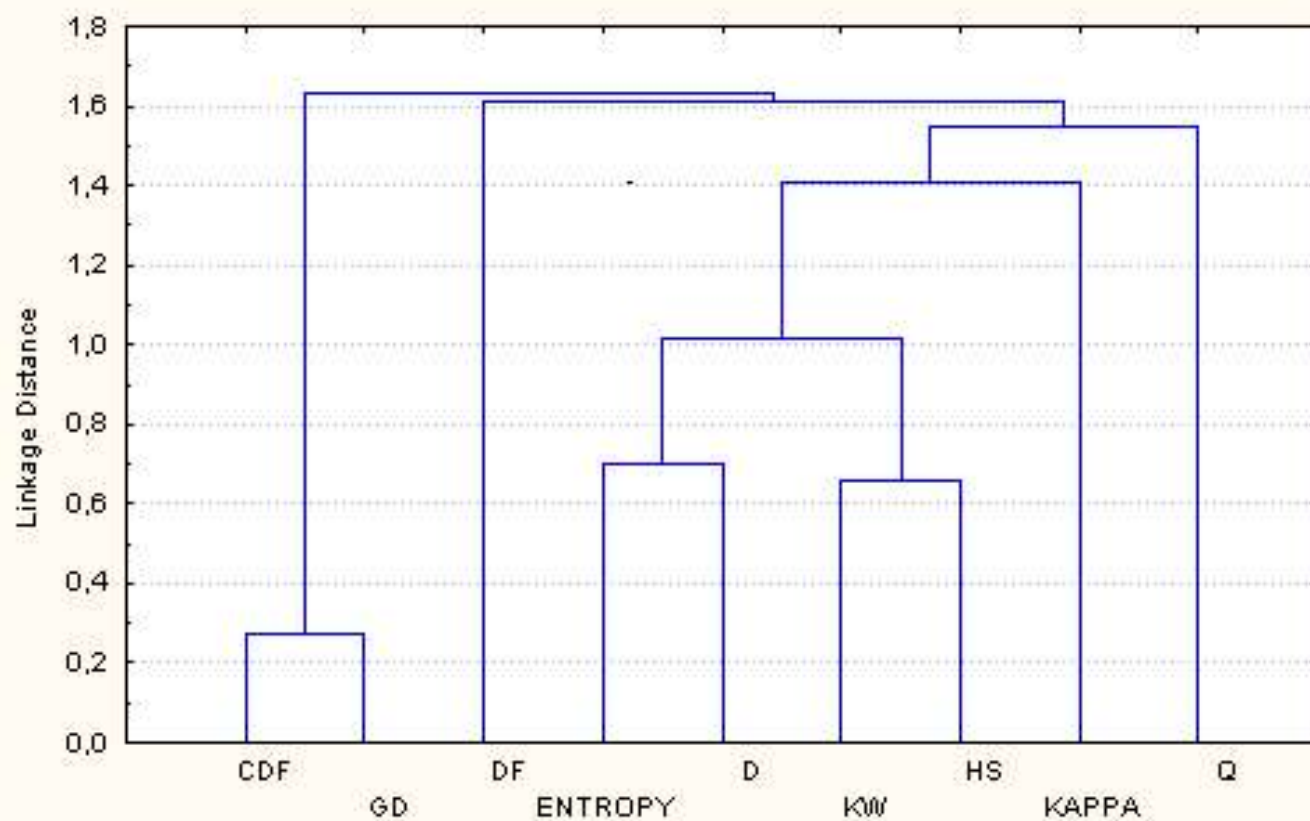
COMBINED FIXED METHODS:

	FROM GROUP				INTO GROUP
MAXIMUM	2	(0,99996	1	0,93312)	2
MINIMUM	2	(0	4E-05	0)	2
AVERAGE	2	(0,19999	0,519782	0,280224)	2
MEDIAN	2	(0	0,66667	0,13467)	1
PRODUCT	2	(0	1,54E-06	0)	2
MAJORITY VOTE	2				2

PP-probabilities a priori proportional to sizes of groups

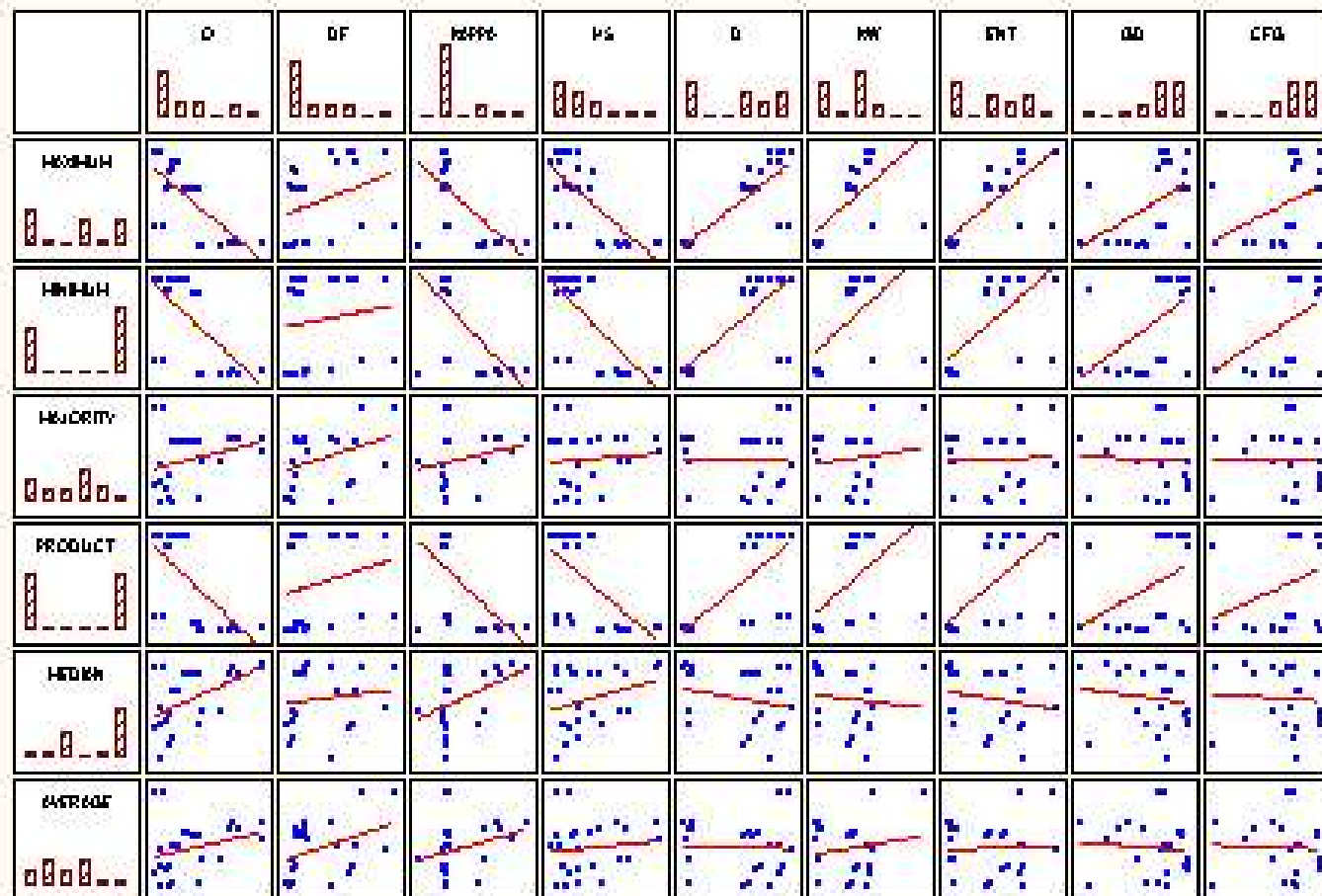
PE-probabilities a priori equal for all groups

Analiza skupień dla metod szacowania zróżnicowania zbioru klasyfikatorów.

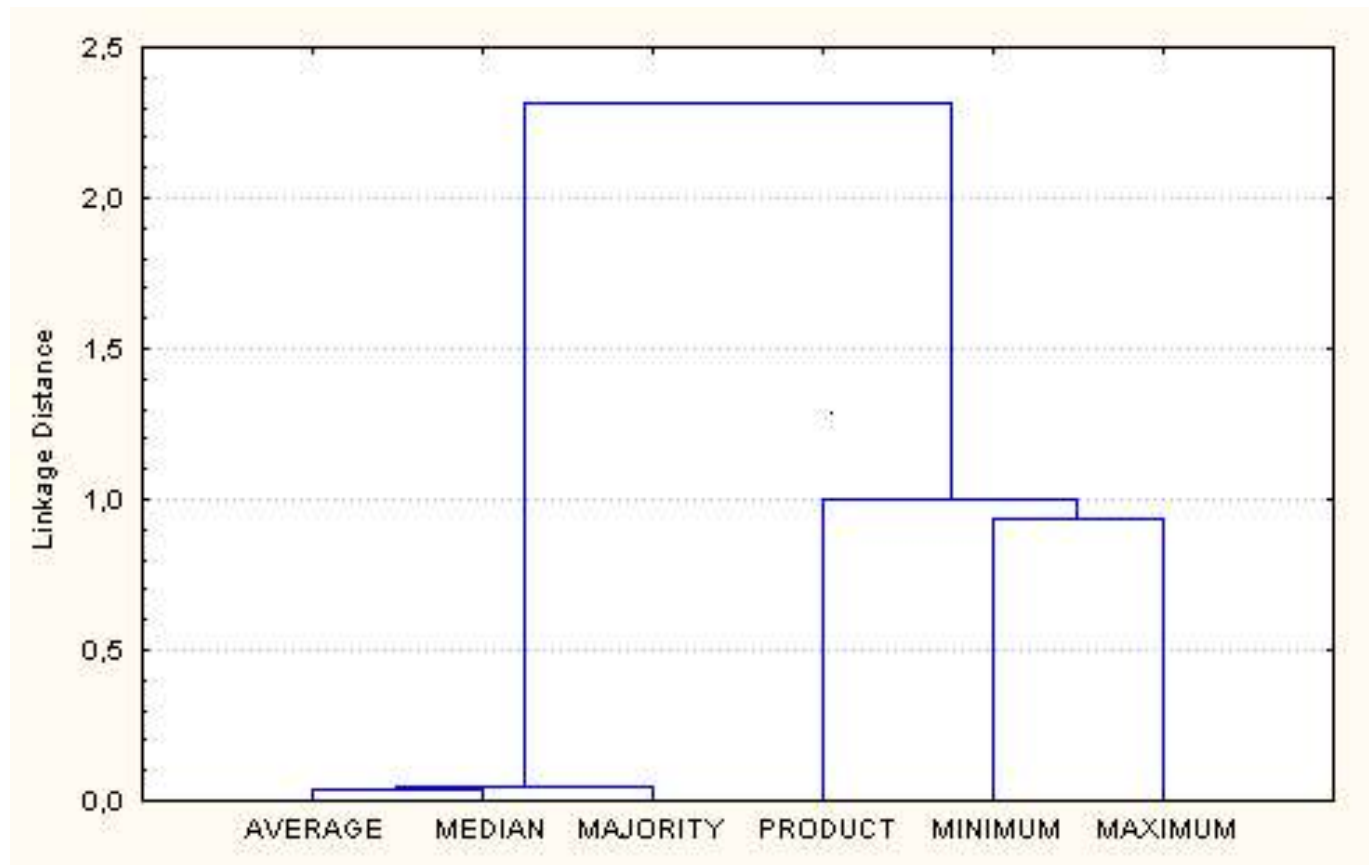


Zależność błędu klasyfikacji metod (maximum, minimum, majority vote, product, median, average) na podstawie próby testującej od miar zróżnicowania klasyfikatorów

Matrix Plot (fixedLeaTestDiversityPop.STA 15v24c)



Analiza skupień dla metod łączenia „fixed”.





Wnioski

Efektywność klasyfikacji poprawia dołączenie do zestawu klasyfikatorów funkcji dyskryminujących z innej grupy, na przykład do zestawu rozmytych klasyfikatorów Bayesowskich-klasyfikatorów „ostrych”: sieci neuronowych i drzew klasyfikacyjnych



Wnioski

Większą efektywność klasyfikacji osiąga się, gdy łączymy klasyfikatory oparte na różnych podzbiorach zestawu cech (parallel combining), niż dla tego samego całego zestawu cech

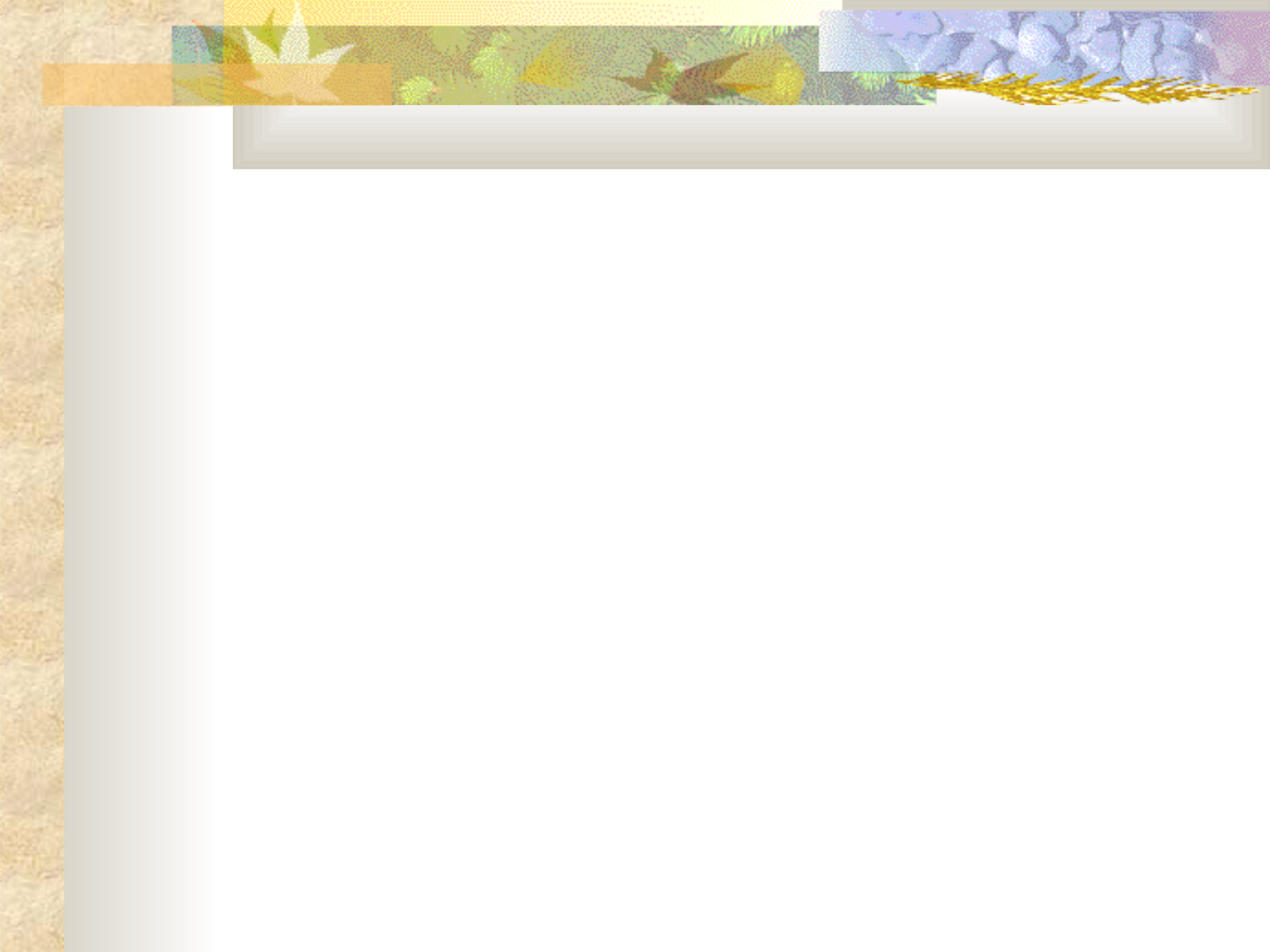


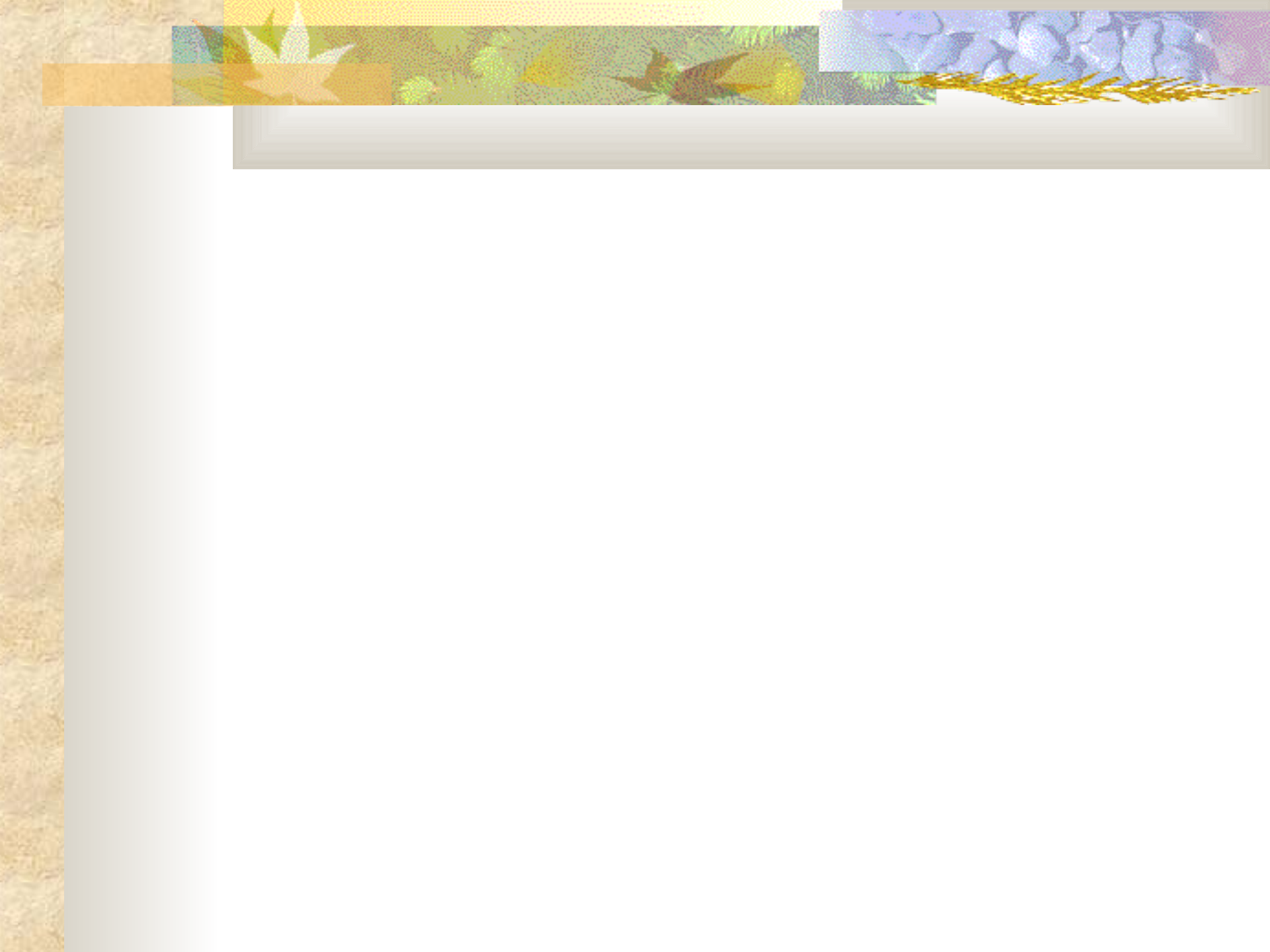
Wnioski

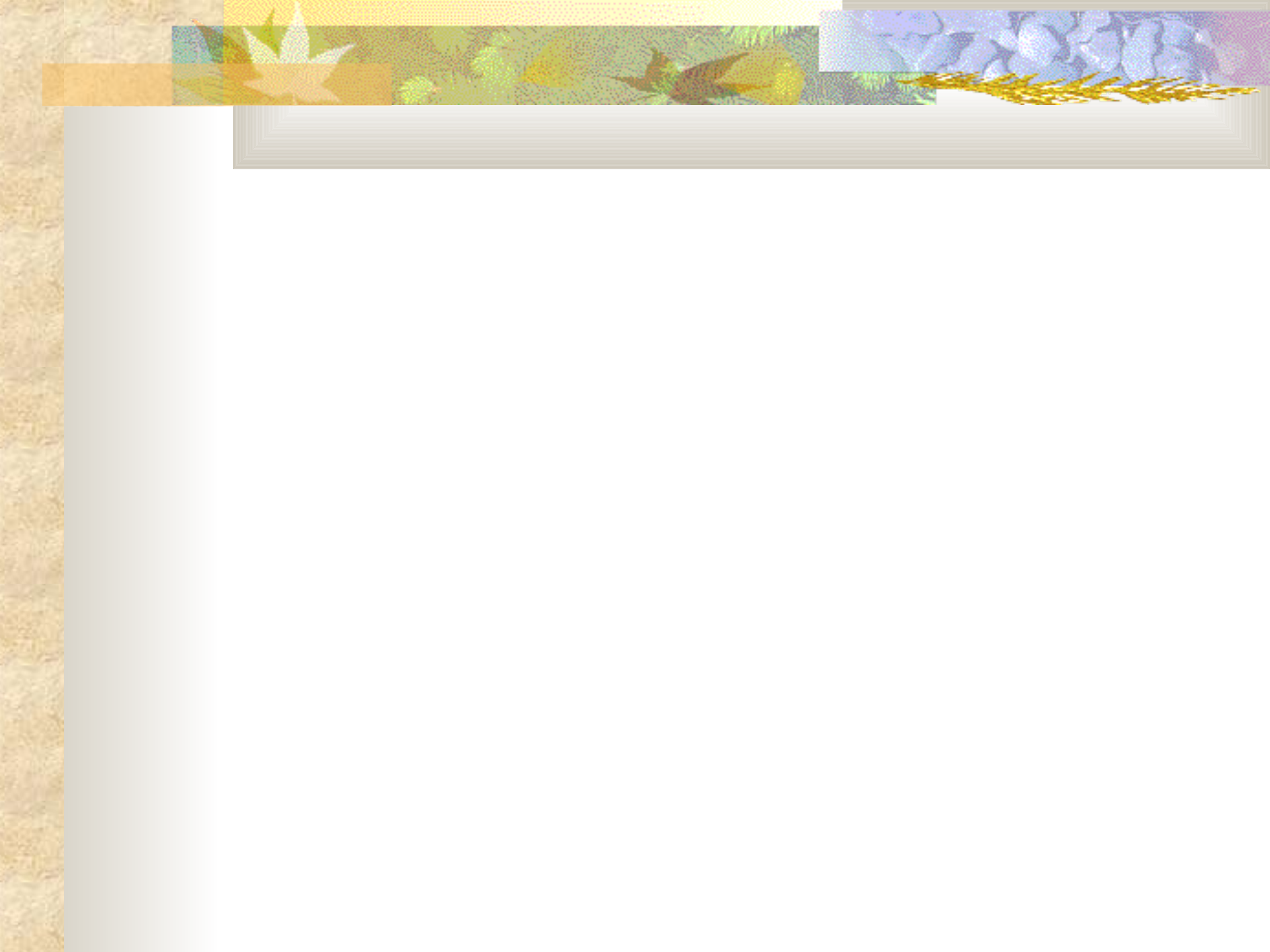
Nawet klasyfikatory bazowe o gorszej efektywności mogą być znaczące w zwiększeniu łącznej efektywności- ze względu na większe zróżnicowanie zbioru klasyfikatorów

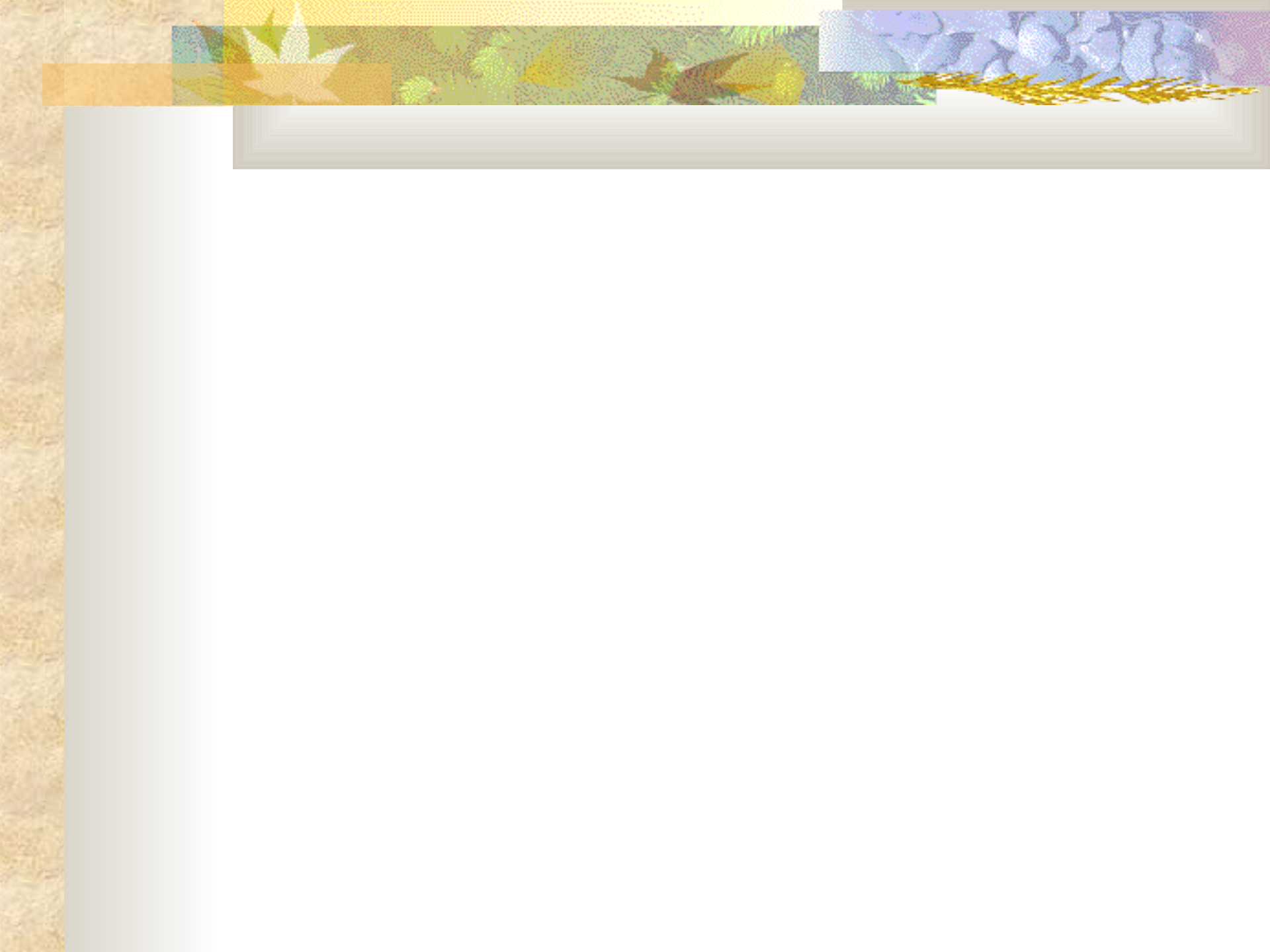
Tab. 3 Spearman correlations between measures of diversity and errors of fixed fusion methods (bolded-significant $p=0.05$)

	Q	DF	KAPP	HS	D	KW	ENT	GD	CFG
Minimum	-0.75	0.41	-0.55	-0.69	0.77	0.74	0.78	0.36	0.36
Maximum	-0.73	0.37	-0.55	-0.63	0.75	0.72	0.75	0.40	0.42
Majority	0.36	0.42	0.49	0.10	-0.09	-0.06	-0.16	-0.22	-0.23
Product	-0.78	0.42	-0.57	-0.64	0.78	0.76	0.78	0.37	0.39
Median	0.41	0.40	0.51	0.21	-0.18	-0.14	-0.22	-0.21	-0.22
Average	0.37	0.42	0.48	0.17	-0.11	-0.07	-0.17	-0.21	-0.22











Wnioski

Mniejsze błędy klasyfikacji otrzymano dla metod „trained” niż „fixed”

Wpływ zróżnicowania metod klasyfikacyjnych na poprawność klasyfikacji jest istotniejszy dla metod „fixed”.