



*Original Contribution*

## СТАТИСТИЧЕСКИ ХАРАКТЕРИСТИКИ НА ЦИФРОВИТЕ ИЗОБРАЖЕНИЯ НА ЗДРАВИ И ЗАБОЛЕЛИ ОТ ФУЗАРИОЗА ЦАРЕВИЧНИ СЕМЕНА

Е. Кирилова, П. Даскалов, Р. Цонев, Ц. Драганова

*Русенски Университет "А. Кънчев"*

### ABSTRACT

This paper presents basic descriptive statistics related to 17 colour features extracted from digital images of sound and Fusarium – damaged corn kernels of 6 varieties. The hypothesis of normal distribution and the correlation between varieties, regarding to refereed colour features were examined.

В доклада са представени основни статистически характеристики на 17 цветови признака на цифровите изображения на 6 сорта здрави и болни от фузариоза царевични семена. Допълнително е проверена хипотеза за нормално разпределение на данните и са оценени корелационните връзки между сортовете по отношение на разглежданите цветови признаци.

**Keywords:** statistical approach, analysis of digital images, corn kernels

**Ключови думи:** Статистически методи, Анализ на цифрови изображения, царевични семена

### УВОД

Определяне качеството на семената за посев е една от първостепенните задачи в селското стопанство, които е необходимо да бъдат решени, тъй като е свързана пряко с качеството и количеството на очакваната реколта. Основен показател за качество на царевичните семена е заболяемостта, а Фузариоза (*Fusarium spp.*) е една от най-значимите болести. Приложението на различни техники за обработка на изображения има основна роля при оценяването на хранителни продукти [1], цветът представлява характерен признак и мощно средство за описание на зърнени храни [2], [3], [6], плодове и зеленчуци [7] и др.

Приложението на автоматизирани системи, използващи компютърно зрение [5] е основна тенденция, свързана с оптимизиране на техниките за

оценяване. При реализирането на една такава система, усилията са насочени към намиране на универсални алгоритми, процедури и класификатори за разпознаването на заболяването. За целта е необходимо да бъдат извършени редица предварителни статистически анализи на опитните данни от обучаващата извадка [4], за да се установи влиянието на основните характеристики на семената върху избраните критерии и процедури за разпознаване. Може да се каже, че това е най-съществената задача при класификацията на образи, тъй като липсата на такъв предварителен анализ може да доведе до получаване на неточни резултати.

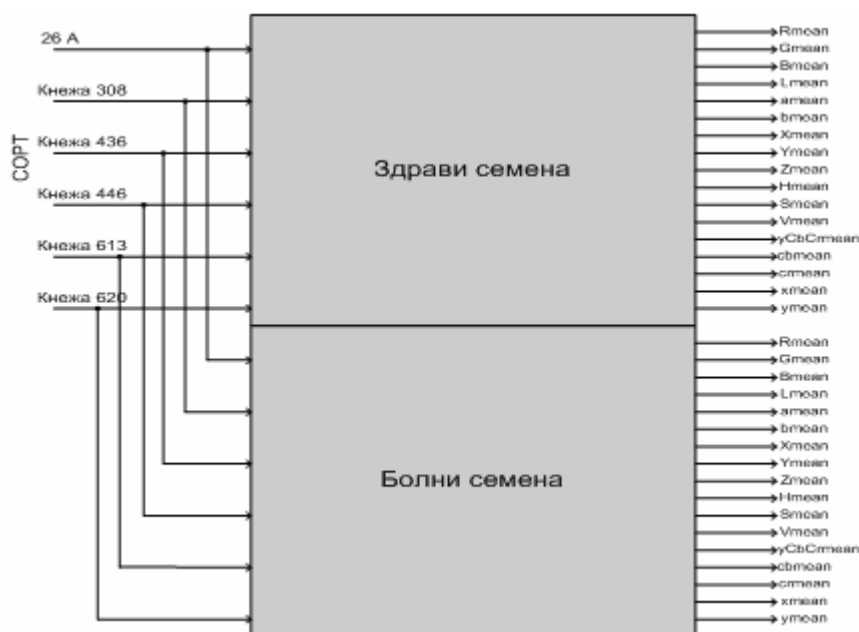
Имайки предвид изложеното дотук, основна цел на настоящия доклад е да се определят статистическите характеристики на цифровите изображения на здрави и болели от *Fusarium Moniliforme* царевичните семена, както и да се оценят съществуващи зависимости между статистическите оценки за шест сорта семена, по отношение на различни цветови признаци.

\* **За контакти:** Пламен И. Даскалов, Русенски Университет "А. Кънчев" 7017, гр. Русе, ул. "Студентска" 8 тел.: 082/ 888 281 e-mail: [daskalov@ru.acad.bg](mailto:daskalov@ru.acad.bg)

## ОБЕКТ И МЕТОДИКА

Обект на изследването са 6 сорта царевични семена – 26А, Кнежа 308, Кнежа 436, Кнежа 446, Кнежа 613, Кнежа 620. Съответно използваните означения на изброените сортове (А1 - А6) се отнасят за здрави, а (В1 – В6) – за болни семена. За всеки сорт са снети изображения на 150 здрави и 150 болни от розова фузариоза семена. Изображенията са получени с размер (352 x 289) пиксела, с 8 битово кодиране за червената (R), зелената (G) и синята (B) компоненти на RGB цветовия модел. Използвана е цветна камера: PИH – 7030 H / IR с разделителна способност 1/3'' CCD, 752 x 582 pix (352 x 289), 460 TV линии, 16 млн. цвята.

По нататъшната обработка на изображенията включва сегментация на изображенията по H (Hue) компонентата на HSV цветовия модел, като целта е да се отделят пикселите принадлежащи на фона, от тези, принадлежащи на обекта. Пикселите, принадлежащи на семето се преобразувани в HSV, Lab, XYZ, YCbCr, xyY цветови модели [8]. Формирани са матрици със стойностите на компонентите на цветовете модели (R, G, B, L, a, b, X, Y, Z, H, S, V, Y, Cb, Cr, x, y) за изображенията на семената. На фиг.1 е показана структурата на първичните данни, за здрави и болни семена от всеки сорт са получени осреднени стойности на компонентите на цветовете модели.



Фигура1. Структура на първичните данни

За да се оцени влиянието на цвета на семето на отделните сортове царевича върху избраните информативни цветови признаци чрез стандартни статистически методи и по-конкретно чрез дисперсионен анализ е необходимо да се провери дали са спазени предпоставките за провеждане на анализа [9]. Необходимо е да се провери условието за нормалност на разпределението на цветовете признаци за различните сортове семена, съответно при здрави и болни семена. Прилагат се два метода :

- метод на Андерсън – Дарлинг, използва т.нар. нормална вероятностна графика, чиито мащаб е избран така, че ако данните са нормално разпределени, то те да се подредят върху една права линия;
- *Lilieforce* тест – Разглежда се

хипотезата, че данните представени във вид на вектора  $X$  имат нормално разпределение, ако резултата е  $H=0$  то хипотезата се приема, ако  $H=1$  хипотезата се отхвърля. Командата в *MATLAB*, която изпълнява този тест има следния вид : *lilietest (X,alpha)*, където *alpha* е ниво на значимост, представлява вероятността, при която дадено събитие практически не може да се случи.

Първичната статистическа обработка на опитните данни, получени от проведеното експериментално изследване, се състои в пресмятането на оценки на основните числени характеристики, т.е цветовете признаци, определящи особеностите на изучаваните случайни величини. За да може по-нататък да бъде подбран подходящ вероятностен модел за

описание на свойствата на изучавания показател се построява неговото емпирично разпределение, в графичен вид то представлява хистограма.

## ИЗЛОЖЕНИЕ

В табл.1 са показани пресметнатите статистическите оценки. Стойностите на

признаците са получени като усреднени стойности на 17-те цетови признака за семената, от всеки сорт по отделно. Стандартните отклонения (std) отразяват разсейването на стойностите на признаците около средноаритметичната им стойност по пиксели, усреднени за всички семена от даден сорт.

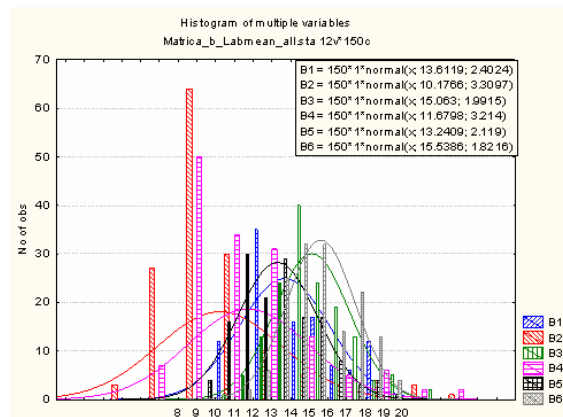
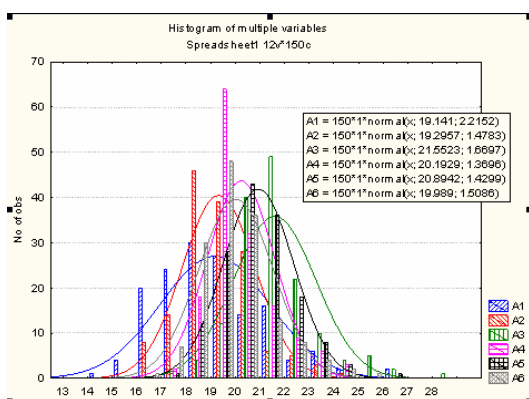
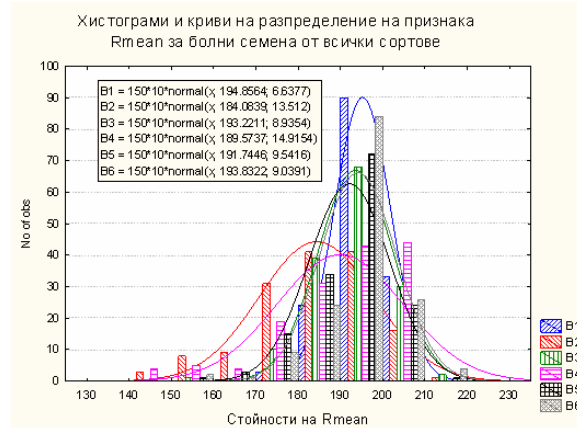
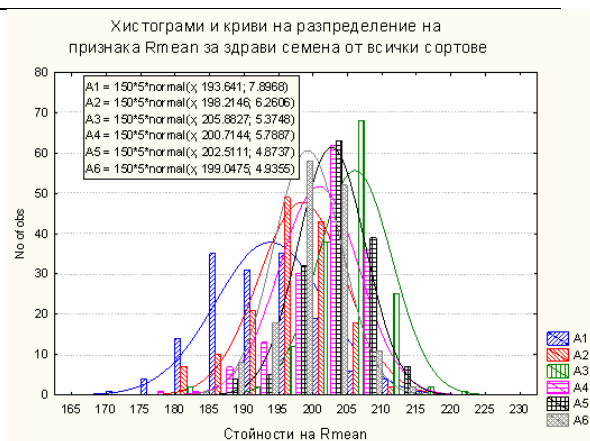
*Таблица 1. Средни стойности и стандартни отклонения (std) на цетовите признаци на семената по сортове*

Средни стойности на статистическите компоненти												
СОРТ	A1 - 26 A		A2 - Кнежа 308		A3 - Кнежа 436		A4 - кнежа 446		A5 - Кнежа 613		A6 - Кнежа 620	
Признаци	здрави	болни	Здрави	болни	здрави	болни	здрави	болни	здрави	болни	здрави	болни
Rmean	193,64	194,86	198,21	184,08	205,88	193,22	200,71	189,57	202,51	191,74	199,05	193,83
std(Rmean)	32,81	32,09	41,58	35,05	35,74	34,22	32,64	33,45	35,78	34,26	37,3	32,75
Gmean	122,45	131,72	123,93	120,44	133,68	128,51	126,4	126,83	129,71	128,07	126,56	131,99
std(Gmean)	15,17	17,55	23,92	22,65	18,63	18,98	18,48	20,42	19,2	20,08	20,8	20,09
Bmean	72,61	94,79	76,67	88,67	77,92	90,32	79,29	91,95	80,99	92,37	79,22	95,59
std(Bmean)	20,33	19,76	15,8	21,5	19,74	17,62	19,2	22,05	18,57	17,63	17,01	17,35
Lmean	77,28	79,34	76,59	79,6	78,87	78,38	78,26	80,24	78,44	78,8	77,28	78,89
std(Lmean)	4,52	4,48	6,03	5,94	4,84	5,03	4,88	5,07	5,01	5,19	5,61	5,22
amean	7,69	6,76	9,44	5,18	7,9	7,78	8,94	5,57	8,67	7	8,79	7,92
std(amean)	5,19	5,13	4,73	6,6	4,02	4,64	5,08	5,86	4,38	5,2	4,53	3,94
bmean	19,14	13,61	19,3	10,18	21,55	15,06	20,19	11,68	20,89	13,24	19,99	15,54
std(bmean)	11,83	8,58	11,29	8,95	11,32	9,23	10,02	9,2	10,25	8,34	10,52	7,82
Xmean	54,18	57,17	55,43	53,42	58,12	56,15	56,37	55,43	57,23	55,99	56,1	57,09
std(Xmean)	6,67	7,58	9,9	8,83	8,01	8,18	7,45	8,18	8,13	8,48	8,66	8,29
Ymean	52,82	56,12	53,75	51,88	57,15	54,96	54,72	54,22	55,84	54,77	54,62	56,13
std(Ymean)	6,35	7,3	9,93	8,96	7,83	7,97	7,39	8,14	7,96	8,32	8,59	8,26
Zmean	34,5	43,15	36,1	40,23	37,12	41,32	37,21	41,81	38,02	42,04	37,18	43,45
std(Zmean)	7,69	7,94	6,32	8,78	7,59	7,07	7,64	8,93	7,29	7,3	6,71	7,17
Hmean	29,47	27,25	27,16	28,99	29,01	26,56	26,12	29,43	26,78	27,15	26,45	25,98
std(Hmean)	34,55	35,22	37,07	42,93	28,75	35,66	26,34	40,04	27,22	36,96	29,15	32,34
Smean	0,6	0,5	0,59	0,5	0,6	0,52	0,59	0,5	0,58	0,5	0,58	0,49
std(Smean)	0,17	0,15	0,15	0,16	0,16	0,14	0,15	0,16	0,15	0,14	0,15	0,13
Vmean	193,75	194,98	198,24	184,43	205,91	193,29	200,75	189,8	202,55	191,88	199,08	193,88
std(Vmean)	32,67	31,95	41,54	34,99	35,67	34,14	32,56	33,33	35,7	34,15	37,24	32,69
yCbCrmean	20,23	22,84	29,61	23,64	31,38	28,35	33,62	22,9	32,82	27,07	36,21	34,29
std(yCbCrmean)	44,88	48,84	53,03	47,52	55,55	52,33	55,35	47,89	55,72	51,28	56,74	56,41
cbmean	-5,02	-4,14	-7,26	-4,2	-8,11	-5,58	-8,14	-4,15	-7,91	-5,04	-8,78	-6,36
std(cbmean)	13,27	10,72	15,27	10,52	16,44	12,21	15,18	10,59	15,43	11,3	15,87	12,2
crmean	6,42	6,01	9,28	6,56	9,2	7,66	10,29	6,15	9,74	7,19	10,83	8,59
std(crmean)	14,13	12,66	16,62	13,44	16,47	14,11	17,28	12,93	16,82	13,59	17,34	14,18
xmean	54,18	57,17	55,43	53,42	58,12	56,15	56,37	55,43	57,23	55,99	56,1	57,09
std(xmean)	6,67	7,58	9,9	8,83	8,01	8,18	7,45	8,18	8,13	8,48	8,66	8,29
ymean	52,82	56,12	53,75	51,88	57,15	54,96	54,72	54,22	55,84	54,77	54,62	56,13
std(ymean)	6,35	7,3	9,93	8,96	7,83	7,97	7,39	8,14	7,96	8,32	8,59	8,26

**Таблица 2** Общи средни стойности и стандартни отклонения на цветовите признаци на семената от всички сортове

	Здрави семена	Болни семена
<b>Цветови признак</b>	Средна ст-т ± отклонение	Средна ст-т ± отклонение
<b>R (RGB)</b>	200 ± 35,97	191,22 ± 33,64
<b>G</b>	127,12 ± 19,37	127,93 ± 19,96
<b>B</b>	77,78 ± 18,44	92,28 ± 19,32
<b>L (Lab)</b>	77,79 ± 5,15	79,21 ± 5,16
<b>a</b>	8,57 ± 4,66	6,70 ± 5,23
<b>b</b>	20,18 ± 10,87	13,22 ± 8,69
<b>X (XYZ)</b>	56,24 ± 8,14	55,88 ± 8,26
<b>Y</b>	54,82 ± 8,01	54,68 ± 8,16

	Здрави семена	Болни семена
<b>Цветови признак</b>	Средна ст-т ± отклонение	Средна ст-т ± отклонение
<b>Z</b>	36,69 ± 7,21	42 ± 7,87
<b>H (HSV)</b>	27,5 ± 30,51	27,56 ± 37,19
<b>S</b>	0,59 ± 0,16	0,50 ± 0,15
<b>V</b>	200,05 ± 35,9	191,38 ± 33,54
<b>Y (YCbCr)</b>	30,65 ± 53,5	26,52 ± 50,71
<b>Cb</b>	-7,54 ± 15,24	-4,91 ± 11,26
<b>Cr</b>	9,29 ± 16,44	7,03 ± 13,49
<b>x (xyY)</b>	56,24 ± 8,14	55,88 ± 8,26
<b>y</b>	54,82 ± 8,01	54,68 ± 8,16



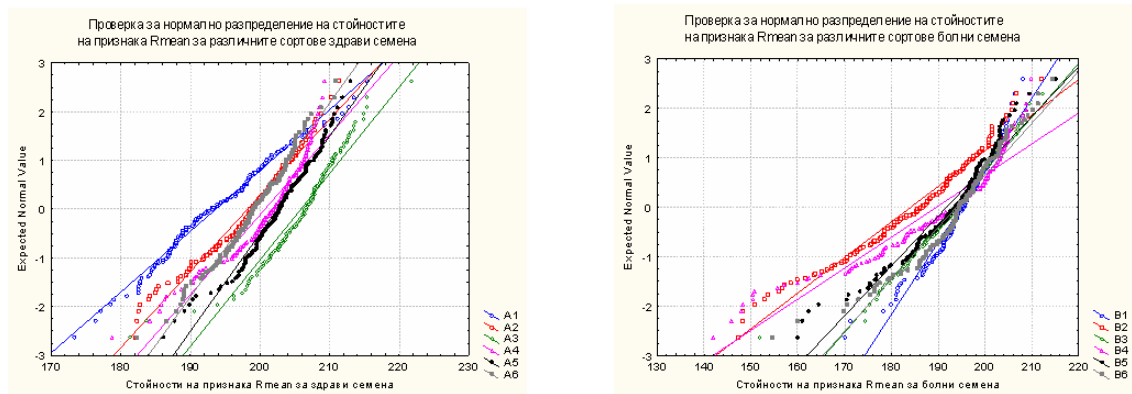
**Фигура 2.** Хистограми и криви на разпределение за признаци Rmean и bmean, за здрави(вляво) и болни(вдясно) семена от всички сортове

Както се вижда от фиг.2 стойностите на признаците са разпределени в голям диапазон, а освен това се характеризират и със значително разсейване спрямо средната стойност, както за здравите, така и за болните семена.

**Оценяване на разпределението на стойностите на цветовите признаци за различните сортове семена.**

Проверява се дали разпределението на

средните стойностите на 17-те цветови признака е нормално, за всеки сорт съответно за здрави и болни семена. Вида на графиките при проверка за нормалност по метода на Андерсън – Дарлинг е даден на фиг.3. В таблица 3 са дадени обобщени резултати от проведен *Lilieforce* тест за нормално разпределение с ниво на значимост  $\alpha=0.05$ .



Фигура 3. Проверка за нормално разпределение по метода на Андерсън – Дарлинг

Таблица 3. Резултати от Lilieforce тест за нормално разпределение

сорт		26А		Кнежа30 8		Кнежа43 6		Кнежа44 6		Кнежаб1 3		Кнежаб2 0	
признаци		здр	бол	здр	бол	здр	бол	здр	бол	здр	бол	здр	бол
1	Rmean	0	1	1	1	1	0	1	1	0	1	0	1
2	Gmean	1	0	1	0	0	0	1	0	0	0	1	0
3	Bmean	1	0	1	1	1	0	0	0	0	0	1	0
4	Lmean	0	0	0	1	0	0	1	1	0	0	0	0
5	amean	0	1	0	1	0	1	0	1	0	1	1	0
6	bmean	1	1	1	1	1	1	1	1	0	1	1	1
7	Xmean	0	0	1	0	1	0	1	1	1	0	1	0
8	Ymean	1	0	1	0	0	0	1	0	1	0	1	0
9	Zmean	1	0	0	1	1	0	1	1	0	0	1	0
10	Hmean	1	1	1	1	0	1	0	1	0	1	0	1
11	Smean	1	1	0	0	1	0	0	1	0	1	1	0
12	Vmean	0	1	1	1	1	0	1	1	0	1	0	1
13	yCbCr mean	0	0	0	1	0	1	1	0	0	1	0	1
14	cbmean	0	0	1	0	0	0	1	1	0	0	0	0
15	crmean	1	0	0	0	0	1	1	0	0	1	0	0
16	xmean	0	0	1	0	1	0	1	1	1	0	1	0
17	ymean	1	0	1	0	0	0	1	0	1	0	1	0

Получените резултати от проверките за нормалност на разпределението сочат, че данните за цветовите признаци като цяло не са разпределени по закона за нормалното разпределение.

#### Оценка на корелацията между сортовете по даден признак.

Зависимостта между сортовете по двойки се оценява чрез изчисляване на

корелационните им коефициенти ( $r_{ij}$ ). В таблици 3 и 4 са показани стойностите на коефициентите за изследваните шест сорта, съответно по признаци R (RGB) и b(Lab).

Разглеждаме хипотезата  $H_0$  – има корелация между данните, при равнище на значимост  $\alpha = 0.05$ . За данните, които са отбелязани с червен цвят хипотеза  $H_0$  е вярна и има връзка между съответните сортове по разглеждания признак.

Таблица 3. Оценки на корелация между сортовете, под влияние на признак Rmean

Correlations (Matrica_Rmean_all.sta) Marked correlations are significant at p < .05000 N=150 (Casewise deletion of missing data)												
Variable	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6
A1	1.00	0.01	-0.08	-0.07	0.06	0.21	-0.06	0.06	-0.10	-0.03	-0.01	0.05
A2	0.01	1.00	0.02	-0.02	0.01	-0.03	-0.15	-0.01	0.01	-0.03	0.15	0.09
A3	-0.08	0.02	1.00	0.12	0.02	-0.02	-0.04	-0.08	0.10	-0.15	-0.08	-0.05
A4	-0.07	-0.02	0.12	1.00	0.04	-0.03	0.00	0.14	0.01	-0.02	0.04	-0.02
A5	0.06	0.01	0.02	0.04	1.00	0.17	0.12	0.26	0.14	0.12	-0.02	-0.04
A6	0.21	-0.03	-0.02	-0.03	0.17	1.00	0.05	0.16	0.11	-0.15	-0.04	0.03
B1	-0.06	-0.15	-0.04	0.00	0.12	0.05	1.00	0.11	0.18	0.06	-0.04	-0.09
B2	0.06	-0.01	-0.08	0.14	0.26	0.16	0.11	1.00	-0.04	-0.02	-0.04	-0.02
B3	-0.10	0.01	0.10	0.01	0.14	0.11	0.18	-0.04	1.00	-0.01	0.03	0.05
B4	-0.03	-0.03	-0.15	-0.02	0.12	-0.15	0.06	-0.02	-0.01	1.00	0.05	0.05
B5	-0.01	0.15	-0.08	0.04	-0.02	-0.04	-0.04	-0.04	0.03	0.05	1.00	-0.11
B6	0.05	0.09	-0.05	-0.02	-0.04	0.03	-0.09	-0.02	0.05	0.05	-0.11	1.00

Таблица 4. Оценки на корелация между сортовете, под влияние на признак bmean

Correlations (Matrica_b_Labmean_all.sta) Marked correlations are significant at p < .05000 N=150 (Casewise deletion of missing data)												
Variable	A1	A2	A3	A4	A5	A6	B1	B2	B3	B4	B5	B6
A1	1.00	0.62	0.56	0.69	0.68	0.74	0.74	0.73	0.38	0.66	0.72	-0.28
A2	0.62	1.00	0.47	0.59	0.52	0.59	0.47	0.62	0.20	0.53	0.48	-0.18
A3	0.56	0.47	1.00	0.57	0.52	0.63	0.61	0.69	0.38	0.62	0.50	-0.16
A4	0.69	0.59	0.57	1.00	0.58	0.69	0.65	0.66	0.36	0.62	0.57	-0.26
A5	0.68	0.52	0.52	0.58	1.00	0.63	0.61	0.68	0.32	0.62	0.61	-0.22
A6	0.74	0.59	0.63	0.69	0.63	1.00	0.68	0.68	0.38	0.65	0.64	-0.21
B1	0.74	0.47	0.61	0.65	0.61	0.68	1.00	0.69	0.47	0.69	0.66	-0.39
B2	0.73	0.62	0.69	0.66	0.68	0.68	0.69	1.00	0.37	0.73	0.65	-0.27
B3	0.38	0.20	0.38	0.36	0.32	0.38	0.47	0.37	1.00	0.45	0.32	-0.19
B4	0.66	0.53	0.62	0.62	0.62	0.65	0.69	0.73	0.45	1.00	0.64	-0.30
B5	0.72	0.48	0.50	0.57	0.61	0.64	0.66	0.65	0.32	0.64	1.00	-0.24
B6	-0.28	-0.18	-0.16	-0.26	-0.22	-0.21	-0.39	-0.27	-0.19	-0.30	-0.24	1.00

От таблици 3 и 4 се вижда, че стойностите на коефициента на корелация между сортове по признак R (RGB) не са големи, т.е. сортовете са независими помежду си, докато по признак b (Lab) съществува значителна зависимост.

## ИЗВОДИ

Разликите в цвета при различните сортове царевични семена оказват влияние върху средните стойности на цветовите параметри и границите, в които тези стойности се изменят. Вижда се, че областите на изменение на повечето цветови признаци се припокриват, т.е. по тях не може точно да се определи към кой клас принадлежи царевичното семе. Областите на изменение на средните стойности на признаците: B(RGB), b(Lab), Z(XYZ) и S(HSV) при шестте изследвани сорта имат най-малко застъпване за двата класа (здрав и заразен) семена. Стойностите на признаците са разпределени в голям диапазон, а освен

това се характеризират и със значително разсейване спрямо средната стойност, както за здравите, така и за болните семена.

В заключение може да се каже, че по признак b (от Lab модела) най-голяма част от данните са корелирани помежду си. На второ място по влиятелност е признака a (от Lab модела). Най-слаба връзка между данните има по признаците G(RGB), v(HSV) и cr(yCrCb). припокриване

## ЛИТЕРАТУРА

1. Du Cheng-Jin, Da-Wen Sun Recent developments in the applications of image processing techniques for food quality evaluation, Trends in Food Science & Technology 15 (2004) 230–249
2. Liu J., M. R. Paulsen Corn whiteness measurement and classification using machine vision, 2000, Transactions of the ASAE VOL. 43(3): 757-763

3. Luo X., D.Jayas, S. Symons., Comparison of statistical and neural network methods for classifying cereal grains using machine vision., Transaction of ASAE, Vol. 42(2), 1999, pp. 413-419.
4. Markou Markos, Sameer Singh Novelty detection: a review—part 1: statistical approaches, Signal Processing 83 (2003) 2481 – 2497
5. Riquelme M.T., P. Barreiro, M. Ruiz-Altisent, C. Valero Olive classification according to external damage using image analysis, Journal of Food Engineering 87 (2008) 371–379
6. Venora G. , O. Grillo, R. Saccone Quality assessment of durum wheat storage centres in Sicily: Evaluation of vitreous, starchy and shrunken kernels using an image analysis system, Journal of Cereal Science (2009), doi:10.1016/j.jcs.2008.12.006
7. Дамянов Ч. Неразрушаващо разпознаване на качеството в системите за автоматично сортиране на хранителни продукти, Академично издателство на УХТ – Пловдив, 2006
8. Драганова Ц., Р. Цонев, Пл. Даскалов. Оценяване на заболяването фузариоза на царевични семена чрез анализ на цветовете им изображения. Научни трудове на РУ”Ангел Кънчев”, 2003, с. 25-32.
9. Митков А., Д.Минков. Статистически методи за изследване и оптимизиране на селскостопанска техника 2 част., София, 1993.