

Подбор на комплекс от информативни цветови признаци за разпознаване на царевични семена, заразени с Фузариоза (*Fusarium Moniliforme*)

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

Abstract: Selection of colour features for recognition of *Fusarium damaged corn seeds:* Assessment of informative color features for recognition of 6 species of *Fusarium damaged corn seeds* is presented in the paper. Vector of 17 color features of 6 colors space RGB, HSV, Lab, YCbCr, XYZ, xyY is defined. The SVM (Support vector machines) method is used for evaluation and ranking of color features. It is found that the SVM classifier gave classification accuracy of over 95% for corn seeds Kneja 620, Kneja 613, Kneja 303 and 26A. For corn seeds Kneja 446 the most informative color feature, with classification accuracy of 91,33% for is B (RGB), and for Kneja 436 this is S(HSV) with 94,67%.

Key words: pattern recognition, Support Vector Machines, corn seeds, Разпознаване на образи, Метод на опорните вектори, царевични семена

ВЪВЕДЕНИЕ

Автоматизираното окачествяване на редица земеделски продукти по техните цифрови изображения [4], [5], [6] налага необходимостта от използване на точни и бързи процедури за получаване, обработване и интерпретация на най-информативния комплекс от признаци, касаещ качествените характеристики на обектите. Задачата за оценка на качеството на самите признаци е тясно свързана с редукция броя на използваните при разпознаването признаци, като по този начин в голяма степен се намаляват необходимите изчислителни ресурси без загуба на информация. Първото условие при ранжирането на признаци е да се намери минимално описание на признаковото пространство [3], а второто условие е да се намерят такива признаци, които са колкото е възможно по-малко зависими един от друг [1].

Основна цел на настоящия доклад е да се определи информационната ценност на цветови признаци от шест цветови модела, при разпознаване на заразени с Розова Фузариоза царевични семена, като се определи точността на класификация по метода на опорните вектори.

ОБЕКТ И МЕТОДИКА

1. Изходни данни

Обект на изследването са цветни цифрови изображения на здрави и заразени с Розова Фузариоза (*Fusarium Moniliforme*) царевични семена от 6 сорта- 26А, Кнежа 308, Кнежа 436, Кнежа 446, Кнежа 613, Кнежа 620, за всеки от сортовете са снети изображения за 150 здрави и 150 заразени семена. Изображенията са получени с размери (352 x 289) пиксела, като е използвана цветна камера: PIN – 7030 H / IR с разделителна способност 1/3" CCD, 752 x 582 pix (352 x 289), 460 TV линии, 16 млн. цвята и обектив Tokina, който позволява ръчна промяна на фокусното разстояние от 6 до 60 mm. Разстоянието между камерата и семето при снемане на изображенията е 26 см. Осветлението е осигурено от 2 бр. кръгли луминисцентни лампи над работната сцена.

По нататъшната обработка на изображенията включва сегментация 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. За целите на разпознаването на заразени царевични семена по цветови признаци, от всеки пиксел от изображенията на семената се извличат по 17 цветови признака, съответстващи на 17-те компоненти (R, G, B, L, a, b, X, Y, Z, H, S, V, Y, Cb, Cr, x, y) от споменатите по горе цветови модели. Базата данни, която се използва при анализа съдържа средноаритметичните стойности на цветовите компоненти от всички пиксели в изображението на семето.

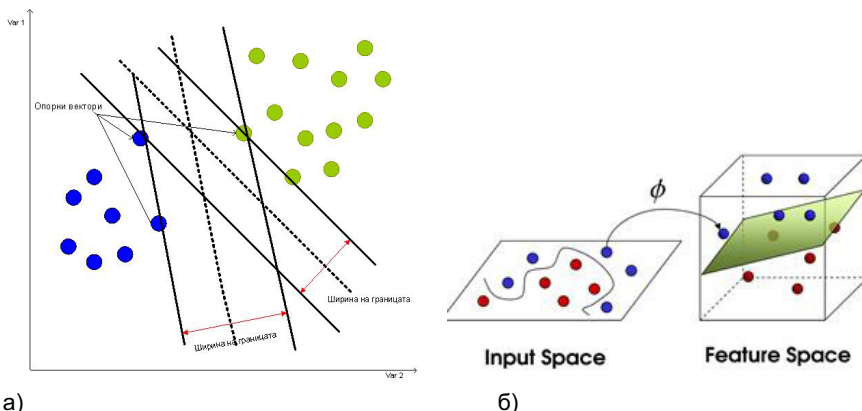
$$\bar{R}_m = \frac{1}{n} \sum_{i=1}^n R_i, \quad (1)$$

където R е стойността на R компонентата от RGB цветовия модел; m е номер на семето (m = 1, 2, ..., 150); i – номер на пиксел от изображението на семето ; n – брой на покселите от изображението на семето.

Всяко царевично семе е оценено от експерт и е класифицирано в клас „здрав“ или клас „заразени“ с Розова фузариоза.

2. Избор на информативен комплекс от цветови признаци

Изборът на метод за определяне информационната ценност на дефинираните цветови признаци при разпознаване на двата класа царевични семена (здрав, заразен) се основава на факта, че данните имат асиметрична структура и не е възможно класовете да се разделят директно с линейни функции. Методът на опорните вектори SVM (Support Vector Machines) е метод за класификация чрез нелинейно преобразуване на оригиналните данни в друго пространство с по-висока размерност, където обектите са линейно разделими (фиг.1.б) [2]. При SVM - метода по т.нар. опорните точки, които представляват гранични точки за даден клас данни в пространството на признаците се построява оптималния хиперплан. Той отговаря на условието, дистанцията между границите за двата класа да е максимална (фиг.1.а).



Фиг.1. Оптимален хиперплан при SVM метода. а) при линейно разделими области; б) при нелинейно разделими области

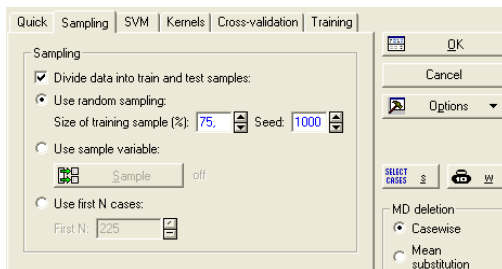
Функцията Φ определя изображение на входното пространство $K(x,y)$ в ново високоразмерно признаково пространство $K(x,y) = (\Phi(x), \Phi(y))$. Както за обучението, така и за класификацията се използва единствено скаларното произведение на

двойките от вектори $\Phi(x)$, $\Phi(y)$. С помощта на кернъл функция може косвено да се дефинира произведението в това пространство, като с използването на различни кернъл функции се конструират различни класификатори – линеен, RBF, полиномиален, невронна мрежа MLP [7,8]. В случая е избрана кернъл функция тип RBF(Radial Basis Function) с ширина σ , тя има следния вид:

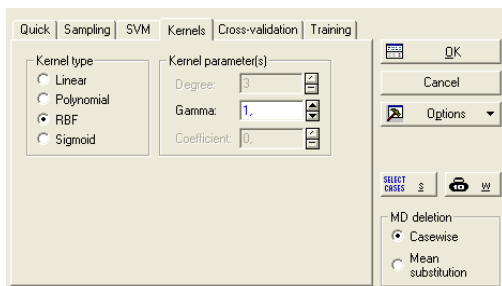
$$K(x, y) = \exp(-\|x - y\|^2 / (2\sigma^2)) \quad (2)$$

3.Получени резултати и анализ

За изпълнението на метода на опорните вектори е използван програмния пакет на StatSoft STATISTICA 8. Реализацията на метода е направена за радиално-базова функция (RBF), (фиг.3). Разгледан е случай за синтез на класификатор за два класа (здрави и заразени семена). В първия етап е извършен предварителен избор на параметри – размер на обучаващата извадка - 75 % от обема на общата извадка и параметъра Gamma, който определя ширината на RBF,по подразбиране има стойност 1,00 (фиг.2).

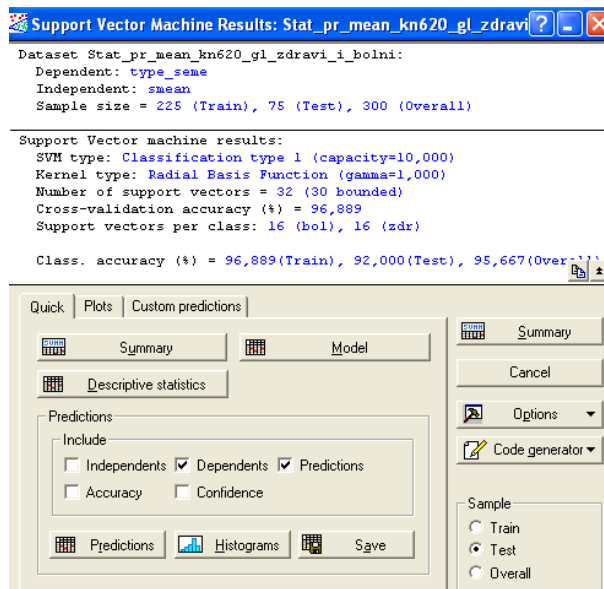


Фиг.2.Прозорец от програмна среда STATISTICA за дефиниране параметрите на SVM класификатора



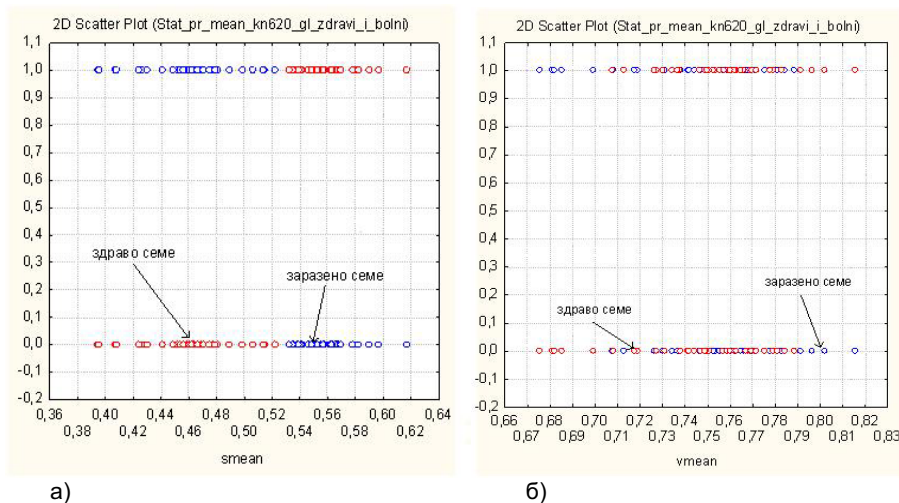
Фиг.3. Прозорец от програмна среда STATISTICA за определяне типа кернел функция на SVM класификатора

Следващата стъпка е да се направи оценка на модела по получената точност за цялата извадка от данни (обучаваща и тестова) поотделно за всеки от 17-те цветови признака. На фиг.4 се вижда етап от изпълнението на метода на опорните вектори, където дефинираната независима променлива е *mean*,формирани са 32 опорни вектора, по 16 за 2-та класа, а точността на крос-валидация е 98,889%.Показани са също и получените точности по отделно за обучаващата и за тестовата извадка.



Фиг.4 Прозорец от STATISTICA с резултатите от изпълнение на SVM класификация за признак *smean*

На фиг.5 в графичен вид са показани разпределенията на образите от тестовата извадка по признаци. С червени точки са отбелязани образи на здрави семена, а със сини точки – на заразени семена. По признак *smean* (фиг.5а) се вижда че образите са добре разделими, докато по признака *vmean* (фиг.5б) има пълно застъпване и може да се каже, че последния не е информативен.



Фиг. 5. Разпределение на образите от тестовата извадка а) по признак *smean*; б) по признак *vmean*

В таблица 1. са систематизирани резултатите за получената със SVM-класификатор, точност за 6- те изследвани сорта. Броя на построените опорни вектори при класификация поотделно за всеки от 17-те цветови признака са дадени в таблица 2.

Таблица 1. Резултати за точността на класификация (в %) по метода на опорните вектори SVM за цялата извадка от данни (обучаваща и тестова)

| Точност на класификация,% | | сорт | | | | | | |
|---------------------------|-----------------|-----------|-----------|-----------|-----------|-----------|-------|--|
| Цветови модел | Цветови признак | Кнежа 620 | Кнежа 613 | Кнежа 446 | Кнежа 436 | Кнежа 308 | 26А | |
| RGB | R | 58,33 | 80,67 | 70,67 | 78,67 | 81,67 | 65,33 | |
| | G | 71,33 | 53,33 | 55,33 | 63,00 | 74,33 | 84,00 | |
| | B | 94,67 | 92,67 | 91,33 | 89,67 | 95,67 | 95,67 | |
| HSV | H | 62,33 | 54,67 | 69,67 | 80,67 | 91,67 | 77,33 | |
| | S | 95,67 | 98,00 | 90,33 | 94,67 | 96,00 | 93,33 | |
| | V | 58,00 | 80,67 | 70,33 | 78,67 | 81,00 | 66,00 | |
| Lab | L | 68,33 | 58,00 | 51,00 | 63,00 | 64,68 | 83,00 | |
| | a | 72,00 | 61,67 | 70,00 | 62,67 | 84,33 | 64,67 | |
| | b | 92,33 | 97,33 | 91,00 | 94,33 | 96,33 | 94,68 | |
| XYZ | X | 68,00 | 62,00 | 53,67 | 62,00 | 57,67 | 82,33 | |
| | Y | 68,67 | 59,00 | 51,33 | 62,33 | 64,33 | 82,33 | |
| | Z | 94,67 | 88,00 | 89,33 | 84,67 | 95,33 | 96,00 | |
| Ycbcr | Y | 68,33 | 59,33 | 53,00 | 63,00 | 62,33 | 82,33 | |
| | cb | 90,67 | 94,33 | 86,33 | 92,33 | 95,33 | 89,67 | |
| | cr | 94,33 | 91,67 | 87,00 | 84,00 | 95,00 | 83,00 | |
| xyY | x | 94,33 | 96,67 | 89,67 | 93,00 | 95,67 | 94,67 | |
| | y | 92,00 | 96,33 | 89,67 | 94,33 | 95,67 | 94,67 | |

Таблица 2. Брой на построените опорни вектори при класификацията

| Сортове семена | Цветови модел | RGB | | | HSV | | | Lab | | | XYZ | | | Ycbcr | | | xyY | |
|----------------|---------------|-----|-----|----|-----|----|-----|-----|-----|----|-----|-----|----|-------|----|----|-----|----|
| | Средни ст-ти | R | G | B | H | S | V | L | a | b | X | Y | Z | Y | cb | cr | x | y |
| | Кнежа 620 | 196 | 157 | 54 | 205 | 32 | 196 | 180 | 152 | 64 | 168 | 180 | 56 | 182 | 82 | 92 | 40 | 58 |
| | Кнежа 613 | 110 | 211 | 44 | 211 | 26 | 110 | 196 | 195 | 24 | 186 | 197 | 62 | 193 | 38 | 56 | 26 | 30 |
| | Кнежа 446 | 160 | 202 | 60 | 162 | 64 | 162 | 215 | 158 | 66 | 211 | 214 | 70 | 214 | 88 | 94 | 68 | 68 |
| | Кнежа 436 | 132 | 198 | 68 | 130 | 44 | 132 | 190 | 196 | 38 | 194 | 192 | 88 | 193 | 52 | 98 | 56 | 36 |
| | Кнежа 308 | 98 | 138 | 26 | 50 | 26 | 104 | 174 | 98 | 30 | 198 | 173 | 30 | 181 | 32 | 38 | 34 | 30 |
| | 26А | 195 | 100 | 30 | 155 | 38 | 194 | 104 | 182 | 38 | 104 | 104 | 32 | 104 | 58 | 94 | 38 | 34 |

От направения анализ се установява, че с нарастване на броя на формираните опорни вектори, точността при класификация намалява.

Таблица 3. Цветови признаци с точност над 95%

| Сортове семена | Цветови модел | RGB | | | HSV | | | Lab | | | XYZ | | | Ycbcr | | | xyY | |
|----------------|---------------|-----|---|---|-----|---|---|-----|---|---|-----|---|---|-------|----|----|-----|---|
| | Средни ст-ти | R | G | B | H | S | V | L | a | b | X | Y | Z | Y | cb | cr | x | y |
| | Кнежа 620 | | | | | + | | | | | | | | | | | | |
| | Кнежа 613 | | | | | + | | | | | | | | | | | + | + |
| | Кнежа 446 | | | | | | | | | | | | | | | | | |
| | Кнежа 436 | | | | | | | | | | | | | | | | | |
| | Кнежа 308 | | | + | | + | | | + | | | + | | + | + | + | + | + |
| | 26А | | | + | | | | | | | | + | | | | | | |

От таблица 3 се вижда, че за сорт Кнежа 308 има най-много информативни признаци- B(RGB), S(HSV), b(Lab), Z(XYZ), cb и cr(Ycbcr), x и y(xyY). Сортове Кнежа 613 и Кнежа 308 са близки, тъй като информативните им признаци съвпадат в голяма степен. По признака S(HSV) се постига задоволителна точност от над 95% за сортове Кнежа 620, Кнежа 613 и Кнежа 308.

ЗАКЛЮЧЕНИЕ

Оценена е информативността на 17 цветови признака, които представляват средните стойности на цветовете компоненти на 6 цветови модела (RGB, HSV, Lab, YCbCr, XYZ и хуY) по отношение на 6-те изследвани сорта царевични семена.

При класифициране на царевичните семена по използвания метод на опорните вектори се постига задоволителна точност от над 95% за следните сортове: Кнежа 620, Кнежа 613, Кнежа 308 и сорт 26А. Като най – информативен се откроява признака S от HSV модела, който е водещ за 3 сорта семена - Кнежа 620, Кнежа 613 и Кнежа 308. При сортове Кнежа 436 и Кнежа 446 няма признаци, при които да се достигне точност над 95% по SVM метода. За сорт Кнежа 446 най-висока точност – 91,33% се постига по признак B(RGB),а за сорт Кнежа 436 – точност 94,67% по признак S(HSV).

ЛИТЕРАТУРА

[1] Дамянов Ч. Неразрушаващо разпознаване на качеството в системите за автоматично сортиране на хранителни продукти, Академично издателство на УХТ – Пловдив, 2006

[2] Начев В.,Ч. Дамянов, Т. Титова Избор на модел с използване на метода на опорните вектори и оценка на грешките, Международна конференция- АВТОМАТИКА И ИНФОРМАТИКА'07, 3 – 6 Октомври 2007г., България, София

[3] Abdullah M.Z., J. Mohamad-Saleh, A.S. Syahir, B.M.W. Azemi, Discrimination and classification of fresh-cut starfruits (Averrhoacarambola L.) using automated machine vision system, *Journal of Food Engineering Volume 76, Issue 4* (2006) 506-523

[4] Luo X., D.S. Jayas, S.J. Symonst. Identification of damaged kernels in wheat using a colour machine vision system, *Journal of cereal science* 30 (1999), 49—59

[5] Manickavasagan A.G. Sathya, D.S. Jayas, N.D.G. White. Wheat class identification using monochrome images, *Journal of cereal science* 47 (2008), 518—527

[6] Venora G., O Grillo, R. Saccone. Quality assessment of durum wheat storage centers in Sicily: Evaluation of vitreous, starchy and shrunken kernels using an image analysis system, *Journal of cereal science* (2009), 1—12

[7] <http://www.cs.rutgers.edu/~mlittman/courses/ml04/svm.pdf>

[8] <http://www.cogs.susx.ac.uk/users/siwu/Neural%20Networks/svm.ppt>

За контакти:

инж. Елеонора Стефанова Кирилова, Катедра “Автоматика, информационна и управляваща техника”, Русенски Университет “Ангел Кънчев”, тел.: 082/888 684, e-mail : ekirilova@ru.acad.bg

Изследванията са подкрепени по договор № **BG051PO001-3.3.04/28**, „Подкрепа за развитие на научните кадри в областта на инженерните научни изследвания и иновациите”. **Проектът се осъществява с финансовата подкрепа на Оперативна програма „Развитие на човешките ресурси” 2007-2013, съфинансирана от Европейския социален фонд на Европейския съюз“.**

Докладът е рецензиран.