Check for updates

Original Article

*Journal of Health and Biomedical Informatics*
*Medical Informatics Research Center*
*2022; 9(3): 106-119*
*10.34172/jhbmi.2022.01*

# Accuracy Improvement of Mood Disorders Prediction using a Combination of Data Mining and Meta-Heuristic Algorithms

**Fariborzi Mana[1], Zeinalnezhad Masoomeh[2]*, Saghaei Abbas[3]**

• **Received:** 25 Aug 2022 • **Accepted:** 20 Nov 2022

**Introduction:** Since the delay or mistake in the diagnosis of mood disorders due to the similarity of their symptoms hinders effective treatment, this study aimed to accurately diagnose mood disorders including psychosis, autism, personality disorder, bipolar, depression, and schizophrenia, through modeling and analyzing patients' data.

**Method:** Data collected in this applied developmental research included 996 records with 130 features obtained by interviewing and completing questionnaires in a mental hospital in the city of Sari, Iran in 2021. After preprocessing, the number of features was reduced to 91, and then through Principal Component Analysis (PCA) reduced to 35 factors. Modeling was done in Python software with K-Nearest Neighbor (KNN), Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), Logistic Regression (LR), and Support Vector Machine (SVM) algorithms. The models were evaluated to select algorithms with higher accuracy. Particle Swarm Optimization (PSO) and Genetic Algorithm (GA) were applied to determine the optimal parameters of the selected algorithms.

**Results:** Among the machine learning algorithms, random forest with 91% accuracy and support vector machine with 90% accuracy showed better performance. The genetic algorithms did not make any notable increase in prediction accuracy. Whereas considering N=30, T=150, W=0.9, c1=2, and c2=2 in the particle swarm optimization algorithm increased the prediction accuracy up to 3.3 %.

**Conclusion:** With less classification error compared to similar studies, the PSO-SVM model designed in this study can be used in patient data monitoring with acceptable accuracy and can be used in intelligent systems in psychiatric centers.

**Keywords:** Mood Disorder, Machine Learning, Meta-Heuristic Algorithms, Data Mining, Prediction

• **Citation:** Fariborzi M, Zeinalnezhad M, Saghaei A. Accuracy Improvement of Mood Disorders Prediction using a Combination of Data Mining and Meta-Heuristic Algorithms. Journal of Health and Biomedical Informatics 2022; 9(3): 106-19. doi: 10.34172/jhbmi.2022.01. [In Persian]

1. M.Sc. in Industrial Engineering, Department of Industrial Engineering, Faculty of Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
2. Ph.D. in Industrial Engineering, Assistant Professor, Department of Industrial Engineering, Faculty of Engineering, West Tehran Branch, Islamic Azad University, Tehran, Iran
3. Ph.D. in Industrial Engineering, Professor, Department of Industrial Engineering, Faculty of Engineering, Science and Research Branch, Islamic Azad University, Tehran, Iran
***Corresponding Author:** Masoomeh Zeinalnezhad
**Address:** Department of Industrial Engineering, West Tehran Branch, Islamic Azad University, Shahid Hasan Azari St., Hemmat Bridge, Ashrafi Esfahani Hwy, Tehran, Iran
•**Tel:** 021-44220677 •**Email:** zeinalnezhad.m@wtiau.ac.ir

[ Downloaded from jhbmi.com on 2026-11-17 ]

[ DOI: 10.34172/jhbmi.2022.01 ]

مقاله پژوهشی

مجله انفورماتیک سلامت و زیست پزشکی
مرکز تحقیقات انفورماتیک پزشکی
دوره نهم شماره سوم. ص ۱۱۹-۱۰۶، ۱۴۰۱
10.34172/jhbmi.2022.01

# بهبود دقت پیش‌بینی اختلالات خلقی با استفاده از ترکیب الگوریتم‌های داده‌کاوی و فراابتکاری

مانا فری برزی[۱]، معصومه زینال نژاد[۲*]، عباس سقایی[۳]

• دریافت مقاله: ۱۴۰۱/۶/۳ • پذیرش مقاله: ۱۴۰۱/۸/۲۹

**مقدمه:** از آنجا که تأخیر یا اشتباه در تشخیص اختلالات خلقی به‌دلیل تشابه علائم، مانع درمان مؤثر می‌شود، هدف تحقیق حاضر، تشخیص دقیق اختلالات خلقی، شامل روان‌پریشی، اوتیسم، اختلال شخصیت، دوقطبی، افسردگی و اسکیزوفرنی، از طریق مدل‌سازی و تحلیل داده‌های بیماران است.

**روش:** داده‌های جمع‌آوری شده در این پژوهش کاربردی-توسعه‌ای دربردارنده ۹۹۶ رکورد با ۱۳۰ ویژگی بود که با مصاحبه و تکمیل پرسشنامه‌ها در یک بیمارستان روان‌پزشکی در شهر ساری، ایران در سال ۱۴۰۰ گردآوری شد. پس از پیش‌پردازش داده‌ها، تعداد ویژگی‌ها با تحلیل همبستگی به ۹۱ و سپس با تحلیل مؤلفه‌های اصلی (PCA) به ۳۵ فاکتور کاهش یافت. مدل‌سازی داده‌ها در نرم‌افزار پایتون با الگوریتم‌های K نزدیک‌ترین همسایه(KNN)، نایو بیز (NB)، درخت تصمیم (DT)، جنگل تصادفی (RF)، رگرسیون لجستیک (LR) و ماشین بردار پشتیبان (SVM) صورت گرفت. عملکرد مدل‌ها ارزیابی گردید و پارامترهای الگوریتم‌های با دقت بالاتر توسط الگوریتم‌های فراابتکاری بهینه‌سازی ازدحام ذرات (PSO) و ژنتیک (GA) تخمین زده شد.

**نتایج:** از بین الگوریتم‌های یادگیری ماشین، دو الگوریتم RF با دقت ۹۱ و SVM با دقت ۹۰ درصد عملکرد بهتری داشتند. GA افزایش دقت قابل‌ملاحظه‌ای ایجاد نکرد، اما درنظرگرفتن مقادیر ۳۰، ۱۵۰، ۰/۹، ۲ و ۲ به‌ترتیب به‌عنوان تعداد ذرات، تعداد تکرار الگوریتم، ضریب اینرسی، ضریب شخصی و ضریب اجتماعی در الگوریتم PSO دقت پیش‌بینی را تا ۳/۳ درصد بهبود بخشید.

**نتیجه‌گیری:** با خطای طبقه‌بندی کمتر نسبت به پژوهش‌های مشابه، مدل PSO-SVM طراحی شده در این پژوهش با دقت قابل قبولی می‌تواند در پایش داده‌های بیماران به‌کار گرفته شده و در سامانه‌های هوشمند مراکز روان‌پزشکی مورد استفاده قرار گیرد.

**کلیدواژه‌ها:** اختلال روانی، یادگیری ماشین، الگوریتم‌های فرا ابتکاری، داده‌کاوی، پیش‌بینی

**ارجاع:** فری برزی مانا، زینال نژاد معصومه. سقایی عباس. بهبود دقت پیش‌بینی اختلالات خلقی با استفاده از ترکیب الگوریتم‌های داده‌کاوی و فراابتکاری. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۱؛ ۹(۳): ۱۹-۱۰۶. doi: 10.34172/jhbmi.2022.01.۱۰۶-۱۹

۱. کارشناس ارشد مهندسی صنایع-سیستم‌های سلامت، گروه مهندسی صنایع، دانشکده فنی و مهندسی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
۲. دکترای مهندسی صنایع، استادیار، گروه مهندسی صنایع، دانشکده فنی و مهندسی، واحد تهران غرب، دانشگاه آزاد اسلامی، تهران، ایران
۳. دکترای مهندسی صنایع، استاد، گروه مهندسی صنایع، دانشکده فنی و مهندسی، واحد علوم و تحقیقات، دانشگاه آزاد اسلامی، تهران، ایران
* **نویسنده مسئول:** معصومه زینال نژاد
**آدرس:** تهران، بلوار اشرفی اصفهانی، خیابان قموشی، بلوار عربشاهی، خیابان شهید حسن آذری، دانشکده فنی و مهندسی دانشگاه آزاد اسلامی واحد تهران غرب، گروه مهندسی صنایع
• **شماره تماس:** ۴۴۲۲۰۶۷۷ – ۰۲۱ • **Email:** m.zeinalnezhad@gmail.com



[ Downloaded from jhbmi.com on 2026-11-17 ]
[ DOI: 10.34172/jhbmi.2022.01 ]

## مقدمه

امروزه نشاط اجتماعی یکی از مهم‌ترین نیازهای جامعه است چراکه شهروندان به‌ویژه ساکنان کلان‌شهرها، فرصت کمتری برای اندیشیدن به خود و نیازهایشان دارند و به‌شدت مستعد اختلالات خلقی هستند [۱]. افراد مبتلابه بیماری‌های روانی در سراسر جهان تا حد زیادی نادیده گرفته می‌شوند [۲]. اختلالاتی مانند افسردگی، اسکیزوفرنی و دوقطبی به ترتیب ۱/۸۴ درصد، ۰/۶۰ درصد، ۰/۳۳ درصد از کل‌سال‌های زندگی با ناتوانی تعدیل‌شده را تشکیل می‌دهند [۳]. مشکلات سلامت روان می‌تواند به پیامدهای جدی مانند افسردگی، خودزنی به‌خصوص برای دانش‌آموزانی که از نظر جسمی و روحی بالغ نیستند، منجر شود [۴]. افسردگی، در صورت عدم‌تشخیص و درمان به‌موقع، می‌تواند منجر به خودکشی شده [۵،۶]. یا تأثیر منفی بر روند بیماری‌های جسمی مانند سکته مغزی، سرطان داشته و هزینه‌های پزشکی را افزایش دهد [۷]. اختلال دوقطبی یک بیماری مزمن شایع و عودکننده است که تأثیر مستقیمی بر سلامت بیمار دارد [۸] و در دو سوم بیماران قابل درمان نیست [۹]. خُلق بر طبق تعریف عبارت است از احساس هیجانی نافذی که درک و نگرش فرد نسبت به خود، دیگران و در کل نسبت به محیط اطراف را عمیقاً تحت تأثیر قرار می‌دهد [۱۰]. خُلق ممکن است طبیعی و یا غیرطبیعی باشد. یک فرد عادی طیف وسیعی از خُلق‌ها را تجربه می‌کند و به همان نسبت مجموعه‌ای از تجلیات عاطفی نیز دارد. این فرد قادر به کنترل اخلاق و عواطف خود است. اختلالات خُلقی حالات بالینی هستند که با اختلال خُلق، فقدان احساس کنترل بر خُلق و تجربه ذهنی ناراحتی شدید همراه است. بر اساس تعریف سازمان بهداشت جهانی، اختلالات خُلقی، مانند اختلال افسردگی عمده و اختلال دوقطبی، دربرگیرنده اختلالات خُلقی مکرر و مزمن است [۳]؛ بدین علت بیماران مبتلابه اختلالات خُلقی از فشار شدید، اختلال در سطح عملکردی و افزایش خطر خودکشی رنج می‌برند [۱۱]. تأخیر در تشخیص و در بیشتر موارد تشخیص اشتباه اختلالات خلقی به دلیل تشابه علائم آن‌ها مانع از درمان مؤثر بیماران می‌شود. پیش‌بینی شروع و سیر اختلالات خلقی و اضطرابی با وجود اهمیت بالا بسیار دشوار است [۱۲]. به‌عنوان نمونه، اختلال دوقطبی از اختلال افسردگی عمده را با دقت بالا نمی‌توان تشخیص داد [۱۳]. تشخیص برای بیمارانی که بیش از یک نوع اختلال روانی دارند، سخت‌تر است [۱۴]. در صورت تشخیص درست بیماری، روان‌درمانی شناختی-رفتاری در کنار درمان‌های معمول، در کاهش علائم و نشانه‌های اختلال خلقی در بیماران مؤثر است و استفاده از رویکرد تیمی مبتنی بر همکاری روان‌شناس و روان‌پزشک می‌تواند نتایج درمانی بهتری داشته باشد [۱۵]. همچنین، آموزش‌های مهارت اجتماعی می‌تواند با اصلاح روابط بین فردی بیماران، زمینه را برای کاهش تعاملات ناکارآمد و در نتیجه کاهش عواطف منفی و سازگاری اجتماعی افراد فراهم کند [۱۶].

با توجه به پیچیدگی تشخیص درست نوع بیماری افراد بیمار، به ابزارهای تشخیصی آسان‌تر و دقیق‌تری نیاز است [۱۷]. در این خصوص، استفاده از علم داده‌کاوی در تحقیقات مهندسی سلامت و پایش اطلاعات بیماران رو به افزایش است [۱۸]. یادگیری ماشین (Machin Learning) به‌عنوان روشی مناسب برای تجزیه‌وتحلیل داده‌های بزرگ شناخته شده که توسط آلن تورینگ (Allan Turing) و آرتور ساموئل (Arthur Samuel) در دهه ۱۹۵۰ ابداع گردید و از اواخر قرن بیستم تاکنون در زمینه‌های مختلف ازجمله حوزه پزشکی کاربرد گسترده داشته است [۱۹]. همچنین، هوش مصنوعی به دلیل قابلیت بالا در خودکارسازی و کار با داده‌های انبوه و قابلیت استنتاج قوی، کاربرد فراوانی در حیطه تشخیص اختلالات روانی داشته است [۲۰]. با گسترش شبکه‌های اجتماعی و ظهور کلان داده‌ها، داده‌های زیادی درباره بیماران روانی در دسترس متخصصین روان‌شناسی است. این داده‌ها منبعی غنی برای یاری‌رسانی به روان‌شناسان در تشخیص اختلالات روانی است [۲۱]. داده‌ها را از منابع مختلفی همچون دستگاه‌های پوشیدنی و گوشی‌های هوشمند می‌توان جمع‌آوری کرد[۲۲،۲۳]. تحلیل و سازمان‌دهی این داده‌ها به‌صورت دستی دشوار است درحالی‌که هوش مصنوعی می‌تواند با پایش داده‌ها و استنتاج و پیش‌بینی به متخصصین روان‌شناسی در تشخیص بیماری‌های روانی بر اساس تصاویر، صوت و یا متن یاری رساند [۲۴]. در میان روش‌های هوش مصنوعی، دو رویکرد یادگیری ماشین و یادگیری عمیق نتایج خوبی در ارزیابی و تشخیص اختلالات روانی از خود نشان داده‌اند [۲۵]. یادگیری ماشین می‌تواند به‌عنوان یک بخش مهم و اساسی در تجزیه‌وتحلیل داده‌های بزرگ، در نظر گرفته شود [۱۷] و در حل مسائل مراقبت‌های بهداشتی مانند تشخیص زودرس بیماری، نظارت بر بیمار در زمان واقعی، مراقبت از بیمار و تقویت درمان نقش داشته باشد [۲۶]. در این راستا، Liu و همکاران [۱۴]، در تحقیقی با استفاده از مدل‌های یادگیری عمیق به بررسی تشخیص ۸ بیماری روانی شاخص شامل بیش‌فعالی، افسردگی، اضطراب، اوتیسم، ناتوانی‌های فکری، اختلال گفتار/زبان، تأخیر در تحولات و



[ Downloaded from jhbmi.com on 2026-11-17 ]
[ DOI: 10.34172/jhbmi.2022.01 ]

اختلال سرپیچی مخالف در بین بیمارانی از اقلیت قومی آفریقایی- آمریکایی پرداختند. داده‌های گردآوری شده شامل ۴۱۷۹ رکورد بود که در آن میان، ۱۳۸۴ بیمار با تشخیص حداقل یک اختلال روانی وجود داشتند. دقت مدل طراحی شده ۶۵ درصد تخمین زده شد. de Siqueira Rotenberg و همکاران [۲۷] با هدف استفاده از رویکردهای یادگیری ماشین به‌عنوان یک پیش‌بینی‌کننده احتمالی در عود افسردگی در دوقطبی مطالعه‌ای انجام دادند که در آن الگوریتم‌های ماشین بردار پشتیبان، جنگل تصادفی، بیزین ساده و پرسپترون چندلایه به مجموعه داده‌ها اعمال شد. در هر دو پژوهش الگوریتم جنگل تصادفی میانگین عملکرد بالاتری را به دست آورد. Bayes و همکاران [۲۸] یک رویکرد یادگیری ماشین را برای تشخیص اختلال دوقطبی (Bipolar Disorder) از اختلال شخصیت مرزی (Borderline Personality Disorder) ارائه کردند. دقت طبقه‌بندی مدل‌های یادگیری ماشین بین ۸۴/۱ درصد تا ۸۷/۸ درصد برای اختلال دوقطبی و ۵۰ تا ۵۷/۷ درصد برای اختلال شخصیت مرزی و با دقت کلی ۷۳/۱ درصد الی ۷۳/۹ درصد بود. با توجه به تأثیر مستقیم اختلال دوقطبی بر رفاه بیمار، Liu و همکاران [۸]، یک مطالعه گذشته‌نگر برای ارزیابی افزایش عملکرد پرسشنامه‌های غربالگری اختلال دوقطبی نسبت به Mood Disorder Questionnaire (پرسشنامه اختلال خلقی) از یک رویکرد یادگیری ماشینی طراحی شده برای ارزیابی مجموعه‌ای از بیماری‌های سلامت روان استفاده کرده‌اند. در مرکز سلامت روانی در غرب کانادا شرکت‌کنندگان (n=۹۵۵؛ ۵۶/۴ درصد زن؛ میانگین سنی ۳۵/۴ و ۱۸/۷ درصد دوقطبی) پرسشنامه را تکمیل کردند و برای دقت تشخیصی تحت یک مصاحبه بالینی با روان‌پزشک قرار گرفتند. با استفاده از امتیازدهی ترکیبی، دقت ۸۰/۶ درصد، با حساسیت ۷۳/۷ درصد و ویژگی ۸۷/۵ درصد بود. در مقایسه با روش امتیازدهی اصلی Mood Disorder Questionnaire، مدل ترکیبی دقت را تا ۶/۹ درصد، حساسیت را تا ۱۴/۵ درصد بهبود بخشید و درعین‌حال ویژگی را حفظ کرد.

van Eeden و همکاران [۱۲] روش‌های یادگیری ماشین احتمالی اولیه و خودکار (Auto-sklearn) را مقایسه کردند. داده‌ها از مطالعه افسردگی و اضطراب در هلند به دست آمده بود. رگرسیون لجستیک چندجمله‌ای، طبقه‌بندی‌کننده ساده بیز و Auto-sklearn تشخیص‌های افسردگی و اضطراب را در یک پیگیری ۲، ۴، ۶ و ۹ ساله پیش‌بینی کردند که به‌عنوان متغیرهای باینری یا طبقه‌بندی عملیاتی شدند. مجموعه‌های پیش‌بینی‌کننده شامل داده‌های جمعیت‌شناختی و خود گزارشی بود که می‌توان آن‌ها را به راحتی در عمل بالینی در دو نقطه زمانی اولیه (پایه و پیگیری ۱ ساله) جمع‌آوری کرد. این سه روش به‌طور مشابه در پیش‌بینی وضعیت سلامت روانی با پیش‌بینی‌های صحیح تا ۷۹ درصد موفق بودند. بااین‌حال، Auto-sklearn هنگام ارزیابی مجموعه داده‌های پیچیده‌تر با امتیازات ویژگی‌ها برتر بود.

Nguyen و همکاران [۲۹] در پژوهشی بیماران اختلالات سلامت روان را به دو گروه اسکیزوفرنی و اختلالات خلقی (دو قطبی و تک قطبی) تقسیم نمودند. آن‌ها از دو الگوریتم‌های یادگیری عمیق استفاده نمودند. داده‌ها با استفاده از دستگاه پوشیدنی جمع‌آوری شدند. دقت به دست آمده در پژوهش ۰/۸۶ و ۰/۷۹ بود.

تمرکز در حیطه فقط یک بیماری روانی در بیشتر پژوهش‌ها و یا عدم استفاده از داده های کافی، همچنین دقت پایین تشخیص با استفاده از الگوریتم‌های محدود و یا عدم به‌کارگیری روش‌های دیگر برای بهبود دقت کار ضعف شناخته‌شده در بررسی تحقیقات پیشین هستند. در پژوهش حاضر تلاش شده است تا با بررسی انواع اختلالات خلقی و شناسایی تمامی ویژگی‌های مؤثر و استفاده همزمان از الگوریتم‌های یادگیری ماشین و فرا ابتکاری مدلی طراحی شود که بتواند شش نوع از اختلالات خلقی شامل روان‌پریشی، اوتیسم، اختلال شخصیت، دوقطبی، افسردگی و اسکیزوفرنی را از یکدیگر تمایز داده و با توجه به علائم، آن‌ها را با دقت قابل قبولی پیش‌بینی نماید.

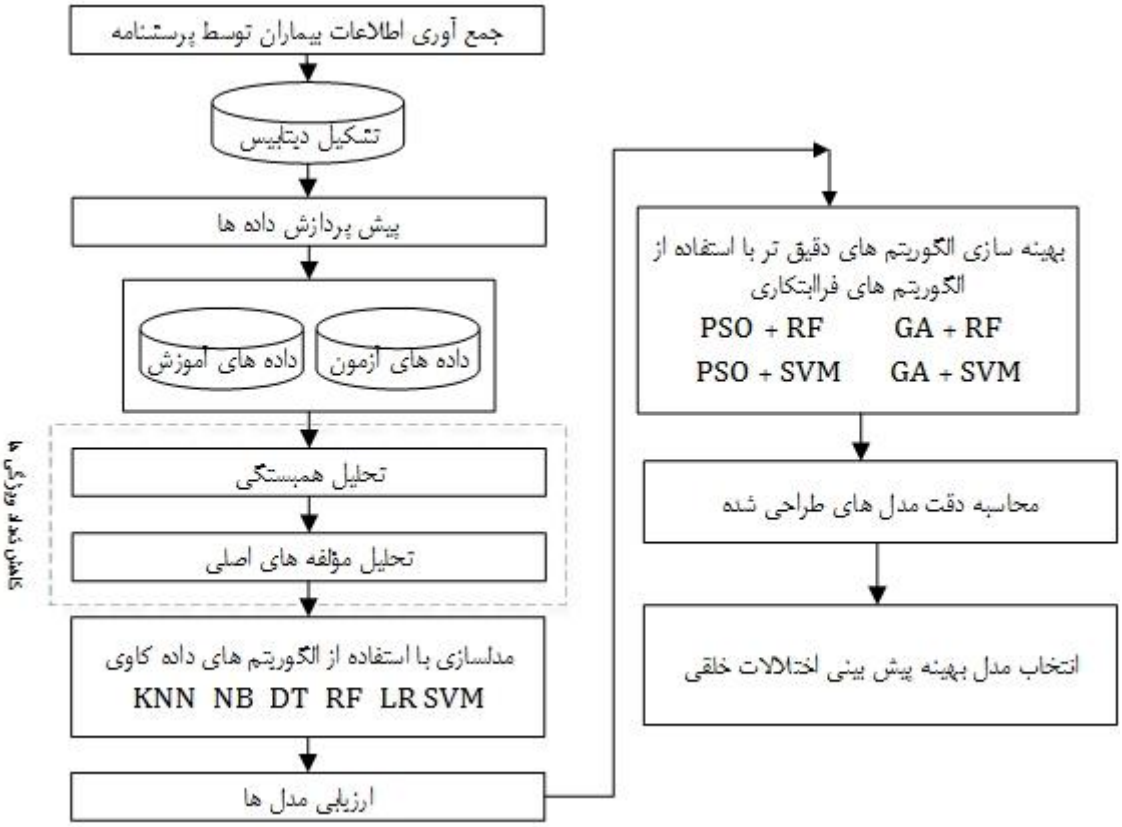
## روش

مجموعه داده مورد تحلیل در این پژوهش کاربردی-توسعه‌ای دربردارنده ۹۹۶ رکورد با ۱۳۰ ویژگی بوده است که با استفاده از پرسشنامه استاندارد موجود، ، از طریق مصاحبه با بیماران یک بیمارستان روان‌پزشکی در شهر ساری در سال ۱۴۰۰ گردآوری شد. این مجموعه داده شامل ۱۷۵ رکورد بیماران دوقطبی، ۱۶۷ رکورد بیماران روان‌پریشی، ۱۶۶ رکورد بیماران اوتیسم، ۱۶۵ رکورد بیماران افسردگی، ۱۶۳ رکورد بیماران اسکیزوفرنی و ۱۶۰ رکورد بیماران اختلال شخصیت بود.

مراحل مختلف انجام تحقیق حاضر در شکل ۱ به‌طور خلاصه تشریح شده‌اند.

[ Downloaded from jhbmi.com on 2026-11-17 ]

[ DOI: 10.34172/jhbmi.2022.01 ]




**شکل ۱: فلوچارت روش تحقیق**

مجموعه داده گردآوری شده شامل اطلاعات دموگرافیک بیماران اعم از سن، جنس، وزن، وضعیت تأهل، قد، ویژگی‌های جسمی مثل شاخص توده بدنی، بیماری جسمی و ویژگی‌های رفتاری مثل اعتیاد، تنبلی، ضعف حافظه، ترومای دوران کودکی و ... است. شش رده (کلاس یا برچسب) اختلال خلقی ۱-دوقطبی، ۲-افسردگی، ۳-اوتیسم، ۴-اسکیزوفرنی، ۵-اختلال شخصیت و ۶-روان‌پریشی برچسب‌گذاری و به‌صورت اسمی ذخیره شدند. این شش رده با تابع OrdinalEncocer پایتون با اعداد ۰ تا ۵ کدگذاری شدند و در محاسبات تحلیلی مورداستفاده قرار گرفتند.

پس از تهیه مجموعه داده، ویژگی‌های اصلی و مؤثر با کمک متخصصان و مرور ادبیات تحقیق مشخص شدند و پیش‌پردازش داده‌ها صورت گرفت. مقادیر مفقوده، داده‌های نامعلومی هستند که مقدار مشخصی برای آن‌ها متصور نیست و می‌توانند الگوریتم‌ها را گمراه کنند. با توجه به این که داده‌ها از طریق پرسشنامه‌های تکمیل شده به دست آمده بودند، باید فاقد مقادیر مفقوده می‌بودند؛ برای اطمینان از صحت داده‌ها، این موضوع بررسی شد. در مرحله بعد برخی فیلدهای اضافی و بی‌تأثیر در طبقه‌بندی بیماری به‌صورت شهودی و با مشورت با متخصصان روان‌پزشکی حذف شده و از ابعاد مسئله کاسته شد. این مرحله، یک مرحله کاهش ویژگی استاندارد نیست و کاهش ویژگی باید در فاز مهندسی ویژگی و با استفاده از روش‌ها و رویکردهای کاهش ویژگی انجام شود. کاری که در این مرحله صورت گرفت، در واقع مقدمه‌ای است بر مرحله اصلی مهندسی ویژگی و باعث بالا رفتن کیفیت کار در مراحل بعدی شد. حذف برخی فیلدها از طریق ایجاد فیلدهای جدید صورت گرفت. به‌عنوان‌مثال تاریخ پذیرش و تاریخ ترخیص حذف و به‌جای آن، ویژگی مدت بستری از تفریق دو تاریخ ایجاد شد. در این مرحله تعداد ویژگی‌ها از ۱۳۰ به ۱۲۲ عدد کاهش پیدا کرد. برای کاهش بیشتر ابعاد داده‌ها با بررسی همبستگی بین دوبه‌دوی ویژگی‌ها و بر اساس آستانه همبستگی با فرض برابر بودن همبستگی قوی با مقدار ۰/۸۵، از متغیرهای همبسته قوی (بالاتر از ۰/۸۵) صرف‌نظر شد. بدین ترتیب، تعداد ویژگی‌ها از ۱۲۲ به ۹۱ ویژگی کاهش یافت که در جدول ۱ تشریح شده‌اند.

[ Downloaded from jhbmi.com on 2026-11-17 ]
[ DOI: 10.34172/jhbmi.2022.01 ]

جدول ۱: ویژگی‌های (متغیرهای) دیتاست بیماران

| ردیف | عنوان فارسی ویژگی | عنوان انگلیسی ویژگی | نوع داده | مقدار و دامنه | توضیح |
|---|---|---|---|---|---|
| ۱ | جنس | sex | Binary | ۰=مرد، ۱=زن | |
| ۲ | سن | age | Numeric | ۱۶-۷۸ سال | |
| ۳ | مدت کل اقامت | the total length of residence | Numeric | ۱-۱۵۰ | |
| ۴ | نوبت بستری | the turn of the hospitalization | Numeric | ۱-۲۲ | |
| ۵ | قد | height | Numeric | ۱۰۰-۱۹۰ | |
| ۶ | شاخص توده بدنی | BMI | Numeric | ۱۵,۵-۷۶ | |
| ۷ | سابقه بستری یا مراجعه به روان‌پزشک | of admission or history referral to psychiatrist | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸ | مراجعه سرپایی به روان‌پزشک | opportunity to psychiatrist | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۹ | چند بار مراجعه سرپایی | outpatient | Numeric | ۰-۲۱ | |
| ۱۰ | تأهل | marital status | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ | ۱= مجرد، ۲=متأهل، ۳=زندگی جدا از همسر، ۴= مطلقه، ۵= همسر فوت‌شده |
| ۱۱ | سن شروع بیماری | age of the start of the disease | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ و ۶ و ۷ و ۸ | ۱= ۱۰-۱ سال، ۲= ۲۰-۱۱ سال، ۳= ۳۰-۲۱ سال، ۴= ۴۰-۳۱ سال، ۵= ۵۰-۴۱ سال، ۶= ۶۰-۵۱ سال، ۷=۷۰-۶۱ سال، ۸=بالای ۷۰ سال |
| ۱۲ | شغل | job | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ و ۶ و ۷ و ۸ | ۱= کارمند، ۲=آزاد، ۳=بیکار، ۴=خانه‌دار، ۵=محصل، ۶=کارگر، ۷=بازنشسته، ۸=سایر |
| ۱۳ | تحصیلات | education | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ و ۶ و ۷ و ۸ | ۱= بی‌سواد، ۲=ابتدایی، ۳=سیکل، ۴=دیپلم، ۵=کاردانی، ۶=لیسانس، ۷=فوق‌لیسانس، ۸=دکتری |
| ۱۴ | نوع ترخیص | type of clearance | Numeric | ۱ و ۲ و ۳ | ۱= دائم، ۲= رضایتی، ۳=فراری |
| ۱۵ | فصل بیماری | season of disease | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ | ۱=بهار، ۲=تابستان، ۳=پاییز، ۴= زمستان |
| ۱۶ | فاصله زمانی شروع علائم و بستری | the time interval of signs and admission | Numeric | ۱ و ۲ و ۳ | ۱=کمتر از یک هفته، ۲=یک ماه، ۳=بیشتر از یک ماه |
| ۱۷ | نوع انتقال بیمار به بیمارستان | patient transfer type to hospital | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ و ۶ | ۱=بستگان درجه اول، ۲=بستگان درجه دوم، ۳=دوست/همکار، ۴=اورژانس، ۵= انتقال از مرکزی دیگر، ۶=نامه قضایی |
| ۱۸ | مصرف مواد | consumption of materials | Numeric | ۱ و ۲ و ۳ و ۴ | ۱=بله ۲=خیر، ۳=مشکوک، ۴=ثبت‌نشده |
| ۱۹ | نوع مواد مصرفی | type of consumables | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ و ۶ و ۷ و ۸ | ۰=ندارد، ۱=تریاک، ۲=سیگار، ۳=شیشه، ۴=متادون، ۵=الکل، ۶=ترامادول، ۷=مورفین، ۸=حشیش، ۹=کراک، ۱۰=مشکوک |
| ۲۰ | سابقه بیماری زمینه‌ای | a history of underlying illness | Numeric | ۱ و ۲ و ۳ | ۱=بله، ۲=خیر، ۳=نامعلوم |
| ۲۱ | نوع بیماری زمینه‌ای | type of underlying disease | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ و ۶ و ۷ و ۸ و ۹ و ۱۰ و ۱۱ و ۱۲ و ۱۳ و ۱۴ و ۱۵ و ۱۶ | ۱=خیر، ۲=مغز و اعصاب، ۳=عفونی، ۴=داخلی، ۵=ترمیمی، ۶=روان، ۷=پوست، ۸=گوش و حلق و بینی، ۹=جراح عمومی، ۱۰=نوروسرجری، ۱۱=جراح عروق، ۱۲=رادیولوژی، ۱۳=توان‌بخشی، ۱۴=چشم، ۱۵=تغذیه، ۱۶=قلب |
| ۲۲ | بیمه | insurance | Numeric | ۱ و ۲ | ۱=بله، ۲=خیر |
| ۲۳ | نوع بیمه | insurance type | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ | ۰=ندارد، ۱=تأمین اجتماعی، ۲=خدمات درمانی، ۳=نیروی مصلح، ۴=کمیته امداد، ۵=آزاد |

[ Downloaded from jhbmi.com on 2026-11-17 ]

[ DOI: 10.34172/jhbmi.2022.01 ]

| | | | | | |
|---|---|---|---|---|---|
| ۲۴ | مشاوره شد | was concealed | Numeric | اعداد صحیح | تعداد مشاوره |
| ۲۵ | نوع مشاوره | consultancy type | Numeric | ۱ و ۲ و ۳ و ۴ و ۵ و ۶ و ۷ و ۸ و ۹ و ۱۰ و ۱۱ و ۱۲ و ۱۳ و ۱۴ و ۱۵ و ۱۶ | ۱=ندارد، ۲=مغز و اعصاب، ۳=عفونی، ۴=داخلی، ۵=ترمیمی، ۶=روان، ۷=پوست، ۸=گوش و حلق و بینی، ۹=جراح عمومی، ۱۰=اعصاب، ۱۱=جراح عروق، ۱۲=رادیولوژی، ۱۳=توان‌بخشی، ۱۴=چشم، ۱۵=تغذیه، ۱۶=قلب |
| ۲۶ | اختلال در تمرکز و توجه حواس‌پرتی | disruption of focus and attention-distance | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۲۷ | اختلال به شکل کم‌خوابی | disturbance in shape | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۲۸ | اعتقادات غیرواقعی در مورد توانمندی خود | unrealistic beliefs about their ability | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۲۹ | بی‌قراری | restlessness | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۰ | افزایش انرژی و میزان فعالیت | increased energy and activity | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۱ | خلق خیلی بالا و خودبزرگ‌بینی | the very high mood and self-important | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۲ | تحریک‌پذیری مفرط | excessive irritability | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۳ | صحبت کردن بی‌وقفه | talking endless | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۴ | خساست | stinginess | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۵ | افزایش تمایلات جنسی | increase sexuality | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۶ | نشان دادن حرکات ضدونقیض | show anti-conflict gestures | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۷ | سوءمصرف مواد | drug abuse | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۸ | رفتارهای اغواگرانه | seductive behaviors | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۳۹ | رفتار متفاوت از حالت معمول که مدت طولانی ادامه داشته | different behavior than the usual state that lasted | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۰ | هذیان | delirium | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۱ | کاهش اشتها | reduce appetite | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۲ | اقدام به خودکشی | suicide | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۳ | دیگرآزاری | Hurting others | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۴ | فرار از منزل | Running from home | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۵ | وراثت | inheritance | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۶ | اختلال در خواب به شکل پرخوابی | disorder in the shape of a drizzle | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۷ | توهم | delusion | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۸ | توهمات بینایی | vision illusions | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۴۹ | توهمات شنوایی | hearing illusions | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۰ | اختلال فکر | thought disorder | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۱ | اختلال حرکت | motion disorder | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۲ | یکنواختی احساسی | emotional uniformity | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۳ | کاهش لذت از زندگی روزمره | reduce the pleasure of everyday life | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۴ | مشکل در آغاز و ادامه فعالیت | problem in and continue the activity | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۵ | کاهش صحبت | reduce the talk | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |

[ Downloaded from jhbmi.com on 2026-11-17 ]
[ DOI: 10.34172/jhbmi.2022.01 ]

| | | | | | |
|---|---|---|---|---|---|
| ۵۶ | عملکرد اجرایی ضعیف | poor executive function | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۷ | اختلال در حافظه فعال | disruption of active memory | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۸ | عدم انگیزه | lack of motivation | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۵۹ | عدم آگاهی از بیماری | lack of knowledge of illness. | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۰ | بیان ضعیف احساسات | poor expression of feelings | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۱ | افزایش اشتها | increased appetite | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۲ | خلق‌وخوی افسرده | depressed mood | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۳ | اضطراب | anxiety | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۴ | شکاکیت | skeptical | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۵ | گوشه‌گیری | calling | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۶ | ناتوانایی در تصمیم‌گیری | disabling in decision making | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۷ | دوری از مسئولیت‌ها | away from responsibilities | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۸ | ترس از ترد شدن | fear of crispy | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۶۹ | حساسیت بیش‌ازحد نسبت به انتقاد | excessive sensitivity to criticism | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۰ | بدبینی | pessimism | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۱ | اجتناب از مخالفت | avoid opposition | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۲ | مشکل با تنهایی | the problem with loneliness | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۳ | ارجح دانستن نیاز دیگران | preferred to know the need for others | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۴ | ناشنوا | deaf | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۵ | رفتارهای تکراری | duplicate behaviors | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۶ | ناتوانی در گفتن اسم | inability to say name | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۷ | ناتوانی از شروع مکالمه | inability to start a conversation | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۸ | تشنج | convulsion | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۷۹ | استعداد خاص | special talent | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۰ | شکست در گذشته | failed in the past | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۱ | کاهش تحمل درد | reduce pain tolerance | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۲ | تمایل به حمایت اجتماعی | tendency to social support | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۳ | عقب‌ماندگی ذهنی | mental retardation | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۴ | گیجی | confusion | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۵ | پراکنده | diffused | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۶ | نقص شناختی | cognitive impairment | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۷ | از دست دادن احساس گرم نسبت به خانواده یا دوستان | loss of warm feelings to family or friends | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۸ | سوءمصرف مواد | drug abuse2 | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۸۹ | هیپوکندریازیس | hypochondriasis | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۹۰ | تکانه خودکشی | suicidal impulses | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |
| ۹۱ | ترومای کودکی | childhood trauma | Binary | ۰ و ۱ | ۰=خیر، ۱=بله |

[ Downloaded from jhbmi.com on 2026-11-17 ]
[ DOI: 10.34172/jhbmi.2022.01 ]

روش تحلیل مؤلفه اصلی (Principal Component Analysis)PCA راهکاری برای کاهش ابعاد مسئله تحقیق بود. ضرورت استانداردبودن داده‌های ورودی مورد بررسی در این روش با هدف دوری از داده‌های پرت و قرارگیری در یک مقیاس عددی، دلیلی بر استفاده از تابع StandardScaler کتابخانه پیش‌پردازش SKlearn بود. با اعمال این تابع روی داده‌ها، تمامی داده‌های جدید از توزیع گوسی پیروی کردند و دارای میانگین صفر و انحراف استاندارد ۱ شدند. در این تحقیق، از تابع PCA کتابخانه Scikit-learn پایتون برای کاهش ابعاد مسئله استفاده شد و مقدار n_components (کم‌ترین تعداد ویژگی‌ها) برابر با ۳۵ در نظر گرفته شد. این مقدار با بررسی نمودار PCA که بیانگر جای‌گیری ۹۰ درصد واریانس در ۳۵ مؤلفه اصلی بود، انتخاب شد. برای svd_solver نیز از مقدار full استفاده شد که پس از انجام کل محاسبات، بردارهای ویژه کم‌اهمیت را حذف کرد. در زبان پایتون پس از کاهش ویژگی، مجموعه داده در اختیار الگوریتم‌های یادگیری ماشین قرار گرفت. الگوریتم‌های یادگیری ماشین روی مجموعه داده‌ها آموزش دیده و آزمایش شدند تا دقت طبقه‌بندی بیماران طبق شش رده اختلال خلقی بر اساس داده‌های موجود روی داده‌های آموزش ندیده بررسی شود. مسئله تعیین اختلالات خلقی بر اساس داده‌های پرسشنامه‌های اختلالات خلقی، از نوع طبقه‌بندی چند رده‌ای (Multi-Class Classification) است. الگوریتم‌های طبقه‌بندی یادگیری ماشین مناسب چند رده‌ای مورد استفاده عبارت‌اند از: K نزدیک‌ترین همسایه (K-Nearest Neighbors)KNN، درخت تصمیم (Decision Tree)، نایو بیز (Naive Bayes)NB، جنگل تصادفی (Random Forest)RF، رگرسیون لجستیک (Logistic Regression)LR، ماشین بردار پشتیبان (Support Vector Machine)SVM. الگوریتم KNN یک الگوریتم غیر پارامتریک است که با محاسبه فاصله بین هر نقطه جدید با نقاط داده و رأی‌گیری رده نقطه جدید را اختصاص می‌دهد. الگوریتم KNN به ازای ۵ همسایه استفاده شد (n_neighbors=۵). در الگوریتم درخت تصمیم نمونه‌ها دسته‌بندی می‌شوند. در ساختار درخت تصمیم، پیش‌بینی به دست آمده از درخت در قالب یک سری قواعد توضیح داده می‌شود. هر مسیر از ریشه تا یک برگ درخت تصمیم، یک قانون را بیان می‌کند و در نهایت برگ با رده‌ای که بیشترین مقدار رکورد در آن تعلق گرفته برچسب می‌خورد. max_depth تعیین‌کننده حداکثر عمق درخت است. از عمق هفت برای ساختن درخت تصمیم خود استفاده شد (max_depth=۷). روش‌های نایو بیز، مجموعه‌ای از الگوریتم‌های یادگیری با نظارت بر اساس به‌کارگیری قضیه بیز با فرض بیز استقلال شرطی بین هر زوج ویژگی و مقدار رده در مجموعه داده است. از تابع GaussianNB از کتابخانه SKLearn پایتون استفاده شد. الگوریتم جنگل تصادفی مجموعه‌ای از الگوریتم‌های درخت تصمیم است. در این الگوریتم هر درخت در مجموعه از نمونه‌ای ساخته می‌شود که با جایگزینی از مجموعه آموزشی ترسیم شده است. max_depth یک درخت در جنگل تصادفی طولانی‌ترین مسیر بین گره ریشه و گره برگ است با استفاده از این پارامتر، می‌توان عمق هر درخت را تعیین نمود که در این جا برابر با ۳ در نظر گرفته شد. رگرسیون لجستیک فرآیند مدل‌سازی احتمال یک خروجی گسسته با داشتن یک متغیر ورودی است. با توجه به این‌که مسئله از نوع چند رده‌ای است، پارامتر multi_class برابر 'multinomial' قرار داده شد. هدف از الگوریتم ماشین بردار پشتیبان به‌عنوان یک طبقه بند، یافتن یک ابر صفحه در فضای n بعدی است (n تعداد ویژگی‌هاست) که با دقت بالا نقاط داده را رده‌بندی کند. برای اجرای الگوریتم ماشین بردار پشتیبان از تابع طبقه‌بندی‌کننده ماشین بردار پشتیبان از کتابخانه SKLearn استفاده شد. از بین این الگوریتم‌ها، دو الگوریتمی که دقت بالای ۰/۹۰ داشتند انتخاب شده و با ترکیب با الگوریتم‌های ژنتیک و بهینه‌سازی ازدحام ذرات، سعی شد دقت در طبقه‌بندی افزایش و خطا کاهش یابد.

الگوریتم ژنتیک یک جستجوی اکتشافی است که از نظریه تکامل طبیعی چارلز داروین الهام گرفته شده است. این الگوریتم فرآیند انتخاب طبیعی را منعکس می‌کند که در آن بهترین افراد برای تولیدمثل انتخاب می‌شوند تا فرزندان نسل بعدی تولید شوند. الگوریتم بهینه‌سازی ازدحام ذرات یک رویکرد تصادفی مبتنی بر جمعیت برای حل مسائل بهینه‌سازی پیوسته و گسسته است. در بهینه‌سازی ازدحام ذرات، عامل‌های نرم‌افزاری ساده‌ای که ذرات نامیده می‌شوند، در فضای جستجوی یک مسئله بهینه سازی حرکت می‌کنند. موقعیت یک ذره یک راه‌حل کاندید برای مسئله بهینه‌سازی در دست است. هر ذره با تغییر سرعت خود بر اساس قوانینی که در اصل از مدل‌های رفتاری دسته پرندگان الهام گرفته شده بود، به‌دنبال موقعیت‌های بهتر در فضای جستجو می‌شود.

## معیارهای ارزیابی مدل

چهار معیار ارزیابی Precision، Recall، F1_score و Accuracy در این تحقیق برای ارزیابی کارایی الگوریتم‌ها



[ Downloaded from jhbmi.com on 2026-11-17 ]
[ DOI: 10.34172/jhbmi.2022.01 ]

استفاده شده است.

Accuracy: دقت، شهودی‌ترین معیار عملکرد است. این معیار نسبت مشاهدات پیش‌بینی شده به کل مشاهدات است. دقت، زمانی معیار مناسبی است که مجموعه داده‌ها از نظر مثبت کاذب و منفی کاذب متقارن باشد. دقت را می‌توان به‌صورت رابطه ذکر شده تعریف کرد؛ که در آن TP، FN، FP و TN به ترتیب تعداد مثبت‌های واقعی، منفی‌های کاذب، مثبت‌های کاذب و منفی‌های واقعی را نشان می‌دهند.

$$Accuracy = (TP + TN)/(TP + FP + FN + TN)$$

Precision: صحت، نسبت مشاهدات مثبت پیش‌بینی شده صحیح به کل مشاهدات مثبت پیش‌بینی شده است. صحت به‌عنوان تعداد اسناد صحیح برگشتی توسط مدل در نظر گرفته می‌شود. صحت از رابطه زیر به دست می‌آید.

$$Precision = TP/(TP + FP)$$

Recall: یادآوری (حساسیت) – یادآوری نسبت مشاهدات مثبت پیش‌بینی شده درست به همه مشاهدات در کلاس واقعی است. رابطه این معیار به شرح زیر است.

$$Recall = TP/(TP + FN)$$

F1_score: معیار F1 معیاری مناسب برای ارزیابی دقت یک آزمایش است. این معیار هر دوی صحت و یادآوری را در نظر می‌گیرد. این معیار در بهترین حالت یک و در بدترین حالت صفر است و از رابطه زیر به دست می‌آید.

$$F1_score = 2 * (Recall * Precision)/(Recall + Precision)$$

## نتایج

نتایج تحقیق حاضر در دو بخش مجزا، شامل نتایج ارزیابی الگوریتم‌های داده‌کاوی و نتایج ترکیب الگوریتم‌های فراابتکاری با الگوریتم‌های داده‌کاوی منتخب ارائه می‌گردد.

### نتایج ارزیابی الگوریتم‌ها

الگوریتم‌های داده‌کاوی به‌کاررفته توسط معیارهای Precision، Recall، و F1_Score ارزیابی شدند. مقادیر حاصل در جدول ۲ برای شش کلاس اوتیسم (C1)، دوقطبی (C2)، افسردگی (C3)، اختلال شخصیت (C4)، روان‌پریشی (C5) و اسکیزوفرنی (C6) به طور مجزا گزارش شده است.

**جدول ۲: نتایج ارزیابی الگوریتم‌ها با معیارهای مختلف**

| معیار | Precision | Recall | F1-score | Precision | Recall | F1-score | Precision | Recall | F1-score |
|---|---|---|---|---|---|---|---|---|---|
| کلاس (رده/برچسب) | الگوریتم NB | | | الگوریتم RL | | | الگوریتم KNN | | |
| اوتیسم | ۰/۸۳ | ۰/۸۱ | ۰/۸۲ | ۰/۸۴ | ۰/۹ | ۰/۸۷ | ۰/۸۹ | ۰/۸۸ | ۰/۸۸ |
| دوقطبی | ۰/۸۹ | ۰/۷۹ | ۰/۸۴ | ۰/۸۷ | ۰/۸۴ | ۰/۸۵ | ۰/۸۷ | ۰/۸۷ | ۰/۸۷ |
| افسردگی | ۰/۹۴ | ۰/۸۳ | ۰/۸۸ | ۰/۹۳ | ۰/۸۵ | ۰/۸۹ | ۰/۸۷ | ۰/۸۵ | ۰/۸۹ |
| اختلال شخصیت | ۰/۹۱ | ۰/۹۳ | ۰/۹۲ | ۰/۹ | ۰/۹۶ | ۰/۹۳ | ۰/۹ | ۰/۹۶ | ۰/۹۳ |
| روان‌پریشی | ۰/۷۸ | ۰/۹۱ | ۰/۸۴ | ۰/۸۷ | ۰/۸۹ | ۰/۸۸ | ۰/۸۵ | ۰/۸۹ | ۰/۸۷ |
| اسکیزوفرنی | ۰/۸ | ۰/۸۹ | ۰/۸۴ | ۰/۸۵ | ۰/۹۳ | ۰/۹۴ | ۰/۹۳ | ۰/۹۳ | ۰/۹۳ |
| کلاس (رده/برچسب) | الگوریتم RF | | | الگوریتم DT | | | الگوریتم SVM | | |
| اوتیسم | ۰/۹۲ | ۰/۹۲ | ۰/۹۲ | ۰/۸ | ۰/۸۵ | ۰/۸۳ | ۰/۸۳ | ۰/۹۲ | ۰/۸۷ |
| دوقطبی | ۰/۸۹ | ۰/۸۷ | ۰/۸۸ | ۰/۷۷ | ۰/۷۵ | ۰/۷۶ | ۰/۸۹ | ۰/۸۶ | ۰/۸۷ |
| افسردگی | ۰/۹۵ | ۰/۸۷ | ۰/۹ | ۰/۹۱ | ۰/۸ | ۰/۸۵ | ۰/۹۴ | ۰/۸۵ | ۰/۸۹ |
| اختلال شخصیت | ۰/۹ | ۰/۹۶ | ۰/۹۳ | ۰/۹۱ | ۰/۹۱ | ۰/۹۱ | ۰/۹۲ | ۰/۹۶ | ۰/۹۴ |
| روان‌پریشی | ۰/۸۵ | ۰/۸۹ | ۰/۸۷ | ۰/۷۶ | ۰/۸۹ | ۰/۸۲ | ۰/۸۷ | ۰/۸۹ | ۰/۸۸ |
| اسکیزوفرنی | ۰/۹۳ | ۰/۹۳ | ۰/۹۳ | ۰/۸۸ | ۰/۸۶ | ۰/۸۷ | ۰/۹۵ | ۰/۹۳ | ۰/۹۴ |



[ Downloaded from jhbmi.com on 2026-11-17 ]

[ DOI: 10.34172/jhbmi.2022.01 ]

برای هر الگوریتم داده‌کاوی، به‌طور مجزا درایه‌های ماتریس درهم‌ریختگی (Confusion Matrix) شش کلاس محاسبه گردید. این ماتریس بیانگر مقادیر TP، FN، FP و TN برای هر کلاس است که با کمک آن‌ها مقدار متوسط دقت برای هر الگوریتم‌ها محاسبه می‌شود. به عنوان نمونه، مقادیر شش ماتریس درهم‌ریختگی مربوط به الگوریتم جنگل تصادفی در جدول ۳ نشان داده شده است.

**جدول ۳: ماتریس‌های درهم‌ریختگی ارزیابی الگوریتم RF**

| پیش‌بینی شده / واقعی | C1 + | C1 - | C2 + | C2 - | C3 + | C3 - | C4 + | C4 - | C5 + | C5 - | C6 + | C6 - |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C1 + | ۴۴ | ۴ | | | | | | | | | | |
| C1 - | ۴ | ۲۲۰ | | | | | | | | | | |
| C2 + | | | ۵۵ | ۸ | | | | | | | | |
| C2 - | | | ۷ | ۲۰۲ | | | | | | | | |
| C3 + | | | | | ۵۲ | ۸ | | | | | | |
| C3 - | | | | | ۳ | ۲۰۹ | | | | | | |
| C4 + | | | | | | | ۵۵ | ۲ | | | | |
| C4 - | | | | | | | ۶ | ۲۰۹ | | | | |
| C5 + | | | | | | | | | ۳۹ | ۵ | | |
| C5 - | | | | | | | | | ۷ | ۲۲۱ | | |
| C6 + | | | | | | | | | | | ۵۳ | ۴ |
| C6 - | | | | | | | | | | | ۴ | ۲۲۱ |

با توجه به جدول ۳، به‌عنوان مثال، مقادیر TP، FN، FP و TN برای کلاس اوتیسم به ترتیب برابر است با ۴۴، ۴، ۴، ۲۲۰. این بدان معناست که در میان ۳۲۹ رکورد آزمایشی که شامل ۴۸ رکورد مربوط به رده اوتیسم هستند، ۴۴ رکورد به درستی برچسب‌گذاری شده‌اند، ۴ رکورد به اشتباهی در رده دیگری پیش‌بینی شده‌اند، ۴ رکورد به اشتباهی در این رده پیش‌بینی شده‌اند و مابقی به تعداد ۲۲۰ رکورد به درستی در این رده قرار نگرفته‌اند.

متوسط دقت محاسبه شده برای هر یک از الگوریتم‌های به‌کار رفته در تحقیق در جدول ۴ گردآوری شده است. ملاحظه می‌گردد دقیق‌ترین الگوریتم‌ها در پیش‌بینی انواع اختلالات خلقی، به ترتیب جنگل تصادفی و بردار پشتیبان تصمیم هستند.

**جدول ٤: مقایسه دقت پیش‌بینی الگوریتم‌های داده‌کاوی**

| | الگوریتم NB | الگوریتم RL | الگوریتم KNN | الگوریتم RF | الگوریتم DT | الگوریتم SVM |
|---|---|---|---|---|---|---|
| دقت (Accuracy) | ۰/۸۶ | ۰/۸۹ | ۰/۸۹ | ۰/۹۱ | ۰/۸۴ | ۰/۹۰ |

## نتایج ترکیب الگوریتم‌های فرابتکاری با الگوریتم‌های داده‌کاوی

همان‌طورکه در جدول ۴ نشان داده شده است، در میان شش الگوریتم پیشنهادی، الگوریتم‌های RF و SVM دقت ۰/۹ و بالاتر را کسب نمودند؛ لذا از الگوریتم ژنتیک و بهینه‌سازی ازدحام ذرات برای تعیین مقادیر بهینه پارامترهای آن‌ها استفاده شد. در ادامه، نتایج حاصل در چهار زیربخش ارائه می‌گردد.

### نتیجه اجرای GA برای تعیین مقادیر بهینه پارامترها

برای تعیین مقادیر بهینه پارامترهای RF، در الگوریتم GA جمعیت اولیه ۳۰ و تعداد نسل‌ها ۱۰ انتخاب گردید. در این رابطه شبه‌کد زیر در پایتون نوشته شد:

```
evolved_estimator = GASearchCV(estimator= RandomForestClassifier),
                    cv=cv,
                    scoring='accuracy',
                    param_grid=param_grid,
                    population_size=30,
                    generations=10,
                    n_jobs=-1,
                    verbose=True)
```



[ Downloaded from jhbmi.com on 2026-11-17 ]

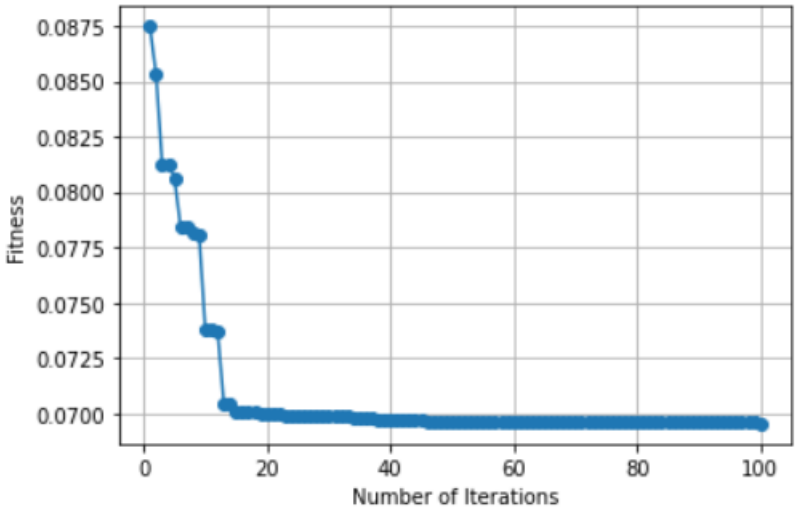
[ DOI: 10.34172/jhbmi.2022.01 ]

در این شـــبه کد، cv به معنای تعداد فولد ها در Cross-validation و scoring نشـــان‌دهنده نوع معیار ارز یابی الگوریتم است. param_grid مجموعه پارامترهای تعیین شده برای الگوریتم و n_jobs= -۱ به معنای اســـتفاده از حداکثر توان CPU و در نهایت verbose نشـــان‌دهنده خروجی با دقت به‌دست‌آمده است. دقت به‌دست‌آمده برابر با ۹۰/۵۷ درصد است.

همچنین، برای تعیین مقادیر بهینه SVM، در الگوریتم GA، جمعیت اولیه ۵۰ و تعداد نسل‌ها ۸۰ انتخاب شد. دقت به دست آمده برابر با ۹۰/۵۷ درصد است.

## نتیجه ترکیب الگوریتم‌های PSO و RF

در بهینه‌ســـازی ازدحام ذرات هدف کمینه کردن خطا در تابع هدف است. در این تحقیق، مقادیر پارامترهای الگوریتم PSO برای تعداد ذرات (N)، تعداد تکرار الگوریتم (T)، ضـــریب اینرسی (W)، ضـریب شـخصـی (C1)، و ضـریب اجتماعی (C2) به ترتیب برابر با ۳۰، ۱۵۰، ۰,۹، ۲، و ۲ تنظیم گردید. نتیجه ترکیب الگوریتم بهینه‌ســـازی ازدحام ذرات و جنگل تصادفی در شکل ۲ نمایش داده شده است.

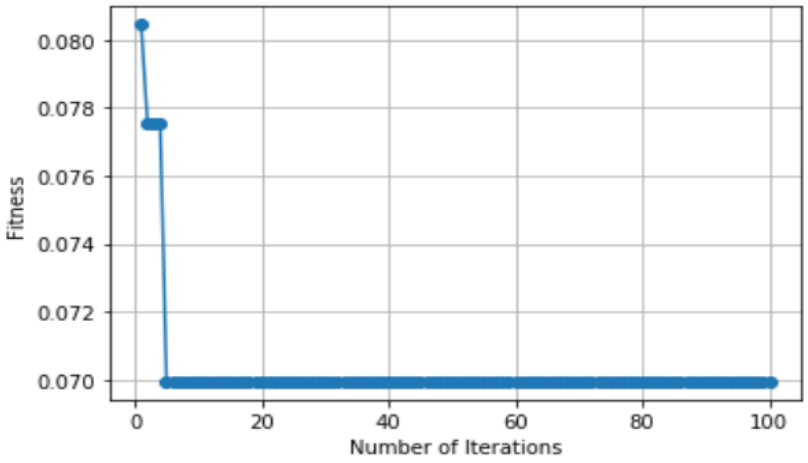



**شکل ۲: نمودار ارزیابی PSO-RF**

برای این طراحی، پس از ۱۰۰ تکرار و با انتخاب ۳۴ ویژگی، دقتی برابر با ۹۱/۳۰۴۳ درصد حاصل گردید.

## نتیجه ترکیب الگوریتم‌های PSO و SVM

مشابه قبل، برای پارامترهای الگوریتم PSO، مقادیر N، T، W، C1 و C2 به ترتیب برابر با ۳۰، ۱۵۰، ۰/۹، ۲، و ۲ در نظر گرفته شد. نتیجه اجرای ترکیب این الگوریتم با الگوریتم ماشین بردار پشتیبان در شکل ۳ به نمایش درآمده است. دقت مدل طراحی شده پس از ۱۰۰ تکرار و با انتخاب ۴۵ ویژگی برابر با ۹۳/۳۰۴۳ درصد بوده است.




**شکل ۳: نمودار ارزیابی PSO-SVM**



[ Downloaded from jhbmi.com on 2026-11-17 ]

[ DOI: 10.34172/jhbmi.2022.01 ]

در شکل‌های ۲ و ۳، مقادیر محور عمودی عبارت است از اختلاف بین مقدار پیش‌بینی شده متغیر هدف و متغیر هدف واقعی در تکرارهای مختلف. به بیان دیگر، محور عمودی، مقادیر تابع هزینه PSO را نشان می‌دهد و بر اساس کمترین مقدار تابع هزینه تعداد ویژگی‌ها به‌دست می‌آید.

## بحث و نتیجه‌گیری

در این پژوهش مسئله طبقه‌بندی بیماران اختلالات خلقی بررسی شده است. با توجه به اهمیت نشاط در جامعه و ضرورت کاهش زمان تشخیص اولیه بیماری‌های روانی خلقی و اتخاذ رویکرد روان‌درمانی مناسب هر یک، موضوع تحقیق از اهمیت ویژه‌ای برخوردار است. برای بهبود دقت از الگوریتم‌های ژنتیک و بهینه‌سازی ازدحام ذرات یاری گرفته شد. ترکیب با ژنتیک دقت دو الگوریتم ماشین بردار پشتیبان و جنگل تصادفی را چندان بهبود نمی‌دهد. این امر می‌تواند به دلیل مشکل بهینه محلی (Local Minimal) باشد که یکی از اشکالات شایع الگوریتم ژنتیک است و زمانی اتفاق می‌افتد که الگوریتم بهترین محلی را با بهترین سراسری اشتباه می‌گیرد. از سوی دیگر، الگوریتم بهینه‌سازی ذرات در ترکیب با هر دو الگوریتم جنگل تصادفی و ماشین بردار پشتیبان باعث بهبود قابل قبولی در دقت پیش‌بینی می‌شود. نتایج ترکیب الگوریتم ماشین بردار پشتیبان و ازدحام ذرات با ۹۳/۳۰ درصد نسبت به ترکیب الگوریتم جنگل تصادفی و ازدحام ذرات با ۹۱/۹۷ درصد دقت عملکرد بالاتری نشان می‌دهد و کاندید مناسبی برای اعمال طبقه‌بندی روی مجموعه داده‌های مسئله است.

مقایسه نتایج تحقیق حاضر با پژوهش‌های مشابه نشان می‌دهد درمیان الگوریتم‌های به‌کار گرفته شده، دو الگوریتم جنگل تصادفی با دقت ۹۱ و ماشین بردار پشتیبان با دقت ۹۰ درصد عملکرد بهتری نشان داده‌اند. دقت مدل یادگیری عمیق ارائه شده توسط Liu و همکاران [۱۴] ۶۵ درصد، مدل جنگل تصادفی مورد استفاده Rotenberg و همکاران [۲۷] ۸۰ درصد، مدل یادگیری ماشین Bayes و همکاران [۲۸] ۷۳/۹ درصد، مدل مورد استفاده Eeden و همکاران [۱۲] ۷۹ درصد، مدل ترکیبی Liu و همکاران ۸۷/۵ درصد [۸] و مدل‌های Nguyen و همکاران [۲۹] ۰/۸۶ و ۰/۷۹، برآورد شده بود.

مدل‌های طراحی شده در این تحقیق، PSO-RF و PSO-SVM، دستاورد مهمی برای طراحی سیستم پشتیبان تصمیم کلینیک‌های روان‌پزشکی است. با توجه به محدودیت‌های مسئله از جمله کوچک بودن ابعاد مجموعه داده، در آینده می‌توان با جمع‌آوری پرسشنامه‌های بیش‌تر، مجموعه داده‌های غنی‌تری تهیه کرد تا نتایج تحقیق تعمیم‌پذیر گردد. همچنین، پیشنهاد می‌گردد الگوریتم‌های یادگیری عمیق و یا الگوریتم‌های داده‌کاوی فازی برای مدل‌سازی به‌کار گرفته شوند.

## تشکر و قدردانی

این مقاله مستخرج از پایان‌نامه کارشناسی ارشد گروه مهندسی صنایع با عنوان «استفاده از روش‌های داده‌کاوی برای پیش‌بینی و تشخیص بیماری دوقطبی» است که در سال ۱۴۰۰ در دانشگاه آزاد اسلامی تهران واحد علوم تحقیقات انجام گردید.

## تعارض منافع

نویسندگان مقاله اعلام می‌کنند که این پژوهش هیچ‌گونه تعارض منافعی ندارد.

**References**

**1.** Parsapour F, Peymani J. Using Data Mining Techniques for Intelligent Diagnosis of Severity of Depressive Disorder. Journal of Health and Biomedical Informatics 2020;7(3):252-62. [In Persian]

**2.** World health organization (WHO). Depression; 2022 [cited 2021 May 1]. Available from: https://www.who.int/news-room/fact-sheets/detail/depression

**3.** Mansourian M, Khademi S, Marateb HR. A comprehensive review of computer-aided diagnosis of major mental and neurological disorders and suicide: a biostatistical perspective on data mining. Diagnostics 2021;11(3):393. 10.3390/diagnostics11030393

**4.** Guo T, Zhao W, Alrashoud M, Tolba A, Firmin S, Xia F. Multimodal Educational Data Fusion for Students'Mental Health Detection. IEEE Access 2022;10:70370-82. 10.1109/ACCESS.2022.3187502

**5.** Vigo D, Thornicroft G, Atun R. Estimating the true global burden of mental illness. The Lancet Psychiatry 2016;3(2):171-8. 10.1016/S2215-0366(15)00505-2

**6.** König H, König HH, Konnopka A. The excess costs of depression: a systematic review and meta-analysis. Epidemiology and psychiatric sciences. Epidemiology and Psychiatric Sciences 2020; 29: e30. 10.1017/S2045796019000180

**7.** Baumeister H, Hutter N, Bengel J, Härter M. Quality of life in medically ill persons with comorbid mental disorders: a systematic review and meta-analysis. Psychotherapy and Psychosomatics 2011;80(5):275-86. 10.1159/000323404

**8.** Liu YS, Chokka S, Cao B, Chokka PR. Screening for bipolar disorder in a tertiary mental health center using

[ Downloaded from jhbmi.com on 2026-11-17 ]

[ DOI: 10.34172/jhbmi.2022.01 ]

EarlyDetect: A machine learning-based pilot study. Journal of Affective Disorders Reports 2021;6:100215.
**9.** Sadr Ameli H. The effect of topiramate on bipolar mood disorder [dissertation]. Tehran: Shahid Beheshti University of Medical Sciences; 2014. [In Persian]
**10.** James SB, Alcott SV. Kaplan and Sadock's Concise Textbook of Clinical Psychiatry. 3rd ed. Philadelphia: Lippincott Williams & Wilkins; 2008.
**11.** Isometsä E. Suicidal behaviour in mood disorders—who, when, and why?. The Canadian Journal of Psychiatry 2014;59(3):120-30. 10.1177/070674371405900303
**12.** van Eeden WA, Luo C, van Hemert AM, Carlier IV, Penninx BW, Wardenaar KJ, et al. Predicting the 9-year course of mood and anxiety disorders with automated machine learning: A comparison between auto-sklearn, naïve Bayes classifier, and traditional logistic regression. Psychiatry Research 2021;299:113823. 10.1016/j.psychres.2021.113823
**13.** Lei Y, Belkacem AN, Wang X, Sha S, Wang C, Chen C. A convolutional neural network-based diagnostic method using resting-state electroencephalograph signals for major depressive and bipolar disorders. Biomedical Signal Processing and Control 2022;72:103370. 10.1016/j.bspc.2021.103370
**14.** Liu Y, Qu HQ, Mentch FD, Qu J, Chang X, Nguyen K, et al. Application of deep learning algorithm on whole genome sequencing data uncovers structural variants associated with multiple mental disorders in African American patients. Molecular Psychiatry 2022;27(3):1469-78.
**15.** Molavi P, Salvat H, Aghdam NS, Mohammadi MN. The efficacy of Cognitive-Behavioral Therapy (CBT+ TAU) on reducing type-I Bipolar disorder symptoms. Shenakht Journal of Psychology and Psychiatry 2018;7(2):92-103. [In Persian] 10.52547/shenakht.7.2.92
**16.** Tari Barazin N, Babapour J, Aghdasi A. Comparison of the effectiveness of social skills training and training of cognitive emotion regulation strategies on social adjustment and positive and negative emotions in Man with type I bipolar disorder. Journal of Applied Family Therapy 2021;2(4):137-57.
**17.** Phillips M, Kupfer D. Bipolar disorder diagnosis: challenges and future directions. Lancet 2013;381(9878):1663-71. 10.1016/S0140-6736(13)60989-7.
**18.** Han J, Kamber M, Mining D. Concepts and techniques. Morgan Kaufmann 2006;340:94104-3205.
**19.** Shailaja K, Seetharamulu B, Jabbar MA. Machine learning in healthcare: A review. Second International Conference on Electronics, Communication and Aerospace Technology (ICECA); 2018 Mar 29; Coimbatore, India: IEEE; 2018. p. 910-4. 10.1109/ICECA.2018.8474918
**20.** Bleidorn W, Hopwood CJ. Using machine learning to advance personality assessment and theory. Personality and Social Psychology Review 2019;23(2):190-203. 10.1177/1088868318772990
**21.** Chancellor S, De Choudhury M. Methods in predictive techniques for mental health status on social media: a critical review. NPJ Digital Medicine 2020;3(1):1.
**22.** Llamocca P, Urgelés D, Cukic M, Lopez V. Bip4Cast: Some advances in mood disorders data analysis. 1st International 'Alan Turing' Conference on Decision Support and Recommender Systems (DSRS-Turing'19); 2019 Nov 21-22; London: The Alan Turing Institute; 2019.
**23.** Čukić M, López V. Progress in Objective Detection of Depression and Online Monitoring of Patients Based on Physiological Complexity. Frontiers in Psychiatry 2022;13:828773.
**24.** Mallick S, Panda M. Application of Deep Learning in Mental Disorder: Challenges and Opportunities. Next Generation Healthcare Informatics 2022:295-308.
**25.** Squarcina L, Villa FM, Nobile M, Grisan E, Brambilla P. Deep learning for the prediction of treatment response in depression. Journal of Affective Disorders 2021;281:618-22. 10.1016/j.jad.2020.11.104
**26.** Najafabadi MM, Villanustre F, Khoshgoftaar TM, Seliya N, Wald R, Muharemagic E. Deep learning applications and challenges in big data analytics. Journal of Big Data 2015;2(1):1-21.
**27.** de Siqueira Rotenberg L, Borges-Júnior RG, Lafer B, Salvini R, da Silva Dias R. Exploring machine learning to predict depressive relapses of bipolar disorder patients. Journal of Affective Disorders 2021;295:681-7. 10.1016/j.jad.2021.08.127
**28.** Bayes A, Spoelma MJ, Hadzi-Pavlovic D, Parker G. Differentiation of bipolar disorder versus borderline personality disorder: A machine learning approach. Journal of Affective Disorders 2021;288:68-73. 10.1016/j.jad.2021.03.082
**29.** Nguyen DK, Chan CL, Li AH, Phan DV, Lan CH. Decision support system for the differentiation of schizophrenia and mood disorders using multiple deep learning models on wearable devices data. Health Informatics Journal 2022;28(4):14604582221137537. 10.1177/1460458222113753

[ Downloaded from jhbmi.com on 2026-11-17 ]

[ DOI: 10.34172/jhbmi.2022.01 ]