



Shahid Bahonar
University of Kerman

Agricultural Biotechnology Journal

p-ISSN 2228-6705

e-ISSN 2228-6500



Iranian Biotechnology
Society

Identification of Superior Biochemical Markers Linked to Salinity Tolerance in Wheat Using Data Mining

Zahra Zinati

Assistant Professor, College of Agriculture and Natural Resources of Darab, Shiraz, Iran.

Email: zahra.zinati@hotmail.com

* **Abbas Alemzadeh**

Associate Professor, Crop Production and Plant Breeding Department, School of Agriculture,

Shiraz University, Shiraz, Iran. Tel: +987132286134, Email: alemzadeh@shirazu.ac.ir

Abstract

Objective

The recognition of pathways involved in salt tolerance mechanisms is one of the interesting topics in plant sciences and the new data-mining techniques provide a new insight for researchers. In this research, various attribute weighting and decision tree (DT) algorithms were executed to discover biochemical markers linked to salinity tolerance.

Materials and methods

In this regard, to assess some biochemical markers such as protein, superoxide dismutase, peroxidase, catalase, ascorbate peroxidase, proline, sodium and potassium in wheat and its wild relative *Aegilops crassa*, a factorial experiment was conducted in a completely randomized design with three replications. The factors in the study were genotypes (Arg (salt tolerance) and Alamout (salt sensitive), and a wild relative (*Ae. crassa*)) and salinity (sodium chloride 0 mM and 150 mM).

Results

According to these approaches, proline, superoxide dismutase and ascorbate peroxidase can be used as biochemical markers to screen wheat genotypes for salt tolerance. Feature selection and decision tree results showed that hierarchy combination of proline, superoxide dismutase and ascorbate peroxidase can be used as biochemical markers for selection of salinity tolerant wheat. The DT Parallel Gain Ratio model when run on the

Info Gain Ratio dataset and the DT Gain Ratio model when run on the Rule dataset were the best model in distinguishing sensitive and tolerant genotypes with 97.5% and 91.67% performances, respectively.

Conclusions

Overall, the results showed that combining bioinformatics, and laboratory studies can lead to identifying pathways associated with salinity and provide a better understanding of the mechanisms underpinning stress tolerance.

Keywords: biochemical markers, decision tree, feature selection, wheat.

Citation: Zinati Z, Alemzadeh A (2019) Identification of superior biochemical markers linked to salinity tolerance in wheat using data mining. *Agricultural Biotechnology Journal* 10(4), 55-74.

Agricultural Biotechnology Journal 10 (4), 55-74.

DOI: 10.22103/jab.2019.2249

Received: July 14, 2018; Accepted: December 5, 2018

© Faculty of Agriculture, Shahid Bahonar University of Kerman-Iranian Biotechnology Society

شناسایی نشانگرهای بیوشیمیایی برتر مرتبط با تحمل به شوری در گندم با استفاده از داده-

کاوی

زهرا زینتی

استادیار، بخش اگرواکولوژی، دانشکده کشاورزی و منابع طبیعی داراب، دانشگاه شیراز، ایران. ایمیل: zahra.zinati@hotmail.com

عباس عالمزاده

* نویسنده مسئول، دانشیار، بخش زراعت و اصلاح نباتات، دانشکده کشاورزی، دانشگاه شیراز، ایران. تلفن: ۰۷۱۳۲۲۸۶۱۳۴،

alemzadeh@shirazu.ac.ir

تاریخ دریافت: ۱۳۹۷/۴/۲۳، تاریخ پذیرش: ۱۳۹۷/۹/۱۴

چکیده

هدف: شناسایی مسیرهای موثر در ایجاد تحمل نسبت به تنش شوری یکی از مباحث جذاب در علوم گیاهی است که در این راستا روش‌های جدید داده‌کاوی نگرش جدیدی برای محققان ایجاد کرده است. در این پژوهش، الگوریتم‌های انتخاب ویژگی (feature selection) و درخت تصمیم‌گیری (decision tree, DT) به منظور شناسایی نشانگرهای بیوشیمیایی در تحمل به شوری استفاده شد.

مواد و روش‌ها: در این راستا، به منظور بررسی برخی نشانگرهای بیوشیمیایی از جمله میزان پروتئین، آنزیم‌های آنتی‌اکسیدانی سوپراکسید دیسموتاز، پراکسیداز، کاتالاز، آسکوربات پراکسیداز، پرولین، میزان سدیم و پتاسیم در ارقام گندم و خویشاوند وحشی *Aegilops crassa* آزمایشی به صورت فاکتوریل در قالب طرح کاملاً تصادفی با سه تکرار انجام شد. فاکتورهای آزمایشی شامل ژنوتیپ (ارقام گندم زراعی ارگ (مقاوم به شوری) و الموت (حساس به شوری)، و یک خویشاوند وحشی (*Ae. crassa*)) و تنش شوری (کلرید سدیم ۱۵۰ mM و ۰ mM) بودند.

نتایج: نتایج این دو راهکار نشان داد که ترکیب سلسله مراتبی پرولین، سوپراکسید دیسموتاز و آسکوربات پراکسیداز می‌تواند به عنوان نشانگرهای بیوشیمیایی برای انتخاب ژنوتیپ متحمل به تنش شوری استفاده گردد. دو درخت تصمیم‌گیری دارای بیشترین کارایی در پیش‌بینی ژنوتیپ حساس و متحمل بودند. این درخت‌های تصمیم‌گیری عبارت بودند از مدل درخت تصمیم‌گیری DT Parallel Gain Ratio با کارایی ۹۷/۵ درصد که روی پایگاه داده‌ای Info Gain Ratio اجرا شد و دیگری DT Gain Ratio با کارایی ۹۱/۶۷ درصد که روی پایگاه داده‌ای Rule اجرا شد.

نتیجه گیری: در مجموع، نتایج حاصله بیانگر آن بود که تلفیق مطالعات بیوانفورماتیک و آزمایشگاهی می‌تواند منجر به شناسایی مسیرهای مرتبط با تنش شوری و فهم بهتر مکانیسم‌های کنترل کننده تحمل به تنش شود.

کلمات کلیدی: انتخاب ویژگی، درخت تصمیم‌گیری، گندم، نشانگرهای بیوشیمیایی.

مقدمه

گیاهان طی چرخه زندگی خود در معرض تنش قرار می‌گیرند. تنش‌های زیستی و غیرزیستی از عواملی هستند که باعث کاهش عملکرد می‌گردند. یکی از مهم‌ترین تنش‌هایی که باعث کاهش عملکرد شدید می‌گردد، تنش شوری است. گیاهان در خاک‌های شور افزون بر تنش خشکی، اثر سمی برخی املاح را تجربه می‌کنند. شوری عمدتاً در مناطق خشک و نیمه خشک به وجود می‌آید. در مناطق خشک و نیمه خشک بیشتر بودن تبخیر و تعرق نسبت به نزولات آسمانی منجر به تجمع املاح در سطح خاک و به وجود آمدن شرایط شور می‌شود. نزدیک به ۲۰ درصد زمین‌های زیر کشت و تقریباً ۴۰ درصد از زمین‌های تحت آبیاری دنیا تحت تأثیر شوری است (Sahi et al. 2006). پژوهش‌هایی در رابطه با تحمل به شوری با هدف اصلاح گیاهان زراعی انجام شده است. اما ماهیت پیچیده و پلی‌ژنی تحمل به تنش شوری باعث مشکلاتی در اصلاح گیاهان متحمل به شوری می‌شود (Zhu et al. 2000). کاهش تورژانس برگ‌ها به دلیل تنش اسمزی منجر به بسته شدن روزنه‌ها می‌شود. بسته شدن روزنه‌ها سبب کاهش تجمع CO_2 در بافت مزوفیلی برگ در نتیجه افزایش تجمع NADPH می‌شود. تحت چنین شرایطی مقدار محدودی NADP برای پذیرش الکترون وجود خواهد داشت در حالی که واکنش نوری و انتقال الکترون به طور طبیعی صورت می‌گیرد. اکسیژن می‌تواند به عنوان یک گیرنده جایگزین برای الکترون NADP شود (Egneus et al. 1975; Cadenas, 1989). این امر منجر به تجمع گونه‌های فعال اکسیژن (Reactive Oxygen Species, ROS) مانند سوپراکسید (O_2^-)، هیدروژن پراکسید (H_2O_2) و هیدروکسیل (OH^-) می‌شود که از میان آن‌ها هیدروژن پراکسید (H_2O_2) دارای طول عمر و خسارت زایی بیشتر می‌باشد. تجمع گونه‌های فعال اکسیژن که طی تنش تولید می‌شوند باعث خسارت به بسیاری از ترکیبات سلولی از جمله DNA، پروتئین، چربی، کلروفیل و از همه مهم‌تر غشای سلولی و در نهایت مرگ سلول می‌شوند (Becana et al. 1998).

از طرفی گیاهان دارای تعدادی سیستم‌های آنزیمی و غیرآنزیمی برای جلوگیری از خسارت رادیکال‌های آزاد اکسیژن و در پی آن سازگاری با تنش هستند. آنتی‌اکسیدانت‌ها مهم‌ترین سیستم آنزیمی و دفاعی گیاهان در مقابله با تنش‌ها را تشکیل می‌دهند که از مهم‌ترین آن‌ها می‌توان به کاتالاز (Catalase, CAT)، پراکسیداز (Peroxidase, POD)، سوپراکسید دیسموتاز (Superoxide dismutase, SOD)، آسکوربات پراکسیداز (Ascorbate peroxidase, APX) و گلوکاتایون ردوکتاز (Glutathione reductase, GR) اشاره کرد و از مهم‌ترین سیستم‌های غیرآنزیمی می‌توان پرولین، بتائین، اسیدآسکوربات و کاروتنوئیدها را نام برد (Reddy et al. 2004). آنتی‌اکسیدانت‌های مختلف و تخریب‌کننده‌های ROS می‌توانند از طریق کاهش

صدمه اکسیداتیو مقاومت گیاهان به شوری یا خشکی را افزایش دهند (Vinocur & Altman 2005). یکی از مؤثرترین و کاربردی‌ترین روش‌های مقابله با شوری، بکارگیری ارقام متحمل می‌باشد. از جمله عواملی که می‌توانند به عنوان معیارهایی جهت شناخت ارقام متحمل مورد ارزیابی قرار گیرند، تغییرات بیوشیمیایی ایجاد شده در گیاهان تحت تنش شوری می‌باشند.

استفاده از روش‌های جدید داده‌کاوی نگرش جدیدی برای دست‌ورزی ژنتیکی گیاهان ایجاد می‌کند. علم داده‌کاوی مجموعه‌ای از علوم مختلف از جمله آمار، یادگیری ماشین و هوش مصنوعی می‌باشد که با تجزیه و تحلیل مجموعه بزرگی از داده‌ها به کشف الگوها و قوانین پنهان درون داده‌ها می‌پردازد (Witten et al. 2005). داده‌کاوی به عنوان تکنیکی ارزشمند برای شناسایی و یافتن مدل‌هایی برای پیش‌بینی سریع‌تر ژنوتیپ‌های حساس و متحمل به تنش‌ها می‌تواند کمک بزرگی باشد. کاهش هزینه‌ها، کاهش منابع انسانی، افزایش دقت شناسایی از مزایای داده‌کاوی می‌باشد. یکی از پرکاربردترین روش‌های مورد استفاده در داده‌کاوی، انتخاب ویژگی (feature selection) می‌باشد. انتخاب ویژگی در حقیقت فرایندی است که طی آن مجموعه خصوصیات مهم و اصلی با کنار گذاشتن داده‌های بی‌ارزش انتخاب می‌شوند. با استفاده از انتخاب ویژگی می‌توان مدل‌های بهتری به دست آورد. بسیاری از الگوریتم‌هایی که در انتخاب ویژگی مورد استفاده قرار می‌گیرند مبتنی بر رتبه‌دهی متغیرها می‌باشند (Bekkerman et al. 2003; Weston et al. 2003). همچنین درخت‌های تصمیم‌گیری یکی دیگر از روش‌های داده‌کاوی برای دسته‌بندی و پیش‌بینی به شمار می‌روند. در پژوهش‌های پیشین از روش‌های انتخاب ویژگی و درخت تصمیم‌گیری برای تعیین ویژگی‌های مهم ساختاری، پیش‌بینی و طبقه‌بندی پایداری حرارتی پروتئین‌ها (Ebrahimi et al. 2009)، پمپ p-گلایکوپروتئین (Hamann et al. 2009)، پروتئین‌های متحمل به شوری (Ebrahimie et al. 2011)، طبقه‌بندی و پیش‌بینی ارقام زیتون (Beiki et al. 2012)، میزان آلفا-لینولینیک اسید (Zinati et al. 2014)، تمایز ژنوتیپ‌ها (Nasiri et al. 2015) و پمپ‌های ATPase (Zinati et al. 2016) استفاده شده است. در پژوهشی که توسط Shaik و Ramakrishna (2013) انجام شد، مدل یادگیری نظارت شده قادر بود با دقت ۱۰۰ درصد با استفاده از مجموعه‌ای از ژن‌های دارای بیان افتراقی، تنش‌های زیستی و غیرزیستی را طبقه‌بندی کند و همچنین ژن‌های دخیل در مقاومت عمومی را شناسایی کند. علاوه بر این، تلفیق یادگیری ماشین و تجزیه و تحلیل چندمتغیره توانست یک ماژول پیش‌بینی پاسخ متابولیکی به تنش را در آرابیدوپسیس شناسایی کند (Furtauer et al. 2018). در پژوهش حاضر داده‌کاوی با استفاده از انتخاب ویژگی و درخت تصمیم‌گیری روی داده‌های حاصل از اندازه‌گیری معیارهای بیوشیمیایی انجام شد.

مواد و روش‌ها

به منظور شناسایی نشانگرهای بیوشیمیایی مرتبط با تحمل به شوری در ارقام گندم و خویشاوند وحشی آن *Aegilops crassa*، آزمایشی به صورت فاکتوریل در قالب طرح کاملاً تصادفی با سه تکرار در دانشکده کشاورزی دانشگاه شیراز در سال ۹۲

انجام شد. فاکتورهای آزمایشی شامل ژنوتیپ (ارقام گندم زراعی ارگ (مقاوم به شوری) و الموت (حساس به شوری))، و یک خویشاوند وحشی (*Ae. crassa*) و تنش شوری (کلرید سدیم ۱۵۰ mM و ۰ mM) بودند. در این تحقیق از کشت هیدروپونیک و محیط کشت مایع هوگلند استفاده شد (Kerepesi & Galiba 2000). لازم به ذکر است که برای حفظ نسبت ۱:۱۰ کلسیم به سدیم، کلرید کلسیم به نسبت مناسب به همراه کلرید سدیم به محیط اضافه شد. گیاهچه‌های ۲۱ روزه به مدت سه هفته تحت تنش شوری قرار گرفته و نمونه‌برداری از بافت شاخساره و ریشه در سه هفته پس از اعمال تنش شوری انجام شد. نمونه‌ها در نیتروژن مایع منجمد شده و در دمای ۸۰- درجه سانتی‌گراد نگهداری شدند.

سنجش میزان پروتئین، فعالیت آنزیم‌های آنتی‌اکسیدانی

به منظور اندازه‌گیری پروتئین، یک گرم بافت تر در حضور ۴ میلی‌لیتر بافر استخراج (تریس اسیدکلریدریک (pH=8) هموژنیزه گردید. مخلوط حاصل را در اپندورف در دمای ۴ درجه سانتیگراد و به مدت ۱۵ دقیقه با دور ۱۳۰۰۰g سانتریفیوژ نموده پس از آن محلول رویی جهت قرائت میزان پروتئین و فعالیت آنزیم‌ها جدا شد. میزان پروتئین محلول طبق روش برادفورد (Bradford, 1976) اندازه‌گیری شد.

به منظور سنجش فعالیت آنزیم‌های آنتی‌اکسیدانی سوپراکسید دیسموتاز، پراکسیداز، کاتالاز، آسکوربات پراکسیداز به ترتیب روش‌های فریدوویچ (Fridovich, 1986)، روش چنس و ماهلی با کمی تغییرات (Chance & Maehly 1954)، روش دیندسا و همکاران (Dhindsa et al. 1981) و روش ناکانو و آسادا (Nakano & Asada 1981) استفاده شد. فعالیت آنزیم سوپراکسید دیسموتاز از طریق سنجش فعالیت آن در جلوگیری از احیای نوری نیترو بلو تترازولیوم کلراید (NBT) اندازه‌گیری شد. سنجش فعالیت آنزیم پراکسیداز با استفاده از گایاکل و میزان جذب تتراکایاکل حاصل از گایاکل در نتیجه فعالیت پراکسیداز در ۴۷۰ نانومتر انجام شد. میزان فعالیت آنزیم کاتالاز با استفاده از محاسبه کاهش جذب H_2O_2 در ۲۴۰ نانومتر سنجش شد. فعالیت آنزیم آسکوربات پراکسیداز با بررسی میزان پراکسیداسیون آسکوربات با کاهش جذب ۲۹۰ نانومتر انجام شد.

اندازه‌گیری پرولین

برای اندازه‌گیری غلظت پرولین از روش بیتس و همکاران استفاده شد (Bates et al. 1973). در این روش با استفاده از دستگاه اسپکتروفتومتر در طول موج ۵۲۰ نانومتر اندازه‌گیری و با استفاده از رابطه
$$\text{Proline } (\mu\text{M g}^{-1} \text{ fresh wt.}) = [(M \cdot T \cdot W)/115.5] \times 1000$$
 محاسبه گردید که در آن M: عدد قرائت شده با دستگاه اسپکتروفتومتر، T: حجم تولوئن مورد استفاده، W: وزن نمونه برگی مورد استفاده بود.

اندازه‌گیری سدیم و پتاسیم

به منظور تعیین میزان سدیم و پتاسیم از روش گودرزی و همکاران استفاده شد (Goudarzi & Pakniyat 2008). میزان سدیم و پتاسیم با دستگاه فلیم‌فتومتر (UK-Jenway) و محلول‌های استاندارد سدیم و پتاسیم تعیین گردید.

انتخاب ویژگی

زمانی که داده‌های مورد بررسی دارای مقیاس‌های مختلفی هستند نیاز هست که آن‌ها را نرمال کرد خصوصاً زمانی که از عملگرهای کلاسترینگ استفاده می‌کنیم. بدین منظور نرمال سازی داده‌ها، و سپس انتخاب ویژگی و ایجاد درخت‌های تصمیم گیری با Rapid Miner (Rapid Miner 5.0.001, Rapid-I GmbH, Stochumer Str. 475, 44227 Dortmund, Germany) انجام شد. الگوریتم final cleaned database (FCdb) داده‌هایی که همبستگی بیشتر از ۹۰ درصد دارند و انحراف معیارشان کمتر یا مساوی ۰/۱ می‌باشد را حذف می‌کند. ده الگوریتم وزن‌دهی برای انتخاب مهمترین معیارهای بیوشیمیایی در تمایز ژنوتیپ حساس و متحمل به شوری روی داده‌های اصلی (FCdb) استفاده شدند. الگوریتم‌های مورد استفاده در این پژوهش عبارت از وزن‌دهی توسط PCA, Uncertainty, Relief, Chi Squared, Gini Index, Deviation, Rule, Gain Ratio, Info Gain و Support Vector Machine (SVM) بودند. الگوریتم PCA، از فاکتورهای تحلیل مؤلفه‌های اصلی به عنوان وزن‌های ویژگی استفاده می‌کند. الگوریتم Uncertainty، از طریق اندازه‌گیری عدم قطعیت متقارن نسبت به متغیر پاسخ، به وزن‌دهی هر ویژگی می‌پردازد. الگوریتم Relief، از طریق نمونه‌برداری از مشاهدات و مقایسه مقدار هر ویژگی در مشاهدات انتخاب شده با مقدار آن در دو مشاهده همسایه که یکی در کلاس یکسان و دیگری در کلاس متفاوت می‌باشد، به وزن‌دهی ویژگی می‌پردازد. الگوریتم Chi Squared از طریق محاسبه آماره کای دو، برای هر ویژگی و با توجه به متغیر پاسخ، به وزن‌دهی هر ویژگی می‌پردازد. الگوریتم Gini Index، وزن‌دهی هر ویژگی را از طریق ضریب Gini انجام می‌دهد. الگوریتم Deviation، بر مبنای محاسبه انحراف معیار ویژگی‌ها می‌باشد. الگوریتم Rule، ارتباط یک ویژگی را با متغیر پاسخ اندازه‌گیری می‌کند و اینکار را از طریق ساختن یک قانون منحصر بفرد برای هر ویژگی و سپس محاسبه خطاها انجام می‌دهد. الگوریتم Gain Ratio، اهمیت یا وزن ویژگی‌ها را بر پایه نسبت بهره اطلاعات برای توزیع کلاس محاسبه می‌کند. الگوریتم Info Gain، اهمیت یا وزن ویژگی‌ها را بر پایه بهره اطلاعات محاسبه می‌کند. الگوریتم SVM، از ضرایب بردار نرمال یک ماشین بردار پشتیبان خطی به عنوان وزن‌های ویژگی استفاده می‌کند (Hosseinzadeh et al. 2012).

همچنین برای ایجاد درخت‌های تصمیم‌گیری ابتدا ده مجموعه داده، توسط فیلتر داده‌های اصلی (FCdb) توسط الگوریتم‌های Rule, Relief, PCA, Info Gain Ratio, Deviation Gini Index, Info Gain, Chi Squared, Decision و SVM ایجاد شدند. برای ایجاد درخت‌های تصمیم‌گیری نیز از چهار مدل درخت تصمیم‌گیری (DT) Tree (DT), DT Parallel, DT Random Forest و DT Stump استفاده شد که هر کدام با چهار شاخص Gini Index, Accuracy, Info Gain و Info Gain Ratio اجرا شدند. در کل شانزده مدل درخت تصمیم‌گیری (Decision Tree (DT) Accuracy, Info Gain, Accuracy, DT Info Gain, DT Gini Index, DT Gain Ratio, DT Parallel Accuracy, DT Parallel Gini Index, Parallel Gain Ratio, DT Stump Accuracy, DT Parallel Info Gain, DT Parallel Gini Index, DT Stump Accuracy, DT Parallel Info Gain, DT Parallel Gini Index, Parallel Gain Ratio

DT Random Forest Accuracy, DT Stump Info Gain, DT Stump Gini Index, Stump Gain Ratio, DT Random Forest Gain, DT Random Forest Info Gain و DT Random Forest Gini Index, Ratio. روی یازده مجموعه داده (Dataset) اجرا شدند. بر اساس پژوهش‌های پیشین رقم ارگ متحمل به شوری و رقم الموت حساس به شوری می‌باشد (Asadi et al. 2012, Niazi et al. 2014). همچنین در *Ae. crassa* صفات زراعی بالقوه از جمله تحمل به شوری گزارش شده است (Liu et al. 2002). در این مطالعه دسته‌بندی ژنوتیپ‌ها بر اساس میزان حساسیت آن‌ها به شوری انجام گرفت. در رابطه با دسته‌بندی ژنوتیپ‌ها بر اساس میزان حساسیت آن‌ها به شوری، رقم ارگ و خویشاوند وحشی *Ae. crassa*، متحمل به شوری و رقم الموت حساس به شوری در نظر گرفته شدند. برای بررسی اعتبار درخت‌های تصمیم‌گیری، اختصاصی بودن^۱، حساسیت^۲، صحت^۳ و سطح زیر منحنی ROC^۴ اندازه‌گیری شدند. حساسیت توانائی یک آزمون در تشخیص درست همه مثبت‌های حقیقی ($Sensitivity = TP / (TP + FN)$) و اختصاصی بودن، توانائی آزمون برای تشخیص درست همه منفی‌های حقیقی ($Specificity = TN / (TN + FP)$) می‌باشد. برای محاسبه صحت، پیش‌بینی‌های اشتباه از کل پیش‌بینی‌ها کم شده و سپس به کل پیش‌بینی‌ها تقسیم شدند $Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$. به عبارتی صحت، میزان دقیق کل پیش‌بینی‌ها را نشان می‌دهد. TP, TN, FP و FN به ترتیب تعداد مثبت‌های حقیقی، تعداد منفی‌های حقیقی، تعداد مثبت‌های کاذب و تعداد منفی‌های کاذب می‌باشند. مشخصه عملکرد گیرنده (ROC^۵) برای ارزیابی کارایی مدل‌ها استفاده شد. منحنی ROC نقشه حساسیت در برابر میزان مثبت کاذب می‌باشد. در یک پیش‌بینی دقیق و قابل اعتماد اختصاصی بودن و حساسیت نزدیک به ۱ می‌باشند. سطح زیر منحنی ROC به عنوان یک معیار معتبر در بحث مدلینگ مطرح است. هر چه میزان AUC به ۱ نزدیکتر باشد، پیش‌بینی قابل اعتمادتر می‌باشد (Bradley, 1997).

نتایج و بحث

بر اساس پژوهش‌های پیشین، رقم ارگ و *Ae. crassa* به عنوان متحمل به شوری و رقم الموت حساس به شوری در نظر گرفته شد (Liu et al. 2002; Asadi et al. 2012, Niazi et al. 2014). یافتن معیارهای مهم بیوشیمیایی که

¹ Specificity

² Sensitivity

³ Accuracy

⁴ AUC, Area Under the ROC Curve

⁵ Receiver Operating Characteristic

ژنوتیپ متحمل را از ژنوتیپ حساس متمایز می‌سازد می‌تواند نقش مهمی در اصلاح نباتات داشته باشد. برای انجام انتخاب ویژگی داده‌های حاصل از اندازه‌گیری معیارهای بیوشیمیایی استفاده شد. معیارهای بیوشیمیایی مورد استفاده شامل پرولین، پروتئین، پراکسیداز، آسکوربات پراکسیداز، کاتالاز، سوپراکسید دیسموتاز، Na^+ ، K^+ ، K^+/Na^+ بودند. از آنجا که داده‌ها نرمال شده‌اند، انتظار می‌رود که وزن‌ها بین ۰ تا ۱ باشند. مقادیر نزدیک‌تر به ۱ اهمیت بیشتری در تمایز ژنوتیپ حساس و متحمل دارند.

نتایج انتخاب ویژگی نشان داد که پرولین، سوپراکسید دیسموتاز، Na^+ و آسکوربات پراکسیداز به ترتیب توسط ۸، ۵، ۴ و ۴ الگوریتم وزن‌دهی، بیشترین اهمیت را در تمایز ژنوتیپ حساس و متحمل دارند (جدول ۱). بر این اساس، ابتدا پرولین و در درجه بعد به ترتیب سوپراکسید دیسموتاز، Na^+ و آسکوربات پراکسیداز از مهمترین معیارهای بیوشیمیایی در پاسخ به شوری و تمایز ژنوتیپ حساس و متحمل در سه هفته پس از تنش می‌باشند و می‌توانند جهت مهندسی گیاه گندم با هدف بهبود تحمل به تنش شوری استفاده شوند. در زمینه استفاده از روش‌های اصلاحی، داشتن اطلاعاتی از نقش مواد بیوشیمیایی مؤثر در تحمل از اهمیت خاصی برخوردار است. یکی از پاسخ‌های عمومی گیاهان برای کاهش تنش شوری و خشکی سنتز و انباشت محلول‌های حفاظت کننده اسمزی مانند پرولین می‌باشد (Iqbal et al. 2014; Kaur & Asthir 2015). پرولین علاوه بر تنظیم اسمزی در محافظت از پروتئین‌های گیاه در شرایط تنش کمک می‌کند (Kishor et al. 2005). زمانی که گیاه تحت تنش قرار می‌گیرد انرژی خود را صرف تولید پرولین نموده و از طریق تنظیم اسمزی موجب تحمل به تنش می‌شود (Pioter Grazyana, 2005). پرولین اثرات گونه‌های فعال اکسیژن را توسط تثبیت سیستم آنتی‌اکسیدانی و حفظ یکپارچگی غشا کاهش می‌دهد (Reddy et al. 2015). همچنین وجود نقش تنظیم‌کنندگی برای پرولین به عنوان یک مولکول پیام‌رسان در تنش‌های زیستی و غیر زیستی پیشنهاد شده است (Szabados & Savoure 2010). یک توافق عمومی وجود دارد که پرولین بیان ژن‌های دخیل در تحمل به شوری و خشکی را افزایش می‌دهد (Naliwajski & Skłodowska 2014; Patade et al. 2014; Teh et al. 2015).

کارایی^۱ شانزده مدل درخت تصمیم‌گیری که روی یازده مجموعه داده اجرا شدند در جدول ۲ آمده است. دو درخت تصمیم‌گیری دارای بیشترین کارایی نسبت به درخت‌های تصمیم‌گیری دیگر بودند (جدول ۲). این درخت‌های تصمیم‌گیری عبارتند از مدل درخت تصمیم‌گیری DT Parallel Gain Ratio با کارایی ۹۷/۵ درصد که روی پایگاه داده‌ای Info Gain Ratio اجرا شد (شکل ۱) و دیگری DT Gain Ratio با کارایی ۹۱/۶۷ درصد که روی پایگاه داده‌ای Rule اجرا شد (شکل ۲). برای بررسی اعتبار دو مدل DT Parallel Gain Ratio و DT Gain Ratio که به ترتیب روی مجموعه داده Info Gain Ratio و Rule اجرا شدند اختصاصی بودن، حساسیت، صحت و سطح زیر منحنی ROC اندازه‌گیری شدند (جدول ۳). بر اساس میزان اعتبار دو مدل می‌توان گفت این دو درخت تصمیم‌گیری مدل‌های کارایی در پیش‌بینی ژنوتیپ‌های متحمل و حساس به شوری می‌باشند.

¹ Performance

در مدل درخت تصمیم‌گیری DT Parallel Gain Ratio که روی پایگاه داده‌ای Info Gain Ratio اجرا شد، اولین خصوصیت میزان سدیم بود. دومین خصوصیت مهم میزان پرولین و سومین خصوصیت مهم آنزیم سوپراکسید دیسموتاز بود. به عبارتی این مدل قادر است با استفاده از میزان سه معیار بیوشیمیایی سدیم، پرولین و فعالیت آنزیم سوپراکسید دیسموتاز حساسیت یا متحمل بودن ژنوتیپ‌ها را پیش بینی کند.

در مدل درخت تصمیم‌گیری DT Gain Ratio که روی پایگاه داده‌ای Rule اجرا شد، پرولین، پروتئین و آسکوربات پراکسیداز، معیارهای ایجاد کننده درخت تصمیم‌گیری بودند. به عبارتی این مدل قادر است با استفاده از میزان پرولین، پروتئین و فعالیت آسکوربات پراکسیداز حساسیت یا متحمل بودن ژنوتیپ‌ها را پیش بینی کند.

نکته قابل توجه این است که پرولین در ایجاد هر دو درخت تصمیم‌گیری نقش دارد و همچنین نتایج حاصل از انتخاب ویژگی را نیز تأیید می‌کند. تنظیم اسمزی از مهمترین فرایندهای مولکولی و فیزیولوژی هستند که می‌توانند منجر به اعطای توانایی حفظ آب به گیاه شوند (Thapa et al. 2011). نکته دیگری که در این پژوهش حائز اهمیت می‌باشد این است که ترکیب سلسله مراتبی تعدادی از نشانگرها بیوشیمیایی می‌تواند منجر به ایجاد تحمل به شوری شود و با توجه به درخت‌های تصمیم نمی‌توان با اکتفا تنها بر روی یک نشانگر (برای مثال پرولین) ارقام متحمل و حساس را متمایز کرد. در تأیید این فرضیه، Flower و Colmer (2008) بیان کردند که میزان مشارکت سدیم و کلر در تنظیم اسمزی گونه‌های هالوفیت متفاوت است و به نظر می‌رسد با غلظت محلول‌های آلی تجمع یافته در ارتباط باشد.

همچنین علاوه بر مد نظر قرار دادن ترکیب سلسله مراتبی نشانگرهای بیوشیمیایی باید به میزان آن‌ها نیز توجه شود. با توجه به مدل درخت تصمیم‌گیری DT Gain Ratio، اگر پرولین بیشتر از $24 \mu\text{g g}^{-1} \text{FW}$ باشد ژنوتیپ حساس، اگر میزان پرولین کمتر یا مساوی $24 \mu\text{g g}^{-1} \text{FW}$ و میزان پروتئین کمتر یا مساوی $1/203 \text{ mg g}^{-1} \text{FW}$ ژنوتیپ حساس و اگر میزان پروتئین بیشتر از $1/203 \text{ mg g}^{-1} \text{FW}$ و میزان پرولین بیشتر از $11 \mu\text{g g}^{-1} \text{FW}$ باشد ژنوتیپ متحمل است. ولی اگر میزان پرولین کمتر یا مساوی $11 \mu\text{g g}^{-1} \text{FW}$ و فعالیت آسکوربات پراکسیداز بیشتر از $330/357$ (واحد آنزیم بر گرم وزن تر) باشد ژنوتیپ حساس و در صورتی که فعالیت آسکوربات پراکسیداز کمتر یا مساوی $330/357$ (واحد آنزیم بر گرم وزن تر) باشد، ژنوتیپ متحمل می‌باشد (شکل ۲). بیشتر گزارشات موجود در رابطه با تغییرات غلظت پرولین، سدیم، فعالیت آنزیم‌های سوپراکسید دیسموتاز و آسکوربات پراکسیداز متناقض هستند و هیچ مدرک مستندی مبنی بر اینکه میزان تجمع پرولین، سدیم و فعالیت آنزیم‌های سوپراکسید دیسموتاز و آسکوربات پراکسیداز در ارقام مقاوم الزاماً بیشتر از ارقام حساس می‌باشد وجود ندارد (Mansour et al. 2005). در بررسی اثرات تأثیر سدیم بر محتوای پرولین دو رقم حساس و مقاوم گندم مشاهده شد که محتوای پرولین در رقم حساس تحت شرایط تنش و کنترل بیشتر از رقم مقاوم بود. همچنین محتوای پرولین در رقم حساس تحت تنش حدود $16 \mu\text{g g}^{-1} \text{FW}$ و در رقم مقاوم تحت تنش حدود 23 FW بود که می‌تواند تأییدی بر نتایج حاصل از درخت‌های تصمیم

باشد (Akbari et al. 2016). با توجه به مدل‌های درخت تصمیم‌گیری، به نظر می‌رسد محدوده خاصی از میزان هر کدام از شاخص‌های بیوشیمیایی برای اعطای تحمل به شوری مورد نیاز می‌باشد و غلظت خارج از این محدوده ممکن است منجر به حساسیت به شوری گردد. به عنوان مثال سدیم نقش مهمی را در تنظیم اسمزی در شرایط تنش شوری ایفا می‌کند و افزایش آن‌ها به کاهش پتانسیل اسمزی کمک خواهد کرد. درحالی که غلظت بالای سدیم باعث برهم زدن تعادل یون‌های درون سلول شده و منجر به عملکرد نامناسب غشا و بازدارندگی رشد و نهایتاً مرگ سلول می‌شود (Wang et al. 2004; Flowers & Colmer 2008; Salma et al. 2015). در پژوهش حاضر نیز میزان غلظت در گیاهان بیشتر از $10/103$ (mg g^{-1}) بافت خشک منجر به حساسیت به شوری گردید.

در مجموع بر اساس ده الگوریتم انتخاب ویژگی (feature selection) از هشت صفت مورد بررسی پرولین، سوپراکسید دیسموتاز، میزان سدیم و آسکوربات پراکسیداز به ترتیب توسط ۸، ۵، ۴ و ۴ الگوریتم وزن‌دهی، به عنوان مهمترین نشانگرهای بیوشیمیایی بیشترین نقش را در تمایز ژنوتیپ حساس و متحمل داشتند. همچنین بر اساس نتایج درخت‌های تصمیم‌گیری، دو درخت تصمیم‌گیری دارای بیشترین کارایی نسبت به درخت‌های تصمیم‌گیری دیگر بودند. هر درخت تصمیم‌گیری تنها با استفاده از سه نشانگر بیوشیمیایی قادر به پیش بینی حساسیت یا متحمل بودن ژنوتیپ‌ها بود. نشانگر بیوشیمیایی پرولین هم توسط انتخاب ویژگی و هم درخت‌های تصمیم‌گیری انتخاب گردید. بر این اساس، شاید بتوان پرولین را به عنوان نشانگر برتر معرفی کرد. بر اساس نتایج درخت‌های تصمیم‌گیری، دامنه خاصی از میزان هر کدام از نشانگرهای بیوشیمیایی برای اعطای تحمل به شوری مورد نیاز می‌باشد و غلظت خارج از این محدوده ممکن است منجر به حساسیت به شوری گردد. این پژوهش علاوه بر اینکه میزان غلظت مورد نیاز برای ایجاد تحمل به شوری را برای نشانگرهای بیوشیمیایی کلیدی تعیین می‌کند ترکیب سلسله مراتبی این شاخص‌های بیوشیمیایی را هم ارائه می‌دهد که این ترکیب سلسله مراتبی در درخت‌های تصمیم‌گیری نشان داده شده است.

تلفیق نتایج حاصل از انتخاب ویژگی و مدل‌های درخت تصمیم‌گیری نشان می‌دهد که نشانگرهای پرولین، سوپراکسید دیسموتاز، میزان سدیم و آسکوربات پراکسیداز از نقاط مهم و کلیدی پاسخ به شوری در گندم می‌باشند. ترکیب سلسله مراتبی این شاخص‌های بیوشیمیایی و غلظت مورد نیاز آن‌ها می‌تواند جهت دست‌ورزی ژنتیکی با هدف بهبود تحمل به تنش شوری در گندم و دیگر گیاهان زراعی مورد استفاده قرار گیرد.

جدول ۱. الگوریتم‌های وزن دهی برای انتخاب مهمترین معیارهای بیوشیمیایی در تمایز ژنوتیپ حساس و متحمل به شوری. پارامترهای بیوشیمیایی با مقادیر نزدیک تر به ۱ اهمیت بیشتری در تمایز ژنوتیپ حساس و متحمل دارند. الگوریتم‌های مورد استفاده عبارتند از PCA (تحلیل مؤلفه‌های اصلی)، SVM (ماشین بردار پشتیبانی)، Relief (امداد)، Uncertainty (عدم قطعیت)، Gini Index (شاخص جینی)، Chi Squared (کای اسکوئر)، Deviation (انحراف)، Rule (قانون)، Info Gain Ratio (نسبت بهره اطلاعاتی) و Gain (بهره اطلاعاتی)

Table 1. The most important biochemical markers, based on different weighting algorithms, to be involved in the discrimination of salt sensitive and tolerant genotypes (Values closer to 1 shows higher effectiveness of features). Weighting algorithms were PCA, Relief, Uncertainty, Gini index, Chi Squared, Deviation, Rule, Correlation, Gain Ratio, and Information Gain.

Attribute	Weight_PCA	Weight_SVM	Weight_Relief	Weight_Uncertainty	Weight_Gini Index	Weight_Chi Squared	Weight_Deviation	Weight_Rule	Weight_Info Gain Ratio	Weight_Info Gain	*Number of more than 0.5
Prolin	0.00	1.00	1.00	1.00	1.00	1.00	0.01	0.67	1.00	1.00	8
SOD	1.00	0.93	0.98	0.39	0.20	0.32	1.00	0.00	0.51	0.26	5
Na	0.00	0.56	0.53	0.43	0.38	0.44	0.00	0.50	0.63	0.38	4
APX	0.20	0.70	0.73	0.25	0.38	0.27	0.43	1.00	0.63	0.38	4
Pr	0.01	0.00	0.90	0.14	0.20	0.14	0.00	0.92	0.51	0.21	3
K	0.01	0.63	0.52	0.08	0.15	0.08	0.01	0.00	0.06	0.11	2
K/Na	0.00	0.02	0.28	0.32	0.03	0.32	0.00	0.00	0.36	0.04	0
CAT	0.01	0.11	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0
POD	0.05	0.37	0.11	0.17	0.05	0.19	0.17	0.00	0.00	0.03	0

*تعداد الگوریتم‌های وزن دهی که به ویژگی مورد نظر وزن بیشتر از ۰/۵۰ اختصاص دادند.

*The number of algorithms which assigned weights more than 0.5 to the given feature.

جدول ۲. مقایسه کارایی ۱۶ الگوریتم درخت تصمیم‌گیری. DT Accuracy (درخت تصمیم صحت)، DT Gain Ratio (درخت تصمیم نسبت بهره)، DT Gini Index (درخت تصمیم شاخص جینی)، DT Info Parallel (درخت تصمیم بهره اطلاعاتی)، DT Parallel Accuracy (درخت تصمیم صحت موازی)، DT Gain Ratio (درخت تصمیم نسبت بهره موازی)، DT Parallel Gini Index (درخت تصمیم شاخص جینی موازی)، DT Parallel Info Gain (درخت تصمیم بهره اطلاعاتی موازی)، DT Stump Accuracy (درخت تصمیم صحت استامپ)، DT Stump Gain Ratio (درخت تصمیم نسبت بهره استامپ)، DT Stump Gini Index (درخت تصمیم شاخص جینی استامپ)، DT Stump Info Gain (درخت تصمیم بهره اطلاعاتی استامپ)، DT Random Forest Accuracy (درخت تصمیم بهره اطلاعاتی استامپ)، DT Random Forest Gain Ratio (درخت تصمیم نسبت بهره جنگل تصادفی)، DT Random Forest Gini Index (درخت تصمیم شاخص جینی جنگل تصادفی) و DT Random Forest Info Gain (درخت تصمیم بهره اطلاعاتی جنگل تصادفی) در پیش‌بینی ژنوتیپ متحمل و حساس که روی ۱۱ مجموعه داده (کای اسکوئر، Info Gain، بهره اطلاعاتی، Deviation (انحراف)، Gini Index (شاخص جینی)، Info Gain Ratio (نسبت بهره اطلاعاتی)، PCA (تحلیل مؤلفه‌های اصلی)، Relief (امداد)، Rule (قانون)، Uncertainty (عدم قطعیت)، FCdb (داده‌های اصلی) و SVM (ماشین بردار پشتیبانی) اجرا شدند.

Table 2. Comparison of performance of sixteen Decision Tree Algorithms such as DT Accuracy, DT Gain Ratio, DT Gini Index, DT Info Gain, DT Parallel Accuracy, DT Parallel Gain Ratio, DT Parallel Gini Index, DT Parallel Info Gain, DT Stump Accuracy, DT Stump Gain Ratio, DT Stump Gini Index, DT Stump Info Gain, DT Random Forest Accuracy, DT Random Forest Gain Ratio, DT Random Forest Gini Index, DT Random Forest Info Gain in the prediction of salt sensitive and tolerant genotypes which were run on eleven datasets including Chi Squared, Info Gain, Deviation, Gini Index, Info Gain Ratio, PCA, Relief, Rule, Uncertainty, FCdb, and SVM.

Data Base	DT Accuracy	DT Gain Ratio	DT Gini Index	DT Info Gain	DT Parallel Accuracy	DT Parallel Gain Ratio	DT Parallel Gini Index	DT Parallel Info Gain
Chi Squared	84.17%	78.33%	78.33%	78.33%	84.17%	78.33%	78.33%	78.33%
Info Gain	84.17%	78.33%	78.33%	78.33%	84.17%	78.33%	78.33%	78.33%
Deviation	66.67%	71.67%	60.83%	71.67%	66.67%	71.67%	60.83%	71.67%
Gini Index	84.17%	78.33%	78.33%	78.33%	84.17%	78.33%	78.33%	78.33%
Info Gain Ratio	84.17%	86.67%	75.00%	89.17%	84.17%	97.50%	83.33%	84.17%
PCA	66.67%	71.67%	60.83%	71.67%	66.67%	71.67%	60.83%	71.67%
Relief	80.83%	80.00%	71.67%	82.50%	85.83%	85.00%	79.17%	60.83%
Rule	84.17%	91.67%	88.33%	85.83%	84.17%	76.67%	76.67%	76.67%
Uncertainty	84.17%	78.33%	78.33%	78.33%	84.17%	78.33%	78.33%	78.33%
FCdb	80.83%	80.00%	71.67%	80.00%	80.83%	81.67%	79.17%	86.67%
SVM	80.83%	83.33%	77.50%	85.83%	80.83%	80.00%	83.33%	76.67%

ادامه جدول ۲.

Table 2 (continued).

Data Base	DT Stump Accuracy	DT Stump Gain Ratio	DT Stump Gini Index	DT Stump Info Gain	DT Random Forest Accuracy	DT Random Forest Gain Ratio	DT Random Forest Gini Index	DT Random Forest Info Gain
Chi Squared	84.17%	84.17%	84.17%	84.17%	84.17%	78.33%	80.83%	78.33%
Info Gain	84.17%	84.17%	84.17%	84.17%	84.17%	78.33%	80.83%	78.33%
Deviation	69.17%	69.17%	66.67%	66.67%	66.67%	69.17%	68.33%	59.17%
Gini Index	84.17%	84.17%	84.17%	84.17%	84.17%	78.33%	80.83%	78.33%
Info Gain Ratio	84.17%	84.17%	84.17%	84.17%	80.00%	89.17%	83.33%	83.33%
PCA	69.17%	69.17%	66.67%	66.67%	66.67%	69.17%	68.33%	59.17%
Relief	84.17%	84.17%	84.17%	84.17%	75.00%	71.67%	77.50%	77.50%
Rule	84.17%	84.17%	84.17%	84.17%	84.17%	80.83%	75.00%	86.67%
Uncertainty	84.17%	84.17%	84.17%	84.17%	84.17%	78.33%	80.83%	78.33%
FCdb	84.17%	84.17%	84.17%	84.17%	72.50%	76.67%	68.33%	80.00%
SVM	84.17%	84.17%	84.17%	84.17%	80.83%	85.83%	72.50%	89.17%

جدول ۳. اختصاصی بودن، حساسیت، صحت و سطح زیر منحنی مدل‌های درخت تصمیم‌گیری با بیشترین کارایی در پیش‌بینی ژنوتیپ حساس و متحمل

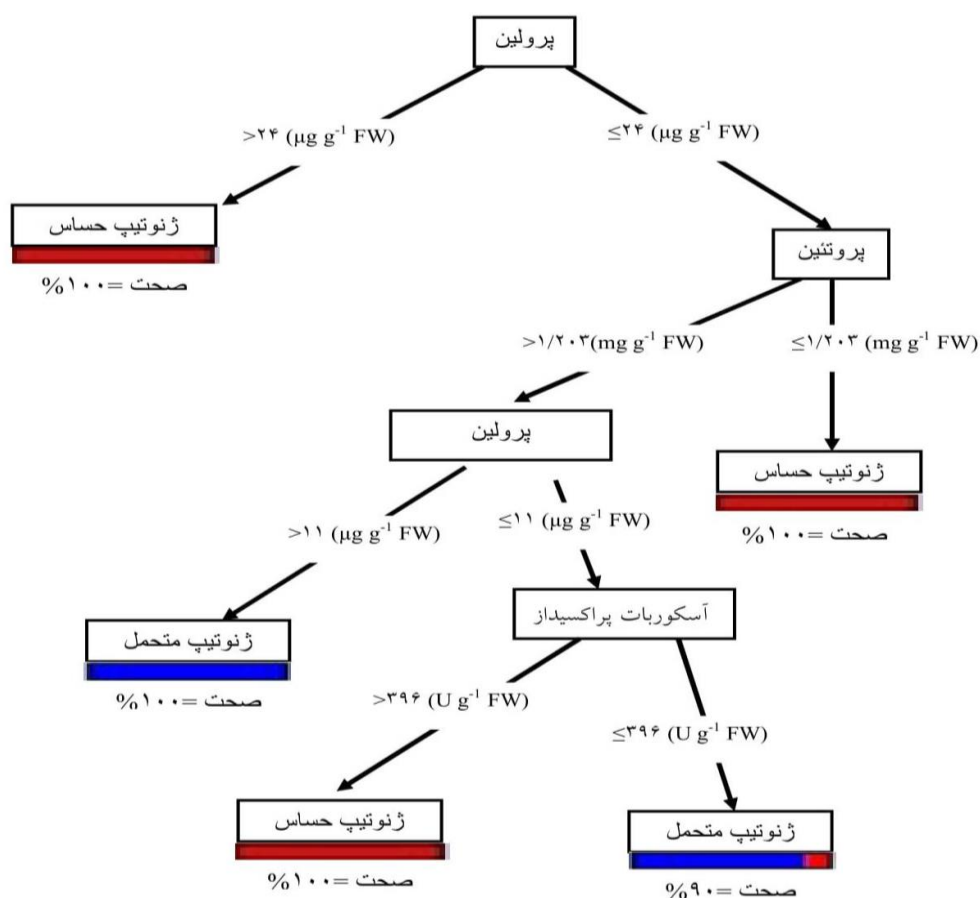
Table 3. Specificity, Sensitivity, Accuracy, and AUC (Area Under the ROC Curve) for Decision Trees with the highest performances in the prediction of salt sensitive and tolerant genotypes

پایگاه داده	مدل	صحت (%)	سطح زیر منحنی	دقت	حساسیت (%)	اختصاصی بودن (%)
Database	Model	Accuracy (%)	AUC	Precision	Sensitivity (%)	Specificity (%)
Info Gain	DT Parallel	90	0.6	1.0	70	100
Ratio	Gain Ratio					
Rule	DT Gain Ratio	90		0.8	80	90



شکل ۱. درخت تصمیم حاصل از اجرای مدل DT parallel Gain Ratio روی پایگاه داده‌ای Info Gain Ratio.

Figure 1. The decision tree generated by the DT parallel Gain Ratio model run on Info Gain Ratio dataset.



شکل ۲. درخت تصمیم حاصل از اجرای مدل DT Gain Ratio روی پایگاه داده‌ای Rule

Figure 2. The decision tree generated by the DT Gain Ratio model run on Rule dataset

سپاسگزاری

هزینه‌های این تحقیق توسط دانشگاه شیراز پرداخت شده است که موجب تشکر و قدردانی است.

References

- Bates LS, Waldren RP, Teare ID (1973) Rapid determination of free proline for water-stress studies. Plant Soil 39, 205-207.
- Becana M, Moran JF, Iturbe-Ormaetxe I (1998) Iron-dependent oxygen free radical generation in plants subjected to environmental stress: toxicity and antioxidant protection. Plant Soil 201, 137-147.
- Beiki AH, Saboor S, Ebrahimi M (2012) A new avenue for classification and prediction of olive cultivars using supervised and unsupervised algorithms. PLoS ONE 7, 44164.

- Bekkerman R, El-Yaniv R, Tishby N, Winter Y (2003) Distributional word clusters vs. words for text categorization. *J Mach Learn Res* 3, 1183-1208.
- Bradford MM (1976) A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal Biochem* 72, 248–254.
- Bradley AP (1997) The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognit* 30, 1145-1159
- Cadenas E (1989) Biochemistry of oxygen toxicity. *Annu Rev Biochem* 58, 79-110.
- Chance B, Maehly AC (1954) Assay of catalase and peroxidase. *Methods Biochem Anal* 1, 357-424.
- Dhindsa RS, Plumb-Dhindsa P, Thorpe TA (1981) Leaf senescence: correlated with increased levels of membrane permeability and lipid peroxidation, and decreased levels of superoxide dismutase and catalase. *J Exp Bot* 32, 93-101.
- Ebrahimi M, Ebrahimie E, Ebrahimi M (2009) Searching for patterns of thermo-stability in proteins and defining the main features contributing to enzyme thermo stability through screening, clustering, and decision tree algorithms. *EXCLI J* 8, 218-233.
- Ebrahimie E, Ebrahimi M, Sarvestani NR (2011) Protein attributes contribute to halo-stability, bioinformatics approach. *Saline Syst* 7, 1.
- Egneus H, Heber U, Matthiesen U, Kirk M (1975) Reduction of oxygen by the electron transport chain of chloroplasts during assimilation of carbon dioxide. *Biochim Biophys Acta* 408, 252-268.
- Fridovich I (1986) Biological effects of the superoxide radical. *Arch Biochem Biophys* 247, 1-11.
- Goudarzi M, Pakniyat H (2008) Evaluation of wheat cultivars under salinity stress based on some agronomic and physiological traits. *JASS* 4, 35-38.
- Hamann F, Gutmann H, Jecklin U et al. (2009) Development of decision tree models for substrates, inhibitors, and inducer of P-glycoprotein. *Curr Drug Metab* 10, 339-346.
- Hosseinzadeh F, Ebrahimi M, Goliaei B, Shamabadi N (2012) Classification of lung cancer tumors based on structural and physicochemical properties of proteins by bioinformatics models. *PLoS ONE* 7, 1-8
- Kerepesi I, Galiba G (2000) Osmotic and salt stress-induced alteration in soluble carbohydrate content in wheat seedlings. *Crop Sci* 40, 482-487.
- Kishor PK, Sangam S, Amrutha R et al. (2005) Regulation of proline biosynthesis, degradation, uptake and transport in higher plants: its implications in plant growth and abiotic stress tolerance. *Curr Sci* 88, 424-438.

- Nakano Y, Asada K (1981) Hydrogen peroxide is scavenged by ascorbate-specific peroxidase in spinach chloroplasts. *Plant Cell Physiol* 22, 867-880.
- Nasiri J, Naghavi MR, Kayvanjoo AH et al. (2015) Precision assessment of some supervised and unsupervised algorithms for genotype discrimination in the genus *Pisum* using SSR molecular data. *J Theor Biol* 368, 122–132.
- Reddy AR, Chaitanya KV, Vivekanandan M (2004) Drought-induced responses of photosynthesis and antioxidant metabolism in higher plants. *J Plant Physiol* 161, 1189-1202.
- Sahi C, Singh A, Blumwald E, Grover A (2006) Beyond osmolytes and transporters: novel plant salt-stress tolerance-related genes from transcriptional profiling data. *Physiol Plant* 127, 1-9.
- Szabados L, Savoure A (2010) Proline: a multifunctional amino acid. *Trends Plant Sci* 15, 89-97.
- Thapa G, Dey M, Sahoo L, Panda S (2011) An insight into the drought stress induced alterations in plants. *Biol Plant* 55, 603-613.
- Vinocur B, Altman A (2005) Recent advances in engineering plant tolerance to abiotic stress: achievements and limitations. *Curr Opin in Biotechnol* 16, 123-132.
- Wang S, Wan C, Wang Y et al. (2004) The characteristics of Na⁺, K⁺ and free proline distribution in several drought-resistant plants of the Alexa Desert, China. *J Arid Environ* 56, 525-539.
- Weston J, Elisseeff A, Scholkopf B, Tipping M (2003) Use of the zero norm with linear models and kernel methods. *J Mach Learn Res* 3, 1439-1461.
- Zhu JK (2000) Genetic analysis of plant salt tolerance using *Arabidopsis*. *Plant Physiol* 124, 941-948.
- Zinati Z, Zamansani F, Kayvanjoo Ah et al. (2014) New layers in understanding and predicting α -linolenic acid content in plants using amino acid characteristics of omega-3 fatty acid desaturase. *Comput Biol Med* 54, 14-23.
- Zinati Z, Alemzadeh A, Kayvanjoo AH (2016) Computational approaches for classification and prediction of P-type ATPase substrate specificity in *Arabidopsis*. *Physiol Mol Biol Plants* 22, 163-174.
- Iqbal N, Umar S, Khan NA, Khan MIR (2014) A new perspective of phytohormones in salinity tolerance: regulation of proline metabolism. *Environ Exp Bot* 100, 34e42.
- Kaur G, Asthir B (2015) Proline: a key player in plant abiotic stress tolerance. *Biol Plant* 59, 609e619.

- Reddy PS, Jogeswar G, Rasineni GK et al. (2015) Proline over-accumulation alleviates salt stress and protects photosynthetic and antioxidant enzyme activities in transgenic sorghum [*Sorghum bicolor* (L.) Moench]. *Plant Physiol Bioch* 94, 104e113.
- Naliwajski MR, Skłodowska M (2014) Proline and its metabolism enzymes in cucumber cell cultures during acclimation to salinity. *Protoplasma* 251, 201e209.
- Pioter Grazyana K (2005) Antioxidant defence in leaves of C3 and C4 plant under salinity stress. *Physiol Plant* 125, 31-40.
- Bollivar DW, Beale SL (1996) The Chlorophyll biosynthetic enzyme Mg-protoporphyrin IX monomethyl ester (oxidative) cyclase (characterization and partial purification from *Chlamydomonas reinhardtii* and *Synechocystis* sp. PCC 6803). *Plant Physiol* 1, 105-114.
- Patade VY, Lokhande VH, Suprasanna P (2014) Exogenous application of proline alleviates salt induced oxidative stress more efficiently than glycine betaine in sugarcane cultured cells. *Sugar Tech* 16, 22e29.
- Teh CY, Mahmood M, Shaharuddin NA, Ho CL (2015) In vitro rice shoot apices as simple model to study the effect of NaCl and the potential of exogenous proline and glutathione in mitigating salinity stress. *Plant Growth Regul* 75, 771-781.
- Mansour MF, Salama KH, Ali FZ, Hadid AF (2005) Cell and plant responses to NaCl in *Zea Mays* L. Cultivars differing in salt tolerance. *Gen App Plant Physiol* 31, 29-41.
- Akbari M, Toorchi M, Shakiba MR (2016) The effects of sodium chloride stress on proline content and morphological characteristics in wheat (*Triticum aestivum* L.). *Biol Forum* 8, 379-385.
- Witten IH, Frank E (2005) *Data mining: practical machine learning tools and techniques*, second edition, Elsevier: San Francisco. ISBN 0 -12-088407-0.
- Asadi M, Mohammadi-Nejad G, Golkar P et al. (2012) Assessment of salinity tolerance of different promising lines of bread wheat (*Triticum aestivum* L.). *Adv Appl Sci Res* 3, 1117-1121.
- Niazi A, Ramezani A, Dinari A (2014) GSTF1 gene expression analysis in cultivated wheat plants under salinity and ABA treatments. *MBRC* 3, 9 -19.
- Liu CG, Wu YW, Hou H et al. (2002) Value and utilization of alloplasmic common wheats with *Aegilops crassa* cytoplasm. *Plant Breed* 121, 407-410.
- Shaik R, Ramakrishna W (2013) Machine learning approaches distinguish multiple stress conditions using stress-responsive genes and identify candidate genes for broad resistance in rice. *Plant Physiol* 164, 481-95.

Furtauer L, Pschenitschnigg A, Scharhosi H et al. (2018) Combined multivariate analysis and machine learning reveals a predictive module of metabolic stress response in *Arabidopsis thaliana*. Mol Omics 14, 437-449.

Flowers TJ, Colmer TD, (2008) Salinity tolerance in halophytes. New Phytol 179, 945–963.

Slama I, Abdelly C, Bouchereau A et al. (2015) Diversity, distribution and the role of osmoprotective compounds accumulated in halophytes under abiotic stress. Ann Bot 115, 433-447.