Journal of Health and Biomedical Informatics 2026; 13(1): 1-11
doi:10.34172/jhbmi.2026.08
Medical Informatics Research Center, Kerman University of Medical Sciences

Original article

Predicting Depression Using Data Mining Techniques and Artificial Intelligence Algorithms

Behzad Rabiee-Ghahfarrokhi [1]*, Zahra Babadi Akashe[2]

1. B.S.c in Psychology, Faculty of Psychology, Payam Noor University, Farrokh Shahr Center, Chaharmahal and Bakhtiari, Iran
2 Assistant Professor, Department of Education, Payame Noor University, Tehran, Iran

ARTICLE INFO:

Article History:

Received: 31 Dec 2025

Accepted: 6 Apr 2026

Published: 21 Jun 2026

Corresponding Author:

Behzad Rabiee-Ghahfarrokhi

Email:
b.rabiee1364@gmail.com

Citation: Rabiee-Ghahfarrokhi B, Babadi Akashe Z. Predicting Depression Using Data Mining Techniques and Artificial Intelligence Algorithms. Journal of Health and Biomedical Informatics 2026; 13(1): 1-11. [In Persian]

Abstract

Introduction: Modern lifestyles have led to various mental health disorders and significant psychological issues—such as anxiety, stress, and depression—in many individuals. Psychologists encounter challenges during the interview process that stem from the process itself, the interviewee, and the interviewer; these errors can be attributed to the intricate and highly complex network of interactions occurring during face-to-face exchanges between the interviewer and the interviewee. Given that human beings and human interactions are inherently prone to error, and that the diagnosis of disorders is often influenced by the emotions and feelings of both the client and the therapist, there is a critical need for a tool capable of predicting depression with high accuracy and without human intervention.

Method: In this study, real data was used to predict depression and demonstrate the proposed method. The dataset comprised 600 individuals, of whom 297 were diagnosed as not having depression and 303 as having depression. The C5.0 decision tree was employed to extract rules, while a genetic algorithm was used to select the optimal rules yielding the highest prediction accuracy. The proposed method was evaluated based on the weighted count of correctly predicted samples.

Results: The results demonstrated that the proposed method offers significantly higher accuracy than many existing machine learning techniques. Furthermore, among all the extracted rules, those exhibiting the greatest impact—along with minimal overlap and conflict—regarding the detection of the presence or absence of depression were selected. Additionally, the algorithm identified the most influential features associated with depression. Based on the analyzed data, suicidal ideation, hypochondriasis, and childhood trauma were identified, in that order, as three significant and influential factors in depression.

Conclusion: The use of data mining techniques and machine learning tools is highly efficient and effective for predicting depression. Extracting the most significant features influencing the output variable can substantially reduce time and costs for clients while enhancing the accuracy of the therapist's diagnosis.

Keywords: Data Mining, Depression, Artificial Intelligence, Genetic Algorithm, Decision Tree

[DOI: 10.34172/jhbmi.2026.08]

مجله انفورماتیک سلامت و زیست پزشکی دوره سیزدهم، شماره اول، سال۱۴۰۵، ص ۱۱-۱
doi: 10.34172/jhbmi.2026.08
مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان

مقاله پژوهشی

پیش‌بینی افسردگی با استفاده از تکنیک‌های داده‌کاوی و الگوریتم‌های هوش مصنوعی

بهزاد ربیعی قهفرخی[1]*، زهرا بابادی عکاشه[2]

۱. کارشناسی روان‌شناسی، دانشکده روان‌شناسی، دانشگاه پیام نور، مرکز فرخ‌شهر، چهارمحال و بختیاری، ایران

۲. استادیار گروه علوم تربیتی، دانشگاه پیام نور، تهران، ایران

اطلاعات مقاله

سابقه مقاله

دریافت: ۱۴۰۴/۱۰/۱۰

پذیرش: ۱۴۰۵/۱/۱۷

انتشار برخط: ۱۴۰۵/۳/۳۱

نویسنده مسئول:

بهزاد ربیعی قهفرخی

ایمیل:

b.rabiee1364@gmail.com

ارجاع: ربیعی قهفرخی بهزاد، بابادی عکاشه زهرا. پیش‌بینی افسردگی با استفاده از تکنیک‌های داده‌کاوی و الگوریتم‌های هوش مصنوعی. مجله انفورماتیک سلامت و زیست پزشکی ۱۴۰۵؛ ۱۳(۱): ۱-۱۱.

چکیده:

مقدمه: سبک زندگی مدرن امروزی باعث ایجاد انواع مختلف اختلالات سلامت روان در بسیاری از افراد و مشکلات روانی زیادی مانند اضطراب، استرس و افسردگی شده است. در انجام مصاحبه، روانشناس با مشکلاتی مواجه است که معلول فرآیند مصاحبه، مصاحبه شونده و مصاحبه‌گر می‌باشند که این اشتباهات را می‌توان معلول شبکه بغرنج و بسیار پیچیده کنش‌هایی دانست که در روابط چهره به چهره بین مصاحبه‌گر و مصاحبه شونده صورت می‌گیرد. با توجه به اینکه انسان و روابط انسانی همیشه مستعد بروز خطا بوده و تشخیص اختلالات، بعضاً تابعی از عواطف و احساسات مراجع و درمانگر می‌باشد، استفاده از ابزاری که بتواند بدون دخالت انسان و با دقت بسیار بالا افسردگی را پیش‌بینی کند، ضروری‌است.

روش کار: در این پژوهش، از داده‌های واقعی برای پیش‌بینی افسردگی و معرفی روش پیشنهادی استفاده شد. این داده‌ها مربوط به ۶۰۰ نفر مراجعه کننده است که از میان آن‌ها ۲۹۷ نفر بدون افسردگی و ۳۰۳ نفر نیز دارای افسردگی تشخیص داده شده‌اند. در این پژوهش از درخت تصمیم C5.0 جهت استخراج قوانین و از الگوریتم ژنتیک برای انتخاب بهترین قوانین با بالاترین دقت پیش‌بینی استفاده گردید. ارزیابی روش پیشنهادی با استفاده از معیار تعداد نمونه‌های به‌درستی پیش‌بینی‌شده با وزن‌دهی انجام شد.

یافته‌ها: نتایج نشان داد که روش پیشنهادی دقتی به مراتب بالاتر از بسیاری از روش‌های یادگیری ماشین ارائه می‌دهد همچنین از میان تمامی قوانین استخراج شده، همچنین از میان تمامی قوانین استخراج‌شده، قوانینی انتخاب گردیدند که بیشترین تأثیر و کمترین همپوشانی و تعارض را در تشخیص افسردگی یا عدم افسردگی داشتند. افزون بر این مؤثرترین ویژگی‌های دخیل در افسردگی نیز توسط الگوریتم مشخص گردید. براساس داده‌های مورد بررسی، به ترتیب انگیزه خودکشی، خود بیمار انگاری یا هیپوکندریازیس و ترومای دوران کودکی، به عنوان سه عامل مهم و مؤثر در افسردگی تشخیص داده شدند.

نتیجه‌گیری: استفاده از تکنیک‌های داده کاوی و ابزارهای یادگیری ماشین در پیش‌بینی افسردگی بسیار کارا و مؤثر است. استخراج مهم‌ترین ویژگی‌های تأثیرگذار بر متغیر خروجی می‌تواند تأثیر بسزایی در کاهش زمان و هزینه مراجعان و همچنین افزایش دقت تشخیص درمانگر داشته باشد.

کلیدواژه‌ها: داده کاوی، افسردگی، هوش مصنوعی، الگوریتم ژنتیک، درخت تصمیم

[DOI: 10.34172/jhbmi.2026.08]

مقدمه

افسردگی به صورت احساس بی‌ارزش بودن، ناراحتی مداوم و فقدان انگیزه برای شرکت در فعالیت‌هایی که پیش‌تر برای فرد لذت‌بخش بوده‌اند، تعریف می‌شود. افسردگی یک احساس غم گذرا که همه ممکن است هر از گاهی آن را تجربه کنند، نیست، بلکه یک بیماری جسمی ذهنی پیچیده است که در عملکرد روزمره فرد اختلال ایجاد می‌کند. سبک زندگی مدرن امروزی باعث ایجاد انواع مختلف اختلالات سلامت روان در بسیاری از افراد و مشکلات روانی زیادی مانند اضطراب، استرس و افسردگی شده است[۱]. وقتی از افسردگی صحبت می‌کنیم، منظور اختلال سلامت روانی است که با غم و اندوه مکرر در طول حداقل ۲ هفته مشخص می‌شود که باعث ناتوانی در انجام فعالیت‌های روزانه شده و افراد مبتلا علاقه خود را به انجام کارهایی که قبلاً از آن لذت می‌بردند از دست می‌دهند. افسردگی شایع ترین وضعیت سلامت روان، با نرخ شیوع جهانی یک ساله بین ۷ تا ۲۱ درصد در سطح جهان است. کیفیت زندگی می‌تواند به طور جدی توسط این اختلال مختل شود، طبق تحقیقات صورت پذیرفته در سال ۲۰۲۰، بیش از یک میلیارد نفر در سراسر جهان دارای اختلالات روانی می‌باشند[۲،۳]. و بیش از ۳۰۰ میلیون نفر در سراسر جهان افسردگی دارند. امروزه افسردگی در افکار خطرناکی همانند خودکشی خود را بروز داده به‌طوری‌که سالانه حدود ۸۰۰۰۰۰ نفر در جهان به‌خاطر افسردگی و عوارض آن دست به خودکشی می‌زنند، بنابراین مقابله و تشخیص این بیماری روانی بیش از پیش ضروری است [۴،۵].

از جمله عواملی که باعث افسردگی می‌شوند می‌توان به عملکرد غیرطبیعی و متابولیسم واسطه‌های سیستم عصبی مرکزی، ژنتیک، سابقه افسردگی، عدم حمایت اجتماعی، جنبه‌های ذهنی، مشکلات جسمی، سوء مصرف مواد و ترومای دوران کودکی اشاره کرد[۸-۶]. از دست دادن حافظه یکی از علائم اصلی افسردگی از دیدگاه بالینی است، همچنین عدم تمرکز، ناتوانی در تصمیم‌گیری، از دست دادن علاقه به فعالیت‌های تفریحی و سرگرمی‌ها از جمله رابطه جنسی، پرخوری و افزایش وزن، کاهش اشتها و کاهش وزن، احساس گناه و بی‌ارزشی، درماندگی، بی‌قراری، تحریک‌پذیری و همچنین افکار خودکشی از مهم‌ترین علائم افسردگی می‌باشند [۹].

در روانشناسی بالینی، تشخیص با مشاهده، آزمون و مصاحبه انجام می‌شود که در بین روش‌های مذکور مصاحبه از اهمیت بالاتری برخوردار است. هدف از مصاحبه تشخیصی، یافتن دقیق علل و عوامل مؤثر در پیدایش اختلال و مشکل روانی در مراجع می‌باشد. در این رابطه مصاحبه باید بتواند کلیه مسائل و عوامل مؤثر در پیدایش اختلال را از کم اهمیت‌ترین تا پراهمیت‌ترین شناسایی کرده و بتواند تشخیص مناسبی در ارتباط با مشکل درمانجو در اختیار مصاحبه گر قرار دهد، زیرا تا زمانی که تشخیصی دقیق و درست از نوع اختلال بیمار در دست نباشد، نمی‌توان درمان مناسبی برای او ارائه نمود [۱۰].

در انجام مصاحبه، روانشناس با مشکلاتی مواجه است که از اعتبار مصاحبه به عنوان یک وسیله ارزشیابی و تشخیصی می‌کاهد. این مشکلات ناشی از اشتباهاتی هستند که معلول فرآیند مصاحبه، مصاحبه شونده و مصاحبه‌گر می‌باشند. اشتباهات ناشی از فرآیند مصاحبه را می‌توان معلول شبکه بغرنج و بسیار پیچیده کنش‌هایی دانست که در روابط چهره به چهره بین مصاحبه‌گر و مصاحبه شونده صورت می گیرد. موقعیتی که در آن مصاحبه صورت می گیرد نیز ممکن است خطاهایی را ایجاد کند. اشتباه دیگر ناشی از ترس‌ها و دلهره‌های مصاحبه شونده و نقش‌هایی است که ایفای آن از وی انتظار می‌رود. اشتباه و مشکل دیگر ناشی از خود مصاحبه‌گر است که بیشتر معلول ویژگی‌های شخصی او، اشتباه در یادداشت‌برداری جریان مصاحبه، استنباط غلط، نداشتن آموزش کافی و مکتبی است که مصاحبه گر از آن پیروی دارد، می‌کند [۱۱]. از طرفی، اختلالات روانی به دلیل گستردگی و داشتن علائم متعدد، پیچیدگی بالایی در شناسایی دارند به ویژه آن که برخی از این اختلالات دارای علائم مشترکی نیز هستند و این مسئله، تشخیص دقیق و در نتیجه برنامه‌ریزی درمانی و آموزشی مناسب را برای آن‌ها دشوار می‌کند. تشخیص بیماری‌های روانی به دلیل نشانه‌های ذهنی و درونی آن پیچیده است و این پیچیدگی با توجه به این حقیقت که نشانه‌ها ممکن است پیشرونده باشند و تشخیص اولیه حتی توسط روانپزشک مجرب هم ممکن نباشد، بیشتر می‌شود. به علاوه یک ننگ اجتماعی وسیع وجود دارد که ممکن است به طرد شدن بیماران روانی شوند. کمبود متخصصان روان آموزش دیده و باتجربه در بسیاری از مناطق جهان باعث نیز مشکلات را حادتر می‌کند. افسردگی مسئله‌ای جدی است که بخش بزرگی از مردم جهان را تحت تأثیر قرار می‌دهد و به علت علائم گسترده و مختلف، تشخیص مشکل به صورت غیرقطعی داده می‌شود [۱۰،۱۲].

[DOI: 10.34172/jhbmi.2026.08]

مجله انفورماتیک سلامت و زیست پزشکی دوره سیزدهم، شماره اول، سال۱۴۰۵، ص ۱۱-۱
doi: 10.34172/jhbmi.2026.08
مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان

با توجه به محدودیت‌های مربوط به مصاحبه، مصاحبه‌گر و مراجع و همچنین ورود سلیقه‌های شخصی به فرآیند مصاحبه و از سوی دیگر به دلیل پیشرفت در قدرت پردازش، حافظه، ذخیره‌سازی و حجم بی‌سابقه‌ای از داده‌ها، در چند سال اخیر استفاده از رایانه‌ها با توجه به موفقیت خیره‌کننده ای که داشته‌اند، به پیش بینی بسیاری از امور اقدام کرده‌اند. رایانه‌ها اکنون بر بازی محبوب پوکر مسلط شده‌اند، قوانین فیزیک را از داده‌های تجربی یاد گرفته‌اند و در بازی‌های ویدیویی متخصص شده‌اند، کارهایی که چندی پیش غیرممکن تلقی می‌شدند. به موازات این پیشرفت‌ها، به دلیل حجم بسیار بالای استفاده از سیستم های هوشمند در پزشکی و موفقیت ماشین‌ها در تشخیص و پیش‌بینی بسیاری از ناهنجاری‌ها و بیماری‌ها، الگوریتم‌های هوش مصنوعی و یادگیری ماشین در علم پزشکی و روانشناسی به‌شدت مورد استقبال قرار گرفته و چه بسا در آینده ای نزدیک، ماشین‌ها جای انسان‌ها را بگیرند[۱۳].

در پژوهش Vanlalawmpuia و همکاران، نتایج نشان می‌دهد که به‌کارگیری روش‌های داده‌کاوی بر روی محتوای کاربران شبکه‌های اجتماعی می‌تواند در پیش‌بینی و شناسایی به‌موقع نشانه‌های افسردگی مؤثر باشد. همچنین نتایج حاکی از آن است که تحلیل الگوهای زبانی و رفتاری در محیط‌های مجازی می‌تواند ابزاری ارزشمند برای متخصصان سلامت روان به‌منظور مداخله زودهنگام باشد[۱۴].

در مطالعه‌ای که Thien و همکاران انجام دادند، از داده‌های نظرسنجی سلامت و تغذیه ملی ایالات متحده برای پیش‌بینی افسردگی با استفاده از شش مدل یادگیری ماشین استفاده شد. داده‌ها به دو مجموعه آموزشی (۸۰ درصد) و آزمایشی (۲۰ درصد) تقسیم شدند و عملکرد مدل‌ها با استفاده از معیارهایی مانند دقت، حساسیت، مساحت زیر منحنی، دقت پیش‌بینی مثبت و غیره بررسی گردید. نتایج نشان داد که XGBoost بهترین عملکرد را داشته است و مهم‌ترین پیش‌بینی‌کننده‌های افسردگی عبارت بودند از: نسبت درآمد خانوار به فقر، جنسیت، فشار خون بالا، سطح سرمی کتونین و هیدروکسی‌کتونین، شاخص توده بدنی، سطح تحصیلات، سطح گلوکز، سن، وضعیت تأهل و عملکرد کلیه‌ها [۱۵].

با توجه به اهمیت تشخیص افسردگی پیش از آن‌که بیماری وارد فاز بحرانی شود، همچنین اهمیت تشخیص دقیق این اختلال، می‌توان از ابزارهای هوش مصنوعی، داده‌کاوی و یادگیری ماشین که با دقتی بسیار بالا افسردگی را تشخیص می‌دهند استفاده کرد. ضمن اینکه این ابزارها مسائل و مشکلات مربوط به مصاحبه حضوری، تأثیر محیط مصاحبه بر طرفین، و حتی مسائل مربوط به مسیر رسیدن مصاحبه‌شونده به دفتر مصاحبه‌گر (مانند ترافیک، تصادف، بحث و بگو مگوهای افراد جامعه، تأخیر وسایل ارتباطی و غیره) را از بین می‌برند، لذا بر آن شدیم تا در مطالعه حاضر، روشی برای پیش‌بینی افسردگی با دقت بالا ارائه دهیم.

روش کار

در این پژوهش، از یک مجموعه داده واقعی از سایت معتبر Kaggle (https://www.kaggle.com/) جهت پیش‌بینی افسردگی و معرفی روش پیشنهادی استفاده شد. این دیتاست مربوط به ۶۰۰ نفر مراجعه‌کننده بوده است که تعداد ۲۹۷ نفر از آن‌ها فاقد افسردگی و ۳۰۳ نفر نیز دارای افسردگی تشخیص‌داده‌شده هستند. پرسشنامه مربوطه دارای ۳۰ متغیر می‌باشد که در جدول ۱ مشاهده می‌شوند. به منظور توضیح بهتر روش پیشنهادی و استفاده راحت‌تر از قوانین استخراج‌شده، متغیرهای مربوطه به صورت A، B، C، ...، AA، AB، AC، AD نشان داده شده‌اند.

در این پژوهش، روشی جدید برای بهبود دقت پیش‌بینی افسردگی ارائه گردید. درخت تصمیم ابزاری مناسب برای کشف روابط بین داده‌ها و روشی قدرتمند برای طبقه‌بندی است که در آن، داده‌های طبقه‌بندی‌شده بر اساس قوانین استخراج می‌شوند. روش‌های مختلفی برای استخراج قوانین از داده‌ها وجود دارد که به نظر می‌رسد بهترین روش، درخت تصمیم C5 باشد. با معرفی یک مجموعه داده به درخت تصمیم، مجموعه‌ای از قوانین طبقه‌بندی به شکل «اگر ... آنگاه ...» استخراج می‌شود. این قواعد از نظر جذابیت متفاوت هستند؛ برخی از آن‌ها زائد و ناسازگار می‌باشند و ممکن است با یکدیگر همپوشانی داشته باشند. بنابراین، استفاده از قوانین برتر برای بهبود سرعت و دقت در بازیابی دانش از داده‌ها ضروری است [۱۶].

[DOI: 10.34172/jhbmi.2026.08]

جدول ۱: متغیرهای پایگاه داده

ویژگی	توضیح	ویژگی	توضیح	ویژگی	توضیح
A	Sadness	K	Poor memory	U	Desire for social support
B	Discouragement	L	Lose libido	V	Psychomotor retardation
C	Low self-esteem	M	Sluggish	W	Confusion
D	Inferiority	N	Crying spells	X	Scatterbrained
E	Guilt	O	Lack of emotional responsiveness	Y	Cognitive impairment
F	Indecisiveness	P	Helplessness	Z	Loss of warm feeling toward family or friends
G	Irritability and frustration	Q	Pessimism	AA	Substance abuse
H	Loss of interest in life	R	Agitation	AB	Hypochondriacs
I	Loss of motivation	S	Past failure	AC	Suicidal impulse
J	Poor self-image	T	Reduced pain tolerance	AD	Childhood trauma

در این پژوهش، از درخت تصمیم C5.0 به عنوان استخراج‌کننده قوانین از مجموعه داده‌ها استفاده شد. برای به‌دست آوردن نتایج قابل اعتماد و معتبر، این الگوریتم در قالب اعتبارسنجی متقابل ۱۰-تایی (۱۰-fold cross-validation) روی مجموعه داده به کار گرفته شد[۱۷]. خروجی الگوریتم C5.0 منجر به چندین مجموعه‌قانون می‌شود که هر مجموعه از این قوانین، دقت ویژه و متمایزی را در مجموعه داده ارائه می‌کند و به دلیل استفاده از داده‌های آموزشی متفاوت (ناشی از به‌کارگیری تکنیک اعتبارسنجی متقابل ۱۰-تایی)، تعداد قوانین متفاوتی استخراج می‌گردد. پس از استفاده از درخت تصمیم C5.0 و انجام تنظیمات مربوطه، پنج مجموعه‌قانون استخراج شد. سپس دقت همه مجموعه‌قوانین را در رابطه با پیش‌بینی افسردگی در داده مربوطه به دست آورده شد و حداکثر دقت به عنوان معیار اعداد شاخه انتخاب شد. جزئیات مجموعه‌قوانین و دقت مربوط به آن‌ها در پایگاه داده، در جدول ۲ نشان داده شده است.

جدول ۲: ویژگی‌های مجموعه قوانین

مجموعه قانون	تعداد قوانین	دقت مجموعه قانون بر روی پایگاه داده
۱	۱۲	۷۸/۳
۲	۸	۸۱/۴
۳	۹	۸۳/۳
۴	۱۴	۸۵/۲
۵	۱۰	۸۳/۳

الگوریتم ژنتیک استفاده شده در این پژوهش جهت انتخاب بهترین قوانین استخراج شده نیز به صورت زیر عمل می کند:
پس از استخراج قوانین، تعداد N قانون طبقه‌بندی در قالب «اگر ... آنگاه«... (If ... Then ...) استخراج می‌شود که در اینجا N مجموع کل قوانین استخراج‌شده است. طول کروموزوم استفاده‌شده برابر با تعداد تمام قوانین است که در پژوهش حاضر برابر با ۵۳ می‌باشد ۵۳ =(۱۲ + ۸ + ۹ + ۱۴ + ۱۰) هر قانون به یک ژن نگاشت می‌شود؛ به این معنی که اولین ژن نشان‌دهنده قانون اول، ژن دوم نشان‌دهنده قانون دوم، و به همین ترتیب تمامی قوانین بر روی ژن‌های مربوطه نگاشت می‌گردند. در ادامه، از شمارنده‌ای به نام K استفاده شد که از ۱ تا N افزایش می‌یابد و K قوانین انتخابی تصادفی را نشان می‌دهد. به عنوان مثال، اگر K= ۵ باشد، در تمام ۵۳ ژن، ۵ ژن به طور تصادفی به عنوان «۱» انتخاب می‌شوند و بقیه به عنوان «۰» در نظر گرفته می‌شوند. قانون مربوط به ژن‌هایی که مقدارشان

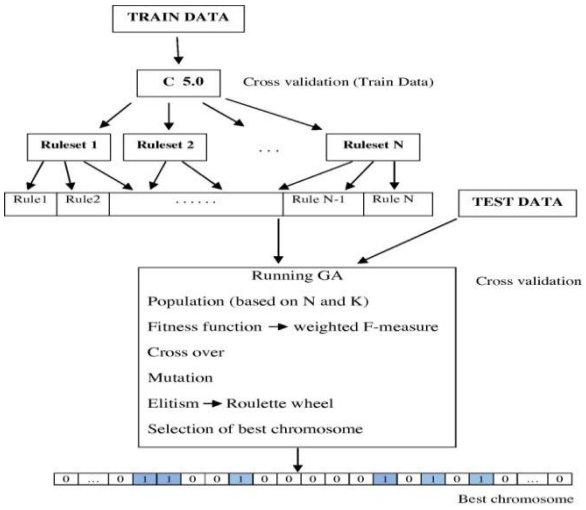
[DOI: 10.34172/jhbmi.2026.08]

مجله انفورماتیک سلامت و زیست پزشکی دوره سیزدهم، شماره اول، سال۱۴۰۵، ص ۱۱-۱
doi: 10.34172/jhbmi.2026.08
مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان

برابر با ۱ است بر روی مجموعه داده اعمال می‌گردد و سایر قوانین نادیده گرفته می‌شوند. ساختار الگوریتم پیشنهادی در شکل ۱ نشان داده شده است.

در الگوریتم پیشنهادی، تمام مراحل الگوریتم ژنتیک در نظر گرفته شده است. در مرحله اول، جمعیت اولیه به طور تصادفی با توجه به اندازه K و N تولید می‌شود. به عنوان مثال، اگر K= ۱۰ و N= ۵۳ باشد، در جمعیت اولیه، در همه کروموزوم‌ها به طور تصادفی ۱۰ ژن دارای مقدار ۱ و ۴۳ ژن دیگر دارای مقدار ۰ خواهند بود. سپس با توجه به قوانین اعمال‌شده بر روی پایگاه داده، عملکرد هر کروموزوم محاسبه می‌گردد. برای ارزیابی عملکرد هر کروموزوم از تابع Weighted F-Measure استفاده شد. این تابع محاسبه می‌کند که هر کروموزوم چه تعداد رکورد را می‌تواند از مجموعه داده در کلاس مربوطه آن، به درستی پیش‌بینی کند.

F-measure معیاری برای ارزیابی صحت طبقه‌بندی است. برای محاسبه F-measure، از دو معیار دیگر precision (دقت) و recall (بازیابی) استفاده می‌شود که هر دو از ماتریس سردرگمی مشتق می‌گردند. جدول ۳ ماتریس سردرگمی را نشان می‌دهد [۱۷].



شکل ۱: ساختار الگوریتم پیشنهادی

جدول ۳: ماتریس سردرگمی

	POSITIVE	NEGATIVE
POSITIVE	TP	FP
NEGATIVE	FN	TN

(TP: True Positive; FP: False Positive; FN: False Negative; TN: True Negative)
ACTUAL CLASS

همبری و جهش نیز با توجه به مقادیر تعریف‌شده برای الگوریتم اعمال می‌شوند. پس از اتمام فرآیند و در آخرین مرحله، کروموزومی که بیشترین مقدار Weighted F-measure را دارد، به عنوان خروجی نشان داده می‌شود. این کروموزوم نشان‌دهنده قوانینی با حداقل ناسازگاری و تضاد و بالاترین دقت پیش‌بینی است.

[DOI: 10.34172/jhbmi.2026.08]

نتایج

در این پژوهش، از تکنیک‌های داده‌کاوی (درخت تصمیم) و ابزارهای هوش مصنوعی (الگوریتم ژنتیک) برای استخراج بهترین قوانین مرتبط با پیش‌بینی افسردگی استفاده شد. پس از تنظیم پارامترهای برنامه و اجرای آن در نرم‌افزار متلب (Matlab R2015a)، تعداد ۱۲ قانون با بالاترین دقت پیش‌بینی، مطابق با جدول ۳ استخراج گردید. دقت پیش‌بینی مجموعه قوانین استخراج‌شده با روش پیشنهادی برابر با ۰/۹۵۷ است که نسبت به تک‌تک مجموعه قوانین مندرج در جدول ۴، دقت به مراتب بالاتری را به دست آورده است.

جدول ٤: مجموعه قوانین استخراجی با روش پیشنهادی

Rule number	Rule description
1	if AC = 0 then No
2	if E in [0 1 2] and AD <= 0 then No
3	if D in [1 2] and AD <= 0 then No
4	if S in [0 1 2] and AA = 0 then No
5	if AC = 1 and AD <= 1 then No
6	if H = 1 and I in [1 2] and J in [0 1 2] and AB = 1 then No
7	if F = 0 and H = 3 and I = 3 and J in [1 2] then No
8	if J = 3 and AA = 1 and AC = 1 then No
9	if F = 3 and AB = 3 and AC in [1 2 3]and AD > 0 then Yes
10	if AC in [1 2 3] then Yes
11	if H = 3 and AC in [2 3] and AD > 2 then Yes
12	if E in [1 2] and AA = 3 and AC in [1 2 3] and AD > 2 then Yes

برای مقایسه دقت پیش‌بینی روش پیشنهادی با دیگر روش‌های طبقه‌بندی موجود، از نرم‌افزار WEKA (نسخه ۳،۷،۹) استفاده شد [۱۸]. WEKA (http://www.cs.waikato.ac.nz/ml/weka/) یک نرم‌افزار متن‌باز است که از مجموعه‌ای از پیشرفته‌ترین الگوریتم‌های یادگیری ماشین و ابزارهای پیش‌پردازش داده تشکیل شده است. این نرم‌افزار توسط دانشگاه Waikato در نیوزیلند توسعه یافته است. سیستم WEKA که به زبان جاوا نوشته شده، می‌تواند برای عملیات مربوط به داده‌کاوی و یادگیری ماشین مورد استفاده قرار گیرد. این نرم‌افزار پیاده‌سازی‌ای از پیشرفته‌ترین الگوریتم‌های یادگیری ماشین ارائه می‌کند که اطلاعات موجود در مجموعه داده‌ها را استخراج نموده و می‌تواند برای مقایسه عملکرد الگوریتم‌های مختلف یادگیری ماشین روی مجموعه داده‌ها به کار رود.

پس از معرفی مجموعه داده به عنوان ورودی به WEKA، تعدادی از مهم‌ترین و معروف‌ترین الگوریتم‌های طبقه‌بندی موجود بر روی مجموعه داده اعمال شدند که نتایج در جدول ۵ نشان داده شده است. این جدول همچنین دقت روش پیشنهادی ما را نشان می‌دهد. واضح است که روش پیشنهادی ما در مقایسه با سایر طبقه‌بندی‌کننده‌های موجود، از دقت بالاتری برخوردار است. دقت بالاتر روش پیشنهادی می‌تواند مربوط به انتخاب بهترین قوانینی باشد که به بهترین نحو ممکن با یکدیگر تعامل داشته و کمترین تعارض را با هم دارند.

[DOI: 10.34172/jhbmi.2026.08]

مجله انفورماتیک سلامت و زیست پزشکی دوره سیزدهم، شماره اول، سال۱۴۰۵، ص ۱۱-۱
doi: 10.34172/jhbmi.2026.08
مرکز تحقیقات انفورماتیک پزشکی، دانشگاه علوم پزشکی کرمان

جدول ۵: دقت کلاس بندی روش‌های مختلف بر روی پایگاه داده

Algorithm	Accuracy (weighted F-measure)	Algorithm	Accuracy (weighted F-measure)
BayesNet	۰/۸۳۴	RotationForest	۰/۸۸۲
NaiveBeyes	۰/۸۹	DecisionTable	۰/۸۳۲
IBK	۰/۷۵۳	Adaboost	۰/۸۲۲
RandomForest	۰/۸۰۷	K-NN	۰/۷۳
RandomTree	۰/۷۴	NBTree	۰/۷۵۳
Bagging	۰/۸۴۵	Dagging	۰/۹۱۸
AttributeSelection	۰/۷۷۳	REPTree	۰/۷۷۸
J48	۰/۷۸۲	**Proposed method**	۰/۹۵۷

علاوه بر انتخاب مهم‌ترین قوانین استخراج‌شده از مجموعه قوانین، و با توجه به اینکه در مجموعه قوانین انتخابی بسیاری از معیارها چندین بار تکرار شده‌اند، بر آن شدیم تا پراهمیت‌ترین ویژگی‌های مؤثر بر خروجی در پایگاه داده بررسی شود. در جدول ۶، ده ویژگی از مهم‌ترین ویژگی‌های تأثیرگذار به ترتیب درج گردیده است. همان‌طور که مشخص می‌باشد، AC (انگیزه خودکشی) در رتبه اول، AB (خودبیمارانگاری یا هیپوکندریازیس) در رتبه دوم و AD (ترومای دوران کودکی) در رتبه سوم قرار دارند. مشخص کردن مهم‌ترین عوامل تأثیرگذار بر خروجی به روان‌درمانگران کمک می‌کند که با دقت بیشتری این عوامل را مد نظر داشته باشند و علاوه بر تشخیص افسردگی مراجع، برای درمان یا تجویز دارو بهتر عمل کنند.

جدول ٦: مهم‌ترین ویژگی‌های تأثیرگذار بر افسردگی

توضیح ویژگی	ویژگی	اهمیت	توضیح ویژگی	ویژگی	اهمیت
خودانگاره ضعیف	J	۶	انگیزه خودکشی	AC	۱
اختلال شناختی	Y	۷	هیپوکندریازیس	AB	۲
بلاتکلیفی	F	۸	ترومای دوران کودکی	AD	۳
احساس گناه	E	۹	از دست دادن علاقه به زندگی	H	۴
سوء مصرف مواد	AA	۱۰	احساس حقارت	D	۵

بحث و نتیجه‌گیری

در این پژوهش، از تکنیک‌های داده‌کاوی (درخت تصمیم) و ابزارهای هوش مصنوعی (الگوریتم ژنتیک) برای استخراج بهترین قوانین مرتبط با پیش‌بینی افسردگی استفاده شد. نتایج به‌دست‌آمده نشان داد که روش پیشنهادی این مطالعه در مقایسه با سایر طبقه‌بندی‌کننده‌های موجود، از دقت بالاتری برخوردار است. دقت بالاتر روش پیشنهادی را می‌توان با انتخاب بهترین قوانینی مرتبط دانست که به بهترین نحو ممکن با یکدیگر تعامل داشته و کمترین تعارض را با هم دارند. برای ارزیابی عملکرد هر کروموزوم در روش پیشنهادی، از بهترین معیار ارزیابی ممکن، یعنی تابع Weighted F-Measure استفاده شد. این تابع محاسبه می‌کند که هر کروموزوم چه تعداد رکورد را می‌تواند از مجموعه داده، در کلاس مربوطه آن، به درستی پیش‌بینی کند.

پس از تنظیم پارامترهای برنامه و اجرای آن در نرم‌افزار متلب، تعداد ۱۲ قانون با بالاترین دقت پیش‌بینی استخراج گردید. دقت پیش‌بینی مجموعه قوانین استخراج‌شده با روش پیشنهادی برابر با ۰/۹۵۷ است که دقتی به مراتب بالاتر از تک‌تک مجموعه قوانین اکتشافی از

طریق درخت تصمیم به دست آورده است. با توجه به این‌که سؤالات پرسشنامه در تشخیص افسردگی تأثیر متفاوتی دارند و پاسخ برخی از سؤالات تأثیر بیشتری در تشخیص خواهد داشت، مهم‌ترین ویژگی‌های تأثیرگذار بر افسردگی نیز استخراج شد که به ترتیب عبارتند از: AC (انگیزه خودکشی)، AB (خودبیمارانگاری یا هیپوکندریازیس)، AD (ترومای دوران کودکی) و سایر ویژگی‌ها در رتبه‌های اول تا سوم قرار دارند. مشخص کردن مهم‌ترین عوامل تأثیرگذار بر خروجی به روان‌درمانگران کمک می‌کند که با دقت بیشتری این عوامل را مد نظر قرار داده و مراجعین با پتانسیل بالاتر را با دقت و وسواس بیشتر تحت نظر داشته باشند.

با توجه به اهمیت تشخیص اختلال افسردگی، مطالعاتی توسط پژوهشگران مختلف در این زمینه انجام گرفته است. در سال‌های اخیر، پژوهش‌های متعددی با استفاده از روش‌های داده‌کاوی، انواع مختلف اختلالات و افسردگی‌ها را با دقتی بالا پیش‌بینی کرده‌اند.

Na و همکاران، شروع افسردگی در آینده جمهوری کره را پیش‌بینی کردند. مدل پیش‌بینی با استفاده از روش Random Forest ساخته شد. این مطالعه به دقت ۸۶/۲۰ درصد دست یافت و همچنین نشان داد که رضایت از روابط اجتماعی–خانوادگی و رضایت از سلامت، از عوامل کلیدی مؤثر در بروز افسردگی هستند[۱۹].

Sau و Bhakta مطالعه خود را بر روی جمعیت سالمندان انجام دادند. آنها با استفاده از ده طبقه‌بندی‌کننده مختلف یادگیری ماشین، ظهور افسردگی را در میان بیماران مسن پیش‌بینی کردند. عوامل اجتماعی–دموگرافیک و مرتبط با سلامت بیماران سالمند برای هدف طبقه‌بندی به کار گرفته شد. در میان ده طبقه‌بندی‌کننده، Random Forest بهترین عملکرد را نشان داد [۲۰].

Zarandi و همکاران، برای تشخیص میزان افسردگی از منطق فازی نوع ۲ استفاده کردند. برای افزایش دقت مطالعه و بهبود پیش‌بینی میزان افسردگی بیماران با تعداد سؤالات کمتر، از روش انتخاب ویژگی مبتنی بر اطلاعات متقابل (Mutual Information Feature Selection) MIFS بهره بردند. رویکرد پیشنهادی آن‌ها تنها از پانزده سؤال برای پیش‌بینی سطح افسردگی استفاده کرد و به دقت ۰/۰۰۸۴ درصد دست یافت[۲۱].

Sau و همکاران، الگوریتم‌های مختلف یادگیری ماشین مانند رگرسیون لجستیک، Catboost، Naïve Bayes، RFT و SVM را برای غربالگری اضطراب و افسردگی در میان دریانوردان به کار گرفتند. در این مطالعه، با ۴۷۰ دریانورد مصاحبه شد و اطلاعات مرتبط در مورد مشاغل، اطلاعات جمعیت‌شناختی و سلامت شرکت‌کنندگان با استفاده از ۱۶ ویژگی شامل سن، مدرک تحصیلی، درآمد ماهانه، وضعیت اشتغال، مدت خدمت، نوع خانواده، وضعیت تأهل، وضعیت فشار خون، دیابت یا بیماری‌های قلبی، مشخصات شغلی، رتبه در سازمان و غیره مورد مطالعه قرار گرفت. در این میان، Catboost بالاترین دقت را در بین همه طبقه‌بندی‌کننده‌ها به دست آورد[۲۲].

نقاط قوت روش پیشنهادی به شرح زیر است:

۱- دقت بسیار بالاتر روش پیشنهادی نسبت به روش‌های موجود

۲- انتخاب بهترین قوانین با کمترین همپوشانی و بیشترین کارایی از بین مجموعه قوانین

۳- امکان تعریف و استفاده از معیارهای ارزیابی متفاوت در روش پیشنهادی و انتخاب یکی از بهترین معیارها در این پژوهش

۴- استخراج مهم‌ترین ویژگی‌های تأثیرگذار بر خروجی جهت کمک بیشتر به درمانگر

۵- حذف محدودیت‌های مربوط به مصاحبه، مصاحبه‌گر و مراجع و همچنین ورود سلیقه‌های شخصی به فرآیند مصاحبه

مهم‌ترین محدودیت‌ها و موانع روش پیشنهادی نیز عبارتند از: مشکلات تهیه پایگاه داده، عدم تمایل بسیاری از مؤسسات مشاوره به ارائه اطلاعات مراجعان، وجود پرسشنامه‌های متفاوت و فرهنگ‌های مختلف در جوامع در رابطه با تهیه پرسشنامه‌های مورد استفاده. همچنین استفاده از روش‌های سنتی و دستی در مراکز مشاوره و مکانیزه نبودن بسیاری از امور مرتبط، از دیگر محدودیت‌های موجود است.

با توجه به اینکه پایگاه داده استاندارد و مورد قبولی در پایگاه داده‌های ایرانی مشاهده نمی‌شود، پیشنهاد می‌گردد تعدادی از مؤسسات مرتبط با همکاری اساتید و مراکز ذی‌صلاح، نسبت به تهیه و در اختیار گذاشتن چنین پایگاه داده‌ای اقدام نمایند.

تعارض منافع

در پژوهش حاضر هیچ‌گونه تعارض منافعی وجود ندارد.

[DOI: 10.34172/jhbmi.2026.08]

حمایت مالی
حمایت مالی از سوی هیچ ارگان یا مؤسسه‌ای صورت نپذیرفته است.

کد اخلاق
این مطالعه در شصت و هشتمین جلسه کمیته اخلاق در پژوهش دانشگاه پیام نور مورخ ۲۷/۰۹/۱۴۰۴ با شماره ۴۲۴۵/۷ /ص به تأیید کارگروه اخلاق در پژوهش دانشگاه رسیده است.

اعلام استفاده از هوش مصنوعی (AI)
نویسندگان اعلام می‌کنند که در انجام پژوهش، تحلیل داده‌ها، تفسیر یافته‌ها و نگارش این مقاله از ابزارها یا سامانه‌های هوش مصنوعی مولد استفاده نشده است.

سهم مشارکت نویسندگان
اجرا و پیش نویس مقاله: بهزاد ربیعی و بازنگری و ویرایش مقاله توسط زهرا بابادی عکاشه انجام شده است.

References
[1]. Kumar P, Garg S, Garg A. Assessment of anxiety, depression and stress using machine learning models. Procedia Computer Science 2020;171:1989-98.
[2]. Phillips MR. World Mental Health Day 2020: Promoting Global Mental Health During COVID-19. 2020. Chinese Center for Disease Control and Prevention 2020;2(3): 844-7.
[3]. World Health Organization (WHO). Mental health atlas 2020: review of the Eastern Mediterranean Region. 2022. [cited 2026 Jun 17]. Available from: https://iris.who.int/server/api/core/bitstreams/6bab42bc-df0f-4f68-a86d-28ebedb85e42/content
[4]. World Health Organization (WHO). Depression and Other Common Mental Disorders: Global Health Estimates. 2017. [cited 2026 Jun 17]. Available from: https://www.who.int/publications/i/item/depression-global-health-estimates
[5]. Wenzel A. The Routledge International Handbook of Perinatal Mental Health Disorders: Taylor & Francis; 2024.
[6]. Xiang J. The causes of depression and its social factor. Proceedings of the 2022 5th International Conference on Humanities Education and Social Sciences (ICHESS 2022); 2022 Oct 14-16: Atlantis Press; 2022. doi:10.2991/978-2-494069-89-3_352
[7]. Remes O, Mendes JF, Templeton P. Biological, psychological, and social determinants of depression: a review of recent literature. Brain Sci 2021;11(12):1633. doi: 10.3390/brainsci11121633
[8]. Kendall K, Van Assche E, Andlauer T, Choi K, Luykx J, Schulte E, et al. The genetic basis of major depression. Psychol Med 2021;51(13):2217-30. doi: 10.1017/S0033291721000441
[9]. Tyshchenko Y. Depression and anxiety detection from blog posts data [dissertation]. Estonia: Institute of Computer Science of University of Tartu; 2018.
[10]. Bakhtiari M, Ismailpour M, Ebrahimi M. Diagnosis of Depression Using Artificial Intelligence. Contemporary Psychology 2017;12:312-09.
[11]. Babbie ER. The practice of Social Research. 15th ed. Cengage Au; 2020.
[12]. Mukherjee S, Ashish K, Hui NB, Chattopadhyay S. Modeling depression data: feed forward neural network vs. radial basis function neural network. Am J Biomed Sci 2014;6(3):166-74. doi: 10.5099/aj140300166
[13]. Deo RC. Machine learning in medicine. Circulation 2015;132(20):1920-30. doi: 10.1161/CIRCULATIONAHA.115.001593
[14]. Vanlalawmpuia R, Lalhmingliana M. Prediction of depression in social network sites using data mining. 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS): EEE; 2020. doi:10.1109/ICICCS48265.2020.9120899
[15]. Vu T, Yamamoto M, Tay JT, Watanabe N, Kuriya Y, Oya A, et al. Prediction of depressive disorder using machine learning approaches: findings from the NHANES. BMC Med Inform Decis Mak 2025;25(1):83. doi: 10.1186/s12911-025-02903-1
[16]. Dreiseitl S, Osl M, Baumgartner C, Vinterbo S. An evaluation of heuristics for rule ranking. Artificial Intelligence in Medicine 2010;50(3):175-80.

[17]. Powers DM. Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:201016061. 2020. https://doi.org/10.48550/arXiv.2010.16061
[18]. Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software: an update. ACM SIGKDD Explorations Newsletter 2009;11(1):10-8.
[19]. Na KS, Cho SE, Geem ZW, Kim YK. Predicting future onset of depression among community dwelling adults in the Republic of Korea using a machine learning algorithm. Neurosci Lett 2020:721:134804. doi: 10.1016/j.neulet.2020.134804
[20]. Sau A, Bhakta I. Predicting anxiety and depression in elderly patients using machine learning technology. Healthcare Technology Letters 2017;4(6):238-43. doi:10.1049/htl.2016.0096
[21]. Zarandi MF, Soltanzadeh S, Mohammadi A, Castillo O. Designing a general type-2 fuzzy expert system for diagnosis of depression. Applied Soft Computing 2019;80:329-41. https://doi.org/10.1016/j.asoc.2019.03.027
[22]. Sau A, Bhakta I. Screening of anxiety and depression among seafarers using machine learning technology. Informatics in Medicine Unlocked 2019;16:100228. https://doi.org/10.1016/j.imu.2019.100228

[DOI: 10.34172/jhbmi.2026.08]