

## КОМБИНИРАНЕ НА AR МОДЕЛИ И SVM КЛАСИФИКАТОРИ ПРИ РАЗПОЗНАВАНЕ НА ЗАБОЛЯВАНЕТО ФУЗАРИОЗА ПО ЦАРЕВИЧНИ СЕМЕНА

**П. Даскалов, В. Манчева, Цв. Драганова**

*Катедра Автоматика, информационна и управляваща техника, Русенски Университет  
ул. Судентска 8, Русе п.к.7017, e-mail: [daskalov@uni-ruse.bg](mailto:daskalov@uni-ruse.bg), тел. (082) 888 281*

## COMBINATION OF AR MODELS AND SVM CLASSIFIERS FOR RECOGNITION OF FUSARIUM DISEASED CORN SEEDS

**P. Daskalov, V. Mancheva, Ts. Draganova**

*Department of Automatics, information and control engineering, University of Ruse,  
8 Studentska Street, 7017 Ruse, BULGARIA; e-mail: [daskalov@uni-ruse.bg](mailto:daskalov@uni-ruse.bg), tel. (082) 888 281*

### ABSTRACT

An approach to identify healthy and Fusarium infected corn seeds by combining AR models and SVM classifiers is presented in the paper. Coefficients of parametric AR models are used as signs to identify Fusarium disease corn seeds. The implementation of SVM method is made for two types kernel functions - linear and radial-base function (RBF) and for two classifier types. Corn seeds of seven varieties were tested with the proposed approach. In varieties with overlapping distributions of AR model coefficients, the accuracy of classification was increased to 80% - for Kneja 308 variety and 87.5 percent - for Kneja 613 and Kneja 620 varieties. Despite improved results with SVM classifiers, the accuracy of classification for some varieties still does not meet the Bulgarian standard.

*Key words: corn seeds, spectral characteristics, Fusarium, AR models, SVM*

### УВОД

Болестите по царевицата са широко разпространени във всички райони на нейното отглеждане, като най-общо могат да бъдат класифицирани в две групи – инфекциозни и неинфекциозни. Втората група болести най-често са в следствие от неблагоприятните външни условия и хранителен режим. По-голямо значение за производството имат инфекциозните болести – и по-конкретно болестта Фузариоза. Експресни методи за нейното диагностициране са чрез анализ на визуални изображения, чрез анализ на спектрални характеристики и чрез анализ на хиперспектрални изображения.

Съвременни и актуални методи за диагностициране на заболяването Фузариоза са хиперспектралните изображения [3,7]. Те са бързо и обективно средство за идентифициране на инфектирани области, но е скъпо струващ метод за работа.

При използването на визуални методи се анализира само повърхността на обекта, или неговата видима част. Направени са изследвания за диагностициране на болестта Фузариум по царевични семена чрез анализ на цифрови изображения, които дават добри резултати [2,8].

Оценката по външни признаци само не е достатъчна поради факта, че те не винаги се проявяват [2]. Разработени са алгоритми за разпознаване на заболели от Фузариоза царевични семена чрез анализ на спектралните им характеристики на базата на тяхното описание с линейни дискретни модели [2,5]. В [5] е предложен подход за разпознаване на здрави и заразени с розова Фузариоза царевични семена посредством спектрални характеристики на интензивност в близката инфрачервена област. Той е базиран на анализ на коефициентите при  $10^{-\text{m}}$  ред на линейни дискретни параметрични модели от типа авторегресия (AR). Разпознаването на семената се основава на критерий, който определя граничната стойност  $A_{sp}$  между групата на здравите и групата на заразените семена. За определянето ѝ е избран един от десетте коефициента, при който е получено максимално

разстояние  $\Delta A$  между двете групи. Получените резултати показват, че метода, базиран на AR модел и използването на първия коефициент ( $A_1$ ) като признак за разпознаване, дава добри резултати при четири (26А, Кнежа 436, ХМ87/136 и Русе 424) от седемте изследвани сорта. За останалите три сорта (Кнежа 308, Кнежа 613 и Кнежа 620) процента на разпознаване се получава доста по-нисък – между 51,25 и 78,75%.

Стремежът на съвременните изследвания е да се създадат конструктивни методи, които да са по-ефективни от тези произтичащи от прякото прилагане на стандартните подходи, приети в теорията. Един от тези методи е метода на опорните вектори (Support Vector Machines), който все повече се налага при разпознаването на отелни класове в различни области [1]. В [6] използват коефициентите, получени от AR модел, като признаци при класификация със SVM метода за разпознаване на четири състояния на човешката дейност – бягане, неподвижност, скачане и ходене.

Целта на статията е да се изследва възможността за разпознаване на здрави и заразени от Фузариоза царевични семена при комбинация на линеен дискретен параметричен модел от типа авторегресия (AR модел) и SVM класификатор.

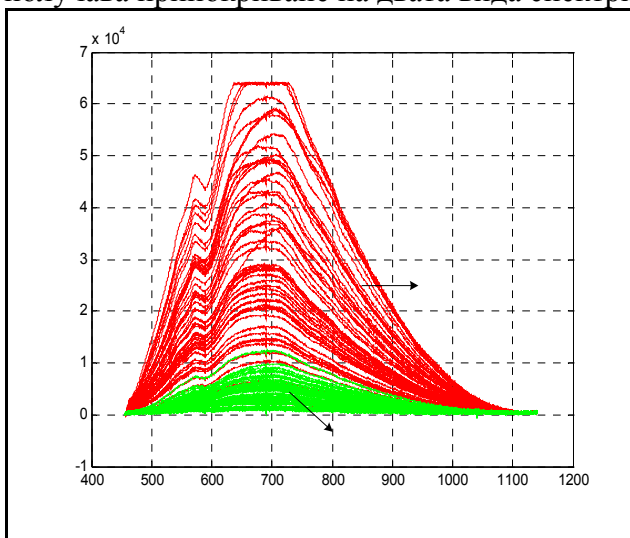
## МАТЕРИАЛ И МЕТОДИ

### *Обект на изследването*

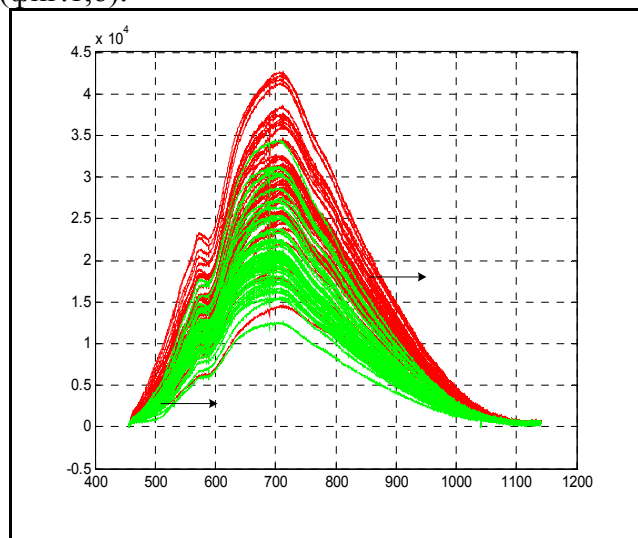
Обект на изследването е заболяването розова Фузариоза (*Fusarium Moniliforme*) и проявата му по царевичните семена. Изследвани са извадки от седем сорта семена - Кнежа 308, Кнежа 436, Кнежа 613, Кнежа 620, 26А, Русе 424 и ХМ87/136. Те са оценени от експерт и са предоставени от Института по царевица – гр. Кнежа.

### *Получаване на спектралните характеристики*

Спектралните характеристики са получени посредством спектрофотометър на фирмата Ocean Optics във видимата и близката инфрачервена област в спектралния диапазон от 456 до 1140,5 nm. За всеки от сортовете са снети спектралните характеристики на интензивност (Intensity) на 50 здрави и 50 заразени семена за двете им страни – гръб (по-долу означени със „г“) и зародиш (по-долу означени със „з“). На фиг.1 са представени спектралните характеристики за два сорта. Установено е, че при някои сортове има съществена разлика между амплитудата на здравите и на заразените семена (фиг.1,а), а при други сортове се получава припокриване на двата вида спектри (фиг.1,б).



а) сорт 26А



б) сорт Кнежа 620

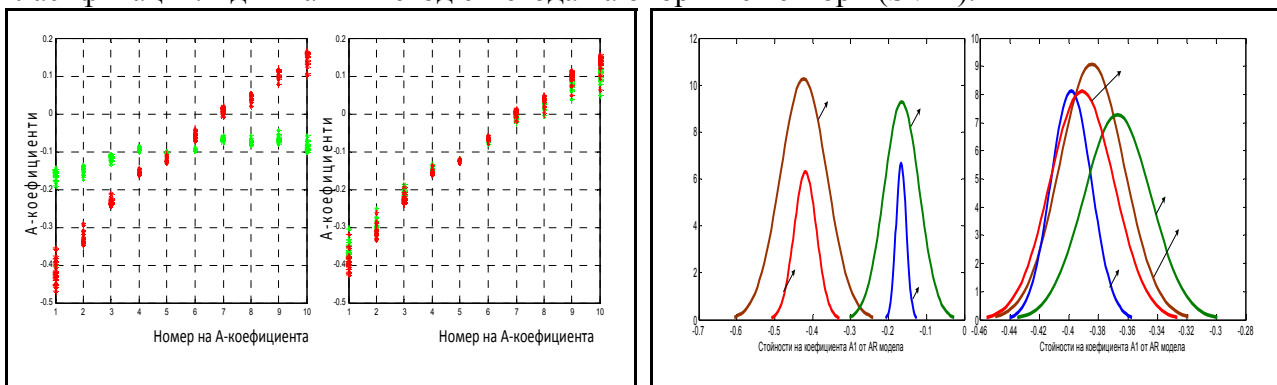
Фиг. 1 Спектрални характеристики на интензивност за 50 здрави и 50 заразени царевични семена

Получените характеристики са много близки по форма и не могат да се определят области от спектъра, които да не се влияят от вида на семената. Съответно не може да се получи и директно разпознаване на царевичните семена. Поради това е необходимо да се извърши допълнителна обработка на данните.

#### **Обработка на спектралните данни**

От предварителни изследвания в [5] при обработка на спектралните данни с линейни дискретни AR модели е получено разпределението на коефициентите при  $10^{\text{ти}}$  ред на моделите. Установено е, че за сортовете, при които има съществена разлика между амплитудата на здравите и на заразените семена, разпределението на коефициентите е съществено значимо (фиг.2,а). Докато за сортовете, при които има припокриване на спектрите от двете групи семена, разпределението на коефициентите е незначително (фиг.2,б). Кривите на разпределение за първия коефициент ( $A_1$ ) от фиг.3 потвърждават хипотезата, че разпределението на коефициентите от AR моделите оказва съществено значение при разпознаването на семената.

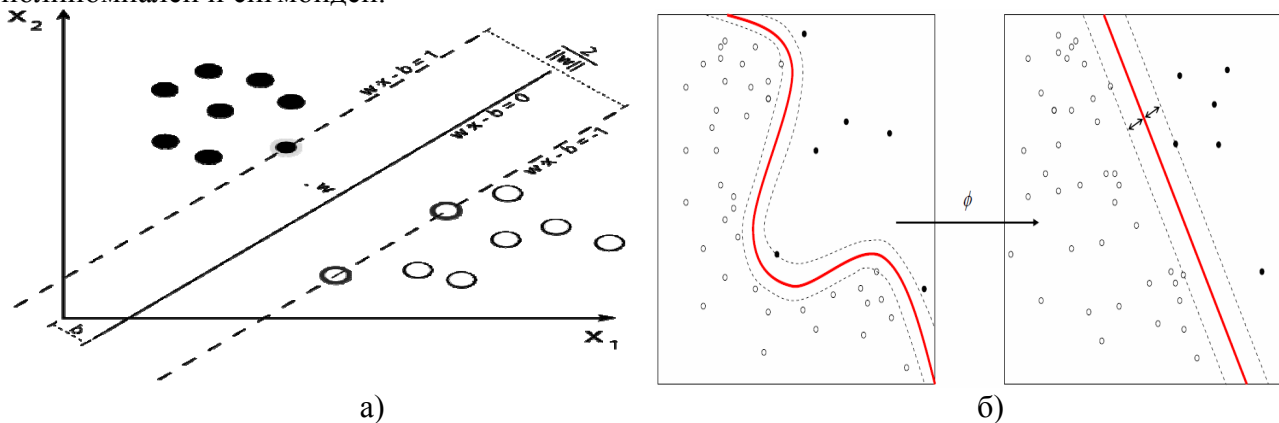
Получените резултати показват, че метода, базиран на AR модел и използването на първия коефициент ( $A_1$ ) като признак за разпознаване, дава добри резултати за сортовете, при които разпределението на коефициентите е от вида на фиг.2,а. Значително по-ниски резултати от разпознаването се получават за сортовете, при които няма ясна разделимост на коефициентите (фиг.2,б). Поради тази причина се налага за тези сортове да се търси метод за разпознаване, който да може да преобразува данните за коефициентите във вид, удобен за класификация. Един такъв метод е метода на опорните вектори (SVM).



SVM е метод за класификация чрез нелинейно преобразуване на оригиналните данни в друго пространство с по-висока размерност, където обектите са линейно разделими (фиг.4,б).

Първоначално методът на опорните вектори е използван като метод за построяване на оптимален линеен класификатор. SVM алгоритъмът създава модел за предсказване към кой от двата класа трябва да бъде отнесена новата проба. Данните за образците са представени геометрично посредством точки в пространството на признаците (фиг.4). При линейно разделимите области (фиг.4,а) SVM моделът намира максимално възможната междина между двете области и строи разделяща равнина, максимално отдалечена от границите на двете области. За всеки нов образец SVM алгоритъмът предсказва принадлежността към една от двете категории в зависимост от това, от коя страна на граничната повърхност попада точката, представляваща образаца. По-късно е предложен обобщен вариант [4], даващ възможност за използване на ядрени (кернал) функции и приложение на метода и за нелинейни задачи (фиг.4,б). В зависимост от вида на използваните кернъл функции се

конструират различни класификатори – линейен, радиална базисна функция (RBF), полиномиален и сигмоиден.



Фиг.4 Оптимален хиперплан при SVM метода – а) при линейно разделими области; б) при нелинейно разделими области

### РЕЗУЛТАТИ И ОБСЪЖДАНЕ

За изпълнението на метода на опорните вектори е използвана библиотека Machine Learning на програмния пакет STATISTICA 8. Реализацията на метода е направена за два вида кернъл функции – линейна и радиално-базова функция (RBF) и два типа класификатори – Type 1 и Type 2. Разгледан е случай за синтез на класификатор за два класа (зdravi и заразени семена). Като входни вектори са използвани веднъж само коефициента  $A_1$  и втори път – всичките десет коефициента ( $A_1 \div A_{10}$ ), получени при  $10^{-th}$  ред на AR-модела от 50 здрави и 50 заразени царевични семена за всеки от седемте сорта. Размера на обучаващата извадка е 60 % от обема на общата извадка, а на тестване - 40%. Резултатите, получени от класификацията с данните за коефициентите от тестовата извадка са представени в табл.1.

Резултати от класификация със SVM и гранична стойност ( $A_{гр}$ ) на параметричен AR модел Таблица 1

класификатор	тип на класификатора	коэф. от AR модела	ТОЧНОСТ ОТ ТЕСТОВАТА ИЗВАДКА, %													
			Кнежа 308		Кнежа 436		Кнежа 613		Кнежа 620		26А		Русе 424		ХМ87/136	
			г	з	г	з	г	з	г	з	г	з	г	з	г	з
Support Vector Machine (SVM)	линеен Type 1	$A_1$	40	40	90	87,5	40	70	77,5	70	95	92,5	95	97,5	100	100
	линеен Type 2		42,5	55	90	87,5	62,5	57,5	72,5	45	95	97,5	95	97,5	100	100
	RBF Type 1		40	40	90	87,5	40	72,5	80	70	95	95	95	97,5	100	100
	RBF Type 2		45	35	90	90	37,5	40	75	77,5	95	97,5	95	97,5	100	100
	линеен Type 1	$A_1 \div A_{10}$	62,5	40	90	90	60	75	87,5	75	95	92,5	97,5	97,5	100	100
	линеен Type 2		80	45	90	87,5	65	65	87,5	85	95	92,5	95	97,5	100	100
	RBF Type 1		40	40	90	87,5	40	72,5	80	75	95	95	95	97,5	100	100
	RBF Type 2		60	70	90	87,5	87,5	85	80	75	95	95	95	97,5	100	100
$A_{гр}$	AR модел	$A_1$	60	45	97,5	97,5	40	62,5	87,5	70	100	95	97,5	100	100	100

Направен е анализ на базата на сравняването на резултатите от SVM класификаторите и използването на граничната стойност  $A_{гр}$  при AR моделите за разделянето на зравите и заразените семена.

Резултатите от табл.1 показват, че за няколко сорта (Кнежа 436, 26А, Русе 424 и ХМ87/136) процента от разпознаване се запазва един и същ както при използването на един признак (коефициента  $A_1$ ), така и при използването на всички признаци (десетте коефициента).

Като цяло – използването на десетте коефициента ( $A_1 \div A_{10}$ ) като признаци за разпознаване дават по-високи резултати и при двата вида кернъл функции, отколкото като се използва само първия коефициент ( $A_1$ ).

За сортове Кнежа 308, Кнежа 613 и Кнежа 620 използването на SVM класификатор от втория тип (линеен Type 2 и RBF Type 2) дава по-добри резултати от тези, получени с параметричния AR модел на базата на граничната стойност  $A_{гр}$  между групата на зравите и групата на заразените семена. Точността от класификацията се е повишила с 20% за гръбната страна и 25% за страната на зародиш при сорт Кнежа 308; 47,5% - за гръбна страна и 22,5% - за страна на зародиш при сорт Кнежа 613; 15% за страна на зародиш при сорт Кнежа 620. За гръбната страна на сорт Кнежа 620 – процента от разпознаването остава същия (87,5%).

За сортове Кнежа 436, 26А и Русе 424 има малък спад в разпознаването при използването на двата типа кернъл функции в сравнение с параметричния AR модел. Но като цяло процентът на разпознаване си остава висок – в границите от 90÷97,5%. Дори при сорт 26А се наблюдава леко покачване на процента (от 95 на 97,5%) с класификаторите Type 2.

С най-висок процент на разпознаване (100%) на здрави и заразни царевични семена и при SVM класификацията, и при параметричните AR модели е сорт ХМ87/136.

### ИЗВОДИ

При комбиниране на AR модели и SVM класификатор, използването на десетте коефициента от AR моделите за всеки сорт дава по-висок процент на разпознаване, отколкото ако се използва само първия коефициент.

При сортове с припокриващо се разпределение на коефициентите (Кнежа 308, Кнежа 613 и Кнежа 620) приложението на подхода с комбинирано SVM и AR модели дава значително по-добри резултати. Докато при останалите сортове (Кнежа 436, 26А и Русе 424) се получава близка точност или тя се запазва (ХМ87/136).

Независимо от подобрените резултати със SVM класификаторите, точността на класификация за някои сортове все още не отговаря на изискванията по стандарт. Поради това се налагат бъдещи изследвания за повишаване на точността.

### ЛИТЕРАТУРА

1. Дамянов, Ч., Неразрушаващо разпознаване на качеството в системите за автоматично сортиране на хранителни продукти, Академично издателство на УХТ – Пловдив, 2006
2. Драганова, Цв., Изследване диагностицирането на заболяването фузариоза (*Fusarium spp.*) по царевични семена чрез използване на цифрови изображения и спектрални характеристики, Докторска дисертация, 2006
3. Bauriegel, E., 2011. Early detection of Fusarium infection in wheat using hyper-spectral imaging, Computers and Electronics in Agriculture, Vol.75(2), p.304-312
4. Boser, B., Guyon, I., Vapnik, V., 1992. A training algorithm for optimal margin classifiers, Fifth Annual Workshop on Computational Learning Theory, ACM Press, Pittsburgh
5. Daskalov, P., V. Mancheva, Ts. Draganova, R. Tsonev, 2010. An approach for *Fusarium* diseased corn kernels recognition using linear discrete models, Agricultural Science and Technology, Vol.2, Number 2, p.90-95
6. He Z., Activity recognition from acceleration data using AR model representation and SVM, Proceedings of the Seventh International Conference on Machine Learning and Cybernetics, Kunming, 12-15 July 2008

7. Williams, P., M. Manley, 2010. Indirect detection of *Fusarium verticillioides* in maize (*Zea mays* L.) kernels by near infrared hyperspectral imaging, *Journal of Near Infrared Spectroscopy*, 18, p.49–58
8. Wiwart, M., I.Koczowska, A.Borusiewicz, 2001. Estimation of *Fusarium* head blight of triticall using digital image analysis of grain, *CAIP, LNCS 2124*, p.563-569