

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ



دانشگاه قم

دانشکده فنی و مهندسی

پایان نامه دوره کارشناسی ارشد مهندسی فن آوری اطلاعات (IT)

عنوان:

شناسایی ژن های مؤثر در بروز بیماری آلزایمر در

مراحل اولیه با استفاده از ابزارهای داده کاوی

استاد راهنما:

دکتر منصور ابراهیمی

نگارنده:

ماریه براتی

تابستان/۹۳



«صورت جلسه دفاع از پایان نامه کارشناسی ارشد»

با تأییدات خداوند متعال و با استعانت از حضرت ولی عصر (عجل الله تعالی فرجه الشریف)

جلسه دفاعیه پایان نامه کارشناسی ارشد آقای / خانم: رسته:

تحت عنوان:

با حضور هیأت داوران در محل دانشگاه قم در تاریخ: / / تشکیل گردید.

در این جلسه، پایان نامه با نمره (به عدد، به حروف و

بدرجه: عالی بسیار خوب خوب قابل قبول مورد دفاع قرار گرفته

نام و نام خانوادگی	سمت	مرتبۀ علمی	امضاء
	استاد راهنما		
	استاد مشاور		
	استاد ناظر		
	استاد ناظر		
	نماینده کمیته تحصیلات تکمیلی		

مدیر امور آموزش و تحصیلات تکمیلی
نام و امضاء

معاون آموزشی و پژوهشی دانشکده
نام و امضاء

تقدیم به سه وجود مقدس:

آمان که ناتوان شد تا ما به توانایی برسیم...

مویشتان سپید شد تا ما رو سفید شویم...

وحاشا که سوختند تا که ما بخش وجود ما و رو سنگر را جان باشند..

پدر عزیز

مادر مهربان

و استاد کران قدر جناب آقای دکتر ابراهیمی

تقدیر و تشکر:

سپاس خدای را که سخنوران، در ستودن او بمانند و شمارندگان، شمردن نعمت‌های او ندانند و کوشندگان، حق او را گزاردن نتوانند. و سلام و دورود بر محمد و خاندان پاک او، طاهران معصوم، هم آنان که وجودمان وامدار وجودشان است؛ و نفرین پیوسته بر دشمنان ایشان تا روز رستاخیز...

بدون شک جایگاه و منزلت معلم، اجل از آن است که در مقام قدردانی از زحمات بی‌شائبه‌ی او، با زبان قاصر و دست ناتوان، چیزی بنگاریم.

اما از آنجایی که تجلیل از معلم، سپاس از انسانی است که هدف و غایت آفرینش را تأمین می‌کند و سلامت امانت‌هایی را که به دستش سپرده‌اند، تضمین؛ بر حسب وظیفه و از باب " من لم یشکر المنعم من المخلوقین لم یشکر الله عزّ و جلّ؛"

از پدر و مادر عزیزم، این دو معلم بزرگوام، که همواره بر کوتاهی و درستی من، قلم عفو کشیده و کریمانه از کنار غفلت‌هایم گذشته‌اند و در تمام عرصه‌های زندگی یار و یآوری بی چشم‌داشت برای من بوده‌اند؛ از استاد با کمالات و شایسته؛ جناب آقای دکتر ابراهیمی که در کمال سعہ صدر، با حسن خلق و فروتنی، از هیچ کمکی در این عرصه بر من دریغ نمودند و زحمت راهنمایی این رساله را بر عهده گرفتند؛ از گرامی، سرکار خانم دکتر فتوحی، مدیریت محترم کرسی گروه، که زحمت مشاوره این رساله را در حالی متقبل شدند که بدون مساعدت ایشان، این پروژه به نتیجه مطلوب نمی‌رسید؛ کمال تشکر و قدردانی رادارم.

باشد که این خردترین، بخشی از زحمات آنان را سپاس گوید.

چکیده:

بیماری آلزایمر یکی از شکل‌های جنون در دوره‌ی پیری است، این بیماری از جمله بیماری‌های غیرقابل درمانی است که به‌طور معمول در دهه‌های هفتم عمر انسان علائم خود را نشان می‌دهد. این بیماری ممکن است، سال‌ها بدون بروز علائم در بدن فرد وجود داشته باشد. با توجه به انجام تحقیقات بسیار زیاد بر روی این بیماری همچنان علل بروز آن مجهول مانده است. بیماری آلزایمر یا بر اساس توارث و یا به‌طور اتفاقی در فرد مسن ایجاد می‌شود. در این پایان‌نامه از کتابخانه‌های ریزآرایه مربوط به بیماران آلزایمر استفاده شد تا ژن‌هایی که در افراد مبتلابه این بیماری دچار افزایش و یا کاهش بیان شده‌اند شناسایی شود، لیست ژن‌های به‌دست‌آمده با سایر تحقیق‌هایی که تاکنون انجام شده مشابهت بسیار بالایی دارد و تنها تفاوت‌هایی در میزان بیان ژن‌ها وجود دارد که با توجه به تفاوت در نمونه‌ها امری طبیعی است. بعد از یافتن ژن‌های دچار تغییر بیان، نوبت به یافتن الگوهای خاص در توالی‌های ژن‌ها می‌رسد که این کار با استفاده از روش‌های وزن دهی انجام شد، توالی‌های AATTG، GAATAT، AGCCTG، TGCCCC با ۸ تکرار به‌عنوان مهم‌ترین توالی‌ها شناخته شدند. در نهایت الگوریتم‌های داده‌کاوی بی‌زین، ماشین بردار پشتیبان، خوشه‌بندی، شبکه‌ی عصبی و درخت تصمیم روی ۱۱ پایگاه داده (۱۰ پایگاه داده حاصل از الگوریتم‌های وزن دهی و FC پایگاه داده‌ی اصلی) اجرا شد. الگوریتم‌های شاخه ماشین بردار پشتیبان بهترین عملکرد را با مقدار ۰/۸۹ داشتند و پایین‌ترین عملکرد را الگوریتم‌های گروه خوشه‌بندی با میانگین عملکرد ۰/۷۱ داشتند. با اجرای الگوریتم‌های درخت تصمیم روی پایگاه داده‌ها به این نتیجه رسیدیم که توالی AATTG در بیماری آلزایمر یک بیومارکر مهم است که درخت تصمیم با استفاده از این توالی با دقت بالای ۰/۹۵ قادر به جدا کردن ژن‌ها به دو گروه افزایش و کاهش بیان می‌شود.

کلمات کلیدی: آلزایمر، بروز بیماری، ژن، داده‌کاوی

فهرست مطالب

عنوان	صفحه
فصل اول : کلیات تحقیق	۲
۱-۱. مقدمه :	۳
۲-۱. موضوع تحقیق :	۴
۳-۱. ضرورت انجام تحقیق:	۵
۴-۱. جنبه جدید بودن و نوآوری طرح در چیست؟	۶
۵-۱. فرضیه‌ها:	۶
۶-۱. قلمرو مکانی و زمانی تحقیق: (جامعه آماری و تعداد نمونه)	۶
۷-۱. هدف‌ها و کاربردهای مورد انتظار از انجام تحقیق:	۶
۸-۱. محدودیت های تحقیق :	۶
۹-۱. ساختار پایان‌نامه :	۷
فصل دوم: شرح موضوع	۸
۱-۲. مقدمه:	۹
۲-۲. بیماری آلزایمر:	۹
۱-۲-۲. آمار :	۱۰
۲-۲-۲. تشریح:	۱۱
۳-۲-۲. مشکلات روانی بیماران :	۱۳
۴-۲-۲. مشکلات داروهای ضدافسردگی:	۱۴
۵-۲-۲. پیشرفت بیماری :	۱۵
۶-۲-۲. تشخیص :	۱۶
۷-۲-۲. درمان و جلوگیری از پیشرفت:	۱۶
۸-۲-۲. گیاه درمانی:	۱۸
۳-۲. داده‌کاوی:	۱۸
۱-۳-۲. مراحل داده‌کاوی:	۲۰
۲-۳-۲. تاریخچه داده‌کاوی:	۲۲
۳-۳-۲. فن‌های کشف دانش در داده‌کاوی:	۲۳

- ۲۳.....۱-۳-۳-۲. شبکه‌های عصبی و کاربرد آن‌ها:.....
- ۲۳.....۲-۳-۳-۲. الگوریتم‌های معماری و برنامه‌های کاربردی آن:.....
- ۲۴.....۳-۳-۳-۲. رویکرد مبتنی بر پیش‌بینی‌های پویا و برنامه‌های کاربردی آن:.....
- ۲۴.....۴-۳-۳-۲. تجزیه و تحلیل معماری سیستم و برنامه‌های کاربردی آن:.....
- ۲۵.....۵-۳-۳-۲. سیستم عامل اطلاعاتی و برنامه‌های کاربردی :.....
- ۲۵.....۶-۳-۳-۲. مدل‌سازی و کاربردهای آن:.....
- ۲۵.....۷-۳-۳-۲. سیستم‌های مبتنی بر دانش و کاربرد آن‌ها:.....
- ۲۶.....۸-۳-۳-۲. علوم و مهندسی:.....
- ۲۶.....۴-۲-۴-۲. داده‌کاوی در پزشکی:.....
- ۲۷.....۱-۴-۲. استفاده از فن‌های داده‌کاوی در پزشکی:.....
- ۲۹.....۲-۴-۲. پیش‌بینی و تشخیص بیماری‌ها :.....
- ۲۹.....۳-۴-۲. اثربخشی معالجات :.....
- ۳۰.....۴-۴-۲. مدیریت خدمات بهداشتی :.....
- ۳۰.....۵-۴-۲. مدیریت ارتباط با مشتری :.....
- ۳۱.....۶-۴-۲. کشف تقلب و سوءاستفاده:.....
- ۳۳..... فصل سوم: مرور ادبیات :.....
- ۳۳.....۱-۳. مقدمه :.....
- ۳۳.....۲-۳. مرور ادبیات آرایه‌ها و ریزآرایه :.....
- ۴۷.....۳-۳. مرور ادبیات آرایه‌ها و داده‌کاوی :.....
- ۵۴..... فصل چهارم: روش انجام تحقیق.....
- ۵۵.....۱-۴. مقدمه:.....
- ۵۶.....۲-۴. کتابخانه‌های ریزآرایه:.....
- ۵۶.....۱-۲-۴. جست‌وجوی کتابخانه‌های ریزآرایه:.....
- ۵۷.....۲-۲-۴. دانلود کتابخانه :.....
- ۵۹.....۳-۴. نرم‌ال‌سازی داده‌ها توسط نرم‌افزار کنسول اکسپرن :.....
- ۵۹.....۱-۳-۴. وارد کردن کتابخانه به نرم‌افزار کنسول اکسپرن:.....
- ۶۰.....۲-۳-۴. افزودن حاشیه علمی به اطلاعات :.....

- ۳-۳-۴. خروجی گرفتن از داده‌ها: ۶۱
- ۴-۴. آنالیز ریزآرایه توسط نرم‌افزار فلکس اری: ۶۲
- ۱-۴-۴. ایمپورت کردن در فلکس اری: ۶۲
- ۲-۴-۴. طراحی آزمایش: ۶۴
- ۳-۴-۴. انجام آزمایش: ۶۵
- ۴-۴-۴. اعمال محدودیت‌ها بر روی نتایج حاصل از آنالیز: ۶۷
- ۵-۴-۴. نام‌گذاری لیست ژن: ۶۸
- ۶-۴-۴. گرفتن خروجی: ۶۸
- ۵-۴. به دست آوردن شناسه‌های ژن‌ها: ۶۹
- ۶-۴. تبدیل به فرمت فستا: ۷۰
- ۷-۴. استخراج توالی‌ها با استفاده از نرم‌افزار دی ان آ: ۷۲
- ۸-۴. کار با نرم‌افزار ریپید ماینر و انجام عملیات داده‌کاوی: ۷۳
- ۱-۸-۴. اجرای الگوریتم‌های وزن دهی: ۷۹
- ۲-۸-۴. اجرای الگوریتم‌های داده‌کاوی: ۷۹
- ۱-۲-۸-۴. الگوریتم‌های خوشه‌بندی: ۸۰
- ۲-۲-۸-۴. الگوریتم‌های درخت تصمیم: ۸۰
- ۳-۲-۸-۴. الگوریتم‌های ماشین بردار پشتیبان: ۸۱
- ۴-۲-۸-۴. الگوریتم‌های بیزین: ۸۲
- ۵-۲-۸-۴. الگوریتم‌های شبکه عصبی: ۸۳
- فصل پنجم: نتایج و بحث ۸۵
- ۱-۵. مقدمه: ۸۶
- ۲-۵. نتایج تحقیق: ۸۶
- ۱-۲-۵. نتیجه نرمال‌سازی داده‌ها توسط کنسول اکسپرشن: ۸۷
- ۲-۲-۵. خروجی نرم‌افزار فلکس اری: ۸۹
- ۱-۲-۲-۵. آزمایش GSE1297: ۸۹
- ۲-۲-۲-۵. آزمایش GSE4757: ۹۱
- ۳-۲-۲-۵. آزمایش GSE28146: ۹۲

- ۹۵.....: GSE32536 آزمایش ۴-۲-۲-۵
- ۹۵.....: GSE6980 آزمایش ۵-۲-۲-۵
- ۹۷.....: ژن‌های دارای بیشترین کاهش و افزایش بیان : ۳-۲-۲-۵
- ۹۷.....: بالاترین افزایش بیان : ۱-۳-۲-۵
- ۹۸.....: بیشترین کاهش بیان : ۲-۳-۲-۵
- ۹۸.....: نمایش فرمت فستا : ۴-۲-۵
- ۹۹.....: خروجی نرم‌افزار دی ان آ : ۵-۲-۵
- ۱۰۰: کاهش بعد: ۶-۲-۵
- ۱۰۱: انتخاب ویژگی : ۷-۲-۵
- ۱۰۱: information gain با وزن دهی با ۱-۷-۲-۵
- ۱۰۱: information gain ratio با وزن دهی با ۲-۷-۲-۵
- ۱۰۳: rule با وزن دهی با ۳-۷-۲-۵
- ۱۰۳: deviation با وزن دهی با ۴-۷-۲-۵
- ۱۰۳: Chi Squared statistic با وزن دهی با ۵-۷-۲-۵
- ۱۰۴: Gini Index با وزن دهی با ۶-۷-۲-۵
- ۱۰۴: Uncertainty با وزن دهی با ۷-۷-۲-۵
- ۱۰۵: Relief با وزن دهی با ۸-۷-۲-۵
- ۱۰۷: Support Vector Machine (SVM) با وزن دهی با ۹-۷-۲-۵
- ۱۰۸: PCA با وزن دهی با ۱۰-۷-۲-۵
- ۱۰۹: مهم‌ترین توالی‌ها: ۸-۲-۵
- ۱۱۰: پایگاه داده‌های حاصل شده از اجرای الگوریتم‌های وزن دهی: ۹-۲-۵
- ۱۱۰: information gain پایگاه داده‌ی ۱-۹-۲-۵
- ۱۱۰: information gain ratio پایگاه داده‌ی ۲-۹-۲-۵
- ۱۱۱: rule پایگاه داده‌ی ۳-۹-۲-۵
- ۱۱۱: deviation پایگاه داده‌ی ۴-۹-۲-۵
- ۱۱۲: Chi Squared statistic پایگاه داده‌ی ۵-۹-۲-۵
- ۱۱۳: Gini Index پایگاه داده‌ی ۶-۹-۲-۵

۱۱۴: Uncertainty	پایگاه داده‌ی ۷-۹-۲-۵
۱۱۴: Relief	پایگاه داده‌ی ۸-۹-۲-۵
۱۱۵: Support Vector Machine (SVM)	پایگاه داده‌ی ۹-۹-۲-۵
۱۱۵: PCA	پایگاه داده‌ی ۱۰-۹-۲-۵
۱۱۶: نتایج اجرای الگوریتم‌های داده‌کاوی:	۱۰-۲-۵
۱۱۶: خروجی الگوریتم‌های بیزین:	۱-۱۰-۲-۵
۱۱۷: پشتیبان بردار پشتیبان:	۲-۱۰-۲-۵
۱۱۹: خروجی الگوریتم‌های شبکه عصبی:	۳-۱۰-۲-۵
۱۲۰: خروجی الگوریتم‌های خوشه‌بندی:	۴-۱۰-۲-۵
۱۲۱: درخت تصمیم:	۵-۱۰-۲-۵
۱۲۲: بررسی عملکرد الگوریتم‌ها:	۱۱-۲-۵
۱۲۲: عملکرد الگوریتم‌های گروه بیزین:	۱-۱۱-۲-۵
۱۲۳: پشتیبان بردار پشتیبان:	۲-۱۱-۲-۵
۱۲۳: گروه شبکه عصبی:	۳-۱۱-۲-۵
۱۲۴: گروه خوشه‌بندی:	۴-۱۱-۲-۵
۱۲۴: درخت تصمیم:	۵-۱۱-۲-۵
۱۲۵: مقایسه‌ی عملکرد الگوریتم‌ها:	۱۲-۲-۵
۱۲۶: الگوریتم‌های دارای بالاترین عملکرد:	۱۳-۲-۵
۱۲۶: الگوریتم‌های دارای ضعیف‌ترین عملکرد:	۱۴-۲-۵
۱۲۷: میزان بازدهی پایگاه‌های داده:	۱۵-۲-۵
۱۲۸: شناسایی بیومارکرها با استفاده از درخت تصمیم:	۳-۵
۱۳۱: بحث:	۴-۵
۱۳۴: منابع و مآخذ:	
۱۴۴: اصطلاح نامه:	

فهرست جداول

عنوان.....	صفحه
جدول ۴-۱: کتابخانه‌های میکرو آرایه دانلود شده.....	۵۸
جدول ۵-۱: خروجی نرم‌افزار کنسول اکسپرشن: داده‌های نرمال شده.....	۸۷
جدول ۵-۲: خروجی نرم‌افزار کنسول اکسپرشن.....	۸۸
جدول ۵-۳: نتیجه مقایسه نمونه سالم و بیمار (شدت ضعیف) در آزمایش GSE1297.....	۸۹
جدول ۵-۴: نتیجه مقایسه نمونه سالم و بیمار (شدت متوسط) در آزمایش GSE1297.....	۹۰
جدول ۵-۵: نتیجه مقایسه نمونه سالم و بیمار (شدت زیاد) در آزمایش GSE1297.....	۹۱
جدول ۵-۶: نتیجه مقایسه نمونه سالم و بیمار در آزمایش GSE4757.....	۹۲
جدول ۵-۷: نتیجه مقایسه نمونه سالم و بیمار در آزمایش GSE28146.....	۹۳
جدول ۵-۸: نتیجه مقایسه نمونه سالم و بیمار در آزمایش GSE28146.....	۹۴
جدول ۵-۹: نتیجه مقایسه نمونه سالم و بیمار در آزمایش GSE28146.....	۹۴
جدول ۵-۱۰: نتیجه مقایسه نمونه سالم و بیمار در آزمایش GSE32536.....	۱۰۰
جدول ۵-۱۱: نتیجه مقایسه نمونه سالم و بیمار (قشر تمپورال) در آزمایش GSE6980.....	۹۶
جدول ۵-۱۲: نتیجه مقایسه نمونه سالم و بیمار (قشر پیشانی) در آزمایش GSE6980.....	۹۶
جدول ۵-۱۳: نتیجه مقایسه نمونه سالم و بیمار (هیپوکامپ) در آزمایش GSE6980.....	۹۷
جدول ۵-۱۴: لیست ژن‌های دارای بالاترین کاهش بیان.....	۹۸
جدول ۵-۱۵: لیست ژن‌های دارای بالاترین کاهش بیان.....	۹۸
جدول ۵-۱۶: نمایش فرمت فستا.....	۹۹
جدول ۵-۱۷: خروجی نرم‌افزار دی ان آ.....	۱۰۰
جدول ۵-۱۸: نتایج حاصل از الگوریتم information gain.....	۱۰۱
جدول ۵-۱۹: نتایج حاصل از الگوریتم information gain ratio.....	۱۰۲
جدول ۵-۲۰: نتایج حاصل از الگوریتم rule.....	۱۰۳
جدول ۵-۲۱: نتایج حاصل از الگوریتم deviation.....	۱۰۳
جدول ۵-۲۲: نتایج حاصل از الگوریتم Chi Squared statistic.....	۱۰۴
جدول ۵-۲۳: نتایج حاصل از الگوریتم Gini Index.....	۱۰۴

۱۰۵	جدول ۵-۲۴: نتایج حاصل از الگوریتم Uncertainty
۱۰۶	جدول ۵-۲۵: نتایج حاصل از الگوریتم Relief
۱۰۸	جدول ۵-۲۶: نتایج حاصل از الگوریتم Support Vector Machine (SVM)
۱۰۹	جدول ۵-۲۷: نتایج حاصل از الگوریتم PCA
۱۰۹	جدول ۵-۲۸: مهم‌ترین توالی‌ها
۱۱۰	جدول ۵-۲۹: پایگاه داده information gain
۱۱۱	جدول ۵-۳۰: پایگاه داده information gain ratio
۱۱۱	جدول ۵-۳۱: پایگاه داده‌ی rule
۱۱۲	جدول ۵-۳۲: پایگاه داده‌ی deviation
۱۱۳	جدول ۵-۳۳: پایگاه داده‌ی Chi Squared statistic
۱۱۳	جدول ۵-۳۴: پایگاه داده‌ی Gini Index
۱۱۵	جدول ۵-۳۶: وزن دهی با Relief
۱۱۵	جدول ۵-۳۷: وزن دهی با Support Vector Machine (SVM)
۱۱۶	جدول ۵-۳۸: وزن دهی با PCA
۱۱۷	جدول ۵-۳۹: خروجی الگوریتم‌های بیزین
۱۱۷	جدول ۵-۴۰: خروجی الگوریتم‌های ماشین بردار پشتیبان
۱۱۸	جدول ۵-۴۱: خروجی الگوریتم‌های ماشین بردار پشتیبان
۱۱۹	جدول ۵-۴۲: خروجی الگوریتم‌های ماشین بردار پشتیبان
۱۱۹	جدول ۵-۴۳: خروجی الگوریتم‌های ماشین بردار پشتیبان
۱۲۰	جدول ۵-۴۴: خروجی الگوریتم‌های شبکه عصبی
۱۲۰	جدول ۵-۴۵: خروجی الگوریتم‌های خوشه‌بندی
۱۲۱	جدول ۵-۴۶: خروجی الگوریتم‌های درخت تصمیم
۱۲۱	جدول ۵-۴۷: خروجی الگوریتم‌های درخت تصمیم
۱۲۲	جدول ۵-۴۸: خروجی الگوریتم‌های درخت تصمیم
۱۲۲	جدول ۵-۴۹: بررسی کلی عملکرد الگوریتم‌های بیزین
۱۲۳	جدول ۵-۵۰: بررسی کلی عملکرد الگوریتم‌های ماشین بردار پشتیبان
۱۲۴	جدول ۵-۵۱: بررسی کلی عملکرد الگوریتم‌های شبکه عصبی

- جدول ۵-۵۲: بررسی کلی عملکرد الگوریتم‌های خوشه‌بندی..... ۱۲۴
- جدول ۵-۵۳: بررسی کلی عملکرد الگوریتم‌های درخت تصمیم..... ۱۲۵
- جدول ۵-۵۴: بررسی عملکرد شاخه‌های اصلی الگوریتم‌ها..... ۱۲۵
- جدول ۵-۵۵: الگوریتم‌های دارای بالاترین عملکرد..... ۱۲۶
- جدول ۵-۵۶: الگوریتم‌های دارای ضعیف‌ترین عملکرد..... ۱۲۶
- جدول ۵-۵۷: میزان بازدهی پایگاه‌های داده..... ۱۲۸

فهرست شکل‌ها

عنوان.....	صفحه
شکل ۴-۱: سایت ان سی بی آی.....	۵۶
شکل ۴-۲: پلت فرم یک آزمایش ریزآرایه.....	۵۷
شکل ۴-۳: توجه به پسوند نمونه‌ها و دانلود کتابخانه ریزآرایه.....	۵۸
شکل ۴-۴: ایمپورت کردن کتابخانه ریزآرایه در کنسول اکسپرشن.....	۵۹
شکل ۴-۵: نرمال کردن داده‌ها در کنسول اکسپرشن.....	۶۰
شکل ۴-۶: افزودن حاشیه‌نویسی علمی به ژن‌ها.....	۶۰
شکل ۴-۷: انتخاب حاشیه‌های موردنظر.....	۶۱
شکل ۴-۸: خروجی گرفتن از نرم‌افزار کنسول اکسپرشن.....	۶۲
شکل ۴-۹: ایمپورت کردن داده‌ها در نرم‌افزار فلکس اری.....	۶۲
شکل ۴-۱۰: ایمپورت کردن داده‌ها در نرم‌افزار فلکس اری.....	۶۳
شکل ۴-۱۱: ایمپورت کردن داده‌ها در نرم‌افزار فلکس اری.....	۶۳
شکل ۴-۱۲: ایمپورت کردن داده‌ها در نرم‌افزار فلکس اری.....	۶۴
شکل ۴-۱۳: تقسیم‌بندی سمپل‌ها و طراحی آزمایش.....	۶۵
شکل ۴-۱۴: اجرای آزمون بیز تی.....	۶۶
شکل ۴-۱۵: اجرای آزمون بیز تی.....	۶۶
شکل ۴-۱۶: نتایج آزمون بیز تی و اعمال محدودیت‌ها روی نتایج.....	۶۷
شکل ۴-۱۷: نام‌گذاری لیست ژن.....	۶۸
شکل ۴-۱۸: لیست نتایج و گرفتن خروجی.....	۶۸
شکل ۴-۱۹: سایت کلون.....	۶۹
شکل ۴-۲۰: انتخاب شناسه‌های موردنظر.....	۷۰
شکل ۴-۲۱: یک نمونه فایل فستا با دو رکورد.....	۷۱
شکل ۴-۲۲: بخش بچ انترز سایت ان سی بی آی.....	۷۱
شکل ۴-۲۳: نرم‌افزار دی ان آ.....	۷۲
شکل ۴-۲۴: ایجاد مخزن داده در ریپد ماینر.....	۷۳

- شکل ۴-۲۵: مراحل ایجاد مخزن داده ۷۳
- شکل ۴-۲۶: مراحل ایجاد مخزن داده ۷۴
- شکل ۴-۲۷: خواندن داده‌ها در ریپید ماینر ۷۴
- شکل ۴-۲۸: خواندن فایل اکسل ۷۵
- شکل ۴-۲۹: انتخاب فایل اکسل مورد نظر ۷۵
- شکل ۴-۳۰: ایمپورت کردن فایل اکسل در ریپید ماینر ۷۶
- شکل ۴-۳۱: ایمپورت کردن فایل اکسل در ریپید ماینر ۷۶
- شکل ۴-۳۲: ایمپورت کردن فایل اکسل در ریپید ماینر ۷۷
- شکل ۴-۳۳: ایمپورت کردن فایل اکسل در ریپید ماینر ۷۷
- شکل ۴-۳۴: ایمپورت کردن فایل اکسل در ریپید ماینر ۷۸
- شکل ۴-۳۵: اجرای دستور `read data` ۷۸
- شکل ۵-۱: نتیجه کلاس‌بندی با استفاده از توالی `AATTG` ۱۲۹
- شکل ۵-۲: نتیجه کلاس‌بندی با استفاده از توالی `AATTG` ۱۲۹
- شکل ۵-۳: درخت تصمیم‌گیری با استفاده از توالی `AAATTG` ۱۳۰
- شکل ۵-۴: درخت تصمیم‌گیری با استفاده از توالی `AAATTG` ۱۳۰
- شکل ۵-۵: درخت تصمیم‌گیری با استفاده از توالی `CCTG` ۱۳۱
- شکل ۵-۶: درخت تصمیم‌گیری با استفاده از توالی `GAATAT` ۱۳۱

فهرست نشانه‌ها و اختصارها

BDNF	Brain-derived neurotrophic factor
cDNA	Complementary DNA
cRNA	Certified Registered Nurse Anesthetists
DNA	Deoxyribonucleic acid
mi-RNA	microRNA
MPTP- 1	Methyl-4-phenyl-1,2,3,6-tetrahydropyridine
MRI	Magnetic resonance imaging
RMA	Return merchandise authorization
RNA	Ribonucleic acid
SVM	Support Vector Machine

فصل اول: کلیات تحقیق

۱-۱. مقدمه:

در سال‌های اخیر، شاهد پیوند گسترده‌ای بین حوزه‌های مختلف دانش بوده‌ایم، به‌خصوص داده‌کاوی^۱ و علوم مختلف. داده‌کاوی در زمینه‌های گسترده‌ای با علوم مختلف ارتباط دارد. کشف تقلب، تشخیص الگو، پردازش زبان طبیعی، بیوانفورماتیک^۲ برخی مسائلی هستند که با داده‌کاوی پیوند وسیعی دارند. در حوزه‌ی علوم پزشکی تشخیص بیماری‌ها، پیش‌بینی بیماری‌ها، میزان پاسخ‌دهی روش‌های درمان از نمونه‌های بارز استفاده از روش‌های داده‌کاوی در پزشکی است. با توجه به اهمیت بیماری آلزایمر^۳ و عدم وجود روش درمان قطعی برای این بیماری بر آن شدیم تا با استفاده از روش‌های داده‌کاوی اقدام به شناسایی الگوهای پنهان در ژن‌های مؤثر در بروز این بیماری پردازیم. روش‌های داده‌کاوی در این تحقیق به شناخت و استخراج الگوهای تکراری و میزان اهمیت آن‌ها در بروز بیماری آلزایمر نقش مهمی ایفا کرد. هدف از انجام این نمونه تحقیق‌ها کمک به شناخت بیشتر و جلوگیری از پیشرفت و گسترش بیماری‌هایی مانند آلزایمر است که هیچ روش قطعی برای درمان آن‌ها وجود ندارد.

¹ Data mining

² bioinformatic

³ Alzheimer disease