Journal of Health Administration

Spring 2022, Vol 25, Issue 1

Original Article

Diagnosis of Coronary Artery Disease by Bat and Harris Hawk Meta-Heuristic Optimization Algorithms and Machine Learning Methods

Sarina Maleki [1], Yahia Zare Mehrjerdi [2*]

[1] *M.Sc student, Department of Industrial Engineering, Faculty of Industry, Yazd University, Yazd, Iran.*
[2] *Professor, Department of Industrial Engineering, Faculty of Industry, Yazd University, Yazd, Iran.*

ARTICLE INFO

Corresponding Author:
Yahia Zare Mehrjerdi
e-mail addresses:
yzare @yazd.ac.ir

Received: 18/Dec/2021
Modified: 12/Mar/2022
Accepted: 19/Mar/2022
Available online: 13/Jul/2022

Keywords:
Coronary Artery Disease
Decision tree
k-nearest Neighbors
Harris Hawk

ABSTRACT

Introduction: Methods of detecting Coronary Artery Disease (CAD) are often prone to error and are also expensive and painful for the patient; therefore, the development and introduction of accurate machine learning-based methods for diagnosing this condition is of high importance. This research aimed to help detect coronary artery disease using the Harris Hawks Optimization (HHO) algorithm and machine learning techniques.

Methods: In this research, a novel approach based on feature selection was employed through a combination of HHO and machine learning techniques such as a Decision Tree (DT) and k-Nearest Neighbors algorithm (k-NN). To evaluate the proposed approach, we used two datasets (Cleveland & Z-Alizadeh-Sani) with medical records of 303 patients, and the evaluation was conducted by means of python 2016.

Results: On the basis of the findings of this research, feature selection by using the Harris hawks optimization algorithm in combination with machine learning methods resulted in an increase in the accuracy of the results in such a way that in the case of Z-Alizadeh-Sani dataset, the percentage of accuracy in combination with a decision tree was equal to 0.98 and in combination with the k-nearest neighbors algorithm was equal to 0.78. Furthermore, the results of the Cleveland dataset showed that using the HHO in combination with a decision tree led to 88 percent accuracy and in combination with the k-nearest neighbors algorithm led to 77 percent accuracy. However, in the case of using all of the features (HHO only mode), accuracy was lower in all cases. Therefore, the HHO algorithm in combination with the decision tree was able to achieve the highest accuracy in diagnosing CAD in the feature selection mode compared to using all of the features.

Conclusion: The results from this study showed that the Harris hawk optimization algorithm in combination with machine learning techniques can have a positive role in the process of selecting effective features in diagnosing coronary artery disease.

نشریه مدیریت سلامت

بهار ۱۴۰۱، دوره ۲۵، شماره ۱

مقاله اصلیل

تشخیص بیماری عروق کرونر قلبی با استفاده از الگوریتم فراابتکاری شاهین هریس و روش‌های یادگیری ماشین

سارینا ملکی [۱]، یحیی زراع مهرجردی [۲*]

[۱] دانشجوی کارشناسی ارشد، گروه مهندسی صنایع، دانشکده صنایع، دانشگاه یزد، یزد، ایران.

[۲] استاد، گروه مهندسی صنایع، دانشکده صنایع، دانشگاه یزد، یزد، ایران.

اطلاعات مقاله

نویسنده مسئول:
یحیی زراع مهرجردی
رایانامه:
yzare @yazd.ac.ir

وصول مقاله: ۱۴۰۰/۰۹/۲۷
اصلاح نهایی: ۱۴۰۰/۱۲/۲۱
پذیرش نهایی: ۱۴۰۰/۱۲/۲۸
انتشارآنلاین: ۱۴۰۱/۰۴/۲۲

واژه‌های کلیدی:
بیماری عروق کرونر
درخت تصمیم
کای نزدیک‌ترین همسایه
شاهین هریس

چکیـــــده

مقدمه: روش‌های تشخیص بیماری‌های عروق کرونر معمولاً در معرض خطا بوده و برای بیمار رنج‌آور و هزینه بر است. بنابراین، توسعه و ارائه روش‌های یادگیری ماشین با دقت بالا در تشخیص بیماریهای عروق کرونر از اهمیت خاصی برخوردار است. هدف این پژوهش کمک به تشخیص بیماری عروق کرونری با استفاده از الگوریتم شاهین هریس و روش‌های یادگیری ماشین می‌باشد.

روش‌ها: در این پژوهش از روش جدید مبتنی بر انتخاب ویژگی توسط شاهین هریس در حالت ترکیبی با روش یادگیری ماشین شامل درخت تصمیم و کای نزدیک‌ترین همسایه استفاده شد. که جهت ارزیابی روش پیشنهادی از دو مجموعه داده با پرونده پزشکی ۳۰۳ بیمار در مجموعه داده Cleveland و Z-Alizadeh-Sani استفاده گردید. ارزیابی مدل پیشنهادی در پایتون نسخه ۲۰۱۶ انجام شد.

یافته‌ها: بر اساس یافته‌ها انتخاب ویژگی با استفاده از الگوریتم شاهین هریس در حالت ترکیبی با روش یادگیری ماشین منجر به افزایش دقت در نتایج شد. به‌طوری‌که در مجموعه داده Z-Alizadeh-Sani درصد دقت در ترکیب با درخت تصمیم برابر با ۰/۹۸ و در ترکیب با کای نزدیک‌ترین همسایه برابر با ۰/۷۸ بود. همچنین، در مجموعه داده Cleveland درصد دقت در ترکیب با درخت تصمیم برابر با ۰/۸۸ و در ترکیب با کای نزدیک‌ترین همسایه برابر با ۰/۷۷ بود. این در حالی است که هریک از این مقادیر در حالت تمام ویژگی‌ها مقادیر کمتری از دقت را دارا بود. بنابراین، الگوریتم شاهین هریس در ترکیب با درخت تصمیم توانست به بالاترین دقت تشخیص بیماری عروق کرونر در حالت انتخاب ویژگی نسبت به کل ویژگی‌ها دست یابد.

نتیجه‌گیری: نتایج این مطالعه به خوبی نقش فرآیند انتخاب ویژگی‌های مؤثر در تشخیص بیماری عروق کرونر توسط الگوریتم شاهین هریس در ترکیب با روش‌های یادگیری ماشین را نشان داد.

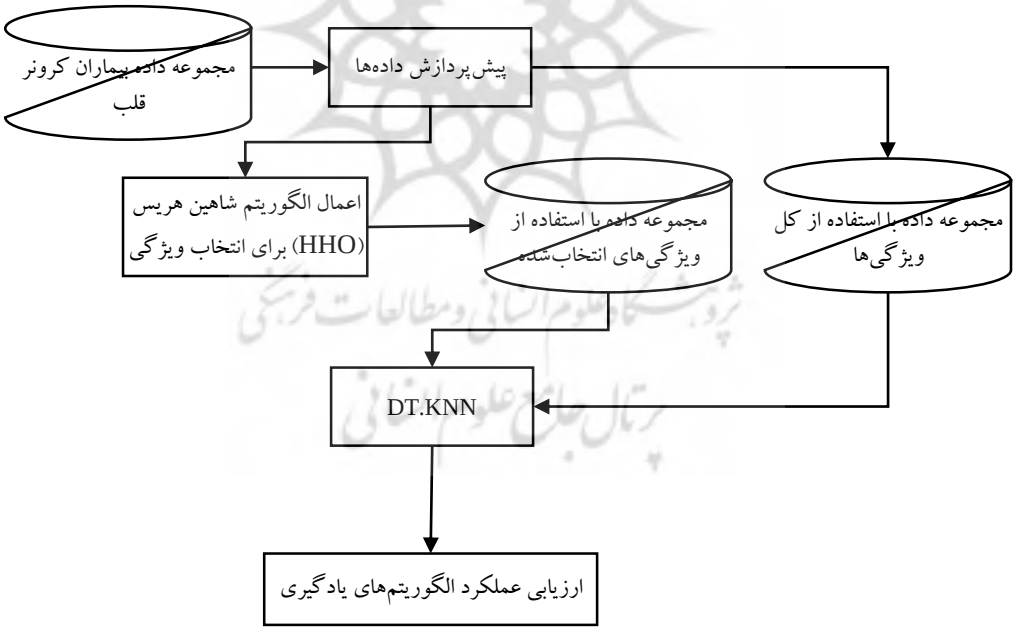
مقدمه

امروزه با گذار سریع سلامت، تغییرات جمعیتی و فرآیند افزایش سن و تغییرات سریع شیوه زندگی به همراه تحولات اقتصادی و اجتماعی روند ابتلا، ناتوانی و مرگ‌ومیر ناشی از بیماری‌های غیر واگیر افزایش‌یافته است. [۱] بیماری قلبی علت اصلی مرگ‌ومیر در جهان در ده سال گذشته بوده است. در این میان، بیماری عروق کرونر (Coronary Artery Disease (CAD))، شایع‌ترین نوع بیماری قلبی و عامل اصلی مرگ‌ومیر در کشورهای صنعتی محسوب می‌شود که به‌سرعت در کشورهای درحال‌توسعه نیز رو به گسترش است. [۲،۳] بیماری عروق کرونر به دلیل تنگ شدن یا اسپاسم عروق کرونر ایجاد می‌شود که عروق اصلی خون‌رسانی به عضله قلبی می‌باشند و درنتیجه، باعث کاهش خون‌رسانی به قلب می‌شود. [۴] بیماری‌های قلبی از اصلی‌ترین عوامل تهدیدکننده سلامت افراد به‌حساب می‌آید. درصورتی‌که بیماری‌های قلبی، ازجمله بیماری عروق کرونر قلب در مراحل اولیه بروز بیماری، تشخیص و درمان نشوند، ممکن است منجر به مرگ بیمار و از دست رفتن بهترین زمان برای درمان شود. به‌کارگیری الگوریتم‌های فراابتکاری بهینه‌سازی شده می‌تواند یک سیستم نظارت بهداشتی مؤثر برای پیش‌بینی زودهنگام بیماری قلبی ارائه کند. الگوریتم بهینه‌سازی شاهین هریس (Harris Hawks Optimization (HHO))، یکی از الگوهای بهینه‌سازی مبتنی بر جمعیت و با الهام گرفتن از طبیعت است که ایده اصلی آن، رفتار مشارکتی و سبک تعقیب شاهین هریس در طبیعت است که به نام هجوم غافلگیرانه شناخته می‌شود. در این استراتژی هوشمند، چندین شاهین به‌صورت مشارکتی و با همکاری یکدیگر تلاش می‌کنند تا یک طعمه را از جهات مختلف پرتاب می‌کنند تا آن را غافلگیر کنند. شاهین هریس می‌تواند الگوهای تعقیب و گریز متنوعی را بر اساس ماهیت پویای سناریوها و الگوهای شکار از خود نشان دهد. [۵] در سالیان اخیر تحقیقات مختلفی در خصوص تشخیص CAD و روش‌های انجام آن مطرح‌شده است. برای مثال، رانی [۶] در مطالعه خود داده‌های CAD را با استفاده از شبکه‌های عصبی تک لایه و چند لایه تحلیل کرده است که در این پژوهش مقادیر دقت مدل به ترتیب ۸۷ و ۸۳ درصد به‌دست‌آمده است. دژالود و همکاران، [۷] از الگوریتم بهینه‌سازی ملخ باینری و یادگیری ماشین کای نزدیک‌ترین همسایه استفاده کردند و CAD را در مجموعه داده ۲۷۰ بیمار با دقت ۸۹،۸ درصد تشخیص دادند. ولا و همکاران [۸] از شبکه‌های بیزی برای تشخیص بیماری‌های قلبی استفاده کردند. مدل پیشنهادی در این مطالعه موارد عادی را با دقت ۹۱ درصد شناسایی کرد. مطالعه آبدار و همکاران [۹] با استفاده از الگوریتم ژنتیک ((Genetic Algorithms (GA))، الگوریتم بهینه‌سازی ازدحام ذرات و ماشین بردار پشتیبان CAD (Support Vector Machines (SVM)) با دقت ۹۳/۰۸ درصد تشخیص داده شد. التاشی و همکاران [۱۰] با استفاده از روش‌های بهینه‌سازی گرگ خاکستری و SVM توانستند به دقت ۸۹/۸۳ درصدی در تشخیص CAD دست یابند. بنابراین، می‌توان استدلال کرد که الگوریتم‌های فراابتکاری در حوزه تمرکز برای کاهش تعداد ویژگی‌ها و انتخاب موارد مؤثرتر نبوده‌اند؛ درحالی‌که این ابتکار می‌تواند دقت تشخیص را افزایش دهد. علاوه بر این، هیچ مطالعه‌ای در مورد کاربرد ترکیبی HHO و الگوریتم‌های یادگیری ماشین برای انتخاب ویژگی انجام‌نشده است. بنابراین، هدف این مطالعه افزایش دقت تشخیص CAD از طریق انتخاب ویژگی HHO است. با توجه به این‌که عدم تشخیص به‌موقع و سریع بیماری عروق کرونر قلب منجر به عوارض بی‌شمار و حتی مرگ بیمار خواهد شد و نیز تعداد متغیرهای مورد نیاز برای تشخیص این بیماری زیاد است، انتخاب مدل مناسب برای تشخیص بیماری عروق کرونر در افراد نقش بسیار بااهمیتی در تشخیص صحیح این بیماری و روند درمان آن ایفا می‌کند. بنابراین، این پژوهش به انتخاب ویژگی‌های مؤثر در تشخیص بیماری عروق کرونر با استفاده از الگوریتم فراابتکاری HHO و روش‌های یادگیری ماشین (درخت تصمیم Decision Tree (DT)) و کای نزدیک‌ترین همسایه (K-Nearest Neighbor (KNN)) انجام‌شده است. با انتخاب مجموعه ویژگی‌های با تعداد کمتر نسبت به ویژگی‌های موجود در

داده اصلی، به‌منظور بهبود دقت تشخیص بیماری می‌پردازد.

روش‌ها

این پژوهش از نوع کاربردی بود که در سال ۱۳۹۹ تا ۱۴۰۰ در دانشگاه یزد انجام شد در این پژوهش، به‌منظور بهبود تشخیص بیماری عروق کرونر قلب، از روش مبتنی بر انتخاب ویژگی با استفاده از الگوریتم HHO و الگوریتم‌های طبقه‌بندی یادگیری ماشین (DT,KNN) استفاده شد. برای این کار زیرمجموعه‌ای از مجموع ویژگی‌های مجموعه داده CAD با استفاده از الگوریتم فراابتکاری HHO انتخاب شد. این ویژگی‌های انتخاب‌شده در مقایسه با مواردی که از همه ویژگی‌ها برای تشخیص استفاده می‌کنند، CAD را با دقت بیشتری تشخیص می‌دهند. در این پژوهش از داده‌های بیماران عروق کرونر قلب موجود در مجموعه داده یادگیری ماشین دانشگاه کالیفرنیا در اروین آمریکا (*University of California*, Irvine (*UCI Irvine*)) و مجموعه داده بیماران CAD بیمارستان Cleveland استفاده شد. این مجموعه داده شامل ۳۰۳ مورد با ۱۴ ویژگی بود که ۱۶۵ و ۱۳۸ مورد به ترتیب بیمار مبتلابه عروق کرونر و سالم تشخیص داده شدند. در مرحله اول از این مطالعه، پیش‌پردازش داده‌ها انجام شد که شامل نرمال‌سازی داده‌ها بود. در مرحله بعد، با استفاده از روش انتخاب ویژگی مبتنی بر الگوریتم HHO به‌منظور انتخاب ویژگی و کاهش ابعاد، زیرمجموعه‌ای از ویژگی‌های مؤثر در تشخیص بیماری عروق کرونر قلب انتخاب شد. مجموعه داده جدید در مقایسه با مجموعه داده اصلی دارای ویژگی‌های کمتری بود. در انتها، پارامترهای ارزیابی عملکرد (پارامترهای صحت، حساسیت، شفافیت، دقت و F-Score) الگوریتم‌های مختلف یادگیری ماشین در حالت با مرحله انتخاب ویژگی و نیز بدون آن به‌دست‌آمده و با یکدیگر مقایسه شدند. مراحل ذکرشده در شکل یک نمایش داده‌شده است.



شکل ۱: فلوچارت مراحل مدل پیشنهادی تشخیص بیماری عروق کرونر قلب

به‌منظور بررسی قدرت روش انتخاب ویژگی در این پژوهش، از دو پایگاه داده Cleveland و Z-Alizadeh-Sani با تعداد ویژگی‌های متفاوت استفاده شد که از مخزن یادگیری ماشین UCI گرفته‌شده است. [۱۱] جداول یک و دو، ویژگی‌های این دو مجموعه داده را نشان می‌دهد. مجموعه داده پایگاه Cleveland، شامل اطلاعات ۳۰۳ بیمار با ۱۳ ویژگی است که یکی از این ویژگی‌ها، وضعیت نهایی فرد مراجعه‌کننده است که می‌تواند سالم یا مبتلابه بیماری عروق کرونر باشد؛ از این مجموعه داده، تعداد ۱۶۵ مورد مبتلابه بیماری عروق کرونر قلب و ۱۳۸ مورد سالم هستند. مجموعه داده دوم، داده

Z-Alizadeh-sani که شامل اطلاعات ۳۰۳ بیمار است که هرکدام دارای ۵۴ ویژگی می‌باشد. از این مجموعه داده، تعداد ۲۱۶ مورد مبتلابه بیماری عروق کرونری و ۸۷ مورد سالم هستند. مجموعه داده بیمارستان Cleveland دارای ۱۳ ویژگی و مجموعه داده Z-Alizadeh-sani دارای ۵۴ ویژگی است که این موضوع باعث می‌شود در مرحله انتخاب ویژگی ورودی‌های متفاوت به الگوریتم HHO وجود داشته باشد.

جدول ۱: ویژگی‌های مجموعه داده Cleveland

ردیف	ویژگی	شرح	بازه ویژگی
۱	Age	سن	۲۹-۷۷
۲	Sex	جنسیت	مرد-زن
۳	Cp	نوع درد قفسه سینه	۰-۳
۴	trestbps	فشارخون در حالت استراحت برحسب میلی‌متر جیوه	۹۴-۲۰۰
۵	chol	کلسترول سرم خون	۱۲۶-۵۶۴
۶	fbs	قند خون ناشتا	۰-۱
۷	restecg	نتایج نوار قلب در حال استراحت	۰-۲
۸	thalach	حداکثر ضربان قلب	۷۱-۲۰۲
۹	exang	آنژین القاشده توسط فعالیت ورزشی	۰-۱
۱۰	oldpeak	القاشده از ST افت ورزش در مقایسه با حالت استراحت	۰-۶/۲
۱۱	slope	ST شیب قله موج	۰-۲
۱۲	ca	تعداد عروق بزرگ مشخص‌شده توسط فلوروسکوپی	۰-۴
۱۳	thal	سینتیگرافی تالیم	۰-۱

جدول ۲: ویژگی‌های مجموعه داده Z-Alizadeh-sani

ردیف	ویژگی	شرح	بازه ویژگی
۱	Age	سن	۳۰-۸۶
۲	Weight	وزن	۴۸-۱۲۰
۳	Length	قد	۱۴۰-۱۸۸
۴	Sex	جنسیت	مرد - زن
۵	BMI (body mass index Kg/m^2)	شاخص توده بدنی	۱۸-۴۱
۶	DM (Diabetes Mellitus)	دیابت	بلی - خیر
۷	HTN (hyper tension)	سابقه فشارخون بالا	بلی - خیر
۸	Current smoker	سیگاری فعلی	بلی - خیر
۹	Ex-Smoker	سیگاری سابق	بلی - خیر
۱۰	FH (family history)	سابقه خانوادگی	بلی - خیر
۱۱	Obesity	چاقی	بلی؛ اگر BMI>25 - خیر
۱۲	CRF (chronic renal failure)	نارسایی مزمن کلیه	بلی - خیر
۱۳	CVA (Cerebrovascular Accident)	صدمه مغزی و عروقی	بلی - خیر
۱۴	Airway disease	بیماری راه هوایی	بلی - خیر
۱۵	Thyroid Disease	بیماری تیروئید	بلی - خیر
۱۶	CHF (congestive heart failure)	نارسایی احتقانی قلب	بلی - خیر
۱۷	DLP (Dyslipidemia)	اختلالات چربی خون	بلی - خیر
۱۸	BP (blood pressure: mmHg)	فشارخون	۹۰-۱۹۰

۱۹	PR (pulse rate) (ppm)	ضربان نبض	۵۰-۱۱۰
۲۰	Edema	ورم	بلی - خیر
۲۱	Weak peripheral pulse	نبض محیطی ضعیف	بلی - خیر
۲۲	Lung rales	صدای غیرعادی ریه	بلی - خیر
۲۳	Systolic murmur	سوفل سیستولیک	بلی - خیر
۲۴	Diastolic murmur	سوفل دیاستولیک	بلی - خیر
۲۵	Typical Chest Pain	درد قفسه سینه	بلی - خیر
۲۶	Dyspnea	تنگی نفس	بلی - خیر
۲۷	Function class	کلاس عملکرد	۱-۲-۳-۴
۲۸	Atypical	غیرمعمول	بلی - خیر
۲۹	Nonanginal CP	CP غیر آنژینال	بلی - خیر
۳۰	Exertional CP (Exertional Chest Pain)	CP عملی (درد قفسه سینه)	بلی - خیر
۳۱	Low Th Ang (low Threshold angina)	آنژین آستانه پایین	بلی - خیر
۳۲	Q Wave	فشارخون	بلی - خیر
۳۳	ST Elevation	ضربان نبض	بلی - خیر
۳۴	ST Depression	ورم	بلی - خیر
۳۵	T inversion	نبض محیطی ضعیف	بلی - خیر
۳۶	LVH (left ventricular hypertrophy)	رال ریه	بلی - خیر
۳۷	Poor R progression (poor R wave progression)	سوفل سیستولیک	بلی - خیر
۳۸	FBS (fasting blood sugar) (mg/dl)	قند خون ناشتا	۶۲-۴۰۰
۳۹	Cr (creatine) (mg/dl)	کراتینین	۰/۵-۲/۲
۴۰	TG (triglyceride) (mg/dl)	تری گلیسرید	۳۷-۱۰۵۰
۴۱	LDL (low density lipoprotein) (mg/dl)	لیپوپروتئین با چگالی کم	۱۸-۲۳۲
۴۲	HDL (high density lipoprotein) (mg/dl)	لیپوپروتئین با چگالی بالا	۱۵-۱۱۱
۴۳	BUN (blood urea nitrogen) (mg/dl)	نیتروژن اوره خون	۶-۵۲
۴۴	ESR (erythrocyte sedimentation rate) (mm/h)	میزان رسوب گلبول‌های قرمز	۱-۹۰
۴۵	HB (hemoglobin) (g/dl)	هموگلوبین	۸/۹-۱۷/۶
۴۶	K (potassium) (mEq/lit)	پتاسیم	۳/۰-۶/۶
۴۷	Na (sodium) (mEq/lit)	سدیم	۱۲۸-۱۵۶
۴۸	WBC (white blood cell) (cells/ml)	گلبول سفید	۳۷۰۰-۱۸۰۰۰
۴۹	Lymph (Lymphocyte) (%)	لنفوسیت	۷-۶۰
۵۰	Neut (neutrophil) (%)	نوتروفیل	۳۲-۸۹
۵۱	PLT (platelet) (1000/ml)	پلاکت	۲۵-۷۴۲
۵۲	EF (ejection fraction) (%)	کسر جهشی	۱۵-۶۰
۵۳	Region with RWMA (Regional Wall Motion Abnormality)	ناهنجاری حرکت دیواره منطقه‌ای	۰-۱-۲-۳-۴
۵۴	VHD (valvular heart disease)	بیماری دریچه‌ای قلب	طبیعی، خفیف، متوسط، شدید

انتخاب ویژگی با استفاده از الگوریتم بهینه‌سازی HHO: هوش جمعی یکی از قوی‌ترین روش‌های بهینه‌سازی است که بر مبنای رفتارهای گروهی و مشارکتی است. HHO نوعی الگوریتم الهام گرفته از طبیعت مبتنی بر جمعیت است. [۱۲] لوئیز لوفور در سال ۱۹۹۷، رویکردی را برای اندازه‌گیری بهره هوشی پرندگان بر اساس نوآوری‌های مشاهده‌شده در رفتارهای تغذیه آن‌ها معرفی کرد. شاهین‌های هریس پرندگان شکارچی معروفی هستند که آن‌ها را می‌توان در گروه‌های نسبتاً پایدار در نیمه جنوبی آریزونا آمریکا یافت. به‌طوری‌که شاهین‌ها را می‌توان در زمره باهوش‌ترین پرندگان طبیعت قرار داد. از مزایای بالقوه الگوریتم شاهین هریس برای انتخاب ویژگی می‌توان به موارد زیر اشاره کرد [۱۳]:

-الگوریتم شاهین هریس تا زمانی که راه‌حل بهینه پیدا نشود از قابلیت اکتشاف قدرتمند برخوردار است؛ زیرا شاهین‌های مصنوعی مختلف می‌توانند قسمت‌های مختلف فضای مسئله را کشف کنند.

-الگوریتم شاهین هریس برای انتخاب ویژگی‌ها از آنجایی جذاب است که شاهین‌ها در این الگوریتم حافظه دارند.

-الگوریتم شاهین هریس، به جای کار کردن با یک راه‌حل، با مجموعه‌ای از راه‌حل‌های بالقوه برای حل مسئله کار می‌کند.

-الگوریتم HHO با تعداد اندکی از پارامترها و به‌صورت ساده و قابل‌فهم صورت می‌گیرد و نتایج امیدوارکننده‌ای را ارائه می‌دهد. همچنین، الگوریتم HHOتحت تأثیر ابعاد مسئله قرار نمی‌گیرد.

در این پژوهش به‌منظور انتخاب ویژگی توسط الگوریتم بهینه‌سازی HHO، مقادیر اولیه پارامترهای الگوریتم HHO با تعداد جمعیت شاهین‌ها ۵۰، تعداد تکرارها ۱۰۰، ابعاد ۵۴ و ۱۳ و $\beta = 1.5$ در نظر گرفته‌شده‌اند.

تابع هدف انتخاب ویژگی توسط الگوریتم HHO: به‌منظور انتخاب ویژگی‌های مؤثر در تشخیص بیماری عروق کرونر قلب توسط الگوریتم HHO، تابع هدف به‌صورت رابطه یک انتخاب‌شده است.

$$x = [x_1, x_2, x_3, \dots, x_d] \quad \text{(رابطه یک)}$$
$$F(x) = Accuracy(x)$$

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$

هدف مطالعه حاضر، بالا بردن صحت تشخیص بیماری عروق کرونر است. با توجه به رابطه، ورودی‌های تابع هدف ویژگی هایی هستند که الگوریتم HHO آن‌ها را انتخاب می‌کنند و خروجی تابع هدف صحت است. مقادیر X ویژگی‌هایی هستند که با الگوریتم HHO انتخاب می‌شوند. در هر بار تکرار، الگوریتم HHO ویژگی‌های مختلفی را در تابع هدف قرار می‌دهد و میزان صحت به دست می‌آید. سرانجام پس از خاتمه تکرارها، ویژگی‌هایی که با الگوریتم HHO انتخاب‌شده‌اند منجر به دست‌یابی به ماکزیمم مقدار صحت شده‌اند. در این مطالعه به‌منظور تشخیص بیماری عروق کرونر از روش‌های یادگیری ماشین درخت تصمیم و کای نزدیک‌ترین همسایه استفاده شد که جزئیات هرکدام در ادامه ارائه‌شده است. درخت تصمیم (Decision Tree(DT)) از نوع درخت تصمیم طبقه‌بندی و رگرسیونی (Classification and Regression Trees (CART)) با معیار انتخاب ویژگی برای شکست با شاخص Gini، حداقل تعداد نمونه مورد نیاز برای تقسیم یک گره داخلی به میزان دو، حداقل تعداد نمونه مورد نیاز برای قرار گرفتن در یک گره برگ به میزان یک، داده‌های آموزش به میزان ۸۰ درصد و داده‌های آزمون به میزان ۲۰ درصد به‌منظور تشخیص بیماری عروق کرونر قلب استفاده‌شده است. روش کای نزدیک‌ترین همسایه، نوعی روش یادگیری موردی است و ازجمله ساده‌ترین الگوریتم‌های یادگیری ماشین است. در این الگوریتم، یک نمونه با رأی اکثریت از همسایه‌هایش دسته‌بندی می‌شود [۱۴] در این پژوهش، کای نزدیک‌ترین همسایه با پارامترهای پنج همسایه، نوع تابع فاصله مینکوفسکی (Minkowski)، پارامتر توان q به میزان دو، داده‌های آموزش به میزان ۸۰ درصد و داده‌های آزمون به میزان ۲۰ درصد به‌منظور تشخیص بیماری عروق کرونر قلب استفاده‌شده است.

مدل‌سازی، نتایج و ارزیابی: برای مدل‌سازی، مدل پیشنهادی از نرم‌افزار پایتون نسخه 2016 استفاده‌شده است. بر اساس روش‌های داده‌کاوی، اولین مرحله در ایجاد هر مدلی مرحله پیش‌پردازش است، که جهت بهبود کیفیت داده‌های واقعی برای داده‌کاوی لازم است. ابتدا هر دو دسته داده Cleveland و Z-Alizadeh-sani با استفاده از رابطه دو نرمال شده‌اند؛ به‌طوری‌که داده‌ها حداقل صفر و حداکثر یک هستند. نرمال‌سازی داده‌ها منجر به کسب نتایج بهتری می‌شود.

$$XN = \frac{X - X_{min}}{X_{max} - X_{min}} \quad \text{(رابطه دو)}$$

$$XN = \frac{X - X_{min}}{X_{max} - X_{min}}$$

در این رابطه، XN, X_{max}, X_{min}، به ترتیب، مقدار مینیمم و ماکزیمم داده‌های ورودی و داده نرمال شده است. نتایج متعادل‌سازی برای داده‌های Cleveland و Z-Alizadeh-san به ترتیب برابر با ۵۵ به ۴۵ و ۷۰ به ۳۰ می‌باشد که حاکی از نرمال بودن داده‌های این مطالعه است. سپس، نمونه‌های این پژوهش، در دو گروه آموزش و آزمون قرار گرفت. ۸۰ درصد داده‌ها به‌عنوان داده‌های آموزش و ۲۰ درصد آن، به عنوان داده آزمون در نظر گرفته‌شد. در این پژوهش برای آموزش مدل‌های یادگیری ماشین از روش اعتبار سنجی متقابل با K=5 استفاده شد. برای ارزیابی عملکرد مدل پیشنهادی معیارهای دقت، صحت، حساسیت، شفافیت و معیار F بررسی شد که این معیارها مبتنی بر معیارهای پایه زیر است [۱۵]:

- مثبت صحیح (TP): نمونه عضو دسته مثبت باشد و عضو همین کلاس تشخیص داده شود.
- منفی اشتباه (FN): نمونه عضو کلاس مثبت باشد و عضو کلاس منفی تشخیص داده شود.
- منفی صحیح (TN): نمونه عضو کلاس منفی باشد و عضو همین کلاس تشخیص داده می‌شود.
- مثبت اشتباه (FP): نمونه عضو کلاس منفی باشد و عضو کلاس مثبت تشخیص داده شود.

سنجه‌های ارزیابی مورداستفاده برای تشخیص بیماری عروق کرونر قلبی مطابق روابط سه تا هفت محاسبه شد که عبارت‌اند از: دقت: نشان‌دهنده تعداد نمونه‌های درست به تعداد کل نمونه‌ها است.

$$Precision = \frac{TP}{TP + FP} \quad \text{(رابطه سه)}$$

حساسیت: حساسیت به معنی نسبتی از موارد مثبت است که آزمایش آن‌ها را به درستی به‌عنوان نمونه مثبت تشخیص داده است.

$$Sensitivity\ (TPR) = \frac{TP}{TP + FN} \quad \text{(رابطه چهار)}$$

شفافیت: در نقطه مقابل پارامتر حساسیت، ممکن است در مواقعی دقت پیش‌بینی کلاس منفی حائز اهمیت باشد. از متداول‌ترین پارامترها که معمولاً در کنار حساسیت بررسی می‌شود، پارامتر شفافیت است که به آن نرخ پاسخ‌های منفی درست (True Negative Rate (TNR)) نیز می‌گویند.

$$Specificity(TNR) = \frac{TN}{TN + FP} \quad \text{(رابطه پنج)}$$

صحت: به نسبت تعداد داده‌هایی که به درستی در کلاس مربوطه طبقه‌بندی‌شده‌اند، به تعداد کل داده‌ها در مجموعه داده گفته می‌شود.

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN} \quad \text{(رابطه شش)}$$

معیار اف: برای ارزیابی عملکرد دسته‌بندی‌ها استفاده می‌شود، این پارامتر از ترکیب دو پارامتر حساسیت (Recall) و پارامتر دقت (Precision)، که در واقع میانگین هارمونیک این دو پارامتر است.

$$F - Score = 2 \times \frac{(Recall \times Precision)}{(Recall + Precision)} \quad \text{(رابطه هفت)}$$

یافته‌ها

تشخیص بیماری عروق کرونر قلب با استفاده از الگوریتم HHO و DT : تشخیص بیماری عروق کرونری توسط روش درخت تصمیم، در دو حالت با و بدون انتخاب ویژگی توسط الگوریتم شاهین هریس انجام شد که نتایج و معیارهای ارزیابی در این دو حالت بررسی شد. نتایج مدل پیشنهادی برمبنای انتخاب ویژگی در جدول سه نشان داده‌شده است. زمانی که انتخاب ویژگی توسط الگوریتم شاهین در دو مجموعه صورت گرفته است صحت، شفافیت و

سایر معیارها نسبت به حالت بدون ویژگی افزایش‌یافته است.

جدول ۳: نتایج تشخیص بیماری عروق کرونر قلب توسط روش DT و الگوریتم HHO و کل ویژگی‌ها

پارامتر	(مجموعه داده Z-Alizadeh-Sani)**		(مجموعه داده Cleveland)*	
	مقدار بهینه به‌دست‌آمده توسط الگوریتم HHO	تشخیص توسط تمام ویژگی‌ها	مقدار بهینه به‌دست آمده توسط الگوریتم HHO	تشخیص توسط تمام ویژگی‌ها
صحت (Accurcy)	۰/۹۷۳	۰/۹۵۳	۰/۹۲۷	۰/۹۰۴
حساسیت (Sensitivity)	۰/۹۸۱	۰/۹۸۱	۱	۱
شفافیت (Specificity)	۰/۹۵۴	۰/۸۸۵	۰/۸۴	۰/۷۸۹
دقت (Precision) کلاس بیمار	۰/۹۸	۰/۹۵	۰/۸۸	۰/۸۵
دقت (Precision) کلاس سالم	۰/۹۵	۰/۹۵	۱	۱
پارامتر (Recall) کلاس بیمار	۰/۹۸	۰/۹۸	۱	۱
پارامتر (Recall) کلاس سالم	۰/۹۵	۰/۸۹	۰/۸۴	۰/۷۹
پارامتر (F-Score) کلاس بیمار	۰/۹۸	۰/۹۷	۰/۹۴	۰/۹۲
پارامتر (F-Score) کلاس سالم	۰/۹۵	۰/۹۲	۰/۹۱	۰/۸۸
مساحت زیر منحنی ROC	۰/۹۶۷	۰/۸۵۶	۰/۹۲۳	۰/۹۰۲
تعداد ویژگی‌های انتخاب‌شده توسط الگوریتم HHO	۱۷	۵۴	۶	۱۳
ویژگی‌های انتخاب‌شده توسط الگوریتم HHO	۱ ۲ ۳ ۴ ۵ ۶ ۷ ۸ ۱۲ ۱۳ ۱۴ ۱۵ ۱۶ ۱۹ ۲۱ ۲۲ ۲۴	–	۲ ۶ ۷ ۹ ۱۱ ۱۲	–

* جدول یک

**جدول دو

تشخیص بیماری عروق کرونر قلب با استفاده از الگوریتم HHO و KNN: تشخیص بیماری عروق کرونر قلب توسط روش KNN، در دو حالت با و بدون انتخاب ویژگی توسط الگوریتم HHO انجام شد که نتایج مدل پیشنهادی بر برمبنای انتخاب ویژگی در جدول چهار نشان داده‌شده است. زمانی که انتخاب ویژگی توسط الگوریتم شاهین در دو مجموعه صورت گرفته است صحت، شفافیت و سایر معیارها نسبت به حالت بدون ویژگی افزایش‌یافته است.

جدول ۴: نتایج تشخیص بیماری عروق کرونر قلب توسط روش KNN و الگوریتم HHO و کل ویژگی‌ها

پارامتر	(مجموعه داده Cleveland)*		(مجموعه داده Z-Alizadeh-Sani)**	
	مقدار بهینه به‌دست‌آمده توسط الگوریتم HHO	تشخیص توسط تمام ویژگی‌ها	مقدار بهینه به‌دست‌آمده توسط الگوریتم HHO	تشخیص توسط تمام ویژگی‌ها
صحت (Accurcy)	۰/۸۳۱	۰/۷۰۶	۰/۷۸۲	۰/۷۲۹
حساسیت (Sensitivity)	۰/۹۷۵	۰/۹۰۹	۰/۹۵۸	۰/۹۲۵
شفافیت (Specificity)	۰/۶۵۹	۰/۴۶۳	۰/۳۴۴	۰/۲۴۱
دقت (Precision) کلاس بیمار	۰/۷۷	۰/۶۷	۰/۷۸	۰/۷۵
دقت (Precision) کلاس سالم	۰/۹۶	۰/۸۱	۰/۷۷	۰/۵۷

پارامتر (Recall) کلاس بیمار	۰/۹۸	۰/۹۱	۰/۹۶	۰/۹۳
پارامتر (Recall) کلاس سالم	۰/۶۶	۰/۴۶	۰/۳۴	۰/۲۴
پارامتر (F-Score) کلاس بیمار	۰/۸۶	۰/۷۷	۰/۸۶	۰/۸۳
پارامتر (F-Score) کلاس سالم	۰/۷۸	۰/۵۹	۰/۴۸	۰/۳۴
مساحت زیر منحنی ROC	۰/۹۱۸	۰/۷۸۵	۰/۸۳۸	۰/۶۹۵
تعداد ویژگی‌های انتخاب‌شده توسط الگوریتم HHO	۶	۱۳	۹	۵۴
ویژگی‌های انتخاب شده توسط الگوریتم HHO	۲ ۶ ۸ ۱۰ ۱۱ ۱۲	-	۱ ۵ ۶ ۷ ۱۰ ۲۰ ۲۱ ۲۴ ۲۹	-

*جدول یک

**جدول دو

این پژوهش، با هدف تشخیص بیماری عروق کرونر با به‌کارگیری روش یادگیری ماشین نظارت‌شده DT و KNN انجام شد که این فرآیند توسط الگوریتم بهینه‌سازی HHO بر روی دو مجموعه داده Alizadeh-Sani و Cleveland انجام شد. همان‌طور که مشاهده شد دقت مدل در روش DT برای هر دو مجموعه داده نسبت به الگوریتم KNN بیشتر بوده است.

بحث

براساس یافته‌های این مطالعه انتخاب ویژگی‌ها با استفاده از الگوریتم HHO در حالت ترکیبی با روش‌های یادگیری ماشین منجر به افزایش هریک از پارامترهای صحت، حساسیت، دقت، شفافیت و F-Score شده است. نتایج دقت در ترکیب الگوریتم HHO با درخت تصمیم در دو مجموعه داده Cleveland و Z-Alizadeh-Sani به ترتیب برابر با ۸۸ و ۹۸ درصد بود. این در حالی است که در حالت تمام ویژگی‌ها، هریک از این دو مجموعه داده به ترتیب دقت ۸۵ و ۹۵ درصد به دست آورند که این مقدار کمتر از نتایج مدل پیشنهادی انتخاب ویژگی با الگوریتم شاهین هریس بوده است. همچنین، نتایج دقت در حالت ترکیبی با الگوریتم KNN در دو مجموعه داده Cleveland و Z-Alizadeh-Sani به ترتیب برابر با ۷۷ و ۷۸ درصد بود؛ اما در مقایسه با تمام ویژگی‌ها هریک از این دو مجموعه داده با دقت ۶۷ و ۷۵ درصد همراه بوده است که این مقادیر کمتر از نتایج مدل پیشنهادی انتخاب ویژگی با الگوریتم شاهین هریس می‌باشد. نتایج این مطالعه با ترکیب HHO با هریک از روش‌های یادگیری ماشین در دو مجموعه داده بررسی شد. درخت تصمیم در مجموعه داده Z-Alizadeh-Sani در پژوهش حاضر توانست به بالاترین میزان دقت (۹۸ درصد) دست یابد که در مقایسه با مطالعه آبدار و همکاران [۹] دقت تشخیص CAD روی همین مجموعه داده به ۹۳/۰۸ درصد دست یافت. بنابراین، مدل پیشنهادی در این پژوهش دارای دقت بالاتری می‌باشد. همچنین، نتایج مطالعات مشابه دیگری توسط ساپرا و همکاران [۱۶] و ویجایاشری و همکاران [۱۷] با روش SVM انجام شد که به ترتیب با دقت ۸۴ و ۸۴/۳۶ درصد همراه بود و در مقایسه با دقت مطالعه حاضر بسیار کمتر بود. علاوه براین، نتایج درخت تصمیم در مجموعه داده Cleveland با دقت ۸۸ درصد همراه بود که حاکی از بهینه‌سازی هرچه بیشتر پیش‌بینی CAD با استفاده از HHO در ترکیب با الگوریتم درخت تصمیم است. این نتیجه در مقایسه با دقت تشخیص CAD روی همین مجموعه داده در مطالعه التاشی و همکاران [۱۰] کمتر بود که توانستند با استفاده از الگوریتم‌های بهینه سای گرگ خاکستری و ماشین بردار پشتیبان به دقت ۸۹/۸۳ درصد دست پیدا کنند؛ اما مطالعات دیگر توسط شومن و همکاران [۱۸] و انگوین و همکاران [۱۹] روی همین مجموعه داده به ترتیب با دقت ۸۴/۱ و ۸۰ درصد با نتایج دقت کمتری نسبت به مطالعه حاضر همراه بود. همان‌طور که گزارش شد، نتایج HHO در ترکیب با کای نزدیک‌ترین همسایه نیز امیدوارکننده بود؛ اما نسبت به درخت تصمیم به مقادیر کمتری دست یافت. نتایج روش

پیشنهادی در ترکیب با هریک از دو الگوریتم یادگیری ماشین توانسته است به دقت بالاتری دست یابد. ملکی و همکاران [۲۰] نیز در مطالعه خود با استفاده از HHO در ترکیب با روش SVM به دقت ۹۰ درصدی پیش‌بینی CAD دست یافتند که این نشان‌دهنده بهبود پیش‌بینی بیماری قلبی با استفاده از الگوریتم HHO در ترکیب با روش‌های یادگیری ماشین است. از نتایج این پژوهش می‌توان نتیجه گرفت که ترکیب الگوریتم HHO در الگوریتم DT با انتخاب ویژگی نسبت به کل ویژگی‌ها، باعث افزایش دقت تشخیص بیماری عروق کرونر قلبی می‌شود. بنابراین، می‌توان از مدل پیشنهادی این پژوهش به‌منظور پیش‌بینی مؤثرتر CAD بهره جست. در ضمن نتایج این مطالعه بر روی مجموعه داده استاندارد UCI انجام شد؛ اما با بررسی نتایج این مطالعه و کشف ویژگی‌های مهم و کمتر می‌توان مجموعه داده واقعی را از مراکز درمانی تهیه نمود و افراد مشکوک به بیماری عروق کرونر قلبی را برمبنای استدلال انتخاب ویژگی، تشخیص داد.

ملاحظات اخلاقی

رعایت دستورالعمل‌های اخلاقی: در مطالعه حاضر به تمامی ملاحظات اخلاقی ازجمله شرط امانت و صداقت توجه شده است. این مقاله، حاصل بخشی از پایان‌نامه با عنوان تشخیص بیماری عروق کرونر قلب با انتخاب ویژگی مؤثر توسط الگوریتم شاهین هریس و روش‌های یادگیری ماشین، در مقطع کارشناسی ارشد، مصوب دانشگاه یزد در سال ۱۳۹۹ است.

حمایت مالی: پژوهش حاضر، تحت حمایت مالی هیچ سازمانی نبوده است.

تضاد منابع: نویسندگان اعلام می‌دارند که در این پژوهش، هیچ‌گونه تضاد منافعی وجود ندارد.

تشکر و قدردانی: از جناب آقای دکتر مسعود میرزایی، استاد مرکز تحقیقات مدل‌سازی دانشگاه علوم پزشکی شهید صدوقی یزد، سپاسگزاری و قدردانی می‌شود که در انجام این پژوهش ما را یاری نمودند.

References

1. Reddy KS. Cardiovascular diseases in the developing countries: Dimensions, determinants, dynamics and directions for public health action. Public Health Nutr. 2002;5(1a):231-7.
2. Squeri A. Coronary artery disease - new insights and novel approaches [Internet]. London: IntechOpen; 2012 [cited 2022 Apr 30]. Available from: https://www.intechopen.com/books/660
3. Nahar J, Imam T, Tickle KS, Phoebe Chen YP. Association rule mining to detect factors which contribute to heart disease in males and females. Expert Syst Appl. 2013;40(4):1086-93.
4. Giri D, Acharya UR, Martis RJ, Sree SV, Lim TC, VI TA, et al. Automated diagnosis of coronary artery disease affected patients using LDA, PCA, ICA and discrete wavelet transform. Knowl Base Syst. 2013;37:274-82.
5. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H. Harris hawks optimization: Algorithm and applications. Future Generat Comput Syst. 2019;97:849-72.
6. Rani KU. Analysis of heart diseases dataset using neural network approach. International Journal of Data Mining & Knowledge Management Process. 2011;1(5):1-8.
7. DezhAloud N, Soleimanian Gharehchopogh F. Diagnosis of heart disease using binary Grasshopper optimization algorithm and K-Nearest neighbors. Journal of Health Administration. 2020;23(3):42-54. [In Persian]
8. Vila-Frances J, Sanchis J, Soria-Olivas E, Serrano AJ, Martinez-Sober M, Bonanad C, et al. Expert system for predicting unstable angina based on Bayesian networks. Expert Syst Appl. 2013;40(12):5004-10.
9. Abdar M, Ksiazek W, Acharya UR, Tan RS, Makarenkov V, Pławiak P. A new machine learning technique for an accurate diagnosis of coronary artery disease. Comput Methods Programs Biomed. 2019;179:1-11.

10. Al-Tashi Q, Rais H, Jadid S. Feature selection method based on grey wolf optimization for coronary artery disease classification. In: Saeed F, Gazem N, Mohammed F, Busalim A, editors. Recent trends in data science and soft computing. IRICT 2018. Advances in intelligent systems and computing. Kuala Lumpu: Springer; 2019 Sep. p. 257-66.

11. Data Sets - UCI Machine Learning Repository [Internet]. Amherst:The University of Massachusetts Amherst; c2007- [cited 2021 Feb]. Available from: https: //archive.ics.uci.edu/ml/datasets.php

12. Heidari AA, Mirjalili S, Faris H, Aljarah I, Mafarja M, Chen H. Harris hawks optimization: Algorithm and applications. Future Gener Comput Syst. 2019;97:849-72.

13. Balamurugan R, Ratheesh S, Venila YM. Classification of heart disease using adaptive Harris hawk optimization-based clustering algorithm and enhanced deep genetic algorithm. Soft comput. 2022;26:2357-73.

14. Coomans D, Massart DL. Alternative k-nearest neighbour rules in supervised pattern recognition: Part 1. k-Nearest neighbour classification by using alternative voting rules. Anal Chim Acta. 1982;136:15-27.

15. Ray S. Understanding support vector machine (SVM) algorithm from examples (along with code) [Internet]. Analytics Vidhya, 2015 Oct [updated 2017 Sep]. Available from: https://www.analyticsvidhya.com/blog/2017/09/understaing-support-vector-machine-example-code/

16. Sapra L, Sandhu JK, Goyal N. Intelligent method for detection of coronary artery disease with ensemble approach. In: Hura G, Singh A, Siong Hoe L, editors. Advances in communication and computational technology. Singapore: Springer; 2021. p. 1033-42.

17. Vijayashree J, Sultana HP. A machine learning framework for feature selection in heart disease classification using improved particle swarm optimization with support vector machine classifier. Program Comput Software. 2018;44(6):388-97.

18. Shouman M, Turner T, Stocker R. Using decision tree for diagnosing heart disease patients. Proceedings of the Ninth Australasian Data Mining Conference. 2011 Dec 1-2; Ballarat Australia. 2011. p. 23-30.

19. Nguyen T, Khosravi A, Creighton D, Nahavandi S. Medical data classification using interval type-2 fuzzy logic system and wavelets. Appl Soft Comput. 2015;30:812-22.

20. Maleki S, Zare Mehrjerdi Y, shishebori D, Mirzaei M. Predicting coronary artery diseases using effective features selected by Harris Hawks optimization algorithm and support vector machine. Journal of Industrial and Systems Engineering. 2022 Jan;14:40-47.

پژوهشگاه علوم انسانی و مطالعات فرهنگی
پرتال جامع علوم انسانی