بیستمین کنفرانس اپتیک و فوتونیک ایران
و ششمین کنفرانس مهندسی و فناوری فوتونیک ایران
۸ تا ۱۰ بهمن ماه۱۳۹۲ - دانشگاه صنعتی شیراز

ارزیابی عملکرد شبکه عصبی پرسپترون درطبقه بندی طیف های رامان حاصل از نمونه های زیستی: تشخیص سم آفلاتوکسین درپسته

رضا محمدی گل[۱]، محمدهادی خوش تقاضا[۱]، رسول ملک فر[۲]، منصوره میرابوالفتحی[۳]، علی محمد نیکبخت[۴] و هادی دارابی[۵]
۱- گروه مکانیک ماشینهای کشاورزی، دانشگاه تربیت مدرس، ص پ ۱۱۱-۱۴۱۵
۲ - بخش فیزیک دانشگاه تربیت مدرس
۳- موسسه تحقیقات گیاهپزشکی کشور ۴- گروه مکانیک ماشینهای کشاورزی دانشگاه ارومیه
۵- دانشگاه تربیت مدرس

چکیده- دراین پژوهش امکان تشخیص آفلاتوکسین در پسته با تکنیک طیف‌سنجی رامان و به کارگیری شبکه عصبی پرسپترون بررسی شده است. بعد از طیف‌برداری، با توجه به نتایج آنالیز مولفه های اصلی(PCA) پیش‌پردازش مناسب، اعمال مشتق دوم روی داده های طیفی انتخاب شد و به دنبال آن برای کاهش ابعاد داده‌های طیفی استخراج مولفه‌های اصلی نیز صورت پذیرفت. برای طبقه بندی نمونه ها شبکه پرسپترون تک لایه(۸ نرون در لایه پنهان) و دو لایه (۲ و۴ نرون به ترتیب در لایه اول و دوم) با قانون یادگیری پس انتشار خطا مورد مقایسه قرار گرفتند. میانگین دقت طبقه بندی شبکه تک لایه ۹۲ درصد و شبکه دو لایه ۸۲ درصد به دست آمد. در هر دو شبکه نمونه های سالم با ۱۰۰ درصد دقت از هر دو سطح نمونه های آلوده جدا شده و لذا عملکرد شبکه های عصبی در جداسازی نمونه های سالم از نمونه های دیگر موفقیت آمیز ارزیابی شد.

کلید واژه- آفلاتوکسین، آنالیز مولفه اصلی، طیف سنجی، پسته، شبکه عصبی، طیف سنجی رامان

Evaluation of perseptron artificial neural networks(ANN) performance in biological samples Raman spectra classification; Pistachio aflatoxin detection

Reza Mohammadigol[1], Mohammad Hadi Khoshtaghaza[1], Rasoul Malekfar[2], Mansooreh Mirabolfathi[3], Ali Mohammad Nikbakht[4] and Hadi Darabi[5]

1-TMU Machinery mechanic engineering department
2- TMU physics department 3- IRIPP
4-Uromia University Assistant professor, biosystem mechanic engineering department
5- TMU

Abstract- The objective of this study was to detect pistachio aflatoxin contamination applying Raman spectroscopy technique and artificial neural networks. After spectra acquisition considering to principal components analysis(PCA) results, Second Derivative preprocessing method was selected and then principal components(PCs) were extracted to reduce the data dimensions. To classify samples, two kind of Multilayer perseptron feed forward back propagation topologies (including 8 neurons in hidden layer as first one and including 2 and 4 neurons in the first and second hidden layers respectively as a second one) were used. On average our classifiers performed with 92(one layer ANN) and 82(two layers ANN) percent accuracy . Both classifiers performance, for no contaminated samples recognition (as they achieved to 100 percent accuracy) was successful.

Keywords: Aflatoxin, ANN, PCA, Pistachio, Raman Spectroscopy

این مقاله در صورتی دارای اعتبار است که در سایت www.opsi.ir قابل دسترسی باشد.

۱- مقدمه

در سال های اخیر تحقیقات وسیعی در حوزه های گوناگون علم پزشکی و داروسازی به منظور به کارگیری طیف سنجی رامان و بخصوص روش‌های تقویت طیف های رامان انجام شده است. تمرکز این تحقیق‌ها روی شناسایی و تشخیص سریع و زودهنگام بافت های سرطانی ، باکتری ها، قارچ ها و مواد شیمیایی ـ دارویی بوده است. در تحقیق‌های مذکور معمولاً از مدل های طبقه بندی یا رگرسیونی چند متغیره برای نگاشت و تفسیر اطلاعات طیفی استفاده شده است[۱].

آفلاتوکسین به عنوان عضو خانواده مایکوتوکسین‌ها، از متابولیت‌های ثانویه کپک‌هایی چون آسپرژیلوس فلاوس و آ. پارازیتیکوس می‌باشد. آفلاتوکسین‌ها به دلیل فراوانی قابل ملاحظه آنها در طبیعت و خاصیت سمی و سرطان زا بودن، به عنوان سردسته مایکوتوکسین ها شناخته شده اند. تا کنون چندین نوع آفلاتوکسین شناسایی شده است که از بین آنها آفلاتوکسین های B1، B2، G1 و G2 دارای بیشترین اهمیت هستند [۲].

مهمترین مزیت استفاده از روش طیف‌سنجی رامان، در تحلیل زیستی غنای اطلاعات موجود در هر طیف می باشد. طیف رامان یک سلول را میتوان بعنوان "اثر انگشت" ترکیبات بیوشیمیایی آن به حساب آورد. بنابراین اگر یک عامل سمی باعث تغییرات بیوشیمیایی شود طیف رامان حاصله دست‌خوش تغییر خواهد شد [۳].

گرایش جامعه جهانی به سمت تولید و عرضه مواد غذایی سالم و بی‌خطر باعث شده است صنعت پسته همواره با چالش تکنولوژی های جدید به منظور سازش با آگاهی و تقاضای رو به افزایش مصرف کننده در خصوص تضمین سلامت محصول مواجه باشد. رویکرد رو به رشد در اغلب محصولات کشاورزی، افزایش استفاده از روش‌های سریع، قابل اطمینان و غیرمخرب ارزیابی کیفی می‌باشد. با توجه به جایگاه ممتاز پسته در صادرات غیر نفتی ایران و به منظور جلوگیری از به مخاطره افتادن صادرات این محصول با ارزش و همچنین برای حضور در تجارت جهانی و رقابت با دیگر کشورها، اولویت تحقیقات بر روی روش‌های نوین تشخیص و سنجش آلودگی پسته به آفلاتوکسین انکار ناپذیر بوده و جایگاه ویژه ای دارد.

۲- مواد و روش ها

سامانه طیف سنج رامان مدل SENTERRA (2009) از نوع پاشنده ساخت شرکت BRUKER آلمان برای به دست آوردن طیف های رامان مورد استفاده قرار گرفت. طول موج لیزر این سامانه ۷۸۵ نانومتر و توان حداکثرآن ۱۰۰میلی وات می باشد. در تحقیقات زیادی در رابطه با طیف سنجی رامان نمونه های زیستی از لیزر با طول موج ۷۸۵ نانومتر به عنوان منبع تهییج استفاده شده است[۴].

نمونه های پسته به صورت خمیری در ۳ سطح زیر تهیه شدند:

c- بدون آلودگی

a- آلودگی با مجموع آفلاتوکسین های B1، B2، G1 و G2 ۲۰ نانوگرم در گرم(ppb)(هر کدام ۵ppb)

A- آلودگی با مجموع آفلاتوکسین های B1، B2، G1 و G2 ۱۰۰ نانوگرم در گرم(ppb)(هر کدام ۲۵ppb)

۲-۱- پیش پردازش طیف ها(Preprocessing)

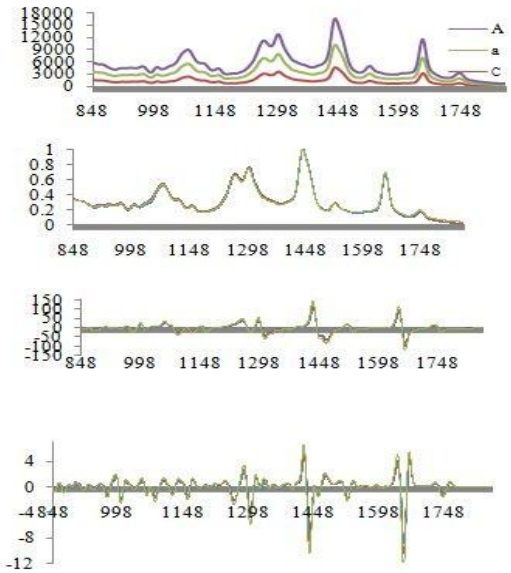
طیف های رامان برداشت شده (c:۱۵، a:۱۷و A:۱۸عدد و مجموعا ۵۰ طیف) علاوه بر اطلاعات نمونه حامل اطلاعات ناخواسته از قبیل نوفه و نور پس زمینه می باشد لذا برای رسیدن به مدل واسنجی دقیق و قابل اعتماد پیش پردازش داده های طیفی قبل از مدل سازی ضرورت دارد. روش های پیش پردازش بسیار متنوع هستند که معمولا هر یک از آنها برای منظور خاصی ابداع شده‌اند و با توجه به شرایط متفاوت نمونه ها و روش ها و ابزار آزمایشگاهی ارائه یک روش پیش پردازش مشخص امکان ندارد. از این رو معمولا در تحقیقات طیف سنجی روش های رایج برتر مورد ارزیابی قرار گرفته و بهترین روش مورد استفاده قرار می‌گیرد. در این تحقیق تاثیر روش های زیر در تفکیک و تمایز طیف های اخذ شده مورد بررسی قرار گرفتند.

هموارسازی ساویتزکی- گولای(SG) : در این روش، الگوریتمSG بر هر بخش متوالی طیف، یک چند جمله ای برازش و مقدارهای اصلی طیف را با مقدارهای برازش شده (تغییرات منظم تر) جایگزین می کند.

هنجار سازی: انجام هنجارسازی طیفی برای حذف تغییرات سیستماتیک در میزان فوتون های پراکنده شده که به صورت محور عمودی در طیف ها ظاهر می شود، ضروری است .

مشتق اول و مشتق دوم: با هر بار مشتق گیری از طیف (افزایش درجه مشتق)، پیک های تیز نسبت به پیک های پهن بیشتر می شوند. به همین دلیل، مشتق دوم نسبت به مشتق اول قدرت تفکیکی بهتری معمولا ارائه می دهد[7]. با این حال مشتق گیری سبب کاهش نسبت سیگنال به نویز(به ویژه در مشتق های بالاتر) می شود که مطلوب نیست.

در شکل ۱ وضعیت طیف ها در عبور از مراحل مختلف پیش پردازشی شامل حذف خط زمینه و هموارسازی، هنجار سازی، اعمال مشتق اول و دوم روی طیف های متوسط هر گروه ($۸۴۸-۱۸۶۰cm^{-1}$) نمونه مشاهده می شود.



شکل **۱**- عملکرد روش‌های پیش پردازش اعمال شده روی متوسط طیف های هرگروه نمونه در نوار $۸۴۸-۱۸۶۰\ cm^{-1}$، به ترتیب از بالا به پایین: حذف خط زمینه و هموارسازی، هنجار سازی، اعمال مشتق اول، اعمال مشتق دوم

۲-۲- آنالیز مولفه های اصلی(PCA)

بطور کلی تعداد مشاهدات (طیف ها) برای آموزش موفق یک Machine-Learning Classifier با تعداد داده ها(متغیرهای طیفی) توسط رابطه نمایی افزایش می یابد [۵] لذا دراغلب تحقیقات مرتبط با طیف سنجی به دلیل اینکه تعداد متغیرهای طیفی بسیار زیاد است، قبل از فرآیند مدل سازی یا طبقه بندی، استخراج مولفه های اصلی(توسط الگوریتم کاهش داده) با استفاده از روشPCA انجام شده است. تحلیل مؤلفه‌های اصلی یکی از انواع روش‌های تحلیل داده‌های چند متغیره است که هدف اصلی آن تقلیل بعد مسئله مورد مطالعه است. با استفاده از تحلیل مؤلفه‌های اصلی می‌توان تعداد زیادی متغیر توضیحی (متغیر مستقل) را با تعداد محدودی متغیر توضیحی جدید که با یکدیگر وابستگی خطی ندارند و مؤلفه‌های اصلی نامیده می‌شوند، جایگزین نمود. با توجه به حجم بالای متغیر های طیف رامان(عدد موج ها) از این روش برای کاهش متغیرها استفاده شد.

۲-۳- مدل‌سازی

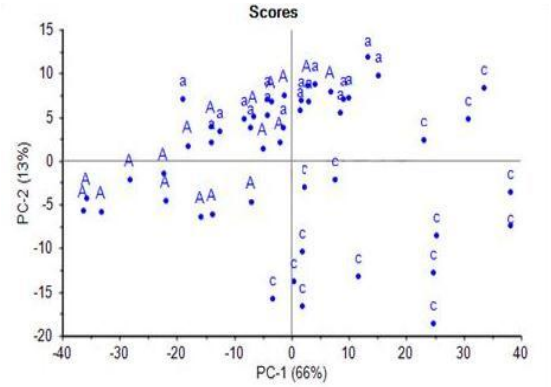
یک شبکه عصبی این قابلیت رادارد که تمایز ها و تفاوت های مرموز را که در شکل کلی و وضعیت نوار های مختلف رامان وجود دارد و به راحتی توسط بشر درک نمی شوند، را فهمیده وبر مبنای آن آموزش ببیند[۵]. در این تحقیق از شبکه پرسپترون یک و دولایه‌ای(MLP) و قانون یادگیری پس انتشار خطا برای طبقه بندی طیف ها استفاده شد[۶]. در این پژوهش نرم افزار MATLAB در مدل سازی شبکه عصبی و نرم افزار Unscrambler برای پیش پردازش طیف ها و انجام PCA مورد استفاده قرار گرفتند.

۳- نتایج

پس از بررسی شکل طیف ها استنباط می شود که در بازه $۸۴۸ - ۱۸۶۰cm^{-1}$ شدت(Raman Intensity) نمونه های گروه A بیشتر از گروه a و آن هم بیشتر از گروه c است، بنابراین مطلب و همچنین مقایسه آنالیز مولفه های اصلی کل محدوده طیف، نوار مذکور در مراحل بعدی مورد تحلیل و بررسی قرار گرفت.

پس از عمل PCA خوشه ای قرار گرفتن طیف ها در مختصات مولفه های اصلی شاخص مناسبی از عملکرد مدل در طبقه بندی و جداسازی است[۵]. در شکل ۲ پراکندگی طیف ها در مختصات مولفه های اصلی اول و دوم مشاهده می شود، همانگونه که مشهود است روش پیش پردازش مشتق دوم به گونه ای طیف ها را پراکنده نموده است که بیشتر از هم قابل تمایز به نظر می-رسند(خوشه‌ای) و همان‌طور که قبلا نیز اشاره شد در منابع این تمایزپذیری جمعیت داده ها شرط موفقیت مدل های واسنجی و از جمله شبکه های عصبی

پرسپترون در طبقه بندی نمونه ها قلمداد شده است[۶]. لذا روش مذکور به عنوان پیش پردازش داده های طیفی در نظر گرفته شد و پس از اعمال این روش بر کل داده ها تعداد ۷ مولفه اصلی اول به عنوان ورودی های شبکه عصبی(تک لایه و دو لایه) در نظر گرفته شد.



شکل ۲- نمایش اثر اعمال مشتق دوم بر پراکندگی طیف ها در مختصات مولفه اصلی اول و دوم(آنالیزPCA)

برنامه تدوین شده بر اساس معیار خطای کمترتعداد نرون بهینه لایه پنهان برای شبکه تک لایه را ۸ و برای شبکه دوم ۲ و۴ نرون به ترتیب در لایه های اول و دوم محاسبه نمود. ماتریس اغتشاش با فراهم نمودن اینکه کلاس نمونه چقدر به اشتباه طبقه بندی شده، دقت (تعداد نمونه های درست پیش بینی شده بر تعداد کل نمونه ها) الگوریتم های طبقه بندی را نمایش می دهد و خلاصه عملکرد طبقه بندی با شبکه های عصبی در قالب این ماتریس ارائه می شود. در جدول های ۱ و ۲ ماتریس اغتشاش عملکرد شبکه عصبی یک لایه(با ۸ نرون در لایه پنهان) و شبکه عصبی دو لایه (با ۲ نرون در لایه اول و ۴ نرون در لایه دوم)در طبقه بندی نمونه ها آمده است.

جدول **۱**- ماتریس اغتشاش خلاصه عملکرد شبکه عصبی یک لایه در طبقه بندی نمونه ها

پیش بینی شده / واقعی	(c)	(a)	(Á)	درصد صحت جداسازی
آلودگی ۰ (c)	۱۵	۰	۰	۱۰۰
آلودگی۲۰ ppb(a)	۰	۱۴	۳	۸۲/۴
آلودگی۱۰۰ ppb(Á)	۰	۱	۱۷	۹۴/۴

جدول**۲**- ماتریس اغتشاش خلاصه عملکرد شبکه عصبی دو لایه در طبقه بندی نمونه ها

پیش بینی شده / واقعی	(c)	(a)	(Á)	درصد صحت جداسازی
آلودگی ۰ (c)	۱۵	۰	۰	۱۰۰
آلودگی۲۰ ppb(a)	۰	۱۳	۴	۷۶/۵
آلودگی۱۰۰ ppb(Á)	۰	۵	۱۳	۷۲/۵

۴- نتیجه گیری

میانگین دقت طبقه بندی شبکه تک لایه مورد استفاده (با ۸ نرون در لایه پنهان) ۹۲ درصد و شبکه دو لایه (با ۲ و ۴ نرون به ترتیب در لایه پنهان اول و دوم) ۸۲ درصد به دست آمد. نکته قابل توجه این است که در هر دو شبکه نمونه های سالم با ۱۰۰ درصد دقت از نمونه های آلوده(مجموع دو سطح ppb ۱۰۰ و ۲۰) جدا شده-اند و بنابراین عملکرد شبکه ها در جداسازی نمونه های سالم از آلوده‌ها کاملاً موفقیت آمیز بوده، اما بیشترین دقت جداسازی نمونه های دو سطح ppb ۱۰۰ و ppb ۲۰ به ترتیب ۹۴/۴ و۸۲/۴ (شبکه یک لایه) به دست آمد.

مراجع

[1] Huang, Z., Teh, S.K., Zheng, W., Lin, K., Ho, K.Y., Teh, M. and Yeoh, K.G. 2010. *In vivo detection of epithelial neoplasia in the stomach using image-guided Raman endoscopy*. *Biosensors and Bioelectronics*, **26**(2): 383-389.

[2] Arrus, K., Blank, G., Abramson, D., Clear, R. and Holley, R. 2005. Aflatoxin *production by Aspergillus flavus in Brazil nuts*. Journal of Stored Products Research, 41(5): 513-527.

[3] Notingher, I. 2007. *Raman spectroscopy cell-based biosensors*. Sensors, 7(8): 1343-1358.

[4] Dennis, A. 2013b. *Automatic Baseline Correction* (Thechnical note). from PerkinElmer company www.perkinelmer.com

[5] Ishikawa, S.T. and Gulick, V.C. 2013. *An automated mineral classifier using Raman spectra*. Computers & Geosciences, 54: 259-268.

[6] Hagan, M.T., Demuth, H.B. and Beale, M.H. 1996. *Neural network design*: Pws Pub. Boston.

[7] Nicolai, B. M., K. Beullens, E. Bobelyn, A. Peirs, W. Saeys, K. I. Theron and J. Lammertyn. 2007. Nondestructive measurement of fruit and vegetable quality by means of NIR spectroscopy: *A review*. *Postharvest Biology and Technology* 46: 99-118.