

طراحی مدل‌های پیش‌بینی، جهت تعیین ساختار آنزیم استریول آسیل کریر پروتئین دیسچوراز ۱

علی سالاری^۱، سید مرتضی رضوی^۲، منصور ابراهیمی^{۳*}

چکیده

زمینه و هدف: اسیدهای چرب ضروری مانند آلفا لینولئیک اسید به‌عنوان یک امگا-۳ و آلفا لینولئیک اسید به‌عنوان یک امگا-۶ در سلول‌های انسانی تولید نمی‌شوند. بنابراین، به‌وسیله منابع غذایی مانند ماهی، روغن سویا، دانه کتان و دانه‌های آفتابگردان تأمین می‌شوند. مراحل تولید آلفا لینولئیک اسید توسط ۶ آنزیم صورت می‌گیرد که یکی از آنها آسیل کریر پروتئین دیسچوراز است. این مطالعه با هدف طراحی مدل‌های پیش‌بینی جهت تعیین ساختار ارگانیزمی آنزیم استریول آسیل کریر پروتئین دیسچوراز ۱ انجام شد.

روش بررسی: تعدادی از ابزارهای بیوانفورماتیک جهت تعیین مهم‌ترین مشخصه‌های ژنی آنزیم آسیل کریر پروتئین دیسچوراز به جهت توسعه مدل‌های پیش‌بینی ارگانیزمی اجرا شد. سپس از تکنیک‌های داده‌کاوی مانند

(Feature selection, Decision tree, Classification models) در جهت تولید الگوریتم‌های پیش‌بینی دقیق و کارآمد برپایه خصوصیات ژنی آنزیم (S-ACP-DES1) از ارگانیزم‌های مختلف استفاده شد.

یافته‌ها: مهم‌ترین متغیر ژنی در تشخیص ساختارهای ارگانیزمی آنزیم (S-ACP-DES1)، فراوانی طول بوده است. همچنین مدل‌های پیش‌بینی طراحی شده نشان داد مدل Naive Bayse با معیار FCdb با دقت ۹۷/۳۸٪ می‌تواند ارگانیزم آنزیم‌های جدید را براساس خصوصیات ژنی پیش‌بینی کند. دو یافته فوق برای اولین بار در این مطالعه گزارش شدند.

نتیجه‌گیری: نتایج این مطالعه نشان داد با استفاده از ابزارهای بیوانفورماتیک می‌توان به‌سهولت نسبت به دسته‌بندی آنزیم DESA1 براساس ارگانیزم آن اقدام کرد و متغیر طول ژن‌ها، بهترین شاخص برای این دسته‌بندی است. همچنین بهترین ماشین یادگیری مدل Bayesian در Naive Bayse با دقت بالای ۹۷/۳۸٪ جهت تعیین آنزیم DESA1 برای اولین بار در این مطالعه گزارش شد.

کلید واژه‌ها: زیست‌شناسی با استفاده از کامپیوتر؛ اسیدهای چرب امگا-۳؛ پروتئین استریول آسیل کریر.

^۱ دانشجوی کارشناس ارشد سلولی و مولکولی، باشگاه پژوهشگران جوان و نخبه، واحد بروجرد، دانشگاه آزاد اسلامی، بروجرد، ایران.

^۲ دانشجوی کارشناس ارشد ژنتیک، دانشکده علوم پایه، دانشگاه خوارزمی، تهران، ایران.

^۳ دانشیار فیزیولوژی، گروه پژوهشی بیوانفورماتیک، پژوهشکده سبز، دانشگاه قم، ایران.

* نویسنده مسئول مکاتبات:

منصور ابراهیمی، گروه پژوهشی بیوانفورماتیک، پژوهشکده سبز، دانشگاه قم، ایران؛

آدرس پست الکترونیکی:
mansoureb@chmail.ir

تاریخ دریافت: ۹۳/۵/۲۶

تاریخ پذیرش: ۹۳/۱۲/۱۴

لطفاً به این مقاله به‌صورت زیر استناد نمایید:

Salari A, Razavi SM, Ebrahimi M. Designing prediction models to determine the structure of Stearoyl-acyl carrier protein desaturase 1. Qom Univ Med Sci J 2015;9(8):13-20. [Full Text in Persian]

مقدمه

آنزیم استریول آسیل کریر پروتئین دیسچوراز، یکی از مهم‌ترین آنزیم‌هایی است که در تولید آلفا لینولئیک اسید (یکی از انواع مهم امگا-۳) نقش دارد (۱). عملکرد این آنزیم به‌صورتی است که زنجیره اشباع کربنی را به‌صورت دوگانه کردن به حالت غیراشباع درمی‌آورد (۲).

انسان قابلیت تولید اسیدهای چرب ضروری را ندارد، به‌همین علت به این نوع اسیدهای چرب، اسیدهای چرب ضروری گفته می‌شود که از جمله آنها می‌توان به linoleic acid (LA; C18:2 ω -6) و α -linolenic acid (ALA; C18:3 ω -3) اشاره کرد که به ترتیب تحت عنوان اسید چرب ضروری امگا-۶ و اسید چرب ضروری امگا-۳ نامگذاری می‌شوند (۳)، بیشترین مقدار آن، از روغن گل آفتابگردان و روغن ذرت لینولئیک اسید به دست می‌آید، همچنین در روغن دانه کتان، میزان آلفا لینولئیک اسید به‌وفور یافت می‌شود (۴). اسیدهای چرب به زنجیره‌ای گفته می‌شود که در یک انتهای آن کربوکسیل قرار دارد و به دو دسته اسید چرب اشباع و غیراشباع طبقه‌بندی می‌شود. بیشتر اسیدهای چرب غیراشباع براساس موقعیت اولین پیوند دوگانه‌شان به امگا-۳، ۶ و ۹ تقسیم‌بندی می‌شوند (۵). اسیدهای چرب امگا-۳ از اهمیت زیادی برخوردار هستند از جمله می‌توان به افزایش زمان خونریزی اشاره کرد که به‌علت وجود خاصیت ضدانعقادی اتفاق می‌افتد، همچنین باعث کاهش تجمع پلاکت‌ها، سیالیت خون، فیبروز و افزایش شکل‌پذیری و انعطاف گلبول‌های قرمز می‌شوند (۷)، که در نتیجه منجر به کاهش انعقاد خون می‌گردد. در هیچ‌یک از آزمایش‌های بالینی مانند جراحی‌های پیوند رگ‌های کرونری، مدرکی مبنی بر افزایش خونریزی به‌علت مصرف اسیدهای چرب امگا-۳ وجود ندارد. بیشتر مطالعات نشان می‌دهد اثرات امگا-۳ بر روی سرم چربی‌ها، به نوع بیماری و یا مقدار اسیدهای چرب اشباع‌شده در رژیم غذایی وابسته است. در بیمارانی با چربی خون بالا؛ امگا-۳ باعث کاهش کلسترول LDL (یک لیپوپروتئین با چگالی پایین) می‌شود (۶). در روند تشکیل آلفا لینولئیک اسید، آنزیم DESA1 با ایفا کردن نقش دوگانه کردن پیوند کربن شماره ۹ زنجیره طویل ۱۸ کربن، اولین قدم را برای تبدیل زنجیره اشباع به زنجیره غیراشباع آلفا لینولئیک اسید برمی‌دارد (۱۰-۷).

اهمیت این آنزیم در عملکرد تخصصی آن؛ یعنی اولین مرحله بیوسنتز اسیدهای چرب غیراشباع از استریک اسید است. هدف اصلی این مطالعه، یافتن مهم‌ترین مؤلفه‌های ژنی اختصاصی ارگانیزی آنزیم استریول آسیل کریر پروتئین دیسچوراز S-ACP-(DES1) با استفاده از ابزارهای بیوانفورماتیک و طراحی مدل‌های پیش‌بینی (برای اولین بار) جهت تعیین ارگانیزم‌های این آنزیم بوده که می‌توان با مدل‌های به‌دست آمده، عملکرد آنزیم‌های مجهول را از نظر غیراشباع‌سازی پیش‌بینی کرد که این نوع پیش‌بینی برای اولین بار گزارش شده است. بیوانفورماتیک یک چشم‌انداز جدید از رشته موردنظر پس از انقلاب قرن بیستم در حوزه‌های زیست‌شناسی مولکولی و محاسباتی، ارائه می‌دهد. تمرکز در علم بیوانفورماتیک برای استفاده از ابزارهای داده‌پردازی و الگوریتم‌های کامپیوتری برای آنالیز داده‌های بزرگ ژنی و پروتئینی است. چالش اساسی در زیست‌شناسی مولکولی، علاقه به محاسبه داده‌های مربوط به ساختار ژن، پروتئین و داده‌های مربوط به توالی ژن بوده که از داده‌های پر حجم ژنومی به دست می‌آید. داده‌های پر حجم زیستی توسط پروژه‌های توالی‌یابی ژنومی ایجاد شده و برای محاسبه و بررسی در داده‌های مذکور از مدل‌های مناسب کامپیوتری همچون درخت‌های تصمیم‌گیری استفاده می‌شود (۱۱). در این مطالعه نشان داده شد در صورتی که تمام جنبه‌های توالی‌های ژنومی توسط تعداد زیادی از خصوصیت آنها پوشش داده شود، تکنیک‌های داده‌پردازی Classification (Feature selection, Decision tree, methods) می‌تواند مدل‌های دقیق و کارآمد جدیدی تولید کنند (۱۷-۱۲).

روش بررسی

داده‌های ژنی آنزیم DESA1 از تمام ارگانیزم‌های کدکننده آنزیم به‌وسیله سایت KEGG (www.genome.jp/kegg/) استخراج گردید. سپس براساس تقسیم‌بندی کلی، جانداران به سه گروه اصلی گیاهان، آغازیان و باکتری‌ها تقسیم‌بندی شدند؛ بدین صورت که تعداد ۱۱۸ ژن به باکتری‌ها و ۶۶ ژن به گیاهان و تنها یک ژن به آغازیان تعلق گرفت. هیچ جانوری ژن آنزیم مورد مطالعه را کد نکرد و اساس انتخاب جامعه آماری در این مطالعه تمام ژن‌های موجود در همه ارگانیزم‌ها بود، که به‌وسیله ابزارهای

۹- اپراتوری که از ضرایب بردار نرمال که از SVM خطی

گرفته‌شده به‌عنوان ویژگی وزن‌دهی استفاده می‌کند

Weight by SVM (Support Vector Machine)

۱۰. اپراتوری که از فاکتورهای PCA اولیه، به‌عنوان ویژگی وزن‌دهی استفاده می‌کند

Weight by PCA (Principle Component Analysis)

پس از اجرا کردن این اپراتورها بر روی داده‌ها، به هر مشخصه ژنی براساس اهمیتی که در داده‌ها داشت عددی بین ۱-۰ تعلق گرفت. سپس از بین تمام مشخصه‌های ژنی که عددی بالاتر از ۰/۵ گرفته بودند، ویژگی‌های مشترک و مهم بین ژن‌ها انتخاب شدند.

آنگاه Supervise و Unsupervised Clustering بر روی آنها اجرا شد. چهار مدل درخت القایی (Induction) به‌نام‌های Decision Tree، Decision Tree Parallel، Decision Stump و Random Forest بر روی dataset‌هایی که در بالا توضیح داده شد، اجرا گردید (۱۴، ۱۸)، همچنین سه نوع از ماشین‌های یادگیری (SVM، Bayesian Model، Neural Decision Tree Network) نیز به‌منظور یافتن مدلی مناسب برای پیش‌بینی طبقه‌بندی‌های ناشناخته در آنزیم DESA1 براساس مشخصه‌های ژنی محاسبه‌شده برای گروه‌های جانداران اجرا شد (۱۲، ۱۵، ۱۷، ۱۸).

یافته‌ها

قبل از اجرای مدل‌ها، داده‌های موجود نرمال شدند، بدین صورت که تمام داده‌ها براساس اهمیت‌شان، امتیازی بین ۱-۰ گرفتند که اعداد آنها بدین صورت بود:

PCA=۳، SVM=۱۲، Relief=۱، Uncertainty=۴، Gini =۵،
index=۹، Chi squared=۴، Deviation=۶، Rule=۵، Info gain=۵ و
Info gain ratio=۵

از سوی دیگر، مهم‌ترین مشخصه‌های ژنی که بر این اساس مشاهده شدند عبارت بود از: متغیر Length که تعداد ۹ الگوریتم بالای ۰/۵ را به خود اختصاص داد. پس از آن تعداد Count of CA با ۷ امتیاز و تعداد Count of CT با ۷ امتیاز

متفاوت بیوانفورماتیکی و نرم‌افزارهایی مانند

CLC Bio (CLC Bio Finlandsgade 10-12, Katrinebjerg 8200 Aarhus N (Denmark) فراوانی نوکلئوتیدها، عناصر متفاوت و تأثیرگذار در ژن‌ها (کربن، نیتروژن، فسفر، اکسیژن و هیدروژن)، همچنین طول ژن‌ها محاسبه گردید. برای شناسایی مهم‌ترین مشخصه‌های ژنی و یافتن الگوهای احتمالی در آنها، داده‌ها توسط نرم‌افزار RapidMiner (RapidMiner 5.0.001, Rapid-I GmbH, Stochumer Str.) (475, 44227 Dortmund, Germany) نسخه ۵ در برابر ۱۰ الگوریتم وزنی متفاوت قرار گرفت که عبارت بود از:

۱- اپراتوری که ارتباط یک صفت را با شمارش افزایش اطلاعات در توزیع سه گروه محاسبه می‌کند:
(Weight by Information Gain)

۲- اپراتوری که ارتباط یک صفت را با شمارش نسبت افزایش اطلاعات در توزیع سه گروه محاسبه می‌کند:
(Weight by Information Gain ratio)

۳- اپراتوری که ارتباط یک صفت را با محاسبه میزان خطا از یک مدل بر روی یک مثال بدون در نظر گرفتن آن صفت خاص، حساب می‌کند (Weight by Rule).

۴- اپراتوری که با ملاک قرار دادن انحراف از معیار برای تمام مشخصه‌ها چه میانگین آنها، چه حداقل و حداکثر آنها، وزن‌دهی می‌کند (Weight by Deviation).

۵- اپراتوری که ارتباط یک صفت را با تک‌تک صفت‌های دیگر، از داده واردشده محاسبه می‌کند (Weight by Chi squared statistic).

۶- اپراتوری که ارتباط یک صفت را بر پایه شاخص Gini از توزیع هر گروه سنجش می‌کند (Weight by Gini index).

۷- اپراتوری که ارتباط یک صفت را به‌وسیله Symmetrical Uncertainty با در نظر گرفتن گروه اندازه‌گیری می‌کند (Weight by Uncertainty).

۸- اپراتوری که ارتباط یک صفت را توسط مثال‌های نمونه و مقایسه ارزش صفات صحیح برای مثال نزدیکتر به همان و به گروه متفاوت اندازه گرفته و داده‌ها را بین صفر و یک نرمال می‌کند (Weight by Relief).

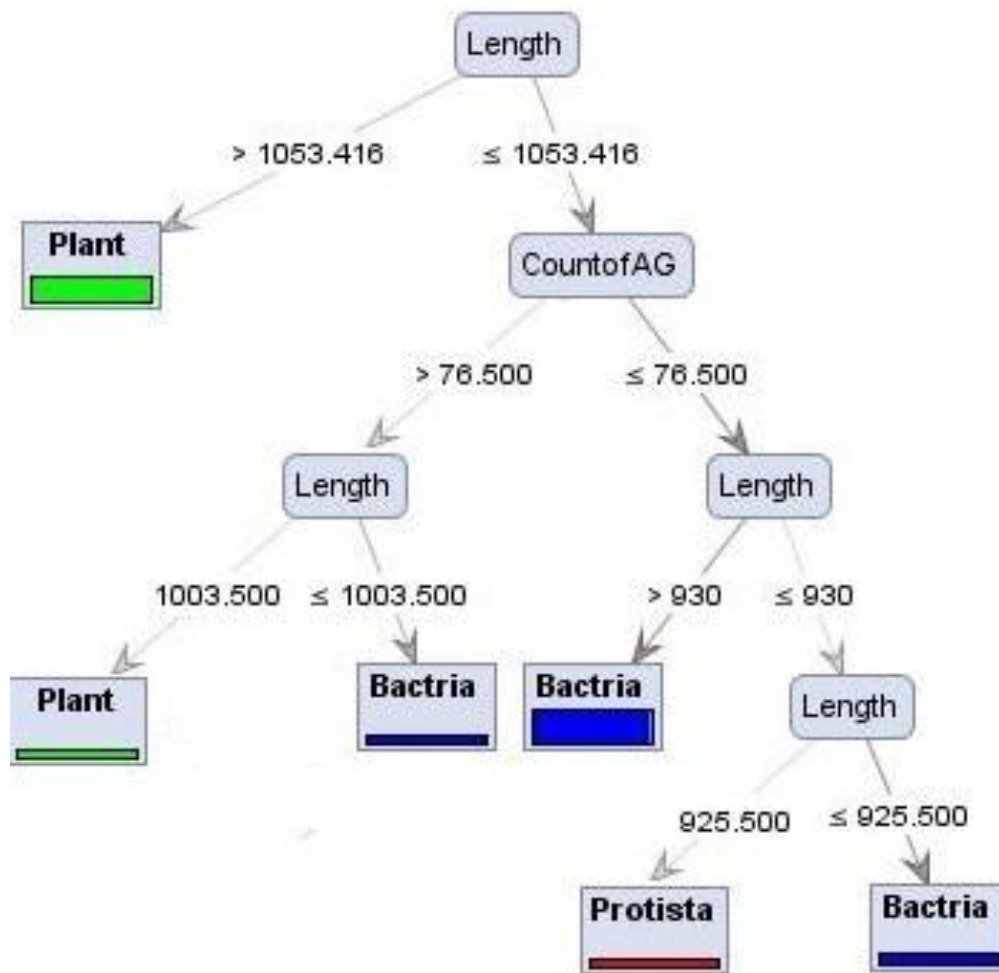
قرار گرفتند و به همین ترتیب تعداد Count of GA و Count of AG با امتیاز ۶ دررتبه بعدی بودند (جدول شماره ۱).

جدول شماره ۱: متغیرهای ژنی که توسط مدل‌های مختلف وزن‌دهی به‌عنوان متغیرهای مهم (وزن بالاتر از ۰/۵) معرفی شده‌اند.

متغیر	تعداد مدل‌های وزن‌دهی که متغیر را مهم ارزیابی کردند
Length	۹
Count of CA	۷
Count of CT	۷
Count of GA	۶
Count of AG	۶
Count of Cytosine	۴
Count of Guanine	۳
salt0.1M	۲

ژنی آنزیم DESA1 از ۱۰۵۳ بزرگتر باشد توالی مربوط به یک گیاه خواهد بود و اگر طول توالی کوچکتر یا مساوی ۱۰۵۳ به اضافه اینکه تعداد AG بزرگتر از ۷۶۵۰۰ و طول توالی نوکلئوتیدی آنزیم، بزرگتر از ۱۰۰۳ باشد، آنزیم DESA1 متعلق به یک گیاه خواهد بود و اگر کوچکتر یا مساوی ۱۰۰۳ باشد آنزیم متعلق به باکتری خواهد بود. اما اگر تعداد نوکلئوتیدهای AG، کوچکتر یا مساوی ۷۶۵۰۰ و صفت طول توالی، بزرگتر از ۹۳۰ باشد توالی آنزیم DESA1 مربوط به یک باکتری می‌باشد همچنین اگر طول توالی کوچکتر یا مساوی ۹۳۰ و دوباره طول توالی نوکلئوتیدی آنزیم DESA1، بزرگتر از ۹۲۵/۵۰۰ باشد توالی آنزیم مربوط به آغازیان خواهد بود و در صورتی که کوچکتر و مساوی ۹۲۵/۵۰۰ باشد، ژن آنزیم DESA1 متعلق به یک باکتری گزارش خواهد شد. با استفاده از این درخت می‌توان ارگانیزم مجهول توالی‌هایی که صفات و مشخصه‌های غیراشباع‌سازی را دارند شناسایی کرد. این ابزار قابلیت آن را دارد که برای هر تشخیصی استفاده شود و احتمالاً می‌توان این روش را در تشخیص بیماری‌ها نیز به کار برد (شکل).

مهم‌ترین مشخصات ژنی توسط الگوریتم‌های متفاوت وزن‌دهی (PCA, SVM, Relief, Uncertainty, Gini index, Chi Squared, Deviation, Rule, Information Gain و Information Gain Ratio)، به‌عنوان مشخصه‌های مهم معرفی شدند. درخت‌های تصمیم‌گیری بر مبنای خصوصیات ژنی که از توالی ژن آنزیم به دست می‌آید در ارگانیزم‌های متفاوت، مدلی را طراحی می‌کنند. برای تشخیص و پیش‌بینی نوع آنزیم مجهول براساس توالی ژنی آن، در این مطالعه از میان ۵۶۰ درختی که به دست آمد تعداد زیادی از آنها توانستند تا حدود زیادی تمام گروه‌ها را از هم تفکیک کنند و اکثر درختان به‌صورت پیچیده و بزرگ ایجاد شده بودند. به دلیل اهمیت صفت طول توالی در آنزیم DESA1، اکثر درخت‌ها براساس صفت طول، کشیده شده و طبیعتاً بهترین درخت‌هایی که توانستند تفکیک خوبی بین گروه‌ها داشته باشند. از مؤلفه طول استفاده کرده بودند. برای نمونه، یکی از درخت‌های ترسیم‌شده براساس خصوصیت طول ارائه شده است. این درخت بر مبنای صفت طول به مهم‌ترین خصوصیت از ۱۸۵ توالی استخراج‌شده از آنزیم DESA1 به‌پیش‌بینی طبقه‌بندی پرداخته است، بدین معنا که اگر طول توالی



شکل: درخت تصمیم‌گیری طراحی شده براساس متغیرهای ژنی برای جداسازی ارگانیزم‌های مختلف

حجم رنگ بستگی به تعداد ژن‌هایی دارد که در هر خانه قرار گرفته است؛ بدین معنا که حجم رنگ آبی باکتری در خانه‌ای بزرگتر از رنگ آبی باکتری دیگر است. بدین جهت تعداد ژن‌های یافت‌شده در نمونه این مطالعه باکتری با توجه به مؤلفه طول در آن خانه بیشتر بوده است؛ یعنی احتمال قرارگیری باکتری مجهول در آن بیشتر بوده، همچنین آنهایی که بزرگتر هستند اهمیت بیشتری برای آن گونه خاص دارند به دلیل اینکه ژن‌های بیشتری را در خود جای داده‌اند.

کمترین درصد دقت در مدل Naive Bayesian از کمترین؛ یعنی ۶۱٪ (برای Relief) تا بهترین؛ یعنی ۹۷/۳۸٪ برای FCdb به دست آمد، درحالی‌که برای Naive Kernel از کمترین؛ یعنی ۶۴٪ برای Relief dataset تا بهترین؛ یعنی ۹۷٪ برای FCdb به دست آمد. میانگین نتایج برای مدل Neural Network در الگوریتم Neural Net، ۹۲/۰۷٪ و برای الگوریتم AutoMLP، ۹۱/۵۷٪ برآورد شد. میزان دقت برای الگوریتم Neural Net از ۶۷/۴۹٪ برای Relief تا ۹۶/۲۳٪ برای FCdb به دست آمد، همچنین در الگوریتم AutoMLP از ۶۴/۲۷٪ برای Relief Dataset تا ۶۹/۲۳٪ برای Dataset Gini index حاصل شد که بالاترین میزان دقت را در مدل Neural Network داشت.

در قسمت ماشین‌های یادگیری (Decision Tree)، بهترین دقت عملکرد (۹۶٪) مربوط به مدل Decision Tree Random Forest بر روی Gini Index و بدترین دقت محاسبه مربوط به مدل Parallel Info Gain Decision tree بر روی قسمت Relief اجرا شده بود. برای SVM، بهترین دقت (۹۴/۰۹٪) بر روی همه الگوها، بجز Relief و SVM اجرا شد و بدترین دقت مربوط به Evolutionary، SVM، Lib SVM و SVM Linear با ۶۳٪ هر چهار مدل بر روی Relief اجرا گردید.

بحث

در مطالعه حاضر، تعدادی الگوریتم و مدل‌های پیش‌بینی برای یافتن بهترین آنها با روش ذکر شده انجام گرفت. در این مطالعه، یافته مهم متغیر به جهت دسته‌بندی آنزیم DESA1 براساس ارگانیزم، متغیر طول ژن‌ها از بهترین شاخص برای این دسته‌بندی بود. این یافته برای اولین بار در این مقاله گزارش گردید، همچنین برای اولین بار سیستم‌های پیش‌بینی براساس ماشین‌های یادگیری مورد استفاده قرار گرفت، نتایج نشان داد از سه مدل ماشین‌های یادگیری استفاده شده (Neural network، SVM، Bayesian) برای مدل‌های Neural network و SVM به ترتیب دقت‌های ۹۶/۲۳٪ و ۹۴/۰۴٪ پیش‌بینی شده است، اما مدل Bayesian بهترین دقت را در Naive Bayse به روی معیار FCdb با دقت بالای ۹۷/۳۸٪ نسبت به تعیین ارگانیزم آنزیم فوق به دست آورده، که نشان‌دهنده میزان خطای کم ماشین یادگیری Naive Bayse هنگام اجرا به روی معیار FCdb می‌باشد که در پیش‌بینی آنزیم استریول آسیل کریر پروتئین دیسچوراز ۱ عمل می‌کند. یافته‌های این مطالعه می‌تواند در طراحی ساختارهای آنزیمی جدید و شناخته‌نشده مورد استفاده قرار گیرد و نقش این آنزیم (نقش تبدیل زنجیره‌ای با پیوندهای یگانه به زنجیره‌ای با پیوندهای دوگانه) که همان تبدیل زنجیره اشباع به غیراشباع و یکی از مهم‌ترین مسئله در روغن‌های امروزی است را پیش‌بینی کند (۷). شاید بتوان از این مؤلفه‌ها، نسبتی با عملکرد آنزیم ایجاد کرد و در طراحی سه بُعدی و ساخت آنزیم‌های مصنوعی از آنان استفاده نمود.

این روش تشخیصی تنها برای این آنزیم مفید نبوده و برای آنزیم‌های دیگر؛ حتی تشخیص مؤلفه‌های ژنی سرطان‌های گوناگون در قدم اول پیش‌بینی و در قدم دوم تشخیص آنها، به محقق کمک خواهد کرد. همچنین در آینده می‌توان از این روش در درمان نیز استفاده کرد، و آن زمانی است که بتوانیم بعد از تشخیص مؤلفه‌های ژنی سرطان، با روش‌هایی، مؤلفه‌های ژنی را تغییر داده و درمان را انجام دهیم (۱۹، ۲۰).

نتیجه‌گیری

نتایج این مطالعه نشان داد با استفاده از ابزارهای بیوانفورماتیکی می‌توان به سهولت نسبت به دسته‌بندی آنزیم DESA1 براساس ارگانیزم آن اقدام کرد که متغیر طول ژن‌ها، بهترین شاخص برای این دسته‌بندی است. این یافته برای اولین بار در این مقاله گزارش گردید، همچنین بهترین ماشین یادگیری مدل Bayesian در Naive Bayse با دقت بالای ۹۷/۳۸٪ جهت تعیین آنزیم DESA1 برای اولین بار در این مطالعه گزارش شده است.

تشکر و قدردانی

بدین وسیله از جناب آقای دکتر بختیارزاده و سرکار خانم شمع‌آبادی و سرکار خانم فاطمه سالاریان و جناب امیرحسین کیوانجو، همچنین پژوهشکده سبز دانشگاه قم به جهت حمایت‌هایی که از طرح فوق کردند کمال تشکر را داریم.

References:

1. Martz F, Kiviniemi S, Palva TE, Sutinen ML. Contribution of omega-3 fatty acid desaturase and 3-ketoacyl-ACP synthase II (KASII) genes in the modulation of glycerolipid fatty acid composition during cold acclimation in birch leaves. *J Exp Bot* 2006;57(4):897-909.
2. McKeon TA, Stumpf PK. Purification and characterization of the stearyl-acyl carrier protein desaturase and the acyl-acyl carrier protein thioesterase from maturing seeds of safflower. *J Biol Chem* 1982;257(20):12141-7.
3. Nelson GJ, Schmidt PS, Bartolini GL, Kelley DS, Kyle D. The effect of dietary docosahexaenoic acid on platelet function, platelet fatty acid composition, and blood coagulation in humans. *Lipids* 1997;32(11):1129-36.
4. Tranchant T, Besson P, Hoinard C, Pinault M, Alessandri JM, Delarue J, et al. Long-term supplementation of culture medium with essential fatty acids alters alpha-linolenic acid uptake in Caco-2 clone TC7. *Can J Physiol Pharmacol* 1998;76(6):621-9.

5. Colin A, Reggers J, Castronovo V, Anseau M. Lipids, depression and suicide. *Encephale* 2003;29(1):49-58.
6. Allayee H, Roth N, Hodis HN. Polyunsaturated fatty acids and cardiovascular disease: Implications for nutrigenetics. *J Nutrigenet Nutrigenomics* 2009;2(3):140-8.
7. Sanders TA, Lewis F, Slaughter S, Griffin BA, Griffin M, Davies I, et al. Effect of varying the ratio of n-6 to n-3 fatty acids by increasing the dietary intake of alpha-linolenic acid, eicosapentaenoic and docosahexaenoic acid, or both on fibrinogen and clotting factors VII and XII in persons aged 45-70 y: The OPTILIP study. *Am J Clin Nutr* 2006;84(3):513-22.
8. Simopoulos AP. Omega-3 fatty acids in health and disease and in growth and development. *Am J Clin Nutr* 1991;54(3):438-63.
9. Schweizer HP, Choi KH. *Pseudomonas aeruginosa* aerobic fatty acid desaturase DesB is important for virulence factor production. *Arch Microbiol* 2011;193(3):227-34.
10. Tilton GB, Shockey JM, Browse J. Biochemical and molecular characterization of ACH2, an acyl-CoA thioesterase from *Arabidopsis thaliana*. *J Biol Chem* 2004;279(9):7487-94.
11. McDonough VM, Stuke JE, Martin CE. Specificity of unsaturated fatty acid-regulated expression of the *Saccharomyces cerevisiae* OLE1 gene. *J Biol Chem* 1992;267(9):5931-6.
12. Niot I, Gresti J, Boichot J, Semporé G, Durand G, Bézard J, et al. Effect of dietary n-3 and n-6 polyunsaturated fatty acids on lipid-metabolizing enzymes in obese rat liver. *Lipids* 1994;29(7):481-9.
13. Yue JC, Murray K. A Similarity measure based on species proportions. *Communications in statistics. Theory Methods* 2005;34(11):2123-31.
14. Du J-X, Wang X-F, Zhang G-J. Leaf shape based plant species recognition. *Appl Mathemat Comput* 2007;185(2):883-93.
15. Ke Y, Quackenbush LJ, Im J. Synergistic use of Quick Bird multispectral imagery and LIDAR data for object-based forest species classification. *Remote Sens Environ* 2010;114(6):1141-54.
16. Parthiban V, Gromiha MM, Schomburg D. Prediction of protein stability upon point mutations. *Nucleic Acids Res* 2006;34(Web Server issue):W239-42.
17. Ebrahimie E, Ebrahimi M, Sarvestani NR, Ebrahimi M. Protein attributes contribute to halo-stability, bioinformatics approach. *Saline Systems* 2011;7(1):1.
18. Breast Feeding, Social Class, and Change Over Time. *Am J Public Health Nations Health* 1959;49(3):365-6.
19. Hosseinzadeh F, Ebrahimi M, Goliaei B, Shamabadi N. Classification of lung cancer tumors based on structural and physicochemical properties of proteins by bioinformatics models. *PLoS One* 2012;7(7):e40017.
20. Ebrahimi M, Lakizadeh A, Agha-golzadeh P, Ebrahimi E, Ebrahimi M. Prediction of thermostability from amino acid attributes by combination of clustering with attribute weighting: A new vista in engineering enzymes. *PLoS One* 2011;6(8):e23146.
21. Ashrafi E, Alemzadeh A, Ebrahimi M, Ebrahimie E, Dadkhodaei N, Ebrahimi M. Amino acid features of P1B-ATPase heavy metal transporters enabling small numbers of organisms to cope with heavy metal pollution. *Bioinform Biol Insights* 2011;5:59-82.
22. Ebrahimi M, Ebrahimi E, Shamabadi N, Ebrahimi M. Are there any differences between features of proteins expressed in malignant and benign breast cancers? *J Res Med Sci* 2010;15(6):299-309.
23. Habashy HO, Powe DG, Glaab E, Ball G, Spiteri I, Krasnogor N, et al. RERG (Ras-like, oestrogen-regulated, growth-inhibitor) expression in breast cancer: A marker of ER-positive luminal-like subtype. *Breast Cancer Res Treat* 2011;128(2):315-26.

Designing Prediction Models to Determine the Structure of Stearoyl-Acyl Carrier Protein Desaturase 1

Ali Salari¹, Seyed Morteza Razavi², Mansour Ebrahimi^{3*}

¹MSc Student of Cellular & Molecular Biology, Young Researchers & Elite Club, Borujerd Branch, Islamic Azad University, Borujerd, Iran.

²MSc Student of Genetics, Faculty of Science, University of Al-Khwarizmi, Tehran, Iran.

³Associate Professor of Physiology, Bioinformatics Research Group, Green Research Institute, University of Qom, Iran.

*Corresponding Author:
Mansour Ebrahimi,
Bioinformatics Research Group, Green Research Institute, University of Qom; Qom, Iran.

Email:
mansoureb@chmail.ir

Received: 17 Aug, 2014

Accepted: 5 Mar, 2015

Abstract

Background and Objectives: Essential fatty acids, such as α -linolenic acid (ALA), as an omega-3 fatty acid, and linoleic acid (LA), as an omega-6 fatty acid cannot be synthesized by human cells. Therefore, they should be provided from food sources, such as fish, soybean oil, flaxseed, and sunflower seeds. The process of α -linolenic acid synthesis is performed by 6 enzymes, one of which is stearoyl-acyl-carrier-protein-desaturase (S-ACP-DES). This study was carried out aiming at designing prediction models to determine the structure of stearoyl-acyl carrier protein desaturase 1.

Methods: Some bioinformatics tools were used to determine the most important gene features of S-ACP-DES enzyme in order to develop some organism prediction models. Then, data mining techniques, such as feature selection, decision tree, and classification models, were used to develop efficient and accurate prediction algorithms based on gene features of enzyme (S-ACP-DES) from various organisms.

Results: The most important gene feature in the identification of organism structures of the enzyme (S-ACP-DES) was frequency of length. Also, the designed prediction models indicated that Naive Bayes model with FCdb criteria and accuracy of 97.83% can predict new organism enzymes based on the gene features. The above findings have been reported for the first time in this study.

Conclusion: The results of this study showed that with the use of bioinformatics tools, we can classify DESA1 enzyme based on its organism, and the variable of the length of genes is the best indicator for this classification. Also, the best learning machine model in Naive Bayes was Bayesian with high accuracy of 38.97% for determination of DESA1 enzyme, which has been reported for the first time in this study.

Keywords: Stearoyl-acyl carrier protein-desaturase; Computational biology; Fatty acids, Omega-3.